UNIVERSIDAD AUTONOMA
DE MADRID

Contribución del Proteasoma y del Polimorfismo de ERAP1 en la Configuración y Patogenicidad de los Peptidomas Constitutivos de MHC-I: Estudios en HLA-A68 y HLA-B27

Memoria para optar al grado de Doctor en Biología Molecular

Presentada por: Noe García Medel

Director: José Antonio López de Castro Álvarez, Profesor de Investigación del C.S.I.C.

Centro de Biología Molecular Severo Ochoa.

Tutor: Balbino Alarcón Sánchez, Profesor de Investigación del C.S.I.C.

Centro de Biología Molecular Severo Ochoa.

A mis Padres, a mi Hermano y a Silvia.

A José Antonio no solo le tengo que agradecer que me aceptase como su mejor esbirro y todo lo que he aprendido de él, sino también la libertad que me ha dado en la forma de trabajar, que haya hecho siempre tan divertidas las horas que pasamos en el laboratorio (aunque a veces cuando estábamos enviando las publicaciones…) y especialmente el esfuerzo que ha supuesto acompañarme en el *sprint* final para que llegásemos a tiempo para registrar en junio.

…y después del jefe, voy por orden cronológico…

Por supuesto a Dani, que no nos merecemos ni una etiqueta de anís del mono y ahora nos dan la Tesis a los dos a la vez, como está el país!!!.

A mis colegas Biofísicos, Mohammad y Pedro, de los que he aprendido un montón de cosas que nunca habría aprendido si me hubiese quedado solamente en el campo de la bioquímica.

A mis primeros compañeros en el grupo, Bego, Carla, Elena, Juanjo, Miguel y Patricia por la diversión, risa, *coconut*, complicidad, lecciones maestras y moderación que me han proporcionado, respectivamente, además por toda su ayuda cuando me ha hecho falta. Los de Boston a ver si volvéis ya eh!. A Vero por los tres días que fuimos compañeros. A Luis, que siempre ha tenido respuesta para mis problemas.

A mis predecesores en este grupo que he tenido oportunidad de conocer, Alberto, Daniel (por supuesto que el agradecimiento se extiende a Elena), Iñaki, Manolo y Miriam.

A mis actuales compañeros, Adrián, por la energía que nos proporciona, Alex, por el trabajo que hemos hecho juntos y por controlar el exceso de energía de Adrían, Carlos, que siempre está en la cresta de la ola y, por imposición del orden alfabético, en último lugar Lorena por su naturalidad y lo divertida que es (Luis y Patri, no tengáis morro que ya habéis salido), la de dentífrico que hemos consumido juntos!!!.

A nuestros vecinos del laboratorio de Balbino, especialmente Aldo, Elena, Enrique, Irene y María (el agradecimiento lleva implícito un vale por un pan de la cafetería, por si a alguno de los citados le interesa), que han sido una agradable compañía y han amenizado muchísimo el trabajo.

A la unidad de proteómica del CNB no se si agradecérselo o hacerles una estatua…gracias Juan Pablo, Sergio y Silvia, solo han sido 10^4 análisis con MALDI-TOF (del TOF/TOF ni hablamos), poca cosa…

La aportación de Juan y Juani podría resumirla en una piscina llena de RMPI, y de la mejor cosecha.

…y ahora me salgo de los laboratorios para agradecerle a mis padres y a mis abuelos, por su incondicional apoyo, y por lo mucho que he aprendido de ellos…*Load “” intro* la primera línea de código que tecleé con mi padre cuando tenía 5 años. A mi hermano, que me enseñó la importancia de usar una buena herramienta.

A Silvia, en la ciencia y fuera de ella, por estar siempre conmigo, por todo lo que me ha ayudado y todo lo que he aprendido de ella.

Y a mis amigos de toda la vida…Luis, Lheman, Tere, Raquel, Patri, Vane, Jose, Rasta, Moe, Marce, David, por el lado norte de Madrid, y por el lado este…Raul, Mariluz, Gema y Vero. También hay algo para el Ciudad de los Poetas: Apa (incluye a Olga y Adrián), Alejandro y Jorge (ojalá pudieras haberlo leído).

…y si me he dejado a alguien es que son las 16:00, y tengo que estar a las 17:00 en la imprenta.

Summary

The proteasome and ERAP1 are two important hydrolytic systems involved in the processing of MHC class I ligands. The modulation of their enzymatic activities has large effects on the surface expression of MCH-I and its bound peptidome. We separately addressed the influence of each system using two different approaches based on the analysis of the features of those HLA-I ligands and their precursors produced in cells treated with proteasome inhibitors or expressing different ERAP1 variants. The influence of ERAP1 polymorphism on the HLA-B27-bound peptidomes has special relevance to explain the molecular mechanism underlying the strong association of HLA-B27 with ankylosing splondylitis.

In this thesis, the susceptibility to proteasome inhibitors of the HLA-A*68:01 peptidome, a particularly inhibitor-resistant alotype, was assessed. As previously observed in HLA-B27, proteasome-inhibitor resistant HLA-A68-bound peptides arose from small basic proteins, except when they corresponded to sequences located at the ends of their precursor proteins. Incomplete inhibition of the proteasome, particularly of its tryptic-like specificity, was observed, but this did not lead to increased epitope production due to preferential inhibition of destructive cleavages. Our results indicate that inhibitor-resistant HLA-I ligands are not necessarily produced by non-proteasomal pathways. However, their generation is not simply explained by decreased epitope destruction upon incomplete proteasomal inhibition and may require additional proteolytic steps acting on incompletely processed proteasomal products.

Comparison of HLA-B*27:04-bound peptidomes from cells expressing different natural variants of ERAP1 revealed significant differences in the size, length and amount of many ligands, and in HLA-B27 stability. Peptide analyses suggested that the mechanism of ERAP1/HLA-B27 interaction is a variant-dependent alteration in the balance between epitope generation and destruction determined by the susceptibility of N-terminal flanking and P1 residues to trimming. ERAP1 polymorphism associated with susceptibility to ankylosing spondylitis ensured efficient peptide trimming and high HLA-B27 stability. Protective polymorphism resulted in diminished ERAP1 activity, less efficient trimming, suboptimal HLA-B27 peptidomes and decreased molecular stability. These results demonstrate that natural ERAP1 polymorphism affects HLA-B27 antigen presentation and stability *in vivo* and suggest a mechanism for the interaction between these molecules in spondyloarthritis.

Abreviaturas

Boc-LRR-amc: Terc-butoxicarbonil-L-leucil-L-arginin-L-arginin aminometil cumarina.

CTL: *Cytotoxic T Lymphocyte*.

DMP: *Deviation from Mean in the Proteome.*

DRIP: *Defective Ribosomal Products*.

DUBs: *De-Ubiquitilation enzimes.*

EA: Espondilitis Anquilosante.

ERAAP: *Endoplasmic Reticulum Associated Aminopeptidase.*

ERAP: *Endoplasmic Reticulum Aminopeptidase.*

HC: *Heavy Chain*.

HLA: *Human Leucocyte Antigen*.

IDE: *Insulin Degrading Enzyme*.

Ig: Inmunoglobulina.

LCL: *Lymphoblastoid Cell Line*.

MALDI-TOF: *Matrix-Assisted Laser Desorption and Ionization Time Of Flight.*

MHC: *Mayor Histocompatibility Complex*.

MS: *Mass Spectrometry.*

MW: *Molecular Weight*.

pI: Punto Isoeléctrico.

pMHC: *peptide-Mayor Histocompatibility Complex*.

PSA: *Puromycin-Sensitive Aminopeptidase*.

RI: Relación de intensidades.

SNP: *Single Nucleótide Polymorphism.*

Suc-LLVY-amc: Succinil-L-leucil-L-leucil-L-valin-L-tirosin aminometil cumarina.

TAP: *Transporter associated with antigen processing.*

TCR: *T Cell Receptor.*

TFA: Ácido trifluoroacético.

TPP II: *Tripeptidyl Peptidase II.*

PSA: *Puromycin-sensitive Aminopeptidase.*

β2M: β2-microglobulina.

Introducción

1.1. El Sistema Principal de Histocompatibilidad Humano.

1.1.1. Presentación antigénica en la respuesta inmune.

La presentación antigénica consiste en la expresión de péptidos procedentes de la degradación de proteínas en la superficie celular. Estos péptidos quedan expuestos a los linfocitos T para su reconocimiento por el receptor específico de antígeno (TCR). Los péptidos se presentan unidos a proteínas codificadas por los genes del Complejo Principal de Histocompatibilidad (MHC), que en el caso del organismo humano se denomina HLA (*Human Leukocyte Antigen*), localizado en el brazo corto del cromosoma 6.

Los péptidos presentados por las moléculas HLA proceden de proteínas de origen endógeno o exógeno. Las proteínas endógenas se degradan fundamentalmente en el núcleo y el citosol por la acción del proteasoma (Pamer and Cresswell, 1998; Rock and Goldberg, 1999). Las proteínas exógenas, que entran por fagocitosis, se degradan en endosomas y lisosomas fundamentalmente por la acción de las catepsinas (Chapman, 2006).

1.1.2. Estructura y función de las moléculas de histocompatibilidad.

Existen dos clases de moléculas HLA presentadoras de péptidos, que están codificadas en dos regiones distintas del MHC y que se diferencian en su estructura, función y tipo de péptidos que presentan. En la región de clase I (HLA-I), existen tres loci clásicos, A, B y C, y en la región de clase II otros tres, DR, DP y DQ, de manera que cada individuo puede presentar entre tres y seis alotipos diferentes de HLA de clase I y otros tantos de clase II. Los antígenos HLA unen péptidos con patrones de secuencia específicos de cada variante alélica. El reconocimiento de péptidos por los linfocitos T está restringido por HLA: un linfocito T reconoce específicamente el conjunto de un péptido unido a un alotipo determinado de HLA.

Las moléculas de clase I se expresan constitutivamente en todas las células nucleadas del organismo y unen principalmente péptidos de 8 a 12 residuos procedentes de proteínas endógenas, fundamentalmente productos defectuosos de síntesis (DRIPs) o proteínas que han llegado al final de su vida útil. Las moléculas de clase I pueden también unir péptidos procedentes de proteínas virales o de bacterias intracelulares, que se degradarían por mecanismos similares a las proteínas celulares. Están formadas por dos cadenas de distinto tamaño (Figura 1A), una cadena α o cadena pesada (HC), de 45 KDa, donde se une el péptido, se produce la interacción con los linfocitos T y se ancla la molécula a la membrana, y una cadena ligera β2-microglobulina (β2M), de 12 KDa, que está codificada en el cromosoma 15. La cadena α se distribuye en tres dominios, α1, α2 y α3 (Orr et al., 1979). Los dominios

α1 y α2 presentan un plegamiento muy similar entre sí, distinto al de las inmunoglobulinas, y juntos conforman una región altamente polimórfica que constituye el sitio de unión de péptidos (Bjorkman et al., 1987). El dominio α3 se une al receptor CD8 de los linfocitos T (Salter et al., 1990) y, al igual que β2M, presenta un plegamiento tipo inmunoglobulina. La función de las moléculas HLA-I es la presentación de antígenos a los linfocitos T citotóxicos CD8+ para la destrucción de la célula presentadora en caso de que sufra un proceso infeccioso o tumoral.

Las moléculas de clase II se expresan constitutivamente en las células presentadoras de antígeno profesionales, linfocitos, macrófagos y células dendríticas, y de forma inducible en otros tipos celulares. Su expresión es inducible por citoquinas en respuesta a un proceso infeccioso y unen péptidos de mayor longitud, procedentes de la degradación lisosomal de proteínas exógenas. Están formadas por dos cadenas de tamaño y estructura similares (Figura 1B), la cadena α, de 33 KDa y la β de 28 KDa. Las dos cadenas presentan polimorfismo y dominios de inserción a la membrana y unión al péptido. El plegamiento determina dos dominios, α1 y α2 en la cadena α y β1 y β2 en la cadena β. Los dominios α1 y β1 contienen el sitio de unión del péptido (McFarland and Beeson, 2002) y β2 el sitio de unión al receptor CD4 del linfocito T (Nag et al., 1993). Estas moléculas presentan antígenos procedentes de patógenos extracelulares a los linfocitos T colaboradores CD4+, que activan la producción de anticuerpos por los linfocitos B y la capacidad microbicida de los macrófagos.

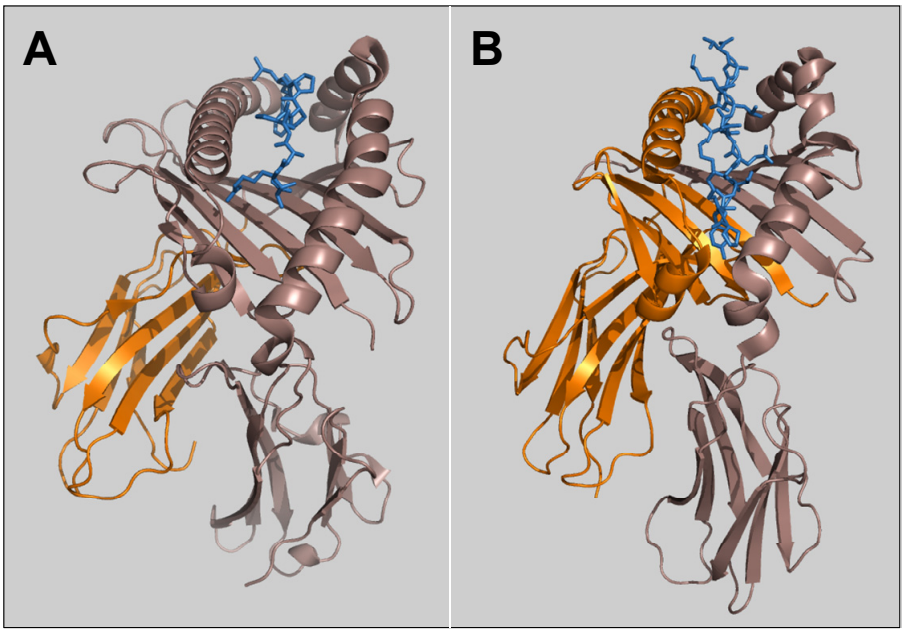


Figura 1: Modelos moleculares de HLA-I y HLA-II. (A) HLA-B27 (*Protein Data Bank* ID: 3BP4), se presenta en marrón la cadena α, en naranja la β2M y en azul el péptido. (B) HLA-DR1 (*Protein Data Bank* ID: 1DLH), se presenta en marrón la cadena α, en naranja la cadena β y en azul el péptido.

1.2. La ruta de procesamiento antigénico de clase I.

Este proceso consta de cuatro etapas: 1) degradación de proteínas por el proteasoma, 2) procesamiento post-proteasómico por aminopeptidasas citosólicas, 3) recorte de precursores peptídicos en el retículo endoplásmico, 4) carga peptídica. En la Figura 2 se muestra un esquema de esta ruta.

Con la degradación de proteínas en el núcleo y el citosol por el proteasoma (Rock et al., 1994) se generan fragmentos proteicos que constituyen precursores con extensiones N-terminales de los péptidos de clase I definitivos (Cascio et al., 2001), si bien excepcionalmente actúan sistemas proteolíticos capaces de reemplazar la acción del proteasoma (Gil-Torregrosa et al., 2000; Shen et al., 2004; Parmentier et al., 2010) o generar el extremo C-terminal del péptido a partir de sus productos de digestión (Kessler et al., 2011; Shen et al., 2011).

Los productos de digestión proteasómica son en su mayor parte degradados por aminopeptidasas citosólicas, las cuales contribuyen asimismo al recorte inicial de los precursores peptídicos e incluso a la generación de algunos ligandos de HLA-I (Yewdell and Bennink, 2001; Rock et al., 2004).

La siguiente etapa de procesamiento se da en el retículo endoplásmico, donde son translocados los péptidos citosólicos mediante el Transportador Asociado con Procesamiento Antigénico (TAP), una proteína heterodimérica anclada a la membrana de este compartimento (Vos et al., 1999) que es capaz de transportar péptidos de hasta 16 residuos con determinados patrones de secuencia (Momburg et al., 1994; Van Endert et al., 1994; Van Endert et al., 1995; Van Endert, 1996). El procesamiento en este compartimento se lleva a cabo por la acción conjunta de las aminopeptidasas ERAP1 y ERAP2 (Saveanu et al., 2005), que eliminan residuos N-terminales en péptidos de longitud superior a 9 residuos (York et al., 2002), generando los péptidos definitivos con el tamaño adecuado para la unión a HLA-I.

En el retículo endoplásmico se lleva también a cabo la unión del péptido al resto de componentes de la molécula HLA-I. Este proceso implica la constitución de un complejo molecular, denominado Complejo de Carga Peptídica, que es necesario para el ensamblaje de la molécula de HLA-I (Wearsch and Cresswell, 2008). La cadena α recién sintetizada se asocia a una chaperona de la familia de las lectinas llamada calnexina, que la estabiliza y proporciona una conformación adecuada para su unión a β2M (Vassilakos et al., 1996; Paulsson and Wang, 2003). Con la unión de β2M se sustituye la calnexina por otra lectina, la calreticulina (Sadasivan et al., 1996; Vassilakos *et al.*, 1996). A continuación se une ERp57 (Hughes and Cresswell, 1998; Lindquist et al., 1998; Morrice and Powis, 1998), que forma un enlace disulfuro intracatenario con la tapasina (Peaper et al., 2005; Vigneron et al.,

2009) la cual interacciona con TAP y con la molécula de HLA-I facilitando la incorporación del péptido (Sadasivan *et al.*, 1996; Ortmann et al., 1997). El complejo de carga peptídica está constituido por las siguientes moléculas: HLA-I, calreticulina, ERp57-tapasina y TAP.

Con la incorporación del péptido, se libera la molécula de clase I con su conformación definitiva y se exporta a la membrana plasmática incluida en la cara luminal de un sistema de vesículas que atraviesan el Aparato de Golgi hasta la superficie celular. A este sistema de vesículas, cuya función es el transporte de moléculas a la membrana celular, se le conoce como ruta exocítica.

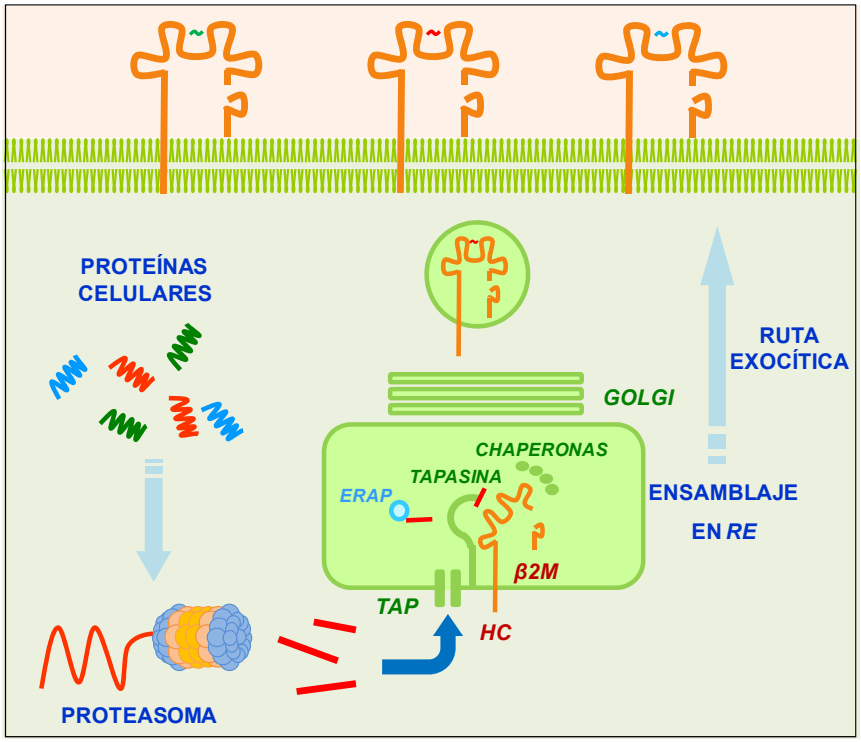


Figura 2: Ruta de procesamiento y presentación antigénica de clase I. Se muestra el proteasoma 26S, en tonos naranjas el núcleo catalítico 20S y en azul las subunidades reguladoras 19S. En el lumen del retículo endoplásmico se presentan en verde TAP, tapasina y las chaperonas que asisten en el plegamiento y carga peptídica de la molécula HLA-I.

1.2.1. Estructura y actividad proteolítica del proteasoma.

El proteasoma es un complejo multienzimático que puede adoptar diferentes estructuras. La más frecuente es conocida como proteasoma 26S, de 2.5 MDa (Glickman and Ciechanover, 2002), que está compuesta por 60-65 polipéptidos distribuidos en un núcleo catalítico, denominado proteasoma 20S, y dos subunidades, denominadas 19S, que regulan el acceso de los sustratos al proteasoma 20S (Peters et al., 1994; Voges et al., 1999; Ferrell et al., 2000; Zwickl, 2002). Éste puede formar complejos con otro tipo de subunidades reguladoras, tales como 11S (PA28) y PI31 (Qian et al., 2006; Voigt et al., 2007) y

formas híbridas compuestas por una subunidad 19S y una 11S (Tanahashi et al., 2000).

El proteasoma 20S está formado por 28 subunidades distribuidas en 4 estructuras anulares heptaméricas que se unen formando un cilindro en cuyo interior tiene lugar la hidrólisis del sustrato. Los anillos situados en los extremos del cilindro están formados por 7 subunidades α, sin actividad catalítica, que interaccionan con las subunidades reguladoras controlando la entrada del sustrato. El mecanismo de control que ejercen estas subunidades se sitúa en un *loop* conservado formado por los residuos Tyr126, Gly127 y Gly128 (Rabl et al., 2008). Los anillos centrales están formados por 7 subunidades β, 3 de las cuales presentan actividades proteolíticas específicas: tríptica, ubicada en la subunidad β_2, rotura preferente en el extremo C-terminal de residuos básicos; quimotríptica, en la subunidad β_5, rotura tras residuos apolares y caspasa, en la subunidad β_1, rotura tras residuos ácidos. Las dos primeras subespecificidades pueden generar directamente el extremo C-terminal de los péptidos de clase I.

La función reguladora de las subunidades 19S es dependiente de ATP. Estas subunidades están formadas por una estructura que forma una *base*, y otra en forma de *tapa* (Glickman et al., 1998; Hanna and Finley, 2007; Hanna et al., 2007). La base contiene 6 subunidades, Rpt1-6, que presentan actividad ATPasa y otras 4, Rpn1, Rpn2, Rpn10 y Rpn13, que no la presentan. La tapa está compuesta por las subunidades Rpn3, Rpn5-9, Rpn11, Rpn12 y Rpn15, todas ellas sin actividad ATPasa. Las subunidades con actividad ATPasa presentan un motivo C-terminal de tres residuos (hidrofóbico-tirosina-otro, HbYX) altamente conservado y necesario para la unión a 20S y la apertura del anillo α.

Existe otra estructura del núcleo catalítico del proteasoma, denominada inmunoproteasoma (Groettrup et al., 1997), que es inducible por IFN-γ, en el que las subunidades catalíticas β_1, β_2 y β_5 son reemplazadas por tres unidades homólogas β_{1i}, β_{2i} y β_{5i}. Se han descrito varias funciones del inmunoproteasoma. Su intervención es crítica en la producción del epítopo HBcAg procedente del virus de la hepatitis B (Sijts et al., 2000) y en la degradación de los agregados moleculares generados en condiciones de estrés oxidativo inducido por interferón (Seifert et al., 2010). Sin embargo, no es capaz de generar algunos epítopos tumorales (Morel et al., 2000). El peptidoma de ratones *knockout* para las subunidades inducibles del proteasoma presento un 50% de diferencias con respecto al fenotipo salvaje (Kincaid et al., 2012).

Se conoce también una estructura específica de tejido que se encuentra en las células corticales del timo y se caracteriza por la sustitución de la subunidad β_5 del proteasoma 20S por una subunidad denominada β_{5T} (Murata et al., 2007). Esta sustitución resulta en una disminución de la actividad de la subespecificidad quimotríptica, y presenta una función crítica en la selección positiva de células T CD8+.

1.2.2. Degradación proteasómica de proteínas.

La mayoría de las proteínas celulares son degradadas por el proteasoma (Pamer *et al.*, 1998; Rock *et al.*, 1999) en un proceso dependiente de ATP y de ubicuitina, una molécula empleada en el marcaje de proteínas intracelulares para su degradación (DeMartino et al., 1994; Knowlton et al., 1997; Glickman et al., 1999; Braun et al., 1999). Los procesos dependientes de ATP son la ubicuitilación de las proteínas y su unión al proteasoma, el desplegamiento del sustrato, la apertura del núcleo catalítico y la translocación del sustrato a su interior.

Para poder ser degradadas por el proteasoma 26S, las proteínas deben llevar asociadas un mínimo de cuatro subunidades de ubicuitina (Thrower et al., 2000), que se añaden a las proteínas formando cadenas poliméricas. Al conjunto formado por la maquinaria implicada en la ubicuitilación selectiva de proteínas para su degradación por el proteasoma 26S se le denomina Sistema Ubicuitina Proteasoma, cuya función es el control del ciclo de vida de las proteínas celulares.

La subunidad 19S reconoce y captura proteínas poliubicuitiladas, elimina las cadenas de poliubiquitina, despliega el sustrato, abre los anillos α e introduce el sustrato desplegado en la cámara catalítica. Rpn10 y Rpn13 son receptores de ubicuitina (Tanaka, 2009) y son las subunidades responsables del reconocimiento del sustrato. Rpn11 elimina las cadenas de poliubicuitina del sustrato cortando en un sitio proximal y liberando las cadenas enteras, que son degradadas por las enzimas de la deubicuitilización (DUBs), las cuales liberan ubicuitina monomérica para su reutilización. Existen dos DUBs, Usp14 y Uch37, que se asocian a la subunidad 19S y cortan la cadena de poliubicuitina en sitios distales. Rpn1, Rpn2, Rpt2, Rpt3 y Rpt5 inducen la apertura del anillo α y translocan el sustrato a la cámara catalítica (Tanaka, 2009). La unión del dominio HbYX de las ATPasas al anillo α induce un cambio conformacional en el que la Gly128 sufre una rotación que está asociada a la apertura del anillo (Rabl *et al.*, 2008).

Una vez en el interior del núcleo 20S, las subunidades de los anillos β con actividad catalítica degradan el sustrato actuando sobre la mayoría de los enlaces peptídicos, pero preferentemente sobre los que constituyen dianas trípticas, quimotrípticas o caspasa, en un proceso no dependiente de ATP.

1.2.3. Procesamiento no proteasómico de péptidos de clase I.

Se conocen proteasas alternativas al proteasoma implicadas en la generación de ligandos de clase I. Algunos ejemplos se presentan en las siguientes líneas.

La Furina, una convertasa de la familia de las subtilisin proteasas cuya función principal es el procesamiento de precursores de diversas proteínas sanguíneas, está involucrada en una vía alternativa

de procesamiento antigénico de clase I que se produce en la ruta secretoria de forma independiente de TAP y de proteasoma (Gil-Torregrosa et al., 1998; Gil-Torregrosa *et al.*, 2000). Esta enzima es capaz de procesar epítopos virales derivados de proteínas del virus de la hepatitis B y del VIH.

La inhibición del proteasoma no afecta a la generación del epítopo Nef(73-82) del VIH, mientras que la inhibición de Tripeptidil Peptidasa (TPP) II disminuye drásticamente su presentación (Seifert et al., 2003), lo cual sugiere un procesamiento citosólico mediado predominantemente por esta enzima. Sin embargo, se descartó una intervención generalizada de TPP II en la producción de ligandos independientes de proteasoma (Marcilla et al., 2008).

La Enzima de Degradación de la Insulina (IDE) es una metalopeptidasa capaz de generar directamente un epítopo de clase I derivado de la proteína tumoral MAGE-A3 (Parmentier *et al.*, 2010).

La acción conjunta de Nardilisina y la Oligopeptidasa Thimet es capaz de generar precursores con extensiones N-terminales del epítopo PRAME, derivado de la proteína EBNA3C del virus de Epstin-Barr, y de otro epítopo clínicamente relevante derivado de la proteína de melanoma MART-1 (Kessler *et al.*, 2011).

Aunque probablemente sea importante en situaciones particulares, la contribución global del procesamiento no proteasómico parece ser minoritaria en términos generales.

1.2.4. Procesamiento post-proteasómico de péptidos de clase I en el citosol.

Las actividades oligopeptidasa y ectopeptidasa en el citosol son muy eficientes, lo que conlleva un efecto general de degradación completa de los productos proteasómicos. No obstante, existen varias aminopeptidasas citosólicas cuya intervención en el procesamiento antigénico ha sido demostrada.

En el anterior epígrafe se han mencionado los casos de la TPP II y nardilisina, que pueden reemplazar al proteasoma en la producción de algunos epítopos de clase I. Estas enzimas actúan también sobre productos de degradación proteasómica. Un ejemplo lo constituye la intervención de TPP II de forma concertada con la aminopeptidasa sensible a puromicina (PSA) en la generación del extremo N-terminal de un epítopo de HLA-B51 derivado del producto de expresión del gen RU1 (Levy et al., 2002). Asimismo, la PSA y la Bleomicin hidrolasa conjuntamente son capaces de generar el extremo N-terminal de los epítopos NP(52-57) y NP(52-59) derivados de la nucleoproteína del virus de la Estomatitis Vesicular (Stoltze et al., 2000). Por otra parte el interferón γ estimula la actividad del proteasoma y de la leucin aminopeptidasa, que generan los extremos C- y N-terminal, respectivamente, del epítopo SIINFEKL derivado de la ovoalbúmina, lo cual no ocurre sin la mediación de esta citoquina (Beninga et al., 1998).

El papel individual de las aminopeptidasas citosólicas en el procesamiento post-proteasómico de ligandos de clase I no suele ser crítico, puesto que normalmente más de una enzima puede llevar a cabo el mismo efecto (Towne et al., 2005; Marcilla *et al.*, 2008).

1.2.5. Procesamiento de péptidos de clase I en el retículo endoplásmico.

Los primeros indicios de la existencia de procesamiento de antígeno en el retículo endoplásmico surgieron de la existencia de péptidos de clase I sin los motivos estructurales de transporte por TAP y de la capacidad de esta proteína de introducir péptidos de tamaño superior al admitido por las moléculas de clase I. La generación de epítopos derivados de precursores N-terminales introducidos directamente en el retículo endoplásmico de células con TAP no funcional confirmó la existencia de procesamiento en este compartimento (Elliott et al., 1995).

Posteriormente se caracterizó una metaloproteasa dependiente de zinc de la familia M1 en el retículo endoplásmico de células de ratón, a la que se llamó Aminopeptidasa Asociada al Procesamiento Antigénico en el Retículo Endoplásmico, ERAAP (Serwold et al., 2002). Salvo por su incapacidad para romper el enlace N-terminal de prolina, no se encontró una especificidad de residuo muy restrictiva y tras su actuación sobre extensiones N-terminales de un péptido de 8 residuos se obtenían siempre especies de este tamaño o superior. Tanto ERAAP como MHC-I son inducibles por interferón γ y se expresan en prácticamente todos los tejidos, pero en aquellos en que la expresión de MHC-I es particularmente alta, la expresión de ERAAP también lo es. Por otro lado, la reducción en la expresión de esta enzima implica una drástica disminución en la expresión de MHC-I. Todas estas características sugerían la existencia de una interacción funcional entre ERAAP, conocida en humanos como ERAP1 (Aminopeptidasa del Retículo Endoplásmico 1), y las moléculas MHC-I.

En posteriores trabajos se determinó la preferencia de ERAP1 por sustratos del tamaño que TAP es capaz de introducir en el retículo endoplásmico (Chang et al., 2005) y su pérdida de eficiencia cuando el tamaño del sustrato se reducía hasta el adecuado para la unión a MHC-I (York *et al.*, 2002). Aunque múltiples aminopeptidasas citosólicas son capaces de eliminar residuos en posiciones alejadas del epítopo, la eliminación de los dos residuos que le preceden ocurre esencialmente en el retículo endoplásmico (Schatz et al., 2008).

Aunque ERAP1 es capaz de eliminar cualquier residuo N-terminal no seguido de prolina, la naturaleza química de los aminoácidos influye en la eficiencia de corte (Hearn et al., 2009). Ésta influencia está también apoyada por estudios de digestión *in vitro* llevados a cabo con librerías de sustratos fluorogénicos (Zervoudi et al., 2011). En otros trabajos se definió también la influencia de la

secuencia del propio epítopo en la actividad enzimática de ERAP1 (Evnouchidou et al., 2008).

En el organismo humano ERAP1 está asociada, al menos parcialmente, a otra aminopeptidasa, ERAP2 (Tanioka et al., 2003), similar en función y estructura, que actúa de manera concertada con ésta y cuya intervención es crítica para la generación de algunos péptidos de clase I (Saveanu *et al.*, 2005). ERAP2 elimina con mayor eficiencia que ERAP1 residuos N-terminales básicos y presenta mayor dificultad para la eliminación de los residuos muy susceptibles a la acción de ERAP1, lo cual sugiere una acción complementaria de ambas enzimas. Contrariamente a ERAP1, la actividad de ERAP2 no está limitada por el tamaño del sustrato (Chang *et al.*, 2005) y su contribución real al procesamiento antigénico *in vivo* se desconoce.

Por otra parte se ha demostrado que una carboxipeptidasa residente en el retículo endoplásmico, la Enzima de Conversión de Angiotensina, es capaz de generar el extremo C-terminal de algunos epítopos y participa en la configuración de los peptidomas de clase I (Shen *et al.*, 2011). La deficiencia de este enzima tiene efectos dependientes de alotipo sobre el peptidoma de clase I, aunque actúa sobre sustratos muy heterogéneos en tamaño y secuencia. Su contribución no es bien conocida, pero no es probable que represente un componente cuantitativamente importante en el procesamiento antigénico.

1.2.6. Estructura y mecanismo de acción de ERAP1.

La estructura tridimensional de ERAP1 (Nguyen et al., 2011; Kochan et al., 2011) consiste en cuatro dominios que experimentan un reordenamiento espacial con la entrada del sustrato, el cual determina una transición desde un estado abierto (Figura 3A), de baja actividad, a un estado cerrado (Figura 3B), que presenta mayor actividad catalítica. El dominio I abarca los residuos 46-254 y adopta una estructura de sandwich β. Este dominio se superpone sobre el dominio IV en la transición al estado cerrado, con lo cual se posiciona sobre el sitio catalítico, proporcionando algunos sitios de unión de la región N-terminal del sustrato. El dominio II (residuos 255-529) contiene el centro catalítico, con actividad termolisina, constituido por los motivos secuenciales GAMEN y $HEXXH(X)_{18}E$, que incluye el sitio de unión del zinc, característicos de las metaloproteasas dependientes de zinc. El dominio III (residuos 530-614) presenta una estructura tipo sandwich β y sirve de conexión entre el dominio catalítico (II) y el dominio IV (residuos 615-940). En éste predominan las hélices α, que generan una estructura cóncava. Junto con el dominio I, el dominio IV contribuye al sellado del centro activo tras la entrada del sustrato.

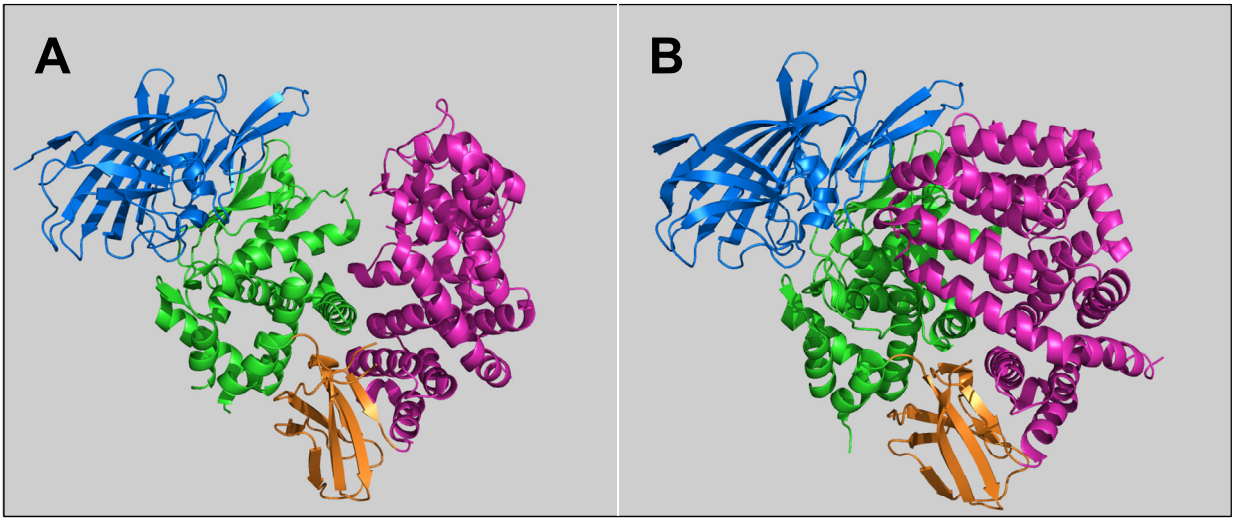


Figura 3: Estructura de ERAP1. Se muestra en azul el dominio I, en verde el dominio II, en naranja el dominio III y en magenta el dominio IV. La entrada del sustrato determina un cambio conformacional en la estructura de la molécula que pasa del estado abierto (A) (*Protein Data Bank* ID: 3QNF), que es poco activo, al estado cerrado (B) (*Protein Data Bank* ID: 2YD0), que presenta una alta capacidad hidrolítica.

La especificidad de sustrato, dependiente de longitud y del residuo C-terminal (York *et al.*, 2002; Chang *et al.*, 2005), supone una particularidad de ERAP1 que no comparten el resto de aminopeptidasas, incluyendo ERAP2. Precisamente esta propiedad confiere a ERAP1 la preferencia por los péptidos que TAP introduce al retículo endoplásmico, liberando frecuentemente productos del tamaño óptimo para la unión a HLA-I. El mecanismo propuesto para la selección de sustratos se define mediante un modelo que se denominó *regla molecular* (Chang *et al.*, 2005) y fue confirmado por los estudios cristalográficos (Gandhi et al., 2011). Este modelo supone la existencia de dos posiciones de anclaje donde se sitúan los extremos del sustrato. El residuo N-terminal que va a ser eliminado se sitúa en el sitio S1, donde reside la actividad aminopeptidasa, y el residuo C-terminal se sitúa en un bolsillo hidrofóbico, lo que dificulta la entrada de aminoácidos polares (Figura 4A). Cuando el residuo N-terminal es eliminado (Figura 4B), se produce la transición al estado abierto, lo que permite la salida del aminoácido libre y el desplazamiento del siguiente residuo hasta S1 (Figura 4C) (Kochan *et al.*, 2011). Puesto que el residuo C-terminal permanece durante todo el proceso en el bolsillo hidrofóbico, la enzima pierde procesividad cuando la longitud del sustrato no permite que el residuo N-terminal alcance el sitio S1 (Figura 4D).

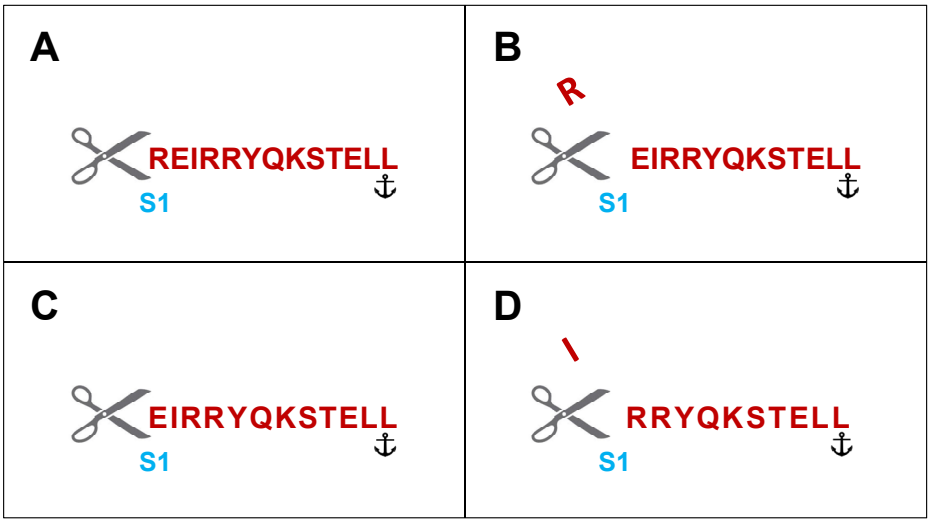


Figura 4: Regla molecular de selección de sustrato por tamaño en ERAP1. El residuo N-terminal se sitúa en el sitio S1, donde se produce la hidrólisis, y el residuo C-terminal permanece anclado al bolsillo hidrofóbico durante todo el proceso (A). Con la eliminación del residuo N-terminal se produce una transición al estado abierto, se libera el aminoácido (B) y se sitúa el nuevo residuo N-terminal en el sitio S1 (C). Cuando la longitud del sustrato se reduce hasta 8-9 residuos, el extremo N-terminal no alcanza el sitio S1 (D), con lo cual se detiene la hidrólisis y se libera el péptido.

Los modelos estructurales indican que M y L encajarían perfectamente en el bolsillo S1, mientras que la unión de otros de mayor tamaño, como W o R, requeriría un cambio conformacional de la cadena lateral, y residuos más pequeños no llenarían completamente este sitio (Nguyen *et al.*, 2011). La incapacidad del enzima para romper enlaces de P se justifica por la desestructuración de la región catalítica que implica el emplazamiento de éste aminoácido en S1', sitio donde se sitúa el residuo P2 del sustrato.

La especificidad de ERAP1 no se restringe a los extremos del sustrato, sino que el resto de la secuencia también influye en la eficiencia de corte del enzima (Evnouchidou *et al.*, 2008). En el trabajo de referencia se encontró una mayor actividad hidrolítica sobre sustratos con residuos internos positivamente cargados e hidrofóbicos, disminuyendo sensiblemente el rendimiento cuando predominaban residuos con carga negativa. La alta densidad de carga negativa en el sitio de unión del péptido explicaría la baja afinidad de ERAP1 por sustratos negativamente cargados.

1.2.7. El polimorfismo de ERAP1 y asociación con enfermedad.

1.2.7.1. Asociaciones históricas de ERAP1 con enfermedades.

El gen que codifica ERAP1 está localizado en el cromosoma 5q15 y su secuencia presenta variantes naturales determinadas por polimorfismos en nucleótidos individuales (SNPs) (Yamamoto et al., 2002). Desde que se describió la implicación de ERAP1 en múltiples procesos fisiológicos, muchos trabajos analizaron la asociación de su polimorfismo con diferencias de actividad enzimática y riesgo de desarrollo de enfermedades.

En los trabajos donde se publicó la existencia del polimorfismo natural de ERAP1 se propuso que algunas de las mutaciones podrían determinar un descenso en la actividad enzimática (Yamamoto *et al.*, 2002). Posteriormente, mediante análisis mutacionales se determinó que la sustitución de K528 disminuye sensiblemente la actividad hidrolítica *in vitro* (Goto et al., 2006). En los trabajos de Evnouchidou *et al*. (2010) se confirmó que los polimorfismos naturales K528R y Q730E disminuyen la actividad de ERAP1 y la expresión de HLA-B27. Estas mutaciones además provocan cambios en la cinética de degradación de algunos sustratos, que sigue comúnmente una cinética de Michaelis-Menten, para ajustarse mejor a una cinética de inhibición por sustrato.

El aparente carácter multifuncional de ERAP1 ha sugerido una posible implicación en procesos inflamatorios no relacionados con el procesamiento antigénico. Una de las funciones descritas, de relevancia inmunológica, es la activación de la capacidad fagocítica de los macrófagos, lo cual se produce por la secreción del enzima al espacio extracelular en respuesta a lipopolisacárico e interferón γ (Goto et al., 2011). Otra función propuesta de ERAP1, de especial importancia en procesos inflamatorios, es el *shedding* o escisión del dominio ectópico de algunos receptores de citoquinas para su liberación. Se reportó que una forma de ERAP1 que se encontraría en la membrana está involucrada en el recorte de IL-1RII, IL-6Rα y TNFR1 (Cui et al., 2002; Cui et al., 2003a; Cui et al., 2003b). La liberación de estos receptores constituiría un mecanismo regulador de la actividad de estas citoquinas proinflamatorias. Estos trabajos no han sido, hasta ahora, confirmados por grupos independientes.

La primera enfermedad asociada con el polimorfismo de ERAP1 fue la hipertensión arterial (Yamamoto *et al.*, 2002) y posteriormente, además del riesgo de espondilitis anquilosante (EA), que se detalla en el siguiente apartado, y de psoriasis (Strange et al., 2010), se encontraron también asociaciones más o menos sólidas con carcinoma de cérvix (Mehta et al., 2008), síndrome hemolítico urémico (Taranta et al., 2009), diabetes tipo I (Fung et al., 2009), melanoma (Kamphausen et al., 2010), y osteoporosis premenopaúsica (Li et al., 2011).

1.2.7.2. ERAP1 y espondilitis anquilosante.

La asociación de ERAP1 con EA fue descrita en 2007 (WTCC Consortium, 2007). En el citado trabajo se publicaron 5 SNPs no sinónimos, localizados en la región codificante del gen de ERAP1, asociados positiva o negativamente al riesgo de enfermedad en poblaciones británicas y norteamericanas de ascendencia europea, rs27044, rs17482078, rs10050860, rs30187 y rs2287987. Estos resultados fueron confirmados por otros estudios llevados a cabo en poblaciones canadienses (Maksymowych et al., 2009), españolas (Szczypiorska et al., 2011) y asiáticas (Davidson et al., 2009).

Posteriormente se demostró que el efecto del polimorfismo de ERAP1 sobre la predisposición a EA se restringe a individuos HLA-B27 positivos, una situación análoga a la reportada en psoriasis, donde la asociación de ERAP1 se restringe a pacientes portadores del alotipo HLA de susceptibilidad C*06:02 (Strange *et al.*, 2010). Además se estimó la importancia de ERAP1 en el riesgo de EA, proponiéndosele como el segundo gen más fuertemente asociado después de HLA-B27 (The TASK and WTCCC2 Consortia, 2011).

Aunque la asociación con EA está sólidamente demostrada, no se conoce el mecanismo por el cual el polimorfismo de ERAP1 contribuye al riesgo, ni las bases de su interacción funcional con HLA-B27. Existen al menos dos posibilidades: 1) ERAP1 tendría un papel proinflamatorio mediado por su capacidad de liberación de receptores solubles de citoquinas; 2) efectuaría cambios en el peptidoma de HLA-B27 que puedan afectar a las características inmunológicas y otras propiedades biológicas de la molécula.

Puesto que los receptores que liberaría ERAP1 corresponden a citoquinas proinflamatorias relacionadas con EA, sería posible que el polimorfismo de esta enzima afectase a la modulación de señalización mediante citoquinas, determinando el desarrollo de la enfermedad. Sin embargo, no se observaron cambios en los niveles séricos de las citoquinas dependientes de la dotación alélica de ERAP1 (Haroon et al., 2010), por lo tanto es muy improbable que sea este el mecanismo mediador.

Dado que el riesgo de enfermedad se restringe a individuos HLA-B27-positivos, resulta lógico pensar en un papel modulador de ERAP1 sobre el peptidoma de HLA-B27. Tal efecto podría tener consecuencias de varios tipos que inciden en las actuales hipótesis sobre el papel patogénico de HLA-B27. En primer lugar, de acuerdo con la hipótesis del péptido artritogénico (Benjamin and Parham, 1990) el polimorfismo de ERAP1 podría condicionar la presentación de un epítopo concreto con potencial autoinmune. Más generalmente, una alteración global del peptidoma podría afectar a la inmunogenicidad y capacidad tolerogénica de HLA-B27, propiedades ambas dependientes de la avidez del TCR (Manz et al., 2011) y por tanto de la densidad de epítopos, alterando su potencial autoinmune.

En segundo lugar, debido al papel esencial de los péptidos en el plegamiento de HLA-I, una alteración del peptidoma que influyese en la estabilidad global de HLA-B27 podría afectar al plegamiento de esta molécula, una propiedad que se ha postulado como patogénica (Mear et al., 1999), así como a su disociación en la superficie celular y subsiguiente formación de homodímeros de HC (Allen et al., 1999b; Bird et al., 2003), cuya función inmunomoduladora debida a su reconocimiento por receptores leucocitarios (Allen and Trowsdale, 2004) se ha propuesto asimismo como potencialmente patogénica (Allen et al., 1999a; Kollnberger and Bowness, 2009).

Estas consideraciones hacen que el análisis de los efectos del polimorfismo de ERAP1 sobre el peptidoma de HLA-B27, que constituye una parte de esta Tesis, adquiera particular importancia para esclarecer el mecanismo patogenético de la EA.

1.3. Métodos computacionales de análisis de datos en peptidómica.

Con la incorporación de las nuevas tecnologías que hacen posible la obtención de grandes volúmenes de datos se hace imprescindible un nuevo planteamiento para la extracción de conclusiones en los trabajos de investigación con carácter científico o de desarrollo tecnológico. La informática representa una base fundamental para el desarrollo de herramientas de asistencia en análisis que el investigador no podría efectuar manualmente. Los estudios relacionados con genómica y proteómica están entre los más afectados por estos avances.

El uso de estos métodos implica el establecimiento de nuevos criterios para evaluar la calidad de los resultados que sustituyan a la observación directa por parte del investigador. Estos criterios se recogen en parámetros de calidad o de confianza y se infieren de la probabilidad de asignación errática. Los parámetros estadísticos, como el *p-value* o el χ^2, las tasas estimadas de error o *False Discovery Rate* (FDR) y los múltiples sistemas de puntuación empleados en proteómica (*Xcore*, *sequest score*, *delta-score*, *p-ratio*) proporcionan un criterio rápido y fiable de validación del análisis automatizado de datos.

Esta Tesis está muy vinculada a la peptidómica, una categoría afín a la proteómica. El volumen de datos correspondió habitualmente a 10^3-10^4 péptidos por ensayo. Los procedimientos empleados siguen una estructura basada en los métodos de trabajo en proteómica de tercera generación o proteómica *shotgun*, que implica el uso de bases de datos y motores de búsqueda. Sin embargo, existen problemas específicos en peptidómica que las herramientas disponibles para proteómica no son capaces de afrontar, principalmente: 1) identificación de secuencias peptídicas y búsqueda de proteínas precursoras, puesto que la cantidad de péptidos representantes de cada proteína es muy limitada y proceden de una actividad hidrolítica inespecífica, 2) clasificación de los epítopos que integran un

peptidoma, extracción de patrones de secuencia y otras propiedades y 3) búsqueda de conjuntos solapantes en peptidomas similares. Esto motivó el desarrollo de un grupo de herramientas específicamente diseñadas para el análisis e interpretación de los resultados obtenidos.

1.3.1. Reconocimiento de patrones por métodos lineales y heurísticos.

El reconocimiento de patrones dominantes en colecciones de datos implica el establecimiento de un umbral de frecuencia de expresión para clasificar el patrón como frecuente o infrecuente. Para evitar el uso de criterios arbitrarios existen varias aproximaciones posibles.

Una posibilidad consiste en la comparación de las observaciones con una colección de valores de referencia, de manera que se clasificará un patrón como frecuente cuando su frecuencia sea significativamente superior a la de referencia, o bien como infrecuente si su frecuencia es significativamente inferior. Esto representó la base para el estudio de la frecuencia de uso de residuos en las posiciones principales de anclaje en peptidomas de HLA-B27 (Lopez de Castro et al., 2004), donde se clasificaron como favorecidos aquellos residuos cuya frecuencia en una posición superaba significativamente su abundancia en el proteoma humano, empleando el test χ^2 para determinar la significación estadística.

Existen otros métodos para la búsqueda de patrones de secuencia en colecciones de péptidos implementados en herramientas libremente disponibles. *IceLogo* emplea un algoritmo basado en teoría de probabilidades para la búsqueda de patrones de secuencia conservados (Colaert et al., 2009). *SMM-Align* (Nielsen et al., 2007) y *NNAlign* (Andreatta et al., 2011) son herramientas basadas en *Redes Neuronales Artificiales* (ver apartado 1.3.2) que permiten la identificación y alineamiento de núcleos de secuencia conservados. *NNAlign* incorpora una prestación adicional para evaluar no solamente la colección de secuencias sino también datos cuantitativos asignados a éstas, lo cual permite una estimación más precisa de la contribución de cada secuencia al patrón encontrado.

Los métodos basados en *redes neuronales* tienen una limitación adicional, puesto que emplean reglas de aprendizaje que están condicionadas a la disponibilidad de un conjunto de datos de entrenamiento con resultados conocidos. Una aproximación que permite obviar el uso de conjuntos de entrenamiento es el aprendizaje por *cross-validación*, consistente en fraccionar los datos de entrada en *n* grupos, entrenar la red con *n*-1 grupos y emplear el grupo restante para la validación de los resultados. Este proceso se repite *n* veces, de forma que cada grupo de datos se empleará *n*-1 veces en el entrenamiento de la red y una vez en la validación.

Otros métodos de clasificación independiente de datos de referencia consisten en una

comparación interna con la propia colección de datos, basada en la lógica heurística o lógica difusa. Este tipo de métodos proporcionan un tratamiento contextualizado de los datos para su clasificación en conjuntos difusos, que no presentan unos límites definidos (Figura 5). Cada dato no se clasifica dentro de un único conjunto difuso, sino que su contribución se distribuye en diferentes conjuntos en función de su valor. Los conjuntos difusos se distribuyen de acuerdo a los valores encontrados en la colección de datos, es decir, el criterio de clasificación depende del contexto. La clasificación de los datos por estos métodos emula los cuantificadores del lenguaje humano, "poco", "mucho", "bastante", "demasiado", por ejemplo, un patrón cuya frecuencia sea del 7% puede clasificarse como muy frecuente si la mayoría de patrones no superan el 3%, o como muy infrecuente si la mayoría superan el 20%. Así, en el estudio de la frecuencia de uso de residuos en peptidomas, un sistema basado en estos métodos sería capaz de detectar las posiciones en las que exista algún residuo favorecido, evitando la necesidad de evaluar cada una de las posiciones para determinar su localización.

Figura 5: Ejemplo propuesto para la explicación de los límites difusos de los conjuntos de clasificación empleando métodos heurísticos. Se observan dos zonas con alta densidad de población de mediciones, por encima de 30 ºC y por debajo de 4 ºC, estableciéndose estos límites para los conjuntos de valores que determinan calor y frío. El establecimiento de límites difusos implica que cada observación no se clasifica dentro de un solo conjunto, sino que tiene una contribución que afecta a 2 conjuntos adyacentes: una temperatura de 29 ºC puede contribuir con un 60% al conjunto templado y un 40% al de calor, o por otro lado, una temperatura de 0 ºC puede contribuir con un 90% al conjunto de frío y un 10% al templado.

1.3.2. Métodos estocásticos para la interpretación de datos experimentales.

La evaluación de datos mediante métodos lineales tiene una serie de limitaciones debido a su intolerancia a errores y a información incompleta en los datos de entrada. Por otra parte, estos métodos pueden ser excesivamente lentos a la hora de buscar soluciones a problemas complejos. Para salvar estas limitaciones se puede recurrir al uso de métodos estocásticos, consistentes en la búsqueda y optimización de soluciones basadas en la probabilidad.

Existen numerosos métodos estocásticos para el tratamiento de datos, muchos de ellos

inspirados en procesos biológicos, como son las *Redes Neuronales Artificiales*, cuyo funcionamiento está basado en el sistema nervioso central de los animales, y los *Algoritmos Genéticos*, basados en el proceso de la evolución.

Las redes neuronales artificiales pueden proporcionar soluciones válidas gracias a su capacidad de aprendizaje. El esquema más básico de red neuronal consiste en un grupo de unidades de computación distribuidas en una capa, llamadas *neuronas*, que pueden almacenar el conjunto de valores de entrada. Las *neuronas* de esta capa, denominada capa de sensores, establecen conexión con las *neuronas* de una segunda capa, denominada capa oculta, las cuales reciben una señal de cada *neurona* de la primera capa que va a depender del valor de entrada y de un parámetro propio de la conexión, denominado *peso sináptico*, que se relacionan mediante la *función de propagación*. Las señales recibidas por cada *neurona* de la capa oculta se integran mediante una *función de activación*, mediante la cual se obtiene el *potencial sináptico*, valor que debe superar el *umbral de activación* para que la *neurona* se active. Las *neuronas* activadas en la capa oculta propagan la señal a las *neuronas* de la capa de salida, que pueden activarse en función de las señales que reciban. El resultado del procedimiento es un vector de valores binarios correspondiente al patrón de activación de las *neuronas* en la capa de salida, que es dependiente de los valores iniciales de entrada.

El aprendizaje de la red consiste en ajustar los *pesos sinápticos* para obtener los valores de salida deseados. Estos ajustes se introducen mediante reglas de aprendizaje, comúnmente basadas en dirigir la salida hacia los valores esperados en un conjunto de entrenamiento, constituido por valores cuya salida es conocida (aprendizaje supervisado), o bien en el grado de familiaridad de la información de entrada (aprendizaje no supervisado o auto-organización).

Los algoritmos genéticos pueden proporcionar soluciones válidas gracias a su capacidad de evolucionar. El uso de estos métodos está condicionado a la posibilidad de codificar las soluciones en secuencias binarias y a la existencia de una función que permita la evaluación de estas soluciones. El tamaño optimo de las secuencias binarias necesario para poder codificar las soluciones se deduce de la Complejidad de Kolmogórov mediante la expresión $n = \Sigma\, n_i \mid (2^{ni} \geq X_i)$, siendo n el mínimo número de *bits* necesario, *i* representa las variables que conforman la solución, X_i el número de valores que puede adoptar cada variable y n_i el número mínimo de *bits* para codificar el valor de la variable. De esta forma, para codificar todos los posibles péptidos de 9 residuos compuestos por los 20 aminoácidos proteicos, tendríamos 9 variables que conforman la solución, los 9 residuos, que podrían adoptar 20 valores, los 20 aminoácidos, con lo que $n=\Sigma^9\, n_i \mid (2^{ni} \geq 20)$, n_i será siempre 5 (2^4=16 codificaciones, no serían suficientes, mientras que 2^5=32, quedando 12 codificaciones sin sentido) y el número mínimo de

bits 45. El proceso se inicia con la generación de una población de secuencias binarias aleatorias, denominadas *cromosomas*, que codifican para un conjunto de soluciones. Estas soluciones son evaluadas, asignando a cada una su puntuación o *fitness*, que debe dar cuenta de su grado de idoneidad como solución al problema planteado. El proceso de optimización se lleva a cabo eliminando las soluciones peor evaluadas y amplificando de nuevo la población con cromosomas procedentes de la recombinación de la información contenida en las soluciones seleccionadas por aplicación del operador *crossover*. Para garantizar la diversidad en las nuevas soluciones, se introducen modificaciones aleatorias en los cromosomas mediante el operador *mutación*. El proceso se repite varias veces acumulándose la información útil en cada iteración, y finaliza cuando se encuentra una condición de término (existencia de una solución aceptable) o bien cuando deja de haber convergencia (el *fitness* ha llegado a un máximo y no se incrementa en cada iteración).

1.3.3. Programación de tareas: *macros* y *scripts*.

Las *macros* son fragmentos de código que incluyen instrucciones para su ejecución dentro de una aplicación. Mediante el empleo de *macros* se pueden programar tareas e implementar funcionalidades adicionales en las aplicaciones que admitan su uso. Las aplicaciones que admiten el uso de *macros* incluyen comúnmente un editor de código para lenguaje *Basic*. MS Excel, para el trabajo con hojas de cálculo, y Data Explorer, software de adquisición de datos en espectrometría de masas, incluyen un sistema de desarrollo basado en Visual Basic 6.0 y unas librerías de objetos que permiten acceder a muchas de las funciones de la aplicación, pudiéndose resumir el trabajo del usuario cuando lleva a cabo tareas repetitivas, búsquedas o clasificación de grandes cantidades de datos.

Los ficheros de procesamiento por lotes o *scripts* pueden emplearse para programar tareas en el sistema operativo, lo que permite llevar a cabo una secuencia de trabajos sin la intervención del usuario y mantener los equipos constantemente operativos. El código se escribe en ficheros de texto plano empleando la sintaxis propia de un intérprete de comandos que se invoca dentro del script. En sistemas *Unix* el intérprete de comandos por defecto es *Bourne-again shell* (bash), cuya sintaxis se basa en otros más primitivos, *Korn shell* (ksh) y *C shell* (csh).

En MS-DOS la programación de tareas se lleva a cabo mediante el uso de *macros*, que se implementan en el sistema operativo como si se tratase de una aplicación, escritas en ficheros con extensión .bat con la sintaxis del sistema y algunas instrucciones específicas.

En esta Tesis se desarrolló una herramienta informática para la interpretación de espectros de fragmentación basada en un algoritmo genético, denominada MSgene cuyo funcionamiento se explica en el apartado 3.13.2 de Materiales y Métodos. La búsqueda de patrones en colecciones de secuencias peptídicas se llevó a cabo con el software MotifSearcher (ver apartado 3.13.3), basado en métodos estadísticos. La extracción y procesamiento de los datos espectrales de MALDI-TOF y MALDI-TOF/TOF se llevó a cabo mediante macros implementadas en el software Data Explorer 4.9. Las macros fueron también empleadas para múltiples análisis de datos con MS Excel.

Objetivos

El presente trabajo se planteó para ampliar el conocimiento de los procesos proteolíticos que intervienen en la ruta de procesamiento antigénico de clase I, centrándose en los sistemas principales que actúan en cada compartimento celular: la degradación de proteínas por el proteasoma en el citosol y el recorte de los precursores N-terminales de los péptidos de clase I por ERAP1 en el retículo endoplásmico. El desarrollo del trabajo se estructuró en cuatro objetivos:

1) Determinar si el patrón de proteínas parentales básicas de baja masa molecular, observado en los péptidos resistentes a inhibidores de proteasoma en HLA-B27 (Marcilla et al., 2007), se reproduce en otros alotipos de HLA-I, ampliando para ello este análisis al repertorio peptídico de HLA-A*68:01.

2) Determinar la eficiencia de los inhibidores de proteasoma sobre las subespecificidades enzimáticas y establecer si la expresión de péptidos resistentes a estos inhibidores puede explicarse por la actividad residual de un proteasoma parcialmente inhibido.

3) Determinar el efecto del polimorfismo natural de ERAP1 en la configuración del repertorio peptídico de HLA-B*27:04.

4) Identificar la base molecular de los efectos del polimorfismo de ERAP1 sobre el peptidoma de HLA-B27, para establecer el mecanismo de la interacción funcional entre ambas moléculas en la patogenia de la espondilitis anquilosante.

Materiales y Métodos

3.1. Líneas celulares y anticuerpos.

C1R es una línea celular linfoide B humana con baja expresión de sus antígenos HLA-I endógenos (Zemmour et al., 1992). El transfectante C1R-A*68:01 fue proporcionado por el Dr. Victor H. Engelhard (Beirne Cancer Center for Immunology Research, University of Virginia). C1R-B*27:04 fue obtenido en nuestro laboratorio (Garcia et al., 1997). Las líneas linfoblastoides humanas (LCL) HLA-B*27:04 positivas empleadas fueron: JSL (HLA-A*11; B*27:04, *48; C*02), WEWAK I (HLA-A*11, *24; B*27:04, *62; C*02, *04) y KNE (HLA-A*01, *02:04; B*27:04, *08).

El anticuerpo monoclonal W6/32 es una IgG2a que reconoce HLA-I en su conformación nativa (Barnstable et al., 1978). ME1 es una IgG1 que reconoce específicamente HLA-B27, B7 y B22 (Ellis et al., 1982). Ambos anticuerpos se purificaron mediante cromatografía de afinidad, en columna de proteína G-sefarosa, a partir del líquido ascítico extraído de la cavidad peritoneal de ratones balb/c inoculados con el correspondiente hibridoma.

3.2. Inhibidores de proteasoma.

La epoxomicina es un tripéptido modificado que se une covalentemente al centro activo del proteasoma inhibiendo su actividad de forma irreversible y específica (Kim et al., 1999). El tratamiento de células HeLa con una concentración 2 μM de epoxomicina redujo la actividad quimotríptica del proteasoma en un 98%, y la tríptica en un 87% (Kisselev et al., 2006). Los altos valores de inhibición combinada de estas subespecificidades hacen de esta droga un excelente método de bloqueo del procesamiento proteasómico.

El carbobenzoxi-L-leucil-L-leucil-L-leucinal (MG-132) es un tripéptido modificado que se une al centro activo del proteasoma de forma covalente pero reversible. Su acción no es específica, puesto que también inhibe otras proteasas (Vinitsky et al., 1994; Rock *et al.*, 1994). Ambos inhibidores se adquirieron a Calbiochem (Schwalbach, Alemania).

3.3. Aislamiento de peptidomas de HLA-I.

Los peptidomas de clase I se aislaron por cromatografía de afinidad (Marcilla *et al.*, 2007), empleando el anticuerpo W6/32 para A*68:01 a partir del transfectante en C1R y ME1 para B*27:04 a partir de Wewak I, JSL, KNE y C1R-B*27:04, ya que la presencia de otros antígenos de HLA-I en las LCL impide el uso de W6/32 y requiere el reconocimiento específico de HLA-B27.

3.4. HPLC y Espectrometría de Masas.

3.4.1. Separación cromatográfica de peptidomas de HLA-I.

Las mezclas peptídicas aisladas se sometieron a un fraccionamiento cromatográfico en un sistema de HPLC Waters Alliance 2690 (Waters, Milford, MA) con una columna C_{18} para cromatografía en fase reversa de 5 μm de tamaño de poro (Vydac, Hesperia, CA). La elución se llevó a cabo a un flujo constante de 100 μL/min, empleando un solvente de composición variable formado por la mezcla de dos soluciones, ácido trifluoroacético (TFA) 0.08% (solución A) y acetonitrilo 80%-TFA 0.075% (solución B), incrementando la proporción del segundo componente con el progreso de la cromatografía de la siguiente forma: 1) solución A 100% durante 15 min, 2) gradiente lineal incrementando de 0-44% la composición de solución B durante 90 min, 3) gradiente lineal incrementando la composición de solución B de 44-100% durante 35 min (Paradela et al., 2000). La abosorbancia se monitorizó a lo largo de toda la cromatografía a 210 y 280 nm. Se colectaron 150 fracciones de 50 μL entre los minutos 40 y 115.

3.4.2. Espectrometría de Masas MALDI-TOF.

Las fracciones de HPLC se secaron en *speedvac* y se reconstituyeron con una solución acuosa de acetonitrilo 33%, TFA 0.1% (TA). Estas muestras se cargaron en una placa Optiplex de 384 *spots*, se añadieron 0.6 μL de α-ciano-4-hidroxicinámico 3 mg/mL en TA y se analizaron en un equipo MALDI-TOF/TOF 4800 Proteomics Analyzer (Applied Biosystems, Foster City, CA) empleando el software Data Explorer 4.9. Los datos se adquirieron en modo reflector positivo a 25 KV, en el rango de masa de 800-2000 Da para las fracciones cromatográficas de los peptidomas y en el rango 700-4000 Da para las digestiones de sustratos sintéticos.

3.4.3. Marcaje Isotópico con Aminoácidos.

La estrategia empleada para la identificación de péptidos sensibles y resistentes a inhibidores de proteasoma se basó en el método de marcaje isotópico no homogéneo con aminoácidos descrito previamente (Marcilla *et al.*, 2007), con algunas modificaciones. En resumen, cada péptido en un espectro de masas de MALDI-TOF no se corresponde con una única señal, sino que se presenta como un grupo de señales correspondientes a especies que incorporan isótopos pesados de manera natural. La distribución de las intensidades de estas señales sigue un patrón característico determinado por la abundancia de cada especie, que depende de la probabilidad de incorporación de isótopos pesados, denominado envoltura isotópica. En cultivos celulares que crecen en presencia de aminoácidos

marcados con isótopos estables, los péptidos que incluyen estos aminoácidos presentan una envoltura isotópica distorsionada por el enriquecimiento en especies más pesadas (Figura 6A). Con la adición simultánea del aminoácido marcado y un inhibidor de proteasoma al cultivo celular, solo se observaría distorsión en la distribución isotópica de los péptidos resistentes, que son producidos en presencia del inhibidor, mientras que los péptidos sensibles dejan de producirse y conservan la distribución isotópica normal (Figura 6B).

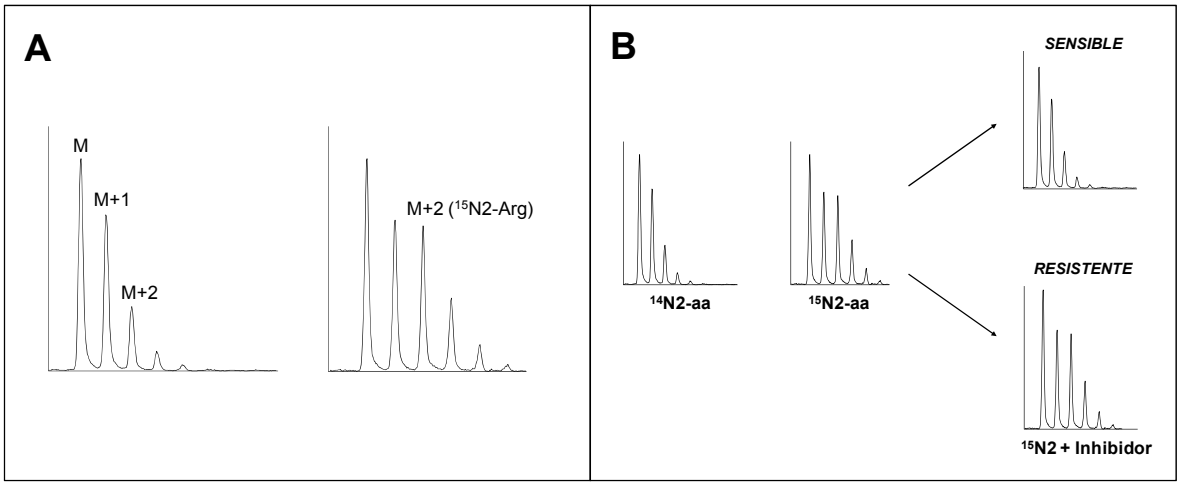


Figura 6: Método empleado para el seguimiento de la incorporación de péptidos al repertorio de HLA-I en presencia de inhibidores de proteasoma por espectrometría de masas MALDI-TOF. (A) La incorporación de péptidos marcados con aminoácidos-^{15}N implica un incremento de la intensidad de la señal isotópica correspondiente a la masa del péptido marcado con respecto a lo observado en ausencia del marcaje. (B) Si el efecto descrito persistía tras el tratamiento con inhibidores, el péptido se clasificó como resistente al inhibidor, mientras que si desaparecía, el péptido se clasificó como sensible.

Aprovechando la preferencia casi exclusiva de HLA-A68 por péptidos con residuos C-terminales básicos, se seleccionaron los aminoácidos marcados L-Arg-guanidino[$^{15}N_2$]·HCl y L-Lys-guanidino[$^{15}N_2$]·HCl (Cambridge Isotope Laboratories, Andover, MA) para marcar la mayoría de su repertorio peptídico.

Los transfectantes C1R-A*68:01 se distribuyeron en grupos de tres botellas por ensayo ($5 \cdot 10^8$ células en cada uno) y se cultivaron en RPMI 1640 sin K o sin R, para los marcajes con K o R respectivamente, suplementado con 10 % de suero fetal bovino dializado (ambos de Gibco, Paisley, UK) a 37 ºC durante 4 ó 12h. Transcurrido este tiempo, se suplementó una de las botellas con 100 µM del aminoácido no marcado y los otros dos, uno de los cuales se había tratado 30 min antes con epoxomicina 2.5 µM ó MG-132 20 µM, con 100 µM del aminoácido marcado. Después de 4h de incubación se lavaron dos veces las células con Tris-HCl 20 mM pH 7.5, NaCl 150 mM y se

conservaron los sedimentos celulares a -70 ºC en presencia de 500 µL de una solución etanólica de fluoruro de fenilmetilsulfonilo (Sigma-Aldrich, St. Louis, MO) a 70 mg/mL.

El marcaje de los péptidos se cuantificó empleando un valor numérico que relaciona las intensidades de señal de la especie isotópica relevante y la monoisotópica en los espectros de masas de los controles con y sin inhibidor, que se calculó mediante la expresión $[(^{15}N+inh)-^{14}N]/(^{15}N-^{14}N)$, donde ^{14}N, ^{15}N y $(^{15}N+inh)$ son los porcentajes de intensidad de la señal marcada relativa a la monoisotópica, en ausencia de marcaje, solo con marcaje y con marcaje e inhibidor, respectivamente.

3.4.4. Fragmentación por MS/MS MALDI-TOF/TOF.

La secuenciación de los péptidos de interés se llevó a cabo por fragmentación por colisión con aire atmosférico en el mismo instrumento que en el apartado 3.4.2 a 1KV con una ventana de masa del precursor de ± 2.5 Da. Los espectros de fragmentación se adquirieron con el software Data Explorer 4.9 con una relación señal/ruido de 3. La interpretación de los datos espectrales fue asistida por varias herramientas informáticas. En primera instancia se trataron los espectros de fragmentación en una hoja de cálculo de Excel con una macro que genera una colección de secuencias parciales tentativas y los datos de entrada para los motores de búsqueda. Con estos datos se llevó a cabo la búsqueda de coincidencias frente a la base de datos Uniprot/Swissprot 14.0 (Julio 2008; 19329 entradas) mediante el software Mascot 2.2 (Matrix Science Inc. Boston, USA; www.matrixscience.com), restringiendo la búsqueda al proteoma humano y con un intervalo de confianza de ±0.5 para el valor de m/z del precursor y los fragmentos iónicos. Las tres secuencias candidatas con las puntuaciones (*Ion Score*) más altas se sometieron a fragmentación teórica mediante la herramienta MS-Product 5.1.8 (University of California, San Francisco, USA; http://prospector.ucsf.edu) y se comparó la lista de masas de los fragmentos teóricos con las especies iónicas del espectro de fragmentación, empleando una segunda macro implementada en la misma hoja de cálculo, para confirmar la asignación de la secuencia. Las fragmentaciones que no pudieron ser interpretadas de esta manera, se trataron de resolver por secuenciación *de novo* mediante el software MSgene desarrollado en esta Tesis, cuyo funcionamiento se explica en el apartado 3.13.2.

3.4.5. Fragmentación en Trampa Iónica Orbitrap.

Las mezclas peptídicas se desalaron y concentraron con columnas micro-tip C_{18} para fase reversa de 200 µL (Harvard Apparatus, Holliston, MA). Las columnas se lavaron con acetonitrilo 80%-TFA 0.1% y se homogeneizaron con TFA 0.1%. Se cargó la mezcla peptídica, se lavó dos veces con

TFA 0.1% y se eluyó con acetonitrilo 80%-TFA 0.1%. El eluído se concentró posteriormente hasta 18 µL por centrifugación a vacío y se analizó por µLC-MS/MS en un espectrómetro de masas Orbitrap XL (Thermo Fisher, San José, CA) acoplado a un HPLC capilar (Eksigent, Dublin, CA). Los péptidos se separaron en una columna C_{18} de 0.3x5 mm (LC-Packings) eluyendo con un gradiente de acetonitrilo 7-95% en ácido fórmico 0.1%.

Los espectros de masas se adquirieron en el rango de m/z 400-2000 y se fragmentaron las siete masas más intensas de cada espectro en los estados de carga +1, +2 y +3 por desintegración inducida por colisión. Los péptidos se identificaron usando los motores de búsqueda Pep-Miner (Beer et al., 2004), Sequest (Thermo Fisher) y Mascot 2.2 sobre la base de datos Uniprot (Enero 2009; 20332 entradas). Se aceptaron las identificaciones de péptidos de masa inferior a 1500 Da con precisión de 0.005 Da y que superaban al menos tres de los siguientes criterios: *score* en Pep-Miner ≥ 80, *Xcore* en Sequest ≥ 2.5, probabilidad en Seaquest ≥ 5, *ionScore/highScore* en Mascot ≥ 0.6, valor esperado en Mascot ≥ 0.1, *delta* hasta el siguiente *score* = 0. Se seleccionaron secuencias concordantes con el motivo de HLA-A*68:01, con un *score* ≥ 20 en el análisis con el software de predicción de unión a HLA (http://www-bimas.cit.nih.gov/molbio/hla_bind/). El FDR se calculó como el porcentaje de asignaciones empleando una base de datos aleatoria y fue inferior al 3.9%. Estos análisis se llevaron a cabo en colaboración con el Profesor Arie Admon (Faculty of Biology, Technion, Israel Institute of Technology, Haifa).

3.5. Análisis de frecuencia de residuos en el peptidoma de HLA-A*68:01.

Éste análisis se llevó a cabo empleando el software MotifSearcher cuya descripción se presenta en el apartado 3.13.3.

3.6. Identificación y análisis de proteínas parentales de ligandos de HLA-A68.

La identificación de proteínas precursoras se llevó a cabo por asignación de los ligandos a secuencias internas de proteínas en la base de datos UniProt 14.0 (2008/07/22) mediante el software Fasta 3 (www.ebi.ac.uk/fasta). Cuando un ligando se podía asignar a varios miembros de una familia de proteínas se seleccionó un miembro representativo, asumiendo que el péptido puede proceder de más de una de estas proteínas.

La masa molecular (MW) y punto isoeléctrico (pI) de las proteínas asignadas se calculó con la herramienta Compute pI/MW (www.expasy.org/tools/pi_tool.html). La localización subcelular de las proteínas se obtuvo de la base de datos UniProt 14.0 (2008/07/22) mediante la herramienta Protein

Information and Knowledge Extractor (Medina-Aunon et al., 2010), libremente disponible en la dirección http://proteo.cnb.csic.es:8080/pike.

3.7. Ensayos de inhibición de proteasoma *in vitro*.

3.7.1. Hidrólisis de sustratos fluorogénicos en solución.

Se incubaron $2 \cdot 10^6$ células en medio IMEM sin rojo fenol suplementado con 10% de suero fetal bovino (ambos de Gibco, Paisley, UK) durante una hora en presencia de diferentes concentraciones de epoxomicina, MG-132 o sin inhibidores. Las células se lavaron dos veces con PBS y se permeabilizaron con 1.4 mL de Tris-HCl 50 mM pH 7.5, sacarosa 250 mM, $MgCl_2$ 5 mM, EDTA 0.5 mM, DTT 1 mM, ATP 1 mM, digitonina 0.025% (solución de homogenización) durante 5 min a 4 ºC. Se eliminaron los sólidos celulares por centrifugación a $2 \cdot 10^4$xg, 10 min y los sobrenadantes con el contenido citosólico se centrifugaron a $3 \cdot 10^5$xg, 2h, obteniendo un sedimento con los complejos moleculares de alta masa molecular, incluyendo el proteasoma 26S, que se resuspendieron en 1.4 mL de tampón de homogenización sin digitonina. En otros ensayos la hidrólisis del sustrato fluorogénico se llevó a cabo directamente con células vivas.

Las actividades tríptica y quimotríptica en estos sedimentos se calcularon en base a su capacidad de hidrolizar los sustratos fluorogénicos terc-butoxicarbonil-L-leucil-L-arginin-L-arginin aminometil cumarina (Boc-LRR-amc) y succinil-L-leucil-L-leucil-L-valin-L-tirosin aminometil cumarina (Suc-LLVY-amc) (Bachem, St. Helens, UK), respectivamente. Se incubaron los extractos celulares o las células vivas con una concentración 100 μM del sustrato fluorogénico a 37 ºC durante 1h, extrayendo alícuotas cada 10 min en las que se paró la reacción por adición de TFA 0.33%.

La intensidad de fluorescencia se midió en un fluorímetro Aminco-Bowman serie 2 (Sim-Aminco Spectronic Instruments, Rochester, NY) a 380 y 460 nm de longitudes de onda de excitación y emisión, respectivamente. La tasa de hidrólisis se calculó como el cociente entre las pendientes de las curvas de progreso de la reacción en presencia y ausencia de los inhibidores.

3.7.2. Hidrólisis de sustratos fluorogénicos en gel nativo.

Se cargó el proteasoma 26S parcialmente purificado en un gel de poliacrilamida 3% preparado en Tris 90 mM/ácido bórico, pH 8.3, $MgCl_2$ 5 mM, EDTA 0.5 mM, ATP-$MgCl_2$ 1 mM, y se corrió durante 3 horas a 100V y 4ºC (Elsasser et al., 2005). Otras muestras se corrieron en un gel con un gradiente de concentración 3-14% de poliacrilamida en la misma solución o invirtiendo la polaridad. Tras la electroforesis, los geles se lavaron con Tris-HCl 50 mM pH 7.4, $MgCl_2$ 5 mM, ATP 1 mM

(solución de desarrollo) y se incubaron en 1 mL de esta solución con una concentración 50 μM del sustrato fluorogénico a 30ºC durante 30 min. Las bandas electroforéticas con actividad hidrolítica se detectaron visualizando los geles en un transiluminador con luz ultravioleta.

3.7.3. Digestión de precursores sintéticos de péptidos de HLA-I *in vitro* por proteasoma 20S.

Se emplearon dos precursores peptídicos de ligandos de HLA-A68, de 30 y 31 residuos, que constituían secuencias internas de la proteína S21 de la subunidad ribosomal 40S.

El proteasoma 20S se aisló de células C1R por cromatografía de intercambio iónico, como se describió previamente (Paradela *et al.*, 2000). Se prepararon alícuotas de 2 μg de proteasoma 20S y se trataron con los inhibidores durante 2h a 30ºC previamente a las digestiones. Se incubaron 20 μg de sustrato con las alícuotas de proteasoma 20S durante 8h a 30ºC. La reacción se paró por adición de TFA 0.067%. Los productos de digestión del 31-mero se separaron del proteasoma por filtrado molecular con centricón de 10 KDa de tamaño de exclusión y se fraccionaron por HPLC de la misma forma que los peptidomas aislados de células. Las fracciones de HPLC se analizaron por espectrometría de masas MALDI-TOF/TOF. Los productos de digestión del 30-mero se purificaron con columnas C_{18} (Agilent Technologies, Santa Clara, CA) previamente lavadas con TFA 0.1% y acetonitrilo 40%-TFA 0.1%, eluyendo los péptidos con acetonitrilo 70%-TFA 0.1%. La mezcla peptídica se analizó directamente por MALDI-TOF/TOF.

Las masas de los productos de digestión se asignaron a secuencias internas del precursor empleando el software *Digestor*, desarrollado en esta Tesis y que se explicará en el apartado 3.13.4. Cuando una masa se podía asignar a más de una secuencia, la ambigüedad se resolvió habitualmente secuenciando la especie por fragmentación en MALDI-TOF/TOF.

3.8. Caracterización de ERAP1 en líneas celulares.

3.8.1. Tipaje de SNPs no sinónimos en el gen de ERAP1.

Para la purificación de DNA se empleó el sistema High Pure PCR Template Preparation (Roche Diagnostics, Barcelona, España) siguiendo las instrucciones del fabricante. Se añadieron alícuotas de 10 ng a una placa de 384 pocillos por duplicado, se secó y se amplificó usando oligonucleótidos específicos para ocho SNPs no sinónimos localizados en la secuencia codificante de ERAP1: rs6653, rs26618, rs27895, rs27044, rs30187, rs17482078 y rs2287987. Las muestras se corrieron en un sistema HT7900 Fast real-Time PCR y se genotiparon empleando el software SDS2.2 (ambos de Applied Biosystems, Life Technologies, Carlsbad, CA) para discriminación alélica.

3.8.2. Secuenciación de las variantes de ERAP1.

Se amplificaron los exones 2-20 de la región codificante de ERAP1 por PCR y se clonaron en el plásmido M13. Para la PCR se empleó AmpliTaq Gold Master Mix (Applied Biosystems) siguiendo procedimientos estándar, y los productos de amplificación se purificaron usando ExoSap (USB Corp., Cleveland, OH) y se secuenciaron en un dispositivo 3730XL (Applied Biosystems). Los *primers* empleados para la secuenciación fueron oligonucleótidos complementarios de M13, para los amplicones que incluían esta extensión en la secuencia de ERAP1, o los propios *primers* específicos cuando no incluían secuencias derivadas de M13.

3.8.3. Análisis de expresión de ERAP1 por PCR cuantitativa en tiempo real (qRT-PCR).

Se extrajo el RNA celular total empleando RNeasy Mini Kit (Qiagen) y se digirió con DNasa I (Invitrogen, Karlsruhe, Alemania). El DNA complementario se sintetizó a partir de 250 ng de RNA total empleando el kit de transcripción inversa de cDNA de alta capacidad (Life Technologies) siguiendo las instrucciones del fabricante. Los *primers* usados para la amplificación se adquirieron a Applied Biosystems (Life Technologies). La comparación cuantitativa de expresión génica se llevó a cabo empleando un dispositivo AB7900HT (Applied Biosystems, Life Technologies) usando sondas TaqMan y Gene Expression Master Mix (Applied Biosystems, Life Technologies). Las amplificaciones se desarrollaron con un calentamiento inicial a 50ºC durante 2 min seguido por desnaturalización a 95ºC, 10 min, 40 ciclos de desnaturalización a 95ºC, 15s y rehibridación y elongación a 60ºC, 60s. Los resultados se presentaron como expresión relativa de mRNA, que se cuantificó con el software RQ Manager normalizando con los niveles de transcripción de la gliceraldehído-3-fosfato deshidrogenasa.

3.8.4. Análisis de expresión de ERAP1 por Western Blot.

Se lisaron $2x10^5$ células en Igepal 0.5% (Sigma-Aldrich, St Louis, MO), Tris HCl 50 mM pH 7.4, $MgCl_2$ 5 mM en presencia de inhibidores de proteasas (Complete Mini, Roche, Mannheim, Alemania). Los lisados se sometieron a separación electroforética SDS-PAGE en un gel de poliacrilamida 10% en condiciones reductoras. Las proteínas se trasladaron a una membrana de nitrocelulosa (Amersham, Hybond-ECL, GE Healthcare, UK) por electrotransferencia a 20V durante una noche usando una solución de metanol 30% en Tris/glicina 50 mM, pH 8.8, SDS 0.04%. ERAP1 se reveló con el anticuerpo monoclonal 6H9, proporcionado por el Dr. Peter van Endert (INSERM, Paris, Francia) usando un anticuerpo policlonal de cabra anti-ratón conjugado con peroxidasa (DakoCytomation, Glostrup, Dinamarca). La γ-tubulina se empleó como estándar interno y se reveló

con el anticuerpo GTU88 (Sigma-Aldrich). Se escanearon las autoradiografías y se cuantificaron con el software de análisis de imagen TINA 2.09 (Raytest Isotopenmessgerate, Straubenhardt, Alemania).

3.9. Análisis comparativo de niveles de expresión de ligandos de B*27:04.

Los niveles relativos de expresión de ligandos de B*27:04 comunes a las LCLs se estimaron a partir de la relación de intensidades (RI) de las señales correspondientes en los espectros de masas MALDI-TOF. Para ello se llevaron a cabo tres comparaciones de peptidomas de LCLs por grupos de dos: Wewak I/JSL, Wewak I/C1R-B*27:04 y C1R-B*27:04/KNE. Para minimizar el error experimental, los peptidomas comparados se fraccionaron en cromatografías desarrolladas en condiciones idénticas y las fracciones cromatográficas se cargaron en una misma placa de MALDI-TOF. Los ligandos se clasificaron en tres subconjuntos atendiendo a su RI en las líneas celulares comparadas: 1) péptidos predominantes, cuya intensidad de señal en una de las líneas supera en más de tres veces la intensidad en la otra línea (RI>3.0), 2) péptidos moderadamente predominantes, con 3.0≥RI>1.5, y 3) péptidos de expresión similar, con 1.5≥RI≥1.0.

De esta forma fue posible determinar diferencias en los niveles de expresión dependientes del tamaño de los ligandos. No obstante, puesto que la capacidad hidrolítica de ERAP1 no está limitada por la masa molecular de sus sustratos sino por su longitud, fue necesario llevar a cabo un cálculo estimado de longitudes en estos peptidomas. La distribución de longitudes peptídicas en cada uno de los subconjuntos descritos se infirió de su distribución de tamaños y de la relación entre el tamaño y la longitud de los ligandos cuyas secuencias se determinaron. Los cálculos se desarrollaron de la manera que se describe a continuación. En primer lugar se clasificaron las secuencias obtenidas en cuatro grupos atendiendo a su longitud (<9, 9, 10 y >10 residuos). El intervalo de m/z de 850-1700, en el que se encontraban los ligandos, se dividió en nueve rangos de 100, a excepción del primer rango, que se estableció en 850-900. Para cada rango de tamaño se determinó el porcentaje de péptidos secuenciados con cada longitud. Los valores obtenidos se trasladaron a una matriz de dimensión 9x4 (9 rangos de tamaño, 4 rangos de longitud). A continuación se convirtieron todos los valores de m/z de todos los péptidos de cada subconjunto en los rangos de longitud establecidos. El resultado es una matriz que indica la cantidad estimada de ligandos de cada longitud en cada rango de m/z.

3.10. Susceptibilidad de los residuos N-terminales flanqueantes y P1 a la acción de ERAP1.

Se estableció un sistema para especificar la susceptibilidad de los residuos flanqueantes N-terminales (P-2 y P-1) y del residuo P1 de los ligandos a la acción hidrolítica de ERAP1. El sistema se basó en un estudio previo (Hearn *et al.*, 2009) en el que se evaluó la eficiencia con que ERAP1 elimina cada aminoácido cuando éste se coloca en el extremo N-terminal de un ligando de MHC-I (SIINFEKL). Los aminoácidos se puntúan con un valor de 0-100 en orden creciente de susceptibilidad a ERAP1 (Apéndice 2, Tabla S1). Dada la limitada capacidad de este enzima para hidrolizar enlaces de P, cualquier residuo precedente a prolina se puntuó con valor cero. Para cada grupo de péptidos se determinó el valor medio de puntuación en P1, P-1, P-2 y la suma de las contribuciones P-1 y P-2. La relación entre la puntuación de susceptibilidad en grupos equivalentes de péptidos de distintas líneas celulares se empleó como una estimación global de su susceptibilidad relativa a ERAP1.

3.11. Contribución de los residuos internos a la susceptibilidad a la hidrólisis por ERAP1.

La influencia de los residuos internos del epítopo se estimó mediante un sistema de calificación basado en un estudio previo (Evnouchidou *et al.*, 2008) en el que se analizó el efecto de diferentes aminoácidos en posiciones posteriores a P1 de los ligandos de MHC-I sobre la actividad hidrolítica de ERAP1, empleando para ello librerías de nonámeros. La puntuación establecida es dependiente del aminoácido y de la posición en el péptido (Apéndice 2, Tabla S1). Debido al uso exclusivo de nonámeros en el trabajo de referencia esta estimación se restringió a los péptidos de esta longitud identificados en esta Tesis. Puesto que en el trabajo de referencia no se evaluaron todos los aminoácidos en cada posición, las puntuaciones correspondientes se dejaron en blanco en nuestras series peptídicas.

3.12. Ensayos de termoestabilidad.

Estos ensayos se llevaron a cabo por un procedimiento descrito previamente (Merino et al., 2008). En forma resumida, se marcaron las células durante 15 min con ^{35}S-Met y se realizaron cazas a varios tiempos. En cada tiempo de caza los lisados celulares se incubaron 1h a varias temperaturas, se inmunoprecipitaron con ME1 y se analizaron por SDS-PAGE. La cantidad de heterodímero precipitado a cada temperatura se determinó por autoradiografía, se expresó como porcentaje relativo a la cantidad precipitada a 4ºC y se representó en función de la temperatura.

3.13. Desarrollo de herramientas informáticas.

3.13.1. Análisis cualitativo de peptidomas y marcajes isotópicos: MSHandler.

Para simplificar la comparación de peptidomas y la localización de péptidos marcados en las series de datos de HPLC-MS se desarrolló una herramienta informática, denominada MSHandler, capaz de llevar a cabo estas tareas de forma automática. Esta herramienta consiste en una aplicación java multiplataforma desarrollada en el entorno NetBeans 6.9 (Oracle, Redwood Shores, CA), que puede ejecutarse en la mayoría de las arquitecturas de PC, en sistemas UNIX y Microsoft Windows, así como en máquinas de Apple.

El método se basa en la probabilidad de que en repertorios peptídicos de alta homología, señales de masas equivalentes con tiempos de retención similares correspondan al mismo péptido. Sin embargo, existen algunas complicaciones que se detallan a continuación. Dependiendo de la resolución de la separación cromatográfica, cada péptido puede eluir en varias posiciones consecutivas. Por otra parte, en los espectros de masas frecuentemente aparecen señales inespecíficas que no corresponden a péptidos, tales como las señales de la matriz, las cuales se pueden identificar en el espectro de masas procedente del análisis de *spots* en los que no se carga muestra, con el cual se puede configurar una lista de exclusión. Para la eliminación de las señales inespecíficas, redundantes, o procedentes de ruido instrumental, se someten los datos a un preprocesado en que se seleccionan las señales que corresponden a máximos en el perfil de elución de cada péptido, tras descartar las que coinciden con las señales de la lista de exclusión y las que no superan un umbral mínimo de intensidad. Otros parámetros de exclusión, determinados por las características de los espectros, pueden ser definidos por el usuario.

Las señales iónicas se distribuyen en un sistema de dos coordenadas correspondientes a su masa y su tiempo de retención (Figura 7A). Los intervalos de confianza para que dos señales puedan ser consideradas como equivalentes dependen de la precisión instrumental y de la resolución de la cromatografía. La escala de los ejes se normaliza para obtener unidades proporcionales al mínimo común múltiplo de los intervalos de confianza para masa y posición de elución, de manera que la distancia entre dos puntos para que puedan considerarse como masas equivalentes es la misma en el eje de masa y el de tiempo de retención. De esta forma, se establece un campo de búsqueda circular, centrado en cada señal, que viene definido por un radiovector de tolerancia proporcional al mínimo común múltiplo de los intervalos de confianza. Las señales presentes dentro del campo de búsqueda de otra señal son asignadas como equivalentes (Figura 7B). La condición aritmética que determina que dos señales de masas, m_i y n_j, coincidan dentro de sus campos de búsqueda viene determinada por la expresión:

$$m_i \in \{P1\} \text{ eq } n_j \in \{P2\} \Rightarrow d(m_i, n_j) \leq \rho$$

donde P1 y P2 son las series de datos de HPLC-MS de dos peptidomas, d(m_i, n_j) es la distancia euclídea entre las dos señales en el plano masa/posición de elución normalizado y ρ es el módulo del radiovector de tolerancia.

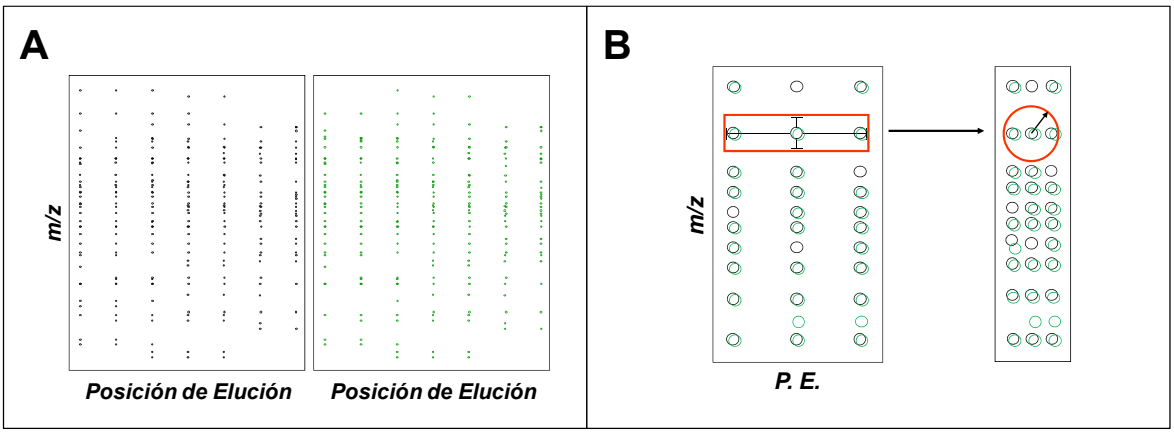


Figura 7: Método de búsqueda de señales solapantes en el software MSHandler. (A) Cada péptido se identifica en un sistema de dos coordenadas correspondientes a su masa y posición de elución. Los puntos negros y verdes representan dos peptidomas diferentes cuyo grado de coincidencia se va a determinar. (B) Con la normalización del valor de los intervalos de confianza mediante la corrección de la escala de los ejes se pasa de un sistema en que se debe evaluar si existe solapamiento en la dimensión de masa y en la de posición de elución independientemente (panel izquierdo), a un sistema que solo depende de la distancia euclídea entre las dos señales y no de su posición relativa en el plano en que se distribuyen (panel derecho), resumiéndose de esta forma una etapa de cálculo. El panel derecho muestra la distribución de m/z frente a posición de elución después del cambio de escala.

Cuando varias asignaciones son posibles debido a la existencia de solapamiento entre más de dos señales de masas, el problema se traslada a un árbol de decisión que se resuelve asignando las señales con relación de intensidad más próxima a la unidad (Figura 8).

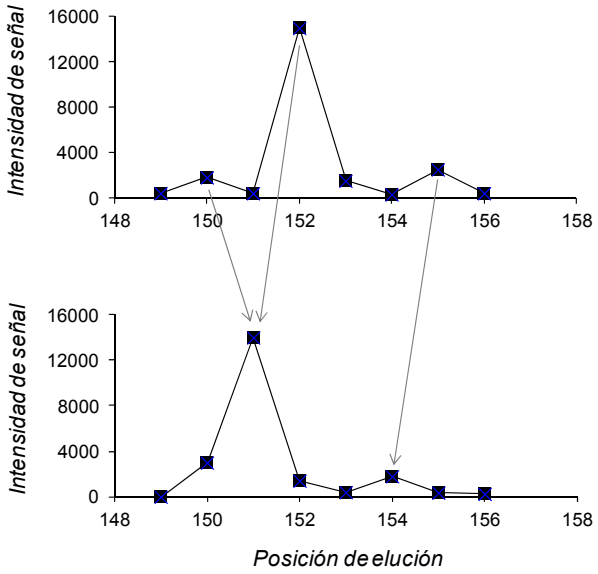


Figura 8: Sistema de restricción de asignaciones. En el perfil de elución que se presenta, la señal en la posición 151 del peptidoma del gráfico inferior puede asignarse a las señales en posiciones 150 y 152 del peptidoma del gráfico superior. Para resolver esta ambigüedad se opta por la asignación de las señales cuya relación de intensidad sea más próxima a la unidad, en este caso se asignaría a la señal de la fracción 152.

La detección de péptidos isotópicamente marcados se lleva a cabo mediante la comparación de los peptidomas de los controles sin marcaje, marcado con el aminoácido-$^{15}N_2$ y con dicho marcaje en presencia de inhibidor (Figura 9). El sistema calcula la relación $^{15}N/^{14}N$ en todas la señales equivalentes y se establece que el péptido incorpora el marcaje cuando esta relación supera un valor definido por el usuario. Se pueden detectar péptidos marcados con hasta tres residuos. Por encima de esta cantidad, las intensidades de las señales isotópicamente relevantes se encuentran predominantemente al nivel del ruido. Una vez identificadas las señales correspondientes a péptidos marcados se recurre a los datos del control con inhibidor para determinar la resistencia a inhibidores. Para ello se calcula el cociente $[(^{15}N+inh)-^{14}N]/(^{15}N-^{14}N)$, que da cuenta de la proporción de marcaje que se incorpora en presencia del inhibidor en relación a la incorporación en ausencia de éste, en función de cuyo valor se determina la susceptibilidad al inhibidor. El usuario puede ajustar el sistema de clasificación para su adecuación a los datos de marcaje observados, para lo cual se establecen dos umbrales del valor $[(^{15}N+inh)-^{14}N]/(^{15}N-^{14}N)$, el primero supone el máximo valor que determina la clasificación de un péptido como sensible al inhibidor, y el segundo el mínimo para la clasificación de un péptido como resistente al inhibidor. Los péptidos marcados cuyo valor sea intermedio entre ambos umbrales no se clasificarán, y en su registro en el informe de resultados aparecerán asignados como no determinados. El correcto ajuste del sistema implica, por una parte, que en los péptidos clasificados como sensibles al inhibidor la

distribución isotópica sea similar en los controles sin marcaje y con marcaje e inhibidor, y por otra, que en los péptidos clasificados como resistentes el incremento de intensidad de la señal correspondiente al péptido marcado sea perceptible en presencia del inhibidor.

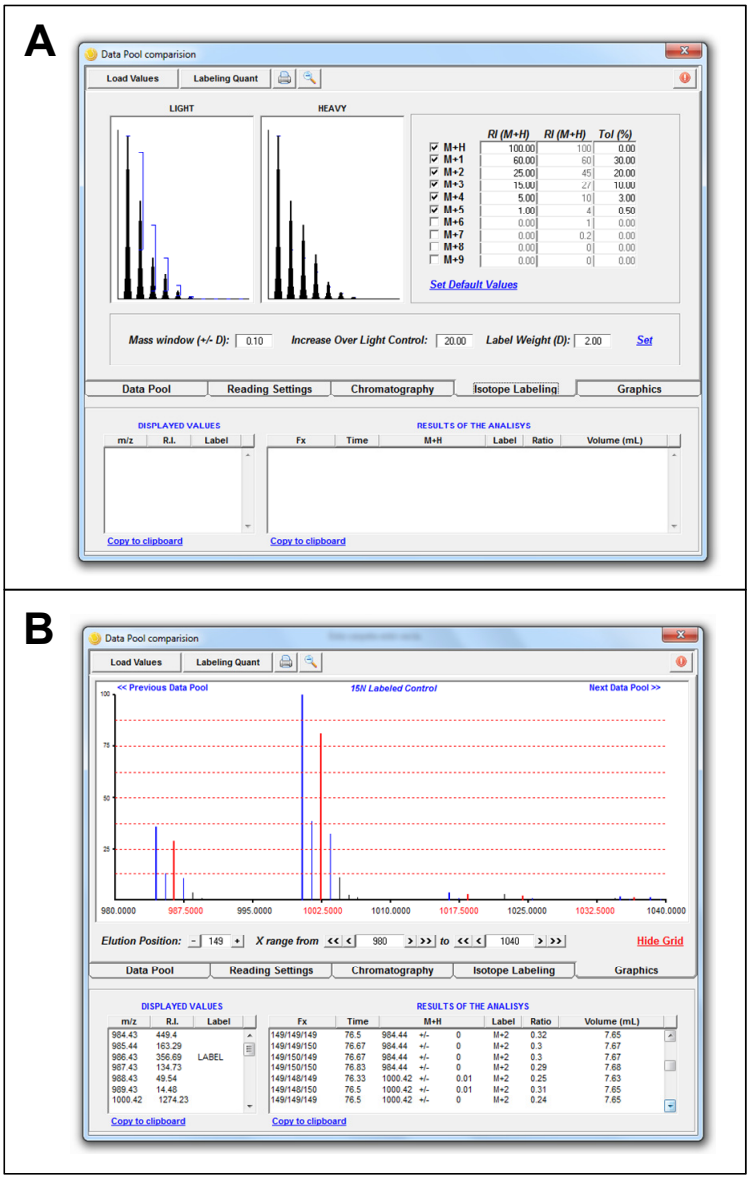


Figura 9: Solución incorporada en el software MSHandler para la identificación de marcajes isotópicos. (A) En la pestaña *Isotope Labeling* el usuario define los parámetros que determinan el marcaje: patrón de distribución de intensidades en la envoltura isotópica, incremento mínimo de intensidad de la señal isotópica, variación de masa del residuo marcado. (B) En la pestaña *Graphics* se visualiza una reconstrucción de los espectros a partir de los datos de m/z e intensidad. Las señales monoisotópicas correspondientes a péptidos marcados se presentan en azul intenso, las señales de las especies que incorporan el residuo marcado se presentan en rojo y el resto de señales pertenecientes a la envoltura isotópica en azul suave. En la Tabla del panel inferior se presentan los resultados de la cuantificación del marcaje.

3.13.2. Interpretación *de novo* de espectros de fragmentación.

La identificación de péptidos cuya fragmentación no se pudo resolver con los motores de búsqueda descritos, se llevó a cabo mediante un sistema basado en un algoritmo genético e implementado en MatLab 7 (Mathworks, Natick, MA), que busca combinaciones aleatorias de residuos compatibles con los datos espectrales, al que se llamó MSgene. La Figura 10 muestra un resumen del funcionamiento de este sistema.

Los datos espectrales se pasan al programa en archivos *.*dta*. El procedimiento se inicia con la generación aleatoria de una población de 100 secuencias binarias en las que se codifican cadenas de aminoácidos. Para la codificación de los 20 residuos proteicos son necesarios grupos de 5 *bits*, quedando 12 codificaciones sin sentido, que proporcionan variabilidad en el tamaño de los productos decodificados. Las secuencias presentan una longitud constante y contienen tantos quintetos como el número máximo de residuos que podría contener una secuencia peptídica de masa igual a la del precursor. La presencia de un quinteto no codificante no implica la interrupción de la decodificación, puesto que el siguiente quinteto codificante será traducido, incorporando el nuevo residuo a continuación del último aminoácido. Este es el mecanismo por el cual se introduce variabilidad en la longitud de las secuencias peptídicas codificadas en secuencias binarias de longitud constante.

Cada una de las secuencias se somete a una evaluación consistente en determinar cómo se ajustaría al espectro de fragmentación problema. Para ello se genera la fragmentación teórica de la secuencia decodificada, obteniendo los iones de las series principales, *b* e *y* y sus modificaciones más frecuentes, descarboxilación, hidratación y desaminación, así como los iones imonio. Se comparan los iones encontrados en la fragmentación teórica con los iones del espectro problema, y se calcula un parámetro de ajuste, el *fitness*, que se define como el cociente entre la cantidad de iones de la fragmentación teórica asignados a iones del espectro problema y la cantidad de iones en el espectro problema.

A continuación se aplica a la población el operador *crowding*, que ordena las secuencias atendiendo a su *fitness* y descarta el 50 % de la población con *fitness* más bajo. La población se amplifica hasta el tamaño inicial por medio del operador *crossover*, consistente en la recombinación de las secuencias organizadas aleatoriamente por parejas. Esta recombinación tiene lugar aleatoriamente en un único punto (*chiasma*). Por último se aplica el operador *mutación*, mediante el cual cada *bit* de la secuencia de cada *cromosoma* puede cambiar su valor con una probabilidad determinada por la tasa de mutación, que es definida por el usuario.

El proceso se repite un número de iteraciones definido por el usuario, acumulándose

información de secuencia que encaja con el espectro problema. En cada iteración se reservan las secuencias de masa coincidente con la del precursor y posteriormente se presentarán como posibles soluciones.

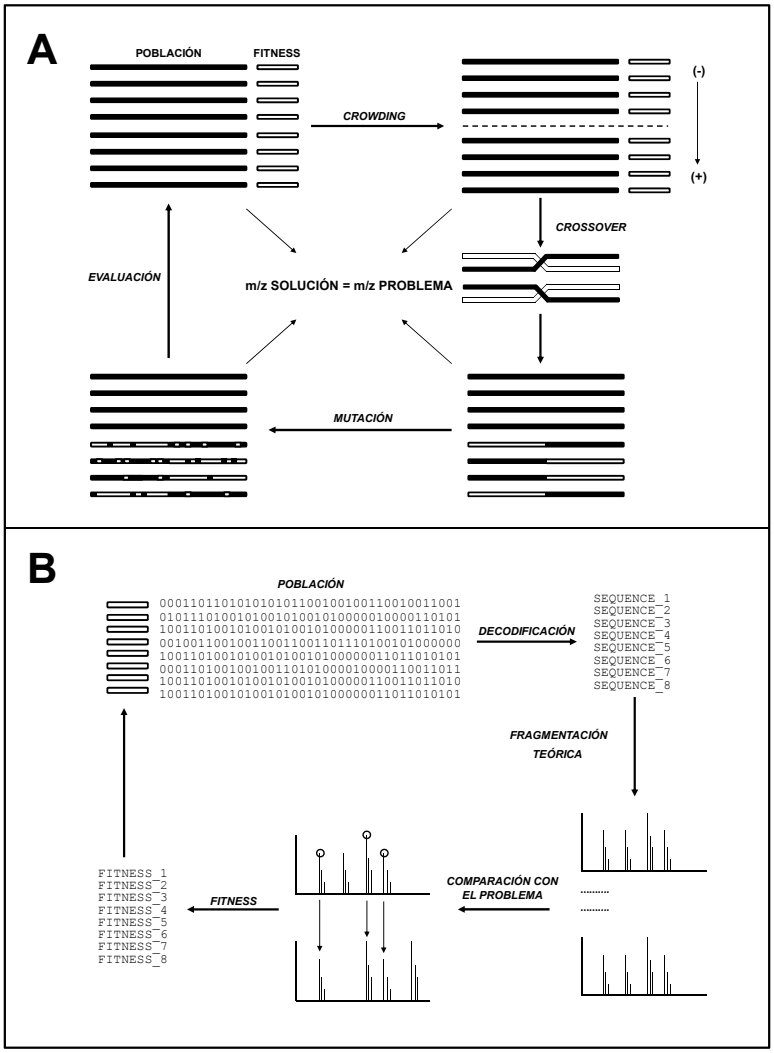


Figura 10: Algoritmo de secuenciación. (A) En cada iteración la población de soluciones se reduce hasta el 50% por eliminación de las secuencias con peores *fitness*. Posteriormente se amplifica la población hasta el tamaño inicial con secuencias procedentes de la recombinación de las que han superado la selección, a las que se aplica el operador mutación y una función de evaluación. Las secuencias con masa igual a la del problema se almacenan como posibles soluciones. (B) Función de evaluación: las secuencias binarias se traducen a secuencias de aminoácidos y se someten a una fragmentación teórica. El *fitness* se calcula en base a la asignación de iones teóricos a iones del espectro problema.

Este procedimiento presenta una serie de ventajas sobre los motores de búsqueda para la identificación de péptidos de HLA-I: 1) la interpretación es *de novo*, con lo que el campo de búsqueda no se reduce a las secuencias incluidas en una base de datos, 2) no se exige que las soluciones encajen con un patrón de digestión específico, 3) la tolerancia a la ausencia de iones representantes de algunos enlaces peptídicos es alta, 4) el sistema de puntuación de las soluciones propuestas depende únicamente de cómo encaja la secuencia candidata con el espectro problema y no de la identificación de otros péptidos procedentes de la misma proteína.

3.13.3. Búsqueda de patrones de secuencia en librerías de péptidos.

La determinación del motivo de unión a un alotipo de HLA-I se llevó a cabo mediante el análisis conjunto de las secuencias de los péptidos identificados. A este efecto se desarrolló una herramienta informática basada en el análisis estadístico de la frecuencia de uso de residuos en cada posición, a la que se llamó MotifSearcher (Figura 11). Ésta es una aplicación para sistemas MS Windows 2000 y versiones posteriores, desarrollada en Visual Basic 6.0 mediante el entorno Visual Studio (Microsoft Corporation). En esta herramienta se implementaron tres soluciones que abordan diferentes aspectos fundamentales para la descripción de patrones de secuencia. La primera solución lleva a cabo un contaje de cada uno de los residuos que aparecen en cada posición y calcula su frecuencia relativa y un parámetro denominado DMP (*Deviation from Mean in the Proteome*), que indica la proporción entre la frecuencia observada de un residuo y la frecuencia que correspondería por su abundancia en el proteoma. Este análisis se puede aplicar a subconjuntos de péptidos restringidos por tamaño o a la serie completa. La segunda solución compara la abundancia de cada residuo en una posición con su abundancia en el proteoma mediante la prueba χ^2, aplicando opcionalmente la corrección de Bonferroni, para buscar residuos significativamente preferentes o desfavorecidos. La tercera solución permite realizar agrupamientos de los residuos de uso preferente atendiendo a sus propiedades químicas o por rangos de tamaño.

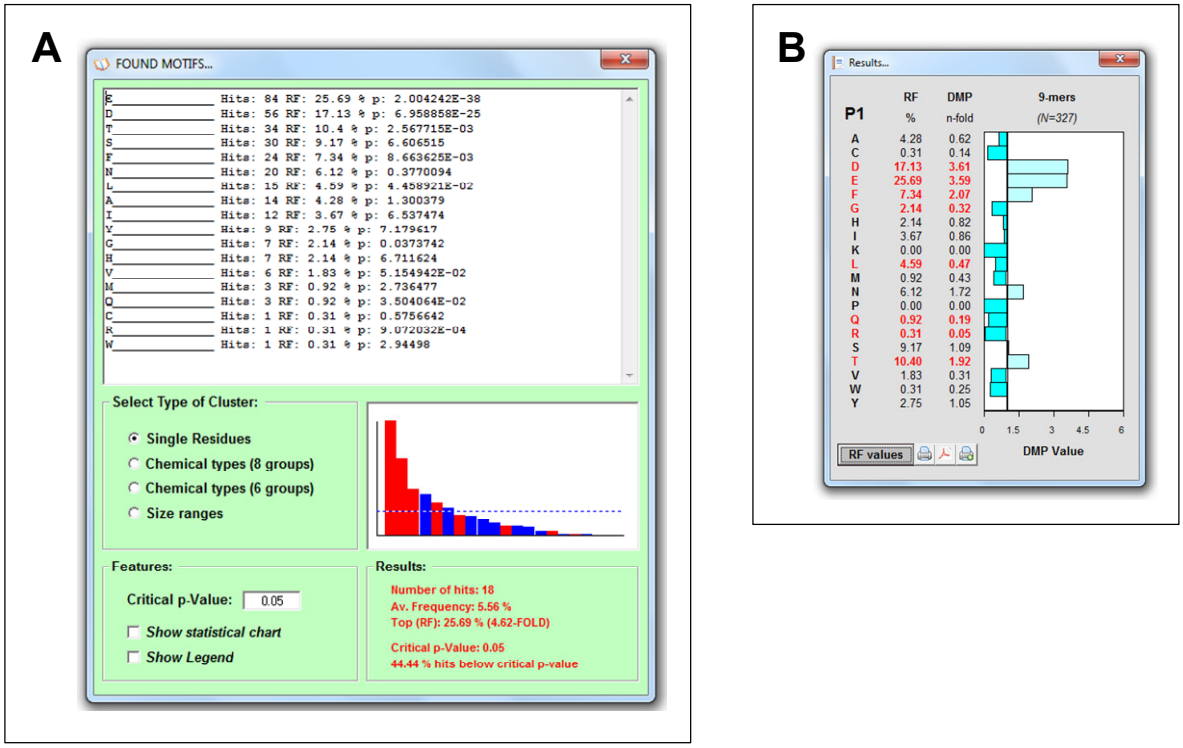


Figura 11: Formularios de resultados del software MotifSearcher. (A) Motivos encontrados en una serie de péptidos. En el panel superior del formulario se muestran los motivos encontrados, número de péptidos que responden a cada motivo, frecuencia relativa y un valor *p* que indica si existen diferencias significativas en la frecuencia de los residuos con respecto a su abundancia en el proteoma. Las barras rojas del gráfico en el panel central derecho representan los valores de frecuencia con diferencias significativas con el proteoma, y las azules las frecuencias sin diferencias significativas. El valor *p* crítico para la determinación de las diferencias significativas es establecido por el usuario en el campo designado en el panel inferior izquierdo. La aplicación de la corrección de Bonferroni es opcional y debe ser activada por el usuario antes del análisis. En el panel central izquierdo se puede seleccionar el tipo de agrupamiento de los residuos, proporcionando cuatro opciones: sin agrupamiento, por características químicas con seis tipos de clasificación (ácidos: D, E; básicos: H, K, R; polares: N, Q, S, T; alifáticos: A, C, I, L, M, V; aromáticos: F, W, Y; otros: G, P), por características químicas con ocho tipos de clasificación (ácidos: D, E; básicos: H, K, R; polares alifáticos: N, Q, S, T; polares aromáticos: Y; apolares alifáticos: A, C, I, L, M, V; apolares aromáticos: F, W; sin cadena lateral: G; Iminoácidos: P) y por tamaño. (B) Frecuencia relativa de residuos (RF) y valor DMP en posiciones individuales. Los valores en rojo indican que existe diferencia significativa con respecto a la abundancia de los residuos en el proteoma. Se puede visualizar el histograma de valores de frecuencia o de DMP opcionalmente.

3.13.4. Asignación de productos de digestión de precursores sintéticos.

El software *Digestor* (Figura 12) mencionado en el apartado 3.7.3 es una aplicación para uso exclusivo en sistemas Windows, desarrollada con Visual Studio 6.0, que tiene como datos de entrada la secuencia del sustrato peptídico sometido a digestión y la lista de masas procedente del análisis por espectrometría de masas del digerido. Esta herramienta asigna las especies iónicas a fragmentos de la secuencia del precursor, u opcionalmente, a modificaciones químicas (oxidaciones, aductos, etc) de estos fragmentos, coincidentes en masa dentro de un intervalo de confianza definido por el usuario. La

búsqueda puede también restringirse a fragmentos procedentes de actividad hidrolítica tríptica, quimotríptica o caspasa.

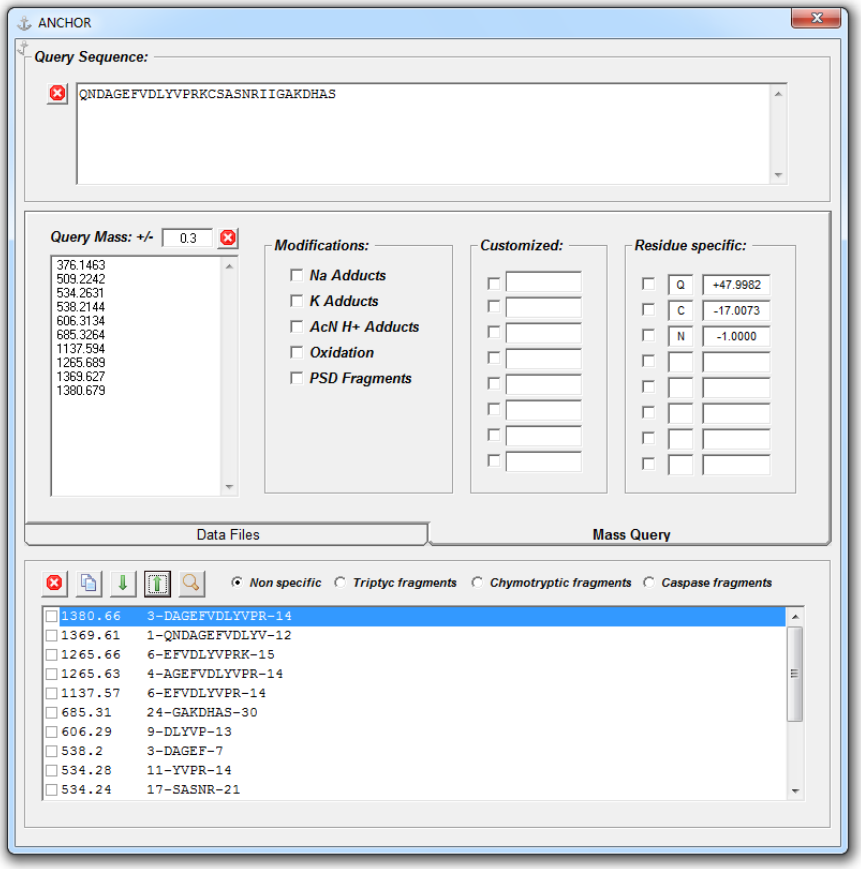


Figura 12: Formulario principal de la aplicación *Digestor*. El usuario introduce la secuencia del sustrato peptídico y una lista de masas de iones procedente del análisis por espectrometría de masas de MALDI-TOF del digerido. Es posible especificar la búsqueda de masas correspondientes con varias modificaciones propuestas por la herramienta u otras definidas por el usuario, incluso se puede definir modificaciones específicas de aminoácido. En la ventana inferior se presenta una lista con todas las asignaciones, donde se clasifica a las especies asignadas según el tipo de actividad hidrolítica que las produce, incluyendo un grupo de clasificación para las procedentes de una actividad inespecífica.

Resultados I

4.1. Peptidoma de HLA-A*68:01 resistente a inhibidores de proteasoma.

La primera parte de esta Tesis abordó el estudio de la naturaleza y el origen de los peptidomas de HLA-I resistentes a inhibidores de proteasoma. La cuestión planteada fue si dicho peptidoma es realmente independiente de proteasoma o es consecuencia de una alteración de la especificidad proteasómica resultante de su inhibición parcial.

4.1.1. Motivo de unión a HLA-A*68:01

Se obtuvieron 816 secuencias, con longitudes entre 6 y 16 residuos, procedentes de tres fuentes diferentes: 35 secuencias publicadas en trabajos previos, disponibles en la base de datos de ligandos de MHC SYFPEITHI (www.syfpeithi.de), 772 secuencias obtenidas por fragmentación MS/MS en electrospray-Orbitrap del peptidoma aislado del transfectante C1R-A*68:01 y 70 secuencias obtenidas por fragmentación por MALDI-TOF/TOF del mismo peptidoma (Apéndice 1, Tabla 1S). La longitud más frecuente de los péptidos fue de 9 residuos, seguido por péptidos de 10, 11 y 12 residuos, por orden decreciente, mientras que péptidos fuera de este rango de longitud eran poco frecuentes (Figura 13).

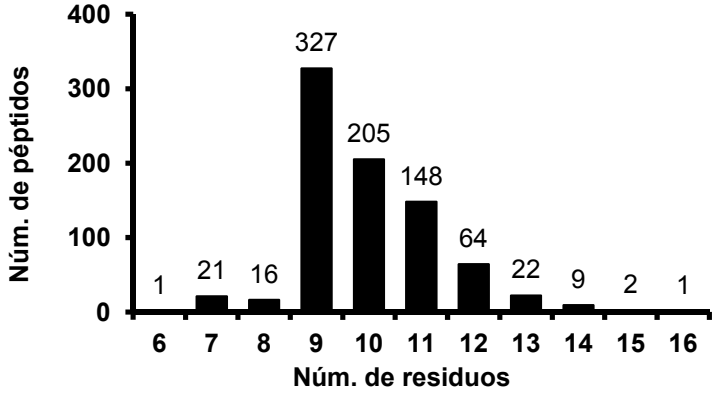


Figura 13: Distribución por longitud de los 816 péptidos de HLA-A68 identificados. En la parte superior de cada barra se muestra el número de péptidos de la longitud indicada.

El motivo de unión se determinó por análisis de la frecuencia de uso de residuos en las 778 secuencias de longitud superior a 8 residuos. Se llevaron a cabo dos búsquedas diferentes empleando el software MotifSearcher. La primera se realizó simultáneamente sobre las 778 secuencias de referencia para las posiciones de anclaje P1, P2, P3, PC-2 y PC, que podían ser adecuadamente alineadas entre péptidos de diferentes longitudes. Para la segunda búsqueda se empleó la opción que permite la restricción por longitud en el MotifSearcher, para analizar por separado nonámeros, decámeros, undecámeros y dodecámeros, evaluando todas las posiciones peptídicas. Los péptidos de longitud

inferior a nueve residuos se analizaron por separado. Para la determinación de la significación estadística se aplicó la corrección de Bonferroni y se estableció un valor *p* crítico de 0.05.

El análisis de los péptidos de longitud superior a 9 residuos reveló una importante selectividad en la unión a HLA-A68 (Figura 14). En P1 se encontraron predominantemente residuos ácidos, con una frecuencia del 50.6%, seguido por T y S. En P2 se encontraron residuos de diferente naturaleza química, pero un 84.4% de ellos con una masa incremental en el rango de 71-103 Da, de lo cual se deduce una restricción prominente por tamaño. En P3 predominaban residuos alifáticos y F. En PC-2 se observó una alta frecuencia de residuos no cargados, con un 54.4% de alifáticos y un 26.1% de polares sin carga. K y R suponían el 97.7% de los residuos en PC. El 2.3% restante carecían del motivo de unión a HLA-C*04 (Apéndice 1, Tabla 1S) que representa una fuente potencial de péptidos contaminantes en este experimento.

La diferencia en la longitud de los péptidos no implicó cambios significativos en el uso de residuos en las posiciones analizadas (Apéndice 1, Figura 2S). En las posiciones no correspondientes a sitios de anclaje no se encontró restricción (Apéndice 1, Tabla 2S).

En los péptidos de longitud inferior a 9 residuos (un hexámero, 21 heptámeros y 16 octámeros) se encontraron diferencias en el uso de residuos en P1 y P3, pero no en PC-2 y PC (Apéndice 1, Figura 3S).

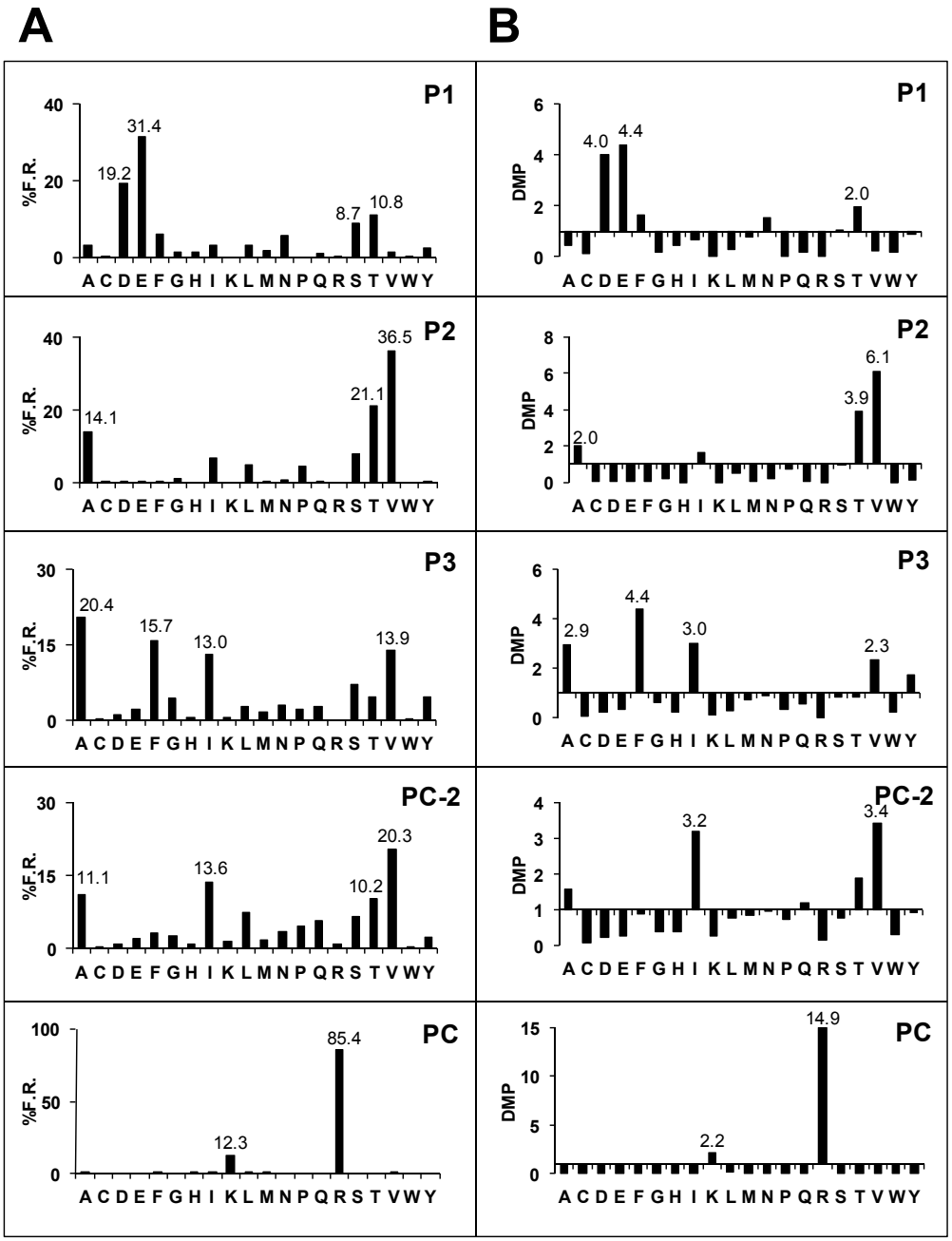


Figura 14: Frecuencia de uso de residuos (A) y DMP (B) en las posiciones P1, P2, P3, PC-2 y PC de 778 secuencias de péptidos de HLA-A68 de longitud superior a 8 residuos. Se indican los valores de frecuencia relativa (F.R.) y DMP de los residuos predominantes en cada posición.

4.1.2. Proteínas precursoras de los péptidos de HLA-A*68:01.

Los 816 péptidos identificados procedían de 672 proteínas diferentes, entre las que se observó una abundancia de proteínas ácidas de masa molecular superior a 30 kDa ligeramente mayor que en el proteoma (49.7% frente a 38.9%), mientras que las proteínas ácidas de masa inferior están ligeramente subrepresentadas, con frecuencias de 7.9% y 13.1%, respectivamente (Figura 15A y D). La distribución subcelular reveló también algunas diferencias con el proteoma, fundamentalmente por una menor frecuencia de proteínas de membrana (Figura 15E y H). Estos resultados sugieren una importante influencia de la disponibilidad de sustratos del proteasoma en la configuración del peptidoma de HLA-A68, donde los componentes de la ruta secretoria tienden a estar subrepresentados.

La comparación con 471 proteínas precursoras de ligandos de HLA-B*27:05 publicadas previamente (Ben Dror et al., 2010) reveló un patrón de distribución de pI/MW y localización subcelular muy similar en ambos alotipos (Figura 15B y F). Por otra parte, se encontraron 74 proteínas que eran fuente común de péptidos de HLA-A68 y B27, lo que constituye el 11.0% y 15.7% de las proteínas parentales identificadas, respectivamente. Este valor es casi cuatro veces superior al esperado por azar, lo que sugiere que las proteínas precursoras de los ligandos de ambos alotipos proceden predominantemente de un subconjunto del proteoma. La distribución de pI/Mw de las proteínas precursoras de péptidos comunes a HLA-A68 y B27 es similar a la de los conjuntos totales pero con una menor abundancia de proteínas ácidas de baja masa molecular (Figura 15C y G).

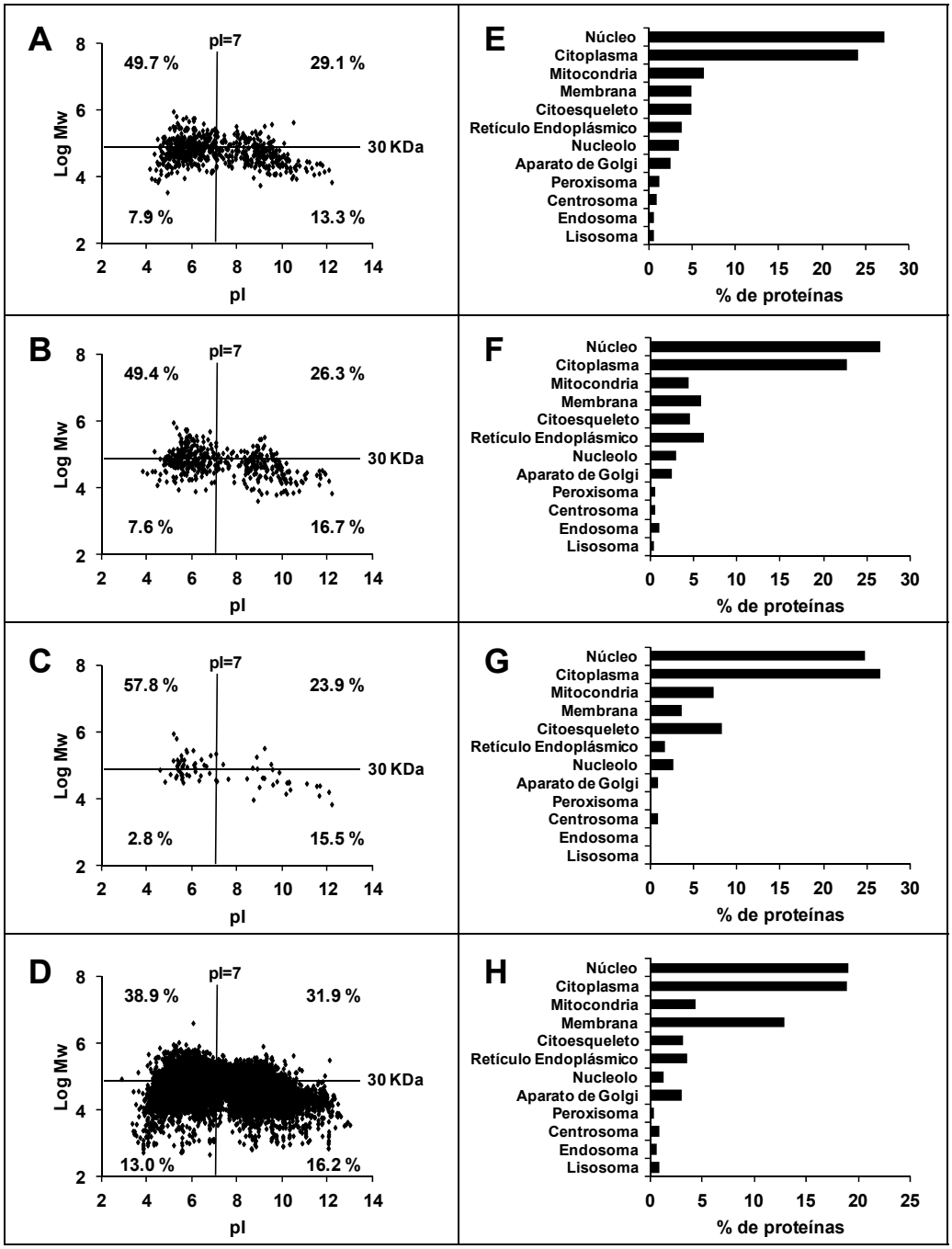


Figura 15: Distribución de pI/Mw y localización subcelular de las proteínas precursoras de ligandos de HLA-A*68:01 (A, E), B*27:05 (B, F), ligandos de ambos subtipos (C, G) y el proteoma humano completo (D, H).

4.1.3. Susceptibilidad de los ligandos de HLA-A*68:01 a la inhibición del proteasoma.

4.1.3.1. Incorporación de ligandos en presencia de inhibidores de proteasoma.

La susceptibilidad de los ligandos a la inhibición del proteasoma se determinó por marcaje isotópico con aminoácidos pesados en presencia de inhibidores. Después de estos tratamientos se aislaron los peptidomas para su fraccionamiento cromatográfico y posterior análisis por espectrometría de masas MALDI-TOF. Los datos se evaluaron mediante el software MSHandler empleando los siguientes parámetros: intensidad mínima de señal 100, rango de masas 800-2000 Da, intervalo de confianza para determinar masas equivalentes, ±0.2 Da y ±1 posición de elución, incremento de intensidad de señal relativa a la monoisotópica en el marcaje 20%.

En un primer grupo de experimentos se llevó a cabo el marcaje con ^{15}N-K o ^{15}N-R en presencia de MG-132 20 μM. Se identificaron 132 péptidos marcados (Apéndice 1, Tabla 3S). En 95 de éstos (71%) el cociente [(^{15}N+inh)-^{14}N]/(^{15}N-^{14}N) (relación de marcaje) fue inferior a 0.35, valor que se estableció como límite superior para determinar sensibilidad al inhibidor. Este valor relativamente alto de la relación de marcaje, que implica que péptidos con hasta 35% de incremento de la intensidad de señal en presencia del inhibidor se consideren sensibles, se estableció para incluir la posibilidad de incorporación de isótopos debido a una actividad proteasómica residual. Un total de 24 péptidos (19%) presentaron una relación de marcaje superior a 0.6, estableciéndose este valor como límite inferior para los péptidos resistentes. Este valor implica que péptidos que hubieran perdido hasta un 40% del marcaje en presencia del inhibidor se consideran resistentes, para incluir péptidos que pudiesen haber disminuido su expresión por efectos indirectos de la droga, tales como la inhibición de la síntesis proteica. Cuatro péptidos con valores superiores a 0.6 no se asignaron como resistentes por presentar mayor susceptibilidad a epoxomicina. Los 9 péptidos restantes presentaron valores intermedios, de forma que un total de 13 péptidos (10%) no se pudieron clasificar como sensibles o resistentes.

En un segundo experimento se llevó a cabo el marcaje con ^{15}N-R en presencia de epoxomicina 2.5 μM, encontrándose 64 péptidos marcados. De éstos, 46 (73%) fueron sensibles al inhibidor, 8 (13%) resistentes y 9 (14%) no pudieron ser clasificados por presentar relaciones de marcaje intermedias o menor inhibición con MG-132 (Apéndice 1, Tabla 4S).

De los 42 péptidos marcados en los experimentos con ambos inhibidores, 31 (74%) fueron sensibles (Tabla 1), 5 (12%) resistentes (Tabla 2) y 6 (14%) no clasificados (Tabla 3). Cuatro de estos últimos mostraron una inhibición sensiblemente mayor con epoxomicina. Estos resultados indican que ambos inhibidores tienen un efecto similar sobre la mayoría de los ligandos de HLA-A68. Sin embargo, el efecto inhibitorio de la epoxomicina fue mayor a juzgar por el porcentaje algo menor de péptidos

resistentes a este inhibidor y por su efecto más drástico sobre algunos péptidos (Tabla 3).

Se determinó la secuencia de 33 péptidos marcados, 24 de ellos sensibles a ambos inhibidores, 7 resistentes y 2 con diferente susceptibilidad a cada inhibidor (Tablas 1-3). No se encontraron diferencias en tamaño, motivo de secuencia o en residuos de la región flanqueante N-terminal entre péptidos sensibles y resistentes.

Sensibles (N=31)		MG-132		Epoxomicina		Secuencia
Fracción	M+H+	Residuos	Ratio	Residuos	Ratio	
92	969.42	R	0.02	R	0.20	
125 / 124	986.43	R	0.19	R	-0.08	
125	991.41	R	0.09	R	0.23	
110	992.42	R	0.15	R	0.10	NAIGGKYNR
111	992.60	K	-0.18			
134 / 135	1000.48	R	-0.13	R	0.29	
112 / 114	1003.41	R	0.11	R	-0.03	DTAAQITQR
120 / 118	1010.51	R	-0.03	R	0.00	
108 / 109	1012.38	R	0.27	R	-0.01	TSETPQPPR
155	1021.51	R	0.27	R	0.18	DVFVVGTER
113 / 114	1025.53	R	-0.13	R	0.06	
144 / 144	1032.47	R	0.04	R	0.48	FPSEIVGKR
145	1032.56	K	0.19			
132	1043.60	R	0.18	R	0.09	ELTAVVQKR
132	1043.60	K	0.15			
152 / 153	1060.57	R	0.01	R	0.08	DVFRDPALK
152	1060.58	K	-0.09			
145 / 146	1070.49	R	0.22	R	0.32	EVAQLIQGGR
126	1075.63	2 R	0.18	2 R	0.09	HPAVVIRQR
129	1077.42	R	-0.06	R	0.20	ETNYGIPQR
150 /152	1089.64	R	0.01	R	0.30	
101 / 99	1108.50	R	0.12	R	0.12	
123 / 124	1122.49	2 R	0.35	2 R	0.11	
204 / 205	1123.63	R	0.14	R	0.13	
158 / 159	1128.57	R	0.12	R	0.02	EVFSHILKR
158	1128.64	K	0.08			
181 / 182	1137.54	R	0.22	R	0.33	EFVDLYVPR
129 / 130	1139.54	2 R	0.19	2 R	0.07	DPFHKAIRR
130	1139.74	K	0.33			
104	1140.5	3 R	0.25	3 R	0.13	EVNDRPVRR
125 / 124	1146.43	2 R	-0.05	2 R	0.06	
114 / 113	1155.67	R	0.21	R	0.26	TVKKGKPELR
114	1155.80	3 K	0.04			
154	1212.52	R	-1.14	R	0.27	FPHNPQFIGR
153	1245.71	R	0.08	R	0.25	
153	1245.72	K	0.13			
126	1246.62	R	0.00	R	0.03	TAADTAAQITQR
152 / 154	1376.76	2 R	0.13	2 R	0.22	NSFRYNGLIHR
142 / 144	1608.91	3 R	0.28	3 R	0.28	

Tabla 1: Ligandos sensibles a MG-132 y a epoxomicina.

Resistentes (N=5)		MG-132		Epoxomicina		Secuencia
Fracción	M+H+	Residuos	Ratio	Residuos	Ratio	
167	1048.58	R	0.62	R	0.64	SVQGIIIYR
127	1200.52	R	1.13	R	0.67	
201 / 203	1207.73	R	1.76	R	0.66	SAFATPFLVVR
129	1209.55	2 R	0.83	2 R	0.73	ETVQLRNPPR
136	1462.75	R	2.93	R	0.82	TVKDVNQQEFVR
136	1462.73	K	2.02			

Tabla 2: Ligandos resistentes a MG-132 y a epoxomicina.

No asignados (N=6)		MG-132		Epoxomicina		Secuencia
Fracción	M+H+	Residuos	Ratio	Residuos	Ratio	
130	1023.50	2 R	0.66	2 R	0.53	
153 / 154	1103.54	2 R	0.51	2 R	0.53	
173 / 174	1132.52	R	1.10	R	0.38	
130	1187.53	R	0.70	R	0.37	NVAEVDKVTGR
130	1187.74	K	0.48			
144	1250.57	2 R	0.51	2 R	0.21	
128 / 127	1443.68	R	0.74	R	0.35	TVINQTQKENLR
127	1443.88	K	0.47			

Tabla 3: Ligandos de susceptibilidad indeterminada a MG-132 y a epoxomicina.

La intensidad media de las señales de masas correspondientes a péptidos resistentes a inhibidor fue inferior a la de los sensibles, lo cual descarta la posibilidad de una asignación artefactual de péptidos resistentes por ser especialmente abundantes. La presencia del inhibidor afectó por igual a péptidos sensibles y resistentes, disminuyendo la intensidad media de señal, lo cual sugiere un efecto global de estos inhibidores en la expresión de proteínas (Tabla 4).

	MG-132			Epoxomicina		
	^{14}N	^{15}N	^{15}N+Inh.	^{14}N	^{15}N	^{15}N+Inh.
Sensibles (N=31)	13461	11092	7026	3001	2489	1644
Resistentes (N=5)	2818	2364	1657	551	356	176
Ratio	4.8	4.7	4.2	5.4	7	9.3

Tabla 4: Intensidad media de las señales correspondientes a péptidos de HLA-A*68:01 sensibles y resistentes a inhibidores de proteasoma en los espectros de masas MALDI-TOF.

4.1.3.2. Proteínas precursoras de péptidos sensibles y resistentes a inhibidores de proteasoma.

La Figura 16 muestra la distribución subcelular y de pI/Mw de las proteínas precursoras de los péptidos marcados cuya secuencia fue determinada. La mayor parte de los péptidos identificados correspondían a secuencias internas de las proteínas de las que derivaban, con la excepción de dos péptidos resistentes, ETVQLRNPPR (Tabla 2) y DPAGVHPPR (Apéndice 1, Tabla 4S), localizados respectivamente en el extremo N- y C-terminal de sus precursores.

Los 5 péptidos resistentes a los inhibidores que derivaban de secuencias internas de sus precursores procedían de proteínas básicas de masa inferior a 30 kDa, mientras que los dos péptidos localizados en los extremos de sus precursores procedían de proteínas que no seguían esta norma. En el caso de los péptidos sensibles a inhibidores solamente un 54.5% procedían de proteínas básicas de masa inferior a 30 kDa. Estos resultados indican que, análogamente a la situación en HLA-B27 (Marcilla *et al.*, 2007), la mayoría de los ligandos de A68 resistentes a inhibidores de proteasoma proceden de proteínas pequeñas y básicas.

Se encontraron dos proteínas de las que derivaban péptidos con diferente susceptibilidad a inhibidores. La proteína S21 de la subunidad ribosomal 40S es fuente de dos péptidos de HLA-A68, EFVDLYVPR, sensible a inhibidores, y NVAEVDKVTGR, no clasificado (Tabla 1), y de otro de HLA-B27, GRFNGQFKTY, que era resistente (Marcilla *et al.*, 2007). De la proteína S3 de la subunidad ribosomal 40S derivan los péptidos de HLA-A68 ELTAVVQKR y ELYAEKVATR, el primero de ellos sensible y el segundo resistente (Apéndice 1, Tabla 3S).

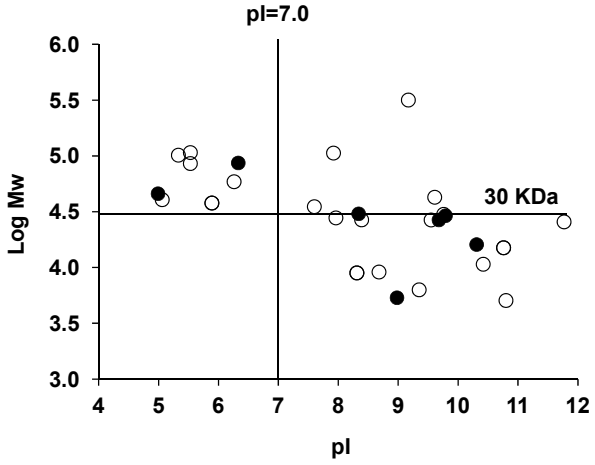


Figura 16: Distribución de pI/Mw de las proteínas precursoras de ligandos de HLA-A*68:01 sensibles (○) y resistentes (●) a inhibidores de proteasoma. Los dos ligandos resistentes situados fuera del cuadrante de pI>7, Mw <30 KDa están localizados en los extremos de las proteínas de las que derivan.

4.1.4. Inhibición selectiva de las especificidades proteolíticas del proteasoma 26S.

Para una observación directa del efecto de los inhibidores en la actividad del proteasoma se llevó a cabo la digestión de sustratos fluorogénicos *in vitro* con proteasoma 26S. La liberación de la aminometil cumarina, componente fluorogénico de estos sustratos, es fácilmente detectable por fluorimetría. La actividad residual del proteasoma se calculó como el cociente entre la curva de progreso de la digestión en presencia y ausencia de inhibidores, y la inhibición como la diferencia hasta el 100% de actividad. El uso de sustratos específicos para las actividades tríptica y quimotríptica permite la monitorización de la eficiencia de cada una de estas especificidades hidrolíticas de manera independiente. Los resultados se compararon con digestiones llevadas a cabo con proteasoma 26S extraído de células previamente tratadas con inhibidores, para determinar si el efecto inhibitorio es diferente cuando el proteasoma se encuentra en el interior celular.

4.1.4.1. Inhibición de la actividad quimotríptica.

La hidrólisis del sustrato quimotríptico se inhibió por encima del 90% tras el tratamiento con epoxomicina 5 μM, tanto en células como en extractos de proteasoma 26S. Por encima de esta concentración se perdió el efecto dosis-respuesta (Figura 17A). A la concentración de 2.5 μM, usada para el marcaje de péptidos, la inhibición del proteasoma no fue la máxima. La saturación del efecto inhibitorio del MG-132 se produjo a una concentración de 10 μM con una inhibición inferior al 80% (Figura 17C).

4.1.4.2. Inhibición de la actividad tríptica.

La saturación de la inhibición de la hidrólisis del sustrato tríptico con epoxomicina se produjo a la concentración de 2.5 μM, con un efecto máximo por debajo del 80% (Figura 17B). El efecto inhibitorio del MG-132, que alcanzó su máximo en 20 μM, no superó el 60% (Figura 17D). En todos los ensayos, el tratamiento de las células con los inhibidores produjo un efecto similar al observado en el tratamiento directo del proteasoma 26S en solución, lo cual descarta que limitaciones en el acceso de la droga al proteasoma en células vivas sean factores determinantes en su efecto.

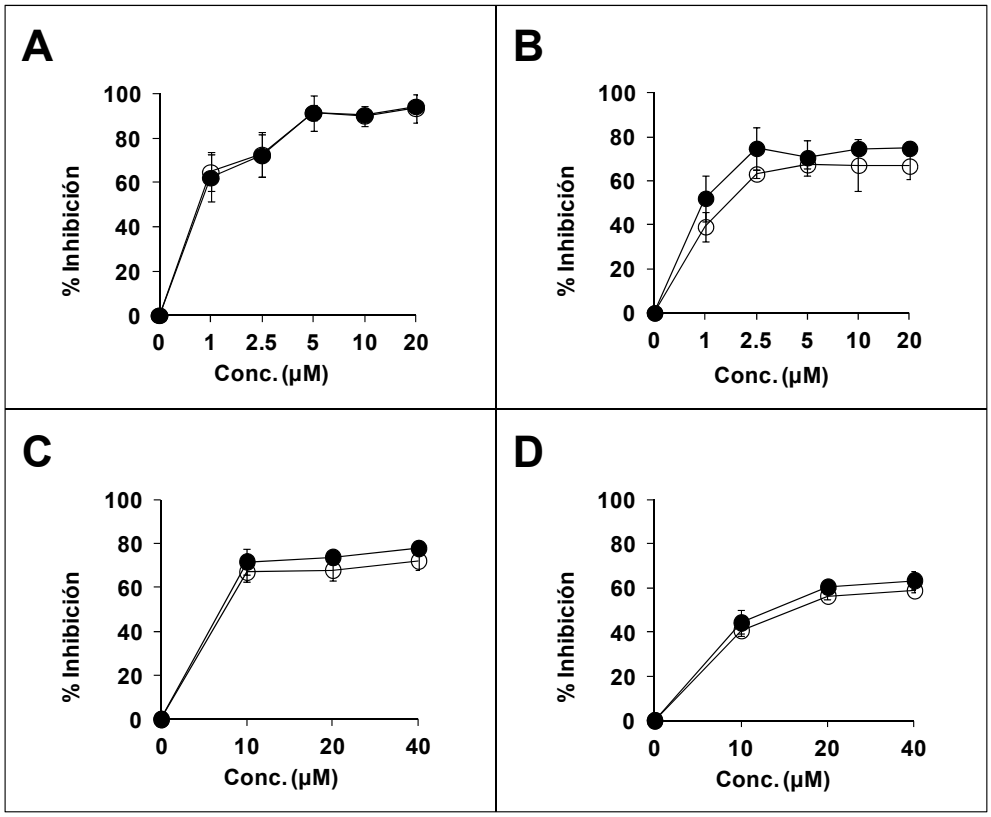


Figura 17: Digestión de sustratos fluorogénicos *in vitro* con proteasoma 26S (●) ó células vivas (○). En los paneles superiores se presentan los datos obtenidos en la digestión del sustrato quimotríptico (A) y tríptico (B) tras el tratamiento con epoxomicina. En los paneles inferiores se presentan los datos obtenidos para el sustrato quimotríptico (C) y tríptico (D) empleando MG-132.

4.1.4.3. Origen de la actividad residual en presencia de inhibidores.

Mediante la separación electroforética de los extractos de proteasoma 26S en gel nativo se detectaron dos bandas con actividad hidrolítica, correspondientes al proteasoma 26S y 20S, respectivamente (Figura 18). No se observó hidrólisis del sustrato quimotríptico tras el tratamiento con cualquiera de los inhibidores (Figura 18C y D). Las bandas de hidrólisis del sustrato tríptico también desaparecieron con el tratamiento con epoxomicina, mientras que con MG-132 se observaron con una intensidad sensiblemente disminuida (Figura 18A y B).

En previsión de la posible presencia de proteasas con carga positiva en los extractos de proteasoma 26S, que tendrían movilidad electroforética inversa en el gel nativo, se desarrolló la electroforesis invirtiendo la polaridad, pero no se detectaron bandas con actividad hidrolítica (datos no mostrados).

En conjunto estos resultados indican que los inhibidores de proteasoma, especialmente MG-132, inhiben solo parcialmente la subespecificidad tríptica, aunque son mucho más efectivos con la

quimotríptica. Los ensayos con geles nativos sugieren que la actividad hidrolítica observada en presencia de inhibidores no se debe a proteasas contaminantes, aunque no se puede descartar formalmente un efecto residual de éstas.

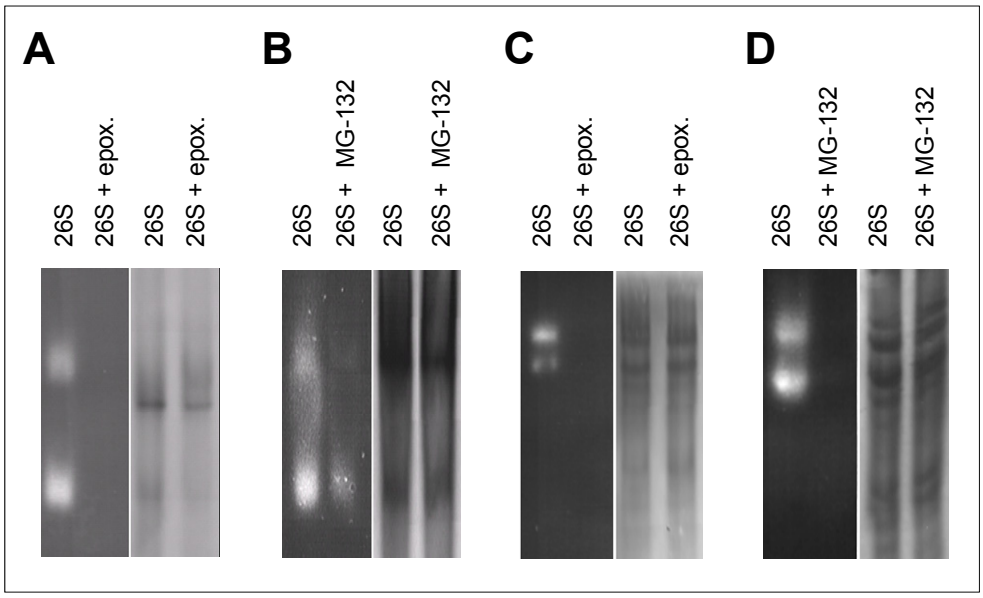


Figura 18: Digestión de sustratos fluorogénicos en gel nativo con proteasoma 26S. En la parte izquierda de cada panel se presenta la vista del gel con transiluminador de luz ultravioleta durante la reacción de hidrólisis y en la parte derecha el revelado con azul Coomassie. La banda superior corresponde al proteasoma 26S y la inferior al núcleo catalítico 20S. El sustrato tríptico se sometió a hidrolisis con proteasoma 26S aislado en gel de poliacrilamida 3.5% después del tratamiento con epoxomicina (A) o MG-132 (B). Para la hidrólisis del sustrato quimotríptico, el proteasoma 26S tratado con epoxomicina (C) o MG-132 (D) se aisló en un gel en gradiente de poliacrilamida 3.5-15.0%.

4.1.5. Efecto de los inhibidores sobre el patrón de corte del proteasoma.

La inhibición incompleta del proteasoma puede alterar sus patrones de corte, resultando en un incremento en la producción de algunos péptidos por bloqueo preferente de sus roturas destructivas, lo que podría explicar su resistencia a inhibidores (Wherry et al., 2006). Para evaluar esta posibilidad, se llevó a cabo la digestión *in vitro* de dos precursores sintéticos con proteasoma 20S. El primero, de 31 residuos, correspondía a la secuencia 29-59 de la proteína S21 de la subunidad ribosomal 40S, e incluía dos secuencias solapantes, NVAEVDKVTGR y GRFNGQFKTY, correspondientes a un péptido de HLA-A68 parcialmente resistente a inhibidores de proteasoma (Tabla 3) y un péptido resistente de B27 (Marcilla *et al.*, 2007), respectivamente (Figura 19A). El segundo precursor, de 30 residuos, correspondía a la secuencia 2-31 de la misma proteína ribosomal y contenía la secuencia EFVDLYVPR, péptido de HLA-A68 sensible a los inhibidores (Tabla 1; Figura 19B).

Los productos de digestión del primer precursor se sometieron a separación cromatográfica y

posterior análisis por espectrometría de masas MALDI-TOF, en condiciones similares a las empleadas para los peptidomas. La abundancia relativa de cada producto de digestión se estimó a partir de dos factores: 1) el cociente entre el área del intervalo del cromatograma en que eluye y el área total del cromatograma y 2) el cociente entre la suma de la intensidad de su señal de masas en todos los espectros en los que aparece dicho producto de digestión y la suma total de la intensidad de todas las señales identificadas en estos espectros.

En ausencia de inhibidores este sustrato fue totalmente degradado y se produjeron cortes en la práctica totalidad de los enlaces peptídicos (Figura 19A). Se recuperaron 80 péptidos con rendimientos superiores al 0.1%, 66 de los cuales requerían cortes duales para su formación (Apéndice 1, Tabla 5S). Aunque se produjeron múltiples cortes en posiciones correspondientes a sus secuencias internas, los epítopos de HLA-A68 (residuos 7-17) y B27 (residuos 16-25) se generaron con rendimientos de 6.4 y 1.0%, respectivamente. Además se encontraron varias especies con extensiones N-terminales de estos ligandos.

En presencia de MG-132 solo se degradó el 28.9% del sustrato. El número de productos de digestión, excluyendo el sustrato no digerido, se redujo a 28 péptidos (Apéndice 1, Tabla 5S) procedentes del corte en 22 posiciones. El ligando de A68 se generó en muy baja cantidad (0.3% del material digerido) y el ligando de B27 no se detectó, pero se encontraron precursores con extensiones N-terminales de ambos (Apéndice 1, Tabla 5S). No obstante, la proporción entre la contribución relativa de cortes destructivos y potencialmente productivos del epítopo se incrementó en presencia del inhibidor con respecto a su ausencia, tanto para el ligando de A68, donde esta proporción pasó de 2.9 a 4.8, como para el de B27, donde pasó de 0.9 a 1.3 (Tabla 5).

Con epoxomicina se digirió el 54.3% del sustrato, y se encontraron 16 productos de digestión procedentes de roturas en 19 posiciones (Apéndice 1, Tabla 5S). No se detectaron ninguno de los ligandos naturales, pero se encontraron algunos precursores con extensiones N-terminales y algunas especies más cortas con el residuo C-terminal de estos epítopos, que pueden proceder de su degradación. La relación destrucción/producción no se alteró en el ligando de A68 (pasó de 2.9 a 3.0) y se incrementó en el de B27 (de 0.9 a 2.7).

Estos resultados indican que: 1) el proteasoma 20S generó ambos ligandos naturales y/o sus extensiones N-terminales, 2) ninguno de los inhibidores bloqueó completamente la actividad proteasómica sobre este sustrato o la generación de especies productivas y 3) la inhibición incompleta del proteasoma no resultó en una sobreproducción de los ligandos o sus precursores.

Ambos inhibidores mostraron efectos globales similares en todas las subespecificidades

proteolíticas del proteasoma, produciendo solamente una ligera alteración de su contribución relativa a la degradación del sustrato (Tabla 7). Sin embargo, el efecto sobre los enlaces individuales fue dependiente del inhibidor empleado (Apéndice 1, Tabla 6S). Por ejemplo, MG-132 inhibió más eficientemente el corte tras K13 que tras R17, mientras que la epoxomicina mostró una mayor inhibición de los cortes tras K13 y R17 que tras K23. En los residuos que constituyen dianas de la actividad quimotríptica, la inhibición de la rotura tras M6 fue más eficiente con epoxomicina, mientras que tras F18 y F22 la inhibición con MG-132 fue mayor. En los residuos no asignados a una actividad específica de corte la inhibición fue, en general, más eficiente con epoxomicina, a excepción de la rotura tras N19 cuya inhibición fue más eficiente con MG-132. En resumen, la inhibición parcial del proteasoma resulta en una alteración de los patrones de corte que es dependiente del inhibidor.

El ensayo llevado a cabo con el segundo sustrato permitió evaluar el efecto de estos inhibidores en un ligando de A68 sensible a inhibidores de proteasoma. En este caso, los productos de digestión se analizaron directamente por espectrometría de masas MALDI-TOF (Tabla 6; Apéndice 1, Tabla 7S). No se encontró el ligando natural, pero en ausencia de inhibidores y con MG-132 se produjo un único precursor N-terminal (residuos 1-14 del sustrato) con rendimientos 16.8% y 9.0%, respectivamente, que no se obtuvo en presencia de epoxomicina (Figura 19B). El rendimiento conjunto de cortes destructivos fue similar en ausencia y en presencia de los inhibidores (Tabla 6). El ligando de A68 sensible a inhibidores en este precursor no se generó en presencia de epoxomicina y mostró una tasa destrucción/producción más baja que el ligando no asignado procedente de la misma proteína en el precursor de 31 residuos, tanto sin inhibidores como con MG-132 (Tabla 5).

En la digestión de este sustrato se observó también una alteración en el patrón de corte del proteasoma en presencia de los inhibidores. MG-132 y epoxomicina tuvieron un efecto global similar, consistente en una mayor inhibición de las actividades tríptica y caspasa que la encontrada para la actividad quimotríptica y otras (Tabla 7). No obstante, se encontraron diferencias dependientes del inhibidor en roturas de algunos enlaces individuales, fundamentalmente tras R14 pero también en otras posiciones (Figura 19B; Apéndice 1, Tabla 7S).

Aunque la extrapolación de los resultados obtenidos *in vitro* al procesamiento proteasómico *in vivo* debe hacerse con precaución, las principales diferencias entre el ligando de A68 no asignado y el sensible a inhibidores fueron que MG-132 tenía un mayor efecto inhibitorio en la generación de especies productivas del ligando sensible, y que la epoxomicina bloqueó completamente la producción de éste.

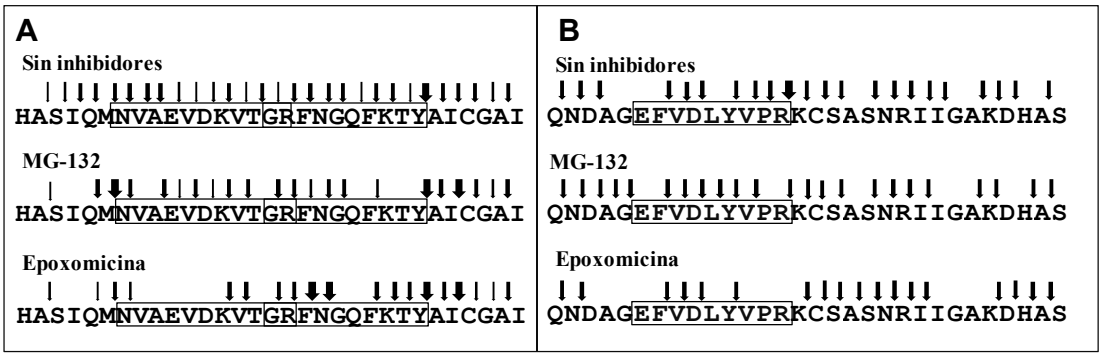


Figura 19: Digestión de precursores sintéticos de ligandos de HLA-A68 y B27 procedentes de la proteína S21 de la subunidad ribosomal 40S. Las flechas más finas representan cortes que generan especies con una abundancia relativa en el rango 0.1-1.0%, las de grosor intermedio 1.0-10.0% y las más gruesas <10.0%.

	Sin inhibidor	MG-132	Epoxomicina
Cantidad digerida	100%	28.9%	54.3%
Productos de digestión	80	28	16
Peptidos que implican corte dual[a]	66 (82.5%)	20 (71.4%)	7 (43.8%)
Enlaces cortados[b]	29 (96.7%)	22 (73.3%)	19 (63.3%)
Especies productivas (péptido de A68)[c]	4 (5.1%)	2 (7.1%)	1 (6.3%)
Rendimiento relativo de cortes productivos[d]	6.9%	5.5%	2.6%
Cortes destructivos (péptido A68)[e]	10 (34.5%)	8 (36.4%)	4 (21.1%)
Especies productivas (péptido de B27)[c]	12 (15.2%)	2 (7.1%)	2 (12.5%)
Rendimiento relativo de cortes productivos[d]	28.2%	14.6%	17.0%
Cortes destructivos (péptido B27)[e]	9 (31%)	6 (27.3%)	7 (36.8%)
Destrucción/producción (péptido A68)[f]	2.9	4.8	3.0
Destrucción/producción (péptido B27)[f]	0.9	1.3	2.7

Tabla 5: Efecto de los inhibidores de proteasoma en la digestión del sustrato de 31 residuos (Figura 19A). ([a]) Porcentaje relativo al número total de productos de digestión. ([b]) Porcentaje relativo al número total de enlaces peptídicos en el sustrato. ([c]) Péptidos con el extremo C-terminal del epítopo y el N-terminal con extensión o primer residuo del epítopo. Porcentaje relativo al número total de productos de digestión. ([d]) Los cortes productivos corresponden a la rotura en C-terminal de R17 para el epítopo de HLA-A68 y de Y25 para el epítopo de HLA-B27. ([e]) Cortes que implican destrucción del epítopo. Porcentaje relativo al número total de enlaces peptídicos cortados. ([f]) Relación entre el rendimiento relativo de cortes que implican destrucción del epítopo y los péptidos potencialmente productivos (Detalles en Apéndice 1, Tablas 5S y 6S).

	Sin inhibidor	MG-132	Epoxomicina
Cantidad digerida	75.8%	56.8%	49.9%
Productos de digestión	21	18	13
Peptidos que implican corte dual[a]	13 (61.9%)	8 (44.4%)	8 (61.5%)
Enlaces cortados[b]	22 (75.9%)	23 (79.3%)	18 (62.1%)
Especies productivas[c]	1	1	0
Rendimiento relativo de cortes productivos[d]	16.8%	9.0%	0%
Cortes destructivos[e]	6 (27.3%)	6 (26.1%)	4 (22.2%)
Destrucción/producción[f]	1.2	2.3	-

Tabla 6: Efecto de los inhibidores de proteasoma en la digestión del sustrato de 30 residuos (Figura 19A). ([a]) Porcentaje relativo al número total de productos de digestión. ([b]) Porcentaje relativo al número total de enlaces peptídicos en el sustrato. ([c]) Péptidos con el extremo C-terminal del epítopo y el N-terminal con extensión o primer residuo del epítopo. Porcentaje relativo al número total de productos de digestión. ([d]) Los cortes productivos corresponden a la rotura en C-terminal de R14. ([e]) Cortes que implican destrucción del epítopo. Porcentaje relativo al número total de enlaces peptídicos cortados. ([f]) Relación entre el rendimiento relativo de cortes que implican destrucción del epítopo y los péptidos potencialmente productivos (Detalles en Apéndice 1, Tabla 7S).

Actividad	**Sin Inhibidor**		**MG-132**		**Epoxomicina**	
31-mero	**Cortes**	**%**	**Cortes**	**%**	**Cortes**	**%**
Tríptica	3	15.4	2	8.3	3	11.6
Quimotríptica	13	65.1	11	70.3	9	70.1
Caspasa	2	1.2	2	0.9	0	0.0
Otros	11	18.3	7	20.5	7	18.4
Total	29	100.0	23	100.0	20	100.0
30-mero	**Cortes**	**%**	**Cortes**	**%**	**Cortes**	**%**
Tríptica	4	27.6	4	17.3	3	12.7
Quimotríptica	8	25.4	9	35.2	6	37.5
Caspasa	3	21.1	2	12.1	2	12.5
Otros	7	25.9	8	35.4	7	37.2
Total	22	100.0	23	100.0	18	100.0

Tabla 7: Efecto de los inhibidores sobre las subespecificidades hidrolíticas del proteasoma 20S.

Resultados II

4.2. Efecto del polimorfismo natural de ERAP1 sobre el peptidoma de HLA-B27 *in vivo*.

En la segunda parte de esta Tesis se estudió el efecto de las variantes alélicas de ERAP1 en la configuración del peptidoma de HLA-B*27:04, las bases moleculares de la interacción funcional de esta molécula con HLA-B27 y sus implicaciones en el mecanismo de asociación con EA.

4.2.1. Variantes alélicas de ERAP1 en líneas celulares HLA-B27 positivas.

El genotipado de SNPs no sinónimos y la secuenciación del gen codificante para ERAP1 revelaron la presencia de cuatro variantes del enzima en las cuatro líneas celulares utilizadas en este estudio (Tabla 8). En JSL la secuencia codificante de ERAP1 era idéntica a la secuencia consenso, que presenta seis cambios de base asociados al riesgo de EA. C1R-B*27:04 presenta cuatro cambios de base con respecto a la secuencia consenso, perdiéndose tres de los cambios asociados a enfermedad. KNE fue la única línea celular heterocigota para el gen de ERAP1, presentando dimorfismo en 3 posiciones. Wewak I presenta 6 cambios no sinónimos que sustituyen a todos los cambios asociados a enfermedad en la secuencia consenso, resultando una variante en la que se ha perdido completamente la asociación con EA.

SNP	Posición (Nucl./AA)	Consenso	*Odds Ratio*	JSL	C1R	KNE	Wewak I
rs26653	380 / 127	g / R*	1.30 (Harvey et al., 2009)	g / R*	**c / P**	g,**c** / R,**P**	**c / P**
rs26618	828 / 276	a / I	0.99 (Harvey *et al.*, 2009)	a / I	**g / M**	a,g / I,M	a / I
rs27895	1037 / 346	g / G	1.07 (Harvey *et al.*, 2009)	g / G	g / G	g**,a** / G,**D**	g / G
rs2287987	1045 / 349	a / M*	0.71 (Brown, 2008)	a / M*	a / M*	a / M*	**g / V**
rs30187	1583 / 528	a / K*	1.40 (Brown, 2008)	a / K*	**g /R**	**g /R**	**g /R**
rs10050860	1723 / 575	g / D*	0.71 (Brown, 2008)	g / D*	g / D*	g / D*	**a / N**
rs17482078	2174 / 725	g / R*	0.70 (Brown, 2008)	g / R*	g / R*	g / R*	**a / Q**
rs27044	2188 / 730	c / Q*	1.40 (Brown, 2008)	c / Q*	**g /E**	**g /E**	**g /E**

Tabla 8: Variantes alélicas de ERAP1 en líneas celulares HLA-B*27:04 positivas. Solamente se muestran los cambios de base no sinónimos correspondientes a la hebra codificante del DNA. Los cambios de base asociados al riesgo de EA (Brown, 2008; Harvey et al., 2009) aparecen marcados con asteriscos, y los que implican variaciones con respecto a la secuencia consenso (Num. Acceso: Q9NZ08-2) están señalados en negrita.

Aunque se sabe que los niveles de expresión de ERAP1 son equiparables en diferentes LCLs (Fruci *et al*. 2006), esto se comprobó en las líneas celulares empleadas. Los niveles de expresión observados mediante Western blot fueron similares en Wewak I, C1R-B*27:04 y KNE (Apéndice 2, Figura 1B). Este análisis no pudo llevarse a cabo en JSL, puesto que esta línea se extinguió en el transcurso de los experimentos. La expresión relativa de mRNA determinada por qRT-PCR mostró solamente diferencias moderadas entre las cuatro líneas celulares (Apéndice 2, Figura 1C).

4.2.2. Análisis comparativo de peptidomas procedentes de diferentes líneas celulares HLA-B27 positivas: planteamiento estratégico.

Se realizaron tres comparaciones de peptidomas, Wewak I/JSL, Wewak I/C1R-B*27:04 y C1R-B*27:04/KNE. Se adquirieron las señales iónicas por encima de un umbral de relación señal/ruido de 3 en el rango m/z de 800-2000. El análisis comparativo se restringió a los ligandos comunes a las dos LCLs en cada caso, empleando el software MSHandler para la obtención de los conjuntos de péptidos compartidos. Los parámetros de análisis empleados fueron los siguientes: umbral de intensidad de detección 200 (300 para C1R-B*27:04/KNE), rango de m/z 850-1700, intervalos de confianza para m/z y posición de elución ±0.2 y ±1 fracción, respectivamente, filtro de masas incompatibles con péptidos activado (masas compatibles en el intervalo de valor decimal 0.2-0.9), exclusión de señales inespecíficas activada.

La abundancia relativa de los péptidos en las líneas celulares comparadas se estimó como la relación de intensidades (RI) de las señales equivalentes en los respectivos espectros de masas MALDI-TOF de las líneas celulares comparadas. Aunque estos parámetros no constituyen un criterio cuantitativo, proporcionan una estimación general de abundancia relativa que es válido cuando se aplica a series grandes de péptidos y constituye un reflejo de tendencias colectivas. Se clasificaron los péptidos en tres grupos atendiendo a su valor de RI: $1.0 \leq RI < 1.5$; $1.5 < RI \leq 3.0$; $RI > 3.0$. Además se trató el conjunto de todos los péptidos con $RI \geq 1.0$ como un grupo más. Se comparó la masa media de los grupos equivalentes de péptidos para de determinar si existe un efecto en el tamaño de los ligandos asociado a la variante de ERAP1. La comparación de las curvas de distribución de tamaño supuso el criterio principal para evaluar la calidad de los resultados.

La distribución de los péptidos por longitud se estimó de la forma descrita en el apartado 3.9 del capítulo de Materiales y Métodos. Los cálculos se realizaron en base a la distribución por longitud de 372 secuencias peptídicas obtenidas (Tabla 9; Anexo 2, Tabla S3).

	Longitud de Péptido							
Mw (Da)	<9-mers		9-mers		10-mers		>10-mers	
	N	%	N	%	N	%	N	%
850-900	3	100.0	0	0.0	0	0.0	0	0.0
900-1000	8	21.1	30	79.0	0	0.0	0	0.0
1000-1110	7	5.9	102	86.4	8	6.8	1	0.9
1100-1200	0	0.0	127	85.8	20	13.5	1	0.7
1200-1300	0	0.0	24	52.2	19	41.3	3	6.5
1300-1400	0	0.0	1	8.3	7	58.3	4	33.3
1400-1500	0	0.0	0	0.0	0	0.0	4	100.0
1500-1600	0	0.0	0	0.0	0	0.0	2	100.0
1600-1700	0	0.0	0	0.0	0	0.0	1	100.0
Total	**18**	**4.8**	**284**	**76.3**	**54**	**14.5**	**16**	**4.3**

Tabla 9: Proporción de ligandos de diferentes longitudes en cada rango de tamaño. Los porcentajes totales presentados en la fila inferior de la Tabla son relativos a las 372 secuencias de ligandos de HLA-B*27:04 empleadas para estos cálculos.

4.2.2.1. Peptidomas de líneas celulares con distintas variantes de ERAP1.

En los peptidomas de Wewak I y JSL se encontraron 3222 péptidos comunes, con una masa media de 1162.0 Da. La masa media de los péptidos de Wewak I con RI≥1.0 fue 33.7 Da mayor que en JSL y este valor ascendía hasta 73.1 Da en los péptidos con RI>3.0 y hasta 49.2 Da para los péptidos con 3.0≥RI>1.5 (Tabla 10). Los péptidos de mayor tamaño, por encima de 1200 Da, fueron predominantes en Wewak I con respecto a JSL, mientras que por debajo de este tamaño se observó el efecto contrario en los subconjuntos de péptidos con 3.0≥RI>1.5 y RI>3.0 (Figura 20A). En estos grupos de péptidos la cantidad estimada de octámeros y nonámeros fue superior en JSL, mientras que los péptidos de mayor longitud fueron más abundantes en Wewak I (Figura 21A).

En la comparación entre Wewak I y C1R-B*27:04 se encontraron 5240 péptidos comunes, de 1154.3 Da de masa media. La masa media en los péptidos de todos los rangos de RI≥1.0 era también mayor en Wewak I, alcanzando una diferencia de masa de hasta 58.6 Da en los de RI>3.0 (Tabla 10). Aunque también se observó un predominio de los péptidos de mayor tamaño en Wewak I, este efecto fue más leve que en la comparación con JSL (Figura 20B). En la distribución de longitud estimada de péptidos se encontró, como en el anterior caso, una menor cantidad de octámeros y nonámeros y mayor cantidad de péptidos de longitud superior en Wewak I (Figura 21B).

Los valores medios de intensidad de las señales de MALDI-TOF de los peptidomas comparados fueron equiparables (Tabla 10). Ello descarta que los resultados puedan estar condicionados

significativamente por diferencias en la cantidad de material cargado en la placa de MALDI-TOF.

4.2.2.2. Peptidomas de líneas celulares con variantes de ERAP1 similares.

En el análisis comparativo de los peptidomas de C1R-B*27:04 y KNE se encontraron 6974 péptidos comunes a ambas líneas, con 1149.9 Da de masa media (Tabla 10). La distribución modal de masas fue similar en todos los grupos de péptidos (Figura 20C). La diferencia de masa observada en péptidos predominantes en cada línea, con RI≥1.0, fue de solamente 5.0 Da y la máxima diferencia encontrada fue de 9.4 Da, en el grupo de péptidos con 3.0≥RI>1.5 (Tabla 10). La distribución de longitudes estimadas de péptido fue equiparable en ambas líneas celulares en todos los rangos de RI (Figura 21C).

Estos resultados indican que el polimorfismo de ERAP1 tiene un efecto global sobre el peptidoma de HLA-B27 que afecta al nivel de expresión de muchos de los péptidos en función de su tamaño. La ausencia de diferencias de expresión dependientes de tamaño en los peptidomas de C1R-B*27:04 /KNE, que presentan un fondo genético similar para ERAP1, sugiere que el polimorfismo de ERAP1 es el responsable cuando se observan este tipo diferencias en otras LCLs. Asumiendo que existe una correlación inversa entre la eficiencia hidrolítica del enzima y el tamaño de los péptidos generados, se puede establecer que la variante más activa sería la que presenta JSL, la menos activa la que presenta Wewak I y C1R-B*27:04 y KNE tendrían enzimas de actividad hidrolítica intermedia y similar entre sí.

Comparación	Intensidad media (MS)	Valor m/z promedio				
Wewak I/JSL		**Total**	**IR≥1.0**	**IR>3.0**	**IR>1.5-3.0**	**IR≥1.0-1.5**
Wewak I	2028	1162.0	1177.6	1197.4	1179.2	1168.6
JSL	1865	1162.0	1143.9	1124.3	1130	1162.7
Δ m/z			33.7	73.1	49.2	5.9
Wewak I/C1R-04						
Wewak I	1181	1154.3	1170.8	1170.6	1176.7	1166.2
C1R-04	1246	1154.3	1154.3	1112	1138.3	1151.5
Δ m/z			30.6	58.6	38.4	14.6
C1R-04/KNE						
C1R-04	2799	1149.4	1151.5	1141.9	1156.3	1152.1
KNE	2083	1149.4	1145.5	1133.3	1146.9	1147.6
Δ m/z			5.0	8.6	9.4	4.5

Tabla 10: Distribución por tamaño y expresión relativa de ligandos de HLA-B*27:04 comunes a líneas celulares con diferentes variantes de ERAP1. El número total de péptidos comunes fue de 3222, 5240 y 6974 para Wewak I/JSL, Wewak I/C1R-B*27:04 y C1R-B*27:04/KNE, respectivamente.

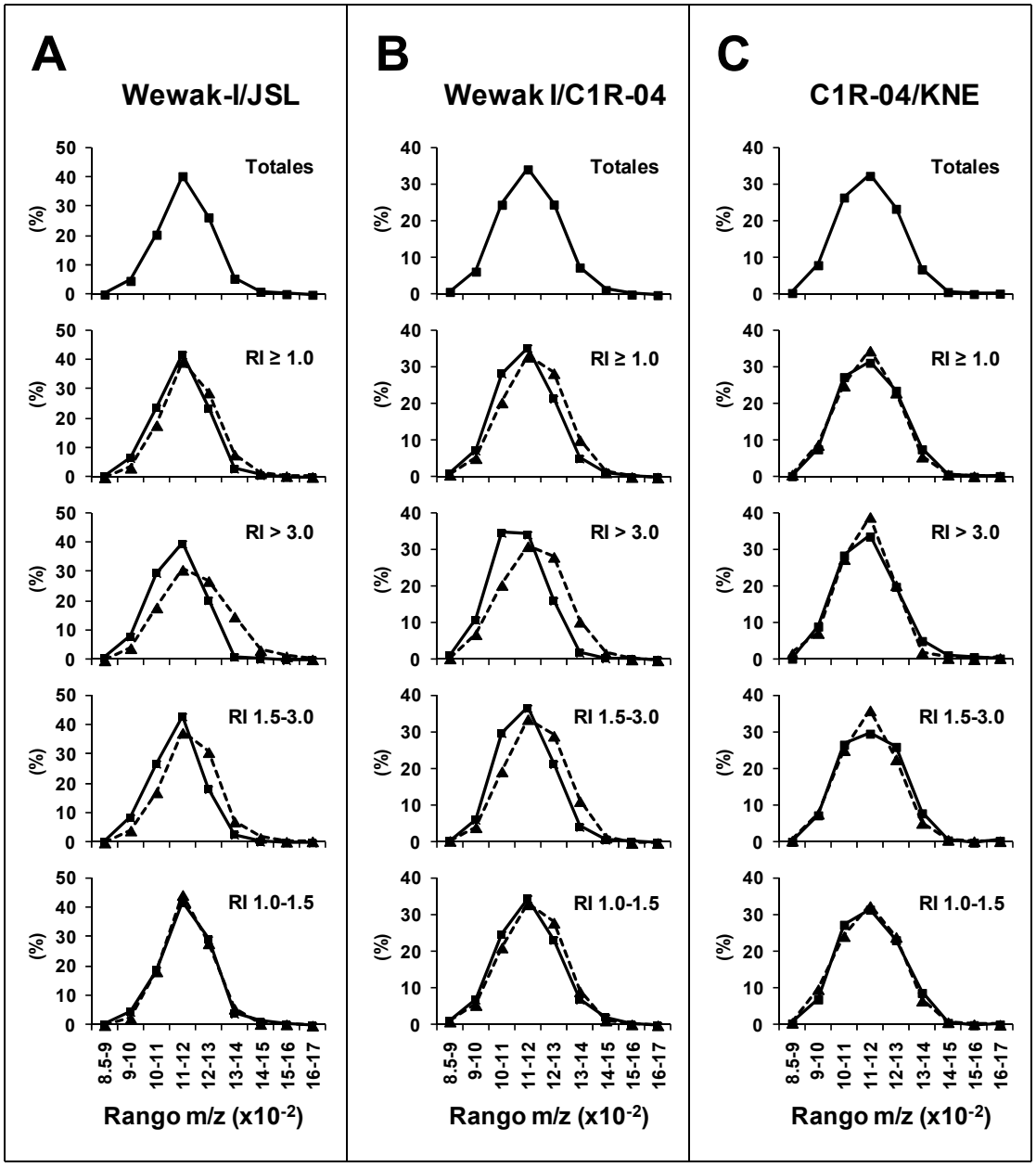


Figura 20: Distribución por tamaño de los ligandos de HLA-B*27:04 en función de su abundancia relativa en varios contextos de ERAP1. (A) Wewak I/JSL. El gráfico superior muestra la distribución de MW de 3222 péptidos compartidos entre ambas líneas. Los otros cuatro histogramas comparan las distribuciones de MW correspondientes a los subconjuntos de péptidos predominantes en Wewak I (línea discontinua) y JSL (línea continua). (B) Wewak I/C1R-B*27:04. En el histograma superior se muestra la distribución de MW de los 5240 péptidos compartidos. En los otros cuatro se compara la distribución de MW de los péptidos predominantes en Wewak I (línea discontinua) y C1R-B*27:04 (línea continua). (C) C1R-B*27:04/KNE. En el histograma superior se muestra la distribución de MW de los 6974 péptidos compartidos. En los otros cuatro se compara la distribución de MW de los péptidos predominantes en Wewak I (línea discontinua) y C1R-B*27:04 (línea continua).

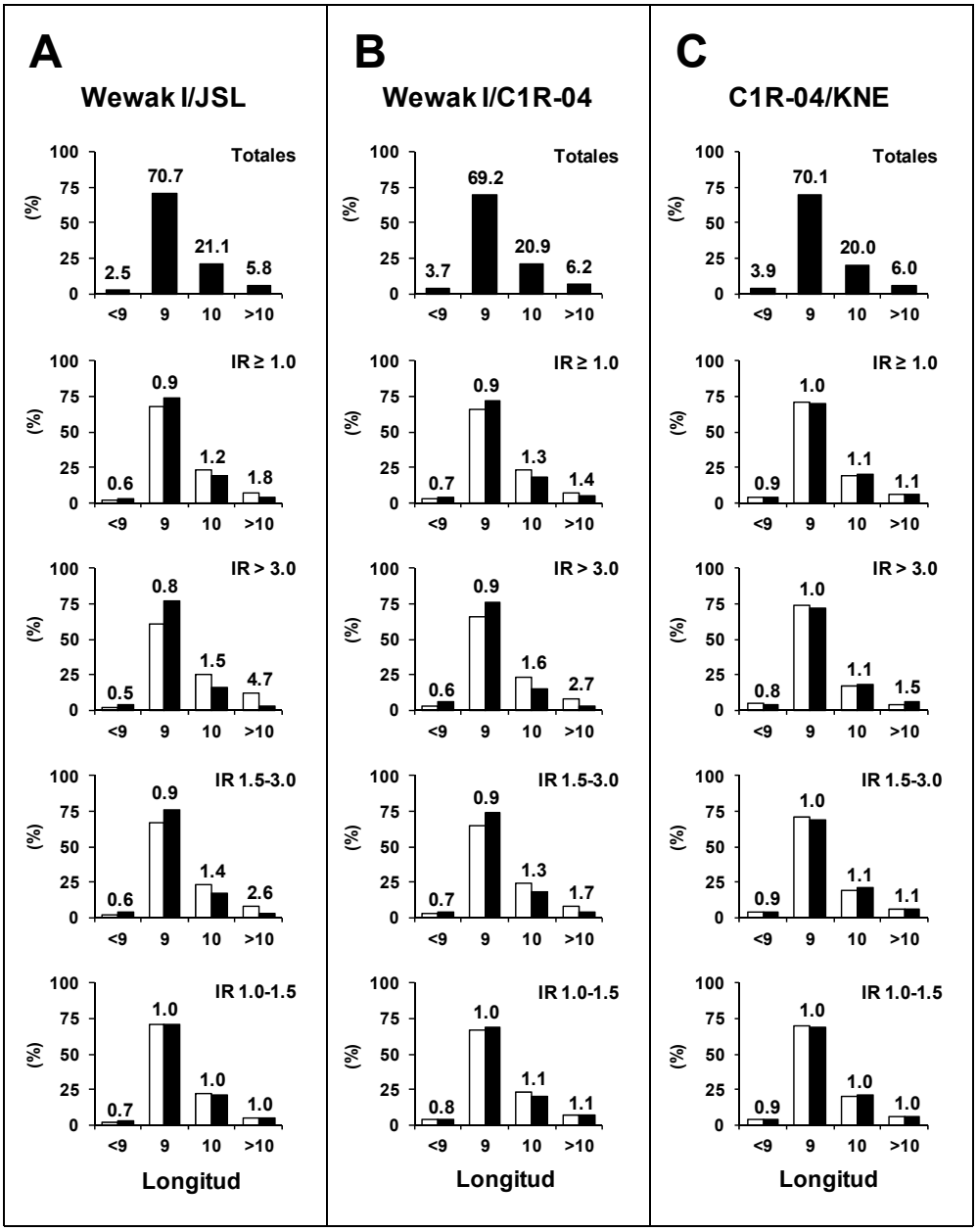


Figura 21: Distribución por longitud de los ligandos de HLA-B*27:04 en función de su abundancia relativa en varios contextos de ERAP1. (A) Wewak I/JSL. El histograma superior muestra la distribución de longitud de 3222 péptidos compartidos entre ambas líneas. Se indica el porcentaje de péptidos de <9, 9, 10 y >10 residuos. Los otros cuatro histogramas comparan las distribuciones correspondientes a los subconjuntos de péptidos predominantes en Wewak I (barras blancas) y JSL (barras negras). El cociente entre el porcentaje de péptidos de la misma longitud en las dos líneas celulares se indica sobre las correspondientes barras. (B) Wewak I/C1R-B*27:04. El histograma superior muestra la distribución de longitud de 5240 péptidos compartidos entre ambas líneas. Los otros cuatro histogramas comparan las distribuciones correspondientes a los subconjuntos de péptidos predominantes en Wewak I (barras blancas) y C1R-B*27:04 (barras negras) siguiendo los mismos criterios que en A. (C) C1R-B*27:04/KNE. El histograma superior muestra la distribución de longitud de 6974 péptidos compartidos entre ambas líneas. Los otros cuatro histogramas comparan las distribuciones correspondientes a los subconjuntos de péptidos predominantes en C1R-B*27:04 (barras negras) y KNE (barras blancas) siguiendo los mismos criterios que en A.

4.2.3. Base molecular de la expresión diferencial de ligandos de HLA-B27 dependiente del polimorfismo de ERAP1.

Los siguientes análisis se llevaron a cabo para buscar patrones estructurales de los péptidos relacionados con su nivel de expresión en el contexto de las distintas variantes alélicas de ERAP1.

La influencia del sustrato en la eficiencia de corte por ERAP1 se analizó por una parte al nivel de los residuos flanqueantes situados en posiciones N-terminales adyacentes al epítopo (P-2 y P-1) y del residuo P1, y por otra parte a nivel de los residuos internos del propio epítopo. La susceptibilidad de los residuos N-terminales flanqueantes afecta directamente al proceso catalítico y determina la eficiencia con la que se genera el epítopo. La susceptibilidad del residuo P1 determina la degradación o la supervivencia del epítopo. La contribución de los residuos internos se reduce a su efecto sobre la afinidad del enzima por el sustrato.

4.2.3.1. Influencia de los residuos en P1 y de la secuencia flanqueante N-terminal.

El sistema empleado para la estimación de la susceptibilidad de los residuos en estas posiciones se ha descrito en el apartado 3.10 del capítulo de Materiales y Métodos. En la Tabla 11 se muestran los promedios de puntuación de susceptibilidad a ERAP1 en cada grupo de péptidos para las tres comparaciones, así como el cociente de estos promedios en los grupos equivalentes de péptidos en las líneas celulares comparadas.

En la comparación Wewak I/JSL se observó una susceptibilidad en P1 2.5 veces mayor en Wewak I para los péptidos con RI>3.0, mientras que las diferencias encontradas para los péptidos con 3.0≥RI>1.5 fueron muy pequeñas. La susceptibilidad en las dos posiciones N-terminales flanqueantes fue también más alta en Wewak I para los péptidos con RI>3.0 (relación P-2+P-1 en ambas líneas celulares: 1.4) y similar para ambas líneas en péptidos con RI≤3.0 (relación P-2+P-1: 1.1). Estas observaciones son compatibles con que Wewak I presenta una variante menos activa del enzima que es capaz de producir abundantemente péptidos procedentes de precursor N-terminales con secuencias muy susceptibles a ERAP1. En el contexto de una variante más activa, en JSL, estos péptidos serían degradados más eficientemente, explicando su predominio en Wewak I, que presenta la variante menos activa.

En Wewak I/C1R-B*27:04 la susceptibilidad de los residuos en P1 fue equiparable para ambas líneas celulares en todos los grupos de péptidos. Sin embargo, la suma de las contribuciones de P-2 y P-1 en el grupo con RI>3.0 la susceptibilidad fue 1.4 veces mayor para los péptidos de Wewak I, resultado también explicable por una variante menos activa en Wewak I, capaz de eliminar residuos N-

terminales muy susceptibles a la vez que preservaría la secuencia del epítopo.

Las diferencias de susceptibilidad en la región N-terminal en C1R-B*27:04 y KNE fueron menores. Las relaciones de susceptibilidad no superaban 1.2 en ningún grupo de péptidos excepto para P-1 en el grupo con 3.0≥RI>1.5. Este resultado, junto con la distribución similar de masa y longitud, sugiere que las diferencias de expresión de péptidos en estas líneas celulares no están relacionadas con ERAP1.

Los resultados de estos análisis conducen a la conclusión de que el balance entre la producción y destrucción de los epítopos, que es dependiente de la variante de ERAP1 y de la secuencia N-terminal de los sustratos, es crítico para la configuración del peptidoma de HLA-B27. Esta constituye la explicación más plausible para el mecanismo de interacción funcional entre ERAP1 y HLA-B27.

Grupo	Puntuación promedio			
Wewak I/JSL	**P-2**	**P-1**	**P-1+P-2**	**P1**
RI >3				
Wewak I	34.2	38.4	72.6	29.9
JSL	14.4	37.9	52.3	12.0
***Ratio* (>3)**	**2.4**	**1.0**	**1.4**	**2.5**
3.0≥RI>1.0				
Wewak I	34.8	44.5	79.2	17.1
JSL	23.3	48.5	71.8	20.4
***Ratio* (≤3)**	**1.5**	**0.9**	**1.1**	**0.8**

Wewak I /C1R-04	**P-2**	**P-1**	**P-1+P-2**	**P1**
RI >3				
Wewak I	38.5	46.3	84.7	23.4
C1R-B*27:04	26.5	36.1	62.6	22.6
***Ratio* (>3)**	**1.5**	**1.3**	**1.4**	**1.0**
3.0≥RI>1.0				
Wewak I	30.3	41.7	72.0	20.6
C1R-B*27:04	33.8	43.6	77.5	19.5
***Ratio* (≤3)**	**0.9**	**1.0**	**0.9**	**1.1**

KNE/C1R-04	**P-2**	**P-1**	**P-1+P-2**	**P1**
RI >3				
KNE	33.6	42.9	75.9	19.4
C1R-B*27:04	30.7	36.1	66.4	23.2
***Ratio* (>3)**	**1.1**	**1.2**	**1.1**	**0.8**
3.0≥RI>1.0				
KNE	27.3	46.3	73.6	20.7
C1R-B*27:04	25.9	36.8	62.8	21.7
***Ratio* (≤3)**	**1.1**	**1.3**	**1.2**	**1.0**

Tabla 11: Influencia del extremo N-terminal en la expresión de péptidos de HLA-B*27:04 en el contexto de diferentes variantes de ERAP1.

4.2.3.2. Residuos internos del epítopo.

El método seguido se describe en apartado 3.11 del capítulo de Materiales y Métodos. Dadas las limitaciones de éste método, el estudio tuvo que restringirse a los nonámeros y puesto que el residuo en P2 de la práctica totalidad de los ligandos de HLA-B27 es R, se evaluaron las posiciones P3 a P9. De forma análoga al anterior análisis, se comparó la puntuación de los residuos en los conjuntos equivalentes de péptidos para cada par de peptidomas. No se encontraron diferencias globales en estas calificaciones para ninguno de los grupos de péptidos (Tabla 12), lo cual sugiere una escasa influencia de la secuencia interna del epítopo en la eficiencia global del enzima sobre el peptidoma de HLA-B27.

Grupo								
Wewak I / JSL	**P3**	**P4**	**P5**	**P6**	**P7**	**P8**	**P9**	**Promedio**
RI >3.0								
Wewak I	3.2	1.4	1.0	3.5	0.7	1.0	5.1	2.3
JSL	2.6	0.6	0.5	5.1	0.7	1.4	6.0	2.4
Ratio (>3)	**1.2**	**2.4**	**1.9**	**0.7**	**1.0**	**0.7**	**0.9**	**1.0**
3.0≥RI>1.0								
Wewak I	3.3	0.9	0.8	3.6	0.5	1.1	6.0	2.3
JSL	2.7	0.9	0.9	5.0	0.4	1.5	5.7	2.4
Ratio (≤3)	**1.2**	**1.0**	**1.0**	**0.7**	**1.5**	**0.8**	**1.1**	**1.0**

Wewak I / C1R-04	**P3**	**P4**	**P5**	**P6**	**P7**	**P8**	**P9**	**Promedio**
RI >3.0								
Wewak I	3.4	1.9	1.1	4.1	0.7	2.4	5.4	2.7
C1R-B*27:04	2.6	1.0	0.7	5.0	0.9	1.1	5.6	2.4
W/C Ratio (>3)	**1.3**	**2.0**	**1.5**	**0.8**	**0.8**	**2.2**	**1.0**	**1.1**
3.0≥RI>1.0								
Wewak I	3.2	1.2	0.9	4.2	0.8	1.8	5.9	2.6
C1R-B*27:04	3.0	1.2	0.6	3.9	1.2	1.9	6.3	2.6
Ratio (≤3)	**1.1**	**1.0**	**1.4**	**1.1**	**0.7**	**0.9**	**0.9**	**1.0**

KNE / C1R-04	**P3**	**P4**	**P5**	**P6**	**P7**	**P8**	**P9**	**Promedio**
RI >3.0								
KNE	2.9	1.0	0.7	3.6	0.5	1.1	5.9	2.2
C1R-B*27:04	2.8	0.7	0.3	3.7	0.9	1.2	5.4	2.1
Ratio (>3)	**1.0**	**1.6**	**2.2**	**1.0**	**0.6**	**0.9**	**1.1**	**1.0**
3.0≥RI>1.0								
KNE	3.1	1.6	1.0	3.6	0.3	1.6	6.0	2.5
C1R-B*27:04	3.0	0.7	0.9	5.0	1.2	0.9	6.0	2.5
Ratio (≤3)	**1.0**	**2.3**	**1.2**	**0.7**	**0.3**	**1.8**	**1.0**	**1.0**

Tabla 12: Influencia de la secuencia interna en la expresión de péptidos de HLA-B*27:04 en el contexto de diferentes variantes de ERAP1. En las columnas segunda a octava se muestran los valores medios de susceptibilidad de los residuos en cada posición. En la novena columna se muestra el promedio de la susceptibilidad de los residuos en las posiciones P3 a P9.

4.2.4. Influencia del polimorfismo de ERAP1 en la termoestabilidad de HLA-B27.

Para evaluar la estabilidad de HLA-B*27:04 en diferentes contextos de ERAP1, se llevaron a cabo estudios de termoestabilidad de la molécula aislada de las cuatro LCLs empleadas en esta Tesis. La mayoría de los subtipos de HLA-B27 asociados con la predisposición a sufrir EA presentan una alta termoestabilidad, lo cual no ocurre en los subtipos no asociados a EA (Galocha and López de Castro, 2010). Se encontró una alta termoestabilidad de la molécula expresada en JSL, C1R-B*27:04 y KNE, mientras que la termoestabilidad de la molécula expresada en Wewak I fue sensiblemente menor incluso a tiempos de caza largos. Este resultado sugiere que el peptidoma de HLA-B*27:04 en Wewak I está menos optimizado que en el resto de LCLs.

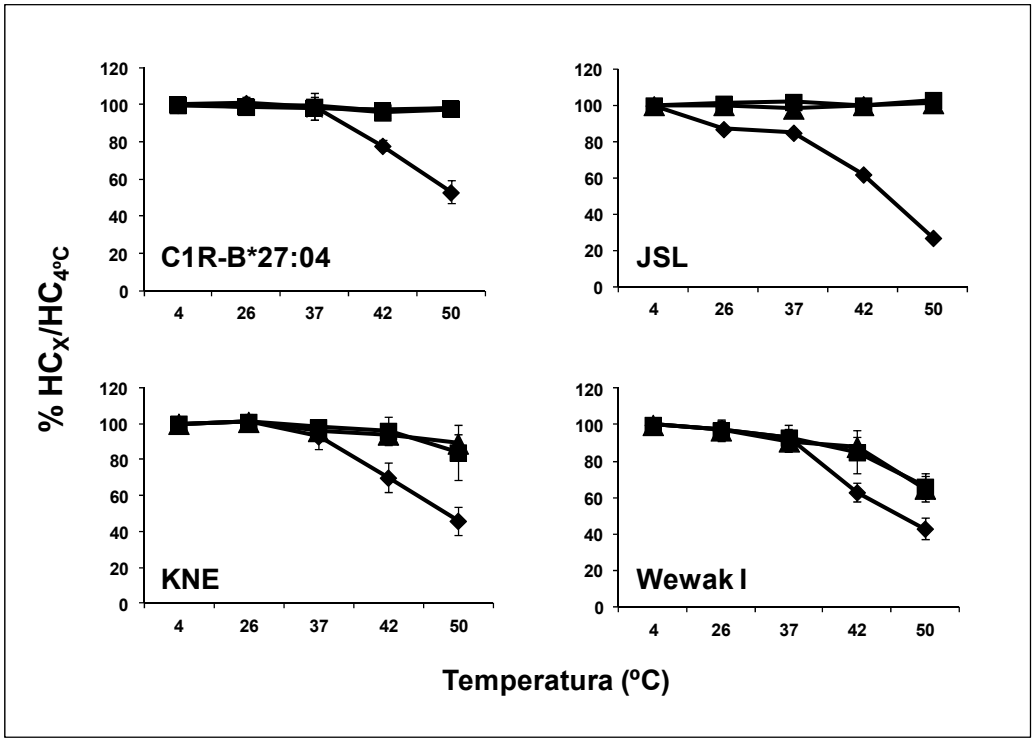


Figura 22: Termoestabilidad de HLA-B*27:04 en varios contextos de ERAP1. Las células se marcaron durante 15 min y se realizaron cazas a 0, 2 y 4h. Se incubaron alícuotas iguales de lisados celulares durante 1h a las temperaturas indicadas antes de la inmunoprecipitación con el anticuerpo ME1, que reconoce complejos HLA-B27/péptido en su conformación nativa pero no la HC desplegada. Posteriormente se separó el material por SDS-PAGE y se analizó por fluorografía. El porcentaje de complejos B27/péptido recuperados en las cazas a 0h (♦), 2h (■) o 4h (▲) tras el calentamiento se representaron como la intensidad de HC a una temperatura dada (HC_x) relativa a la intensidad a 4ºC ($HC_{4ºC}$). Se presentan los valores medios con desviación estándar para 4 experimentos con C1R-B*27:04, 5 con KNE y 8 con Wewak I. Solamente se pudo llevar a cabo un experimento con JSL.

Discusión

5.1. Interacción funcional entre el proteasoma, ERAP1 y HLA-I.

Esta Tesis se ha centrado en analizar el papel de dos de los principales sistemas proteolíticos implicados en la generación de ligandos de HLA-I: el proteasoma y ERAP1. La importancia de ambas enzimas para la generación de peptidomas de clase I está ampliamente documentada y es evidente debido, por ejemplo, a la drástica reducción de la expresión de MHC-I tanto en presencia de inhibidores de proteasoma (Luckey et al., 2001; Marcilla *et al.*, 2007) como en ratones *knockout* de ERAAP (Hammer et al., 2007), el homólogo murino de ERAP1. La estrecha relación funcional entre las enzimas implicadas en el procesamiento antigénico y HLA-I va más allá de su papel en la generación de epítopos específicos y tiene un efecto global en la biología de HLA. Esto es claro cuando se considera que la unión estable de péptidos a HLA-I es crítica para el correcto plegamiento y expresión en superficie de estas moléculas. En condiciones fisiológicas, el dímero HC/β2M libre de péptido no es estable ni detectable y la constante de disociación de ligandos fisiológicos de HLA es extremadamente baja (Bouvier et al., 1998). Por estas razones, los péptidos unidos a HLA-I no tienen un comportamiento típico de un ligando, sino de un componente constitutivo de un complejo trimolecular HC/β2M/péptido.

Como es sabido, el proteasoma actúa en las etapas iniciales de la ruta de procesamiento de clase I degradando proteínas y generando el extremo C-terminal de la inmensa mayoría de los ligandos de MHC-I (Pamer *et al.*, 1998; Rock *et al.*, 1999). Probablemente algunas peptidasas citosólicas, tales como TPP II, IDE y nardilisina, podrían generar adicionalmente algunos extremos C-terminales de ligandos de clase I (Seifert *et al.*, 2003; Parmentier *et al.*, 2010; Kessler *et al.*, 2011). La identificación de una carboxipeptidasa en el retículo endoplásmico (Shen *et al.*, 2011) deja asimismo abierta la posibilidad de un cierto procesamiento C-terminal de péptidos en este compartimento.

En contraste con el proteasoma, ERAP1 se encarga de generar el extremo N-terminal de muchos ligandos de clase I en el retículo endoplásmico (Serwold *et al.*, 2002). Por lo tanto, el proteasoma y ERAP1 están funcionalmente relacionados, actúan secuencial y conjuntamente y son dos de los principales responsables de la configuración de los peptidomas de clase I.

La asociación de ERAP1 con EA en individuos HLA-B27 positivos (The TASK and WTCCC2 Consortia, 2011) añade un interés especial al análisis de la interacción funcional entre ambas proteínas.

5.2. Procesamiento citosólico de péptidos de clase I: origen de los peptidomas de HLA-I resistentes a inhibidores de proteasoma.

Uno de los objetivos de ésta Tesis fue determinar si la observación de que los ligandos de HLA-B27 resistentes a inhibidores de proteasoma proceden de proteínas pequeñas básicas (Marcilla *et al.*, 2007) puede ser generalizable a otras moléculas de HLA-I y si ello es el resultado de un procesamiento independiente de proteasoma o de una inhibición incompleta. El estudio se centró en HLA-A*68:01, ya que este alotipo une de forma casi exclusiva péptidos con residuos C-terminales básicos, lo cual podría hacerle particularmente susceptible a una inhibición parcial de la subespecificidad tríptica del proteasoma (Rock et al., 2002).

5.2.1. Motivo peptídico de HLA-A*68:01 y su relación con HLA-B27.

La aplicación de las técnicas de secuenciación masiva *de novo* basadas en electrospray/Orbitrap, en conjunción con sofisticados criterios de análisis de los datos de MS/MS (Ben Dror *et al.*, 2010) así como el empleo de herramientas informáticas desarrolladas en este Tesis (MSgene) para la interpretación *de novo* de espectros MS/MS de MALDI-TOF posibilitaron la identificación de una gran cantidad de ligandos de A*68:01 y el refinamiento de su motivo de unión. Dicho motivo puede describirse por restricciones significativas en cuatro posiciones de anclaje: en P1 se observó una clara preferencia por residuos ácidos, T y S; en P2 los residuos preferentes fueron V, T y A, que tienen tamaño similar; en P3 residuos no polares y en PC la práctica totalidad de los residuos encontrados fueron básicos. Los péptidos cortos se desviaban de este patrón en las posiciones de la región N-terminal (P1-P3). Algunos de estos, especialmente octámeros, son probablemente ligandos canónicos o pueden unirse de forma no canónica (Khan et al., 2000). Sin embargo, a pesar del uso de inhibidores de proteasas durante la purificación, no es posible descartar formalmente que algunos de los péptidos más cortos procedan de una actividad aminopeptidasa residual.

El motivo de A*68:01 es muy diferente al de B*27:05, donde están favorecidos los residuos básicos y G en P1. El motivo R2 de HLA-B27 no aparece en A68 y solo el 42% de los ligandos de B27 presentan residuos básicos en C-terminal (Ben Dror *et al.*, 2010). Por tanto, las dos moléculas unen peptidomas altamente divergentes. Sin embargo, el hecho de que el 11% de las proteínas parentales de los ligandos de A68 eran también fuente de ligandos de B27 indica que ambas moléculas HLA presentan al sistema inmune dos subconjuntos diferentes del proteoma, pero también, con cierta frecuencia, distintos epítopos de la misma proteína.

5.2.2. Inhibición incompleta del proteasoma.

El criterio para distinguir entre ligandos sensibles y resistentes a inhibidores de proteasoma se basó en el marcaje metabólico y la cuantificación de su inhibición. Puesto que la inhibición del proteasoma disminuye drásticamente la viabilidad celular, los tiempos de marcaje fueron relativamente cortos, lo que impide llevar a cabo ensayos de marcaje homogéneo. El hecho de que la mayoría de los ligandos resistentes a inhibidores de proteasoma procedían de proteínas pequeñas y básicas sugiere que las observaciones sobre HLA-B27 tienen un significado más general. Sin embargo, las siguientes observaciones indican que la inhibición del proteasoma fue incompleta. Primero, el porcentaje de ligandos resistentes a inhibidores fue diferente con epoxomicina y con MG-132. Segundo, los ligandos resistentes a inhibidores procedentes de proteínas que no eran pequeñas y básicas correspondían a los extremos N- o C-terminales de sus proteínas parentales, que no requieren corte dual. Tercero, 5 péptidos se inhibieron más con epoxomicina que con MG-132. Cuarto, la inhibición intermedia del marcaje observado en algunos ligandos de A68 podría reflejar una regulación negativa de la síntesis de proteínas (Ding et al., 2006), pero también una inhibición incompleta del proteasoma. Quinto, la diferente susceptibilidad a los inhibidores de algunos péptidos procedentes de la misma proteína parental podría explicarse por dos mecanismos: 1) una alteración del patrón de corte del proteasoma parcialmente inhibido, que podría redundar en una sobreproducción de uno solo de los ligandos, o 2) el ligando resistente a inhibidores podría generarse a partir de fragmentos proteicos, producidos por un proteasoma parcialmente inhibido, por la acción de nardilisina y/o la oligopeptidasa thimet, como se ha sugerido recientemente (Kessler *et al.*, 2011).

La inhibición incompleta del proteasoma se demostró directamente con sustratos fluorogénicos y con precursores sintéticos de ligandos. A las concentraciones usadas *in vivo*, Los dos inhibidores bloquearon la hidrólisis del sustrato quimotríptico *in vitro* cerca del 70%. Este valor fue menor que el obtenido con células HeLa (Kisselev *et al.*, 2006) y refleja una inhibición incompleta del proteasoma 26S en C1R, ya que concentraciones de epoxomicina más altas inhibieron más eficientemente la hidrólisis de este sustrato. La epoxomicina inhibió la actividad tríptica hasta un máximo del 67%, valor también menor al obtenido en células HeLa (Kisselev *et al.*, 2006). Con MG-132 la inhibición de la actividad tríptica fue incluso menor y la inhibición parcial del proteasoma se observó directamente en los experimentos en gel. La actividad caspasa no se analizó con sustratos fluorogénicos, pero se sabe que la epoxomicina y el MG-262, un análogo de MG-132, tienen menor efecto sobre ésta que sobre las subespecificidades tríptica y quimotríptica (Kisselev *et al.*, 2006). En el mismo estudio se demostró que la eficacia de los inhibidores era dependiente del sustrato, presumiblemente en función de la

composición de aminoácidos. Así, la inhibición incompleta del proteasoma podría explicar el predominio de ligandos resistentes a inhibidores procedentes de proteínas pequeñas y básicas, ya que éstas son comparativamente ricas en residuos básicos y pueden ser particularmente susceptibles a una actividad tríptica residual.

En el trabajo previo de nuestro laboratorio con HLA-B27 (Marcilla *et al.*, 2007) no se observó un efecto dosis-respuesta en la proporción de ligandos resistentes o sensibles al inhibidor al aumentar la concentración de epoxomicina de 1.0 a 2.5 μM, lo que indujo a pensar que los ligandos resistentes a inhibidores eran independientes de proteasoma y a postular un mecanismo de generación distinto. Nuestros estudios demuestran ahora que incluso la concentración más alta usada en el trabajo anterior era insuficiente para inhibir completamente el proteasoma.

Un residuo C-terminal básico no es requisito indispensable para el patrón observado en las proteínas parentales de ligandos resistentes al inhibidor en HLA-B27, puesto que una fracción significativa de dichos ligandos en esta especificidad carecían de tal motivo (Marcilla *et al.*, 2007). Sin embargo, los ligandos de B27 poseen R en posición 2. No se conoce si se conserva el mismo patrón de resistencia a inhibidores de proteasoma en otros alotipos sin motivos de unión básicos.

5.2.3. Alteraciones de especificidad en proteasomas parcialmente inhibidos.

Con la digestión de precursores peptídicos sintéticos se abordó el estudio de las siguientes cuestiones. Primero, si los ligandos resistentes a inhibidores o sus precursores N-terminales eran producidos directamente por el proteasoma *in vivo*. Segundo, si los inhibidores bloqueaban la destrucción o la generación de ligandos de HLA o sus precursores. Tercero, si la inhibición incompleta del proteasoma resultaba en una alteración del patrón de corte y, en ese caso, si esto conduciría a un incremento en la producción de ligandos naturales o sus precursores. El hecho de que cantidades significativas de especies productivas de los ligandos de A68 y B27 fueron generados a partir del 31-mero en ausencia de inhibidores sugiere que éstos podrían ser producidos por el proteasoma *in vivo*. La inhibición incompleta del proteasoma alteró sus patrones de corte, pero estas alteraciones, aunque no bloquearon la generación de especies productivas para el 31-mero, tampoco resultaron en una sobreproducción de ligandos resistentes al inhibidor, en contraste con lo publicado en un ejemplo anterior (Wherry *et al.*, 2006).

5.2.4. Procesamiento post-proteasómico y peptidomas resistentes a inhibidores de proteasoma.

Algunos avances recientes sobre el papel de las aminopeptidasas citosólicas en la generación de ligandos de MHC podrían proporcionar una explicación para la expresión de ligandos resistentes a inhibidores de proteasoma en condiciones de actividad proteasómica disminuida. La bleomicín hidrolasa y la PSA promueven una unión peptídica óptima y mejoran la expresión en superficie de algunos alotipos (Kim et al., 2009). Además, recientemente se sugirió que algunos ligandos resistentes a inhibidores, incluyendo los de HLA-B27, podrían generarse por un procesamiento post-proteasómico de fragmentos proteicos mediante la acción combinada de nardilisina y la oligopeptidasa thimet. Ésta última enzima fue capaz de generar el extremo C-terminal básico de ligandos de HLA-A3 y A11 a partir de precursores sintéticos in vitro (Kessler *et al.*, 2011). La especificidad de corte de la nardilisina en secuencias dibásicas (Chesneau et al., 1994; Chow et al., 2000) podría ser particularmente adecuada para la generación del extremo C-terminal de péptidos que terminan en residuos básicos. De los ligandos de A68 identificados en esta Tesis (Apéndice 1, Tabla S1) 168 (20.7%) tenían un residuo C+1 básico en la proteína parental correspondiente (datos no mostrados). Sin embargo, solamente 1 (14.3%) de los 7 ligandos resistentes a inhibidores mostraban un motivo C/C+1 dibásico (ETVQLRNPPR-RQL), frente a 8 (33.3%) de los 24 ligandos sensibles a inhibidores. Esta observación no excluye la participación de la nardilisina, ya que esta enzima puede también cortar en sitios monobásicos (Chow et al., 2003).

Una posible explicación de nuestros datos que hace compatible la inhibición incompleta del proteasoma y la resistencia a inhibidores de muchos ligandos es que en la generación de éstos, el papel del procesamiento post-proteasómico podría ser especialmente significativo debido a la mayor generación de precursores más largos por proteasomas parcialmente inhibidos.

5.3. Procesamiento post-proteasómico de péptidos en el retículo endoplásmico: papel del polimorfismo de ERAP1.

En la segunda parte de esta Tesis se abordó específicamente el estudio del efecto del polimorfismo de ERAP1 en la configuración del peptidoma de HLA-B27. Este estudio estuvo motivado por el descubrimiento de la asociación de ERAP1 con EA en individuos HLA-B27 positivos (WTCC Consortium, 2007; The TASK and WTCCC2 Consortia, 2011). Aunque el efecto de la ausencia de ERAP1 sobre los peptidomas de clase I es bien conocido (Hammer *et al.*, 2007) y se han efectuado varios estudios, generalmente *in vitro*, con mutantes de ERAP1 en posiciones individuales (Goto *et al.*, 2006; Evnouchidou et al., 2011) el efecto de las variantes naturales sobre los antígenos

HLA-I *in vivo* no ha sido analizado previamente. En el contexto de HLA-B27 este análisis tiene especial importancia por su incidencia directa en el peptidoma de este alotipo y en la patogenia de EA. Se analizaron específicamente los efectos cuantitativos en la expresión de ligandos de HLA-B27, asumiendo que el procesamiento diferencial mediado por las variantes de ERAP1 puede afectar a la abundancia de muchos de estos péptidos. No se buscaron ligandos específicos cuya presencia pudiese depender de un contexto particular de ERAP1, porque la ausencia de un péptido no puede ser formalmente establecida, y no existen epítopos candidatos para la asociación a EA.

5.3.1. Consideraciones metodológicas.

La aproximación experimental planteada, aunque es apropiada para detectar efectos generales sobre el peptidoma, merece algunos comentarios. La intensidad de una señal iónica en un espectro de masas MALDI-TOF está influida por varios factores, incluyendo los parámetros instrumentales, la manipulación de la muestra, la presencia de otros componentes, etc. Por esta razón, a pesar de los estrictos controles sobre las condiciones experimentales, la RI de una señal iónica dada en dos peptidomas representa solo una estimación de cantidades relativas. Sin embargo, cuando se aplica a muchas señales iónicas, esta aproximación puede revelar tendencias generales. Así, nuestros resultados deben interpretarse como tendencias generales en grandes conjuntos de péptidos y no como una cuantificación de péptidos individuales. En trabajos previos se empleó el MALDI-TOF para buscar diferencias de expresión de ligandos de HLA-B27 en presencia o ausencia de tapasina (Purcell et al., 2001) o en diferentes subtipos de HLA-B27 (Ramos et al., 2002). Por otra parte, nuestros resultados constituyen por sí mismos una validación del método, puesto que se encontró que la variante que contenía más cambios cuyo efecto con mutantes *in vitro* era un aumento en la actividad enzimática (ver apartado 5.3.4), en JSL, expresaba predominantemente péptidos más cortos que la variante de Wewak I, que tenía cambios asociados a la disminución de la actividad *in vitro* (ver apartado 5.3.4), mientras que los péptidos más abundantes en esta última línea eran más largos. Este efecto también se observó, aunque con menor contraste, en la comparación de Wewak I con C1R-B*27:04, que refleja una situación intermedia, y desapareció en C1R-B*27:04/KNE, que tenían variantes similares del enzima.

Un segundo aspecto concierne a los sistemas de calificación empleados para evaluar la susceptibilidad de los residuos a ERAP1. El sistema usado para los residuos flanqueantes N-terminales y P1 se basó en un ensayo que analizó la presentación de antígeno en función del procesamiento por ERAP1 en el retículo endoplásmico (Hearn *et al.*, 2009) y no en la digestión *in vitro* de sustratos sintéticos, que puede no reflejar con mucha precisión los efectos *in vivo*. Solamente un estudio

publicado analizó el efecto de los residuos internos de los epítopos sobre la eficiencia de corte por ERAP1 (Evnouchidou *et al.*, 2008). Esto nos permitió examinar la influencia de la secuencia en la expresión los ligandos de HLA-B27 en el contexto del polimorfismo de ERAP1. En este estudio se emplearon librerías de nonámeros y no se analizaron todos los residuos en cada posición, lo cual supone una limitación y restringe el análisis a los nonámeros. Ambos sistemas de evaluación, aunque pueden no ser precisos para ligandos individuales, son, una vez más, adecuados para establecer tendencias generales en conjuntos grandes de péptidos.

5.3.2. Diferencias de actividad enzimática en las variantes alélicas de ERAP1.

Muchos factores independientes de ERAP1 pueden influir en la expresión relativa de ligandos de MHC-I en diferentes líneas celulares. Sin embargo, la mayoría de estos factores no deberían implicar efectos específicos sobre el tamaño de péptido. Si esto fuera así, las células que expresan moléculas de ERAP1 similares generarían peptidomas con distribuciones de tamaño similares. Este fue el caso de C1R-B*27:04 y KNE, lo cual sugiere que las diferencias significativas de tamaño observadas en el contexto de variantes distintas se debe al polimorfismo de ERAP1. En efecto, el orden de actividad deducido de los efectos en la longitud y MW de los ligandos de HLA-B27, JSL>C1R-B*27:04/KNE>Wewak I, se correlaciona con el número de aminoácidos diferentes entre las secuencias de ERAP1 es estas líneas celulares. Es poco probable que las diferencias de expresión de ERAP1 en las líneas celulares justifiquen las diferencias observadas en los peptidomas, puesto que Wewak I y C1R-B*27:04 expresaban cantidades muy similares. Por otra parte, la expresión de mRNA de ERAP1 en JSL fue similar, pero algo más baja que en las otras líneas celulares.

5.3.3. Mecanismo de interacción entre ERAP1 y HLA-B27.

ERAP1 puede eliminar casi cualquier residuo N-terminal de péptidos que superen una longitud mínima, aunque con diferentes eficiencias (Hearn *et al.*, 2009). Por lo tanto, las variantes de ERAP1 con diferente actividad enzimática podrían mostrar diferencias en la generación de ligandos de HLA-I en función de sus extensiones N-terminales flanqueantes. En un estudio previo (Schatz *et al.*, 2008) se demostró que, aunque los residuos en P-3 y posiciones más alejadas podían ser eliminados por una amplia variedad de peptidasas *in vivo*, P-2 y P-1 eran predominantemente eliminados por ERAP1 en el retículo endoplásmico. Por otra parte, las diferencias de actividad entre las variantes de ERAP1 afectarían a la destrucción de epítopos porque el residuo P1 será eliminado más eficientemente por la variante más activa. La mayor susceptibilidad al corte de los residuos flanqueantes y P1 en los

subconjuntos de péptidos predominantes en Wewak I, relativa a su equivalente en JSL, implica que los residuos en estas posiciones serían más eficientemente procesados por la variante más activa de ERAP1 en JSL, de forma que los ligandos correspondientes no solo serían generados más eficientemente, sino que también se destruirían en mayor medida en ésta línea celular. La abundancia de estos péptidos en Wewak I se explicaría por la presencia en estas células de una variante de ERAP1 menos activa, capaz de eliminar residuos flanqueantes susceptibles sin destruir extensivamente el epítopo. Muchos péptidos predominantes en JSL se generarían menos eficientemente en Wewak I debido a la baja susceptibilidad de sus residuos flanqueantes a la acción de ERAP1.

El mismo mecanismo explicaría las diferencias observadas entre Wewak I y C1R-B*27:04, pero en este caso la susceptibilidad similar de los residuos P1 en los subconjuntos de péptidos predominantes de ambas líneas celulares significa que la variante más activa en C1R-B*27:04 aún destruiría ligandos de HLA-B27 más intensamente que en Wewak I, pero menos que en JSL. Esto podría explicar por qué las diferencias de tamaño entre los peptidomas de HLA-B27 de Wewak I y C1R-B*27:04 son menores que entre Wewak I y JSL.

En conclusión, el polimorfismo natural de ERAP1 tiene una influencia significativa sobre la longitud y abundancia de muchos ligandos de HLA-B27. La variante menos activa generó mayores cantidades de péptidos largos y menos péptidos cortos, y también aumentó la expresión de péptidos de todas las longitudes con residuos susceptibles en las posiciones N-terminales flanqueantes y P1. Así, la alteración en el balance entre generación y destrucción de epítopos dependiente de la variante del enzima, en función de estos residuos, constituye un mecanismo básico por el que el polimorfismo de ERAP1 determina diferencias en los peptidomas de HLA-B27. Aunque las secuencias internas de los epítopos no afectaron globalmente su expresión relativa en diferentes contextos de ERAP1, nuestros resultados no excluyen una posible influencia de algunas secuencias individuales sobre la eficiencia de ERAP1, como se demostró *in vitro* (Evnouchidou *et al.*, 2008).

5.3.4. Polimorfismo de ERAP1 y susceptibilidad a espondilitis anquilosante.

Varios estudios analizaron el efecto del polimorfismo natural en la actividad enzimática de ERAP1. Las mutaciones K528R y R725Q, ambas con efecto protector frente a EA, inducen un descenso en la actividad de ERAP1, mientras que R127P, I276M, D575Q y Q730E tenían poco o ningún efecto (Goto *et al.*, 2006; The TASK and WTCCC2 Consortia, 2011; Kochan *et al.*, 2011). Estos estudios se llevaron a cabo con mutantes simples *in vitro* y no se analizaron los posibles efectos cooperativos debidos a la coexistencia de las mutaciones. En un estudio reciente (Evnouchidou *et al.*,

2011) se observaron cambios en la cinética enzimática de ERAP1 en función del sustrato y de la variante del enzima, lo cual podría afectar diferencialmente al procesamiento antigénico por ERAP1 *in vivo*. En este estudio, la mutación Q730E, que tiene también efecto protector frente a la enfermedad, disminuyó la presentación de un epítopo específico de HLA-B27 *in vivo* incluso con mayor eficiencia que K528R, en contraste con la actividad relativa de estos mutantes *in vitro*. La estructura tridimensional de ERAP1 (Nguyen *et al.*, 2011; Kochan *et al.*, 2011) muestra que, entre las posiciones polimórficas en las líneas celulares empleadas en esta Tesis, M349V está localizada en el sitio activo, G346D, R575Q y Q730E están en el sitio de unión del sustrato, R127P, I276M y K528R están localizadas en regiones inter-dominio y D575N está localizada en la superficie del dominio III (Figura 23). Así, estos polimorfismos podrían afectar la función de ERAP1 modulando la hidrólisis, la unión del sustrato o la transición conformacional entre los estados abierto y cerrado, que es crítica para la actividad enzimática.

ERAP1 en JSL difiere del enzima en C1R-B*27:04, KNE y Wewak I en las posiciones 127(R), 528(K) y 730(Q). Por lo tanto, la mayor actividad de ERAP1 en JSL concuerda con los efectos observados en los cambios K528R y Q730E sobre el recorte de péptidos. Aún así, a pesar de las notables diferencias entre los peptidomas de HLA-B*27:04 de C1R-B*27:04 y Wewak I, las moléculas de ERAP1 de estas líneas celulares son idénticas en las posiciones 528(R) y 730(E), lo cual indica que el polimorfismo en otras posiciones también influye en la actividad de ERAP1 *in vivo*. Por ejemplo, Wewak I difiere del resto de líneas celulares en los cambios M349V, D575N y R575Q, lo cual puede afectar la función de ERAP1, como se demostró para el último cambio en ensayos *in vitro* (The TASK and WTCCC2 Consortia, 2011). En esta Tesis no se puede determinar si las diferencias entre los peptidomas de HLA-B*27:04 se deben fundamentalmente a una única mutación o a su efecto combinado. Siempre se debería tener en cuenta que el efecto de una mutación dada en ERAP1 podría estar modulado por cambios concomitantes en otras posiciones. Es muy frecuente la coexistencia de SNPs no sinónimos asociados con EA en una variante dada de ERAP1, lo que conduce a la identificación de haplotipos asociados con la enfermedad (Harvey *et al.*, 2009; Maksymowych *et al.*, 2009; Szczypiorska *et al.*, 2011). Debido al fuerte desequilibrio de ligamiento, es difícil distinguir si la asociación con la enfermedad está determinada por una única mutación o por la coexistencia de polimorfismos en los haplotipos susceptibles. Los análisis genéticos son consistentes con un modelo de dos mutaciones en el cual la asociación de ERAP1 con EA está determinada por un efecto primario de rs30187 (K528R) y un efecto secundario de rs17482078 (R725Q) y/o rs10050860 (Q730E). Estos dos últimos polimorfismos se encuentran en un fuerte desequilibrio de ligamiento. En combinación, los

alelos protectores en estos *loci* (R528, Q725 y E730) confieren un efecto protector particularmente alto (TASC y WTCCC2 2011). De las líneas celulares empleadas en esta Tesis, solo Wewak I contiene estos tres polimorfismos protectores, mientras que C1R-B*27:04 y KNE tienen solo dos (R528 y E730) y JSL presenta los tres SNPs de susceptibilidad (K528, R725 y Q730). De esta forma, sin descartar posibles influencias adicionales debidas a otras posiciones, los efectos del polimorfismo de ERAP1 en HLA-B27 en las líneas celulares empleadas son completamente consistentes con los datos genéticos y apoyan un efecto mayoritario de las posiciones 528, 725 y 730, que es más prominente con la coexistencia en éstas de cambios asociados a la protección o al riesgo de EA.

No se analizó sistemáticamente la posible contribución de ERAP2 a las alteraciones observadas en los peptidomas de HLA-B*27:04. Se sabe que existen diferencias de expresión de este enzima en LCLs humanas (Fruci et al., 2006), sin embargo, a pesar de la menor actividad de recorte peptídico en Wewak I, el nivel de expresión en esta línea celular, analizado por Western blot, fue casi dos veces superior al de C1R-B*27:04 (datos no mostrados). No se puede descartar la posibilidad de que el polimorfismo de ERAP1 influya en la acción concertada de los complejos ERAP1/ERAP2 (Saveanu *et al.*, 2005) alterando la interacción entre ambas moléculas. No obstante, la localización de la mayoría de los SNPs en sitios susceptibles de alterar directamente la actividad hidrolítica de ERAP1 constituye un argumento en contra de una influencia significativa de esta alternativa.

Un genotipado reciente de ERAP2 en las líneas celulares empleadas en esta Tesis para el SNP 2549782 (N392K), un polimorfismo hallado en haplotipos de ERAP1/ERAP2 asociados con EA (Tsui et al., 2010) y asociado también a otras patologías (Johnson et al., 2009; Cagliani et al., 2010; Hill et al., 2011), mostró heterocigosis en C1R, KNE y Wewak I (datos no mostrados), lo que descarta un papel significativo de esta mutación en los efectos observados sobre los peptidomas de HLA-B27 en estas líneas celulares.

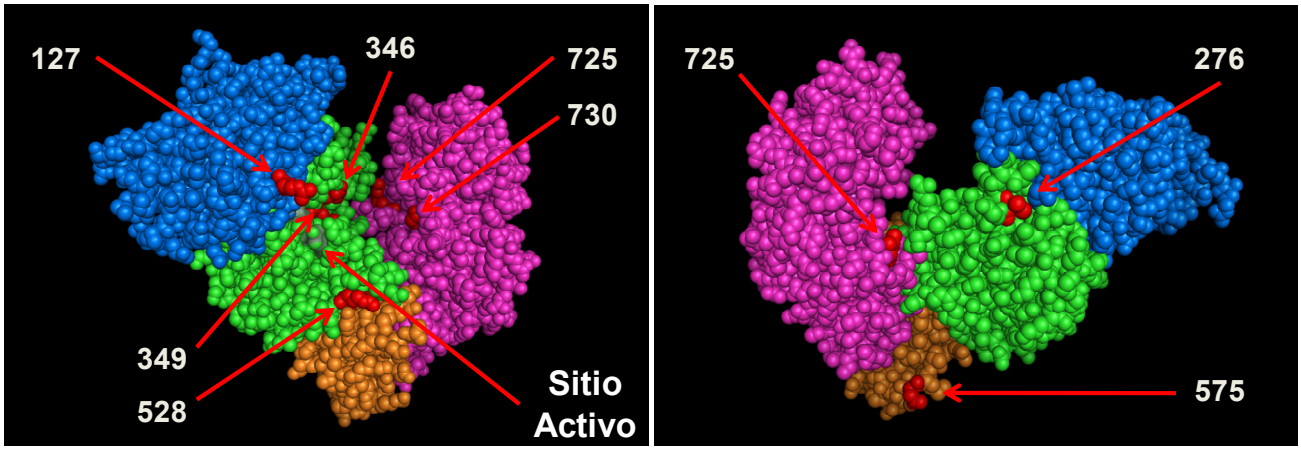


Figura 23: Localización espacial de los polimorfismos de ERAP1 en las líneas celulares usadas en esta Tesis. El residuo 349 está localizado en el sitio activo. Los residuos 346, 725 y 730 se localizan en el sitio de unión del sustrato. Los residuos 127, 276 y 528 se sitúan en zonas implicadas en el reordenamiento de los dominios durante la transición entre los estados abierto y cerrado.

5.3.5. Interacción funcional entre ERAP1 y HLA-B27 y papel patogenético de HLA-B27 en las espondiloartropatías.

Los efectos del polimorfismo de ERAP1 descritos en esta Tesis podrían tener influencia sobre varios aspectos de la biología de HLA-B27 relevantes para la patogenia de EA. En la presente discusión se comentará la incidencia de este estudio en la evaluación de las hipótesis actuales concerniente al papel patogénico de HLA-B27. En primer lugar, el polimorfismo de ERAP1 podría afectar a las propiedades inmunológicas canónicas del heterodímero de HLA-B27 a varios niveles. Los efectos diferenciales en el procesamiento observados con las variantes alélicas de ERAP1 podrían determinar que se produzcan o no algunos de los ligandos de HLA-B27. Debido a que el efecto protector frente a la enfermedad fue previamente asociado con mutaciones en ERAP1 que conllevan un descenso en la actividad enzimática (The TASK and WTCCC2 Consortia, 2011), se podría especular sobre la posibilidad de que epítopos asociados con enfermedad, generados de forma específica en el contexto de las variantes más activas, podrían ser péptidos con residuos flanqueantes y/o residuos P1 que no sean muy susceptibles la acción de ERAP1. Por otra parte, la alteración en los niveles de expresión de muchos ligandos de HLA-B27 probablemente influya en la selección del repertorio de células T, que está determinado por la avidez de la interacciones TCR/pMHC y consecuentemente, por el número de copias de epítopos individuales en la superficie celular. Por ello, el potencial autoinmune de HLA-B27 podría estar modulado por el polimorfismo de ERAP1. Finalmente, la inmunogenicidad de HLA-B27 podría estar igualmente modulada por ERAP1 a través de su efecto cuantitativo sobre el peptidoma, puesto los umbrales de activación de las células T son dependientes del número de

complejos pMHC reconocidos por el TCR durante dicha activación (Manz *et al.*, 2011).

En segundo lugar, aunque no se examinó el efecto de ERAP1 sobre el plegamiento de HLA-B27, la disminución de la termoestabilidad de esta molécula observada en Wewak I indica que las variantes que contienen múltiples polimorfismos protectores generan peptidomas de HLA-B27 subóptimos. En trabajos previos se demostró que la alta estabilidad térmica es característica de tres de los subtipos asociados a EA B*27:02, B*27:04 y B*27:05, expresados en células C1R, pero no de B*27:07 y B*1403, también asociados con enfermedad, ni de B*27:06 y B*27:09, no asociados con EA (Merino *et al.*, 2008; Galocha *et al.*, 2010). La termoestabilidad de B*27:04 en Wewak I sugiere que los polimorfismos de ERAP1 protectores proporcionan una estabilidad molecular a los subtipos de HLA-B27 asociados con EA relativamente similar a la de los subtipos no asociados. El descenso en la termoestabilidad no se correlaciona por sí mismo con la susceptibilidad a EA. Sin embargo, el mecanismo molecular que promueve la disminución de la estabilidad en B*27:06, B*27:07, B*27:09 y B*1403 es probablemente diferente del que opera en B*27:04 de Wewak I. De hecho, los cuatro subtipos anteriores tienen una subcavidad hidrofóbica (subcavidad F) y unen de forma casi exclusiva péptidos con residuos C-terminales apolares (Lopez de Castro *et al.*, 2004; Merino et al., 2005). Previamente se demostró que en subtipos de HLA-B44 la hidrofobicidad de la subacavidad F determina una alta inestabilidad de la molécula en ausencia de ligando, lo que favorece una rápida carga de péptidos hidrofóbicos para proteger las regiones apolares de la exposición al medio acuoso (Williams et al., 2002). Éste mecanismo podría implicar una incorporación de péptidos evitando su edición por tapasina, como ocurre en HLA-B*44:05, lo que promovería un rápido plegamiento pero con una carga de péptidos subóptimos. De hecho, no se observó una mal plegamiento significativo de B*27:06, B*27:07, B*27:09 y B*1403 en el retículo endoplásmico (Merino *et al.*, 2008; Galocha *et al.*, 2010).

Obviamente, la menor estabilidad de B*27:04 en Wewak I no se debe a una hidrofobicidad alterada de su sitio de unión de péptidos, ya que la molécula no está alterada. Presumiblemente el efecto puede deberse a la incapacidad de la variante de ERAP1 en éstas células para generar un repertorio peptídico óptimo. Si así fuese, los factores termodinámicos que favorecen una rápida unión de péptidos no tendrían un papel diferente en relación a otras líneas celulares B*27:04 positivas. Por lo tanto, podría no evitarse la edición por tapasina. La ausencia de un peptidoma óptimo, como consecuencia de la incapacidad de ERAP1 para generarlo, podría suponer una cinética de carga más lenta y, posiblemente, un aumento del plegamiento incorrecto de B*27:04 en el contexto de una variante de ERAP1 con múltiples polimorfismos protectores. Aunque esta predicción debe ser confirmada para HLA-B*27:04 en Wewak I, contradice la hipótesis del plegamiento incorrecto de

HLA-B27 en la patogenia de EA (Colbert, 2000) y es coherente con la asociación de B*27:07 y de B*1403 con esta enfermedad a pesar de sus bajos niveles de mal plegamiento (Merino *et al.*, 2008; Galocha *et al.*, 2010).

La expresión de homodímeros de HC de HLA-B27 en la superficie celular se propuso también como un elemento determinante en la patogenia de EA. Estos homodímeros tendrían un potencial inmunomodulador, puesto que podrían ser reconocidos por receptores leucocitarios (Allen et al., 2001; Kollnberger et al., 2002). La formación de homodímeros en membrana está asociada al reciclaje endosomal de de HLA-B27 (Kollnberger *et al.*, 2002) y presumiblemente surgen de la disociación de los heterodímeros canónicos durante este reciclaje. Si, como se muestra en esta Tesis, los polimorfismos de ERAP1 asociados con la protección frente a EA pueden promover un descenso en la estabilidad de HLA-B27, se podría esperar que, al menos en algunos casos, los cambios en ERAP1 asociados con la susceptibilidad a EA podrían dar lugar a una disminución en la expresión de homodímeros de HC de HLA-B27. La expresión de éstos puede detectarse con el anticuerpo HC10, que reconoce de forma específica HC de HLA-I no plegada. En un trabajo reciente (Haroon et al., 2012) se demostró una disminución en la expresión de moléculas reactivas con HC10 en la superficie de los macrófagos de individuos HLA-B*27:05 positivos homocigotos para el polimorfismo Q730 de ERAP1, asociado con susceptibilidad a EA. Una vez más, los efectos de los polimorfismos de ERAP1 asociados con EA sobre HLA-B27, parecen funcionar en la dirección opuesta a la que se esperaría en base a la hipótesis de la expresión de homodímeros en superficie.

En conclusión, nuestros resultados revelan efectos globales significativos de los polimorfismos naturales de ERAP1 sobre el peptidoma de HLA-B27 *in vivo* y definen sus bases moleculares. Estos efectos, a través de su influencia sobre las características inmunológicas y otras propiedades biológicas de HLA-B27, definen la naturaleza de la interacción funcional entre ambas moléculas, ponen de manifiesto el papel fundamental del peptidoma en la asociación de HLA-B27 con EA y sugieren que esta función tiene lugar, principalmente, a través de la alteración de las características inmunológicas del heterodímero de HLA-B27.

Conclusiones

1) La modulación de la actividad hidrolítica del proteasoma y de ERAP1 tiene efectos globales en la configuración de los peptidomas de HLA-I.

2) El patrón común de punto isoeléctrico y masa molecular encontrado en las proteínas precursoras de péptidos resistentes a inhibidores de proteasoma en HLA-B27 se reproduce en A68, indicando la generalidad de este efecto sobre múltiples moléculas de HLA-I.

3) En las condiciones utilizadas *in vivo*, no se logra una inhibición completa del proteasoma. Por tanto la producción de péptidos en presencia de inhibidores de esta enzima no implica necesariamente la existencia de rutas de procesamiento alternativas.

4) Sin embargo, los péptidos resistentes a inhibidores de proteasoma no se explican por una alteración del patrón de corte que favorezca su producción.

5) Estos resultados sugieren un papel prominente del procesamiento post-proteasómico mediado por proteasas citosólicas, que actuarían sobre los fragmentos proteicos producidos por proteasomas parcialmente inhibidos, en la generación de péptidos de HLA-I resistentes a inhibidores de proteasoma.

6) Las variantes naturales de ERAP1 presentan diferencias significativas en su actividad enzimática.

7) El polimorfismo de ERAP1 tiene una influencia global sobre el peptidoma de HLA-B27, afectando al tamaño, longitud y nivel de expresión de muchos ligandos.

8) Proponemos que el mecanismo de la interacción funcional entre ERAP1 y HLA-B27 es una alteración del balance entre la generación y destrucción de epítopos, que está determinada por la susceptibilidad de los residuos flanqueantes N-terminales y del residuo P1 a la acción hidrolítica de ERAP1. Dicho balance es modulado por el polimorfismo natural de esta enzima.

9) Las variantes de ERAP1 que incluyen polimorfismos asociados con riesgo de espondilitis anquilosante presentan una mayor eficiencia de corte y determinan una optimización del peptidoma, reflejada en una alta estabilidad de la molécula de HLA-B27.

10) Las variantes que incluyen polimorfismos asociados con la protección frente a esta enfermedad presentan una actividad hidrolítica disminuida, que determina un peptidoma subóptimo, péptidos más largos y una menor estabilidad de HLA-B27.

Referencias

1. Allen,R.L., Bowness,P., and McMichael,A. (1999a). The role of HLA-B27 in spondyloarthritis. *Immunogenetics,* **50,** 220-227.
2. Allen,R.L., O'Callaghan,C.A., McMichael,A.J., and Bowness,P. (1999b). Cutting edge: HLA-B27 can form a novel b2-microglobulin-free heavy chain homodimer structure. *J. Immunol.,* **162,** 5045-5048.
3. Allen,R.L., Raine,T., Haude,A., Trowsdale,J., and Wilson,M.J. (2001). Leukocyte receptor complex-encoded immunomodulatory receptors show differing specificity for alternative HLA-B27 structures. *J. Immunol.,* **167,** 5543-5547.
4. Allen,R.L. and Trowsdale,J. (2004). Recognition of classical and heavy chain forms of HLA-B27 by leukocyte receptors. *Curr. Mol. Med.,* **4,** 59-65.
5. Andreatta,M., Schafer-Nielsen,C., Lund,O., Buus,S., and Nielsen,M. (2011). NNAlign: a web-based prediction method allowing non-expert end-user discovery of sequence motifs in quantitative peptide data. *PLoS. One.,* **6,** e26781.
6. Barnstable,C.J., Bodmer,W.F., Brown,G., Galfre,G., Milstein,C., Williams,A.F., and Ziegler,A. (1978). Production of monoclonal antibodies to group A erythrocytes, HLA and other human cell surface antigens. New tools for genetic analysis. *Cell,* **14,** 9-20.
7. Beer,I., Barnea,E., Ziv,T., and Admon,A. (2004). Improving large-scale proteomics by clustering of mass spectrometry data. *Proteomics,* **4,** 950-960.
8. Ben Dror,L., Barnea,E., Beer,I., Mann,M., and Admon,A. (2010). The HLA-B*2705 peptidome. *Arthritis Rheum.,* **62,** 420-429.
9. Beninga,J., Rock,K.L., and Goldberg,A.L. (1998). Interferon-gamma can stimulate post-proteasomal trimming of the N-terminus of an antigenic peptide by inducing leucine aminopeptidase. *J. Biol. Chem.,* **273,** 18734-18742.
10. Benjamin,R. and Parham,P. (1990). Guilt by association: HLA-B27 and ankylosing spondylitis. *Immunol. Today,* **11,** 137-142.
11. Bird,L.A., Peh,C.A., Kollnberger,S., Elliott,T., McMichael,A.J., and Bowness,P. (2003). Lymphoblastoid cells express HLA-B27 homodimers both intracellularly and at the cell surface following endosomal recycling. *Eur. J. Immunol.,* **33,** 748-759.
12. Bjorkman,P.J., Saper,M.A., Samraoui,B., Bennett,W.S., Strominger,J.L., and Wiley,D.C. (1987). Structure of the human class I histocompatibility antigen, HLA-A2. *Nature,* **329,** 506-512.
13. Bouvier,M., Guo,H.C., Smith,K.J., and Wiley,D.C. (1998). Crystal structures of HLA-A*0201 complexed with antigenic peptides with either the amino- or carboxyl-terminal group substituted by a methyl group. *Proteins,* **33,** 97-106.
14. Braun,B.C., Glickman,M., Kraft,R., Dahlmann,B., Kloetzel,P.M., Finley,D., and Schmidt,M. (1999). The base of the proteasome regulatory particle exhibits chaperone-like activity. *Nat. Cell Biol.,* **1,** 221-226.
15. Brown,M.A. (2008). Breakthroughs in genetic studies of ankylosing spondylitis. *Rheumatology. (Oxford),* **47,** 132-137.
16. Cagliani,R., Riva,S., Biasin,M., Fumagalli,M., Pozzoli,U., Lo,C.S., Mazzotta,F., Piacentini,L., Bresolin,N., Clerici,M., and Sironi,M. (2010). Genetic diversity at endoplasmic reticulum aminopeptidases is maintained by balancing selection and is associated with natural resistance to HIV-1 infection. *Hum. Mol. Genet.,* **19,** 4705-4714.

17. Cascio,P., Hilton,C., Kisselev,A.F., Rock,K.L., and Goldberg,A.L. (2001). 26S proteasomes and immunoproteasomes produce mainly N-extended versions of an antigenic peptide. *EMBO J.,* **20,** 2357-2366.
18. Chang,S.C., Momburg,F., Bhutani,N., and Goldberg,A.L. (2005). The ER aminopeptidase, ERAP1, trims precursors to lengths of MHC class I peptides by a "molecular ruler" mechanism. *Proc. Natl. Acad. Sci. U. S. A,* **102,** 17107-17112.
19. Chapman,H.A. (2006). Endosomal proteases in antigen presentation. *Curr. Opin. Immunol.,* **18,** 78-84.
20. Chesneau,V., Pierotti,A.R., Barre,N., Creminon,C., Tougard,C., and Cohen,P. (1994). Isolation and characterization of a dibasic selective metalloendopeptidase from rat testes that cleaves at the amino terminus of arginine residues. *J. Biol. Chem.,* **269,** 2056-2061.
21. Chow,K.M., Csuhai,E., Juliano,M.A., St Pyrek,J., Juliano,L., and Hersh,L.B. (2000). Studies on the subsite specificity of rat nardilysin (N-arginine dibasic convertase). *J. Biol. Chem.,* **275,** 19545-19551.
22. Chow,K.M., Oakley,O., Goodman,J., Ma,Z., Juliano,M.A., Juliano,L., and Hersh,L.B. (2003). Nardilysin cleaves peptides at monobasic sites. *Biochemistry,* **42,** 2239-2244.
23. Colaert,N., Helsens,K., Martens,L., Vandekerckhove,J., and Gevaert,K. (2009). Improved visualization of protein consensus sequences by iceLogo. *Nat. Methods,* **6,** 786-787.
24. Colbert,R.A. (2000). HLA-B27 misfolding: a solution to the spondyloarthropathy conundrum? *Mol. Med. Today,* **6,** 224-230.
25. Cui,X., Hawari,F., Alsaaty,S., Lawrence,M., Combs,C.A., Geng,W., Rouhani,F.N., Miskinis,D., and Levine,S.J. (2002). Identification of ARTS-1 as a novel TNFR1-binding protein that promotes TNFR1 ectodomain shedding. *J. Clin. Invest,* **110,** 515-526.
26. Cui,X., Rouhani,F.N., Hawari,F., and Levine,S.J. (2003a). An aminopeptidase, ARTS-1, is required for interleukin-6 receptor shedding. *J. Biol. Chem.,* **278,** 28677-28685.
27. Cui,X., Rouhani,F.N., Hawari,F., and Levine,S.J. (2003b). Shedding of the type II IL-1 decoy receptor requires a multifunctional aminopeptidase, aminopeptidase regulator of TNF receptor type 1 shedding. *J. Immunol.,* **171,** 6814-6819.
28. Davidson,S.I., Wu,X., Liu,Y., Wei,M., Danoy,P.A., Thomas,G., Cai,Q., Sun,L., Duncan,E., Wang,N., Yu,Q., Xu,A., Fu,Y., Brown,M.A., and Xu,H. (2009). Association of ERAP1, but not IL23R, with ankylosing spondylitis in a Han Chinese population. *Arthritis Rheum.,* **60,** 3263-3268.
29. DeMartino,G.N., Moomaw,C.R., Zagnitko,O.P., Proske,R.J., Chu-Ping,M., Afendis,S.J., Swaffield,J.C., and Slaughter,C.A. (1994). PA700, an ATP-dependent activator of the 20 S proteasome, is an ATPase containing multiple members of a nucleotide-binding protein family. *J. Biol. Chem.,* **269,** 20878-20884.
30. Ding,Q., Dimayuga,E., Markesbery,W.R., and Keller,J.N. (2006). Proteasome inhibition induces reversible impairments in protein synthesis. *FASEB J.,* **20,** 1055-1063.
31. Elliott,T., Willis,A., Cerundolo,V., and Townsend,A. (1995). Processing of major histocompatibility class I-restricted antigens in the endoplasmic reticulum. *J. Exp. Med.,* **181,** 1481-1491.
32. Ellis,S.A., Taylor,C., and McMichael,A. (1982). Recognition of HLA-B27 and related antigens by a monoclonal antibody. *Hum. Immunol.,* **5,** 49-59.
33. Elsasser,S., Schmidt,M., and Finley,D. (2005). Characterization of the proteasome using native gel electrophoresis.

Methods Enzymol., **398,** 353-363.

34. Evnouchidou,I., Kamal,R.P., Seregin,S.S., Goto,Y., Tsujimoto,M., Hattori,A., Voulgari,P.V., Drosos,A.A., Amalfitano,A., York,I.A., and Stratikos,E. (2011). Cutting Edge: Coding single nucleotide polymorphisms of endoplasmic reticulum aminopeptidase 1 can affect antigenic peptide generation in vitro by influencing basic enzymatic properties of the enzyme. *J. Immunol.,* **186,** 1909-1913.
35. Evnouchidou,I., Momburg,F., Papakyriakou,A., Chroni,A., Leondiadis,L., Chang,S.C., Goldberg,A.L., and Stratikos,E. (2008). The internal sequence of the peptide-substrate determines its N-terminus trimming by ERAP1. *PLoS. ONE.,* **3,** e3658.
36. Ferrell,K., Wilkinson,C.R., Dubiel,W., and Gordon,C. (2000). Regulatory subunit interactions of the 26S proteasome, a complex problem. *Trends Biochem. Sci.,* **25,** 83-88.
37. Fruci,D., Ferracuti,S., Limongi,M.Z., Cunsolo,V., Giorda,E., Fraioli,R., Sibilio,L., Carroll,O., Hattori,A., Van Endert,P.M., and Giacomini,P. (2006). Expression of endoplasmic reticulum aminopeptidases in EBV-B cell lines from healthy donors and in leukemia/lymphoma, carcinoma, and melanoma cell lines. *J. Immunol.,* **176,** 4869-4879.
38. Fung,E.Y., Smyth,D.J., Howson,J.M., Cooper,J.D., Walker,N.M., Stevens,H., Wicker,L.S., and Todd,J.A. (2009). Analysis of 17 autoimmune disease-associated variants in type 1 diabetes identifies 6q23/TNFAIP3 as a susceptibility locus. *Genes Immun.,* **10,** 188-191.
39. Galocha,B. and López de Castro,J.A. (2010). Mutational Analysis Reveals a Complex Interplay of Peptide Binding and Multiple Biological Features of HLA-B27. *J. Biol. Chem.,* **285,** 39180-39190.
40. Gandhi,A., Lakshminarasimhan,D., Sun,Y., and Guo,H.C. (2011). Structural insights into the molecular ruler mechanism of the endoplasmic reticulum aminopeptidase ERAP1. *Sci. Rep.,* **1,** 186.
41. Garcia,F., Marina,A., and Lopez de Castro,J.A. (1997). Lack of carboxyl-terminal tyrosine distinguishes the B*2706-bound peptide repertoire from those of B*2704 and other HLA-B27 subtypes associated to ankylosing spondylitis. *Tissue Antigens,* **49,** 215-221.
42. Gil-Torregrosa,B.C., Castano A.R., and Del Val,M. (1998). Major histocompatibility complex class I viral antigen processing in the secretory pathway defined by the trans-Golgi network protease furin. *J. Exp. Med.,* **188,** 1105-1116.
43. Gil-Torregrosa,B.C., Castano,A.R., Lopez,D., and Del Val,M. (2000). Generation of MHC class I peptide antigens by protein processing in the secretory route by furin. *Traffic.,* **1,** 641-651.
44. Glickman,M.H. and Ciechanover,A. (2002). The ubiquitin-proteasome proteolytic pathway: destruction for the sake of construction. *Physiol Rev.,* **82,** 373-428.
45. Glickman,M.H., Rubin,D.M., Fried,V.A., and Finley,D. (1998). The regulatory particle of the Saccharomyces cerevisiae proteasome. *Mol. Cell Biol.,* **18,** 3149-3162.
46. Glickman,M.H., Rubin,D.M., Fu,H., Larsen,C.N., Coux,O., Wefes,I., Pfeifer,G., Cjeka,Z., Vierstra,R., Baumeister,W., Fried,V., and Finley,D. (1999). Functional analysis of the proteasome regulatory particle. *Mol. Biol. Rep.,* **26,** 21-28.
47. Goto,Y., Hattori,A., Ishii,Y., and Tsujimoto,M. (2006). Reduced activity of the hypertension-associated Lys528Arg mutant of human adipocyte-derived leucine aminopeptidase (A-LAP)/ER-aminopeptidase-1. *FEBS Lett.,* **580,** 1833-1838.

48. Goto,Y., Ogawa,K., Hattori,A., and Tsujimoto,M. (2011). Secretion of endoplasmic reticulum aminopeptidase 1 is involved in the activation of macrophages induced by lipopolysaccharide and interferon-gamma. *J. Biol. Chem.,* **286,** 21906-21914.
49. Groettrup,M., Standera,S., Stohwasser,R., and Kloetzel,P.M. (1997). The subunits MECL-1 and LMP2 are mutually required for incorporation into the 20S proteasome. *Proc. Natl. Acad. Sci. U. S. A,* **94,** 8970-8975.
50. Hammer,G.E., Gonzalez,F., James,E., Nolla,H., and Shastri,N. (2007). In the absence of aminopeptidase ERAAP, MHC class I molecules present many unstable and highly immunogenic peptides. *Nat. Immunol.,* **8,** 101-108.
51. Hanna,J. and Finley,D. (2007). A proteasome for all occasions. *FEBS Lett.,* **581,** 2854-2861.
52. Hanna,J., Meides,A., Zhang,D.P., and Finley,D. (2007). A ubiquitin stress response induces altered proteasome composition. *Cell,* **129,** 747-759.
53. Haroon,N., Tsui,F.W., Chiu,B., Tsui,H.W., and Inman,R.D. (2010). Serum cytokine receptors in ankylosing spondylitis: relationship to inflammatory markers and endoplasmic reticulum aminopeptidase polymorphisms. *J. Rheumatol.,* **37,** 1907-1910.
54. Haroon,N., Tsui,F.W., Uchanska-Ziegler,B., Ziegler,A., and Inman,R.D. (2012). Endoplasmic reticulum aminopeptidase 1 (ERAP1) exhibits functionally significant interaction with HLA-B27 and relates to subtype specificity in ankylosing spondylitis. *Ann. Rheum. Dis.,* **71,** 589-595.
55. Harvey,D., Pointon,J.J., Evans,D.M., Karaderi,T., Farrar,C., Appleton,L.H., Sturrock,R.D., Stone,M.A., Oppermann,U., Brown,M.A., and Wordsworth,B.P. (2009). Investigating the genetic association between ERAP1 and ankylosing spondylitis. *Hum. Mol. Genet.,* **18,** 4204-4212.
56. Hearn,A., York,I.A., and Rock,K.L. (2009). The specificity of trimming of MHC class I-presented peptides in the endoplasmic reticulum. *J. Immunol.,* **183,** 5526-5536.
57. Hill,L.D., Hilliard,D.D., York,T.P., Srinivas,S., Kusanovic,J.P., Gomez,R., Elovitz,M.A., Romero,R., and Strauss,J.F., III (2011). Fetal ERAP2 variation is associated with preeclampsia in African Americans in a case-control study. *BMC. Med. Genet.,* **12,** 64.
58. Hughes,E.A. and Cresswell,P. (1998). The thiol oxidoreductase ERp57 is a component of the MHC class I peptide-loading complex. *Curr. Biol.,* **8,** 709-712.
59. Johnson,M.P., Roten,L.T., Dyer,T.D., East,C.E., Forsmo,S., Blangero,J., Brennecke,S.P., Austgulen,R., and Moses,E.K. (2009). The ERAP2 gene is associated with preeclampsia in Australian and Norwegian populations. *Hum. Genet.,* **126,** 655-666.
60. Kamphausen,E., Kellert,C., Abbas,T., Akkad,N., Tenzer,S., Pawelec,G., Schild,H., van Endert,P., and Seliger,B. (2010). Distinct molecular mechanisms leading to deficient expression of ER-resident aminopeptidases in melanoma. *Cancer Immunol. Immunother.,* **59,** 1273-1284.
61. Kessler,J.H., Khan,S., Seifert,U., Le Gall,S., Chow,K.M., Paschen,A., Bres-Vloemans,S.A., de Ru,A., van Montfoort,N., Franken,K.L., Benckhuijsen,W.E., Brooks,J.M., van Hall,T., Ray,K., Mulder,A., Doxiadis,I.I., van Swieten,P.F., Overkleeft,H.S., Prat,A., Tomkinson,B., Neefjes,J., Kloetzel,P.M., Rodgers,D.W., Hersh,L.B., Drijfhout,J.W., van Veelen,P.A., Ossendorp,F., and Melief,C.J. (2011). Antigen processing by nardilysin and thimet oligopeptidase generates cytotoxic T cell epitopes. *Nat. Immunol.,* **12,** 45-53.
62. Khan,A.R., Baker,B.M., Ghosh,P., Biddison,W.E., and Wiley,D.C. (2000). The structure and stability of an HLA-

A*0201/octameric tax peptide complex with an empty conserved peptide-N-terminal binding site. *J. Immunol.,* **164,** 6398-6405.

63. Kim,E., Kwak,H., and Ahn,K. (2009). Cytosolic aminopeptidases influence MHC class I-mediated antigen presentation in an allele-dependent manner. *J. Immunol.,* **183,** 7379-7387.
64. Kim,K.B., Myung,J., Sin,N., and Crews,C.M. (1999). Proteasome inhibition by the natural products epoxomicin and dihydroeponemycin: insights into specificity and potency. *Bioorg. Med. Chem. Lett.,* **9,** 3335-3340.
65. Kincaid,E.Z., Che,J.W., York,I., Escobar,H., Reyes-Vargas,E., Delgado,J.C., Welsh,R.M., Karow,M.L., Murphy,A.J., Valenzuela,D.M., Yancopoulos,G.D., and Rock,K.L. (2012). Mice completely lacking immunoproteasomes show major changes in antigen presentation. *Nat. Immunol.,* **13,** 129-135.
66. Kisselev,A.F., Callard,A., and Goldberg,A.L. (2006). Importance of the different proteolytic sites of the proteasome and the efficacy of inhibitors varies with the protein substrate. *J. Biol. Chem.,* **281,** 8582-8590.
67. Knowlton,J.R., Johnston,S.C., Whitby,F.G., Realini,C., Zhang,Z., Rechsteiner,M., and Hill,C.P. (1997). Structure of the proteasome activator REGalpha (PA28alpha). *Nature,* **390,** 639-643.
68. Kochan,G., Krojer,T., Harvey,D., Fischer,R., Chen,L., Vollmar,M., von Delft,F., Kavanagh,K.L., Brown,M.A., Bowness,P., Wordsworth,P., Kessler,B.M., and Oppermann,U. (2011). Crystal structures of the endoplasmic reticulum aminopeptidase-1 (ERAP1) reveal the molecular basis for N-terminal peptide trimming. *Proc. Natl. Acad. Sci. U. S. A,* **108,** 7745-7750.
69. Kollnberger,S., Bird,L., Sun,M.Y., Retiere,C., Braud,V.M., McMichael,A., and Bowness,P. (2002). Cell-surface expression and immune receptor recognition of HLA-B27 homodimers. *Arthritis Rheum.,* **46,** 2972-2982.
70. Kollnberger,S. and Bowness,P. (2009). The role of B27 heavy chain dimer immune receptor interactions in spondyloarthritis. *Adv. Exp. Med. Biol.,* **649,** 277-285.
71. Levy,F., Burri,L., Morel,S., Peitrequin,A.L., Levy,N., Bachi,A., Hellman,U., Van den Eynde,B.J., and Servis,C. (2002). The final N-terminal trimming of a subaminoterminal proline-containing HLA class I-restricted antigenic peptide in the cytosol is mediated by two peptidases. *J. Immunol.,* **169,** 4161-4171.
72. Li,G.H., Cheung,C.L., Xiao,S.M., Lau,K.S., Gao,Y., Bow,C.H., Huang,Q.Y., Sham,P.C., and Kung,A.W. (2011). Identification of QTL genes for BMD variation using both linkage and gene-based association approaches. *Hum. Genet.,* **130,** 539-546.
73. Lindquist,J.A., Jensen,O.N., Mann,M., and Hammerling,G.J. (1998). ER-60, a chaperone with thiol-dependent reductase activity involved in MHC class I assembly. *EMBO J.,* **17,** 2186-2195.
74. Lopez de Castro,J.A., Alvarez,I., Marcilla,M., Paradela,A., Ramos,M., Sesma,L., and Vazquez,M. (2004). HLA-B27: a registry of constitutive peptide ligands. *Tissue Antigens,* **63,** 424-445.
75. Luckey,C.J., Marto,J.A., Partridge,M., Hall,E., White,F.M., Lippolis,J.D., Shabanowitz,J., Hunt,D.F., and Engelhard,V.H. (2001). Differences in the expression of human class I MHC alleles and their associated peptides in the presence of proteasome inhibitors. *J. Immunol.,* **167,** 1212-1221.
76. Maksymowych,W.P., Inman,R.D., Gladman,D.D., Reeve,J.P., Pope,A., and Rahman,P. (2009). Association of a specific ERAP1/ARTS1 haplotype with disease susceptibility in ankylosing spondylitis. *Arthritis Rheum.,* **60,** 1317-1323.
77. Manz,B.N., Jackson,B.L., Petit,R.S., Dustin,M.L., and Groves,J. (2011). T-cell triggering thresholds are modulated

by the number of antigen within individual T-cell receptor clusters. *Proc. Natl. Acad. Sci. U. S. A,* **108,** 9089-9094.

78. Marcilla,M., Cragnolini,J.J., and Lopez de Castro,J.A. (2007). Proteasome-independent HLA-B27 ligands arise mainly from small basic proteins. *Mol. Cell Proteomics,* **6,** 923-938.
79. Marcilla,M., Villasevil,E.M., and López de Castro,J.A. (2008). Tripeptidyl peptidase II is dispensable for the generation of both proteasome-dependent and proteasome-independent ligands of HLA-B27 and other class I molecules. *Eur. J. Immunol.,* **38,** 631-639.
80. McFarland,B.J. and Beeson,C. (2002). Binding interactions between peptides and proteins of the class II major histocompatibility complex. *Med. Res. Rev.,* **22,** 168-203.
81. Mear,J.P., Schreiber,K.L., Münz,C., Zhu,X., Stevanovic,S., Rammensee,H.G., Rowland-Jones,S.L., and Colbert,R.A. (1999). Misfolding of HLA-B27 as a result of its B pocket suggests a novel mechanism for its role in susceptibility to spondyloarthropathies. *J. Immunol.,* **163,** 6665-6670.
82. Medina-Aunon,J.A., Paradela,A., Macht,M., Thiele,H., Corthals,G., and Albar,J.P. (2010). Protein Information and Knowledge Extractor: Discovering biological information from proteomics data. *Proteomics,* **10,** 3262-3271.
83. Mehta,A.M., Jordanova,E.S., Kenter,G.G., Ferrone,S., and Fleuren,G.J. (2008). Association of antigen processing machinery and HLA class I defects with clinicopathological outcome in cervical carcinoma. *Cancer Immunol. Immunother.,* **57,** 197-206.
84. Merino,E., Galocha,B., Vazquez,M.N., and Lopez de Castro,J.A. (2008). Disparate folding and stability of the ankylosing spondylitis-associated HLA-B*1403 and B*2705 proteins. *Arthritis Rheum.,* **58,** 3693-3704.
85. Merino,E., Montserrat,V., Paradela,A., and Lopez de Castro,J.A. (2005). Two HLA-B14 subtypes (B*1402 and B*1403) differentially associated with ankylosing spondylitis differ substantially in peptide specificity, but have limited peptide and T-cell epitope sharing with HLA-B27. *J. Biol. Chem.,* **280,** 35868-35880.
86. Momburg,F., Roelse,J., Howard,J.C., Butcher,G.W., Hammerling,G.J., and Neefjes,J.J. (1994). Selectivity of MHC-encoded peptide transporters from human, mouse and rat. *Nature,* **367,** 648-651.
87. Morel,S., Levy,F., Burlet-Schiltz,O., Brasseur,F., Probst-Kepper,M., Peitrequin,A.L., Monsarrat,B., Van Velthoven,R., Cerottini,J.C., Boon,T., Gairin,J.E., and Van den Eynde,B.J. (2000). Processing of some antigens by the standard proteasome but not by the immunoproteasome results in poor presentation by dendritic cells. *Immunity.,* **12,** 107-117.
88. Morrice,N.A. and Powis,S.J. (1998). A role for the thiol-dependent reductase ERp57 in the assembly of MHC class I molecules. *Curr. Biol.,* **8,** 713-716.
89. Murata,S., Sasaki,K., Kishimoto,T., Niwa,S., Hayashi,H., Takahama,Y., and Tanaka,K. (2007). Regulation of CD8+ T cell development by thymus-specific proteasomes. *Science,* **316,** 1349-1353.
90. Nag,B., Wada,H.G., Passmore,D., Clark,B.R., Sharma,S.D., and McConnell,H.M. (1993). Purified beta-chain of MHC class II binds to CD4 molecules on transfected HeLa cells. *J. Immunol.,* **150,** 1358-1364.
91. Nguyen,T.T., Chang,S.C., Evnouchidou,I., York,I.A., Zikos,C., Rock,K.L., Goldberg,A.L., Stratikos,E., and Stern,L.J. (2011). Structural basis for antigenic peptide precursor processing by the endoplasmic reticulum aminopeptidase ERAP1. *Nat. Struct. Mol. Biol.,* **18,** 604-613.
92. Nielsen,M., Lundegaard,C., and Lund,O. (2007). Prediction of MHC class II binding affinity using SMM-align, a novel stabilization matrix alignment method. *BMC. Bioinformatics.,* **8,** 238.

93. Orr,H.T., Lopez de Castro,J.A., Lancet,D., and Strominger,J.L. (1979). Complete amino acid sequence of a papain-solubilized human histocompatibility antigen, HLA-B7. 2. Sequence determination and search for homologies. *Biochemistry,* **18,** 5711-5720.

94. Ortmann,B., Copeman,J., Lehner,P.J., Sadasivan,B., Herberg,J.A., Grandea,A.G., Riddell,S.R., Tampe,R., Spies,T., Trowsdale,J., and Cresswell,P. (1997). A critical role for tapasin in the assembly and function of multimeric MHC class I-TAP complexes. *Science,* **277,** 1306-1309.

95. Pamer,E. and Cresswell,P. (1998). Mechanisms of MHC class I-restricted antigen processing. *Annu. Rev. Immunol.,* **16,** 323-358.

96. Paradela,A., Alvarez,I., Garcia-Peydro,M., Sesma,L., Ramos,M., Vazquez,J., and Lopez de Castro,J.A. (2000). Limited diversity of peptides related to an alloreactive T cell epitope in the HLA-B27-bound peptide repertoire results from restrictions at multiple steps along the processing-loading pathway. *J. Immunol.,* **164,** 329-337.

97. Parmentier,N., Stroobant,V., Colau,D., de Diesbach,P., Morel,S., Chapiro,J., van Endert,P., and Van den Eynde,B.J. (2010). Production of an antigenic peptide by insulin-degrading enzyme. *Nat. Immunol.,* **11,** 449-454.

98. Paulsson,K. and Wang,P. (2003). Chaperones and folding of MHC class I molecules in the endoplasmic reticulum. *Biochim. Biophys. Acta,* **1641,** 1-12.

99. Peaper,D.R., Wearsch,P.A., and Cresswell,P. (2005). Tapasin and ERp57 form a stable disulfide-linked dimer within the MHC class I peptide-loading complex. *EMBO J.,* **24,** 3613-3623.

100. Peters,J.M., Franke,W.W., and Kleinschmidt,J.A. (1994). Distinct 19 S and 20 S subcomplexes of the 26 S proteasome and their distribution in the nucleus and the cytoplasm. *J. Biol. Chem.,* **269,** 7709-7718.

101. Purcell,A.W., Gorman,J.J., Garcia-Peydro,M., Paradela,A., Burrows,S.R., Talbo,G.H., Laham,N., Peh,C.A., Reynolds,E.C., Lopez de Castro,J.A., and McCluskey,J. (2001). Quantitative and qualitative influences of tapasin on the class I peptide repertoire. *J. Immunol.,* **166,** 1016-1027.

102. Qian,S.B., Reits,E., Neefjes,J., Deslich,J.M., Bennink,J.R., and Yewdell,J.W. (2006). Tight linkage between translation and MHC class I peptide ligand generation implies specialized antigen processing for defective ribosomal products. *J. Immunol.,* **177,** 227-233.

103. Rabl,J., Smith,D.M., Yu,Y., Chang,S.C., Goldberg,A.L., and Cheng,Y. (2008). Mechanism of gate opening in the 20S proteasome by the proteasomal ATPases. *Mol. Cell,* **30,** 360-368.

104. Ramos,M., Paradela,A., Vazquez,M., Marina,A., Vazquez,J., and Lopez de Castro,J.A. (2002). Differential association of HLA-B*2705 and B*2709 to ankylosing spondylitis correlates with limited peptide subsets but not with altered cell surface stability. *J. Biol. Chem.,* **277,** 28749-28756.

105. Rock,K.L. and Goldberg,A. (1999). Degradation of cell proteins and the generation of MHC class I-presented peptides. *Annu. Rev. Immunol.,* **17,** 739-779.

106. Rock,K.L., Gramm,C., Rothstein,L., Clark,K., Stein,R., Dick,L., Hwang,D., and Goldberg,A.L. (1994). Inhibitors of the proteasome block the degradation of most cell proteins and the generation of peptides presented on MHC class I molecules. *Cell,* **78,** 761-771.

107. Rock,K.L., York,I.A., and Goldberg,A.L. (2004). Post-proteasomal antigen processing for major histocompatibility complex class I presentation. *Nat. Immunol.,* **5,** 670-677.

108. Rock,K.L., York,I.A., Saric,T., and Goldberg,A.L. (2002). Protein degradation and the generation of MHC class I-

presented peptides. *Adv. Immunol.,* **80,** 1-70.

109. Sadasivan,B., Lehner,P.J., Ortmann,B., Spies,T., and Cresswell,P. (1996). Roles for calreticulin and a novel glycoprotein, tapasin, in the interaction of MHC class I molecules with TAP. *Immunity,* **5,** 103-114.
110. Salter,R.D., Benjamin,R.J., Wesley,P.K., Buxton,S.E., Garrett,T.P., Clayberger,C., Krensky,A.M., Norment,A.M., Littman,D.R., and Parham,P. (1990). A binding site for the T-cell co-receptor CD8 on the alpha 3 domain of HLA-A2. *Nature,* **345,** 41-46.
111. Saveanu,L., Carroll,O., Lindo,V., Del Val,M., Lopez,D., Lepelletier,Y., Greer,F., Schomburg,L., Fruci,D., Niedermann,G., and Van Endert,P.M. (2005). Concerted peptide trimming by human ERAP1 and ERAP2 aminopeptidase complexes in the endoplasmic reticulum. *Nat. Immunol.,* **6,** 689-697.
112. Schatz,M.M., Peters,B., Akkad,N., Ullrich,N., Martinez,A.N., Carroll,O., Bulik,S., Rammensee,H.G., van Endert,P., Holzhutter,H.G., Tenzer,S., and Schild,H. (2008). Characterizing the N-terminal processing motif of MHC class I ligands. *J. Immunol.,* **180,** 3210-3217.
113. Seifert,U., Bialy,L.P., Ebstein,F., Bech-Otschir,D., Voigt,A., Schroter,F., Prozorovski,T., Lange,N., Steffen,J., Rieger,M., Kuckelkorn,U., Aktas,O., Kloetzel,P.M., and Kruger,E. (2010). Immunoproteasomes preserve protein homeostasis upon interferon-induced oxidative stress. *Cell,* **142,** 613-624.
114. Seifert,U., Marañón,C., Shmueli,A., Desoutter,J.F., Wesoloski,L., Janek,K., Henklein,P., Diescher,S., Andrieu,M., de la Salle,H., Weinschenk,T., Schild,H., Laderach,D., Galy,A., Haas,G., Kloetzel,P.M., Reiss,Y., and Hosmalin,A. (2003). An essential role for tripeptidyl peptidase in the generation of an MHC class I epitope. *Nature Immunology,* **4,** 375-379.
115. Serwold,T., Gonzalez,F., Kim,J., Jacob,R., and Shastri,N. (2002). ERAAP customizes peptides for MHC class I molecules in the endoplasmic reticulum. *Nature,* **419,** 480-483.
116. Shen,L., Sigal,L.J., Boes,M., and Rock,K.L. (2004). Important role of cathepsin S in generating peptides for TAP-independent MHC class I crosspresentation in vivo. *Immunity.,* **21,** 155-165.
117. Shen,X.Z., Billet,S., Lin,C., Okwan-Duodu,D., Chen,X., Lukacher,A.E., and Bernstein,K.E. (2011). The carboxypeptidase ACE shapes the MHC class I peptide repertoire. *Nat. Immunol.,* **12,** 1078-1085.
118. Sijts,A.J., Ruppert,T., Rehermann,B., Schmidt,M., Koszinowski,U., and Kloetzel,P.M. (2000). Efficient generation of a hepatitis B virus cytotoxic T lymphocyte epitope requires the structural features of immunoproteasomes. *J. Exp. Med.,* **191,** 503-514.
119. Stoltze,L., Schirle,M., Schwarz,G., Schroter,C., Thompson,M.W., Hersh,L.B., Kalbacher,H., Stevanovic,S., Rammensee,H.G., and Schild,H. (2000). Two new proteases in the MHC class I processing pathway. *Nat. Immunol.,* **1,** 413-418.
120. Strange,A., Capon,F., Spencer,C.C., Knight,J., Weale,M.E., Allen,M.H., Barton,A., Band,G., Bellenguez,C., Bergboer,J.G., Blackwell,J.M., Bramon,E., Bumpstead,S.J., Casas,J.P., Cork,M.J., Corvin,A., Deloukas,P., Dilthey,A., Duncanson,A., Edkins,S., Estivill,X., Fitzgerald,O., Freeman,C., Giardina,E., Gray,E., Hofer,A., Huffmeier,U., Hunt,S.E., Irvine,A.D., Jankowski,J., Kirby,B., Langford,C., Lascorz,J., Leman,J., Leslie,S., Mallbris,L., Markus,H.S., Mathew,C.G., McLean,W.H., McManus,R., Mossner,R., Moutsianas,L., Naluai,A.T., Nestle,F.O., Novelli,G., Onoufriadis,A., Palmer,C.N., Perricone,C., Pirinen,M., Plomin,R., Potter,S.C., Pujol,R.M., Rautanen,A., Riveira-Munoz,E., Ryan,A.W., Salmhofer,W., Samuelsson,L., Sawcer,S.J., Schalkwijk,J.,

Smith,C.H., Stahle,M., Su,Z., Tazi-Ahnini,R., Traupe,H., Viswanathan,A.C., Warren,R.B., Weger,W., Wolk,K., Wood,N., Worthington,J., Young,H.S., Zeeuwen,P.L., Hayday,A., Burden,A.D., Griffiths,C.E., Kere,J., Reis,A., McVean,G., Evans,D.M., Brown,M.A., Barker,J.N., Peltonen,L., Donnelly,P., and Trembath,R.C. (2010). A genome-wide association study identifies new psoriasis susceptibility loci and an interaction between HLA-C and ERAP1. *Nat. Genet.,* **42,** 985-990.

121. Szczypiorska,M., Sanchez,A., Bartolome,N., Arteta,D., Sanz,J., Brito,E., Fernandez,P., Collantes,E., Martinez,A., Tejedor,D., Artieda,M., and Mulero,J. (2011). ERAP1 polymorphisms and haplotypes are associated with ankylosing spondylitis susceptibility and functional severity in a Spanish population. *Rheumatology. (Oxford),* **50,** 1969-1975.
122. Tanahashi,N., Murakami,Y., Minami,Y., Shimbara,N., Hendil,K.B., and Tanaka,K. (2000). Hybrid proteasomes. Induction by interferon-gamma and contribution to ATP-dependent proteolysis. *J. Biol. Chem.,* **275,** 14336-14345.
123. Tanaka,K. (2009). The proteasome: overview of structure and functions. *Proc. Jpn. Acad. Ser. B Phys. Biol. Sci.,* **85,** 12-36.
124. Tanioka,T., Hattori,A., Masuda,S., Nomura,Y., Nakayama,H., Mizutani,S., and Tsujimoto,M. (2003). Human leukocyte-derived arginine aminopeptidase: the third memeber of the oxytocinase subfamily of aminopeptidases. *J. Biol. Chem.,* **278,** 32275-32283.
125. Taranta,A., Gianviti,A., Palma,A., De,L., V, Mannucci,L., Procaccino,M.A., Ghiggeri,G.M., Caridi,G., Fruci,D., Ferracuti,S., Ferretti,A., Pecoraro,C., Gaido,M., Penza,R., Edefonti,A., Murer,L., Tozzi,A.E., and Emma,F. (2009). Genetic risk factors in typical haemolytic uraemic syndrome. *Nephrol. Dial. Transplant.,* **24,** 1851-1857.
126. The TASK and WTCCC2 Consortia (2011). Interaction between ERAP1 and HLA-B27 in ankylosing spondylitis implicates peptide handling in the mechanism for HLA-B27 in disease susceptibility. *Nat. Genet.,* **43,** 761-767.
127. Thrower,J.S., Hoffman,L., Rechsteiner,M., and Pickart,C.M. (2000). Recognition of the polyubiquitin proteolytic signal. *EMBO J.,* **19,** 94-102.
128. Towne,C.F., York,I.A., Neijssen,J., Karow,M.L., Murphy,A.J., Valenzuela,D.M., Yancopoulos,G.D., Neefjes,J.J., and Rock,K.L. (2005). Leucine aminopeptidase is not essential for trimming peptides in the cytosol or generating epitopes for MHC class I antigen presentation. *J. Immunol.,* **175,** 6605-6614.
129. Tsui,F.W., Haroon,N., Reveille,J.D., Rahman,P., Chiu,B., Tsui,H.W., and Inman,R.D. (2010). Association of an ERAP1 ERAP2 haplotype with familial ankylosing spondylitis. *Ann. Rheum. Dis.,* **69,** 733-736.
130. Van Endert,P.M. (1996). Peptide selection for presentation by HLA class I: a role for the human transporter associated with antigen processing? *Immunol. Res.,* **15,** 265-279.
131. Van Endert,P.M., Riganelli,D., Greco,G., Fleischhauer,K., Sidney,J., Sette,A., and Bach,J.F. (1995). The peptide-binding motif for the human transporter associated with antigen processing. *J. Exp. Med.,* **182,** 1883-1895.
132. Van Endert,P.M., Tampe,R., Meyer,T.H., Tisch,R., Bach,J.F., and McDevitt,H.O. (1994). A sequential model for peptide binding and transport by the transporters associated with antigen processing. *Immunity.,* **1,** 491-500.
133. Vassilakos,A., Cohen-Doyle,M.F., Peterson,P.A., Jackson,M.R., and Williams,D.B. (1996). The molecular chaperone calnexin facilitates folding and assembly of class I histocompatibility molecules. *EMBO J.,* **15,** 1495-1506.
134. Vigneron,N., Peaper,D.R., Leonhardt,R.M., and Cresswell,P. (2009). Functional significance of tapasin membrane

association and disulfide linkage to ERp57 in MHC class I presentation. *Eur. J. Immunol.,* **39,** 2371-2376.

135. Vinitsky,A., Cardozo,C., Sepp-Lorenzino,L., Michaud,C., and Orlowski,M. (1994). Inhibition of the proteolytic activity of the multicatalytic proteinase complex (proteasome) by substrate-related peptidyl aldehydes. *J. Biol. Chem.,* **269,** 29860-29866.
136. Voges,D., Zwickl,P., and Baumeister,W. (1999). The 26S proteasome: a molecular machine designed for controlled proteolysis. *Annu. Rev. Biochem.,* **68,** 1015-1068.
137. Voigt,A., Salzmann,U., Seifert,U., Dathe,M., Soza,A., Kloetzel,P.M., and Kuckelkorn,U. (2007). 20S proteasome-dependent generation of an IEpp89 murine cytomegalovirus-derived H-2L(d) epitope from a recombinant protein. *Biochem. Biophys. Res. Commun.,* **355,** 549-554.
138. Vos,J.C., Spee,P., Momburg,F., and Neefjes,J. (1999). Membrane topology and dimerization of the two subunits of the transporter associated with antigen processing reveal a three-domain structure. *J. Immunol.,* **163,** 6679-6685.
139. Wearsch,P.A. and Cresswell,P. (2008). The quality control of MHC class I peptide loading. *Curr. Opin. Cell Biol.,* **20,** 624-631.
140. Wherry,E.J., Golovina,T.N., Morrison,S.E., Sinnathamby,G., McElhaugh,M.J., Shockey,D.C., and Eisenlohr,L.C. (2006). Re-evaluating the generation of a "proteasome-independent" MHC class I-restricted CD8 T cell epitope. *J. Immunol.,* **176,** 2249-2261.
141. Williams,A.P., Peh,C.A., Purcell,A.W., McCluskey,J., and Elliott,T. (2002). Optimization of the MHC class I peptide cargo is dependent on tapasin. *Immunity.,* **16,** 509-520.
142. WTCC Consortium (2007). Association scan of 14,500 nonsynonymous SNPs in four diseases identifies autoimmunity variants. *Nat. Genet.,* **39,** 1329-1337.
143. Yamamoto,N., Nakayama,J., Yamakawa-Kobayashi,K., Hamaguchi,H., Miyazaki,R., and Arinami,T. (2002). Identification of 33 polymorphisms in the adipocyte-derived leucine aminopeptidase (ALAP) gene and possible association with hypertension. *Hum. Mutat.,* **19,** 251-257.
144. Yewdell,J.W. and Bennink,J.R. (2001). Cut and trim: generating MHC class I peptide ligands. *Curr. Opin. Immunol.,* **13,** 13-18.
145. York,I.A., Chang,S.C., Saric,T., Keys,J.A., Favreau,J.M., Goldberg,A.L., and Rock,K.L. (2002). The ER aminopeptidase ERAP1 enhances or limits antigen presentation by trimming epitopes to 8-9 residues. *Nat. Immunol.,* **3,** 1177-1184.
146. Zemmour,J., Little,A.M., Schendel,D.J., and Parham,P. (1992). The HLA-A,B "negative" mutant cell line C1R expresses a novel HLA-B35 allele, which also has a point mutation in the translation initiation codon. *J. Immunol.,* **148,** 1941-1948.
147. Zervoudi,E., Papakyriakou,A., Georgiadou,D., Evnouchidou,I., Gajda,A., Poreba,M., Salvesen,G.S., Drag,M., Hattori,A., Swevers,L., Vourloumis,D., and Stratikos,E. (2011). Probing the S1 specificity pocket of the aminopeptidases that generate antigenic peptides. *Biochem. J.,* **435,** 411-420.
148. Zwickl,P. (2002). The 20S proteasome. *Curr. Top. Microbiol. Immunol.,* **268,** 23-41.

Apéndice I

*The Origin of Protesome Inhibitor-Resistant HLA Class I Peptidomes: a Study with HLA-A*68:01.*

Molecular and Cellular Proteomics, Vol. 11: M111 (2012).

The Origin of Proteasome-inhibitor Resistant HLA Class I Peptidomes: a Study With HLA-A*68:01*[S]

Noel García-Medel‡, Alejandro Sanz-Bravo‡, Eilon Barnea§, Arie Admon§, and José A. López de Castro‡¶

Some HLA class I molecules bind a significant fraction of their constitutive peptidomes in the presence of proteasome inhibitors. In this study, A*68:01-bound peptides, and their parental proteins, were characterized through massive mass spectrometry sequencing to refine its binding motif, including the nearly exclusive preference for C-terminal basic residues. Stable isotope tagging was used to distinguish proteasome-inhibitor sensitive and resistant ligands. The latter accounted for less than 20% of the peptidome and, like in HLA-B27, arose predominantly from small and basic proteins. Under the conditions used for proteasome inhibition *in vivo*, epoxomicin and MG-132 incompletely inhibited the hydrolysis of fluorogenic substrates specific for the tryptic or for both the tryptic and chymotryptic subspecificities, respectively. This incomplete inhibition was also reflected in the cleavage of synthetic peptide precursors of A*68:01 ligands. For these substrates, the inhibition of the proteasome resulted in altered cleavage patterns. However these alterations did not upset the balance between cleavage at peptide bonds resulting in epitope destruction and those leading to their generation. The results indicate that inhibitor-resistant HLA class I ligands are not necessarily produced by non-proteasomal pathways. However, their generation is not simply explained by decreased epitope destruction upon incomplete proteasomal inhibition and may require additional proteolytic steps acting on incompletely processed proteasomal products. ***Molecular & Cellular Proteomics 11: 10.1074/mcp.M111.011486, 1–15, 2012.***

Major Histocompatibility complex class I (MHC-I)[1] molecules constitutively bind and present at the cell surface large peptide repertoires for recognition by $CD8^+$ T cells. These peptides arise mainly from proteasomal degradation of endogenous proteins (1). Alternative pathways seem to have a limited contribution, although they generate some of these ligands (2–12). The proteasome is a complex protease located in the cytosol and nucleus. Its catalytic core, designated as the 20S proteasome, consists of a four-ring barrel structure. Each of the two external rings is formed by 7 noncatalytic α subunits, whereas the two inner rings contain seven β subunits, three of which, β1, β2, and β5 are catalytic. These three constitutive subunits can be replaced, upon γ-interferon stimulation, by homologous inducible subunits, to form the immunoproteasome (13). The 20S core binds the PA700 activator to form the 26S proteasome, which carries out the ATP-dependent degradation of ubiquitilated proteins, and can also interact with the PA28 regulator, which increases dual cleavage (14, 15). Although the proteasome can cleave virtually any peptide bond, three major subspecificities are distinguished: tryptic-like, after basic residues, chymotryptic-like, after aliphatic/aromatic residues, and caspase-like, after acidic residues, which can be assigned to β2, β5, and β1, respectively (16–20).

MHC-I molecules differ widely in their capacity to bind and present self-derived ligands in the presence of proteasome-inhibitors (21, 22). We previously reported that a significant fraction of the HLA-B27-bound peptidome was efficiently presented in cells treated with epoxomicin and MG-132 in a way essentially unaffected by the concentration of inhibitor. These inhibitor-resistant peptides predominantly arose from small basic proteins (23). This study was consistent with an undefined proteasome-independent proteolytic pathway with preference for this type of proteins. However, there were several problems with this interpretation. The inhibition of the proteasome was not directly measured in our previous study. Total inhibition of this enzyme is difficult to achieve (24), and degradation of small basic proteins might be particularly unaffected if the tryptic-like subspecificity is incompletely inhibited. Moreover, incomplete inhibition of the proteasome may lead to altered cleavage, because impairment of a catalytic site may allosterically modulate the activity of other sites (25). Because proteasomes not only produce MHC ligands and their precursors, but also degrade them by hydrolyzing their internal peptide bonds, altered cleavage by partially inhibited

From the ‡Centro de Biología Molecular *Severo Ochoa* (Consejo Superior de Investigaciones Científicas and Universidad Autónoma de Madrid), c/ Nicolas Cabrera N.1, Universidad Autónoma, 28049 Madrid, Spain, and §Faculty of Biology, Technion - Israel Institute of Technology, Haifa 32000, Israel

Received May 25, 2011, and in revised form, September 27, 2011
Published, MCP Papers in Press, October 3, 2011, DOI 10.1074/mcp.M111.011486

[1] The abbreviations used are: MHC-I, Major Histocompatibility Complex class I; mAb, monoclonal antibody; FBS, fetal bovine serum; MG-132, carbobenzoxy-L-leucyl-L-leucyl-L-leucinal; TFA, trifluoroacetic acid; *m/z*, mass-to-charge; MALDI-TOF, Matrix-assisted laser desorption/ionization time-of-flight; MW, molecular mass; pI, Isoelectric point; P, position.

proteasomes may increase presentation of some ligands by favoring their production and/or decreasing their degradation (26).

To address these issues, we focused on the origin and proteasome-dependence of the HLA-A*68:01-bound peptidome. Like HLA-B27, the surface expression of HLA-A68 is relatively high in the presence of proteasome inhibitors (22). In addition, this allotype binds almost exclusively peptides with C-terminal basic residues, so that it might be particularly susceptible to incomplete inhibition of the proteasomal tryptic-like activity (24). Our study was designed as follows. First, the peptide motif of A*68:01, currently based on a limited set of ligands (http://www.syfpeithi.de), was refined through extensive sequencing of A*68:01-bound peptides, and their presumed parental proteins were identified. Second, by using stable isotope tagging, the percentage of proteasome inhibitor sensitive and resistant peptides was determined and multiple members of each subset, as well as their parental proteins, were identified. Third, the inhibition of the tryptic and chymotryptic-like proteasomal subspecificities in the conditions used for the analysis of the A*68:01 peptidome was quantified with fluorogenic substrates. Four, the effect of proteasome inhibitors on the generation and destruction of specific A*68:01 ligands and on altering cleavage patterns was analyzed with synthetic peptide substrates.

EXPERIMENTAL PROCEDURES

Cell Lines, Monoclonal Antibodies (mAb), and Inhibitors—C1R is a human lymphoid cell line with low expression of its endogenous HLA class I molecules (27). C1R-A*68:01 transfectants (22) were a kind gift of Dr. V. Engelhard (University of Virginia School of Medicine, USA). Cells were cultured in Roswell Park Memorial Institute 1640 medium supplemented with 10% fetal bovine serum (FBS) (both from Invitrogen, Paisley, UK). The mAb W6/32 (IgG2a; specific for a monomorphic HLA class I determinant) (28) was used. Epoxomicin, an irreversible and specific inhibitor of the proteasome (29) and carbobenzoxy-L-leucyl-L-leucyl-L-leucinal (MG-132) (1), a potent reversible inhibitor of the proteasome and calpains, were from Calbiochem (Schwalbach, Germany).

Electrospray-Orbitrap MS Analysis—Peptide mixtures were desalted and concentrated with Micro-Tip reverse-phase columns (C_{18}, 200 μl, Harvard Apparatus, Holliston, MA). Each C_{18} tip was prepared by washing with 80% acetonitrile in 0.1% trifluoroacetic acid (TFA) followed by pre-equilibration with two washes with 150 μl of 0.1% TFA and then loaded with the peptide mixture. The C_{18} tips were washed twice again with 0.1% TFA, the peptides were eluted with 80% acetonitrile in 0.1% TFA, and concentrated to ~18 μl using vacuum centrifugation.

The HLA class I peptides were analyzed by μLC-MS/MS using an Orbitrap XL mass spectrometer (Thermo Fisher, San Jose, CA) fitted with a capillary high-performance liquid chromatography (HPLC) (Eksigent, Dublin, CA). The peptides were resolved on a C_{18} trap column (0.3 × 5 mm, LC-Packings) connected on-line to a 75 micron I.D. fused silica capillaries (J&W, Folsom, CA) self-packed with 3.5 micron Reprosil C_{18} (Dr. Maisch, GmbH, Germany) as in (30), and eluted at flow rates of 0.25 μl/min, with linear gradients of 7–40% acetonitrile in 0.1% formic acid, during 90 min, followed by 15 min at 95% acetonitrile in 0.1% formic acid. The spectra were collected in the orbitrap mass analyzer using full ion scan mode over the mass-to-charge (*m/z*) range 400–2,000, which was set to 60,000 resolutions. The most intense seven masses from each full mass spectrum, with single, double and triple charge states, were selected for fragmentation by collision-induced disintegration in the linear ion-trap

Database Searches—Pep-Miner (31) was used for generation of the peak-lists based on the μLC-MS/MS data. The peptides were identified using multiple search engines: Pep-Miner, Sequest (Thermo-Fisher) (32) and Mascot (server 2.2, Matrix Science Inc. Boston, MA) (33) searched against the human part of the Uniprot database (http://www.uniprot.org, Jan 2009) including 20,332 proteins. The Sequest and Mascot search results were combined into one report by Proteome Discoverer 1.0 SP1 (Thermo-Fisher). The search was not limited by enzymatic specificity, the peptide tolerance was set to 0.01 Da. and the fragment ion tolerance was set to 0.5 Da. Oxidized methionine was searched as a variable modification.

Peptide identifications were accepted if their masses were below 1500 Da. and mass accuracy was better than 0.005 Da. Other required criteria were: Pep-Miner Score ≥80, Sequest Xcore ≥2.5, Sequest Probability ≥5, Mascot ionScore/Mascot high score ≥0.6, Mascot Expectation value ≥0.1, delta to next score = 0. Peptides were filtered also according to their fitness to the HLA-A68 consensus and were accepted if their HLA-A68 score was ≥20 (according to http://www-bimas.cit.nih.gov/molbio/hla_bind/) (34). Some peptides were also observed with different charge-states and as overlapping, longer or shorter peptides by one amino acid, listed in the table as variants. The false discovery rate of the A68 peptides adhering to at least three or more of these criteria was <3.9%. This was calculated by running the same data against the corresponding randomized database and calculating the ratio between the results from the random and Uniprot databases.

Matrix-assisted Laser Desorption-Ionization Time-of-Flight (MALDI-TOF) MS—This was done in a MALDI-TOF/TOF instrument (4800 Proteomics Analyzer, Applied Biosystems, Foster City, CA) as previously described (35). MS data were acquired in the mass range 800–2000 Da (or 700–4000 for synthetic substrate digests) in reflector positive mode at 25kV and analyzed using the Data Explorer software version 4.9 (Applied Biosystems). Each spectrum was externally calibrated with the Peptide Calibration Standard Mixture (Brucker, Product # 206195) to reach a typical mass measurement accuracy of <25 ppm.

MALDI-TOF/TOF MS/MS spectra were acquired with the same software at 1Kv, using collision-induced dissociation with atmospheric air, and a precursor mass window of ±2.5 Da. A signal-to-noise ratio of three was used for processing data. Interpretation of the MS/MS spectra was done manually, but assisted by various tools as follows. Manual inspection of the spectrum, which was facilitated by using a simple tool implemented as an Excel macro, usually allowed us to derive a tentative sequence. This was used to screen the human proteome in the human protein entries of the Uniprot/Swiss-prot database (Release 14.0: 2008/07/22, with 19329 entries), using a window of 0.5 *m/z* units for both precursor and fragment ions, for a possible match using the Mascot server 2.2 software. For those sequences showing the highest scores in this preliminary search, the MS-product tool (version 5.1.8) at http://prospector.ucsf.edu (University of California, San Francisco), which generates a list of theoretical fragment ions from a given peptide sequence, was used to match the candidate sequences to our experimental MS/MS spectra. Alternatively, a new software (designated as MSgene) was used for *de novo* generation of candidate sequences. This tool is based on a genetic algorithm in which an initial population of random peptide sequences encoded in 5-bit codon units, is subjected to an iterative process of optimization, by means of recombination and mutation, leading to convergence toward candidate sequences with improved fit to the experimental MS/MS spectrum (supplemental Fig. S1). These candi-

date sequences were subsequently analyzed with the MS-product tool as above. Some of these sequences were confirmed with the corresponding synthetic peptides.

Isotopic Labeling—This was done as previously described (23) with minor modifications. Briefly, C1R-A*68:01 transfectants (about 5×10^8 cells per flask) were incubated for 4 h in Roswell Park Memorial Institute 1640 medium without Arg or, in some experiments, Lys, supplemented with 10% dialyzed FBS. One flask was then supplemented with standard (^{14}N) Arg or Lys (100 μg/ml), a second one with 100 μg/ml of L-Arg-guanido[$^{15}N_2$] ·HCl (Cambridge Isotope Laboratories, Andover, MA), or L-Lys-[$^{15}N_2$]·HCl, both reagents containing two ^{15}N atoms, and the third flask was treated with 20 μM MG-132 for 30 min prior to the addition of the ^{15}N-tagged amino acid, and left for the entire labeling period. After 5 h, the cells were washed twice in 20 mM Tris-HCl, 150 mM NaCl, pH 7.5. Pellets were stored at −70 °C in the presence of PMSF (50 μl of a 70 mg/ml solution) for further processing. In other experiments 2.5 μM epoxomicin was used in the same conditions, except that starving of cells in the absence of Arg before labeling was carried out for 12 h. All incubations were done at 37 °C. Peptide labeling was quantified by the labeling ratio, which was defined as follows. Ratio = [(^{15}N + inh.)- ^{14}N]/(^{15}N-^{14}N), where ^{14}N, ^{15}N, and (^{15}N + inh.) are the percent intensities of the relevant isotopic peak, relative to the monoisotopic one, in the MALDI-TOF spectrum of the peptide, in the absence of labeling, upon labeling in the absence of inhibitor, and in its presence, respectively.

*Isolation of HLA-A*68:01-bound Peptides*—This was carried out as previously described (36). Briefly, cells were lysed in 1% Igepal CA-630 (Sigma-Aldrich, St Louis, MO) in the presence of a mixture of protease inhibitors. The soluble fraction was subjected to affinity chromatography using the W6/32 mAb. HLA-A68 -bound peptides were eluted with 0.1% TFA at room temperature, filtered through Centricon 3 (Amicon, Beverly, MA), concentrated and either used as a peptide pool for sequencing, or subjected to HPLC fractionation as previously described (37). The peptides to be sequenced from individual HPLC fractions were identified as those with the same monoisotopic mass and retention time as the labeled peptides in a parallel chromatography of an unlabeled peptide pool obtained from 10^{10} cells, by comparing the MALDI-TOF spectra of correlative and highly matched HPLC fractions from this peptide pool and those from the labeling experiments (23).

*Analysis of Residue Frequencies Among A*68:01 Ligands*—This was automatically done using an updated version of the previously described MSearcher software (38).

Assignment and Analysis of the Parental Proteins of HLA-A68 Ligands—This was done on the basis of unambiguous matching with a single human protein in the UniProtKB database release 14.0: 2008/07/22 using the Fasta 3 software (htpp://www.ebi.ac.uk/fasta), after taking into account the database redundancy because of multiple entries for the same protein. When a peptide ligand matched several closely related members of a protein family, a single entry for a representative member was chosen, with the understanding that the same ligand can arise from more than one member of such families.

The molecular mass (MW) and theoretical isoelectric point (pI) of the assigned proteins was calculated with the Compute pI/Mw tool (htpp://www.expasy.org/tools/pi_tool.html). Subcellular localization of the proteins was obtained from the UniProtKB release 14.0: 2008/07/22 database using the Protein Information and Knowledge Extractor software (39), available at http://proteo.cnb.csic.es:8080/pike. Human proteome analysis was performed with the 19329 entries in the UniProtKB/Swiss-Prot database using the Compute pI/Mw tool as above. Statistical analyses were carried out using the χ^2 test.

Proteasome Purification and Hydrolysis of Fluorogenic Substrates—Preparation of cell extracts containing 26S proteasome was performed as previously described (40). Briefly, about 2×10^6 cells were cultured for 1 h in phenol red-free Ivecco's modified eagle's medium (Invitrogen), supplemented with 10% FBS, in the presence or absence of proteasome inhibitors. Cells were washed twice with phosphate-buffered saline, permeabilized with 1.4 ml of 50 mM Tris-HCl, pH 7.5, 250 mM sucrose, 5 mM $MgCl_2$, 0.5 mM EDTA, 1 mM dithiothreitol, 1 mM ATP, 0.025% digitonin (homogenization buffer), for 5 min at 4 °C, and centrifuged (2×10^4 g, 10 min). Supernatants containing the cytosol were further centrifuged (3×10^5 g, 2 h) to separate large protein complexes, including the 26S proteasome, and these pellets were resuspended in 1 ml of digitonin-free homogenization buffer. Analysis of the tryptic and chymotryptic activity was performed with 100 μM Boc-LRR-amc or Suc-LLVY-amc (both from Bachem, St. Helens, UK), respectively. The cell extracts were incubated with the fluorogenic substrates at 37 °C for 1 h, and aliquots were taken every 10 min. The reaction was stopped by adding 0.33% TFA. Fluorescence was measured in an Aminco-Bowman series 2 fluorimeter (Sim-Aminco Spectronic Instruments, Rochester, NY) with excitation and emission wavelengths of 380 and 460 nm, respectively. The hydrolysis rate was calculated as the ratio between the slope of the reaction progress curves in the presence and absence of inhibitor. 20S proteasomes were purified from C1R cells by ion exchange chromatography and centrifugation in glycerol gradient as previously described (37).

In-gel Digestion Assays—Cell extracts containing 26S proteasome from 2×10^6 cells were loaded into a 3% polyacrylamide gel, in 90 mM Tris/90 mM boric acid pH 8.3, 5 mM $MgCl_2$, 0.5 mM EDTA, 1 mM ATP-$MgCl_2$ (resolving buffer), and run for 3 h at 110 V and 4 °C as previously described (41). Other samples were run in a 3–14% gradient gel in the same buffer. This process was also carried out changing the electrophoretic polarity in some experiments. After electrophoresis, the gels were washed with 50 mM Tris-HCl, pH 7.4, 5 mM $MgCl_2$, 1 mM ATP (developing buffer), and incubated in 1 ml of this buffer with 50 μM of the fluorogenic substrate for 30 min at 30 °C. Gels were visualized in a UV Transiluminator.

In Vitro Digestions of Synthetic Precursors—Two synthetic peptide substrates, a 31-mer and a 30-mer, were used. They were obtained using N-(fluorenyl)-methocarbonyl chemistry and purified by reverse phase HPLC. Their purity and correct molecular mass was established by MALDI-TOF MS. About 20 μg of the substrate were incubated for 8h at 37 °C with 2 μg of purified 20S proteasome with and without inhibitors, in 25 mM HEPES buffer, pH 7.6. The proteasome was incubated with the inhibitors for 2 h, previous to the addition of the substrate. The reaction was stopped with 0.067% TFA. For the 31-mer the digestion products were separated from the proteasome by molecular filtrating using 10 kDa pore size Centricon and fractionated by HPLC as described (37). The HPLC fractions were analyzed by MALDI-TOF/TOF MS. For the 30-mer the digestion mixture was purified with OMIX C18 tips (Agilent technologies, Santa Clara. CA), after washing with 0.2% TFA, then with 40% acetonitrile in 0.1% TFA, and finally eluting the peptides with 70% acetonitrile in 0.1% TFA. The digestion mixture was directly analyzed by MALDI-TOF MS/MS. The digestion products were identified using a software tool, designated as *Digestor,* developed as a Windows application. This tool assigns each ion peak to possible internal sequences in the precursor substrate on the basis of its molecular mass. If an ion peak matched several possible internal sequences, the ambiguity was either resolved by sequencing or not taken into account. The yield of individual digestion products was estimated as follows. The percentage of the total absorbance at 210 nm corresponding to a given chromatographic peak was determined. When multiple digestion products co-eluted in a same peak a percentage of its absorbance was assigned to each peptide on the basis of the relative intensity of the ion peaks in the MALDI-TOF spectrum of the corresponding HPLC frac-

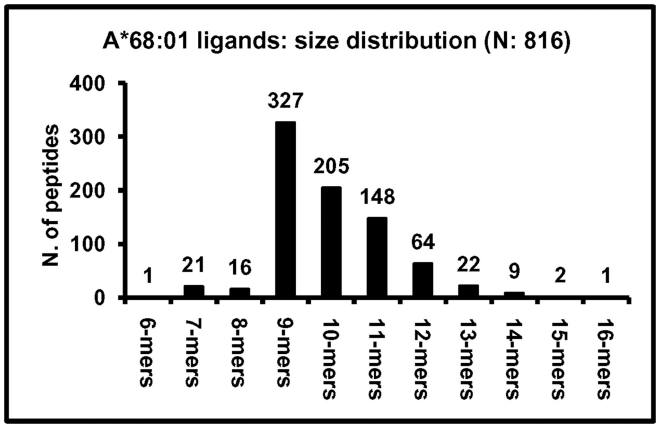


FIG. 1. **Size distribution among natural A*68:01 peptide ligands.** The total of 816 A*68:01 ligands, includes 35 previously available (htpp://www.syfpeithi.de) and 781 newly determined sequences.

tions. Alternatively, the relative yields of individual digestion products were calculated directly from the relative intensities of the ion peaks in the MALDI-TOF MS of the unfractionated digestion mixture.

RESULTS

*The A*68:01 Peptide Motif*—This was determined from a database of natural self-ligands, built from 3 sources. First, 35 sequences were available from the SYFPEITHI database of MHC ligands (http://www.syfpeithi.de). Second, 772 sequences were obtained from the A*68:01-bound peptide pool isolated from C1R transfectants by electrospray-Orbitrap MS/MS. The false discovery rate of this data set was 3.9%. Third, 70 sequences were obtained from the same peptide pool by MALDI-TOF/TOF MS/MS. Since some peptides were sequenced by more than one procedure, their total number was 816 (Table 1S), with a size range of 6 to 16 amino acid residues (Fig. 1). The A*68:01 peptide motif was determined from the 778 nonamer and longer peptides. The analysis was performed with this total set for the 5 anchor positions (P)1, P2, P3, PC-2, and PC (Fig. 2), because these can be properly aligned among peptides with different sizes. In addition, residue usage at all positions was analyzed separately for 9-mers, 10-mers, 11-mers and 12-mers (supplemental Fig. S2 and supplemental Table S2).

At all anchor positions some residues were predominant (Fig. 2). At P1, acidic residues were found in 50.6% of the ligands, followed by T and S. P2 was restricted by size rather than polarity, since residues in the range of 71–103 Da accounted for 84.4%, but both polar (T) and aliphatic residues (A, V) were tolerated. At P3, aliphatic residues and F were predominant. PC-2 showed preference for aliphatic residues (54.4%), but uncharged polar ones were also frequent (26.1%). At PC, K and R jointly accounted for 97.7% of the ligands, with a minor allowance for nonpolar residues. When residue frequencies were made relative to the human proteome a good correlation was found between the frequency of a given residue at any given anchor position and an increase or decrease relative to the proteome, indicating positive or negative selection of that residue among A68 ligands. The few peptides with nonpolar C-terminal residues (2.3%) are unlikely to be contaminant HLA-C*04 ligands, expressed on C1R cells, since they have features typical of A*68:01, but not of C*04 ligands, such as acidic P1 residues, and lack the P2 F/Y motif of C*04 (supplemental Table S1). Residue frequencies among peptides of different length were small and not statistically significant (supplemental Fig. S2). At non-anchor positions residue restrictions were low (supplemental Table S2).

One hexamer, 21 heptamers and 16 octamers were sequenced from the A*68:01-bound pool. Residue usage among these peptides deviated significantly from longer ones at P1-P3, but not at PC or PC-2 (supplemental Fig. S3). Longer variants, frequently consisting of N-terminal extensions, were found for many of these ligands.

*Parental Proteins of A*68:01 Ligands*—The 816 A*68:01 ligands arose from 672 parental proteins whose MW and pI distribution was similar as in HLA-B*27:05 (42) and showed, relative to the human proteome, a higher percentage of large (>30 K_D) acidic proteins and a decreased percentage of small (<30 K_D) acidic ones (Figs. 3*A*–3*D*). These differences were statistically significant. The majority of proteins were from the cytoplasm and nucleus and much lower percentages were from other subcellular compartments (Fig. 3*E*). This was similar to B*27:05 (Fig. 3*F*) and other MHC class I molecules (43).

A total of 74 proteins were the source of both A*68:01 or B*27:05 ligands (11% and 15.7%, respectively). This is about fourfold higher than expected if each set of parental proteins were independently selected in an unbiased way, suggesting that the ligands of both MHC molecules arise from a subset of the proteome. For example, membrane proteins were underrepresented, relative to the whole proteome, among the sources of A68 and B27 ligands (Figs 3*E*–3*H*). Other factors such as abundance, turnover rate, degree of cotranslational degradation of defective ribosomal products, etc., may influence the contribution of a given protein to the MHC-I-bound peptidomes. The parental proteins common to A68 and B27 showed similar distribution of MW and pI as the total sets, except for a lower percentage of small acidic proteins (Fig. 3*C*).

*Susceptibility and Resistance of A*68:01 Ligands to Proteasome Inhibitors*—The expression of A*68:01 ligands was analyzed by stable isotope labeling, using ^{15}N-tagged Arg or Lys, in the absence or presence of MG-132 or epoxomicin. The A*68:01-bound peptide pool was fractionated by HPLC, and each fraction was analyzed by MALDI-TOF MS. The peptides isolated from cells treated with ^{15}N-tagged amino acids showed a different isotopic distribution compared with the same unlabeled peptides. As the tagged residues are 2 Da heavier than their nontagged counterpart, labeling was detected as an increase of the intensity of the A+2, A+4, etc. peaks (A being the monoisotopic peak), depending on the number of tagged peptide residues. The labeling of inhibitor-

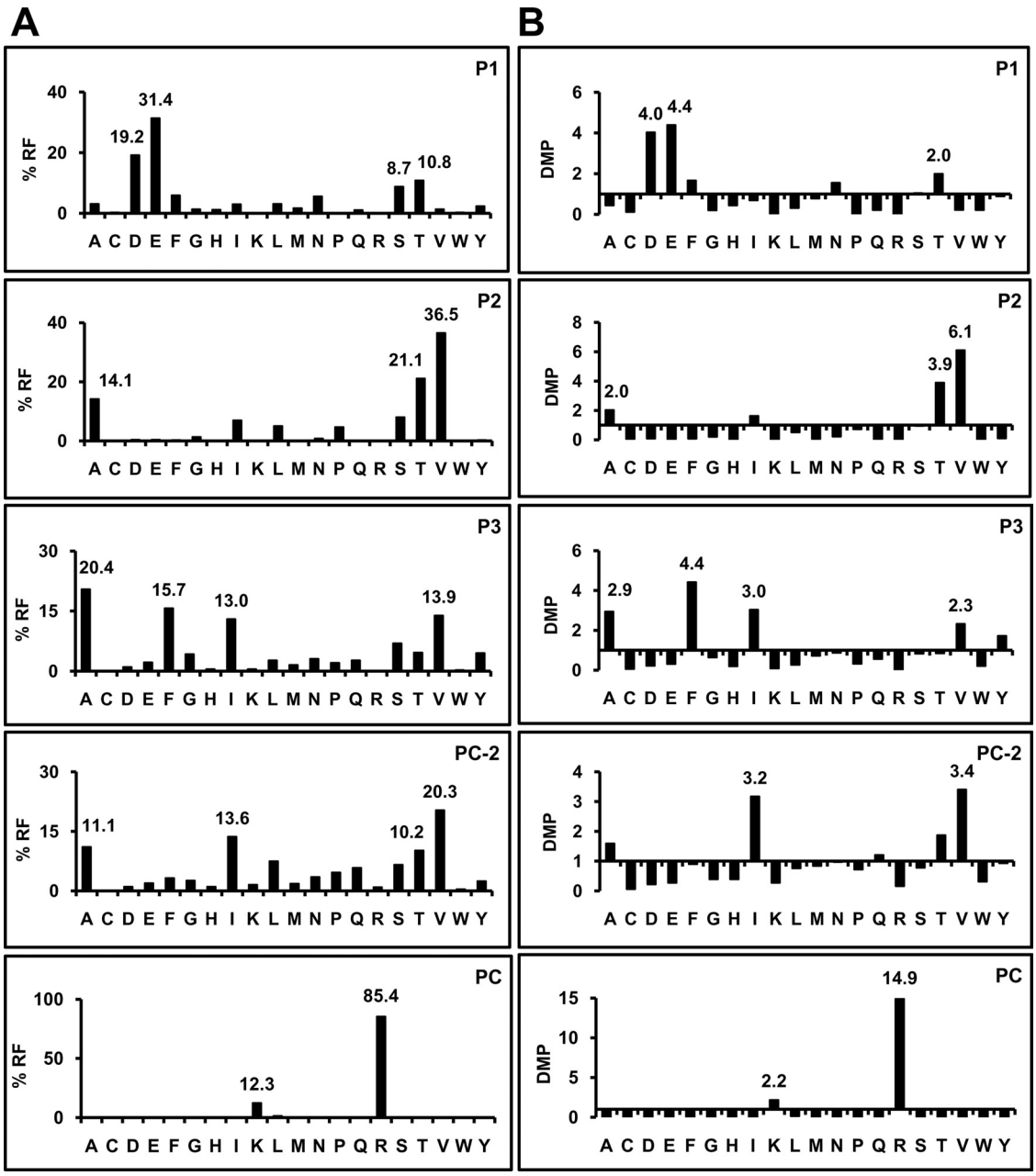


FIG. 2. **Residue usage among A*68:01 ligands longer than octamers.** *A,* percent residue frequencies (% RF) and *B,* deviation relative to the mean frequency in the human proteome (DMP) among 778 A*68:01 ligands ranging from 9 to 16 residues were determined at positions (P)1, P2, P3, PC-2, and PC. At each position the % RF and DMP values of the predominant residues is highlighted.

resistant ligands should be unaltered or only slightly decreased in the presence of inhibitor, because their generation would not be abrogated by inhibition of the proteasome. In contrast, little or no labeling should be detected in inhibitor-sensitive ligands isolated from cells treated with inhibitor plus ^{15}N-tagged amino acids.

In a first set of experiments cells were labeled with ^{15}N-tagged Arg or Lys in the presence of 20 μM MG-132. A total of 132 ion peaks (supplemental Table S3) were amenable to analysis on the following basis: (1) they showed sufficient intensity for good detection of the isotopic envelope, and (2) the relevant isotopic peak increased at least 20% upon ^{15}N labeling in the absence of inhibitor, relative to unlabeled cells. A given ligand was assigned as inhibitor-resistant if its labeling ratio was $>$0.6, that is, if the increased intensity of the corresponding isotopic peak in the presence of the inhibitor was more than 60% of the increase obtained in its absence. This threshold was adopted because decreased label in the presence of inhibitor may be because of indirect effects, such as down-regulation of protein synthesis (44). The inhibitor-

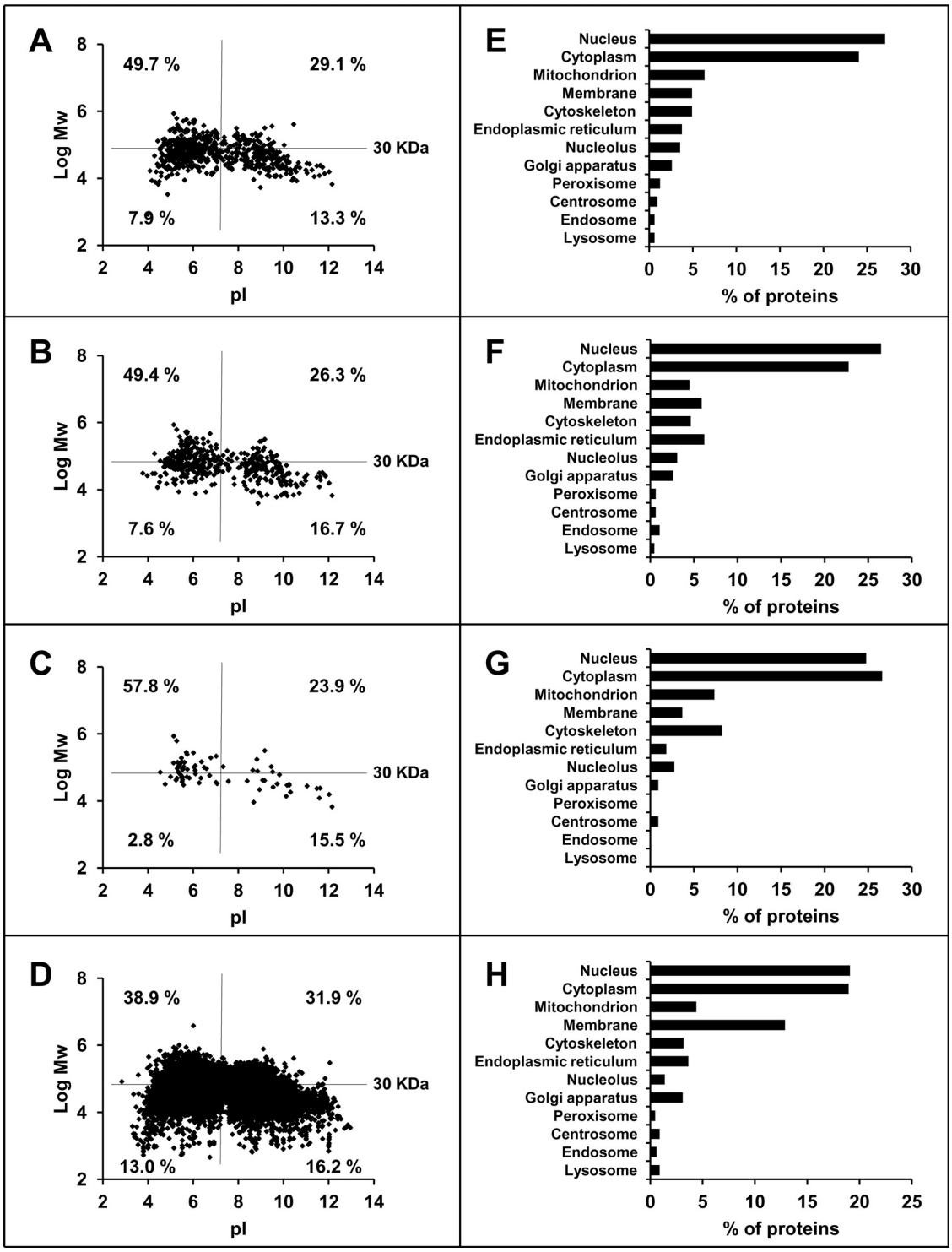


FIG. 3. **Parental proteins of A*68:01 ligands and their relationship to those of HLA-B27 and to the human proteome.** *A–D*, the molecular mass (log Mw) was plotted *versus* the isoelectric point (pI) for (*A*) the parental proteins ($n = 672$) of A*68:01 ligands, (*B*) those of B*27:05 ligands ($n = 471$) (42), (*C*) parental proteins ($n = 74$) common to A*68:01 and B*27:05 ligands, and (*D*) the human proteome (UniProtKB database release 14.0: 2008/07/22). In each panel, the proteins are classified by size (using a 30KD threshold) and pI, and the percentage of each subset is specified. *E–H*, subcellular distribution of (*E*) the parental proteins of A*68:01 ligands, (*F*) those of B*27:05 ligands, (*G*) parental proteins common to A*68:01 and B*27:05 ligands, and (*H*) the human proteome. The analysis was carried out using the Protein Information and Knowledge Extractor software (39) at http://proteo.cnb.csic.es/pike/analysis.html.

Sensitive (N=31)		MG-132		Epoxomicin		
Fraction N.	M+H$^+$	Residues	Ratio	Residues	Ratio	Sequence
92	969.42	R	0.02	R	0.20	
125 / 124	986.43	R	0.19	R	-0.08	
125	991.41	R	0.09	R	0.23	
110	992.42	R	0.15	R	0.10	NAIGGKYNR
111	992.60	K	-0.18			
134 / 135	1000.48	R	-0.13	R	0.29	
112 / 114	1003.41	R	0.11	R	-0.03	DTAAQITQR
120 / 118	1010.51	R	-0.03	R	0.00	
108 / 109	1012.38	R	0.27	R	-0.01	TSETPQPPR
155	1021.51	R	0.27	R	0.18	DVFVVGTER
113 / 114	1025.53	R	-0.13	R	0.06	
144 / 144	1032.47	R	0.04	R	0.48	FPSEIVGKR
145	1032.56	K	0.19			
132	1043.60	R	0.18	R	0.09	ELTAVVQKR
132	1043.60	K	0.15			
152 / 153	1060.57	R	0.01	R	0.08	DVFRDPALK
152	1060.58	K	-0.09			
145 / 146	1070.49	R	0.22	R	0.32	EVAQLIQGGR
126	1075.63	2 R	0.18	2 R	0.09	HPAVVIRQR
129	1077.42	R	-0.06	R	0.20	ETNYGIPQR
150 /152	1089.64	R	0.01	R	0.30	
101 / 99	1108.50	R	0.12	R	0.12	
123 / 124	1122.49	2 R	0.35	2 R	0.11	
204 / 205	1123.63	R	0.14	R	0.13	
158 / 159	1128.57	R	0.12	R	0.02	EVFSHILKR
158	1128.64	K	0.08			
181 / 182	1137.54	R	0.22	R	0.33	EFVDLYVPR
129 / 130	1139.54	2 R	0.19	2 R	0.07	DPFHKAIRR
130	1139.74	K	0.33			
104	1140.50	3 R	0.25	3 R	0.13	EVNDRPVRR
125 / 124	1146.43	2 R	-0.05	2 R	0.06	
114 / 113	1155.67	R	0.21	R	0.26	TVKKGKPELR
114	1155.80	3 K	0.04			
154	1212.52	R	-1.14	R	0.27	FPHNPQFIGR
153	1245.71	R	0.08	R	0.25	
153	1245.72	K	0.13			
126	1246.62	R	0.00	R	0.03	TAADTAAQITQR
152 / 154	1376.76	2 R	0.13	2 R	0.22	NSFRYNGLIHR
142 / 144	1608.91	3 R	0.28	3 R	0.28	
Resistant (N=5)		MG-132		Epoxomicin		
Fraction N.	M+H$^+$	Residues	Ratio	Residues	Ratio	Sequence
167	1048.58	R	0.62	R	0.64	SVQGIIIYR
127	1200.52	R	1.13	R	0.67	
201 / 203	1207.73	R	1.76	R	0.66	SAFATPFLVVR
129	1209.55	2 R	0.83	2 R	0.73	ETVQLRNPPR
136	1462.75	R	2.93	R	0.82	TVKDVNQQEFVR
136	1462.73	K	2.02			
Not assigned (N=6)		MG-132		Epoxomicin		
Fraction N.	M+H$^+$	Residues	Ratio	Residues	Ratio	Sequence
130	1023.5	2 R	0.66	2 R	0.53	
153 / 154	1103.54	2 R	0.51	2 R	0.53	
173 / 174	1132.52	R	1.10	R	0.38	
130	1187.53	R	0.70	R	0.37	NVAEVDKVTGR
130	1187.74	K	0.48			
144	1250.57	2 R	0.51	2 R	0.21	
128 / 127	1443.68	R	0.74	R	0.35	TVINQTQKENLR
127	1443.88	K	0.47			

FIG. 4. **Isotopic labeling of HLA-A*68:01 ligands in the presence of proteasome inhibitors.** A total of 42 ion peaks were analyzed with both 20 μM MG-132 and 2.5 μM epoxomicin. Their elution position in HPLC (Fraction N), monoisotopic mass (M+H$^+$), the number of Arg (R) or Lys (K) residues, which was inferred from the isotopic peak that increased upon ^{15}N-labeling, and the labeling ratio (Ratio: see Experimental Procedures), are indicated. Peptides were classified as inhibitor-sensitive or -resistant when the labeling ratio was <0.35 or >0.6, respectively, with both inhibitors. Underlined labeling ratios for some inhibitor-sensitive ligands indicate distorted values due to inhibition of labeling close to background levels. When the labeling ratio showed intermediate values or was significantly different with both inhibitors, the peptides were not assigned. Within each set, the peptides are ordered by their M+H$^+$. Ion peaks that were sequenced are indicated (see Tables 3S and 4S for full labeling data).

sensitive ligands were those showing labeling ratios <0.35. This threshold was adopted to allow for residual proteasome activity in the presence of inhibitors, as well as for inherent experimental error. On this basis 95 peptides (71%) were sensitive to proteasome inhibition, 24 (19%) were inhibitor-resistant, and 13 (10%) were not assigned, either because they showed intermediate labeling ratios, were more inhibited with epoxomicin, or both.

In a second experiment the cells were labeled with ^{15}N-tagged Arg in the presence of 2.5 μM epoxomicin. Of 63 peptides amenable to this analysis (supplemental Table S4), 46 (73%) were inhibitor-sensitive, 8 (13%) were inhibitor-resistant and 9 (14%) were not assigned because of intermediate labeling ratios, less inhibition with MG-132 or both.

Of 42 peptides analyzed with both inhibitors 31 (74%) were sensitive, 5 (12%) were resistant and 6 (14%) were not assigned (Fig. 4). In the latter subset, 5 of the 6 peptides showed higher inhibition with epoxomicin.

These results indicate that both inhibitors have a similar effect on most A*68:01 ligands. However, a higher inhibition by epoxomicin was suggested by the smaller percentage of resistant ligands, relative to MG-132, and by its more drastic effect on some peptides (Fig. 4).

The average intensity of the monoisotopic (M+H$^+$) MALDI-TOF ion peaks corresponding to the peptides in Fig. 4 was significantly higher for inhibitor-sensitive than for inhibitor-resistant ligands (Table I), indicating that the resistance was not an artifact due to high peptide abundance. Indeed, many inhibitor-sensitive peptides showed ion peak intensities much higher than any of the resistant peptides. The lower intensity values observed in the presence of inhibitors affected equally to inhibitor-sensitive and -resistant peptides. It results from nonspecific global effects affecting peptide yields, such as decreased cell viability in the presence of inhibitors, and perhaps also from down-regulation of multiple cell biological processes following proteasome inhibition, including MHC-I maturation and export.

*Identification of Inhibitor-resistant and Inhibitor-sensitive A*68:01 Ligands and Their Parental Proteins*—The sequence of 24 inhibitor-sensitive, seven inhibitor-resistant, and two peptides showing different susceptibility with both inhibitors was determined. All of the inhibitor-sensitive, the two nonassigned and five inhibitor-resistant peptides corresponded to internal sequences of their presumed parental proteins. Two

TABLE I

Mean intensity of the monoisotopic ($M+H^+$) ion peaks of inhibitor-sensitive and resistant peptides. This analysis was performed for the peptides in Figure 4

	MG-132			Epoxomicin		
	^{14}N	^{15}N	^{15}N+Inh.	^{14}N	^{15}N	^{15}N+Inh.
Sensitive (n = 31)	13461	11092	7026 (63%)[c]	3001	2489	1644 (66%)[c]
Resistant (n = 5)	2818	2364[a]	1657 (70%)[c]	551	356[b]	176 (49%)[c]
Ratio	4.8	4.7	4.2	5.4	7.0	9.3

[a] Range: 537 to 5030. As many as 20 inhibitor-sensitive peptides (64.5%) had higher intensity.
[b] Range: 56 to 602. As many as 17 inhibitor-sensitive peptides (54.8%) had higher intensity.
[c] Percent values are relative to peptides labeled in the absence of inhibitor (^{15}N).

inhibitor-resistant peptides corresponded to N-terminal (ETVQLRNPPR) or C-terminal (DPAGVHPPR) protein sequences (Fig. 5). There were no obvious differences between inhibitor-sensitive and resistant peptides in size, sequence motifs, or frequency of flanking residues (data not shown). In addition, abundance of the source protein was unrelated to inhibitor-resistance. For instance, abundant ribosomal proteins were the source of both types of ligands (Fig. 5). However, all 5 inhibitor-resistant peptides corresponding to internal protein sequences arose from small (30 kDa or smaller) and basic proteins. The two peptides arising from N- or C-terminal protein ends did not follow this rule. For inhibitor-sensitive ligands the parental proteins with MW<30 kDa and pI>7 accounted for 54.5% (Fig. 6).

We found two examples of ligands derived from the same protein but showing differential sensitivity to proteasome inhibitors. First, two peptides (EFVDLYVPR and NVAEVDKVTGR) arising from the 40S ribosomal protein S21 behaved as inhibitor-sensitive and not assigned, respectively. A previously reported B27 ligand (GRFNGQFKTY) derived from the same protein was inhibitor-resistant (23). Second, ELTAVVQKR and ELYAEKVATR, from the 40S ribosomal protein S3 were inhibitor-sensitive and -resistant, respectively (Fig. 5).

Inhibition of the 26S Proteasome Subspecificities—The effect of inhibitors on the chymotryptic- and tryptic-like activities of the proteasome from C1R cells was tested with specific fluorogenic substrates. When fractionated in native gels the 26S proteasome samples (Fig. 7*A*) contained a mixture of 26S and 20S proteasome, and some additional bands that were best observed in gradient electrophoresis. Hydrolysis of both substrates was observed in the 20S and 26S proteasome bands, but not elsewhere. Upon inverting the electrophoretic polarity, additional contaminants or hydrolytic activity were not observed (data not shown). These results suggest that the hydrolysis of the fluorogenic substrates is due to the proteasome. In the presence of 2.5 μM epoxomicin, all detectable hydrolysis of both substrates was inhibited. In contrast, with MG-132 the hydrolysis of the tryptic substrate was only partially inhibited and residual chymotryptic activity was also observed (Fig. 7*A*).

In other experiments C1R-A*68:01 cells were incubated with or without inhibitor at various concentrations. After 1h, the cytosolic 26S proteasome was partially purified and its hydrolytic activity for the tryptic and chymotryptic substrates was determined by measuring the generation of fluorescence as a function of time. In parallel experiments, the 26S proteasome was purified from untreated C1R cells, incubated with inhibitors and tested in the same way. The results in both conditions were very similar, indicating that proteasomal inhibition was not significantly affected by incubating either whole cells or purified proteasome with MG-132 or epoxomicin (Figs. 7*B*–7*C*). The chymotryptic-like activity was totally inhibited (94 ± 5%) with 20 μM epoxomicin. At the 2.5 μM concentration used for the analysis of A*68:01 ligands, the inhibition of this subspecificity was high (72 ± 7%), but incomplete. The tryptic-like activity was inhibited to a maximum of about 67% at 5 μM or higher concentrations of epoxomicin, and nearly as much (63 ± 2%) at 2.5 μM. With MG-132 the maximal inhibition of the chymotryptic-and tryptic-like activities was obtained at 40 μM (72 ± 4% and 59 ± 1%, respectively), but similar values (68 ± 5% and 56 ± 2%, respectively) were observed at 20 μM, used for A*68:01 ligands.

These results indicate that: (1) inhibition of the tryptic- and chymotryptic-like specificities of the proteasome was higher with epoxomicin than with MG-132, (2) this inhibition was not complete, but was close to 100% with epoxomicin for the chymotryptic-like activity, (3) the residual hydrolysis of fluorogenic substrates is probably because of incomplete inhibition of the proteasome, rather than to non-proteasomal activity (Fig. 7*A*), and (4) in the conditions used for A*68:01 ligands both the tryptic- and the chymotryptic-like activities of the proteasome were incompletely inhibited.

Effect of Proteasome Inhibition on the Generation of Inhibitor-resistant Ligands In Vitro—Incomplete inhibition may alter proteasomal cleavage patterns, resulting in decreased destruction and/or increased production of some MHC-I ligands, which might explain their resistance to inhibitors (26). To explore this possibility, a synthetic 31-mer spanning residues 29–59 from the 40S ribosomal protein S21 was digested *in vitro* with purified 20S proteasome in the absence or presence of inhibitors (Fig. 8*A*, Table II and supplemental Table S5). This substrate includes two overlapping sequences (NVAEVDKVTGR and *GR*FNGQFKTY) that correspond to an A*68:01 ligand showing partial resistance to

FR. N.	M+H	Sequence	A. N.	Protein name	AA Length	Sequence Location	Mw	pI
SENSITIVE								
110	992.61	NAIGGKYNR	P14324	Farnesyl pyrophosphate synthase	419	120-130	40507	5.1
113/114	1003.41	DTAAQITQR	**P30685**	**HLA class I histocompatibility antigen, B35 alpha chain**	362	161-169	37804	5.9
108/109	1012.38	TSETPQPPR	Q9UBU8	Mortality factor 4-like protein 1	362	169 177	26716	8.4
145	1018.45	DTPGFIVNR	Q16836	Hydroxyacyl-coenzyme A dehydrogenase, mitochondrial	314	219-227	27842	8.0
154/155	1021.44	DVFVVGTER	P78347	General transcription factor II-I	949	53-63	107163	5.5
142	1032.47	FPSEIVGKR	**P62081**	**40S ribosomal protein S7**	194	135-143	8945	8.3
152	1038.41	DAFFNPATR	Q4KMQ2	Anoctamin-6	910	200-208	106062	7.9
130	1043.53	ELTAVVQKR	**P23396**	**40S ribosomal protein S3**	243	70-78	26688	9.7
150	1060.47	DVFRDPALK	P61353	60S ribosomal protein L27	136	101-109	5058	10.8
143	1070.49	EVAQLIQGGR	Q9Y5U8	Brain protein 44-like protein	109	90-99	6308	9.4
126	1075.63	HPAVVIRQR	Q12870	Transcription factor 15	140	79-87	15026	10.8
126/127	1077.42	ETNYGIPQR	P63244	Guanidine Nucleotide-binding protein subunit beta-2-like	317	49-57	35089	7.6
145	1089.49	EPSEVFINR	Q8WXF1	Paraspeckle component 1	523	109-117	58706	6.3
157	1128.57	EVFSHILKR	P33076	MHC class II transactivator	1130	498-506	101512	5.3
181	1137.54	EFVDLYVPR	**P63220**	**40S ribosomal protein S21**	83	7-15	9106	8.7
141	1139.48	ESFPGSFRGR	Q9BWT1	Cell division cycle-associated protein 7	371	222-231	42629	9.6
128/129	1139.54	DPFHKAIRR	Q9H6E4	Coiled-coil domain-containing protein 134	204	140-148	25646	11.8
104	1140.50	EVNDRPVRR	Q9Y520	Protein BAT2-like 2	2897	1036-1044	316845	9.2
110/111	1155.59	TVKKGKPELR	Q9BTD8	RNA-binding protein 42	140	66-75	15026	10.8
153	1212.52	FPHNPQFIGR	Q9H9Y2	Ribosome production factor 1	349	275-284	29962	9.8
129	1221.53	DAAHPTNVQRL	P35222	Catenin beta-1	781	114-124	85442	5.5
177	1244.63	LVFPSEIVGKR	**P62081**	**40S ribosomal protein S7**	194	137-147	8945	8.3
123	1246.55	TAADTAAQITQR	**P30685**	**HLA class I histocompatibility antigen, B35 alpha chain**	362	158-169	37804	5.9
152	1376.59	NSFRYNGLIHR	P46779	60S ribosomal protein L28	137	36-46	10694	10.4
RESISTANT								
86/87	969.34	EPAPGTNQR*	Q9UBW8	COP9 signalosome complex subunit 7a	275	244-252	30267	8.3
118	945.51	DPAGVHPPR*	Q08J23	tRNA (cytosine-5-)-methyltransferase NSUN2	767	759-767	86471	6.3
166/167	1048.54	SVQGIIIYR	P05141	ADP/ATP translocase 2	298	184-194	29103	9.8
137	1179.61	ELYAEKVATR*	**P23396**	**40S ribosomal protein S3**	243	87-96	26688	9.7
203	1207.65	SAFATPFLVVR	P15954	Cytochrome c oxidase subunit 7C, mitochondrial	63	47-57	5356	9.0
129/130	1209.55	ETVQLRNPPR	Q13043	Serine/threonine-protein kinase 4	487	2-11	45822	5.0
136	1462.68	TVKDVNQQEFVR	P39019	40S ribosomal protein S19	145	5-16	16051	10.3
NOT ASSIGNED								
130	1187.53	NVAEVDKVTGR	**P63220**	**40S ribosomal protein S21**	83	35-45	9106	8.7
128	1443.68	TVINQTQKENLR	P42766	60S ribosomal protein L35	123	59-70	14420	11.0

FIG. 5. **Amino acid sequences of inhibitor-sensitive and inhibitor-resistant HLA-A*68:01 ligands.** Peptides are classified as in Fig. 4 and ordered by their monoisotopic mass ($M+H^+$) within each set. The corresponding parental proteins, their accession numbers (A.N.) in the Swiss-Prot database, amino acid (AA) length, location of the ligand in their sequence (sequence location), MW and theoretical pI are indicated. Polypeptides giving rise to more than one ligand are indicated in boldface. Resistant peptides marked with an asterisk (*) were assigned only on the basis of their lack of inhibition with MG-132.

proteasome inhibitors (Fig. 5) and to an inhibitor-resistant HLA-B27 ligand (23), respectively. In the absence of inhibitors, the substrate was totally degraded and cleavage occurred at virtually all peptide bonds. A total of 80 peptides, 66 of them requiring dual cleavage, were recovered with yields ≥ 0.1% (supplemental Table S5). Despite extensive cleavage within their sequences, the A68 (residues 7–17) and B27 (residues 16–25) epitopes were generated by the proteasome with 6.4 and 1% yield, respectively. Multiple N-terminally extended species of these ligands were also found.

With MG-132 only 28.9% of the substrate was cleaved. The number of digestion products, excluding the undigested substrate, was reduced to 28 peptides (Table 5S), which arose from cleavage at 22 peptide bonds. The A68 ligand was generated in very low amounts (0.3% of the digested material) and the B27 ligand was not produced, but N-terminally extended precursors of both ligands were observed (supplemental Table S5). Yet the ratio between the relative contribution of the cleavages implying epitope destruction and potential production was increased for the A68 (from 2.9 to 4.8) and the B27 ligand (from 0.9 to 1.3) in the presence of this inhibitor, relative to its absence (Table II). With epoxomicin 54.3% of the substrate was digested, and only 16 digestion products (supplemental Table S5), arising from cleavage

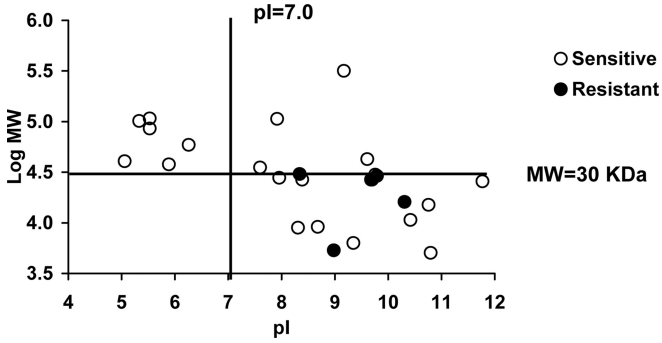


FIG. 6. **Inhibitor-resistant A*68:01 ligands predominantly arise from small and basic proteins.** The molecular mass (log MW) of the parental proteins from inhibitor-sensitive and -resistant A*68:01 ligands was plotted *versus* their theoretical pI. The value of 30 kDa corresponds to the highest molecular mass (30267 Da) observed among parental proteins of inhibitor-resistant ligands derived from internal sequences. The two parental proteins of ligands matching N- or C-terminal sequences were not included.

at 19 peptide bonds, were obtained. None of the natural ligands was generated, but potential precursors with the right C-terminal residues were observed. The ratio between the relative yield of destructive and productive cleavages was unaltered (from 2.9 to 3) or increased (from 0.9 to 2.7) for the A68 and B27 ligands, respectively.

These results indicate that: (1) the 20S proteasome generated both natural ligands and their N-terminally extended precursors, (2) neither inhibitor abolished the proteasomal activity on this substrate or the generation of productive species, and (3) incomplete proteasomal inhibition did not result in increased production of the natural ligands or their precursors.

Both inhibitors showed a comparable global effect on the various subspecificities, altering only slightly their relative contribution to cleavage (Table III). However, the effect on individual peptide bonds within each subspecificity, was inhibitor-dependent (supplemental Table S6). For instance, MG-132 inhibited more efficiently cleavage after K13 than after R17, whereas epoxomicin showed higher inhibition on cleavage after K13 and R17 than after K23. For residues assigned to the chymotryptic-like activity the relative inhibition of cleavage after M6 was more efficient with epoxomicin, whereas after F18 and F22 was more efficient with MG-132. For residues not assigned to a particular subspecificity the inhibition was generally more efficient with epoxomicin than with MG-132, but the opposite was true for N19. Thus, incomplete inhibition of the proteasome resulted in an inhibitor-dependent alteration of cleavage patterns.

The ligands analyzed above derived from the same protein as an inhibitor-sensitive A68 ligand: EFVDLYVPR (Fig. 5). Thus, we analyzed the effect of proteasome inhibitors on the *in vitro* generation of this epitope from a synthetic 30-mer with the sequence of its parental protein (residues 2–31) in this region (Fig. 8*B*, Table IV, supplemental Table S7). The natural ligand was not detected in the digest, but a productive cleavage, generating a single digestion product corresponding to residues 1–14 of the substrate, occurred with a relative yield of 16.8% in the absence of inhibitor, at lower yield (9%), with MG-132, and not with epoxomicin. The joint relative yield of cleavages implying epitope destruction was similar in the absence or in the presence of inhibitors (Table IV). Compared with the non-assigned one, the inhibitor-sensitive A68 ligand from the same protein showed a lower destruction/production ratio without inhibitor and with MG-132 and was not generated in the presence of epoxomicin.

Alteration of cleavage patterns in the presence of inhibitors was also observed with this substrate. Both MG-132 and epoxomicin had a similar global effect, consisting in the preferential inhibition of the tryptic and caspase subspecificities, at the expense of the chymotryptic and non-assigned ones (Table III). However, inhibitor-dependent differences on cleavage of individual peptide bonds were observed, most notably after R14, but also at other positions (Fig. 8*B* and supplemental Table S7).

Although extrapolating *in vitro* results to proteasomal processing and inhibition *in vivo* must be done with great caution, the main differences between the non-assigned and the sensitive A68 ligand were that MG-132 had a larger inhibitory effect on the generation of the productive species of the second ligand and that epoxomicin abrogated its generation. This is consistent with the sensitivity of this ligand to inhibitors observed *in vivo*.

DISCUSSION

This study was undertaken to address whether the observation that proteasome inhibitor-resistant HLA-B27 ligands predominantly arose from small and basic proteins (23) could be generalized to other MHC-I molecules and whether this results from proteasome-independent processing or incomplete proteasomal inhibition.

The canonical A*68:01 motif can be described by significant restrictions at four anchor positions: a preference for acidic residues, T and S at P1 (joint frequency: 70.1%), V, T, and A at P2 (joint frequency: 71.7%), nonpolar residues, mainly A, F, I, and V (joint frequency: 63%) at P3 and an almost absolute restriction for C-terminal basic residues. Short peptides deviated significantly from this pattern at P1-P3. Some, especially 8-mers, are probably canonical ligands or may be bound in noncanonic ways (45). Yet, we cannot rule out that, despite the protease inhibitors used during purification, some of the shortest species might result from residual amino peptidase activity.

The A*68:01 motif is very different from that of B*27:05, where basic residues and G are favored, and acidic residues are very rare at P1. The R2 motif of HLA-B27 is absent in A68, and C-terminal basic residues account only for 42% of B*27:05 ligands (42). Thus, the two molecules bind largely divergent peptidomes. However, that 11% of the parental proteins of A68 ligands were also the source of B27-bound

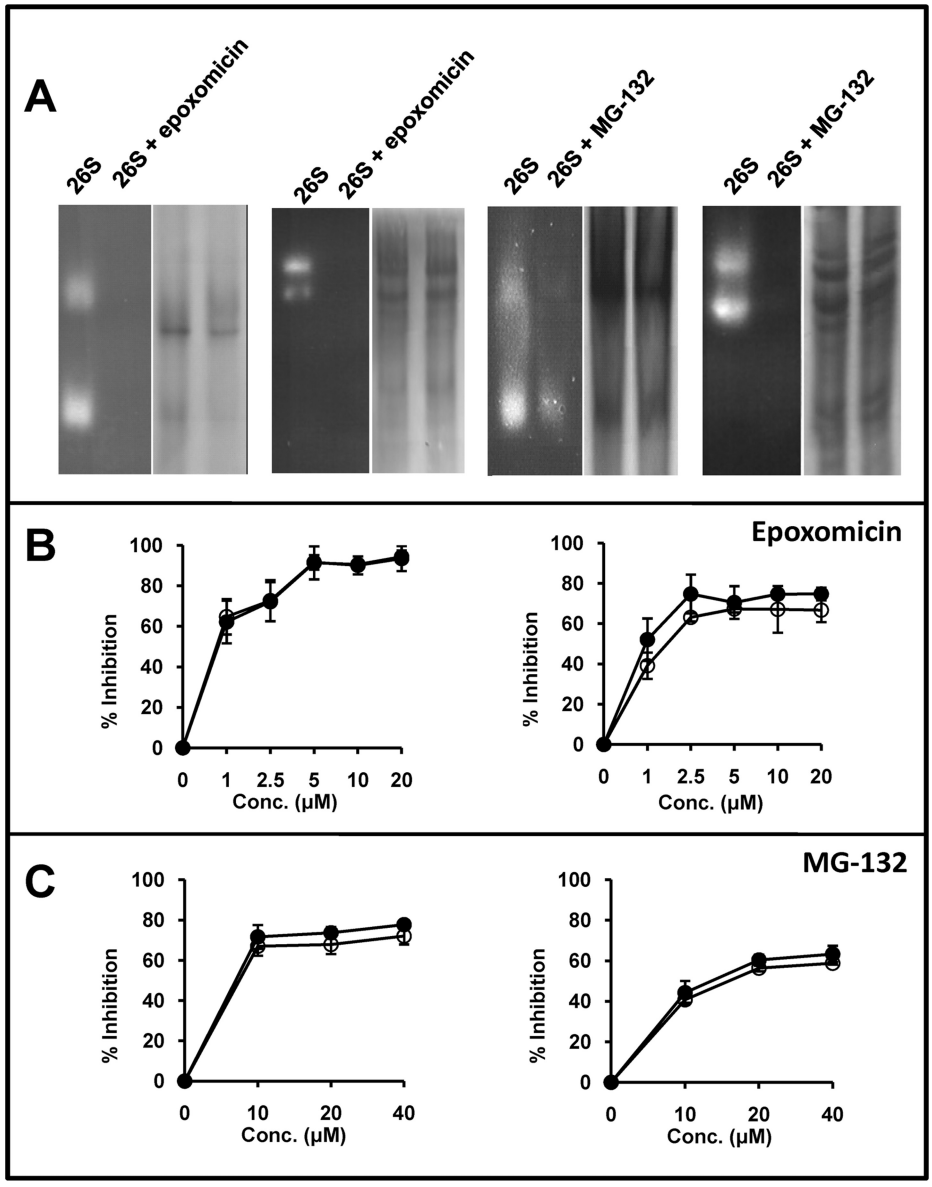


FIG. 7. **Activity of the 26S proteasome in the presence or absence of proteasome-inhibitors.** *A,* In-gel hydrolysis of site-specific fluorogenic substrates by partially purified 26S proteasomes, which include significant amounts of 20S proteasomes, incubated or not with 2.5 μM epoxomicin or 20 μM MG-132. The experiments were performed in 3% polyacrylamide gel or 3–15% polyacrylamide gradient gels for the tryptic (Boc-LRR-amc) and chymotryptic (Suc-LLVY-amc) substrates, respectively. For each experiment the in-gel digestion analysis is shown on the left side. The same gel, washed and stained with Coomassie blue, is shown at right. *B,* Inhibition of the hydrolysis of fluorogenic substrates specific for the chymotryptic (*left*) or tryptic (*right*) activity by partially purified 26S proteasomes in the presence of various concentrations of epoxomicin. Empty circles (o) show the results obtained with proteasome purified from cells treated with the inhibitor. Black circles (●) show results obtained with proteasome treated with the inhibitor after its purification from cells. *C,* Inhibition of the hydrolysis of fluorogenic substrates specific for the chymotryptic (*left*) or tryptic (*right*) activity by partially purified 26S proteasomes in the presence of various concentrations of MG-132. Conventions are as in panel B.

peptides indicates that both MHC molecules sample for the immune system both distinct subsets of the proteome and distinct epitopes of the same protein.

The rationale to distinguish between proteasome inhibitor sensitive and resistant A68 ligands was based on their metabolic labeling and the quantization of its inhibition. Since proteasomal inhibition severely impairs cell viability, the labeling times were relatively short, precluding uniform labeling. The observation that most of the inhibitor-resistant-peptides arose from small and basic proteins suggests that our previous observations on HLA-B27 (23) have a more general significance. Yet, the following facts hinted at an incomplete inhibition of the proteasome. First, the percentage of inhibitor-resistant ligands was smaller with ep-

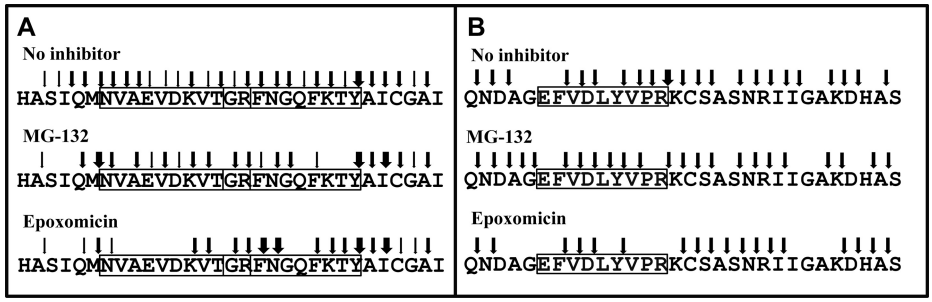


FIG. 8. **Digestion of synthetic precursors of inhibitor-resistant A*68:01 and B*27:05 ligands by purified 20S proteasome.** *A*, Cleavage of a 31-mer substrate, including the overlapping sequences of an A68 and a B27 ligand (boxed) in the absence or in the presence of the indicated inhibitors. Thin, medium and thick arrows indicate peptide bonds whose cleavage result in a joint yield of digestion products of 0.1 to 1%, 1.1 to 10% and >10%, respectively (see supplemental Table S6). *B,* Cleavage of a 30-mer substrate, containing the sequence of an inhibitor sensitive A68 ligand. The arrows follow the same convention as in panel A (see supplemental Table S7).

TABLE II
Effect of proteasome inhibitors on the digestion of the 31-mer substrate

	No Inhibitor	MG-132	Epoxomicin
Amount digested	100%	28.9%	54.3%
N. of digestion products	80	28	16
Peptides involving dual cleavage[a]	66 (82.5%)	20 (71.4%)	7 (43.8%)
N. of cleaved peptide bonds[b]	29 (96.7%)	22 (73.3%)	19 (63.3%)
N. of productive species for the A68 ligand[c]	4 (5.1%)	2 (7.1%)	1 (6.3%)
Relative yield of productive cleavages[d]	6.9%	5.5%	2.6%
N. of destructive cleavages for the A68 ligand[e]	10 (34.5%)	8 (36.4%)	4 (21.1%)
Relative yield of the destructive cleavages[d]	20.3%	26.3%	7.8%
N. of productive species for the B27 ligand[c]	12 (15.2%)	2 (7.1%)	2 (12.5%)
Relative yield of productive cleavages[d]	28.2%	14.6%	17.0%
N. of destructive cleavages for the B27 ligand[e]	9 (31%)	6 (27.3%)	7 (36.8%)
Relative yield of the destructive cleavages[d]	25.9%	18.3%	45.8%
Destruction/production ratio for the A68 ligand[f]	2.9	4.8	3.0
Destruction/production ratio for the B27 ligand[f]	0.9	1.3	2.7

[a] Percent relative to the total number of digestion products.
[b] Percent relative to the total number of peptide bonds in the substrate.
[c] Peptides with the right C-terminus and right or extended N-terminus. Percent values are relative to the total number of digestion products (see Table VS).
[d] Added values from Table VIS. For productive cleavages the yields correspond to the relative contribution of cleavage after R17 or Y25 for the A68 and B27 ligands, respectively.
[e] Cleaved peptide bonds implying destruction of the epitope. Percent values are relative to the total number of cleaved peptide bonds.
[f] Ratio between the joint relative yield of cleavages implying destruction of the natural ligand and that of the potentially productive peptides.

oxomicin than with MG-132. Second, inhibitor-resistant ligands from proteins other than small and basic ones corresponded to N- or C-terminal protein ends, which do not require dual cleavage. Third, five peptides were more inhibited with epoxomicin than with MG-132. Fourth, the intermediate inhibition of labeling observed on some A68 ligands might reflect down-regulation of protein synthesis (44), but also incomplete proteasomal inhibition. Fifth, the differential susceptibility to inhibitors of some peptides from the same parental protein could be explained by two mechanisms. Altered cleavage by a partially inhibited proteasome might result in increased production of only one of the ligands. Alternatively, the inhibitor-resistant ligand might be generated from protein fragments produced by a partially inhibited proteasome by nardylisin and/or thimet oligopeptidase, as recently reported (46).

Because replicate isotope-labeling experiments were not performed, we cannot rule out some inaccuracy in the assignments of inhibitor-sensitive and resistant ligands. However, the consistency in the assignment of most peptides amenable to analysis with both inhibitors, or with the same inhibitor upon K and R tagging, underlines the reliability of our analysis.

Incomplete inhibition of the proteasome was directly demonstrated both with fluorogenic substrates and synthetic peptide precursors. At the concentrations used *in vivo* both inhibitors impaired the hydrolysis of the chymotryptic substrate *in vitro* by about 70%. This value was lower than in HeLa cells (47), and reflects incomplete inhibition of the proteasome in C1R, because higher concentrations of epoxomicin, inhibited the hydrolysis of this substrate by >90%. Epoxomicin inhibited the tryptic activity by a maximum of about 67%, also less

TABLE III
Effect of inhibitors on the cleavage of synthetic peptide precursors by proteasome 20S

Activity	No inhibitor		MG-132		Epoxomicin	
31-mer	N. of Cleavages	% Contribution[a]	N. of Cleavages	% Contribution[a]	N. of Cleavages	% Contribution[a]
Tryptic	3	15.4	2	8.3	3	11.6
Chymotryptic	13	65.1	11	70.3	9	70.1
Caspase	2	1.2	2	0.9	0	0
Other	11	18.3	7	20.5	7	18.4
Total	29	100	23	100	20	100
30-mer	N. of Cleavages	% Contribution[b]	N. of Cleavages	% Contribution[b]	N. of Cleavages	% Contribution[b]
Tryptic	4	27.6	4	17.3	3	12.7
Chymotryptic	8	25.4	9	35.2	6	37.5
Caspase	3	21.1	2	12.1	2	12.5
Other	7	25.9	8	35.4	7	37.2
Total	22	100	23	100	18	100

[a] Added values from Table VIS.
[b] Added values from Table VIIS.

TABLE IV
Effect of proteasome inhibitors on the digestion of the 30-mer substrate

	No Inhibitor	MG-132	Epoxomicin
Amount digested	75.8%	56.8%	49.9%
N. of digestion products	21	18	13
Peptides involving dual cleavage[a]	13 (61.9%)	8 (44.4%)	8 (61.5%)
N. of cleaved peptide bonds[b]	22 (75.9%)	23 (79.3%)	18 (62.1%)
N. of productive species for the ligand[c]	1	1	0
Relative yield of productive cleavages[d]	16.8%	9.0%	0
N. of destructive cleavages for the ligand[e]	6 (27.3%)	6 (26.1%)	4 (22.2%)
Relative yield of the destructive cleavages[d]	19.8%	21.0%	25.9%
Destruction/production ratio for the ligand[f]	1.2	2.3	–

[a] Percent relative to the total number of digestion products.
[b] Percent relative to the total number of peptide bonds in the substrate.
[c] Peptides with the right C-terminus and right or extended N-terminus. Percent values are relative to the total number of digestion products.
[d] Added values from Table VIIS. Productive cleavages correspond to cleavage after R14.
[e] Cleaved peptide bonds implying destruction of the epitope. Percent values are relative to the total number of cleaved peptide bonds.
[f] Ratio between the joint relative yield of cleavages implying destruction of the natural ligand and that of the potentially productive peptides.

than in HeLa cells (47). With MG-132 the inhibition of the tryptic activity was still lower and partial inhibition of the proteasome was clear in gel experiments. The caspase activity was not tested with fluorogenic substrates, but it was reported that epoxomicin and MG-262, a boronate analog of MG-132, have lower effect on this than on the chymotryptic and tryptic subspecificities (47). In the same study the efficacy of inhibitors was protein substrate dependent, presumably as a function of the amino acid composition. Thus, incomplete inhibition of the proteasome might explain the predominance of inhibitor-resistant ligands arising from small basic proteins, because these are comparatively rich in basic residues and might be particularly susceptible to a residual tryptic-like activity.

That a C-terminal basic residue is not a pre-requisite for the observed pattern of parental proteins among inhibitor-resistant ligands was clear from our previous study on HLA-B27 (23) because a significant portion of B27 ligands lack a basic C-terminal motif. However, B27 ligands have R2. Whether the same pattern holds for allotypes lacking basic residues at any anchor position is not known.

The digestion of synthetic peptide substrates addressed the following issues. First, whether inhibitor-resistant ligands or their N-terminal precursors were directly produced by the proteasome *in vitro*. Second, whether the inhibitors abrogated cleavage and the generation of the MHC ligands or their precursors. Third, whether incomplete inhibition of the proteasome resulted in altered cleavage and, if so, whether this would lead to increased production of the natural ligand or their precursors. That significant amounts of productive species for the A68 and B27 ligands were generated from the 31-mer in the absence of inhibitors suggests that they might be produced by the proteasome *in vivo*. Incomplete inhibition of the proteasome altered the cleavage patterns, but these alterations, although failing to abrogate the generation of productive species for the 31-mer, did not result in increased production of the inhibitor-resistant or non-assigned ligands, in contrast to a previously reported example (26).

Some recent advances on the role of cytosolic peptidases in the generation of MHC-bound peptides might provide a plausible explanation for the presentation of inhibitor-resistant

ligands under conditions of decreased proteasomal activity. Bleomycin hydrolase and puromycine-sensitive amino peptidase promote optimal peptide binding and enhanced surface expression of some allotypes (48). In addition, it was recently suggested that some inhibitor-resistant ligands, including those from HLA-B27, could be generated by post-proteasomal processing of protein fragments by the combined action of nardilysin and thimet oligopeptidase, and that the former enzyme could generate the basic C-terminal end of HLA-A3 and -A11 ligands from synthetic precursors *in vitro* (46). The cleavage specificity of nardilysin at dibasic sequences (49, 50) could be particularly suited to generate the C terminus of peptides ending at basic residues. Of the A68 ligands reported in this study (supplemental Table S1) 168 (20.7%) had a basic C+1 residue in the corresponding parental protein (data not shown). However, only 1 of 7 (14.3%) of the inhibitor-resistant ligands showed a dibasic C/C+1 motif (ETVQLRNPPR-RQL), as opposed to 8 of 24 (33.3%) inhibitor-sensitive ligands. Yet this does not exclude a role of nardilysin, because this enzyme can also cleave at monobasic sites (51).

In conclusion, because of incomplete inhibition of the proteasome, inhibitor-resistant ligands are not necessarily generated by non-proteasomal pathways. Yet, their presence is not explained by decreased epitope destruction. It may require further processing of intermediate proteasomal products by other cytosolic enzymes, whose contribution would become more prominent under conditions of proteasome impairment.

Acknowledgments—We thank the staff of the Proteomics facilities at the Centro Nacional de Biotecnología, Madrid (member of the ProteoRed network) for help in MS.

* This work was supported by grants SAF2008/00461 and RD08/0075 from the Ministry of Science and Innovation, and an institutional grant of the Fundación Ramón Areces to the Centro de Biología Molecular *Severo Ochoa* and by the Israel Science Foundation (ISF 916/05 to AA).

S This article contains supplemental Figs. S1 to S3 and Tables S1 to S7.

¶ To whom correspondence should be addressed: Centro de Biología Molecular *Severo Ochoa,* c/ Nicolás Cabrera N. 1, Universidad Autónoma, 28049 Madrid, Spain. Tel.: 34-91-196 4554; Fax: 34-91-196 4420; E-mail: aldecastro@cbm.uam.es.

REFERENCES

1. Rock, K. L., Gramm, C., Rothstein, L., Clark, K., Stein, R., Dick, L., Hwang, D., and Goldberg, A. L. (1994) Inhibitors of the proteasome block the degradation of most cell proteins and the generation of peptides presented on MHC class I molecules. *Cell* **78,** 761–771
2. Henderson, R. A., Michel, H., Sakaguchi, K., Shabanowitz, J., Appella, E., Hunt, D. F., and Engelhard, V. H. (1992) HLA-A2.1-associated peptides from a mutant cell line: a second pathway of antigen presentation. *Science* **255,** 1264–1266
3. Wei, M. L., and Cresswell, P. (1992) HLA-A2 molecules in an antigen-processing mutant cell contain signal sequence-derived peptides. *Nature* **356,** 443–446
4. Snyder, H. L., Bacik, I., Bennink, J. R., Kearns, G., Behrens, T. W., Bächi, T., Orlowski, M., and Yewdell, J. W. (1997) Two novel routes of transporter associated with antigen processing (TAP)-independent major histocompatibility complex class I antigen processing. *J. Exp. Med.* **186,** 1087–1098
5. Gil-Torregrosa, B. C., Raúl Castano, A., and Del Val, M. (1998) Major histocompatibility complex class I viral antigen processing in the secretory pathway defined by the trans-Golgi network protease furin. *J. Exp. Med.* **188,** 1105–1116
6. Gil-Torregrosa, B. C., Castaño, A. R., López, D., and Del Val, M. (2000) Generation of MHC class I peptide antigens by protein processing in the secretory route by furin. *Traffic.* **1,** 641–651
7. Zhang, Y., Kida, Y., Kuwano, K., Misumi, Y., Ikehara, Y., and Arai, S. (2001) Role of furin in delivery of a CTL epitope of an anthrax toxin-fusion protein. *Microbiol. Immunol.* **45,** 119–125
8. Seifert, U., Marañón, C., Shmueli, A., Desoutter, J. F., Wesoloski, L., Janek, K., Henklein, P., Diescher, S., Andrieu, M., de la Salle, H., Weinschenk, T., Schild, H., Laderach, D., Galy, A., Haas, G., Kloetzel, P. M., Reiss, Y., and Hosmalin, A. (2003) An essential role for tripeptidyl peptidase in the generation of an MHC class I epitope. *Nature Immunology* **4,** 375–379
9. Shen, L., Sigal, L. J., Boes, M., and Rock, K. L. (2004) Important role of cathepsin S in generating peptides for TAP-independent MHC class I crosspresentation in vivo. *Immunity.* **21,** 155–165
10. Leonhardt, R. M., Keusekotten, K., Bekpen, C., and Knittler, M. R. (2005) Critical role for the tapasin-docking site of TAP2 in the functional integrity of the MHC class I-peptide-loading complex. *J. Immunol.* **175,** 5104–5114
11. Guil, S., Rodríguez-Castro, M., Aguilar, F., Villasevil, E. M., Antón, L. C., and Del Val, M. (2006) Need for tripeptidyl-peptidase II in major histocompatibility complex class I viral antigen processing when proteasomes are detrimental. *J. Biol. Chem.* **281,** 39925–39934
12. Parmentier, N., Stroobant, V., Colau, D., de Diesbach, P., Morel, S., Chapiro, J., van Endert, P., and Van den Eynde, B. J. (2010) Production of an antigenic peptide by insulin-degrading enzyme. *Nat. Immunol.* **11,** 449–454
13. Griffin, T. A., Nandi, D., Cruz, M., Fehling, H. J., Kaer, L. V., Monaco, J. J., and Colbert, R. A. (1998) Immunoproteasome assembly: cooperative incorporation of interferon gamma (IFN-gamma)-inducible subunits. *J. Exp. Med.* **187,** 97–104
14. Dick, T. P., Ruppert, T., Groettrup, M., Kloetzel, P. M., Kuehn, L., Koszinowski, U. H., Stevanović, S., Schild, H., and Rammensee, H. G. (1996) Coordinated dual cleavages induced by the proteasome regulator PA28 lead to dominant MHC ligands. *Cell* **86,** 253–262
15. Shimbara, N., Nakajima, H., Tanahashi, N., Ogawa, K., Niwa, S., Uenaka, A., Nakayama, E., and Tanaka, K. (1997) Double-cleavage production of the CTL epitope by proteasomes and PA28: role of the flanking region. *Genes Cells* **2,** 785–800
16. Heinemeyer, W., Fischer, M., Krimmer, T., Stachon, U., and Wolf, D. H. (1997) The active sites of the eukaryotic 20 S proteasome and their involvement in subunit precursor processing. *J. Biol. Chem.* **272,** 25200–25209
17. Dick, T. P., Nussbaum, A. K., Deeg, M., Heinemeyer, W., Groll, M., Schirle, M., Keilholz, W., Stevanović, S., Wolf, D. H., Huber, R., Rammensee, H. G., and Schild, H. (1998) Contribution of proteasomal beta-subunits to the cleavage of peptide substrates analyzed with yeast mutants. *J. Biol. Chem.* **273,** 25637–25646
18. Nussbaum, A. K., Dick, T. P., Keilholz, W., Schirle, M., Stevanović, S., Dietz, K., Heinemeyer, W., Groll, M., Wolf, D. H., Huber, R., Rammensee, H. G., and Schild, H. (1998) Cleavage motifs of the yeast 20S proteasome beta subunits deduced from digests of enolase 1. *Proc. Natl. Acad. Sci. U.S.A.* **95,** 12504–12509
19. Kisselev, A. F., Akopian, T. N., Castillo, V., and Goldberg, A. L. (1999) Proteasome active sites allosterically regulate each other, suggesting a cyclical bite-chew mechanism for protein breakdown. *Mol. Cell* **4,** 395–402
20. Kisselev, A. F., Garcia-Calvo, M., Overkleeft, H. S., Peterson, E., Pennington, M. W., Ploegh, H. L., Thornberry, N. A., and Goldberg, A. L. (2003) The caspase-like sites of proteasomes, their substrate specificity, new inhibitors and substrates, and allosteric interactions with the trypsin-like sites. *J. Biol. Chem.* **278,** 35869–35877
21. Benham, A. M., Grommé, M., and Neefjes, J. (1998) Allelic differences in the relationship between proteasome activity and MHC class I peptide loading. *J. Immunol.* **161,** 83–89

22. Luckey, C. J., Marto, J. A., Partridge, M., Hall, E., White, F. M., Lippolis, J. D., Shabanowitz, J., Hunt, D. F., and Engelhard, V. H. (2001) Differences in the expression of human class I MHC alleles and their associated peptides in the presence of proteasome inhibitors. *J. Immunol.* **167,** 1212–1221
23. Marcilla, M., Cragnolini, J. J., and López de Castro, J. A. (2007) Proteasome-independent HLA-B27 ligands arise mainly from small basic proteins. *Mol. Cell Proteomics* **6,** 923–938
24. Rock, K. L., York, I. A., Saric, T., and Goldberg, A. L. (2002) Protein degradation and the generation of MHC class I-presented peptides. *Adv. Immunol.* **80,** 1–70
25. Kisselev, A. F., Akopian, T. N., Castillo, V., and Goldberg, A. L. (1999) Proteasome active sites allosterically regulate each other, suggesting a cyclical bite-chew mechanism for protein breakdown. *Mol. Cell* **4,** 395–402
26. Wherry, E. J., Golovina, T. N., Morrison, S. E., Sinnathamby, G., McElhaugh, M. J., Shockey, D. C., and Eisenlohr, L. C. (2006) Re-evaluating the generation of a "proteasome-independent" MHC class I-restricted CD8 T cell epitope. *J. Immunol.* **176,** 2249–2261
27. Zemmour, J., Little, A. M., Schendel, D. J., and Parham, P. (1992) The HLA-A,B "negative" mutant cell line C1R expresses a novel HLA-B35 allele, which also has a point mutation in the translation initiation codon. *J. Immunol.* **148,** 1941–1948
28. Barnstable, C. J., Bodmer, W. F., Brown, G., Galfre, G., Milstein, C., Williams, A. F., and Ziegler, A. (1978) Production of monoclonal antibodies to group A erythrocytes, HLA and other human cell surface antigens. New tools for genetic analysis. *Cell* **14,** 9–20
29. Kim, K. B., Myung, J., Sin, N., and Crews, C. M. (1999) Proteasome inhibition by the natural products epoxomicin and dihydroeponemycin: insights into specificity and potency. *Bioorg. Med. Chem. Lett.* **9,** 3335–3340
30. Ishihama, Y., Rappsilber, J., Andersen, J. S., and Mann, M. (2002) Microcolumns with self-assembled particle frits for proteomics. *J. Chromatogr. A* **979,** 233–239
31. Beer, I., Barnea, E., Ziv, T., and Admon, A. (2004) Improving large-scale proteomics by clustering of mass spectrometry data. *Proteomics* **4,** 950–960
32. Eng, J. K., McCormack, A. L., and Yates, J. R. (1994) An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass. Spectrom.* **5,** 976–989
33. Perkins, D. N., Pappin, D. J., Creasy, D. M., and Cottrell, J. S. (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20,** 3551–3567
34. Parker, K. C., Bednarek, M. A., and Coligan, J. E. (1994) Scheme for ranking potential HLA-A2 binding peptides based on independent binding of individual peptide side-chains. *J. Immunol.* **152,** 163–175
35. Cragnolini, J. J., Garcia-Medel, N., and Lopez de Castro, J. A. (2009) Endogenous processing and presentation of T-cell Epitopes from chlamydia trachomatis with relevance in HLA-B27-associated reactive arthritis. *Mol. Cell. Proteomics* **80,** 1850–1859
36. Paradela, A., Garcia-Peydro, M., Vázquez, J., Rognan, D., and López de Castro, J. A. (1998) The same natural ligand is involved in allorecognition of multiple HLA-B27 subtypes by a single T cell clone: role of peptide and the MHC molecule in alloreactivity. *J. Immunol.* **161,** 5481–5490
37. Paradela, A., Alvarez, I., Garcia-Peydró, M., Sesma, L., Ramos, M., Vázquez, J., and López de Castro, J. A. (2000) Limited diversity of peptides related to an alloreactive T cell epitope in the HLA-B27-bound peptide repertoire results from restrictions at multiple steps along the processing-loading pathway. *J. Immunol.* **164,** 329–337
38. Gómez, P., Mavian, C., Galocha, B., Garcia-Medel, N., and López de Castro, J. A. (2009) Presentation of cytosolically stable peptides by HLA-B27 is not dependent on the canonic interactions of N-terminal basic residues in the A pocket. *J. Immunol.* **182,** 446–455
39. Medina-Aunon, J. A., Paradela, A., Macht, M., Thiele, H., Corthals, G., and Albar, J. P. (2010) Protein Information and Knowledge Extractor: Discovering biological information from proteomics data. *Proteomics* **10,** 3262–3271
40. Kisselev, A. F., and Goldberg, A. L. (2005) Monitoring activity and inhibition of 26S proteasomes with fluorogenic peptide substrates. *Methods Enzymol.* **398,** 364–378
41. Elsasser, S., Schmidt, M., and Finley, D. (2005) Characterization of the proteasome using native gel electrophoresis. *Methods Enzymol.* **398,** 353–363
42. Ben Dror, L., Barnea, E., Beer, I., Mann, M., and Admon, A. (2010) The HLA-B*2705 peptidome. *Arthritis Rheum.* **62,** 420–429
43. Hickman, H. D., Luis, A. D., Buchli, R., Few, S. R., Sathiamurthy, M., VanGundy, R. S., Giberson, C. F., and Hildebrand, W. H. (2004) Toward a definition of self: proteomic evaluation of the class I peptide repertoire. *J. Immunol.* **172,** 2944–2952
44. Ding, Q., Dimayuga, E., Markesbery, W. R., and Keller, J. N. (2006) Proteasome inhibition induces reversible impairments in protein synthesis. *FASEB J.* **20,** 1055–1063
45. Khan, A. R., Baker, B. M., Ghosh, P., Biddison, W. E., and Wiley, D. C. (2000) The structure and stability of an HLA-A*0201/octameric tax peptide complex with an empty conserved peptide-N-terminal binding site. *J. Immunol.* **164,** 6398–6405
46. Kessler, J. H., Khan, S., Seifert, U., Le Gall, S., Chow, K. M., Paschen, A., Bres-Vloemans, S. A., de Ru, A., van Montfoort, N., Franken, K. L., Benckhuijsen, W. E., Brooks, J. M., van Hall, T., Ray, K., Mulder, A., Doxiadis, I. I., van Swieten, P. F., Overkleeft, H. S., Prat, A., Tomkinson, B., Neefjes, J., Kloetzel, P. M., Rodgers, D. W., Hersh, L. B., Drijfhout, J. W., van Veelen, P. A., Ossendorp, F., and Melief, C. J. (2011) Antigen processing by nardilysin and thimet oligopeptidase generates cytotoxic T cell epitopes. *Nat. Immunol.* **12,** 45–53
47. Kisselev, A. F., Callard, A., and Goldberg, A. L. (2006) Importance of the different proteolytic sites of the proteasome and the efficacy of inhibitors varies with the protein substrate. *J. Biol. Chem.* **281,** 8582–8590
48. Kim, E., Kwak, H., and Ahn, K. (2009) Cytosolic aminopeptidases influence MHC class I-mediated antigen presentation in an allele-dependent manner. *J. Immunol.* **183,** 7379–7387
49. Chesneau, V., Pierotti, A. R., Barré, N., Créminon, C., Tougard, C., and Cohen, P. (1994) Isolation and characterization of a dibasic selective metalloendopeptidase from rat testes that cleaves at the amino terminus of arginine residues. *J. Biol. Chem.* **269,** 2056–2061
50. Chow, K. M., Csuhai, E., Juliano, M. A., St Pyrek, J., Juliano, L., and Hersh, L. B. (2000) Studies on the subsite specificity of rat nardilysin (N-arginine dibasic convertase). *J. Biol. Chem.* **275,** 19545–19551
51. Chow, K. M., Oakley, O., Goodman, J., Ma, Z., Juliano, M. A., Juliano, L., and Hersh, L. B. (2003) Nardilysin cleaves peptides at monobasic sites. *Biochemistry* **42,** 2239–2244

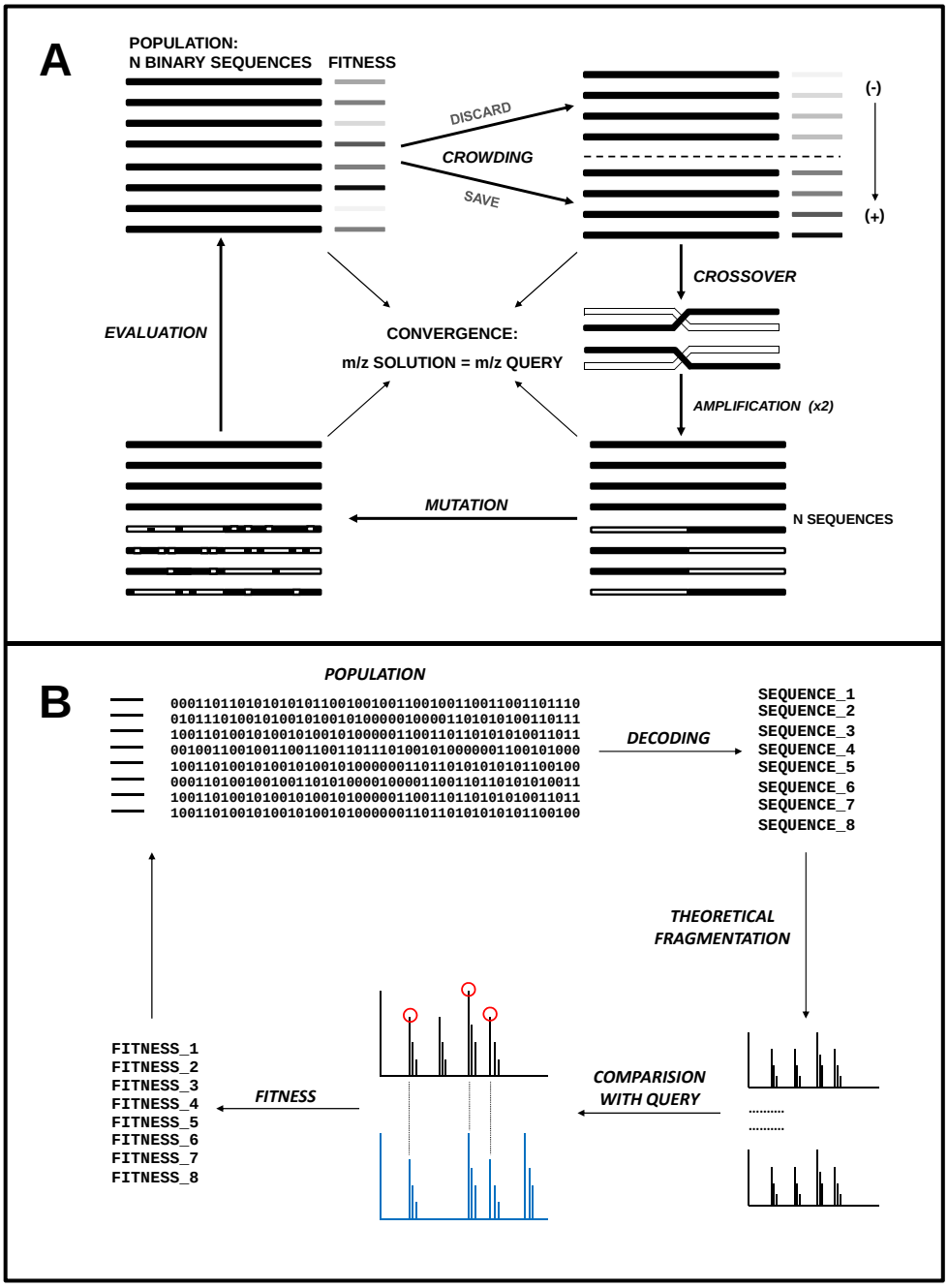

A
POPULATION:
N BINARY SEQUENCES
FITNESS
DISCARD
CROWDING
SAVE
(-)
(+)
CROSSOVER
EVALUATION
CONVERGENCE:
m/z SOLUTION = m/z QUERY
AMPLIFICATION (x2)
MUTATION
N SEQUENCES
B
POPULATION
0001101101010101011001001001100100110011001101110
0101110100101001010010100000100001101010100110111
1001101001010010100101000001100110110101010011011
0010011001001100110011011101001010000001100101000
1001101001010010100101000000110110101010101100100
0001101001001001101010000100001100110110101010011
1001101001010010100101000001100110110101010011011
1001101001010010100101000000110110101010101100100
SEQUENCE_1
SEQUENCE_2
SEQUENCE_3
SEQUENCE_4
SEQUENCE_5
SEQUENCE_6
SEQUENCE_7
SEQUENCE_8
DECODING
THEORETICAL
FRAGMENTATION
COMPARISION
WITH QUERY
FITNESS
FITNESS_1
FITNESS_2
FITNESS_3
FITNESS_4
FITNESS_5
FITNESS_6
FITNESS_7
FITNESS_8

Fig. 1S

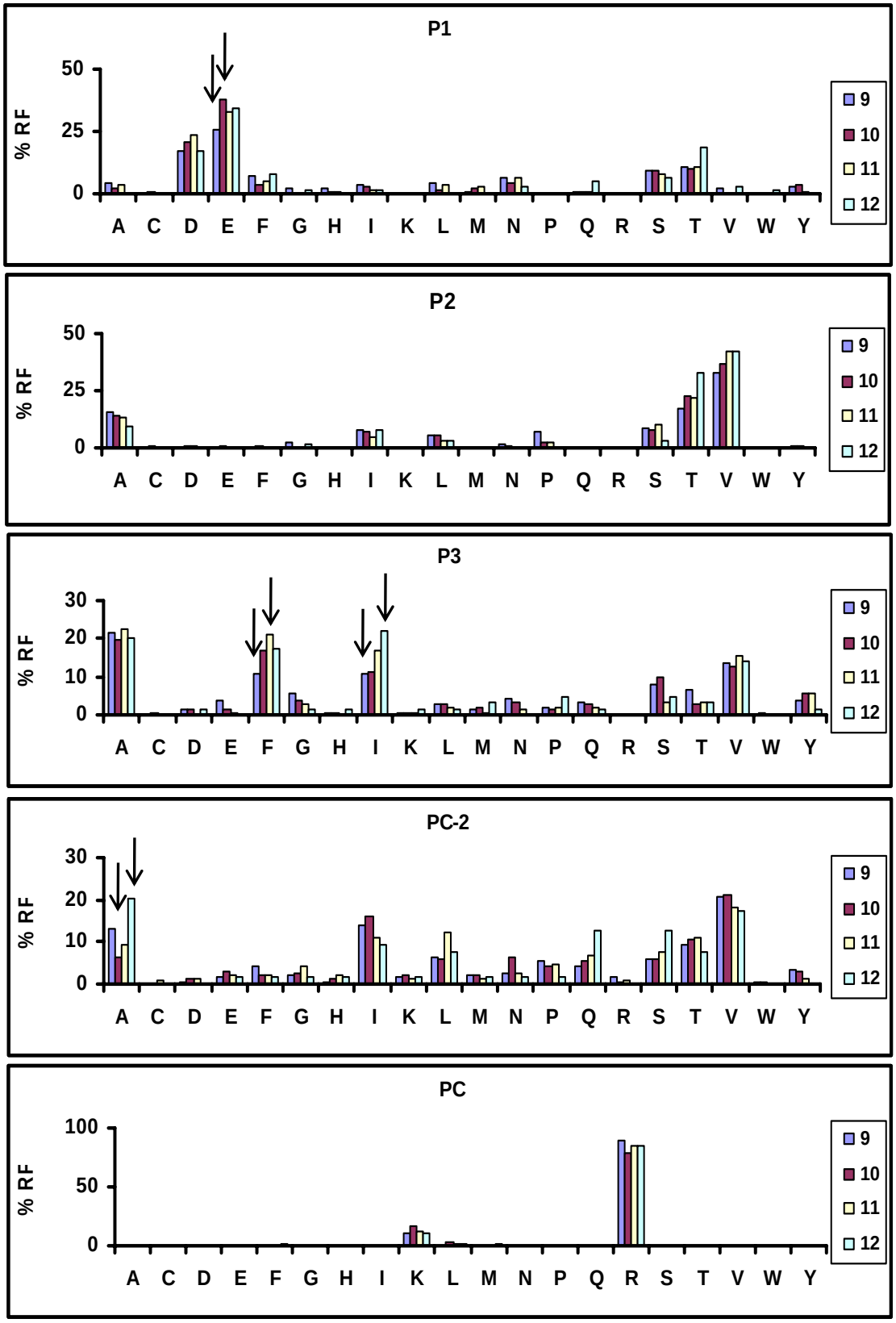

P1
% RF
50
25
0
A C D E F G H I K L M N P Q R S T V W Y
9
10
11
12
P2
% RF
50
25
0
A C D E F G H I K L M N P Q R S T V W Y
9
10
11
12
P3
% RF
30
20
10
0
A C D E F G H I K L M N P Q R S T V W Y
9
10
11
12
PC-2
% RF
30
20
10
0
A C D E F G H I K L M N P Q R S T V W Y
9
10
11
12
PC
% RF
100
50
0
A C D E F G H I K L M N P Q R S T V W Y
9
10
11
12

Figure 2S

A

Sequence	Size	M+H	Variants
DYAIAR	6	708.37	EVT**DYAIAR**
YPKVIER	7	904.48	VV**YPKVIER**
GPSIVHR	7	765.40	ES**GPSIVHR**/EA**GPSIVHR**/ ES**GPSIVHR**K
GSHIIQR	7	810.44	EA**GSHIIQR**
FTKAYQR	7	913.47	
AAVAARR	7	714.43	AV**AAVAARR**
ADIMAKR	7	804.43	SL**ADIMAKR**
AALYKFR	7	868.47	GV**AALYKFR**
VDIATKR	7	802.46	II**VDIATKR**/EII**VDIATKR**
PAVAKLR	7	754.46	
PAVVVGR	7	697.42	
FEKAIQR	7	891.49	VV**FEKAIQR**
RALAPLR	7	796.49	
PVLRLPR	7	850.53	
AAQITQR	7	787.46	DT**AAQITQR**/ TAADT**AAQITQR**
KPPTILR	7	824.48	ELI**KPPTILR**
ASVLQRR	7	829.50	
VSKFLNR	7	863.47	
VEGLITR	7	787.43	TT**VEGLITR**
VSGVVSK	7	675.37	EV**VSGVVSK**
AFGHVQR	7	814.43	
PAVVSER	7	757.40	
TADVHFER	8	974.46	G**TADVHFER**
VFLENVIR	8	989.55	
ILIDPFHK	8	982.54	EV**ILIDPFHK**
PSIVGRPR	8	881.50	
AEITPKGR	8	871.45	LV**AEITPKGR**
PVGIFQKR	8	944.55	Y**PVGIFQKR**/ EVY**PVGIFQKR**
FPSEIVGK	8	876.44	**FPSEIVGK**R/ LV**FPSEIVGK**R
SPLGGIKR	8	827.46	
YGGKSTIR	8	881.43	
TQRELVSR	8	988.54	
IVFGEVDR	8	934.50	
NTDEMVEL	8	950.40	
PSGLLPQR	8	867.49	
MSFLKVDK	8	967.50	
EAFEAIPR	8	932.46	
FSDFTLAR	8	956.48	

B

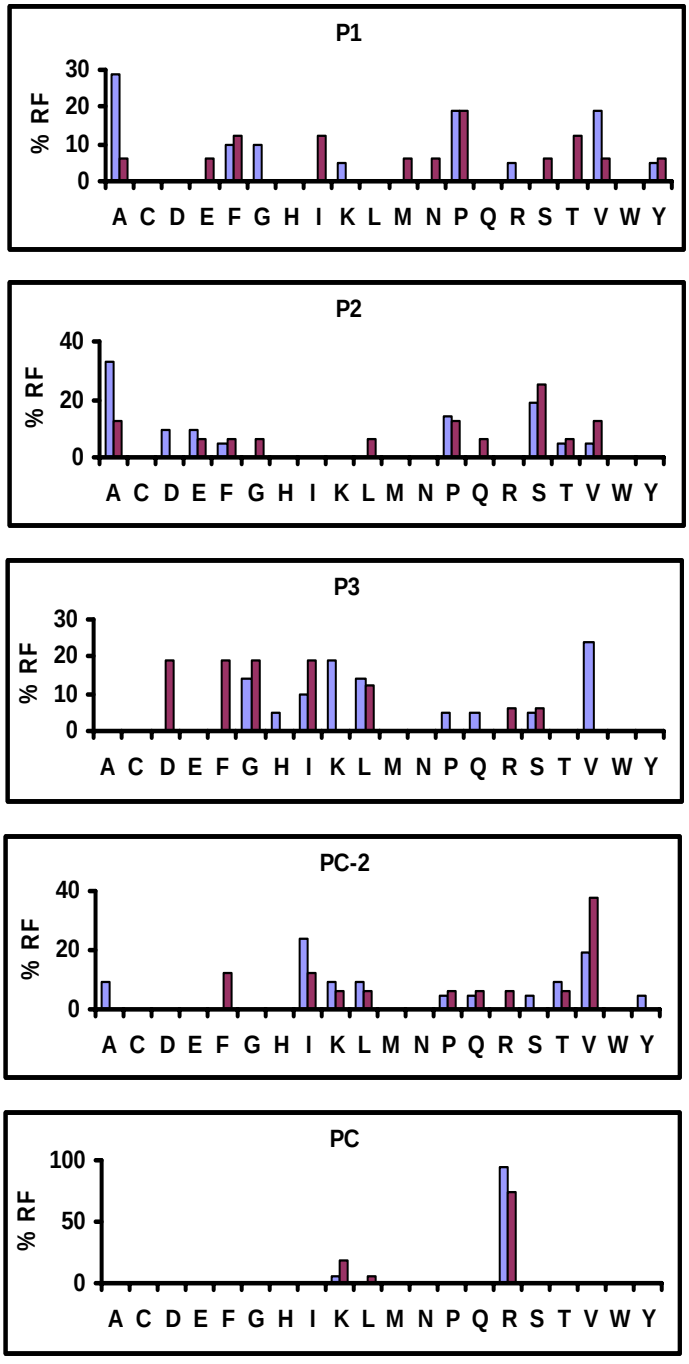


Figure 3S

Supplementary Material

Figure Legends

Fig. 1S. **Flow chart of the MSgene algorithm.** (A) A pool of random binary sequences encoding for peptide sequences are evaluated for their fit with the query spectrum. The 50% of sequences with lowest score are discarded, and the best scored sequences are selected for amplifying the population up to 100 % of the original size with sequences arising from the crossover and mutation of the selected ones. This process is repeated n times, preserving those sequences matching the query mass. (B) Evaluation Function. Each sequence is subjected to a theoretical fragmentation, which provides a list of expected MS/MS ions. These are compared with the query ion list with following criteria: 1) "b" ions and their modifications are matched to query ions by identity, 2) “y” ions are assigned to values resulting from mass of the query ions plus a shift factor consisting of the difference between the mass of the theoretical sequence and the query mass.

Fig. 2S. **Residue usage among A*68:01 ligands at anchor positions as a function of peptide length.** Comparison of the percent residue frequencies (% RF) among 9-mers (N=327), 10-mers (N=205), 11-mers (N=148) and 12-mers (N=64). Arrows indicate differences showing a $p<0.01$ before Bonferroni’s correction, using the χ^2 test. After correction, no statistically significant differences were found.

Fig. 3S. **Amino acid sequence and residue usage of A68:01 ligands shorter than nonamers.** *A,* amino acid sequence of natural A*68:01 ligands shorter than nonamers. The number of residues, theoretical mass (M+H), and related peptides (variants) sequenced from the same peptide pool are indicated (see Table 1S). *B,* Percent residue frequency (% RF) at the anchor positions P1, P2, P3, PC-2 and PC among 7-mers (N=21, blue bars) and 8-mers (N=16, red bars).

Table 1S. A*68:01 ligands

Sequence	AN	Parental Protein	Charge	Score	Source
AAAAAGAFAGR	Q8N697	Solute carrier family 15 member 4	2	4	(1)
AAANPVYGR	P25205	DNA replication licensing factor MCM3	2	3	(1)
AALPGIVHR	Q9Y3I0	UPF0027 protein C22orf28	2	4	(1)
AALYKFR	P09669	Cytochrome c oxidase polypeptide VIc	2	4	(1)
AAQITQR	P30501	HLA class I histocompatibility antigen, Cw-2 alpha chain	2	4	(1)
AAVAARR	Q14697	Neutral alpha-glucosidase AB	2	5	(1)
ACNSLAGIQR	Q58A45	PAB-dependent poly(A)-specific ribonuclease subunit 3	2	3	(1)
ADIMAKR	P83731	60S ribosomal protein L24	2	5	(1)
AEITPKGR	Q9NPA8	Enhancer of yellow 2 transcription factor homolog	2	5	(1)
AFGHVQR	P36551	Coproporphyrinogen III oxidase	2	3	(1)
AITQVVVSR	Q9H9B4	Sideroflexin-1	2	6	(1)
AIVDVVANR	P20073	Annexin A7	2	6	(1)
ALLAVGATK	P40967	Melanocyte protein Pmel 17	--	--	(3)
ALNFPGSQK	P40967	Melanocyte protein Pmel 17	--	--	(3)
AQEVLARAR	Q9H211	DNA replication factor Cdt1	2	5	(1)
ASVLQRR	P31321	cAMP-dependent protein kinase type I-beta regulatory subunit	2	3	(1)
ATAELIMSR	Q9BVK8	Transmembrane protein 147	2	7	(1)
ATAGDGLIELR	P35232	Prohibitin	2	6	(1)
ATAGIIGVNR	Q00610	Clathrin heavy chain 1	2	8	(1)
ATSPLQENR	P52292	Importin subunit alpha-2	2	3	(1)
AVAAVAARR	Q14697	Neutral alpha-glucosidase AB	2	7	(1), (2), (3)
AVANIVNSVK	Q9BRQ8	Apoptosis-inducing factor 2	2	6	(1)
AVFPSIVGR	P63261	Actin, cytoplasmic 2	2	4	(1)
AVIDVGINR	P13995	Bifunctional methylenetetrahydrofolate dehydrogenase/cyclohydrolase	2	4	(1)
AVIQVSQIVAR	Q7Z4W1	L-xylulose reductase	2	7	(1)
AVISPPTVPK	Q9H9Y6	DNA-directed RNA polymerase I subunit RPA2	2	5	(1)
AVSGLGVVGR	P11388	DNA topoisomerase 2-alpha	2	7	(1)
AVTDAIMSR	Q86TM6	E3 ubiquitin-protein ligase synoviolin	2	7	(1)
AVVALAVGR	Q15021	Condensin complex subunit 1	2	5	(1)
AVVDFKAVQGR	P26358	DNA (cytosine-5)-methyltransferase 1	3	5	(1)
AVVGVVAGGGR	P62917	60S ribosomal protein L8	2	7	(1)
CLAAMAVANR	Q9P2H3	Intraflagellar transport protein 80 homolog	2	4	(1)
CTKIESSSV	O75928	E3 SUMO-protein ligase PIAS2	2	5	(1)
DAAALVAGR	Q13395	Probable methyltransferase TARBP1	2	6	(1)
DAAEPDVQR	P57764	Gasdermin-D	2	6	(1)
DAAFIIGSGR	Q14146	Uncharacterized protein KIAA0133	2	7	(1)
DAAGKTTILYK	P61204	ADP-ribosylation factor 3	2	4	(1)
DAAHPTNVQR	P35222	Catenin beta-1	3	4	(1), (3)
DAAHPTNVQRL	P35222	Catenin beta-1	--	--	(2)
DAAIVGAVR	Q9UGQ3	Solute carrier family 2, facilitated glucose transporter member 6	2	7	(1)
DAAMFYTNR	Q01518	Adenylyl cyclase-associated protein 1	2	6	(1)
DAATFSSLR	Q8TD16	Protein bicaudal D homolog 2	2	4	(1)
DAEAIYVTR	Q8NFH3	Nucleoporin Nup43	2	6	(1)
DAFADAVQR	Q92945	Far upstream element-binding protein 2	2	7	(1)
DAFAYVQER	Q8WUJ0	Serine/threonine/tyrosine-interacting protein	2	7	(1)
DAFFNPATR	Q4KMQ2	Anoctamin-6	2	6	(1), (2)

DAFGTGIVEK	P43005	Excitatory amino acid transporter 3	2	6	(1)
DAFNLPSSHLR	Q9NZL9	Methionine adenosyltransferase 2 subunit beta	3	3	(1)
DAGSVPHVNR	Q460N5	Poly [ADP-ribose] polymerase 14	3	5	(1)
DANGNSFATR	P22090	40S ribosomal protein S4, Y isoform 1	2	7	(1)
DASLLIVNR	Q5VTR2	E3 ubiquitin-protein ligase BRE1A	2	5	(1)
DATLLIVNR	O75150	E3 ubiquitin-protein ligase BRE1B	2	6	(1)
DAVDVFETR	Q9BVL2	Nucleoporin p58/p45	2	7	(1)
DAVGFTFPNR	Q9UBB6	Neurochondrin	2	4	(1)
DAVGSNIVVSTR	P28331	NADH-ubiquinone oxidoreductase 75 kDa subunit, mitochondrial	2	7	(1)
DAVNLAFGR	Q8WW24	Tektin-4	2	7	(1)
DAVPCSTTINR	Q12824	SWI/SNF-related matrix-associated actin-dependent regulator of chromatin	2	3	(1)
DAVQLNVIATR	Q8WVX9	Fatty acyl-CoA reductase 1	2	7	(1)
DAVQTVTGGLR	Q9H223	EH domain-containing protein 4	2	7	(1)
DAYGSELVER	P33121	Long-chain-fatty-acid--CoA ligase 1	2	6	(1)
DGIGFAIAR	Q6PKH6	Dehydrogenase/reductase SDR family member 4-like 2	2	6	(1)
DIDSPPITAR	P14618	Pyruvate kinase isozymes M1/M2	2	3	(1)
DIFIDGVVAR	Q8IYB7	DIS3-like exonuclease 2	2	5	(1)
DIFNPQTVGVGR	Q9UMY4	Sorting nexin-12	2	7	(1), (2)
DIFNQVVPR	Q96EB6	NAD-dependent deacetylase sirtuin-1	2	6	(1)
DIFSEVVGR	Q8TDD1	ATP-dependent RNA helicase DDX54	2	7	(1)
DIFSTGTGSQSVER	Q9Y4E1	Protein FAM21C	2	7	(1)
DIFYKAIQK	Q16850	Cytochrome P450 51A1	2	3	(1), (2)
DIIATPGPR	O60216	Double-strand-break repair protein rad21 homolog	2	5	(1)
DIIPFGNNPIFR	Q15392	24-dehydrocholesterol reductase	2	7	(1)
DIISQQLVER	P39210	Protein Mpv17	2	7	(1)
DIMAAVSGR	O75179	Ankyrin repeat domain-containing protein 17	2	6	(1)
DIQGHPFFR	P41743	Protein kinase C iota type	--	--	(2)
DIVKINNQLR	P24928	DNA-directed RNA polymerase II subunit RPB1	2	6	(1)
DIYFNPKTR	Q15029	116 kDa U5 small nuclear ribonucleoprotein component	2	5	(1), (2)
DLASPLIGR	O95067	G2/mitotic-specific cyclin-B2	2	5	(1)
DLEPPVVKR	O14974	Protein phosphatase 1 regulatory subunit 12A	3	6	(1)
DLFDIDPVVAR	Q14669	Probable E3 ubiquitin-protein ligase TRIP12	2	4	(1)
DLFPTSDTPR	Q9H6T3	RNA polymerase II-associated protein 3	2	5	(1)
DLFYNIATR	P40692	DNA mismatch repair protein Mlh1	2	7	(1)
DLLAASQVVNR	Q8N1F7	Nuclear pore complex protein Nup93	2	8	(1)
DLYQGLLLQR	Q63HN8	RING finger protein 213	2	6	(1)
DMFATPQYR	Q86V24	Adiponectin receptor protein 2	2	7	(1)
DPAGVHPPR	Q08J23	tRNA (cytosine-5-)-methyltransferase NSUN2	2	7	(1), (2)
DPFFGGMTR	O00165	HCLS1-associated protein X-1	2	4	(1)
DPFHKAIRR	Q9H6E4	Coiled-coil domain-containing protein 134	--	--	(2)
DPFYGPWYR	Q5U3C3	Transmembrane protein 164	2	4	(1)
DPGEEPPIVRR	Q9P2Q2	FERM domain-containing protein 4A	3	3	(1)
DSADLVAQLR	P78371	T-complex protein 1 subunit beta	2	6	(1)
DSALLPAVR	Q01892	Transcription factor Spi-B	2	3	(1)
DSAPNEMVYR	O00139	Kinesin-like protein KIF2A	2	4	(1)
DSAPPSSVAR	Q9BZE4	Nucleolar GTP-binding protein 1	2	3	(1)
DSFIGENSR	Q99661	Kinesin-like protein KIF2C	2	6	(1)
DSGDVAALR	P40763	Signal transducer and activator of transcription 3	2	7	(1)
DSINVIDVNK	P53611	Geranylgeranyl transferase type-2 subunit beta	2	7	(1)
DSSGTVLHR	Q9Y2J4	Angiomotin-like protein 2	2	4	(1)

DSVFVIVNK	Q8IWA0	WD repeat-containing protein 75	2	6	(1)
DSVLMEQPPLR	Q14204	Cytoplasmic dynein 1 heavy chain 1	2	7	(1)
DTAALSHARL	O60307	Microtubule-associated serine/threonine-protein kinase 3	2	5	(1)
DTAAQITQR	P30501	HLA class I histocompatibility antigen, Cw-2 alpha chain	2	8	(1), (2), (3)
DTAAWTVGR	Q14974	Importin subunit beta-1	2	7	(1)
DTAGKTTILYK	Q8IVW1	ADP-ribosylation factor-like protein 17	2	6	(1)
DTAGTEQFTAMR	A6NIZ1	Ras-related protein Rap-1b-like protein	2	7	(1)
DTAVQYAIGR	Q05048	Cleavage stimulation factor 50 kDa subunit	2	8	(1)
DTFQPEIMER	Q96RT1	Protein LAP2	2	4	(1)
DTFSYNFYNR	Q9HC16	DNA dC- _gt_ dU-editing enzyme APOBEC-3G	2	6	(1)
DTFSYPIER	Q9H568	Actin-like protein 8	2	5	(1)
DTGEKMILIGR	Q9H9B4	Sideroflexin-1	3	4	(1)
DTGNIGQER	Q01082	Spectrin beta chain, brain 1	--	--	(2)
DTIDERTINKK	P13796	Plastin-2	3	3	(1)
DTIDSSEVNVAK	Q53R41	FAST kinase domain-containing protein 1	2	7	(1)
DTIEIITDR	P22626	Heterogeneous nuclear ribonucleoproteins A2/B1	2	6	(1), (3)
DTITTDTGFMR	Q13868	Exosome complex exonuclease RRP4	2	3	(1)
DTIVSGTLPQR	P49327	Fatty acid synthase	2	7	(1)
DTIVSGTLPQRM	P49327	Fatty acid synthase	2	4	(1)
DTMSVISGISSR	Q8N3U4	Cohesin subunit SA-2	2	5	(1)
DTPGFIVNR	Q16836	Hydroxyacyl-coenzyme A dehydrogenase, mitochondrial	2	3	(1), (2)
DTPGIQIGR	P11310	Medium-chain specific acyl-CoA dehydrogenase, mitochondrial	2	3	(1)
DTQELTQILR	Q5SQN1	t-SNARE coiled-coil homology domain-containing protein C1orf142	2	7	(1)
DTQSGSLLFIGR	P50454	Serpin H1	2	7	(1)
DTSFTTLVER	Q06210	Glucosamine--fructose-6-phosphate aminotransferase [isomerizing] 1	2	6	(1)
DTSGGTTALR	Q16531	DNA damage-binding protein 1	2	8	(1)
DTSHIPPLNR	Q9H9L3	Interferon-stimulated 20 kDa exonuclease-like 2	3	3	(1)
DTSSLPTVIMR	Q12802	A-kinase anchor protein 13	2	7	(1)
DTTVGTLSQR	P51665	26S proteasome non-ATPase regulatory subunit 7	2	8	(1)
DTVAQITQR	P30511	HLA class I histocompatibility antigen, alpha chain F	2	7	(1)
DTVDLTSVLR	Q9NX46	Poly(ADP-ribose) glycohydrolase ARH3	2	7	(1)
DTVFVGTAGTVGR	Q86VY4	Testis-specific Y-encoded-like protein 5	2	8	(1)
DTVTGLLER	Q9H8H0	Nucleolar protein 11	2	7	(1)
DTYPALPPSAR	Q96S16	JmjC domain-containing protein 8	2	3	(1)
DTYSFQVINK	Q9H583	HEAT repeat-containing protein 1	2	7	(1)
DVAAAQTGVSR	Q9BV10	Dolichyl-P-Man:Man(7)GlcNAc(2)-PP-dolichyl-alpha-1,6-mannosyltransferase	2	6	(1)
DVAEEIANYR	Q01082	Spectrin beta chain, brain 1	2	7	(1)
DVAGAEALLDR	Q13813	Spectrin alpha chain, brain	2	7	(1)
DVAGGTGDIAFR	Q5HYK3	Ubiquinone biosynthesis methyltransferase COQ5, mitochondrial	2	7	(1)
DVAMTGEITLR	Q86WA8	Peroxisomal Lon protease homolog 2	2	7	(1)
DVAQEIAER	Q8NBL1	KTEL motif-containing protein 1	2	4	(1)
DVAQGTQVTGR	Q9UQ80	Proliferation-associated protein 2G4	2	7	(1)
DVAQPDQTR	P22087	rRNA 2 _quote_ -O-methyltransferase fibrillarin	2	4	(1), (2)
DVAQQFQYAVR	Q96CW5	Gamma-tubulin complex component 3	2	6	(1)
DVAQWNIGSLR	O94953	JmjC domain-containing histone demethylation protein 3B	2	4	(1)
DVASPDGLGRL	O95365	Zinc finger and BTB domain-containing protein 7A	2	4	(1)
DVASVIVTK	P61011	Signal recognition particle 54 kDa protein	2	7	(1)
DVASVIVTKL	P61011	Signal recognition particle 54 kDa protein	2	6	(1)
DVAVVAGGLGR	Q15904	V-type proton ATPase subunit S1	2	7	(1)
DVAYINVETK	Q9P0R6	GSK3-beta interaction protein	2	6	(1)

DVFDPHGTLSGGAR	O95347	Structural maintenance of chromosomes protein 2	2	8	(1), (2)
DVFGNYVIQK	Q8TB72	Pumilio homolog 2	2	3	(1)
DVFGQATAVLR	Q8WXF8	DNA-binding death effector domain-containing protein 2	2	6	(1)
DVFIIGAVAK	O43808	Peroxisomal membrane protein PMP34	2	6	(1)
DVFIPKPFR	Q13148	TAR DNA-binding protein 43	--	--	(2)
DVFISAAER	P20618	Proteasome subunit beta type-1	2	4	(1)
DVFRDPALK	P61353	60S ribosomal protein L27	3	5	(1), (2), (3)
DVFRDPALKR	P61353	60S ribosomal protein L27	3	3	(1), (2)
DVFSGVKPLTR	Q12768	Strumpellin	3	3	(1)
DVFTDIDIFR	Q9H4H8	Protein FAM83D	2	4	(1)
DVFVVGTER	P78347	General transcription factor II-I	2	6	(1), (2)
DVFVVGTERGR	P78347	General transcription factor II-I	2	6	(1)
DVFYLQPER	Q5VZL5	Zinc finger MYM-type protein 4	2	3	(1)
DVIAINYNEK	Q92890	Ubiquitin fusion degradation protein 1 homolog	2	7	(1)
DVIGQTITISR	Q6UN15	Pre-mRNA 3 _quote_ -end-processing factor FIP1	2	5	(1)
DVIKTEGATLR	Q53H82	Beta-lactamase-like protein 2	2	4	(1)
DVINPMALR	O43290	U4/U6.U5 tri-snRNP-associated protein 1	2	4	(1)
DVISIDKTGENFR	P62701	40S ribosomal protein S4, X isoform	3	8	(1)
DVISNINITR	Q13144	Translation initiation factor eIF-2B subunit epsilon	2	7	(1)
DVITIDKATGK	Q9Y230	RuvB-like 2	2	4	(1)
DVKTTDGYLLR	P61247	40S ribosomal protein S3a	3	5	(1)
DVMTALSQGYR	P07948	Tyrosine-protein kinase Lyn	2	5	(1)
DVMVGPFKLR	Q9Y2D5	A-kinase anchor protein 2	--	--	(3)
DVNQQEFVR	P39019	40S ribosomal protein S19	2	8	(1)
DVSAAWSGSGR	Q9H4I9	UPF0466 protein C22orf32, mitochondrial	2	7	(1)
DVSHPIIYR	Q9UK73	Protein fem-1 homolog B	3	3	(1), (2)
DVSNPQTVGVGR	O60493	Sorting nexin-3	2	7	(1)
DVSPVKPLSR	Q8WYQ5	Protein DGCR8	2	6	(1)
DVVDEKISAMGK	Q9UL33	Trafficking protein particle complex subunit 2-like protein	2	3	(1)
DVVFVIDTGR	Q7Z478	Putative ATP-dependent RNA helicase DHX29	2	5	(1)
DVYDAFPPLR	Q9Y3D9	28S ribosomal protein S23, mitochondrial	2	5	(1)
DVYENLYAGR	P49137	MAP kinase-activated protein kinase 2	2	7	(1)
DVYQEMPAQLR	P12694	2-oxoisovalerate dehydrogenase subunit alpha, mitochondrial	2	7	(1)
DVYSGTPTK	Q14493	Histone RNA hairpin-binding protein	1	4	(1)
DVYTDQVIKR	Q75QN2	Integrator complex subunit 8	2	4	(1)
DVYVPEASR	P49757	Protein numb homolog	2	3	(1)
DYAIAR	Q9Y2G3	Probable phospholipid-transporting ATPase IF	1	3	(1)
EAAEVLQNNR	P61081	NEDD8-conjugating enzyme Ubc12	2	7	(1)
EAAEYIAQAR	P56589	Peroxisomal biogenesis factor 3	2	6	(1)
EAAPSEAALVSR	Q8IZY2	ATP-binding cassette sub-family A member 7	2	4	(1)
EAASYLDQISR	P61289	Proteasome activator complex subunit 3	2	5	(1)
EAEVASLNR	P06753	Tropomyosin alpha-3 chain	2	5	(1)
EAFDFVKQR	Q05923	Dual specificity protein phosphatase 2	3	6	(1)
EAFEAIPR	P50990	T-complex protein 1 subunit theta	2	3	(1)
EAFEHPNVVR	P11802	Cell division protein kinase 4	2	6	(1), (2)
EAFETFINKR	Q13620	Cullin-4B	2	7	(1)
EAFKPILSTR	O95456	Proteasome assembly chaperone 1	2	4	(1)
EAFNMIDQNR	P24844	Myosin regulatory light polypeptide 9	2	7	(1)
EAFRNLVPVTR	Q13342	Nuclear body protein SP140	2	6	(1)
EAFSQFGPIER	P23246	Splicing factor, proline- and glutamine-rich	2	7	(1)

EAFSVFGQVER	Q15233	Non-POU domain-containing octamer-binding protein	2	5	(1), (2)
EAFVIDAVR	Q7L2H7	Eukaryotic translation initiation factor 3 subunit M	2	5	(1)
EAGAGSATEFQFR	P46783	40S ribosomal protein S10	2	6	(1)
EAGLGALPR	Q9NR77	Peroxisomal membrane protein 2	2	5	(1)
EAGPSIVHR	P62736	Actin, aortic smooth muscle	2	3	(1), (3)
EAGSHIIQR	P10319	HLA class I histocompatibility antigen, B-58 alpha chain	2	8	(1)
EAIEGTYIDK	P62280	40S ribosomal protein S11	2	6	(1)
EAIGNEVTR	P42356	Phosphatidylinositol 4-kinase alpha	2	6	(1)
EAISLLNSER	Q9UPN3	Microtubule-actin cross-linking factor 1, isoforms 1/2/3/5	2	4	(1)
EALNQVTQR	Q05655	Protein kinase C delta type	2	4	(1)
EALPSDIAAEAR	Q9GZR7	ATP-dependent RNA helicase DDX24	2	6	(1)
EAMEWVIHK	Q9Y3B1	Protein slowmo homolog 2	2	4	(1)
EATAGFTIGR	Q93074	Mediator of RNA polymerase II transcription subunit 12	2	4	(1)
EATGAYVPGR	Q14653	Interferon regulatory factor 3	2	4	(1)
EAVAPEFTNR	Q9H0U6	39S ribosomal protein L18, mitochondrial	2	3	(1)
EAVATAVQR	O60645	Exocyst complex component 3	2	7	(1)
EAVAVDVAVPK	Q6PII5	Hydroxyacylglutathione hydrolase-like protein	2	5	(1)
EAVGGLQTVR	Q03519	Antigen peptide transporter 2	2	7	(1)
EAVTPSPSFQQR	Q9UKG1	DCC-interacting protein 13-alpha	2	4	(1)
EAWVSSSLFGVSR	Q9Y255	PRELI domain-containing protein 1, mitochondrial	2	6	(1)
EAYGAVTQTVR	Q15126	Phosphomevalonate kinase	2	5	(1)
EFVDLYVPR	P63220	40S ribosomal protein S21	2	6	(1), (2)
EGAGILLTR	Q99595	Mitochondrial import inner membrane translocase subunit Tim17-A	2	3	(1)
EGFDPTGPAGLGR	Q9H9B1	Histone-lysine N-methyltransferase, H3 lysine-9 specific 5	2	6	(1)
EGFFLDASR	Q9NVK5	FGFR1 oncogene partner 2	2	3	(1)
EGTVVVTER	Q14126	Desmoglein-2	2	6	(1)
EGVGILLTR	O60830	Mitochondrial import inner membrane translocase subunit Tim17-B	2	5	(1)
EIADLSNIINK	Q9GZL7	WD repeat-containing protein 12	2	4	(1)
EIADTMVQIFK	P14868	Aspartyl-tRNA synthetase, cytoplasmic	2	5	(1)
EIAMATVTALR	P09972	Fructose-bisphosphate aldolase C	2	8	(1), (3)
EIAQAEEQAR	Q969G3	SWI/SNF-related matrix-associated actin-dependent regulator of chromatin	2	5	(1)
EIASSLPAAQR	A2VDJ0	Transmembrane protein 131-like	2	5	(1)
EIFGTIGMR	Q99873	Protein arginine N-methyltransferase 1	2	6	(1)
EIFQNEVAR	Q9BS26	Thioredoxin domain-containing protein 4	2	6	(1)
EIFYTAIVNR	Q9NRZ9	Lymphoid-specific helicase	2	7	(1)
EIHGGAGGPSGR	Q16763	Ubiquitin-conjugating enzyme E2 S	--	--	(2)
EIINIMQDR	O96000	NADH dehydrogenase [ubiquinone] 1 beta subcomplex subunit 10	2	3	(1)
EIIQSTPYGR	Q7Z401	C-myc promoter-binding protein	2	7	(1)
EIIVDIATKR	Q96IY1	Kinetochore-associated protein NSL1 homolog	2	7	(1)
EILELAGNAAR	O75367	Core histone macro-H2A.1	2	5	(1)
EILGIANNR	Q86UX7	Fermitin family homolog 3	2	3	(1)
EINAILQKR	Q99707	Methionine synthase	2	4	(1)
EIPALSVSR	P43686	26S protease regulatory subunit 6B	2	4	(1)
EITPPVVLR	P06748	Nucleophosmin	2	6	(1)
EIVGNLPSAMR	Q8NBE8	Kelch-like protein 23	2	4	(1)
EIVGVVNVR	Q96AA3	Protein RFT1 homolog	2	7	(1)
EIVPGDIVSIGR	Q9HD20	Probable cation-transporting ATPase 13A1	2	3	(1)
EIYRPPASR	Q6PJT7	Zinc finger CCCH domain-containing protein 14	--	--	(2)
ELAQVIVSR	Q49A26	Nuclear protein NP60	2	7	(1)
ELDPALAPVVSR	O00170	AH receptor-interacting protein	2	8	(1)

ELFEDSQLTTR	Q8NEY8	Periphilin-1	2	7	(1)
ELFHQQGTPR	P20700	Lamin-B1	2	4	(1), (2)
ELFHVEPGLQR	Q96T88	E3 ubiquitin-protein ligase UHRF1	2	5	(1), (2)
ELFIPPKINR	O95163	Elongator complex protein 1	2	3	(1)
ELFPAPILR	Q7L513	Fc receptor-like A	2	3	(1), (2)
ELIDNSLSATSR	A6NHR9	Structural maintenance of chromosomes flexible hinge domain-containing protein	2	7	(1)
ELIKPPTILR	P53677	AP-3 complex subunit mu-2	3	8	(1), (2), (3)
ELIQDITQR	P26038	Moesin	2	6	(1)
ELISTTANYR	Q13123	Protein Red	2	4	(1)
ELMGYIMGK	P61599	N-acetyltransferase 5	2	3	(1)
ELNPSTGNR	Q29RF7	Sister chromatid cohesion protein PDS5 homolog A	2	5	(1)
ELPIVTPALR	P52292	Importin subunit alpha-2	2	7	(1)
ELQSALLLR	P62253	Ubiquitin-conjugating enzyme E2 G1	2	4	(1)
ELSNVLAAMR	Q9Y3U8	60S ribosomal protein L36	2	6	(1)
ELTAVVQKR	P23396	40S ribosomal protein S3	2	6	(1), (2)
ELVNVLQER	Q6ZYL4	General transcription factor IIH subunit 5	2	5	(1)
ELYAEKVATR	P23396	40S ribosomal protein S3	2	--	(1)
ELYTISVER	A3KN83	Protein strawberry notch homolog 1	2	3	(1)
ENTAITIGR	Q92973	Transportin-1	2	3	(1)
EPAPGTNQR	Q9UBW8	COP9 signalosome complex subunit 7a	--	--	(2)
EPLPPTIGR	Q9Y6C9	Mitochondrial carrier homolog 2	2	3	(1)
EPSEVFINR	Q8WXF1	Paraspeckle component 1	--	--	(2)
EPYANPTKR	P52272	Heterogeneous nuclear ribonucleoprotein M	--	--	(2)
ESAAINQILGR	Q9UJU2	Lymphoid enhancer-binding factor 1	2	6	(1)
ESAIAIATGK	P09668	Cathepsin H	2	5	(1)
ESAPEGQAQQR	P67809	Nuclease-sensitive element-binding protein 1	2	7	(1), (2)
ESFASDPILYR	P14618	Pyruvate kinase isozymes M1/M2	2	7	(1)
ESFGPQPISR	Q06609	DNA repair protein RAD51 homolog 1	2	7	(1)
ESFPELTIITR	Q2KHR3	Glutamine and serine-rich protein 1	2	3	(1)
ESFPGSFRGR	Q9BWT1	Cell division cycle-associated protein 7	--	--	(2)
ESFPSIGPQRL	Q9Y6J0	Calcineurin-binding protein cabin-1	2	4	(1)
ESFSDYPPLGR	P68104	Elongation factor 1-alpha 1	2	7	(1)
ESFTMASSPAQR	P49736	DNA replication licensing factor MCM2	2	7	(1)
ESGGASVISTR	Q9UH99	Protein unc-84 homolog B	2	3	(1)
ESGPSIVHR	P63261	Actin, cytoplasmic 2	2	8	(1), (2), (3)
ESGPSIVHRK	Q6S8J3	ANKRD26-like family C member 1A	--	--	(3)
ESIAAFIQR	P46379	Large proline-rich protein BAT3	2	5	(1)
ESIDPQFTIR	Q9Y2L9	Leucine-rich repeat and calponin homology domain-containing protein 1	2	4	(1)
ESQSPDTTIQR	Q92973	Transportin-1	2	5	(1)
ESSAGGSFTVR	Q14568	Putative heat shock protein HSP 90-alpha A2	2	7	(1)
ESTGSIAKR	P04075	Fructose-bisphosphate aldolase A	--	--	(3)
ESVGTNADTPVLR	Q15468	SCL-interrupting locus protein	2	5	(1)
ESVPEFPLSPPK	P16949	Stathmin	2	7	(1)
ESVPPGTTISR	Q6P1K2	Polyamine-modulated factor 1	2	5	(1)
ETADFKVYTK	P21333	Filamin-A	2	3	(1)
ETAGGVMTVLIKR	P11142	Heat shock cognate 71 kDa protein	3	3	(1)
ETANADIVTLLR	Q15027	ARFGAP with coiled-coil, ANK repeat and PH domain-containing protein 1	2	7	(1)
ETASNEVVYR	Q99661	Kinesin-like protein KIF2C	2	7	(1)
ETESVVSLR	O14641	Segment polarity protein dishevelled homolog DVL-2	2	6	(1)
ETFDFQAMGR	Q14204	Cytoplasmic dynein 1 heavy chain 1	2	5	(1)

ETFGSSQLPDK	Q14005	Interleukin-16	2	6	(1)
ETFNYNAIDR	Q8TCG2	Phosphatidylinositol 4-kinase type 2-beta	2	7	(1)
ETFSGVYKK	P62081	40S ribosomal protein S7	2	6	(1)
ETFSGVYKKL	P62081	40S ribosomal protein S7	2	4	(1)
ETIGEILKK	P61978	Heterogeneous nuclear ribonucleoprotein K	--	--	(3)
ETILASSGTDR	Q16576	Histone-binding protein RBBP7	2	7	(1)
ETILTSGENLAR	O43813	LanC-like protein 1	2	7	(1)
ETIPLTAEKL	P24385	G1/S-specific cyclin-D1	--	--	(3)
ETIYQNYQR	P49736	DNA replication licensing factor MCM2	2	3	(1)
ETLGSASTSTPGR	Q96FF9	Sororin	2	6	(1)
ETLPPVGVNR	P49753	Acyl-coenzyme A thioesterase 2, mitochondrial	2	3	(1)
ETLSAVTNNR	Q9NRA8	Eukaryotic translation initiation factor 4E transporter	2	7	(1)
ETNEIANANSR	P84098	60S ribosomal protein L19	2	5	(1)
ETNTGEVYR	P62318	Small nuclear ribonucleoprotein Sm D3	2	7	(1)
ETNYGIPQR	Q86UW1	Organic solute transporter subunit alpha	--	--	(2)
ETPSTSNVSGR	Q9HCM1	Uncharacterized protein C12orf35	2	6	(1)
ETQGQQPPQR	P67809	Nuclease-sensitive element-binding protein 1	2	3	(1), (2)
ETQPPVALK	Q8NC51	Plasminogen activator inhibitor 1 RNA-binding protein	2	6	(1)
ETSSLESFVR	Q9ULT8	E3 ubiquitin-protein ligase HECTD1	2	3	(1)
ETTPLTIEKL	P30281	G1/S-specific cyclin-D3	2	5	(1)
ETTVYVIVR	Q6ZUJ8	Phosphoinositide 3-kinase adapter protein 1	2	3	(1)
ETVDPTYIFHPR	O95155	Ubiquitin conjugation factor E4 B	3	7	(1), (2)
ETVGFGMLK	Q9Y2Q5	Mitogen-activated protein-binding protein-interacting protein	2	3	(1)
ETVGSIFSR	Q9Y3D9	28S ribosomal protein S23, mitochondrial	2	6	(1)
ETVGTGIMGR	P32019	Type II inositol-1,4,5-trisphosphate 5-phosphatase	2	8	(1)
ETVNVEPLTK	Q14203	Dynactin subunit 1	2	6	(1)
ETVPATQTQTTTR	Q9NTW7	Zinc finger protein 64 homolog, isoforms 3 and 4	2	3	(1)
ETVQLRNPPR	Q13043	Serine/threonine-protein kinase 4	3	3	(1), (2)
ETYEEIYKSTK	P62310	U6 snRNA-associated Sm-like protein LSm3	2	5	(1)
ETYKYLYGR	Q15404	Ras suppressor protein 1	--	--	(2)
ETYLNENLR	Q9UQE7	Structural maintenance of chromosomes protein 3	2	6	(1)
ETYTEIAGMQR	Q7Z7G8	Vacuolar protein sorting-associated protein 13B	2	7	(1)
EVAGFGESR	P27708	CAD protein	2	5	(1)
EVAGIINDAK	Q14008	Cytoskeleton-associated protein 5	2	3	(1)
EVAGIVAAR	O00154	Cytosolic acyl coenzyme A thioester hydrolase	2	6	(1)
EVAGLQVGR	Q8TCU6	Phosphatidylinositol 3,4,5-trisphosphate-dependent Rac exchanger 1 protein	2	4	(1)
EVAKLDGTQR	P14543	Nidogen-1	2	3	(1)
EVAPPASGTR	Q9NR46	Endophilin-B2	2	6	(1)
EVAPPEYHR	Q9UBU8	Mortality factor 4-like protein 1	2	4	(1), (2), (3)
EVAPPEYHRK	Q9UBU8	Mortality factor 4-like protein 1	2	3	(1), (3)
EVAQFLTGR	Q6ICB0	UPF0326 protein FAM152B	2	5	(1)
EVAQLIQGGR	Q9Y5U8	Brain protein 44-like protein	2	6	(1), (2), (3)
EVCGQDITTK	P30153	Serine/threonine-protein phosphatase 2A 65 kDa regulatory subunit A alpha isoform	2	3	(1)
EVDAAMAAR	P22626	Heterogeneous nuclear ribonucleoproteins A2/B1	2	6	(1)
EVENIISSMR	Q9UFH2	Dynein heavy chain 17, axonemal	2	6	(1)
EVENPQNQLR	O43556	Epsilon-sarcoglycan	2	3	(1)
EVFEYSSSTR	Q460N3	Poly [ADP-ribose] polymerase 15	2	3	(1)
EVFLDTIISR	Q96CW5	Gamma-tubulin complex component 3	2	6	(1)
EVFPTAGAGTDLR	P28331	NADH-ubiquinone oxidoreductase 75 kDa subunit, mitochondrial	2	6	(1)
EVFSHILKR	P33076	MHC class II transactivator	--	--	(2)

EVFSIAGTVKR	Q9P2K5	Myelin expression factor 2	2	5	(1)
EVFSMAGVVVR	P52272	Heterogeneous nuclear ribonucleoprotein M	2	7	(1)
EVGDIVKVTR	P46109	Crk-like protein	2	4	(1)
EVGDIVVGR	Q13868	Exosome complex exonuclease RRP4	2	7	(1)
EVGDIVVGRI	Q13868	Exosome complex exonuclease RRP4	2	6	(1)
EVGEFPVVVQR	Q7L591	Docking protein 3	2	4	(1)
EVGSYPILR	Q5QP82	WD repeat-containing protein 32	2	3	(1)
EVIALFDQTR	Q5SRE5	Nucleoporin NUP188 homolog	2	5	(1)
EVIAYILER	Q9ULK4	Mediator of RNA polymerase II transcription subunit 23	2	6	(1)
EVIEIMTDR	P09651	Heterogeneous nuclear ribonucleoprotein A1	2	3	(1)
EVIGLLGGR	Q5VVJ2	Protein MYSM1	2	3	(1)
EVILIDPFHK	P61313	60S ribosomal protein L15	2	8	(1), (2), (3)
EVIPITDSELR	Q5MIZ7	Serine/threonine-protein phosphatase 4 regulatory subunit 3B	2	4	(1)
EVIPYTPAMQR	P09601	Heme oxygenase 1	2	4	(1)
EVIQLQGDQR	O60739	Eukaryotic translation initiation factor 1b	2	7	(1)
EVIQNPATITR	Q9Y4X5	Protein ariadne-1 homolog	2	7	(1)
EVISLSQVTPK	O60313	Dynamin-like 120 kDa protein, mitochondrial	2	7	(1)
EVISSAVDALQR	P42695	Condensin-2 complex subunit D3	2	7	(1)
EVISTPGVVAR	P52294	Importin subunit alpha-1	2	8	(1)
EVISTPGVVARF	P52294	Importin subunit alpha-1	2	5	(1)
EVITEEAAAALR	Q9Y4D8	Probable E3 ubiquitin-protein ligase KIAA0614	2	7	(1)
EVLDRVFER	Q02548	Paired box protein Pax-5	3	3	(1)
EVLKIMPVQK	P15880	40S ribosomal protein S2	2	3	(1)
EVLNGAAVVR	Q9UBN7	Histone deacetylase 6	2	6	(1)
EVLQGGSQR	P22415	Upstream stimulatory factor 1	2	3	(1)
EVLQPLQQR	Q9C0E2	Exportin-4	2	3	(1)
EVMGTGAVR	Q969K3	E3 ubiquitin-protein ligase RNF34	2	3	(1)
EVNDPSLTIK	P00403	Cytochrome c oxidase subunit 2	2	3	(1)
EVNDRPVRR	Q9Y520	Protein BAT2-like 2	--	--	(2)
EVNGAPVYVVR	Q9Y2G3	Probable phospholipid-transporting ATPase IF	2	3	(1)
EVNQETIQR	Q9UHA2	SS18-like protein 2	2	4	(1)
EVNQETIQRL	Q9UHA2	SS18-like protein 2	2	5	(1)
EVNQPKFNQR	Q9HAU5	Regulator of nonsense transcripts 2	--	--	(2)
EVPAAIAPFQGR	Q15393	Splicing factor 3B subunit 3	2	5	(1)
EVPDVTATPARL	Q07820	Induced myeloid leukemia cell differentiation protein Mcl-1	2	4	(1)
EVPFSNENPR	Q92835	Phosphatidylinositol-3,4,5-trisphosphate 5-phosphatase 1	2	4	(1)
EVPGGGIITGIGR	Q9HCC0	Methylcrotonoyl-CoA carboxylase beta chain, mitochondrial	2	6	(1)
EVPSFLVER	Q9NXB0	Meckel syndrome type 1 protein	--	--	(2)
EVPSSSGINSTK	Q9Y5Q9	General transcription factor 3C polypeptide 3	2	7	(1)
EVQDTSGGTTALR	Q16531	DNA damage-binding protein 1	2	8	(1)
EVQPKPHHR	Q13469	Nuclear factor of activated T-cells, cytoplasmic 2	--	--	(2)
EVQVASQSSQR	P12270	Nucleoprotein TPR	2	3	(1)
EVSDVEIITR	Q76FK4	Nucleolar protein 8	2	6	(1)
EVSLQGINTR	Q9Y6W5	Wiskott-Aldrich syndrome protein family member 2	2	3	(1)
EVSPMYGVQK	Q9H3K2	Growth hormone-inducible transmembrane protein	2	3	(1)
EVSQLTGEVAK	O60610	Protein diaphanous homolog 1	2	4	(1)
EVSVVPPSR	Q2TAY7	WD40 repeat-containing protein SMU1	2	3	(1)
EVTDYAIAR	Q14566	DNA replication licensing factor MCM6	2	6	(1)
EVTDYTTGR	O60488	Long-chain-fatty-acid--CoA ligase 4	2	6	(1)
EVTELAGYTAR	Q9UBJ2	ATP-binding cassette sub-family D member 2	2	3	(1)

EVTGQETVAQIK	P35544	Ubiquitin-like protein FUBI	2	7	(1)
EVTPSSTTASVAR	Q8IZT6	Abnormal spindle-like microcephaly-associated protein	2	3	(1)
EVTPVTGASLR	P55265	Double-stranded RNA-specific adenosine deaminase	2	3	(1)
EVTQDFMMYR	P09668	Cathepsin H light chain	2	5	(1)
EVTRILDGK	Q9H299	SH3 domain-binding glutamic acid-rich-like protein 3	--	--	(3)
EVTSGGSLQPK	Q9UJA5	tRNA (adenine-N(1)-)-methyltransferase non-catalytic subunit TRM6	2	6	(1)
EVVAGEFGSR	P19447	TFIIH basal transcription factor complex helicase XPB subunit	2	3	(1)
EVVDSNPYSR	Q9GZZ9	Ubiquitin-like modifier-activating enzyme 5	2	3	(1)
EVVDVMVIGR	Q9NQX3	Gephyrin	2	7	(1)
EVVEFLKNPQK	Q96TA2	ATP-dependent metalloprotease YME1L1	3	8	(1)
EVVEIQIDR	Q9Y230	RuvB-like 2	2	6	(1)
EVVEPEGPVAQR	Q96EK5	KIF1-binding protein	2	7	(1)
EVVEQGITR	P51398	28S ribosomal protein S29, mitochondrial	2	4	(1)
EVVGNMAVSR	Q9Y573	Actin-binding protein IPP	2	6	(1)
EVVNYVQKR	Q9H9A7	RecQ-mediated genome instability protein 1	2	3	(1)
EVVPEPVAR	P27707	Deoxycytidine kinase	2	4	(1)
EVVQSQQVAVGR	Q92888	Rho guanine nucleotide exchange factor 1	2	4	(1)
EVVSGVVSK	Q14008	Cytoskeleton-associated protein 5	1	4	(1)
EVVSLLLDR	Q6S8J7	ANKRD26-like family A member 1	2	6	(1)
EVVSQEAELR	Q7Z7B0	Filamin-A-interacting protein 1	2	7	(1)
EVVTGVIGQR	Q02153	Guanylate cyclase soluble subunit beta-1	2	6	(1)
EVVTPGSGPGTGR	Q86V97	Kelch repeat and BTB domain-containing protein 6	2	3	(1)
EVVTPNIFNSR	**P26639**	**Threonyl-tRNA synthetase, cytoplasmic**	**2**	**3**	**(1)**
EVVTSMGTQR	Q53G59	Kelch-like protein 12	2	7	(1)
EVVVSGKLR	P23396	40S ribosomal protein S3	2	5	(1), (2)
EVVVSTGVLGR	Q5T1V6	Probable ATP-dependent RNA helicase DDX59	2	7	(1)
EVYDIAFSRA	P61962	WD repeat-containing protein 68	2	4	(1)
EVYPVGIFQK	Q9UI95	Mitotic spindle assembly checkpoint protein MAD2B	2	5	(1)
EVYPVGIFQKR	Q9UI95	Mitotic spindle assembly checkpoint protein MAD2B	2	3	(1)
EVYQILEKGAAK	P52732	Kinesin-like protein KIF11	3	6	(1)
EYFPGTGDLR	Q92769	Histone deacetylase 2	2	3	(1)
EYVDATNLTPR	P07686	Beta-hexosaminidase subunit beta	2	3	(1)
FAAESIIKR	O95503	Chromobox protein homolog 6	2	5	(1)
FAGAIFQGR	Q9H8Y5	Ankyrin repeat and zinc finger domain-containing protein 1	2	6	(1)
FAITAIKGVGR	P62269	40S ribosomal protein S18	3	7	(1)
FASDPILYR	P14618	Pyruvate kinase isozymes M1/M2	2	3	(1)
FASLPQVER	O75083	WD repeat-containing protein 1	2	6	(1)
FDASDTLALPR	Q5JXC2	Invasion-inhibitory protein 45	2	6	(1)
FEITPPVVLR	P06748	Nucleophosmin	2	3	(1)
FEKAIQR	Q9UFH2	Dynein heavy chain 17, axonemal	2	4	(1)
FNETPINPR	Q9UBF2	Coatomer subunit gamma-2	2	3	(1)
FPANVPALR	O15055	Period circadian protein homolog 2	2	3	(1)
FPHNPQFIGR	Q9H9Y2	Ribosome production factor 1	--	--	(2)
FPSEIVGK	P62081	40S ribosomal protein S7	1	4	(1)
FPSEIVGKR	P62081	40S ribosomal protein S7	2	6	(1), (2)
FPSSVAALR	Q9BQ83	GIY-YIG domain-containing protein 1	2	5	(1)
FSAGLSDLR	Q9BRV3	RAG1-activating protein 1	2	7	(1)
FSAPENAVR	Q99622	Protein C10	2	4	(1)
FSDFTLAR	Q6J4K2	Sodium/potassium/calcium exchanger 6	--	--	(2)
FSGEGQSLR	Q92890	Ubiquitin fusion degradation protein 1 homolog	2	5	(1)

FSGLEEAVYR	P50990	T-complex protein 1 subunit theta	2	6	(1)
FSQPGTISR	Q8NEZ4	Histone-lysine N-methyltransferase MLL3	2	4	(1)
FSVPVIYER	Q9NQC3	Reticulon-4	2	4	(1)
FSVSDLAPPR	Q96DE5	UPF0448 protein C10orf104	2	5	(1)
FTADGDQVFAGR	Q96SB8	Structural maintenance of chromosomes protein 6	2	7	(1)
FTAFEEAQLPR	Q96CT7	Coiled-coil domain-containing protein 124	2	7	(1)
FTAPSTVGK	O95453	Poly(A)-specific ribonuclease PARN	2	3	(1)
FTASPASSTTAR	O00151	PDZ and LIM domain protein 1	2	6	(1)
FTGSIFIGR	Q8IZH2	5 _quote_ -3 _quote_ exoribonuclease 1	2	7	(1)
FTITPGSEQIR	Q99798	Aconitate hydratase, mitochondrial	2	5	(1)
FTKAYQR	P52701	DNA mismatch repair protein Msh6	2	5	(1)
FTLDINTAYAR	Q16891	Mitochondrial inner membrane protein	2	6	(1)
FTSGSPEETAFR	O60313	Dynamin-like 120 kDa protein, mitochondrial	2	3	(1)
FTSQLIVER	O15498	Synaptobrevin homolog YKT6	2	7	(1)
FTYEGNSNDIR	Q9UJU6	Drebrin-like protein	2	3	(1)
FVANLPQELK	P04090	Prorelaxin H2 B chain	2	4	(1)
FVFDNTEAIAQR	Q9UP83	Conserved oligomeric Golgi complex subunit 5	2	6	(1)
FVFDRPTNDK	Q6ZN66	Guanylate-binding protein 6	2	3	(1)
FVFSEGSEASGR	Q9UFC0	Leucine-rich repeat and WD repeat-containing protein 1	2	7	(1)
FVIDAVRTK	Q7L2H7	Eukaryotic translation initiation factor 3 subunit M	2	4	(1)
FVIGGPQGDAGLTGR	P31153	S-adenosylmethionine synthetase isoform type-2	2	7	(1)
FVNGDGTGVASIYR	O43447	Peptidyl-prolyl cis-trans isomerase H	2	7	(1)
FVNVVPTFGK	P62861	40S ribosomal protein S30	2	3	(1)
FVQGIFVEK	A6NIZ1	Ras-related protein Rap-1b-like protein	2	3	(1)
FVSAPEVSR	Q96Q27	Ankyrin repeat and SOCS box protein 2	2	6	(1)
FVSTVPSKLR	P26010	Integrin beta-7	2	5	(1)
FVTPVTNQR	Q8NEZ4	Histone-lysine N-methyltransferase MLL3	2	3	(1)
FVVEKVLDR	Q13185	Chromobox protein homolog 3	2	7	(1)
FVVEPSEATNR	Q03188	Centromere protein C 1	2	3	(1)
FVYGEPVTR	P36956	Sterol regulatory element-binding protein 1	2	3	(1)
FVYGSGVYR	Q96BW9	MMP37-like protein, mitochondrial	2	5	(1)
FVYQDPQGR	Q16531	DNA damage-binding protein 1	2	4	(1)
GGFEITPPVVLR	P06748	Nucleophosmin	2	6	(1)
GIGAIGGTPPAFNR	Q15233	Non-POU domain-containing octamer-binding protein	2	3	(1)
GPISSGSSASSVTVTR	P02545	Lamin-A/C	2	6	(1)
GPSIVHR	P63261	Actin, cytoplasmic 2	2	6	(1)
GSHIIQR	P10319	HLA class I histocompatibility antigen, B-58 alpha chain	2	5	(1)
GTADVHFER	Q86V81	THO complex subunit 4	2	6	(1)
GTASLFTNR	Q7RTX9	Monocarboxylate transporter 14	2	7	(1)
GTSSVIVSR	Q9UKB1	F-box/WD repeat-containing protein 11	2	5	(1)
GVAALYKFR	P09669	Cytochrome c oxidase polypeptide VIc	2	8	(1)
GVFGFPLGR	Q460N5	Poly [ADP-ribose] polymerase 14	--	--	(2)
GVSPVITVR	Q7Z7B0	Filamin-A-interacting protein 1	2	5	(1)
GVTDVVITR	Q6PK18	PKHD domain-containing transmembrane protein FLJ22222	2	6	(1)
HAADPIITR	Q9BW91	ADP-ribose pyrophosphatase, mitochondrial	2	4	(1)
HAVGIVVNK	P46778	60S ribosomal protein L21	2	4	(1)
HPAVVIRQR	Q12870	Transcription factor 15	--	--	(2)
HVIETLIGK	P48643	T-complex protein 1 subunit epsilon	2	7	(1)
HVPEVEVQVKR	Q9Y675	SNRPN upstream reading frame protein	3	4	(1)
HVSLITPTK	Q9H8U3	AN1-type zinc finger protein 3	2	3	(1)

HVSLITPTKR	Q9H8U3	AN1-type zinc finger protein 3	3	4	(1)
HVVDIVALK	P08579	U2 small nuclear ribonucleoprotein B	2	3	(1)
HVVSEGVPR	Q9NUP1	Protein cappuccino homolog	2	3	(1)
IAAGEVIQR	P40692	DNA mismatch repair protein Mlh1	2	6	(1)
IASLIPTPR	Q9NRH3	Tubulin gamma-2 chain	2	5	(1)
IIQNVPVHQR	O94915	Protein furry homolog-like	3	3	(1)
IIVDIATKR	Q96IY1	Kinetochore-associated protein NSL1 homolog	2	3	(1)
ILIDPFHK	P61313	60S ribosomal protein L15	2	5	(1)
INQTQKENLR	P42766	60S ribosomal protein L35	2	7	(1)
IPAAIATAK	Q92575	UBX domain-containing protein 4	1	3	(1)
IPFGNNPIFR	Q15392	24-dehydrocholesterol reductase	2	7	(1)
IPFPTPTTPLTGR	Q14157	Ubiquitin-associated protein 2-like	2	6	(1)
IPSEIISPK	Q8NCN5	Pyruvate dehydrogenase phosphatase regulatory subunit, mitochondrial	2	6	(1)
IPVLVVQAQR	Q01629	Interferon-induced transmembrane protein 2	2	6	(1)
ISFSGIATR	Q5QJE6	Deoxynucleotidyltransferase terminal-interacting protein 2	2	6	(1)
ISSGAVVHR	Q9BXK1	Krueppel-like factor 16	2	5	(1)
ITAGGSLRSIQR	Q96BI3	Gamma-secretase subunit APH-1A	2	6	(1)
ITAIKGVGR	P62269	40S ribosomal protein S18	2	8	(1)
ITDPRTVFV	Q96NB2	Sideroflexin-2	2	3	(1)
ITEGAQAPR	P27918	Properdin	2	7	(1)
ITEGTASVLQR	P31321	cAMP-dependent protein kinase type I-beta regulatory subunit	2	7	(1)
ITSPPNALPR	P30260	Cell division cycle protein 27 homolog	2	3	(1)
ITTGNSPPR	P27694	Replication protein A 70 kDa DNA-binding subunit	2	4	(1)
IVFGEVDR	O75306	NADH dehydrogenase [ubiquinone] iron-sulfur protein 2, mitochondrial	2	3	(1)
IVNENLVER	Q7LG56	Ribonucleoside-diphosphate reductase subunit M2 B	2	7	(1)
IVVFGDEPPVFSR	Q99816	Tumor susceptibility gene 101 protein	2	5	(1)
IVVGTVFIIR	P05536	HLA class II histocompatibility antigen, DQ(W3) alpha chain	2	6	(1)
IVVPKAAIVAR	Q9H3M7	Thioredoxin-interacting protein	3	7	(1)
KPPTILR	P53677	AP-3 complex subunit mu-2	2	3	(1)
LAASQVVNR	Q8N1F7	Nuclear pore complex protein Nup93	2	7	(1)
LAQTGSLGR	Q9Y3L3	SH3 domain-binding protein 1	2	5	(1)
LATDVQTVAQR	P49720	Proteasome subunit beta type-3	2	6	(1)
LATMISTMR	O75533	Splicing factor 3B subunit 1	2	5	(1)
LELDPALAPVVSR	O00170	AH receptor-interacting protein	2	7	(1)
LIIGGVAAR	Q92499	ATP-dependent RNA helicase DDX1	2	5	(1)
LLAGISIER	O15552	Free fatty acid receptor 2	2	6	(1)
LPAPAAVGR	Q8ND56	Protein LSM14 homolog A	2	5	(1)
LPFGSSPHR	Q86WC4	Osteopetrosis-associated transmembrane protein 1	2	4	(1)
LPIVTPALR	P52292	Importin subunit alpha-2	2	6	(1)
LPSPVTAQK	P13639	Elongation factor 2	2	6	(1)
LSFEGATKR	Q9NQT5	Exosome complex exonuclease RRP40	2	6	(1)
LSFETTDESLR	P09651	Heterogeneous nuclear ribonucleoprotein A1	2	7	(1)
LTAGVSAGR	Q15782	Chitinase-3-like protein 2	2	5	(1)
LTGPVMPVR	P26373	60S ribosomal protein L13	2	5	(1)
LTIQSGEQPYK	O43920	NADH dehydrogenase [ubiquinone] iron-sulfur protein 5	2	7	(1)
LVAEITPKGR	Q9NPA8	Enhancer of yellow 2 transcription factor homolog	3	4	(1)
LVAEQVTYQR	P49736	DNA replication licensing factor MCM2	2	7	(1)
LVASQTVQR	Q9NUT2	ATP-binding cassette sub-family B member 8, mitochondrial	2	6	(1)
LVFCGEPPAIK	Q4LDE5	Sushi, von Willebrand factor type A, EGF and pentraxin domain-containing protein	2	3	(1)
LVFPSEIVGKR	P62081	40S ribosomal protein S7	3	3	(1), (2)

LVGPVPAGR	Q9NVP2	Histone chaperone ASF1B	2	5	(1)
LVIPFTIKK	P21333	Filamin-A	3	3	(1)
LVMSQANVSR	Q13765	Nascent polypeptide-associated complex subunit alpha	2	5	(1)
MAAAAFAEK	Q9Y3B1	Protein slowmo homolog 2	2	6	(1)
MAAESLLVTVR	Q9UKV5	Autocrine motility factor receptor, isoform 2	2	4	(1)
MAGDIYSVFR	Q99541	Adipophilin	2	6	(1), (3)
MIVDLDNVVK	Q16891	Mitochondrial inner membrane protein	2	4	(1)
MPFTASPASSTTAR	O00151	PDZ and LIM domain protein 1	2	7	(1)
MPIVLELIPLK	P00403	Cytochrome c oxidase subunit 2	2	5	(1)
MPVGPDAILR	P46379	Large proline-rich protein BAT3	2	6	(1)
MSFLKVDK	O95470	Sphingosine-1-phosphate lyase 1	2	3	(1)
MTFGIEPFKTK	Q9BPX3	Condensin complex subunit 3	2	6	(1)
MTSALPIIQK	Q99541	Adipophilin	2	8	(1), (3)
MVANLLYEK	P49720	Proteasome subunit beta type-3	2	4	(1)
MVIGIPVYVGR	Q6ZMZ0	E3 ubiquitin-protein ligase RNF19B	2	3	(1)
MVLQVSAAPR	P04440	HLA class II histocompatibility antigen, DP(W4) beta chain	2	7	(1)
MVPSVLWPR	P32942	Intercellular adhesion molecule 3	2	5	(1)
NAAGLLAAR	Q9NPQ8	Synembryn-A	2	6	(1)
NAIGGKYNR	P14324	Farnesyl pyrophosphate synthase	--	--	(2)
NAMGFSQGGQFLR	P50897	Palmitoyl-protein thioesterase 1	2	6	(1)
NANPLIYQR	Q9Y3D0	UPF0195 protein FAM96B	2	5	(1)
NASQLITQR	Q99623	Prohibitin-2	2	6	(1)
NGAGVTVLR	Q9NVF7	F-box only protein 28	2	4	(1)
NIFNPPIIAR	P00451	Coagulation factor VIII	2	3	(1)
NLTDIVAQR	Q0P5N6	ADP-ribosylation factor-like protein 16	2	3	(1)
NPADVSVISSR	P08865	40S ribosomal protein SA	2	4	(1)
NSFDPSKITR	Q13352	Centromere protein R	2	4	(1)
NSFRYNGLIHR	P46779	60S ribosomal protein L28	--	--	(2)
NTAISEVIGK	O43264	Centromere/kinetochore protein zw10 homolog	2	4	(1)
NTAVSQLTK	Q9NTJ3	Structural maintenance of chromosomes protein 4	2	3	(1)
NTDEMVEL	P61978	Heterogeneous nuclear ribonucleoprotein K	1	3	(1)
NTFEFAKPVQR	Q8NCL4	Polypeptide N-acetylgalactosaminyltransferase 6	2	4	(1)
NTFLGTPVQK	P46013	Antigen KI-67	2	5	(1)
NTFSSGSDLNAVK	Q9NTX5	Enoyl-CoA hydratase domain-containing protein 1	2	6	(1)
NTIDQKIVER	Q9NRZ9	Lymphoid-specific helicase	2	8	(1)
NTLTNIAMR	P33992	DNA replication licensing factor MCM5	2	4	(1)
NTNDFVTLR	Q92673	Sortilin-related receptor	2	7	(1)
NTSPYSVQQR	Q9H2X6	Homeodomain-interacting protein kinase 2	2	6	(1)
NTTSAVTVK	P61513	60S ribosomal protein L37a	2	4	(1)
NTVDEVYVSR	P52630	Signal transducer and activator of transcription 2	2	7	(1)
NTVGLVVER	P46952	3-hydroxyanthranilate 3,4-dioxygenase	2	7	(1)
NTVHDIVNR	Q9UKF6	Cleavage and polyadenylation specificity factor subunit 3	2	7	(1)
NVAALPGIVHR	Q9Y3I0	UPF0027 protein C22orf28	2	7	(1)
NVAANVTSR	Q9H6E5	U6 snRNA-specific terminal uridylyltransferase 1	2	3	(1)
NVAATFKGIVR	P53396	ATP-citrate synthase	2	7	(1)
NVADLTSLR	Q9UL46	Proteasome activator complex subunit 2	2	8	(1)
NVADLVVILK	Q12905	Interleukin enhancer-binding factor 2	2	7	(1)
NVAEVDKVTGR	P63220	40S ribosomal protein S21	2	8	(1), (2)
NVAGSTLSR	Q5JWF8	Putative uncharacterized actin family protein C20orf134	2	3	(1)
NVEEIAFYR	Q9UBJ2	ATP-binding cassette sub-family D member 2	2	7	(1)

NVIALVGGATAR	Q9Y2Z4	Tyrosyl-tRNA synthetase, mitochondrial	2	7	(1)
NVIEPSFGLGR	P41250	Glycyl-tRNA synthetase	2	3	(1)
NVIEQYIGR	Q6ZRQ5	Uncharacterized protein C6orf167	2	3	(1)
NVIESKAYVGR	O00399	Dynactin subunit 6	2	7	(1)
NVIGLQMGTNR	P37802	Transgelin-2	2	7	(1)
NVMTDLKNAQER	Q96B42	Transmembrane protein 18	2	3	(1)
NVNEVISNR	P07954	Fumarate hydratase, mitochondrial	2	7	(1)
NVSHTVVLR	P43308	Translocon-associated protein subunit beta	--	--	(3)
NVTDDTYYYR	Q9H819	DnaJ homolog subfamily C member 18	2	6	(1)
NVTYVVKER	O75369	Filamin-B	2	5	(1)
NVVEYIPNAER	Q8TCN5	Zinc finger protein 507	2	3	(1)
PAVAKLR	Q86WA8	Peroxisomal Lon protease homolog 2	2	4	(1)
PAVVSER	P62851	40S ribosomal protein S25	1	3	(1)
PAVVVGR	O75352	Mannose-P-dolichol utilization defect 1 protein	1	4	(1)
PSGLLPQR	Q6ZSA8	Putative uncharacterized protein FLJ45684	2	3	(1)
PSIVGRPR	P63261	Actin, cytoplasmic 2	2	5	(1)
PVGIFQKR	Q9UI95	Mitotic spindle assembly checkpoint protein MAD2B	2	5	(1)
PVLRLPR	Q9H840	Gem-associated protein 7	2	4	(1)
QAVDVSPLR	P46782	40S ribosomal protein S5	2	4	(1)
QIAEGMAYIER	P51451	Tyrosine-protein kinase BLK	2	3	(1)
QTIQGILER	P12270	Nucleoprotein TPR	2	5	(1)
QTIQPPQPTGGR	Q02930	cAMP-responsive element-binding protein 5	2	5	(1)
QTMSDQIIGR	O75506	Heat shock factor-binding protein 1	2	3	(1)
QVAPPVLKR	Q8N766	Uncharacterized protein KIAA0090	2	6	(1)
QVIEGNAAVVLR	O95544	NAD kinase	2	4	(1)
QVIGIGAGQQSR	P31939	Bifunctional purine biosynthesis protein PURH	2	7	(1)
RALAPLR	Q96F44	Tripartite motif-containing protein 11	2	4	(1)
RPVASQLPR	P35232	Prohibitin	2	5	(1)
SAADVVVVHGR	P09110	3-ketoacyl-CoA thiolase, peroxisomal	2	6	(1)
SAAEVGFVR	P43246	DNA mismatch repair protein Msh2	2	5	(1)
SAALPQSQR	Q9P2Q2	FERM domain-containing protein 4A	2	6	(1)
SAFATPFLVVR	P15954	Cytochrome c oxidase subunit 7C, mitochondrial	2	5	(1)
SAIGYLTSR	Q9UJ83	2-hydroxyacyl-CoA lyase 1	2	6	(1)
SAVDITKVAR	Q9NQX3	Gephyrin	3	7	(1)
SAVDPSSGHPR	Q9P2B7	UPF0501 protein KIAA1430	2	3	(1)
SDAGIFFTR	Q9NR46	Endophilin-B2	2	6	(1)
SGASVVAIR	Q5SSJ5	Heterochromatin protein 1-binding protein 3	2	6	(1)
SIELVAVER	Q92545	Transmembrane protein 131	2	3	(1)
SIFDGRVVAK	Q9UM00	Transmembrane and coiled-coil domain-containing protein 1	--	--	(3)
SIFTGAQDPGLQR	A8K0Z3	WAS protein family homolog 1	2	4	(1)
SIIGVTEPR	Q8IY37	Probable ATP-dependent RNA helicase DHX37	2	6	(1)
SLADIMAKR	P83731	60S ribosomal protein L24	2	5	(1), (3)
SLSSVTYVPR	Q9Y271	Cysteinyl leukotriene receptor 1	2	6	(1)
SNDIPHVVR	Q8NI36	WD repeat-containing protein 36	2	3	(1)
SPASIGQPR	Q8NFH5	Nucleoporin NUP53	2	3	(1)
SPIGASGVAHR	O95155	Ubiquitin conjugation factor E4 B	2	6	(1)
SPLGGIKR	Q9H0D6	5 _quote_ -3 _quote_ exoribonuclease 2	2	4	(1)
SSAGGSFTVR	Q14568	Putative heat shock protein HSP 90-alpha A2	2	8	(1)
SSAPVVVTR	P85037	Forkhead box protein K1	2	4	(1)
SSAWSTVFR	Q9H2D1	Mitochondrial folate transporter/carrier	2	3	(1)

SSSTVVSTQR	P31483	Nucleolysin TIA-1 isoform p40	2	3	(1)
STADNVLTLSLR	Q14690	Protein RRP5 homolog	2	3	(1)
STARHLYLR	P39019	40S ribosomal protein S19	3	4	(1)
STASVASASTTR	Q99607	ETS-related transcription factor Elf-4	2	7	(1)
STASVNLGDPGSTR	Q96RT1	Protein LAP2	2	4	(1)
STDHIPILYR	Q06210	Glucosamine--fructose-6-phosphate aminotransferase	3	4	(1)
STFGELTVEPR	Q8N264	Rho GTPase-activating protein 24	2	3	(1)
STFNQVVLKR	Q07020	60S ribosomal protein L18	2	4	(1)
STFPAGVPVSER	Q13490	Baculoviral IAP repeat-containing protein 2	2	3	(1)
STIEYVIQR	Q15437	Protein transport protein Sec23B	--	--	(3)
STITSREIQTAVR	Q99877	Histone H2B type 1-N	3	5	(1)
STSEAPQPPR	Q15014	Mortality factor 4-like protein 2	2	7	(1)
STSETPQPPR	Q9UBU8	Mortality factor 4-like protein 1	2	8	(1)
STVEEAVATR	Q9UKV3	Apoptotic chromatin condensation inducer in the nucleus	2	7	(1)
STVENFGQLGR	O60287	Nucleolar pre-ribosomal-associated protein 1	2	8	(1)
STVGAGAYAYK	O95831	Apoptosis-inducing factor 1, mitochondrial	2	4	(1)
STVGIIGLGR	Q9UBQ7	Glyoxylate reductase/hydroxypyruvate reductase	2	7	(1)
STVGSAISR	O43399	Tumor protein D54	2	5	(1)
STYYGSFVTR	O15372	Eukaryotic translation initiation factor 3 subunit H	2	6	(1)
SVAEIIRNH	Q8WV28	B-cell linker protein	2	6	(1)
SVAEISSNSLER	Q5T4S7	E3 ubiquitin-protein ligase UBR4	2	7	(1)
SVAEMPQNYR	Q14258	Tripartite motif-containing protein 25	2	7	(1)
SVAGLLPRR	Q96K49	Transmembrane protein 87B	2	6	(1)
SVAGLTAAAYR	Q86Y39	NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 11	2	7	(1)
SVAGMPATSR	Q13761	Runt-related transcription factor 3	2	5	(1)
SVALPAIMR	Q9Y490	Talin-1	2	6	(1)
SVESIISPK	O60216	Double-strand-break repair protein rad21 homolog	2	7	(1)
SVFEDPVISK	Q9Y2R9	28S ribosomal protein S7, mitochondrial	2	4	(1)
SVFQEVTLR	Q9BTE0	N-acetyltransferase 9	2	7	(1)
SVHLQPITR	O75643	U5 small nuclear ribonucleoprotein 200 kDa helicase	2	4	(1)
SVINPGAIYR	P13612	Integrin alpha-4	2	7	(1)
SVINSVSTSR	Q01664	Transcription factor AP-4	2	7	(1)
SVITQAPSTNR	Q02930	cAMP-responsive element-binding protein 5	2	4	(1)
SVNELIYKR	P18124	60S ribosomal protein L7	2	6	(1)
SVPEFPLSPPK	P16949	Stathmin	2	8	(1)
SVQGAGTSYR	O15519	CASP8 and FADD-like apoptosis regulator	2	6	(1)
SVQGIIIYR	P05141	ADP/ATP translocase 2	2	8	(1), (2)
SVQGYPTQR	P10124	Serglycin	2	4	(1)
SVSDLAPPR	Q96DE5	UPF0448 protein C10orf104	2	7	(1)
SVSEGLNVPR	Q8TAP8	UPF0683 protein C7orf47	2	6	(1)
SVSGIPNAR	Q96GE4	Coiled-coil domain-containing protein 45	2	3	(1)
SVSSLISGR	Q9BRK4	Leucine zipper putative tumor suppressor 2	2	4	(1)
SVTALAVSR	Q6Q0C0	E3 ubiquitin-protein ligase TRAF7	2	7	(1)
SVTPEGVIKQR	Q9Y4G8	Rap guanine nucleotide exchange factor 2	2	6	(1)
SVVEIASLR	Q96EZ8	Microspherule protein 1	2	6	(1)
SVVGIPVFR	Q9BTF0	THUMP domain-containing protein 2	2	5	(1)
SVYSPSGPVNR	Q9NS56	E3 ubiquitin-protein ligase Topors	2	3	(1)
TAAAAASIR	Q8WVM8	Sec1 family domain-containing protein 1	2	4	(1)
TAAAIGATPR	Q14980	Nuclear mitotic apparatus protein 1	2	6	(1)
TAADLGSVVR	O95235	Kinesin-like protein KIF20A	2	6	(1)

TAADTAAQITQR	P30501	HLA class I histocompatibility antigen, Cw-2 alpha chain	3	7	(1), (2), (3)
TAAGPLFQQR	O60828	Polyglutamine-binding protein 1	2	7	(1)
TAAGQGILSR	O94833	Bullous pemphigoid antigen 1, isoforms 6/9/10	2	6	(1)
TADVHFER	Q86V81	THO complex subunit 4	2	7	(1), (2)
TAFEVTSGGSLQPK	Q9UJA5	tRNA (adenine-N(1)-)-methyltransferase non-catalytic subunit TRM6	2	7	(1)
TAGIIGVNR	Q00610	Clathrin heavy chain 1	2	4	(1)
TAGPPDPPR	Q96HB5	Coiled-coil domain-containing protein 120	2	3	(1)
TAIDILTTVVR	Q96P70	Importin-9	2	5	(1)
TAMDVVYALK	P62805	Histone H4	2	7	(1)
TAVQYAIGR	Q05048	Cleavage stimulation factor 50 kDa subunit	2	7	(1)
TDLEPPVVKR	O14974	Protein phosphatase 1 regulatory subunit 12A	3	3	(1)
TFDVYKQNPR	Q9UHD4	Cell death activator CIDE-B	--	--	(2)
TGIFLPILR	P33993	DNA replication licensing factor MCM7	2	6	(1)
TIIDIITHR	P08133	Annexin A6	--	--	(3)
TIIDILTKR	P04083	Annexin A1	--	--	(3)
TIIDLPGITR	P20592	Interferon-induced GTP-binding protein Mx2	2	6	(1)
TIQSGEQPYK	O43920	NADH dehydrogenase [ubiquinone] iron-sulfur protein 5	2	3	(1)
TITTKELGTVMR	P62158	Calmodulin	2	3	(1)
TIVNILTNR	P07355	Annexin A2	2	5	(1), (3)
TLAGIMGMR	Q9P0S9	Transmembrane protein 14C	2	5	(1)
TLGDIVFKR	P07148	Fatty acid-binding protein, liver	--	--	(3)
TLIDPETLLPR	Q86VP6	Cullin-associated NEDD8-dissociated protein 1	2	3	(1)
TLIESGIIGR	P28288	ATP-binding cassette sub-family D member 3	2	5	(1)
TNADIIETLR	Q14814	Myocyte-specific enhancer factor 2D	2	6	(1)
TNNIGSIAR	Q8TE04	Pantothenate kinase 1	2	4	(1)
TPAGGGFPR	Q07352	Butyrate response factor 1	2	7	(1)
TPFNYNPSPR	P25054	Adenomatous polyposis coli protein	2	6	(1)
TQRELVSR	P35232	Prohibitin	2	4	(1)
TSAALPQSQR	Q9P2Q2	FERM domain-containing protein 4A	2	8	(1)
TSEAPQPPR	Q15014	Mortality factor 4-like protein 2	2	7	(1), (2)
TSETPQPPR	Q9UBU8	Mortality factor 4-like protein 1	2	7	(1), (2)
TSFFPTAGGFR	Q9HAW4	Claspin	2	3	(1)
TSFPSGQGPGGGVSR	O43189	PHD finger protein 1	2	4	(1)
TSIQEVVSR	Q8IWC1	MAP7 domain-containing protein 3	2	7	(1)
TSISSLFER	P23258	Tubulin gamma-1 chain	2	6	(1)
TSSGNPIFR	Q9HAV4	Exportin-5	2	6	(1)
TSTPVIVMR	Q96RG2	PAS domain-containing serine/threonine-protein kinase	2	7	(1)
TSYPFDTVRR	P05141	ADP/ATP translocase 2	2	4	(1)
TTAAGPLFQQR	O60828	Polyglutamine-binding protein 1	2	8	(1)
TTAEREIVR	P63261	Actin, cytoplasmic 2	2	3	(1), (3)
TTAGAVTQCYR	Q02543	60S ribosomal protein L18a	2	7	(1)
TTALGSLANIAR	Q92797	Symplekin	2	7	(1)
TTAQEVLAR	Q9H211	DNA replication factor Cdt1	2	6	(1)
TTASVVIIR	O15297	Protein phosphatase 1D	2	7	(1)
TTATMATSGSAR	P38919	Eukaryotic initiation factor 4A-III	2	3	(1)
TTAVITNKDVR	P60228	Eukaryotic translation initiation factor 3 subunit E	2	5	(1)
TTFDLGGHEQAR	Q9NR31	GTP-binding protein SAR1a	2	7	(1)
TTFDLGGHVQAR	Q9Y6B6	GTP-binding protein SAR1b	2	8	(1)
TTGKVVGVFR	Q13685	Angio-associated migratory cell protein	2	3	(1)
TTGWGVSPR	P60510	Serine/threonine-protein phosphatase 4 catalytic subunit	2	5	(1)

TTIFSPEGR	P25789	Proteasome subunit alpha type-4	2	3	(1)
TTISKYFSER	Q9UL46	Proteasome activator complex subunit 2	2	7	(1)
TTITAPPSR	O00767	Acyl-CoA desaturase	2	6	(1)
TTQGLDGLSER	P09972	Fructose-bisphosphate aldolase C	2	7	(1)
TTSTPGASPAPR	Q504T8	Midnolin	2	3	(1)
TTVDESQVVTR	O94966	Ubiquitin carboxyl-terminal hydrolase 19	2	6	(1)
TTVEGLITR	O75312	Zinc finger protein ZPR1	2	8	(1)
TTVGTLSQR	P51665	26S proteasome non-ATPase regulatory subunit 7	2	8	(1)
TTVLAAVLR	Q9Y3T9	Nucleolar complex protein 2 homolog	2	5	(1)
TTVNFINQNLR	P55957	BH3-interacting domain death agonist	2	6	(1)
TTVTPFTAQQR	Q9BQG0	Myb-binding protein 1A	2	4	(1)
TTVTSMTTPR	Q53G59	Kelch-like protein 12	2	4	(1)
TVAEKLDWAR	P39023	60S ribosomal protein L3	2	4	(1)
TVATVMALR	P49588	Alanyl-tRNA synthetase, cytoplasmic	2	5	(1)
TVENFGQLGR	O60287	Nucleolar pre-ribosomal-associated protein 1	2	8	(1)
TVFDAKRLIGR	P48741	Putative heat shock 70 kDa protein 7	--	--	(2), (3)
TVFGATGFLGR	Q16795	NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 9	2	7	(1)
TVFNAHEPFRR	Q99942	E3 ubiquitin-protein ligase RNF5	--	--	(2)
TVFSPTLPAAR	Q7Z2W4	Zinc finger CCCH-type antiviral protein 1	2	7	(1)
TVFVGTAGTVGR	Q86VY4	Testis-specific Y-encoded-like protein 5	2	8	(1)
TVGGEQISAIGR	Q8TEA8	D-tyrosyl-tRNA(Tyr) deacylase 1	2	8	(1)
TVIFVASQK	Q9NVH2	Integrator complex subunit 7	2	3	(1)
TVILSPTAR	Q8NCJ5	SPRY domain-containing protein 3	2	3	(1)
TVINQTQKENLR	P42766	60S ribosomal protein L35	2	8	(1), (2)
TVITVDTKAAGK	P21333	Filamin-A	3	8	(1)
TVKDVNQQEFVR	P39019	40S ribosomal protein S19	2	5	(1), (2)
TVKKGKPELR	Q9BTD8	RNA-binding protein 42	--	--	(2)
TVSSLIIVR	Q92545	Transmembrane protein 131	2	7	(1)
TVSSVNEAFGR	Q13610	Periodic tryptophan protein 1 homolog	2	5	(1)
TVTLIPAFR	Q9H3H5	UDP-N-acetylglucosamine--dolichyl-phosphate N-acetylglucosaminephosphotransferase	2	3	(1)
TVVEALFQR	P57740	Nuclear pore complex protein Nup107	2	6	(1)
TVVGENGSVLR	Q9NQS7	Inner centromere protein	2	7	(1)
TVVQNLMER	Q01581	Hydroxymethylglutaryl-CoA synthase, cytoplasmic	2	4	(1)
VAVDVAVPK	Q6PII5	Hydroxyacylglutathione hydrolase-like protein	2	6	(1)
VAVGVARAR	P01833	Polymeric immunoglobulin receptor	--	--	(3)
VDIATKR	Q96IY1	Kinetochore-associated protein NSL1 homolog	2	4	(1)
VEGLITR	O75312	Zinc finger protein ZPR1	1	3	(1)
VFLENVIR	P62805	Histone H4	2	6	(1)
VLELDPALAPVVSR	O00170	AH receptor-interacting protein	2	5	(1)
VSGVVSK	Q14008	Cytoskeleton-associated protein 5	1	3	(1)
VSKFLNR	P62140	Serine/threonine-protein phosphatase PP1-beta catalytic subunit	2	3	(1)
VTDPRVVDL	P56556	NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 6	2	3	(1)
VTQEIVTER	Q14126	Desmoglein-2	2	7	(1)
VTVGGEQISAIGR	Q8TEA8	D-tyrosyl-tRNA(Tyr) deacylase 1	2	8	(1)
VVFEKAIQR	Q9UFH2	Dynein heavy chain 17, axonemal	2	7	(1)
VVFGTSGSEVQR	Q8NBF2	NHL repeat-containing protein 2	2	5	(1)
VVFGYEAGTKPR	Q8IUX1	Transmembrane protein 126B	2	8	(1)
VVYPKVIER	P30566	Adenylosuccinate lyase	2	3	(1)
WAVEGTAVALLR	Q14165	Malectin	2	6	(1)
WVIGSVVAR	P22102	Trifunctional purine biosynthetic protein adenosine-3	2	4	(1)

YAFNSGNVER	Q9UNM6	26S proteasome non-ATPase regulatory subunit 13	2	3	(1)
YAGGPSGVAR	Q7L591	Docking protein 3	2	4	(1)
YANEVGEAFR	Q9UDX5	Mitochondrial 18 kDa protein	2	6	(1)
YEPLPPTIGR	Q9Y6C9	Mitochondrial carrier homolog 2	2	3	(1)
YGGKSTIR	Q8WWY3	U4/U6 small nuclear ribonucleoprotein Prp31	2	4	(1)
YPAPIPPPR	A1KXE4	UPF0541 protein FAM168B	2	4	(1)
YPKVIER	P30566	Adenylosuccinate lyase	2	6	(1)
YPVGIFQKR	Q9UI95	Mitotic spindle assembly checkpoint protein MAD2B	--	--	(2)
YPYSPPAFR	P49427	Ubiquitin-conjugating enzyme E2 R1	2	6	(1)
YSFDPVGSYQR	P20618	Proteasome subunit beta type-1	2	5	(1)
YSFTTTAER	P63261	Actin, cytoplasmic 2	2	7	(1)
YSPGIVSTR	Q12789	General transcription factor 3C polypeptide 1	2	6	(1)
YSWDIVVQR	O15371	Eukaryotic translation initiation factor 3 subunit D	2	5	(1)
YTFYSISQER	Q9Y2R0	Coiled-coil domain-containing protein 56	2	6	(1)
YTYEHDPITK	P42224	Signal transducer and activator of transcription 1-alpha/beta	2	5	(1)
YVANLTELVR	P51532	Probable global transcription activator SNF2L4	2	7	(1)
YVFPKPFNR	O95453	Poly(A)-specific ribonuclease PARN	--	--	(2)
YVGEIIGKR	O75964	ATP synthase subunit g	2	6	(1)
YVHMFLMAR	P32245	Melanocortin receptor 4	2	3	(1)
YVYEVGPGPGGITR	Q8WVM0	Dimethyladenosine transferase 1	2	3	(1)

Source: (1) Electrospray-Orbitrap MS/MS; (2) MALDI-TOF/TOF, (3) http://www.syfpeithi.de

Table 2S. Percent residue frequencies at non-anchor positions among A*68:01 ligands as a function of peptide length[a].

9-mers					**10-mers**						**11-mers**							**12-mers**						
N: 327					**N: 205**						**N: 148**							**N: 64**						
	P4	**P5**	**P6**	**P8**		**P4**	**P5**	**P6**	**P7**	**P9**		**P4**	**P5**	**P6**	**P7**	**P8**	**P10**		**P4**	**P5**	**P6**	**P7**	**P8**	**P9**	**P11**
A	9.8	5.2	**10.1**	5.2	**A**	5.4	3.4	3.9	**11.7**	5.9	**A**	8.1	**14.2**	7.4	8.8	**10.8**	6.8	**A**	3.1	6.3	**10.9**	**18.8**	**10.9**	**20.3**	**15.6**
C	0.0	0.0	0.0	0.0	**C**	0.0	0.0	0.0	0.0	0.0	**C**	0.7	0.7	0.0	0.0	0.0	0.0	**C**	0.0	0.0	0.0	0.0	0.0	0.0	0.0
D	9.2	1.5	1.5	2.1	**D**	**13.2**	5.4	3.4	3.4	1.5	**D**	**10.8**	2.7	6.8	2.0	0.0	2.0	**D**	**18.8**	1.6	7.8	0.0	3.1	0.0	0.0
E	9.2	4.0	3.1	9.8	**E**	**10.7**	2.9	7.3	2.9	5.4	**E**	**14.9**	8.1	7.4	3.4	2.7	3.4	**E**	7.8	7.8	**10.9**	6.3	4.7	6.3	4.7
F	2.8	4.0	3.1	2.8	**F**	2.0	2.9	2.4	7.8	2.4	**F**	1.4	4.1	4.1	1.4	4.7	1.4	**F**	0.0	1.6	1.6	0.0	1.6	7.8	4.7
G	**17.7**	7.7	6.1	**10.7**	**G**	**12.7**	9.8	7.8	6.8	**11.2**	**G**	**16.2**	5.4	9.5	**16.2**	**10.8**	**15.5**	**G**	**17.2**	**15.6**	12.5	**15.6**	**10.9**	6.3	**21.9**
H	1.2	1.2	0.9	3.1	**H**	2.0	1.0	0.0	1.0	0.5	**H**	1.4	0.0	0.7	0.0	0.0	2.0	**H**	0.0	0.0	0.0	0.0	3.1	0.0	0.0
I	2.8	**16.5**	**16.5**	0.9	**I**	2.9	**11.2**	9.8	**10.2**	1.5	**I**	0.0	7.4	4.1	4.1	**12.2**	2.0	**I**	0.0	6.3	3.1	9.4	3.1	6.3	1.6
K	0.3	2.5	0.6	6.4	**K**	3.4	1.0	3.9	2.0	5.4	**K**	0.7	2.7	2.0	4.7	1.4	3.4	**K**	0.0	1.6	1.6	1.6	4.7	0.0	0.0
L	4.6	**10.1**	9.2	9.2	**L**	3.9	**11.2**	4.9	3.4	**10.2**	**L**	1.4	8.8	7.4	4.7	8.8	**11.5**	**L**	3.1	4.7	4.7	9.4	4.7	3.1	**10.9**
M	0.9	0.0	2.5	1.8	**M**	0.0	2.4	2.4	1.0	1.0	**M**	1.4	1.4	2.7	0.7	0.0	2.0	**M**	0.0	3.1	0.0	0.0	0.0	0.0	3.1
N	3.7	4.0	0.6	6.7	**N**	9.8	3.9	4.9	6.3	7.3	**N**	2.0	2.7	6.1	4.7	2.7	4.7	**N**	6.3	4.7	4.7	1.6	4.7	6.3	0.0
P	**11.0**	8.9	**11.3**	7.3	**P**	8.3	**13.7**	**13.2**	8.8	7.8	**P**	8.8	**10.1**	8.1	9.5	8.1	6.8	**P**	**10.9**	12.5	4.7	4.7	9.4	9.4	6.3
Q	6.1	3.4	4.0	**10.4**	**Q**	5.4	6.3	5.9	6.3	**11.2**	**Q**	6.8	3.4	2.0	8.1	7.4	**11.5**	**Q**	4.7	3.1	6.3	9.4	3.1	1.6	**12.5**
R	1.2	1.5	0.0	1.2	**R**	0.5	0.5	1.0	0.0	3.4	**R**	1.4	0.0	0.7	0.7	0.0	3.4	**R**	0.0	0.0	0.0	0.0	1.6	0.0	4.7
S	9.8	6.4	4.3	7.3	**S**	**10.7**	5.9	6.3	5.9	5.9	**S**	**12.8**	6.8	**12.2**	6.8	5.4	4.7	**S**	**12.5**	**14.1**	**15.6**	9.4	**14.1**	7.8	7.8
T	3.4	3.4	6.7	8.0	**T**	4.4	4.4	**11.2**	5.9	9.3	**T**	6.8	8.1	**12.8**	**14.9**	5.4	6.8	**T**	**12.5**	7.8	**12.5**	9.4	3.1	**12.5**	4.7
V	3.7	**15.0**	**18.0**	4.6	**V**	2.9	**10.7**	8.8	**12.2**	4.4	**V**	4.7	**10.1**	4.7	8.8	**14.2**	6.8	**V**	3.1	7.8	3.1	3.1	**17.2**	**12.5**	1.6
W	0.6	0.6	0.0	0.0	**W**	0.0	0.0	0.0	0.0	0.0	**W**	0.0	0.7	0.7	0.0	0.0	0.0	**W**	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Y	2.1	4.3	1.5	2.5	**Y**	2.0	3.4	2.9	4.4	5.9	**Y**	0.0	2.7	0.7	0.7	5.4	5.4	**Y**	0.0	1.6	0.0	1.6	0.0	0.0	0.0

[a]Residue frequencies >10% are in boldface.

Table 3S. Isotopic labeling of HLA-A*68:01 ligands in the presence of MG-132[a].

MG-132 Sensitive Ligands							
Fraction	**M+H$^+$**	**Residues**	**^{14}N**	**^{15}N**	**^{15}N+Inh**	**Ratio**	**Sequence**
172	926.50	K	41.88	112.96	45.70	0.05	
113	884.56	2 R	0.12	0.70	0.15	0.04	
134	941.46	R	39.49	65.60	38.42	-0.04	
195	949.56	R	35.43	88.39	49.39	0.26	
155	956.36	R	21.54	86.96	35.49	0.21	
92	969.42	R	30.69	71.21	31.68	0.02	
156	974.42	R	26.78	95.92	36.57	0.14	
109	982.59	K	35.02	73.73	27.78	-0.19	
108	982.38	R	41.38	107.64	32.20	-0.14	
125	986.43	R	25.37	59.78	32.00	0.19	
135	986.57	K	14.00	79.31	21.00	0.11	
136	986.53	R	30.83	70.81	36.76	0.15	
125	991.41	R	29.23	57.18	31.62	0.09	
111	992.60	K	26.36	55.84	20.93	-0.18	NAIGGKYNR
110	992.42	R	15.00	49.00	20.00	0.15	
134	1000.48	R	0.50	0.97	0.44	-0.13	
112	1003.41	R	20.00	81.00	27.00	0.11	DTAAQITQR
215	1003.57	R	29.23	57.18	34.82	0.20	
165	1005.51	R	22.55	227.95	35.07	0.06	
120	1010.51	R	13.06	28.73	12.52	-0.03	
220	1010.63	R	37.88	57.65	31.70	-0.31	
186	1011.56	K	30.63	50.26	24.70	-0.30	
108	1012.38	R	18.00	64.00	27.00	0.27	TSETPQPPR
150	1018.45	R	24.81	47.68	29.00	0.18	DTPGFIVNR
133	1021.42	R	24.09	60.82	19.00	-0.14	
155	1021.51	R	24.36	46.16	30.28	0.27	DVFVVGTER
164	1024.58	K	28.31	72.64	39.25	0.25	
113	1025.53	R	49.00	79.00	45.00	-0.13	
145	1032.56	K	22.44	55.03	28.62	0.19	FPSEIVGKR
144	1032.47	R	22.64	45.77	23.48	0.04	
155	1038.41	R	21.00	50.35	27.21	0.21	DAFFNPATR
132	1043.60	K	20.29	59.77	26.10	0.15	ELTAVVQKR
132	1043.60	R	20.05	70.51	29.32	0.18	
152	1060.58	K	61.10	72.26	60.07	-0.09	DVFRDPALK
152	1060.57	R	40.83	82.39	41.38	0.01	
188	1069.59	R	34.93	112.38	43.46	0.11	
145	1070.49	R	20.00	115.00	41.00	0.22	EVAQLIQGGR
173	1075.50	R	27.88	62.65	32.07	0.12	
126	1075.63	2 R	8.34	140.92	31.67	0.18	HPAVVIRQR
129	1077.42	R	13.29	25.57	12.55	-0.06	ETNYGIPQR
163	1080.47	K	31.29	67.00	25.00	-0.18	
88	1084.69	K	49.43	71.60	49.53	0.00	
128	1085.51	2 R	11.49	77.02	34.85	0.36	
145	1089.49	R	43.12	64.69	46.50	0.16	EPSEVFINR
150	1089.64	R	0.41	1.08	0.41	0.01	
156	1106.43	R	25.91	64.86	29.45	0.09	
193	1106.60	R	63.14	102.20	72.33	0.24	
101	1108.50	R	33.53	69.62	37.76	0.12	
163	1112.41	K	38.52	63.40	38.53	0.00	
152	1114.49	K	25.59	52.88	23.00	-0.09	

109	1119.49	R	31.85	56.71	34.00	0.09	
123	1122.49	2 R	3.89	18.62	9.05	0.35	
204	1123.63	R	28.10	239.85	57.57	0.14	
173	1126.56	R	0.24	1.49	0.55	0.25	
158	1128.64	K	24.65	64.50	27.74	0.08	EVFSHILKR
158	1128.57	R	24.65	51.52	27.88	0.12	
152	1136.47	R	33.28	63.70	33.34	0.00	
181	1137.54	R	15.64	77.60	29.01	0.22	EFVDLYVPR
130	1139.74	K	32.12	59.41	41.25	0.33	DPFHKAIRR
129	1139.54	2 R	4.33	38.13	10.81	0.19	
143	1139.55	2 R	6.49	39.76	9.68	0.10	ESFPGSFRGR
199	1139.70	R	43.76	88.19	55.80	0.27	
104	1140.50	3 R	2.98	9.06	6.41	0.25	EVNDRPVRR
189	1141.65	2 R	8.82	93.69	13.22	0.05	
125	1146.43	2 R	8.00	30.00	7.00	-0.05	
155	1147.45	R	68.27	277.60	44.90	-0.11	
169	1147.63	2 K	17.41	65.88	23.40	0.12	
114	1155.80	3 K	7.19	52.28	8.99	0.04	TVKKGKPELR
114	1155.67	R	28.30	130.47	49.62	0.21	
160	1163.48	R	26.50	102.68	52.00	0.33	
152	1168.41	K	27.07	59.25	24.00	-0.10	
144	1169.57	K	32.75	60.94	32.86	0.00	
132	1180.54	R	40.49	81.41	40.92	0.01	
153	1183.46	2 R	21.26	43.13	12.15	-0.42	
147	1199.58	K	39.00	84.61	37.00	-0.04	
146	1205.54	R	46.78	105.76	55.29	0.14	
154	1212.52	R	49.28	58.00	39.37	-1.14	FPHNPQFIGR
146	1216.66	K	4.18	92.40	8.05	0.04	
148	1216.58	2 R	8.45	76.73	10.03	0.02	
163	1218.67	K	33.86	61.25	35.27	0.05	
131	1221.53	R	33.18	99.28	35.76	0.04	DAAHPTNVQRL
150	1234.57	2 R	7.47	37.78	19.67	0.40	
147	1238.64	K	40.26	82.23	48.73	0.20	
147	1238.56	2 R	9.49	58.20	11.55	0.04	
178	1244.77	K	34.80	69.55	44.89	0.29	LVFPSEIVGKR
179	1244.64	R	37.16	54.75	38.14	0.06	
153	1245.72	K	45.42	91.83	51.38	0.13	
153	1245.71	R	25.37	45.12	26.85	0.08	
126	1246.62	R	39.94	209.92	39.65	0.00	TAADTAAQITQR
188	1267.67	R	51.91	117.28	44.44	-0.11	
147	1270.51	2 R	34.08	47.65	11.54	-1.66	
152	1280.68	K	41.19	70.64	38.00	-0.11	
159	1299.60	R	47.59	69.48	50.19	0.12	
151	1311.19	2 R	0.08	1.03	0.09	0.01	
152	1336.69	K	43.00	57.00	38.00	-0.36	
136	1351.61	R	0.38	0.92	0.52	0.25	
129	1353.87	2 K	31.17	73.38	40.87	0.23	
122	1358.74	K	42.75	71.86	39.73	-0.10	
122	1358.53	R	33.27	70.76	33.92	0.02	
152	1376.76	2 R	0.14	1.57	0.33	0.13	NSFRYNGLIHR
154	1444.73	4 R	0.00	0.32	0.09	0.28	
142	1608.91	3 R	0.01	0.82	0.24	0.28	
132	1719.86	2 R	14.15	78.47	30.26	0.25	

MG-132 Resistant Ligands

Fraction	M+H$^+$	Residues	^{14}N	^{15}N	^{15}N+Inh	Ratio	Sequence
145	933.49	K	0.25	0.64	0.52	0.69	
86	969.34	R	34.15	95.58	108.95	1.22	EPAPGTNQR
136	992.41	R	22.00	55.16	91.11	2.08	
167	1048.58	R	22.84	56.15	43.54	0.62	SVQGIIIYR
140	1104.48	2 R	23.83	80.38	87.18	1.12	
131	1114.50	2 R	20.48	71.43	59.82	0.77	
132	1134.46	R	32.84	231.08	187.63	0.78	
137	1179.61	K	36.85	69.16	73.25	1.13	ELYAEKVATR
137	1179.60	R	36.81	83.69	73.04	0.77	
169	1198.59	2 R	7.12	19.68	38.94	2.53	
127	1200.52	R	14.09	34.55	37.15	1.13	
201	1207.73	R	52.23	138.28	203.38	1.76	SAFATPFLVVR
129	1209.55	2 R	10.83	45.49	39.54	0.83	ETVQLRNPPR
148	1241.54	2 R	13.55	53.30	50.70	0.93	
162	1254.54	R	53.62	69.45	67.57	0.88	
144	1299.58	R	39.41	77.58	108.78	1.82	
140	1304.71	K	36.38	87.50	73.76	0.73	
140	1304.66	2 R	11.81	91.47	88.27	0.96	
188	1320.73	R	42.81	222.41	408.39	2.04	
146	1332.68	K	44.10	86.58	69.95	0.61	
146	1332.59	R	45.69	87.72	72.25	0.63	
181	1372.66	2 R	11.73	66.02	74.00	1.15	
136	1462.73	K	41.87	92.20	143.46	2.02	TVKDVNQQEFVR
136*	1462.75	R	35.96	95.90	211.32	2.93	
157	1493.65	R	53.23	92.75	80.07	0.68	
156	1500.73	R	40.63	78.80	70.59	0.78	

Not assigned

Fraction	M+H$^+$	Residues	^{14}N	^{15}N	^{15}N+Inh	Ratio	Sequence
135	948.42	R	16.57	52.04	35.47	0.53	
130	1023.50	2 R	5.78	89.65	61.11	0.66*	
129	1041.64	K	35.97	63.32	47.08	0.41	
153	1103.54	2 R	25.27	62.67	44.30	0.51	
193	1109.59	R	58.87	162.38	102.16	0.42	
173	1132.52	R	23.92	64.47	68.45	1.10*	
130	1187.74	K	26.12	79.69	51.64	0.48	NVAEVDKVTGR
130	1187.53	R	26.44	84.93	67.65	0.70*	
162	1254.64	K	54.17	73.18	64.62	0.55	
152	1201.53	R	36.28	63.90	46.35	0.36	
144	1250.57	2 R	6.91	34.16	20.87	0.51*	
127	1443.88	K	36.53	78.14	56.00	0.47	TVINQTQKENLR
128	1443.68	R	43.72	97.26	83.20	0.74*	
155	1550.65	R	55.19	126.34	86.67	0.44	
137	1640.83	3 R	7.16	66.44	29.13	0.37	

Table 3S

Footnote to Table 3S:

[a]A total of 132 ion peaks were analyzed with 20 μM MG-132. Their elution position in HPLC (Fraction N), monoisotopic mass (A: $M+H^+$), the number of R or K residues, the % relative intensity of the corresponding A+2, A+4 or A+6 peak in the absence of labeling (^{14}N), upon ^{15}N-labeling (^{15}N), or upon labeling in the presence of MG-132 (^{15}N+Inh), the labeling ratio (Ratio: see Experimental Procedures), and the amino acid sequence are indicated. Peptides were classified as inhibitor-sensitive or inhibitor-resistant when the labeling ratio was <0.35 (except when the ratio was altered by values too close to background in the presence of inhibitor, underlined) or >0.6, respectively. Peptides were not-assigned when their labeling ratios were intermediate between these values or when they showed lower ratios with epoxomicin (marked with *)

Table 4S. Isotopic labeling of HLA-A*68:01 ligands in the presence of epoxomicin[a].

Epoxomicin Sensitive Ligands							
Fraction	**M+H$^+$**	**Residues**	**^{14}N**	**^{15}N**	**^{15}N+Inh**	**Ratio**	**Sequence**
92	969.46	R	16.35	124.00	38.26	0.20	
124	986.52	R	43.27	72.30	41.03	-0.08	
125	991.51	R	20.54	72.45	32.52	0.23	
110	992.53	R	24.63	50.43	27.23	0.10	NAIGGKYNR
145	1000.42	R	17.93	95.28	54.90	0.48	
135	1000.42	R	20.18	95.47	42.31	0.29	
149	1000.43	R	19.30	93.53	43.83	0.33	
114	1003.51	R	16.26	78.34	14.32	-0.03	DTAAQITQR
118	1010.56	R	16.09	94.28	15.77	0.00	
109	1012.52	R	15.11	49.69	14.81	-0.01	TSETPQPPR
155	1021.50	R	15.70	44.29	20.83	0.18	DVFVVGTER
114	1025.49	R	17.04	75.00	20.63	0.06	
118	1032.53	R	19.60	103.43	35.29	0.19	
145	1032.57	R	38.20	104.17	69.57	0.48	FPSEIVGKR
132	1043.61	R	14.05	102.14	22.22	0.09	ELTAVVQKR
135	1054.33	R	26.23	77.87	37.21	0.21	
149	1054.34	R	28.04	84.40	41.12	0.23	
153	1060.57	R	18.08	77.03	22.97	0.08	DVFRDPALK
146	1070.58	R	17.23	94.77	42.42	0.32	EVAQLIQGGR
126	1075.67	2 R	1.85	307.48	29.40	0.09	HPAVVIRQR
129	1077.52	R	17.32	51.88	24.08	0.20	ETNYGIPQR
152	1089.58	R	20.27	62.29	32.67	0.30	
99	1108.53	R	17.62	63.10	23.18	0.12	
124	1122.59	2 R	1.61	13.72	2.97	0.11	
205	1123.58	R	23.58	383.06	70.13	0.13	
159	1126.66	R	23.42	47.31	23.79	0.02	
159	1128.66	R	23.42	47.31	23.79	0.02	EVFSHILKR
182	1137.60	R	43.22	389.42	157.43	0.33	EFVDLYVPR
138	1138.51	R	21.98	50.78	21.99	0.00	
130	1139.63	2 R	4.07	67.66	8.27	0.07	DPFHKAIRR
104	1140.60	3 R	7.84	27.99	10.47	0.13	EVNDRPVRR
124	1146.53	2 R	2.13	22.89	3.32	0.06	
113	1155.69	R	18.43	301.93	93.19	0.26	TVKKGKPELR
152	1180.57	R	26.38	56.83	36.23	0.32	
154	1212.60	R	27.00	87.00	43.00	0.27	FPHNPQFIGR
153	1245.72	R	31.86	112.60	52.42	0.25	
126	1246.66	R	22.49	746.40	40.76	0.03	TAADTAAQITQR
154	1376.69	2 R	12.00	175.00	48.67	0.22	NSFRYNGLIHR
121	1393.58	R	42.86	100.00	57.23	0.25	
176	1412.78	R	32.93	68.04	31.88	-0.03	
144	1608.80	3 R	0.00	74.00	21.00	0.28	
200	1629.82	R	45.70	179.60	84.32	0.29	
199	1651.80	R	77.00	166.79	78.26	0.01	
181	1713.84	2 R	7.28	497.75	80.00	0.15	
195	1719.83	R	58.52	95.08	56.83	-0.05	
166	1765.91	R	46.58	119.12	63.22	0.23	

Epoxomicin Resistant Ligands							
Fraction	**M+H$^+$**	**Residues**	**^{14}N**	**^{15}N**	**^{15}N+Inh**	**Ratio**	**Sequence**
118	945.51	R	0.17	1.00	1.21	1.25	DPAGVHPPR
167	1048.64	R	23.23	61.79	47.81	0.64	SVQGIIIYR
190	1132.59	R	37.24	88.54	76.47	0.76	
127	1200.56	R	25.23	121.81	90.10	0.67	
203	1207.65	R	26.36	341.32	235.59	0.66	SAFATPFLVVR
129	1209.63	2 R	19.65	64.36	52.22	0.73	ETVQLRNPPR
190	1441.71	R	34.96	83.65	78.44	0.89	
136	1462.74	R	0.67	2.75	2.37	0.82	TVKDVNQQEFVR

Not assigned							
Fraction	**M+H$^+$**	**Residues**	**^{14}N**	**^{15}N**	**^{15}N+Inh**	**Ratio**	**Sequence**
149	984.44	R	23.36	89.32	48.03	0.37	
115	1007.50	2 R	6.62	69.96	41.69	0.55	
130	1023.59	2 R	28.63	185.71	112.30	0.53*	
154	1103.65	2 R	6.00	74.19	42.44	0.53	
174	1132.64	R	23.89	107.32	55.26	0.38*	
130	1187.63	R	25.65	160.51	75.00	0.37*	NVAEVDKVTGR
143	1250.64	2 R	14.66	56.41	23.33	0.21*	
127	1443.81	R	39.82	155.56	80.07	0.35*	TVINQTQKENLR
159	1755.89	3 R	11.00	68.00	35.00	0.42	

Table 4S

Footnote to Table 4S:

[a]A total of 63 ion peaks were analyzed with 2.5 μM epoxomicin. See footnote to Table 3S for conventions.

Table 5S. Digestion products of the 31-mer substrate by 20S proteasome in the absence or presence of inhibitors[a].

HASIQM**NVAEVDKVTGRFNGQFKTY**AICGAI

No inhibitor

Residues	% Yield	Residues	% Yield	Residues	% Yield	Residues	% Yield	Residues	% Yield
1_25	8.4	6_28	1.2	10_31	0.1	12_26	1.7	14_22	2.7
1_23	0.3	6_20	0.4	10_28	0.1	12_25	0.8	15_27	0.5
1_18	1.9	6_17	0.4	10_27	1.8	12_23	1.0	15_25	0.3
1_17	0.6	7_27	1.0	10_25	4.8	13_31	2.1	15_22	1.8
1_14	0.3	7_25	1.9	10_24	0.2	13_30	1.5	16_31	0.4
1_12	0.8	7_19	0.2	10_22	0.6	13_29	0.4	16_27	0.3
1_11	0.3	7_18	0.5	10_21	0.3	13_27	2.1	**16_25**	**1.0**
1_10	0.8	**7_17**	**6.4**	10_20	2.1	13_26	0.3	16_22	2.0
1_9	2.9	8_30	0.8	10_19	0.4	13_25	4.2	17_27	0.3
1_8	4.9	8_27	0.6	10_18	1.6	13_24	0.1	17_25	2.1
3_16	0.5	8_25	0.9	10_17	0.7	13_19	0.6	17_22	1.6
3_14	0.2	9_27	0.2	11_31	0.2	14_29	0.2	18_31	0.2
4_16	0.2	9_25	0.1	11_27	0.4	14_28	2.7	18_27	0.3
5_25	0.9	9_22	0.2	11_26	0.1	14_27	0.3	18_25	4.6
5_24	0.6	9_19	0.5	11_25	0.8	14_26	2.0	19_27	0.3
5_17	0.4	9_17	0.6	12_27	0.4	14_25	2.0	19_25	4.3

MG132

Residues	% Yield	Residues	% Yield	Residues	% Yield
1_27	11.1	6_28	3.8	10_19	3.3
1_26	0.1	6_26	1.6	11_20	0.9
1_25	13.0	7_31	10.7	12_28	0.7
1_22	0.6	7_27	1.2	12_26	4.9
1_19	4.9	7_20	3.3	13_26	0.5
1_18	0.9	7_19	0.1	14_26	4.3
1_17	8.1	**7_17**	**0.3**	15_30	9.8
3_29	0.5	8_30	0.5	17_25	6.0
3_15	0.2	8_25	3.5		
6_29	0.2	10_27	6.2		

Epoxomicin

Residues	% Yield	Residues	% Yield
1_27	22.0	7_20	2.3
1_25	18.6	8_25	0.6
1_24	1.7	14_26	3.8
1_23	7.0	15_30	3.7
1_22	9.3	17_25	1.2
1_19	7.2	19_31	16.3
1_17	3.1		
3_29	0.3		
6_28	1.0		
7_31	2.0		

[a]Only peptides obtained with ≥0.1 % yield are shown. In the presence of MG132 and epoxomicin only 28.9% and 54.3 % of the substrate, respectively, was digested. The % yield of the digestion products in the presence of these inhibitors is relative to the amount of digested material, excluding the undigested substrate. The natural A68 (7-17) and B27 (16-25) ligands are highlighted in boldface.

Table 6S. *Effect of inhibitors on the cleavage at individual peptide bonds of the synthetic 31-mer by proteasome 20S*

	% contribution per residue[a]		
Tryptic	**No inhibitor**	**MG-132**	**Epoxomicin**
K13	7.5	2.8	3.2
R17	6.9	5.5	2.6
K23	1.0	0	5.8
Total	**15.4**	**8.2**	**11.6**
Chymotryptic	**No inhibitor**	**MG-132**	**Epoxomicin**
A2	0.5	0.4	0.3
I4	1.4	0	0.0
M6	7.6	10.1	3.6
V8	3.7	0	0
A9	2.2	6.1	0
V11	0.2	3.6	0
V14	0.4	6.4	3.1
F18	2.7	0.5	13.6
F22	6.8	0.4	7.8
Y25	28.2	14.6	17.0
A26	3.1	7.4	3.2
I27	6.5	12.1	18.4
A30	1.8	8.7	3.1
Total	**65.1**	**70.4**	**69.9**
Caspase	**No inhibitor**	**MG-132**	**Epoxomicin**
E10	0.6	0.6	0
D12	0.6	0.3	0
Total	**1.2**	**0.9**	**0**
Other	**No inhibitor**	**MG-132**	**Epoxomicin**
S3	0.2	0	0
Q5	1.5	2.6	0.8
N7	1.8	2.6	0.5
T15	2.8	0	0
G16	0.5	3.9	1.0
N19	5.2	5.3	13.6
G20	1.9	2.7	0
Q21	0.2	0	0
T24	0.7	0	1.4
C28	3.0	3.0	0.8
G29	0.5	0.4	0.3
Total	**18.3**	**20.5**	**18.4**

Footnote to Table 6S:

[a]Joint yield of digestion products resulting from cleavage at these peptide bonds. Cleavage at a given peptide bond i/i+1 was estimated from the yield of all the peptides whose C-terminal residue was i, disregardings the yields of peptides whose N-terminal residue was i+1. If no peptides with C-terminal residue i were found, then the yields of peptides starting at i+1 were assigned to that peptide bond. Values were subsequently normalized to 100%.

Table 7S. *Effect of inhibitors on the cleavage at individual peptide bonds of the30-mer substrate by proteasome 20S*

	% Contribution per residue		
Tryptic	**No inhibitor**	**MG-132**	**Epoxomicin**
R14	16.8	9.0	0.0
K15	2.6	1.5	4.0
R21	4.5	4.8	4.9
K26	3.7	2.1	3.8
Total	27.6	17.3	12.7
Chymotryptic			
A4	0.0	4.0	0.0
F7	2.4	2.5	3.7
V8	4.2	5.4	10.0
L10	0.0	3.3	0.0
Y11	3.5	2.1	3.8
V12	1.6	2.9	0.0
A18	0.0	0.0	3.3
I22	4.2	5.4	10.0
I23	3.7	0.0	0.0
A25	2.3	3.4	0.0
A29	3.5	6.2	6.8
Total	25.4	35.2	37.5
Caspase			
D3	7.5	7.3	0.0
E6	0.0	0.0	0.0
D9	6.0	4.8	8.5
D27	7.7	0.0	4.0
Total	21.1	12.1	12.5
Other			
Q1	3.6	6.0	7.6
N2	3.5	6.2	6.8
G5	0.0	2.9	0.0
P13	2.1	0.0	0.0
C16	8.3	3.1	3.2
S17	3.9	6.2	5.2
S19	2.4	4.3	6.4
N20	2.1	2.7	4.8
G24	0.0	0.0	0.0
H28	0.0	4.0	3.3
Total	25.9	35.4	37.2

Footnote to Table 7S:

[a]Joint yield of digestion products resulting from cleavage at these peptide bonds. See footnote to Table 6S.

Apéndice II

Functional Interaction of the Ankylosing Splondylitis Associated ERAP1 Polymorphism and HLA-B27 in vivo.

Manuscript submitted.

TITLE:

Functional interaction of the ankylosing spondylitis associated ERAP1 polymorphism and HLA-B27 *in vivo*

Authors:

Noel García-Medel[1*], Alejandro Sanz-Bravo[1*], Dung Van Nguyen[1], Begoña Galocha[1,2], Patricia Gómez-Molina[1], Adrián Martín-Esteban[1], Carlos Alvarez-Navarro[1], and José A. López de Castro[1**].

Affiliations:

[1]Centro de Biología Molecular Severo Ochoa, CSIC-UAM, 28049 Madrid, SPAIN.

[2]Current address: Instituto de Salud Carlos III, 28220 Madrid, SPAIN

Contact:

**Correspondence: aldecastro@cbm.uam.es

Footnotes:

*The contribution of N. García-Medel and A. Sanz-Bravo to this work was equal.

Running Title:

ERAP1 and the HLA-B27 peptidome

Abstract

The association of ERAP1 with ankylosing spondylitis (AS) among HLA-B27-positive individuals suggests that ERAP1 polymorphism may affect pathogenesis through altering peptide-dependent features of the HLA-B27 molecule. Comparison of HLA-B*27:04-bound peptidomes from cells expressing different natural variants of ERAP1 revealed significant differences in the size, length and amount of many ligands, and in HLA-B27 stability. Peptide analyses suggested that the mechanism of ERAP1/HLA-B27 interaction is a variant-dependent alteration in the balance between epitope generation and destruction determined by the susceptibility of N-terminal flanking and P1 residues to trimming. ERAP1 polymorphism associated with AS susceptibility ensured efficient peptide trimming and high HLA-B27 stability. Protective polymorphism resulted in diminished ERAP1 activity, less efficient trimming, suboptimal HLA-B27 peptidomes and decreased molecular stability. This study demonstrates that natural ERAP1 polymorphism affects HLA-B27 antigen presentation and stability *in vivo* and proposes a mechanism for the interaction between these molecules in AS.

The mechanism underlying the strong association of HLA-B27 with ankylosing spondylitis (AS) remains unknown. Three main possibilities, each one based on a different molecular feature of HLA-B27, are currently being investigated. The *arthritogenic peptide* hypothesis (1), based on the canonic antigen-presenting properties of Major Histocompatibility Complex class I (MHC-I) molecules, assumes that a peptide epitope of external origin would activate HLA-B27-restricted T-cells whose cross-reactivity with a self-derived HLA-B27 ligand would result in autoimmune damage. The *misfolding* hypothesis (2) is based on the slow folding and tendency to misfold of HLA-B27 (3,4). Accumulation of misfolded heavy chain (HC) in the endoplasmic reticulum (ER) would elicit a misfolded protein response and activate pro-inflammatory pathways. The *surface homodimer* hypothesis (5,6) is based on the expression of HLA-B27 HC homodimers at the cell surface and their recognition by leukocyte receptors (7), leading to immunomodulation of inflammatory responses. Since the constitutive binding of endogenous peptides by MHC-I molecules determines not only their antigen-presenting specificity, but also their folding and stability, it was proposed that the HLA-B27 peptidome, through its global influence on the biological behavior of the molecule, is critical to its pathogenetic role (8). This idea found strong support from the discovery of the association of the ER amino peptidase (ERAP)1 with AS (9) in HLA-B27-positive, but not B27-negative disease (10). With an estimated population attributable risk of 26%, ERAP1 is the non-MHC gene most strongly associated to AS. Since ERAP1 is involved in the N-terminal trimming of peptides to their optimal size for MHC-I binding (11-13) its association with AS suggests a pathogenetic mechanism of functional interaction with HLA-B27 through influencing peptide binding and antigen presentation. ERAP1 trimming is limited by peptide size, becoming highly inefficient for 8-mers and shorter peptides (13,14). The three-

dimensional structure of this molecule revealed a substrate binding cavity close to the catalytic site, as well as 4 domains, whose conformational rearrangement between an open and a closed conformation, presumably induced upon substrate binding, regulates its enzymatic activity (15,16). The polymorphic residues found among natural ERAP1 variants (17) and often co-occurring in complex allotypes, are located in various topological regions, including close proximity to the catalytic site, the substrate binding cavity or close to domain junctions. Therefore, they may alter ERAP1 activity by directly affecting catalysis, altering substrate binding or modulating domain rearrangements. The association of ERAP1 with AS does not by itself reveal the specific feature(s) determining the pathogenetic role of HLA-B27. Indeed, ERAP1 may influence the generation of specific pathogenetic epitopes, have a general effect on the HLA-B27 peptidome, altering the stability or other features of the molecule, or both. This study investigated general effects of ERAP1 polymorphism on the HLA-B27 peptidome, by comparing the size distribution, molecular features and N-terminal flanking sequences of peptides from human cells expressing the AS-associated B*27:04 subtype and different natural variants of ERAP1.

EXPERIMENTAL PROCEDURES

Cell lines and antibodies-The following B*27:04-positive cell lines were used. JSL (HLA-A*11; B*27:04, *48; C*02), WEWAK I (WE-I: HLA-A*11, *24; B*27:04, *62; C*02, *04) and KNE (HLA-A*01, *02:04; B*27:04, *08) are lymphoblastoid cell lines. C1R04 is a transfectant of the HLA class I-defective Hmy2.C1R cell line (18) expressing B*27:04 (19). The cells were cultured in RPMI 1640 medium supplemented with 2 mM L-glutamine and 10% FBS (Gibco, Life Technologies, Paisley, UK). ME1

(IgG1), an anti-HLA-B7/B27/B22 monoclonal antibody (mAb) that recognizes HC/β_2m/peptide complexes(20), was used for immunopurification of HLA-B27.

Typing of non-synonymous SNPs in the ERAP1 gene-DNA purification was performed using the High Pure PCR Template Preparation system (Roche Diagnostics, Barcelona, Spain) following the instructions of the manufacturer. Aliquots of 10 ng were added onto 384-well plates in duplicate, dried and amplified using specific oligonucleotides for 8 non-synonymous SNPs located in the coding sequence of the ERAP1 gene: rs6653, rs26618, rs27895, rs27044, rs30187, rs10050860, rs17482078, and rs2287987. Samples were run in a HT7900 Fast real-Time PCR System and genotyped using the SDS2.2 software (both from Applied Biosystems, Life Technologies, Carlsbad, CA) for allelic discrimination.

Sequencing of ERAP1 variants-The exons 2-20, encompassing the coding region of ERAP1 were separately amplified by PCR and cloned into M13 for sequencing. PCR products were generated using AmpliTaq Gold PCR Master Mix (Applied Biosystems) following standard procedures, purified using ExoSap (USB Corp., Cleveland, OH) and sequenced in a 3730XL instrument (Applied Biosystems). Sequencing primers were either M13-complementary oligonucleotides, for amplicons which included this extension along the ERAP1 sequence, or the specific primers themselves when the amplicons lacked M13-derived sequences. Both strands of amplicons were routinely sequenced.

Quantitative real time(RT)-PCR-Total cellular RNA was extracted employing the RNeasy Mini Kit (Qiagen) followed by digestion with DNase I (Invitrogen, Karlsruhe, Germany). Complementary DNA was synthesized from 250 ng of total RNA employing the High Capacity cDNA reverse transcription kit (Life Technologies) according to the manufacturer instructions. The primers used for the amplification were purchased from

Applied Biosystems (Life Technologies). Comparative quantification of gene expression was performed by qRT-PCR using a AB7900HT instrument (Applied Biosystems, Life Technologies) using TaqMan probes and Gene Expression Master Mix (Applied Biosystems, Life Technologies). Amplifications were carried out by an initial hold at 50°C for 2 min followed by denaturation at 95°C for 10 min, 40 cycles of denaturation at 95°C for 15s and annealing and extension at 60°C for 60s. The results were expressed as relative mRNA expression quantified with the RQ Manager software and normalized to glyceraldehyde-3-phospahate dehydrogenase (GAPDH) transcript levels.

Western blot-About 2 X 10^5 cells were lysed in 0.5% Igepal CA-630 (Sigma-Aldrich, St Louis, MO), 50mM Tris HCl, 5 mM $MgCl_2$, pH 7.4, containing protease inhibitors (Complete Mini; Roche, Mannheim, Germany). After SDS-PAGE (10% slab gels) of whole lysates under reducing conditions, the separated components were electroblotted onto a nitrocellulose membrane (Amersham Hybond-ECL, GE Healthcare, UK) at 20V overnight using 30% methanol in 50 mM Tris/Gly buffer, pH 8.8, 0.04% SDS. ERAP1 and γ-tubulin, used as an internal standard, were revealed with the 6H9 mAb (a kind gift of Peter van Endert, INSERM, Paris, France) or GTU88 (Sigma-Aldrich), respectively, using peroxidase-conjugated goat anti-mouse Ig polyclonal antibody (DakoCytomation, Glostrup, Denmark). The scanned autoradiograms were quantified using TINA 2.09e image analyzer software (Raystest Isotopenmessgeräte, Straubenhardt, Germany).

Isolation of HLA-B27-bound peptides-This was carried out as previously described (21). Briefly, cells were lysed in 1% Igepal CA-630 with a cocktail of protease inhibitors (Roche). The soluble fraction was subjected to affinity chromatography using the ME1 mAb. HLA-B27-bound peptides were eluted with 0.1%

aqueous TFA at room temperature, filtered through Centricon 3 (Amicon, Beverly, MA) or Vivaspin 2 Hydrosart (VS02H11, Sartorious Stedim Biotech, Goettingen, Germany), concentrated and subjected to HPLC fractionation in a Waters Alliance system (Waters, Milford, MA) using a Vydac 218TP52-C18 column (Vydac, Hesperia, CA) at a flow rate of 100 µl/min, as described (22). Fractions of 50 µl were collected and stored at -20ºC.

Mass spectrometry-Individual HPLC fractions were analyzed by matrix-assisted laser desorption-ionization time-of-flight (MALDI-TOF) mass spectrometry (MS) using a 4800 Proteomics Analyzer (Applied Biosystems), as described elsewhere (23). The mass spectra were acquired in reflector positive mode at 25kV in the mass-to-charge (m/z) range 800-2000, using a signal-to-noise ratio (s/n) cutoff of 3, and processed using the Data Explorer software version 4.9 (Applied Biosystems). Sample handling and acquisition parameters were tightly controlled to minimize differences in the experimental conditions among samples to be compared in each experiment.

Peptide sequencing was carried out by MALDI-TOF/TOF MS/MS, as previously described (23). Interpretation of the MS/MS spectra was assisted by various tools as follows. Manual inspection of the spectrum usually allowed us to derive a tentative sequence. This was used to screen the human proteome in the human protein entries of the Uniprot/Swiss-prot database (Release 57.6: 2009/06/28, with 20331 entries), using a window of 0.5 m/z units for both precursor and fragment ions, for a possible match using the Mascot server 2.2 software. For those sequences showing the highest scores in this preliminary search, the MS-product tool (version 5.9.4) at http://prospector.ucsf.edu (University of California, San Francisco, USA), was used to match the candidate sequences to our MS/MS spectra.

Trimming susceptibility of N-terminal flanking and P1 residues-The susceptibility of flanking (P-2, P-1) and P1 residues to ERAP1 trimming was estimated by assigning a score to each amino acid, ranging from 0 to 100, based on a previously reported assay that measured presentation of the SIINFKEL peptide from ER-targeted precursors (24). The score assigned to each residue (s**upplemental Table S1**) corresponded to the mean percent of SIINFKEL expressed at the cell surface from XX-SIINFKEL precursors, relative to the most susceptible residue in this assay (Y: score 100). Proline and any residue immediately preceding it was assigned score zero.

Contribution of internal peptide residues to ERAP1-mediated trimming-This was estimated based on a previous study (25) in which a 9-mer library was used to assess the influence of different residues at each peptide position on trimming of P1 by ERAP1 *in vitro*. Each residue at a given position, from P3 to P9, was assigned a score corresponding to the percentage of depleted substrate in that experiment (s**upplemental Table S1**). Residues not directly tested in that study were not scored.

Thermostability assay-This was performed as previously described (26). Briefly, the cell lines were pulse-labeled for 15 min and chased at various times. At each time point the cell lysates were incubated for 1 h at various temperatures, immunoprecipitated with ME1 and analyzed by SDS-PAGE. The amount of heterodimer precipitated at each temperature at any given time was expressed as a percentage of the amount precipitated at 4ºC, and plotted as a function of the temperature.

RESULTS

*ERAP1 polymorphism in HLA-B*27:04 positive cell lines*-Four B*27:04-positive cell lines, including 3 LCL and C1R transfectants, were selected on the basis of their

expression of 8 single-nucleotide polymorphisms (SNPs) encoding non-synonymous substitutions in ERAP1, including 6 AS-associated ones (9,17). Sequencing of all the exons of this gene in the 4 cell lines confirmed these polymorphisms (**Table I**). Whereas the coding sequence of ERAP1 from JSL was identical to the consensus and carried all 6 polymorphisms associated with increased susceptibility to AS, that from WE-I showed 6 non-synonymous changes, including all those associated with protection from AS. ERAP1 from C1R04 showed 4 non-synonymous changes relative to the consensus. KNE was heterozygous for rs26653, rs26618 and rs27895. This variability allowed us to examine the effect of various combinations of amino acid changes in natural ERAP1 variants on the HLA-B27-bound peptidome.

Western blot analyses of WE-I, C1R04, and KNE revealed that these cell lines expressed similar ERAP1 protein levels (**Fig. 1A-B**). Although these could not be assessed for JSL due to difficulties in maintaining this cell line in the course of this study, the relative mRNA expression of ERAP1, as determined by quantitative RT-PCR (**Fig. 1C**) showed little difference among the four cell lines. Indeed, ERAP1 protein expression was reported to be fairly invariant among LCL (27).

Automated comparison of HLA-B27-bound peptide repertoires-B*27:04-bound peptide pools were fractionated by HPLC, and each fraction was analyzed by MALDI-TOF MS. Three pairwise comparisons, WE-I/JSL, WE-I/C1R04 and C1R04/KNE, were carried out, following a strategy previously used for HLA-B27 subtype-bound peptidomes (28-30). For each cell line pair, the MS spectra of correlative HPLC fractions from consecutive chromatographic runs performed under identical conditions were automatically compared, using a newly developed software, Mshandler (**Fig. 2**). All the ion peaks in the m/z range 800-2000 and s/n>3 were collected from the MS spectra of all HPLC fractions from each peptide pool, and identified by their m/z and

HPLC fraction number. Prior to each comparison, a series of filters were applied to remove irrelevant signals and maximize the selection of B27-related ion peaks. The filtered ion peak sets from the two cell lines in each experiment were compared: ion peaks with identical (±0.2) m/z in the correlative ±1 HPLC fraction of the other cell line were considered as shared peptides in the two cell lines and subjected to further analysis. All other peaks were disregarded. A total of 3222, 5240, and 6974 shared B*27:04 ligands were obtained from WE-I/JSL, WE-I/C1R04 and C1R04/KNE, respectively (**supplemental Table 2S**).

*Relative expression of B*27:04 ligands in distinct ERAP1 contexts*-We reasoned that if there were not a general influence of ERAP1 polymorphism on the HLA-B27 peptidome, or in the context of similar ERAP1 molecules, the size distribution of shared peptides in two cell lines would be essentially identical independently of their relative abundance in each cell line. Conversely, if ERAP1 polymorphism has a general influence on peptide trimming, the size distribution of shared B27 ligands would be shifted towards higher molecular mass (Mw) values in the cell line with the less efficient ERAP1 variant, to an extent that would depend on the relative abundance of the peptides compared. Thus, we calculated the intensity ratio (IR) between shared ion peaks in each of the WE-I/JSL, WE-I/C1R04 and C1R04/KNE comparisons, as an estimation of relative peptide abundance. The peptides showing IR≥1 were classified in 3 subsets: those showing >3-fold, >1.5 to 3-fold, and ≥1 to 1.5-fold intensity in one relative to the other cell line **(Table II)**. We next compared the Mw distribution within equivalent subsets in both cell lines. The results are described below.

WE-I/JSL (**Fig. 3A**). Their shared ion peaks showed a gaussian size distribution with mean m/z: 1162.0. For peptides with IR ≥1 in each cell line there was a shift of the gaussian curve from WE-I, relative to JSL, towards higher Mw values (Δm/z: 33.7).

When the different IR subsets were compared, the magnitude of the shift was highest between the IR>3 subsets (Δm/z: 73.1), somewhat smaller between those with IR>1.5-3 (Δm/z: 49.2), and virtually overlapping (Δm/z: 5.9) in the 1-1.5-fold subsets.

WE-I/C1R04 (**Figure 3B**). The shared ion peaks in these two cell lines showed a gaussian size distribution with mean m/z: 1154.3. The curve of the peptides with IR≥1 in WE-I was shifted towards higher Mw, relative to C1R04, similarly as for JSL (Δm/z: 30.6). When IR subsets were compared a similar curve shift than in WE-I/JSL was observed for the IR>3 and >1.5-3 subsets, although the mean Mw differences were somewhat smaller (Δm/z: 58.6 and 38.4, respectively) (**Table II**). These results indicate that B27 ligands with high Mw tend to be more abundant in WE-I than in JSL or C1R04, and the opposite is true for ligands with low Mw, but differences are smaller with C1R04 than with JSL.

C1R04/KNE (**Figure 3C**). The ERAP1 molecules in these two cell lines were identical except at positions 127, 276 and 346, where KNE, but not C1R, showed heterozygosis (**Table I**). The size distribution and average Mw of shared B27 ligands were very similar in both cell lines (Δm/z <10 Da) for all the IR subsets compared, indicating little or no effect of the ERAP1 differences between C1R04 and KNE on these features**.**

The results indicate that allelic ERAP1 polymorphism has a global influence on the B*27:04 peptidome, affecting the relative expression of many B*27:04 ligands as a function of peptide size. The observed effects are consistent with the following ERAP1 activity ranking: JSL>C1R04/KNE>WE-I.

*Relationship between size and length distribution of B*27:04 ligands*-Due to the wide Mw range among peptides of the same length, translating the observed peptide size into length differences, which best reflect the relative activity among ERAP1

variants, was not straightforward. Thus, the amino acid sequence of 372 B*27:04 ligands showing various degrees of differential expression between the cell lines compared was determined (**supplementary Table S3**), and the percentage of peptides of any given length within each Mw range was calculated (**Table III**). These values were applied to translate the Mw differences described in the previous paragraph into peptide length differences (**Fig. 4**).

WE-I/JSL (**Fig. 4A**). The shared B*27:04 ligands between these cell lines was estimated to include 2.5% 8-mers or smaller peptides, 70.7% 9-mers, 21.1% 10-mers and 5.8% longer peptides. Peptides with IR ≥1, showed a skewing against short peptides (9-mers and shorter) and towards long ones (10-mers and longer) in WE-I, whose magnitude depended strictly on their expression relative to JSL, being highest for the IR>3 subsets, intermediate for IR>1.5-3 subsets, and low among peptides with IR≥1-1.5

WE-I/C1R04 (**Fig. 4B**). As in the previous comparison, peptides with IR≥1 showed a moderate skewing against short peptides and towards long ones in WE-I. Length skewing was again more prominent between the IR>3 subsets, decreased in the IR>1.5-3 subsets and was marginal in the IR≥1-1.5 subsets. The pattern was similar to that in WE-I/JSL, but the differences were smaller.

C1R04/KNE (**Fig. 4C**). Length differences among the B*27:04 ligands of these cell lines were only noticeable among the peptide subsets with IR>3 (C1R04/KNE ratio 0.8 and 1.5 for peptides <9-mers and >10-mers, respectively).

Thus, in WE-I long B*27:04 ligands are more abundant and short ones less abundant compared to JSL and, to less extent, to C1R04. The results indicate that ERAP1 polymorphism in WE-I leads to decreased trimming and affects the expression level of a substantial percentage of the B27 peptidome. That the observed differences

are mainly due to ERAP1, rather to unrelated factors, is strongly supported by the similarity between C1R04 and KNE.

*The N-terminal flanking and P1 residues determine the ERAP1-dependent expression of B*27:04 ligands*-Amino acid residues differ widely in their susceptibility to ERAP1 trimming. Susceptible residues in the N-terminal flanking sequences favor the generation of B*27:04 ligands, whereas susceptible P1 residues may favor their destruction. Thus, the N-terminal flanking residues of numerous B*27:04 ligands showing distinct relative expression among cell lines were assigned from the sequence of their parental proteins (**supplemental Table S3**). The sequenced peptides in each pairwise comparison were classified in those with IR>3 and those with IR ≥ 1-3 in one cell line, relative to the other. The flanking (P-2 and P-1) and P1 residues of each peptide were assigned a trimming susceptibility score (**supplemental Table S1**) based on a previously reported assay (24). For each peptide subset, a mean score was determined for P-2, P-1, P-2+P-1, and P1. The ratio between the mean scores of equivalent peptide subsets in the cell lines compared was used as a global estimation of their relative susceptibility to trimming. The following observations emerged from this analysis (**Table IV**).

WE-I/JSL. Both the flanking and P1 residues of the peptides in the IR>3 subsets were more susceptible to trimming in WE-I than in JSL (added P-2+P-1 score ratio: 1.4; P1 score ratio: 2.5), but those in the IR<3 subsets showed smaller differences. These results strongly suggest that the predominance of B*27:04 ligands in these cell lines is determined by higher or lower susceptibility of the flanking and P1 residues to trimming, in the context of the less (WE-I) or more active (JSL) ERAP1 variant, respectively.

WE-I/C1R04. The flanking residues of the peptides in the IR>3 subsets were more susceptible to trimming in WE-I than in C1R04 (added P-2+P-1 score ratio: 1.4).

In contrast, the average susceptibility of P1 residues to trimming was very similar in both cell lines for all peptide subsets, regardless of their relative expression (**Table IV**). Thus, again, the predominance of B*27:04 ligands in these cell lines correlated with higher susceptibility of the flanking residues to trimming in the context of the less active variant (WE-I). The similar P1 susceptibility suggests that differential epitope destruction contributes less to peptide predominance between these two cell lines. Yet, similarly susceptible P1 residues should be more efficiently trimmed by the most active ERAP1 variant.

C1R04/KNE. This analysis revealed smaller differences in trimming susceptibility of the flanking and P1 residues between the peptides predominant in either cell line, relative to the previous comparisons (**Table IV**). Score ratios for any given position or peptide subset were no higher than 1.3 and their magnitude did not correlate with IR. Thus, peptide expression differences between C1R04 and KNE seem largely unrelated to ERAP1, in agreement with the similar activity of this enzyme in both cell lines suggested by the size and length distribution of their HLA-B27 peptidomes.

*The internal sequence features of B*27:04 ligands do not determine their differential ERAP1-dependent expression*-We investigated whether peptide residues other than P1 might account for the prevalence of some B*27:04 ligands in a given ERAP1 context. This analysis was carried out only for nonamers, and was based on the reported role of residues downstream of P1 in modulating trimming by ERAP1 (25). Each residue at positions P3 to P9, since P2 is nearly invariant among B*27:04 ligands, was assigned a score related to its effect on ERAP1-mediated trimming (**supplemental Table S1**). For each peptide subset the mean score of each position, and the mean of the P3-P9 scores, was calculated. Although position-dependent differences were observed

(**supplemental Table S4**), the joint contribution of P3-P9 to trimming was virtually identical in all cases, independently of the relative abundance of the peptide subsets compared (**Table V**). These results strongly suggest that, globally, the internal peptide positions do not have a significant influence on the differential expression of B*27:04 ligands in distinct ERAP1 contexts.

ERAP1 polymorphism influences HLA-B27 stability-The molecular stability of B*27:04 expressed in various ERAP1 contexts was analyzed by measuring the thermostability of HLA-B27 isolated from the 4 cell lines in this study (**Fig. 5**). We previously reported that B*27:04 expressed on C1R cells shows very high thermostability, a feature that was shared with 2 other AS-associated subtypes B*27:02 and B*27:05, but not with the non AS-associated B*27:06 and B*27:09 (31). An equally high thermostability of B*27:04 was observed in JSL, and nearly as much in KNE, but WE-I showed lower thermostability even at long chase times. This result strongly suggests that the B*27:04-bound peptidome in WE-I is less optimized than in the other cell lines analyzed.

DISCUSSION

The role of natural ERAP1 polymorphism in shaping the HLA-B27 peptidome was addressed in this study. We determined quantitative effects on the expression of B27 ligands, assuming that differential processing among ERAP1 variants may affect many peptides. We did not look for specific ligands whose presence might depend on a particular ERAP1 context, because the absence of a peptide cannot be formally established, and there are no likely candidates for AS-specific epitopes.

Our experimental approach, which was a powerful one to detect general effects on the peptidome, deserves some comments. The intensity of an ion peak in MALDI-TOF MS is influenced by many factors, including instrument settings, sample handling,

the presence of other components, etc. For this reason, although the acquisition parameters and sample preparation were strictly controlled, the IR of a given ion peak from two peptide pools is only an estimation of relative amounts. Yet, when applied to many ion peaks, collective tendencies can be revealed. Thus, our data should be interpreted as general patterns within large peptide sets and not on an individual peptide basis. MALDI-TOF MS was used in previous studies to estimate expression differences of HLA-B27 ligands in the absence or presence of tapasin (32) or among HLA-B27 subtypes (28). A second issue concerns the scores used. We based our score system for flanking and P1 residues on an assay that analyzed antigen presentation as a function of ERAP1 processing in the ER (24), rather than *in vitro* trimming of synthetic substrates, which may not accurately reflect trimming *in vivo*. To our knowledge, only one systematic study (25) analyzed the effect of internal peptide residues on ERAP1 trimming. It allowed us to examine the sequence of B27 ligands for their influence on ERAP1-dependent peptide expression. That study used nonamer libraries and did not include all residues at each position, which precluded an exhaustive scoring and restricted our analysis to nonamers. Again, although not fully accurate for individual ligands, this was suitable for globally assessing extensive peptide sets.

Many ERAP1-independent factors may influence the relative expression of MHC-I ligands among cell lines. However, most of these factors should not imply cell-specific effects on peptide size. Thus, cells expressing similar ERAP1 molecules might be expected to generate B*27:04 peptidomes with similar size distribution. That this was the case between C1R04 and KNE, strongly suggests that the significant size differences observed in the context of more distinct variants were due to ERAP1 polymorphism. Indeed, the ERAP1 activity ranking deduced from the effects on the size

and length of B27 ligands, JSL>C1R04/KNE>WE-I, correlated with the number of amino acid differences among the ERAP1 molecules in these cell lines.

Differences in ERAP1 expression levels among cell lines are very unlikely to account for the observed differences among the HLA-B27 peptidomes, since WE-I and C1R04 expressed very similar ERAP1 amounts. Moreover, ERAP1 mRNA expression in JSL, although similar to the other cell lines, actually showed the lowest value.

What is the mechanism by which ERAP1 polymorphism modulates the quantitative expression of B*27:04 ligands?, what peptides are more likely to be affected?. ERAP1 can remove almost any N-terminal residue of peptides longer than a minimum length, albeit with different efficiencies (24). Therefore, ERAP1 variants with different enzymatic activity might show differences in the generation of MHC-I ligands as a function of their N-terminal flanking extensions. A previous study (33) showed that, although residues P-3 and more distant can probably be removed by a variety of peptidases *in vivo*, P-2 and P-1 are predominantly removed in the ER. Moreover, activity differences among ERAP1 variants will affect epitope destruction because the P1 residue will be more efficiently cleaved by the most active variant. We have directly observed this effect with synthetic peptides *in vitro* (our unpublished observations). The higher susceptibility to trimming of flanking and P1 residues in the peptide subset most predominant in WE-I, relative to its equivalent in JSL, implies that these positions would be trimmed more efficiently by the more active ERAP1 variant in JSL, so that the corresponding ligands would be not only generated more efficiently, but also destroyed to larger extent, in the latter cell line. The abundance of these peptides in WE-I would be explained by the presence in this cell line of a less active ERAP1, capable of removing susceptible flanking residues without extensively destroying the epitope.

Many peptides predominant in JSL would be generated less efficiently in WE-I due to their lower susceptibility of their flanking residues to trimming.

The same mechanism would explain the observed differences between WE-I and C1R04, but here the similar susceptibility of P1 residues to trimming in the predominant peptide subsets from both cell lines means that the more active ERAP1 in C1R04 would still destroy B27 ligands more extensively than in WE-I, but less than in JSL. This may explain why the size differences between the B27 peptidomes from WE-I/C1R04 are smaller than in WE-I/JSL.

In conclusion, natural ERAP1 polymorphism has a significant influence on the length and abundance of many HLA-B27 ligands. The less active variant generated higher numbers of long peptides and less short ones, and also increased the expression of peptides of any length with susceptible flanking and/or P1 residues. Thus, the variant-dependent alteration in the balance between epitope generation and destruction, as a function of these residues, is a basic mechanism determining ERAP1-dependent differences among HLA-B27 peptidomes. Although the internal sequence features of B*27:04 ligands did not globally affect relative peptide expression in distinct ERAP1 contexts, our results do not exclude an influence of individual sequences on ERAP1 trimming, as reported *in vitro* (25).

Several studies analyzed the effect of natural polymorphisms on the enzymatic activity of ERAP1. The K528R and R725Q mutations, both of which are protective from AS, decreased ERAP1 activity, whereas R127P, I276M, D575N, and Q730E had smaller or no effect (10,16,34). These studies were performed with single mutants *in vitro*, where the mutual influence of co-existing mutations, or their effects *in vivo,* were not addressed. In a recent study (35) the kinetics of ERAP1 activity was consistent with a substrate-inhibition model that was both substrate and allele specific, and might

differentially affect antigen processing *in vivo* by ERAP1 variants. In that study the Q730E mutation, which is also protective from AS, decreased the presentation of a specific B27 epitope *in vivo* even more effectively than K528R, in contrast with the relative activity of these mutants *in vitro*. In another study (36) polymorphism at residue 730 affected the surface expression of free MHC-I HC in monocytes of HLA-B27-positive individuals. The three-dimensional structure of ERAP1 (15,16) shows that, among the polymorphic positions in the cell lines used in our study, M349V is located in the active site, G346D, R725Q and Q730E are in the substrate binding cavity, R127P, I276M and K528R are located at or near domain junctions and D575N is located at the surface of domain III. Thus, these polymorphisms could affect ERAP1 function by modulating hydrolysis, substrate binding or the conformational transition between the open and closed states, which is critical for the enzymatic activity.

ERAP1 in JSL differs from both C1R/KNE and WE-I at positions 127(R), 528(K), and 730(Q). Therefore, the higher activity of ERAP1 from JSL is in agreement with the reported effects of the K528R and Q730E changes on peptide trimming. Yet, in spite of the significant differences between the B*27:04 peptidomes of C1R04 and WE-I, their ERAP-1 molecules are identical at positions 528(R) and 730(E), indicating that polymorphism at other positions also influences ERAP1 trimming *in vivo*. For instance, WE-I differs from C1R and all other cell lines by the M349V, D575N, and R725Q, which may affect ERAP1 function, as demonstrated for the latter change *in vitro* (10). Whether the observed differences among B*27:04 peptidomes are predominantly due to a single mutation or to their combined effect cannot be determined from our study. That the effect of a given mutation in ERAP1 might be modulated by concomitant changes at other positions should always be taken into account. The co-occurrence of AS-associated non-synonymous SNPs in a given ERAP1 variant is very frequent, leading to

identification of AS-associated ERAP1 haplotypes (17,37,38). Because of strong linkage disequilibrium, it is difficult to distinguish whether association to AS is determined by a single position or by co-occurring polymorphisms in the susceptible haplotypes. In the absence of definitive evidence for the former possibility, the latter might seem a plausible alternative.

The possible role of ERAP2 in the observed alterations of the B*27:04 peptidome was not analyzed. Expression differences of this enzyme were reported among human LCL (27). However, ERAP2 expression in WE-I, as assessed by Western blot, was about 2-fold higher than in C1R04 (our unpublished observations) in spite of the lower trimming observed in the former cell line. We cannot rule out the possibility that ERAP1 polymorphism influences the concerted trimming by ERAP1/ERAP2 complexes (39) through altering the interaction between both molecules. Yet, the location of most of the natural ERAP1 changes in sites susceptible to directly alter trimming may argue against a major influence of this alternative.

The effects of ERAP1 polymorphism described here may influence HLA-B27 biology in various aspects relevant to AS pathogenesis. First, the differential effects on trimming observed among ERAP1 variants may result in the production or not of some B27 ligands. Since protection from AS has been linked to loss-of-function ERAP1 mutations (10), one might speculate that putative disease-associated epitopes specifically produced in the context of the most active variants might be peptides with flanking and/or P1 residues not highly susceptible to ERAP1 trimming. Second, altering the expression level of many B27 ligands is likely to influence T-cell repertoire selection, which is determined by the avidity of TCR/pMHC interactions and, therefore, by individual epitope amounts at the cell surface. Thus, the autoimmune potential of HLA-B27 might be modulated by ERAP1 polymorphism. Third, HLA-B27

immunogenicity would also be modulated by ERAP1 through its quantitative effect on the peptidome, since T-cell triggering thresholds are dependent on the number of pMHC complexes within individual TCR clusters during T cell activation (40). Fourth, the less active ERAP1 variants may generate suboptimal B27 peptidomes, decreasing the molecular stability of HLA-B27, as observed in WE-I. We have previously shown that high thermostability is a feature of 3 AS-associated subtypes (B*27:02, B*27:04 and B*27:05) expressed on C1R cells that is not shared by the non-AS-associated B*27:06 and B*27:09 subtypes (31). The thermostability of B*27:04 in WE-I was closer to the latter subtypes, indicating that ERAP1 polymorphism protective from AS can alter the molecular stability phenotype of an AS-associated B27 subtype to that shown by non AS-associated ones.

In conclusion, our results reveal large effects of natural ERAP1 polymorphism on the HLA-B27 peptidome *in vivo* and their molecular basis. These effects, through their influence on the immunological and other features of HLA-B27, define the nature of the functional interaction between both molecules and underline the fundamental role of peptides in the pathogenesis of HLA-B27-associated disease.

Acknowledgements. This work was supported by grants SAF2008/00461 and SAF2011/25681 from the Plan Nacional de I+D+i, RD08/0075 (RIER) of the Instituto de Salud Carlos III, and an institutional grant of the Fundación Ramón Areces to the Centro de Biología Molecular *Severo Ochoa.* We thank the staff of the Proteomics facility at the Centro Nacional de Biotecnología, Madrid and Ricardo Ramos (Parque Científico de Madrid) for help in MS and DNA analyses, respectively.

Reference List

1. Benjamin, R. and P. Parham. 1990. Guilt by association: HLA-B27 and ankylosing spondylitis. *Immunol.Today* **11,** 137-142.

2. Colbert, R. A. 2000. HLA-B27 misfolding: a solution to the spondyloarthropathy conundrum? *Mol.Med.Today* **6,** 224-230.

3. Mear, J. P., K. L. Schreiber, C. Münz, X. Zhu, S. Stevanovic, H. G. Rammensee, S. L. Rowland-Jones, and R. A. Colbert. 1999. Misfolding of HLA-B27 as a result of its B pocket suggests a novel mechanism for its role in susceptibility to spondyloarthropathies. *J.Immunol.* **163,** 6665-6670.

4. Dangoria, N. S., M. L. DeLay, D. J. Kingsbury, J. P. Mear, B. Uchanska-Ziegler, A. Ziegler, and R. A. Colbert. 2002. HLA-B27 misfolding is associated with aberrant intermolecular disulfide bond formation (dimerization) in the endoplasmic reticulum. *J.Biol.Chem.* **277,** 23459-23468.

5. Allen, R. L., C. A. O'Callaghan, A. J. McMichael, and P. Bowness. 1999. Cutting edge: HLA-B27 can form a novel β2-microglobulin-free heavy chain homodimer structure. *J.Immunol.* **162,** 5045-5048.

6. Bowness, P. 2002. HLA B27 in health and disease: a double-edged sword? *Rheumatology.(Oxford)* **41,** 857-868.

7. Kollnberger, S., L. Bird, M. Y. Sun, C. Retiere, V. M. Braud, A. McMichael, and P. Bowness. 2002. Cell-surface expression and immune receptor recognition of HLA-B27 homodimers. *Arthritis Rheum.* **46,** 2972-2982.

8. Marcilla, M. and J. A. Lopez de Castro. 2008. Peptides: the cornerstone of HLA-B27 biology and pathogenetic role in spondyloarthritis. *Tissue Antigens* **71,** 495-506.

9. WTCC Consortium. 2007. Association scan of 14,500 nonsynonymous SNPs in four diseases identifies autoimmunity variants. *Nat.Genet.* **39,** 1329-1337.

10. The TASK and WTCCC2 Consortia. 2011. Interaction between ERAP1 and HLA-B27 in ankylosing spondylitis implicates peptide handling in the mechanism for HLA-B27 in disease susceptibility. *Nat.Genet.* **43,** 761-767.

11. Serwold, T., S. Gaw, and N. Shastri. 2001. ER aminopeptidases generate a unique pool of peptides for MHC class I molecules. *Nat.Immunol.* **2,** 644-651.

12. Serwold, T., F. Gonzalez, J. Kim, R. Jacob, and N. Shastri. 2002. ERAAP customizes peptides for MHC class I molecules in the endoplasmic reticulum. *Nature* **419,** 480-483.

13. York, I. A., S. C. Chang, T. Saric, J. A. Keys, J. M. Favreau, A. L. Goldberg, and K. L. Rock. 2002. The ER aminopeptidase ERAP1 enhances or limits antigen presentation by trimming epitopes to 8-9 residues. *Nat.Immunol.* **3,** 1177-1184.

14. Chang, S. C., F. Momburg, N. Bhutani, and A. L. Goldberg. 2005. The ER aminopeptidase, ERAP1, trims precursors to lengths of MHC class I peptides by a "molecular ruler" mechanism. *Proc.Natl.Acad.Sci.U.S.A* **102,** 17107-17112.

15. Nguyen, T. T., S. C. Chang, I. Evnouchidou, I. A. York, C. Zikos, K. L. Rock, A. L. Goldberg, E. Stratikos, and L. J. Stern. 2011. Structural basis for antigenic peptide precursor processing by the endoplasmic reticulum aminopeptidase ERAP1. *Nat.Struct.Mol.Biol.* **18,** 604-613.

16. Kochan, G., T. Krojer, D. Harvey, R. Fischer, L. Chen, M. Vollmar, F. von Delft, K. L. Kavanagh, M. A. Brown, P. Bowness, P. Wordsworth, B. M. Kessler, and U. Oppermann. 2011. Crystal structures of the endoplasmic reticulum aminopeptidase-1 (ERAP1) reveal the molecular basis for N-terminal peptide trimming. *Proc.Natl.Acad.Sci.U.S.A* **108,** 7745-7750.

17. Harvey, D., J. J. Pointon, D. M. Evans, T. Karaderi, C. Farrar, L. H. Appleton, R. D. Sturrock, M. A. Stone, U. Oppermann, M. A. Brown, and B. P. Wordsworth. 2009. Investigating the genetic association between ERAP1 and ankylosing spondylitis. *Hum.Mol.Genet.* **18,** 4204-4212.

18. Zemmour, J., A. M. Little, D. J. Schendel, and P. Parham. 1992. The HLA-A,B "negative" mutant cell line C1R expresses a novel HLA-B35 allele, which also has a point mutation in the translation initiation codon. *J.Immunol.* **148,** 1941-1948.

19. Garcia, F., A. Marina, and J. A. Lopez de Castro. 1997. Lack of carboxyl-terminal tyrosine distinguishes the B*2706-bound peptide repertoire from those of B*2704 and other HLA-B27 subtypes associated to ankylosing spondylitis. *Tissue Antigens* **49,** 215-221.

20. Ellis, S. A., C. Taylor, and A. McMichael. 1982. Recognition of HLA-B27 and related antigens by a monoclonal antibody. *Hum.Immunol.* **5,** 49-59.

21. Paradela, A., M. Garcia-Peydro, J. Vazquez, D. Rognan, and J. A. Lopez de Castro. 1998. The same natural ligand is involved in allorecognition of multiple HLA-B27 subtypes by a single T cell clone: role of peptide and the MHC molecule in alloreactivity. *J.Immunol.* **161,** 5481-5490.

22. Paradela, A., I. Alvarez, M. Garcia-Peydro, L. Sesma, M. Ramos, J. Vazquez, and J. A. Lopez de Castro. 2000. Limited diversity of peptides related to an alloreactive T cell epitope in the HLA-B27-bound peptide repertoire results from restrictions at multiple steps along the processing-loading pathway. *J.Immunol.* **164,** 329-337.

23. Garcia-Medel, N., A. Sanz-Bravo, E. Barnea, A. Admon, and J. A. Lopez de Castro. 2012. The Origin of Proteasome-inhibitor Resistant HLA Class I Peptidomes: a Study With HLA-A*68:01. *Mol.Cell Proteomics.* **11,** M111.

24. Hearn, A., I. A. York, and K. L. Rock. 2009. The specificity of trimming of MHC class I-presented peptides in the endoplasmic reticulum. *J.Immunol.* **183,** 5526-5536.

25. Evnouchidou, I., F. Momburg, A. Papakyriakou, A. Chroni, L. Leondiadis, S. C. Chang, A. L. Goldberg, and E. Stratikos. 2008. The internal sequence of the peptide-substrate determines its N-terminus trimming by ERAP1. *PLoS.ONE.* **3,** e3658.

26. Merino, E., B. Galocha, M. N. Vazquez, and J. A. Lopez de Castro. 2008. Disparate folding and stability of the ankylosing spondylitis-associated HLA-B*1403 and B*2705 proteins. *Arthritis Rheum.* **58,** 3693-3704.

27. Fruci, D., S. Ferracuti, M. Z. Limongi, V. Cunsolo, E. Giorda, R. Fraioli, L. Sibilio, O. Carroll, A. Hattori, P. M. Van Endert, and P. Giacomini. 2006. Expression of endoplasmic reticulum aminopeptidases in EBV-B cell lines from healthy donors and in leukemia/lymphoma, carcinoma, and melanoma cell lines. *J.Immunol.* **176,** 4869-4879.

28. Ramos, M., A. Paradela, M. Vazquez, A. Marina, J. Vazquez, and J. A. Lopez de Castro. 2002. Differential association of HLA-B*2705 and B*2709 to ankylosing spondylitis correlates with limited peptide subsets but not with altered cell surface stability. *J.Biol.Chem.* **277,** 28749-28756.

29. Sesma, L., V. Montserrat, J. R. Lamas, A. Marina, J. Vazquez, and J. A. Lopez de Castro. 2002. The peptide repertoires of HLA-B27 subtypes differentially associated to spondyloarthropathy (B*2704 and B*2706) differ by specific changes at three anchor positions. *J.Biol.Chem.* **277,** 16744-16749.

30. Gomez, P., V. Montserrat, M. Marcilla, A. Paradela, and J. A. López de Castro. 2006. B*2707 differs in peptide specificity from B*2705 and B*2704 as much as from HLA-B27 subtypes not associated to spondyloarthritis. *Eur.J.Immunol.* **36,** 1867-1881.

31. Galocha, B. and J. A. López de Castro. 2010. Mutational Analysis Reveals a Complex Interplay of Peptide Binding and Multiple Biological Features of HLA-B27. *J.Biol.Chem.* **285,** 39180-39190.

32. Purcell, A. W., J. J. Gorman, M. Garcia-Peydro, A. Paradela, S. R. Burrows, G. H. Talbo, N. Laham, C. A. Peh, E. C. Reynolds, J. A. Lopez de Castro, and J. McCluskey. 2001. Quantitative and qualitative influences of tapasin on the class I peptide repertoire. *J.Immunol.* **166,** 1016-1027.

33. Schatz, M. M., B. Peters, N. Akkad, N. Ullrich, A. N. Martinez, O. Carroll, S. Bulik, H. G. Rammensee, P. van Endert, H. G. Holzhutter, S. Tenzer, and H. Schild. 2008. Characterizing the N-terminal processing motif of MHC class I ligands. *J.Immunol.* **180,** 3210-3217.

34. Goto, Y., A. Hattori, Y. Ishii, and M. Tsujimoto. 2006. Reduced activity of the hypertension-associated Lys528Arg mutant of human adipocyte-derived leucine aminopeptidase (A-LAP)/ER-aminopeptidase-1. *FEBS Lett.* **580,** 1833-1838.

35. Evnouchidou, I., R. P. Kamal, S. S. Seregin, Y. Goto, M. Tsujimoto, A. Hattori, P. V. Voulgari, A. A. Drosos, A. Amalfitano, I. A. York, and E. Stratikos. 2011. Cutting Edge: Coding single nucleotide polymorphisms of endoplasmic

reticulum aminopeptidase 1 can affect antigenic peptide generation in vitro by influencing basic enzymatic properties of the enzyme. *J.Immunol.* **186,** 1909-1913.

36. Haroon, N., F. W. Tsui, B. Uchanska-Ziegler, A. Ziegler, and R. D. Inman. 2012. Endoplasmic reticulum aminopeptidase 1 (ERAP1) exhibits functionally significant interaction with HLA-B27 and relates to subtype specificity in ankylosing spondylitis. *Ann.Rheum.Dis.* **71,** 589-595.

37. Maksymowych, W. P., R. D. Inman, D. D. Gladman, J. P. Reeve, A. Pope, and P. Rahman. 2009. Association of a specific ERAP1/ARTS1 haplotype with disease susceptibility in ankylosing spondylitis. *Arthritis Rheum.* **60,** 1317-1323.

38. Szczypiorska, M., A. Sanchez, N. Bartolome, D. Arteta, J. Sanz, E. Brito, P. Fernandez, E. Collantes, A. Martinez, D. Tejedor, M. Artieda, and J. Mulero. 2011. ERAP1 polymorphisms and haplotypes are associated with ankylosing spondylitis susceptibility and functional severity in a Spanish population. *Rheumatology.(Oxford)* **50,** 1969-1975.

39. Saveanu, L., O. Carroll, V. Lindo, M. Del Val, D. Lopez, Y. Lepelletier, F. Greer, L. Schomburg, D. Fruci, G. Niedermann, and P. M. Van Endert. 2005. Concerted peptide trimming by human ERAP1 and ERAP2 aminopeptidase complexes in the endoplasmic reticulum. *Nat.Immunol.* **6,** 689-697.

40. Manz, B. N., B. L. Jackson, R. S. Petit, M. L. Dustin, and J. Groves. 2011. T-cell triggering thresholds are modulated by the number of antigen within individual T-cell receptor clusters. *Proc.Natl.Acad.Sci.U.S.A* **108,** 9089-9094.

41. Ben Dror, L., E. Barnea, I. Beer, M. Mann, and A. Admon. 2010. The HLA-B*2705 peptidome. *Arthritis Rheum.* **62,** 420-429.

Figure legends

Fig. 1. ERAP1 expression among cell lines. *(A)* Western blots were performed for the indicated cell lines using the anti-ERAP1 6H9 mAb and the anti-γ-tubulin GTU88 mAb as internal control. A representative experiment is shown. *(B)* ERAP1 protein expression in the indicated cell lines, as established by Western blot. Data are expressed as the ERAP1/γ-tubulin ratio and are means ± S.D. of 5 experiments. *(C)* Quantitative RT-PCR showing the expression of ERAP1mRNA in the indicated cell lines, relative to C1R04. Data are expressed as Relative Quantity (RQ) values and are means ± S.D. of 3 experiments.

Fig. 2. Flow chart for the automatic pairwise comparison of B*27:04-bound peptidomes. The upper scheme shows the filtering of ion peaks from the MALDI-TOF MS spectra of HPLC fractions from each peptide pool. The following filters were applied. A: signals with m/z ≤ 850 and >1700 are usually abundant in matrix-related and other MHC-I unrelated ion peaks. B: signals with low intensity, since many of them are too close to background. The intensity cutoff was adopted depending on the total number and overall intensity of the ion peaks in each experiment. It was set at 200 for WE-I/JSL and WE-I/C1R04, and at 300 for KNE/C1R04. C: ion peaks with decimal part of the m/z ($M+H^+$) value ranging from 0.0 to 0.2 and from 0.9 to 0.0 were removed based on the extremely low frequency (1 of 1268) of HLA-B27 ligands in these ranges (41). D: matrix-related ion peaks were identified from MS spectra of matrix without peptide. E: The program computed identical ion peaks (±0.2 Da) along consecutive HPLC fractions and selected only the one with maximal intensity. The lower scheme shows the criteria used for the assignment of shared peptides.

Figure 3. Size distribution of B*27:04 ligands as a function of their relative abundance in various ERAP1 contexts. *(A)* WE-I/JSL. The upper histogram shows the Mw distribution of 3222 shared ligands between both cell lines. The other 4 histograms compare the Mw distribution of the peptide sets predominant (IR ≥1.0) in WE-I (dotted line) or JSL (solid line), as well the corresponding subsets showing IR>3.0, IR >1.5-3.0 and IR ≥1.0-1.5 in each cell line; *(B)* WE-I/C1R04. The upper histogram shows the Mw distribution of 5240 shared ligands between both cell lines. The other 4 histograms compare the Mw distribution of the peptide sets predominant in WE-I (dotted line) or C1R04 (solid line), following the same conventions as in panel *A*. *(C)* C1R04/KNE. The upper histogram shows the Mw distribution of 6974 shared ligands between both cell lines. The other 4 histograms compare the Mw distribution of the peptide sets predominant (IR ≥1.0) in C1R04 (solid line) or KNE (dotted line), following the same conventions as in panel *A*.

Fig. 4. Length distribution of B*27:04 ligands as a function of their relative abundance in various ERAP1 contexts. *(A)* WE-I/JSL. The upper histogram shows the length distribution of 3222 shared ligands between both cell lines. The percent of peptides with <9, 9, 10, or >10 residues is indicated. The other 4 histograms compare the length distribution of the peptide sets predominant (IR ≥1.0) in WE-I (white bars) or JSL (black bars), and those of the corresponding subsets showing IR>3.0, IR >1.5-3.0 and IR ≥1.0-1.5 in each cell line. The ratio between the percentage of peptides of the same length in both cell lines (WE-I/JSL ratio) is indicated over the corresponding bars; *(B)* WE-I/C1R04. The upper histogram shows the length distribution of 5240 shared ligands between both cell lines. The other 4 histograms compare the length distribution of the peptide sets predominant in WE-I (white bars) or C1R04 (black bars), following the same conventions as in panel *A*, and showing the WE-I/C1R04 ratios. *(C)*

C1R04/KNE. The upper histogram shows the length distribution of 6974 shared ligands between both cell lines. The other 4 histograms compare the length distribution of the peptide sets predominant (IR ≥1.0) in C1R04 (black bars) or KNE (white bars), following the same conventions as in panel *A* , and showing the C1R04/KNE ratios.

Fig. 5. **Thermostability of HLA-B*27:04 in various ERAP1 contexts.** The indicated cell lines were labeled for 15 min and chased at 0, 2 and 4h. Equal aliquots of the lysates were kept at 4 ºC or heated at the indicated temperatures for 1h prior to immunoprecipitation with ME1, a mAb that recognizes undissociated HLA-B27/peptide complexes, but not unfolded HC, separated by SDS-PAGE, and analyzed by fluorography. The percentage of ME1-reactive HLA/peptide complexes recovered at 0h (♦) , 2h (■) or 4h (▲) chase after heating, was plotted as the intensity of the class I HC at any given temperature (HC_X) relative to that at 4 ºC ($HC_{4ºC}$). The data are means ± SD of 4 (C1R04), 5 (KNE) and 8 (WE-I) experiments. Only 1 experiment could be performed with JSL. The data concerning C1R04, except, for one additional experiment included here, were previously reported (31) and are shown only for comparison.

Table I

*ERAP1 polymorphism in B*27:04 positive cell lines*[a]

	Nucl./AA						
SNP	Position N.	OR (ref.)	Consensus	JSL	C1R	KNE [b]	WE-I
rs26653	380 / 127	1.30 (Harvey et al., 2009)	g / R	g / R	**c / P**	g, **c** / R, **P**	**c / P**
rs26618	828 / 276	0.99 (Harvey et al., 2009)	a / I	a / I	**g / M**	a, **g** / I, **M**	a / I
rs27895	1037 / 346	1.07 (Harvey et al., 2009)	g / G	g / G	g / G	g, **a** / G, **D**	g / G
rs2287987	1045 / 349	0.71 (Brown, 2008)	a / M	a / M	a / M	a / M	**g / V**
rs30187	1583 / 528	1.40 (Brown, 2008)	a / K	a / K	**g /R**	**g /R**	**g /R**
rs10050860	1723 / 575	0.71 (Brown, 2008)	g / D	g / D	g / D	g / D	**a / N**
rs17482078	2174 / 725	0.70 (Brown, 2008)	g / R	g / R	g / R	g / R	**a / Q**
rs27044	2188 / 730	1.40 (Brown, 2008)	c / Q	c / Q	**g /E**	**g /E**	**g /E**

[a]Only non-synonymous changes in the coding strands of ERAP1 are shown. Nucleotide and amino acid residue numbering and consensus sequence are from Human ERAP1 Isoform 2 (Accession N.: Q9NZ08-2). Deviations from the consensus sequence are highlighted in boldface. Polymorphisms associated with increased risk for AS are underlined. All polymorphic positions were determined by SNP typing and confirmed by genomic sequencing. OR: odds ratio.

[b]KNE was heterozygous for rs26653, rs26618 and rs27895, which was confirmed by genomic sequencing.

1. Brown,M.A. (2008). Breakthroughs in genetic studies of ankylosing spondylitis. *Rheumatology. (Oxford),* **47,** 132-137.

2. Harvey,D., Pointon,J.J., Evans,D.M., Karaderi,T., Farrar,C., Appleton,L.H., Sturrock,R.D., Stone,M.A., Oppermann,U., Brown,M.A., and Wordsworth,B.P. (2009). Investigating the genetic association between ERAP1 and ankylosing spondylitis. *Hum. Mol. Genet.*, **18,** 4204-4212.

Table II

*Relative expression and mean size of shared B*27:04 ligands in various ERAP1 contexts*[a]

Comparison	All		IR≥1.0		IR >3.0		IR >1.5-3.0		IR ≥1.0-1.5	
1	N (%)	m/z	N (%)	m/z	N (%)	m/z	N (%)	m/z	N (%)	m/z
WE-I	3222 (100)	1162.0	1729 (54)	1177.6	289 (9)	1197.4	686 (21)	1179.2	754 (23)	1168.6
JSL	3222 (100)	1162.0	1496 (46)	1143.9	260 (8)	1124.3	554 (17)	1130.0	682 (21)	1162.7
Δ m/z				33.7		73.1		49.2		5.9
2										
WE-I	5240 (100)	1154.3	2450 (47)	1170.8	380 (7)	1170.6	888 (17)	1176.7	1182 (23)	1166.2
C1R04	5240 (100)	1154.3	2801 (53)	1154.3	439 (8)	1112.0	1089 (21)	1138.3	1273 (24)	1151.5
Δ m/z				30.6		58.6		38.4		14.6
3										
C1R04	6974 (100)	1149.4	4562 (65)	1151.5	1009 (14)	1141.9	1832 (26)	1156.3	1721 (25)	1152.1
KNE	6974 (100)	1149.4	2421 (35)	1145.5	323 (5)	1133.3	804 (12)	1146.9	1294 (19)	1147.6
Δ m/z				5.0		8.6		9.4		4.5

[a] IR: intensity ratio between the corresponding ion peaks in the MS spectra from the cell lines compared; N: number and % of peptides; m/z: mean value in the peptide set. In MALDI-TOF MS this is equivalent to the mean Mw ($M+H^+$).

Table III

*Relationship between Mw and length among 372 B*27:04 ligands*[a]

	Peptide length							
Mw	<9-mers [b]		9-mers		10-mers		>10-mers[c]	
	N	%	N	%	N	%	N	%
850-900	3	100	0	0	0	0	0	0
900-1000	8	21.1	30	79.0	0	0	0	0
1000-1100	7	5.9	102	86.4	8	6.8	1	0.9
1100-1200	0	0	127	85.8	20	13.5	1	0.7
1200-1300	0	0	24	52.2	19	41.3	3	6.5
1300-1400	0	0	1	8.3	7	58.3	4	33.3
1400-1500	0	0	0	0	0	0	4	100
1500-1600	0	0	0	0	0	0	2	100
1600-1700	0	0	0	0	0	0	1	100
Total	**18**	**4.8**	**284**	**76.3**	**54**	**14.5**	**16**	**4.3**

[a]Percent values are relative to each Mw range, except in the Total line, where the values are relative to the total number of peptide sequences. [b] All these peptides were 8-mers, except one 7-mer. [c]11-mers: 10, 12-mers: 4, 13-mers: 1, 15-mers: 1.

Table IV

*Influence of N-terminal flanking and P1 residues of B*27:04 ligands on ERAP1-mediated trimming*[a]

Subset [b]	Mean Score			
WE- I/JSL				
IR> 3	**P-2**	**P-1**	**P-1+P-2**	**P1**
WE- I	34.2	38.4	72.6	29.9
JSL	14.4	37.9	52.3	12.0
Ratio	**2.4**	**1.0**	**1.4**	**2.5**
IR≥1-3				
WE- I	34.8	44.5	79.2	17.1
JSL	23.3	48.5	71.8	20.4
Ratio	**1.5**	**0.9**	**1.1**	**0.8**
WE- I/C1R04				
IR> 3	**P-2**	**P-1**	**P-1+P-2**	**P1**
WE- I	38.5	46.3	84.7	23.4
C1R04	26.5	36.1	62.6	22.6
Ratio	**1.5**	**1.3**	**1.4**	**1.0**
IR≥1-3				
WE- I	30.3	41.7	72.0	20.6
C1R04	33.8	43.6	77.5	19.5
Ratio	**0.9**	**1.0**	**0.9**	**1.1**
KNE/C1R04				
IR> 3	**P-2**	**P-1**	**P-1+P-2**	**P1**
KNE	33.6	42.9	75.9	19.4
C1R04	30.7	36.1	66.4	23.2
Ratio	**1.1**	**1.2**	**1.1**	**0.8**
IR≥1-3				
KNE	27.3	46.3	73.6	20.7
C1R04	25.9	36.8	62.8	21.7
Ratio	**1.1**	**1.3**	**1.2**	**1.0**

[a]The P-2, P-1 and P1 residues of each peptide were assigned a score (**supplemental Table S1**) related to their susceptibility to ERAP1 trimming *in vivo* (24). The mean score for each individual position, or for P-1+P-2, are shown for each peptide subset. The flanking sequences are shown in **supplemental Table S3.** [b]Peptide subsets whose ion peaks show the indicated intensity ratio (IR) relative to the other cell line.

Table V

*The internal sequence features of B*2704-bound nonamers do not determine their ERAP1-dependent expression*[a]

Peptide subset[b]	Mean P3-P9 Score		
WE- I/JSL	**WE-I**	**JSL**	**Ratio**
IR>3	2.3	2.4	1.0
IR≥1-3	2.3	2.4	1.0
WE- I/C1R04	**WE-I**	**C1R04**	
IR>3	2.7	2.4	1.1
IR≥1-3	2.6	2.6	1.0
KNE/C1R04	**KNE**	**C1R04**	
IR>3	2.2	2.1	1.0
IR≥1-3	2.5	2.5	1.0

[a]Each residue at a given position was assigned a score (**supplemental Table S1**) related to its effect on ERAP1-mediated trimming (25). The mean added score of positions P3 to P9 for each peptide subset, as well as the corresponding ratios, are given (see **supplemental Table S4** for full details). [b]Peptide subsets whose ion peaks show the indicated intensity ratio (IR) relative to the other cell line.

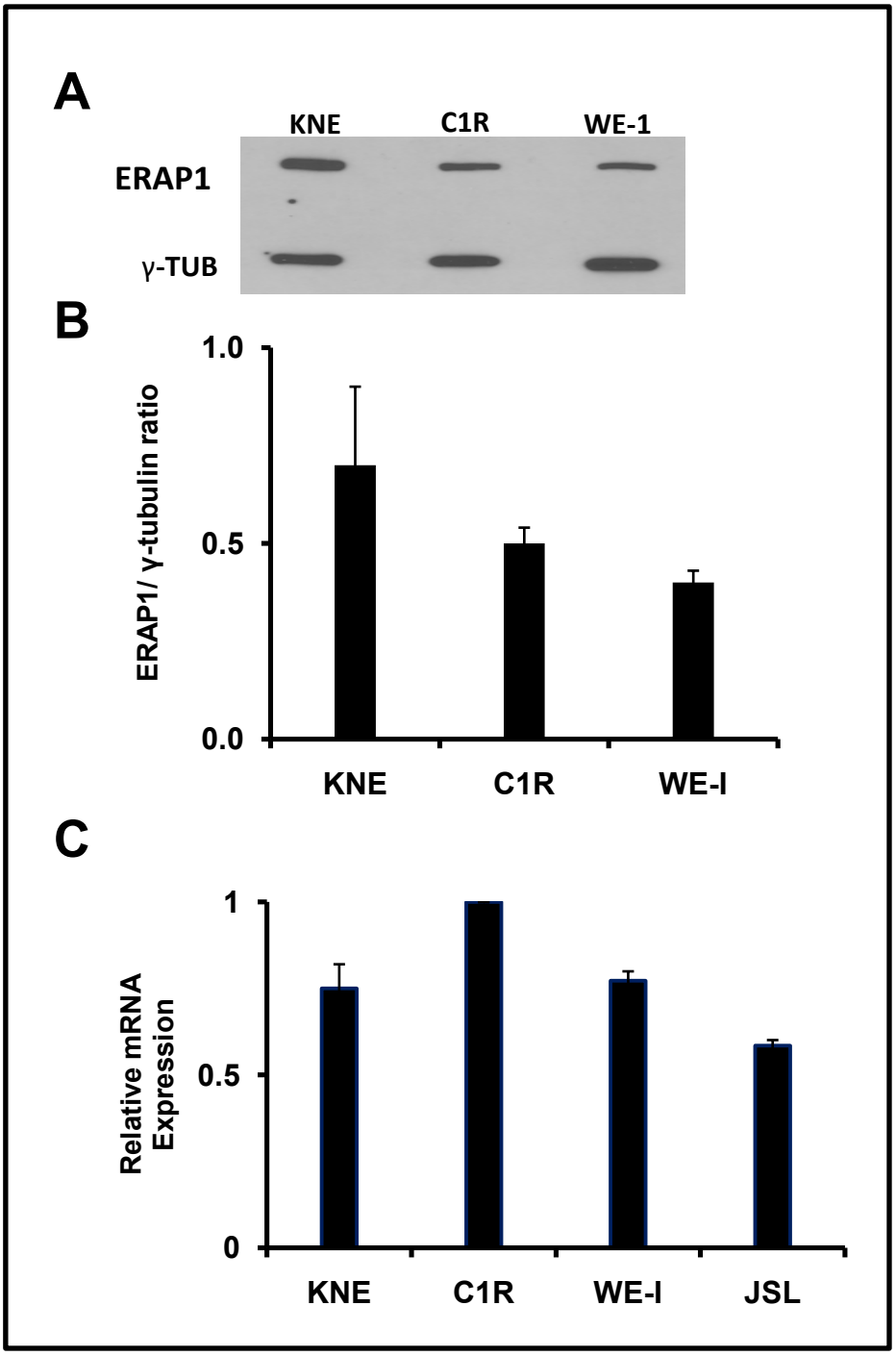

A
KNE
C1R
WE-1
ERAP1
γ-TUB
B
1.0
0.5
0.0
ERAP1/ γ-tubulin ratio
KNE
C1R
WE-I
C
1
0.5
0
Relative mRNA Expression
KNE
C1R
WE-I
JSL

Figure 1

Ion peak filtering

All ion peaks with m/z 800-2000 and s/n > 3

↓

A. m/z range selection

Removal of signals outside the m/z 850-1700 range

↓

B. Intensity threshold selection

Removal of signals with intensity value below a given cut off

↓

C. Selection of peptide-related signals

Removal of signals with m/z decimal part close to 0

↓

D. Removal of matrix-related signals

Removal of the ion peaks matching an exclusion list obtained from matrix alone

↓

E. Removal of redundant signals

For each ion peak, its intensity along consecutive HPLC fractions was computed and only the signal showing maximal intensity was taken into account

Assignment of shared peptides

Fitered signals from each peptide pool are matched according to identity of m/z (± 0.2) and chromatographic retention time (± 1 HPLC fraction).

When ambiguities arose the ion peak pair with an intensity ratio closest to 1 was selected.

Figure 2

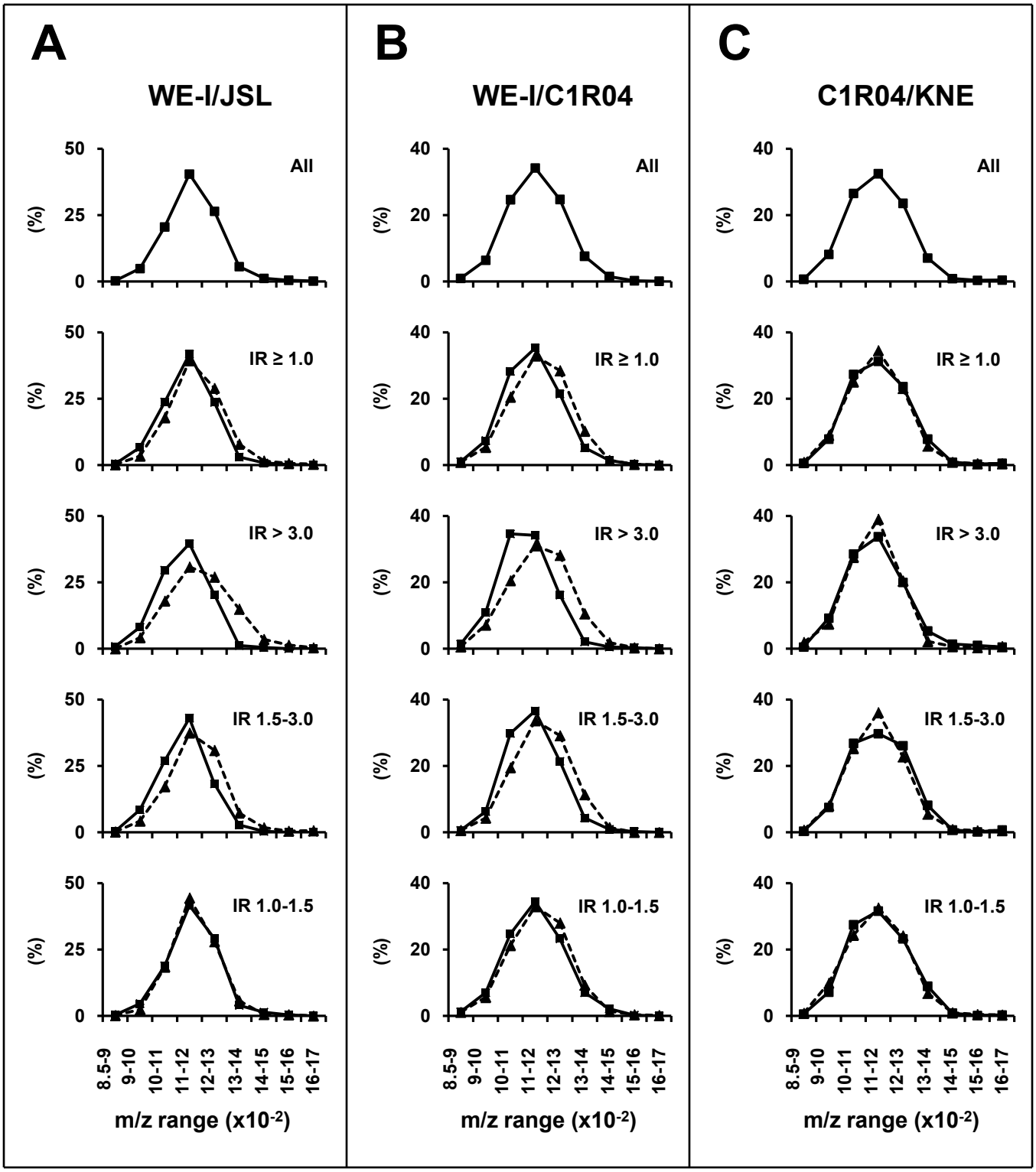

A
WE-I/JSL
B
WE-I/C1R04
C
C1R04/KNE
All
IR ≥ 1.0
IR > 3.0
IR 1.5-3.0
IR 1.0-1.5
(%)
8.5-9
9-10
10-11
11-12
12-13
13-14
14-15
15-16
16-17
m/z range (x10-2)

Figure 3

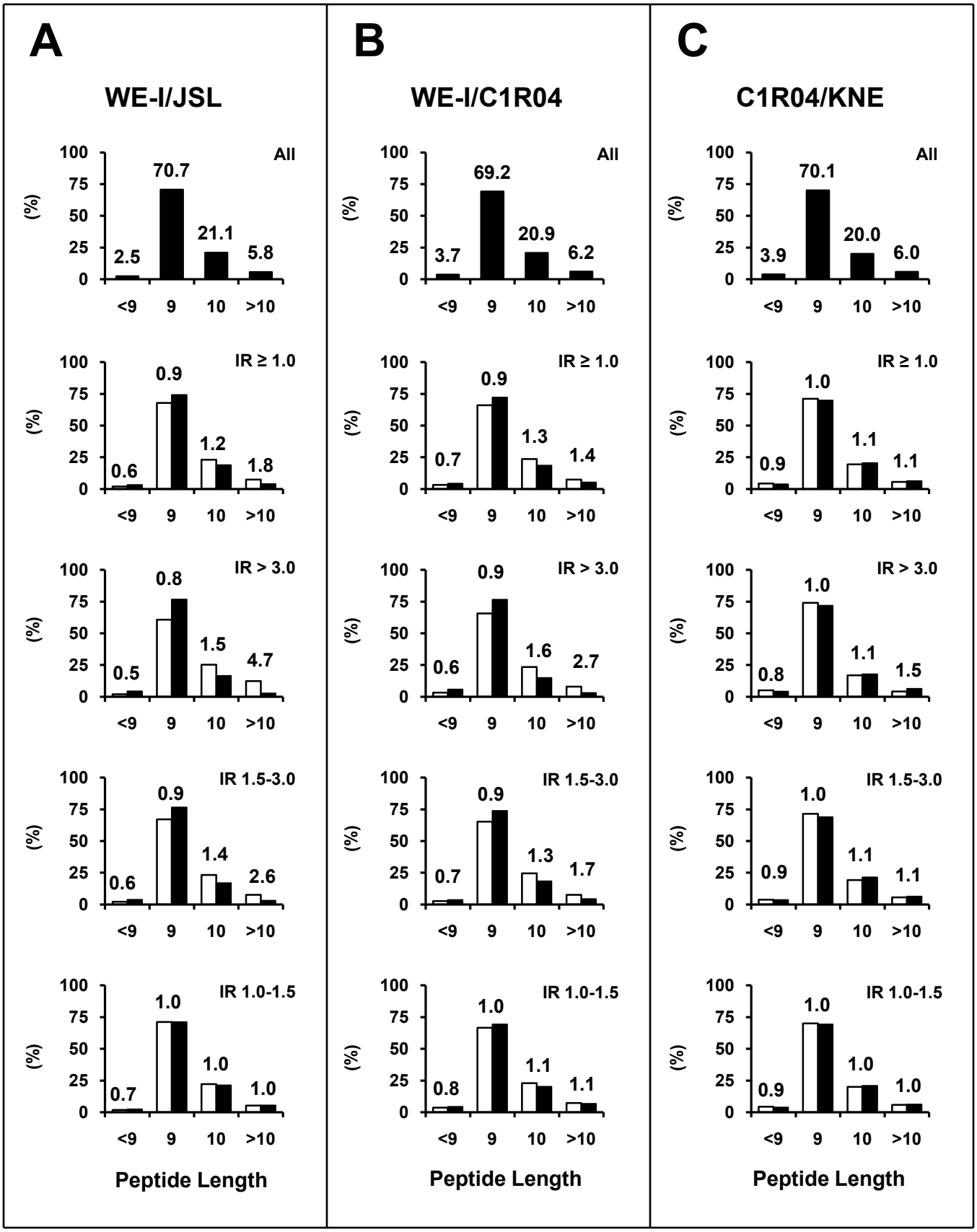

A
WE-I/JSL
B
WE-I/C1R04
C
C1R04/KNE
All
IR ≥ 1.0
IR > 3.0
IR 1.5-3.0
IR 1.0-1.5
(%)
100
75
50
25
0
<9
9
10
>10
2.5
70.7
21.1
5.8
0.6
0.9
1.2
1.8
0.5
0.8
1.5
4.7
0.6
0.9
1.4
2.6
0.7
1.0
1.0
1.0
3.7
69.2
20.9
6.2
0.7
0.9
1.3
1.4
0.6
0.9
1.6
2.7
0.7
0.9
1.3
1.7
0.8
1.0
1.1
1.1
3.9
70.1
20.0
6.0
0.9
1.0
1.1
1.1
0.8
1.0
1.1
1.5
0.9
1.0
1.1
1.1
0.9
1.0
1.0
1.0
Peptide Length

Figure 4

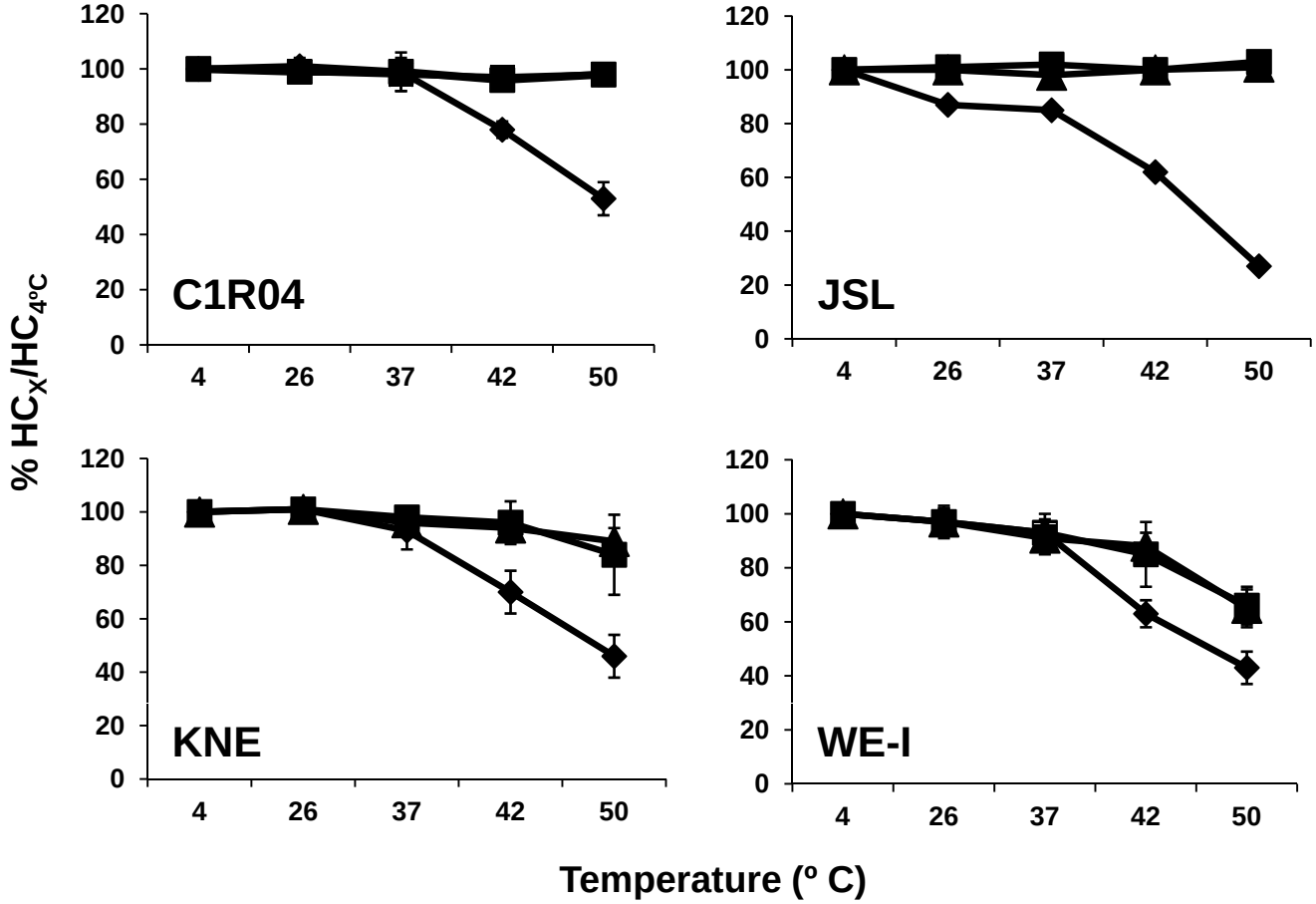

% HCX/HC4°C
120
100
80
60
40
20
0
4
26
37
42
50
C1R04
JSL
KNE
WE-I
Temperature (º C)

Figure 5

Table S1. Score systems for amino acid residues

Trimming susceptibility scores for N-terminal flanking and P1 residues[a]		*Internal peptide sequence scores*[b]						
Residue	***Score***	**P3**	**P4**	**P5**	**P6**	**P7**	**P8**	**P9**
P	0	2.5	0	0	4.0	0	0.5	1.0
R	4					2.5		8.5
K	4	4.0	2.5	2.0	2.0	1.0	5.5	
V	5	4.0	2.0	0		0	1.0	6.0
D	5							
W	8							
E	8	1.5	0	0	0.5	0	0	
N	13							
T	19	4.0	0.5		4.0	0	1.0	
G	20	4.0	0	0	4.0	0	1.0	
H	20							
Q	22	5.0	1.5	0	2.5	0	1.0	1.0
S	32							
F	34	2.0	3.5	2.5	7.0	2.5	0.5	6.0
I	47							
C	55							
A	76						1.0	
L	88							
M	91	6.0	1.5	0	8.0		1.5	4.0
Y	100							

[a]Based on data from Hearn et al. (2009), J.Immunol. 183: 5526-5536.

[b]Based on data from Evnouchidou et al. (2008), PLoS One 3 (11): e3658. Empty cells mean that the corresponding residues at those positions were not scored.

Table S2. Automated comparison of B*27:04-bound peptidomes expressed in various ERAP1 contexts.

Comparison	Ion peaks	Filter[a]					Comparison	
1	Unfiltered	A	B	C	D	E	Shared	Other
WE-I	14326	12990	12292	9666	9343	5760	3222	2538
JSL	13507	12230	11750	8885	8541	5072	3222	1850
2								
WE-I	54185	49716	17262	15146	14771	8504	5240	3264
C1R04	57118	51534	18120	16476	15860	9988	5240	4748
3								
C1R04	53612	47760	26236	25492	24773	13597	6974	6223
KNE	49854	42069	20354	19161	18501	10005	6974	3031

[a]Figures indicate the number of ion peaks remaining after applying the following filters: A, Mw cutoff ≤850, >1700; B, intensity cutoff 200 or, for C1R04/KNE , 300; C, removal of ions with decimal m/z value in the ranges 0.0-0.2 and 0.9-0.0; D, removal of matrix-related ion peaks; E, removal of redundancies: among ion peaks with identical (±0.2) m/z eluting in consecutive HPLC fractions only the one showing the maximal intensity was selected.

Table S3. Amino acid sequence of 372 B*27:04 ligands.[a]

Flanking Seq.	*Peptide Seq.*	*A. N.*		*Intensity Ratio*		
				C1R04/WE-I	*JSL/WE-I*	*C1R04/KNE*
TW	RRQNKVF	Q5NDL2	7	*	0.3	*
LY	ARFKPVAV	Q8WVK7	8	*	*	0.3
VV	GRYPDPNF	Q92466	8	*	1.9	*
VS	HRRPPIGY	Q9H6H4	8	*	*	2.4
YG	IRKGNLLF	O15205	8	*	*	0.2
NF	KRFHGRAF	Q92598	8	0.3	*	*
NH	KRYHPIYI	P78317	8	*	34.1	0.2
EE	QRYQALKI	O75410	8	*	*	5.2
TL	RRAVPLAL	Q13200	8	*	3.0	*
DD	RRGRPVGF	P61978	8	*	5.9	*
RG	RRLGLALG	Q15293	8	*	1.7	*
EK	RRNEPIKY	Q9UHA3	8	20.5	2.7	1.1
VR	RRVLRPVL	Q9H9V9	8	*	0.5	2.3
LN	RRYLSPKY	P60842	8	*	*	0.1
IF	SRFRPISV	Q9Y4B6	8	*	0.2	*
VG	SRGFPLRL	Q9UI30	8	*	*	8.3
QH	SRIWPLYL	Q9HCS7	8	*	*	5.4
DT	SRVQPIKL	P62857	8	*	3.0	*
AG	ARARFALGL	O95336	9	0.1	*	*
FI	ARFFKQPTF	P17252	9	0.3	1.4	0.5
TD	ARFGHGSDY	Q9NWH9	9	*	*	4.7
FL	ARFKERVGY	Q96AT1	9	0.3	0.2	0.3
LA	ARFLFTTGF	O00507	9	0.1	0.7	0.3
NR	ARFRGRKGY	Q86WW8	9	*	0.1	0.7
EA	ARHEHQVML	Q15233	9	*	*	0.2
RL	ARHGGVKRI	Q99525	9	*	*	1.4
AR	ARIAKGLRL	P47914	9	0.0	0.3	1.8
FE	ARIALLPLL	Q9P0J0	9	1.7	*	*
CT	ARIEIGLHY	O43325	9	0.4	*	0.3
KY	ARILFPVTF	P48169	9	0.4	*	*
EL	ARKERQLHM	Q9NQB0	9	*	0.2	0.2
RG	ARKGRGLVV	P83876	9	*	*	0.0
DF	ARKHFVLLL	Q9BPX3	9	0.1	0.0	1.6
SM	ARKIGARVY	P13796	9	0.4	*	*
AI	ARLKEVLEY	P14324	9	4.1	*	3.9
YK	ARLPLRLFL	O00767	9	0.4	0.5	5.0
M	ARLQTALLV	O00626	9	0.5	1.1	0.7
AR	ARTPRLLLL	Q6PCB8	9	*	*	6.9
QG	ARVPSSLAY	Q9NQS7	9	*	*	7.6
KY	ARVVQKLGF	P20226	9	0.1	*	*

SL	ARYGKSPYL	P50395	9	*	1.0	0.3
EM	ARYKRKLLI	Q14493	9	*	1.1	*
AA	ERLQYVFGY	Q96MG7	9	2.7	*	*
EM	ERTLALLAF	Q9NWU2	9	*	*	5.3
RQ	FRFDPQFAL	P01909	9	*	*	9.2
SY	FRLDTPLYF	Q32P28	9	10.5	*	12.4
SR	FRQQRPLEF	Q15773	9	0.2	0.6	0.5
PT	FRRPKTLRL	P62750	9	3.0	*	*
NS	FRYNGLIHR	P46779	9	2.7	*	*
TL	FRYQGHVGA	P40306	9	2.6	*	*
ML	FRYQGYIGA	Q99436	9	3.1	*	6.3
VC	GRAFIFPSY	Q8TBZ8	9	0.3	0.3	0.5
KR	GRAPQVLVL	Q9NR30	9	0.4	*	1.6
QN	GRDTHMRQI	Q9UM13	9	*	*	3.8
QA	GRFARGLRL	Q13393	9	0.0	*	*
SR	GRFGGVITI	O43315	9	*	1.6	*
ER	GRFHGGNLF	P55789	9	*	*	4.7
AT	GRFKDILAL	Q8N1W1	9	*	*	0.1
AT	GRFKDVLVL	Q92974	9	2.9	*	0.2
LT	GRFSHRIYV	P46977	9	*	*	5.0
ML	GRHGVFLEL	Q13200	9	2.6	0.8	1.2
TQ	GRHSTPLHL	Q9H2K2	9	*	*	0.3
KL	GRHVRALVL	P22314	9	0.3	*	*
RP	GRIDRKIEF	P62195	9	0.6	*	0.1
NW	GRIVTIFAF	Q16548	9	*	0.2	*
NW	GRIVTLISF	Q07820	9	1.1	*	0.9
AL	GRKKNGEEY	Q8WV28	9	*	*	0.1
VF	GRKKTATAV	P62249	9	*	*	0.1
PS	GRLCLLTIV	Q96DB9	9	*	*	0.6
GV	GRLGHILEY	Q8IXQ5	9	4.2	14.5	7.5
TS	GRLKDILAI	Q6ZSZ5	9	*	*	3.9
CP	GRLRGVRAV	P49207	9	0.3	*	0.1
RR	GRLTKHTKF	Q9Y3U8	9	2.2	3.6	0.7
AL	GRNKPHTPF	O00139	9	0.5	*	0.3
SR	GRNRQPLVL	Q9NTI7	9	0.2	*	0.1
KF	GRNSNICHY	Q96CG3	9	18.5	*	18.2
DD	GRRHGFLIL	Q10570	9	45.3	*	*
RY	GRRPYGVGL	P25786	9	*	*	3.6
NS	GRRYGLIGL	Q9UG63	9	*	*	9.1
NL	GRSFHILAI	Q96EE3	9	3.8	*	*
RL	GRSRWQGEL	Q12789	9	*	*	4.3
KM	GRTKIFIRF	O00159	9	*	1.2	*
CY	GRVAPRSGL	P33316	9	*	0.4	*
LD	GRWFGGRQI	O43719	9	*	*	4.4
TY	GRWQYPLIY	Q9BX95	9	*	*	0.3

GC	GRYAGGQGY	O75592	9	*	0.5	*
SQ	GRYEVAVPL	Q07866	9	*	*	3.2
LN	GRYFGGRVV	Q96I25	9	2.3	*	*
RA	GRYGGFLLA	P18858	9	*	1.3	*
AY	GRYNPRRSL	P33993	9	*	0.7	*
QQ	GRYPGVSNY	Q13564	9	2.3	4.5	2.1
ED	GRYRDPTTV	P63151	9	0.4	*	19.1
VS	GRYRGSVHF	Q9C0C9	9	*	*	0.5
LA	GRYSGRKAV	P61353	9	*	1.1	0.3
GI	GRYSRSAMY	Q02878	9	*	8.5	*
VT	GRYTILIKY	P21333	9	6.3	*	*
RG	HRHGKPFVF	O43187	9	3.1	*	*
RA	HRIGQQNEV	P51531	9	*	*	6.2
RA	HRIGQTKTV	O60264	9	*	*	9.9
WA	HRLFPKLQF	Q9BVW5	9	5.1	*	1.9
KD	HRNKAMINL	Q9NZ45	9	*	*	1.7
NI	IRAGLIPKF	P52292	9	0.4	3.3	0.4
NY	IRCIPSQAV	P16260	9	6.8	*	14.5
VT	IRKSKNILF	Q13765	9	1.2	*	2.6
EV	IRNDGVLLL	O60763	9	*	*	12.0
QS	IRSSLLLGF	P42704	9	0.6	1.5	4.4
SF	IRTRLLLQF	Q96K31	9	*	1.2	*
LR	IRYGQTKKM	P49591	9	*	1.5	*
TL	KQYPRKNLF	O60667	9	*	23.5	*
YA	KRASVFVKL	Q8IZP2	9	0.2	*	*
SL	KRFDDKYTL	Q15005	9	*	*	2.5
HL	KRFEHSAKL	Q9UPT9	9	0.1	1.1	0.0
LF	KRFEKQKEV	Q9Y6A5	9	*	*	0.1
EC	KRFGKAYNL	Q86XP3	9	*	*	0.2
DP	KRFPHGIPF	P17050	9	2.2	*	*
PG	KRFRPMDNL	Q14683	9	*	1.4	*
KD	KRFSGTVRL	P62906	9	0.7	*	0.2
CK	KRFSHSGSY	P37275	9	*	*	0.1
DM	KRHRIQFKY	P11388	9	0.2	*	0.1
TH	KRILKFLKL	P54619	9	0.2	*	*
TH	KRILKFLQL	Q9UGJ0	9	0.2	*	*
DA	KRINWRTVL	P54368	9	*	*	4.0
DK	KRKGQVIQF	Q969Q0	9	2.6	7.9	1.3
VF	KRKKPQLRL	Q5EE01	9	0.8	*	*
AF	KRLGASLAF	P78527	9	*	*	0.2
RG	KRLGILVVF	Q6UWH6	9	*	0.4	3.8
EQ	KRLGLLAGF	P55061	9	9.4	*	5.3
EQ	KRLKTVLEL	Q14444	9	*	*	0.3
LS	KRNPRQINW	P83731	9	*	*	3.4
KG	KRRDDGLSA	P84101	9	*	*	6.1

GY	KRRGGVPSL	Q9Y6G1	9	*	*	0.1
QR	KRRKALPTF	Q00653	9	*	0.5	*
DA	KRRQQGSSL	Q14161	9	*	*	0.3
QT	KRSSRNPVF	Q969W3	9	*	*	0.3
AK	KRTNIIPVI	P22087	9	0.1	*	1.1
AK	KRTNIIPVL	A6NHQ2	9	0.1	*	1.1
NI	KRYHIAKVY	P12081	9	0.5	*	*
NS	KRYKEAQKI	O43242	9	*	*	8.1
TE	KRYKSIVKY	P14324	9	2.0	1.0	1.2
DT	KRYQEALHL	O00231	9	*	*	3.5
MA	LRAGLVLGF	P06340	9	*	*	0.2
TC	LRFPGQLNA	Q9BVA1	9	2.2	1.8	1.3
HH	LRHGGRMQY	P49643	9	*	*	13.7
VN	NRFKGVKYF	Q6V1X1	9	*	*	6.1
VC	NRYLVVLYY	Q9Y3B4	9	3.3	*	*
SI	QRFPKILVL	O75604	9	12.4	*	18.3
FY	QRFPLSFGF	P17050	9	*	1.5	*
RG	QRKGAGSVF	P62917	9	*	0.4	*
AD	QRKKAYADF	P09669	9	3.5	*	1.5
FL	QRLDLQIKL	Q8IV50	9	0.3	*	0.5
MA	QRLVRVLQI	Q9NSP4	9	0.2	0.1	*
QI	QRNDYVHAL	Q9NR22	9	2.8	*	4.1
SF	QRNLFILAY	O95573	9	*	*	0.6
QV	QRSRFIVVV	Q9BRQ8	9	*	0.7	*
FS	QRYAKVPLF	Q9BQ52	9	0.5	0.1	*
NG	RRAFIGIGF	Q9NVZ3	9	*	8.9	*
LA	RRAIYQATY	P28074	9	*	0.5	0.4
VR	RRAKGSVSL	Q13115	9	*	*	0.1
LR	RRARGPPAA	Q05923	9	*	*	0.2
M	RRDVRILLL	Q8IXI1	9	14.3	*	*
DV	RRESRLLQL	P82930	9	5.4	*	*
GL	RRFAFPLSL	Q13144	9	*	0.8	*
RE	RRFEKPLEE	Q8NC51	9	*	2.1	*
YL	RRFFKAKKL	P49755	9	0.7	0.4	*
LY	RRFFKSPHF	Q8NEG7	9	0.1	2.5	*
YS	RRFFPYYVY	P20618	9	1.1	2.3	1.0
QL	RRFGDKLNF	Q13794	9	1.1	4.2	0.4
PA	RRFGGPVHH	Q15717	9	*	0.9	0.2
VR	RRFKGQILM	P62910	9	0.3	0.2	1.2
AL	RRFMQTFVL	Q13283	9	*	*	1.4
GF	RRFRASPLF	Q2NL82	9	*	1.6	*
TL	RRFRGDVTL	Q96F44	9	*	*	0.1
EK	RRFRKFLVY	P50395	9	*	*	0.1
RS	RRFRPRSEF	Q14999	9	0.1	0.7	0.2

VK	RRFSDFLGL	Q13596	9	*	4.6	*
VY	RRFTEIYEF	A6NI72	9	*	4.1	1.4
CL	RRFTRPEHL	A6NNY8	9	1.6	0.3	0.6
KA	RRFWGKIVA	P55040	9	*	*	5.5
YS	RRFWQGDTF	Q9BUX1	9	*	*	13.1
RA	RRGLPRLAV	Q96AZ6	9	0.1	0.9	*
PA	RRGPFPLAL	P49755	9	*	2.2	*
HS	RRGPPPRRY	Q01130	9	*	0.5	8.8
PK	RRGRPPSKF	Q9UIF9	9	0.1	*	*
VC	RRHFPPAVI	O43164	9	*	*	0.3
EI	RRHWGGNVL	P62424	9	*	*	1.9
RD	RRISGVDRY	O15239	9	*	*	5.2
HE	RRISQIQQL	O60306	9	*	2.6	*
DR	RRIYLILEY	Q96GD4	9	2.1	*	1.9
SY	RRKDGVFLY	P62829	9	1.4	*	4.5
EG	RRKEQERSF	P37198	9	*	*	4.7
EE	RRKERQEAL	Q13111	9	*	0.5	*
CG	RRKGEQIYY	O95711	9	2.0	0.3	0.2
MD	RRKKALTDY	Q5TA31	9	3.6	*	*
PK	RRKKPPIQY	Q96JP5	9	0.5	1.8	*
EL	RRKQFHVLL	Q6I9Y2	9	0.2	*	*
EE	RRKQLQNYL	Q8TEQ0	9	*	*	0.2
QC	RRLDRSAHF	O15294	9	*	*	0.3
EW	RRLGVQQSL	P33552	9	*	0.9	*
NS	RRLPIFSRL	Q07352	9	0.5	10.6	1.9
AW	RRLPQAFRP	P13598	9	*	*	3.9
NR	RRLSELLRY	P08238	9	0.4	6.6	0.7
DT	RRLSFLVSY	P47897	9	*	3.1	*
VR	RRLSLFLNV	Q99836	9	*	*	0.1
HE	RRNDKIIVF	P19447	9	*	*	3.0
GI	RRNEKIAVH	P62913	9	2.6	*	*
WE	RRNKVLNHF	P35269	9	*	1.0	*
ES	RRNREIAQL	Q7Z4S6	9	*	*	0.3
EL	RRQQEQLRY	Q14151	9	*	*	0.2
PQ	RRRDIQQTL	Q9HAV4	9	*	*	0.2
EE	RRREKAEEL	Q04323	9	*	*	0.1
RV	RRRFKGQIL	P62910	9	2.1	*	*
QL	RRRGKTNHY	Q13291	9	*	0.3	*
TY	RRRLSYNTA	P49207	9	0.3	0.4	8.8
YR	RRRPENPKP	P67809	9	0.2	0.1	3.4
EM	RRRREEEGL	P52294	9	*	*	0.2
EE	RRRRSFQDY	Q9NP64	9	*	*	0.3
GD	RRRSNTLDI	O94915	9	*	*	3.3
CC	RRRSSVAFF	O15297	9	*	0.5	*
HF	RRSGQHPRY	Q9NWZ8	9	4.0	0.6	0.3

HL	RRSPFLQVF	Q15437	9	1.0	1.6	0.5
NS	RRSQPLISL	Q9H967	9	9.2	*	*
HL	RRSSFLQVF	Q15436	9	*	2.1	*
SD	RRSSIPITV	P33992	9	0.3	*	*
QV	RRSSSVAQV	O15205	9	*	3.3	*
LD	RRTGYLKGY	Q9Y5S9	9	6.9	*	*
SL	RRTRLILFV	Q96TA2	9	*	2.2	*
EA	RRVARQAQL	O60239	9	*	7.2	0.0
LC	RRVLVQVSY	Q00266	9	*	*	0.1
AE	RRWEKSSRY	Q9HCG8	9	*	*	0.3
LA	RRYGDVFQI	Q16678	9	*	18.5	0.1
TG	RRYGRRHAY	O43164	9	*	1.0	*
TA	RRYIGIVKQ	P52597	9	0.5	1.1	0.2
LN	RRYLSPKYI	P60842	9	0.6	0.9	0.7
VY	RRYNDFVVF	Q9Y5X2	9	*	4.6	*
PF	RRYNGGVGR	P18621	9	*	1.1	*
HL	RRYNIIPVL	Q9UI12	9	0.3	1.1	0.4
LR	RRYPQIYTT	Q6PL18	9	*	*	3.0
EI	RRYQKSTEL	Q6NXT2	9	0.6	0.4	0.8
RT	RRYVRKFVL	O43633	9	0.1	0.5	0.7
RS	SRAGLQFPV	P0C0S8	9	2.6	0.3	0.2
NV	SRAKAVRAL	Q9H009	9	4.3	*	42.7
VL	SRANSLFAF	P61009	9	4.2	*	*
NS	SRAPVFLQF	Q4ZIN3	9	3.7	4.8	2.1
ST	SRFARLQKL	P04035	9	*	*	6.7
QA	SRFDIPLGL	O94913	9	0.3	*	1.7
NS	SRFGKFIQL	Q96H55	9	12.0	*	*
NS	SRFGKFIQV	B2RTY4	9	0.7	1.4	1.3
NS	SRFGKFIRI	P35749	9	0.5	1.0	1.0
NS	SRFGKFVEI	Q9UM54	9	13.0	*	*
KK	SRFGNAFHL	Q6P2Q9	9	*	*	0.1
RG	SRFIKFQEM	P33993	9	*	*	14.6
KF	SRFPEALRL	Q13200	9	7.7	*	2.3
AF	SRFPSFLKA	Q9BTD8	9	2.9	*	*
DL	SRFRGFLEH	Q9P241	9	*	3.4	*
IF	SRFRPISVF	Q9Y4B6	9	0.1	0.7	0.3
FC	SRHSKALKL	Q92974	9	0.4	*	0.2
RA	SRIRKLFNL	P62753	9	0.5	0.2	1.4
QN	SRLGLPLLL	Q9BSV6	9	0.5	*	*
VT	SRLGSVFPF	Q13217	9	*	5.5	*
EA	SRLKSILKL	Q9BRT6	9	0.1	*	0.5
AG	SRLPRQLFL	Q9BU23	9	3.0	*	5.1
YL	SRLRNQSVF	P37268	9	2.8	0.1	6.0
SL	SRMIRKMKL	Q07020	9	0.1	*	*

YH	SRMQRGGLR	Q9BTL4	9	*	0.8	*
NS	SRNLFVLGF	Q9UHI7	9	0.9	*	4.9
HF	SRNPKVLNP	Q9ULX3	9	3.8	*	*
SS	SRRAGGEHY	Q5T4S7	9	*	*	6.5
ET	SRRKAGHQF	Q96EY7	9	*	*	7.8
NY	SRTGRHLAF	O15213	9	2.8	2.0	0.3
AI	SRTPVLMNF	Q9H3K2	9	*	*	18.2
KS	SRTSVQPTF	Q5T4S7	9	0.3	0.9	0.3
LH	SRVKLILEY	Q7L5N1	9	0.2	0.4	0.2
TT	SRVLKVLSV	P35222	9	10.9	0.1	1.2
SF	SRVPLFPVF	Q9H6F2	9	0.9	2.0	5.1
YR	SRVRPCVVY	O15523	9	0.2	0.2	*
VA	SRYFKGPEL	P68400	9	0.6	*	1.1
RC	SRYLYTLVI	P63173	9	0.1	*	1.6
RR	SRYQDILVF	P19623	9	0.5	*	0.7
VY	SRYRPQYGY	Q8IYM9	9	0.1	0.1	*
LS	SRYSQKQHY	Q8WUB2	9	*	*	3.7
FK	SRYTGINQF	O75417	9	4.1	*	*
LQ	SRYYRAPEI	Q9H2X6	9	26.3	0.1	*
IC	SRYYRAPEL	P49840	9	26.3	0.1	*
NK	TRGRFELAF	O14925	9	0.4	*	0.8
FS	TRLFAVLRF	P46977	9	4.0	0.1	2.2
NL	TRMRHVISY	O14949	9	0.8	*	0.4
FR	TRSAIILHL	P53350	9	*	0.8	0.3
CL	TRTGLFLRF	Q9NX02	9	3.0	*	1.0
AW	TRVDTILEF	O14980	9	*	0.3	*
DA	TRVYLILEY	O14965	9	1.1	0.3	0.6
EY	TRYLFALQL	O94887	9	0.1	0.6	2.4
HR	TRYPTILQL	Q8WW24	9	0.1	*	*
RI	TRYQGVNLY	P11940	9	2.3	*	*
TD	VRAQIRLQF	Q92878	9	*	7.3	0.1
YY	VRAVLHLLL	P13498	9	0.3	*	*
SR	VRFGQQKRY	Q68E01	9	*	*	4.3
SH	VRFLGNLVL	Q7L523	9	*	*	3.5
EV	VRLPYPLQM	Q9NPA0	9	*	*	6.7
MA	WRALHPLLL	O95944	9	*	*	4.9
DE	WRFKPIEQL	Q96DE5	9	*	8.1	*
KD	WRSGKPVRV	Q96T88	9	20.1	*	8.0
YY	YRYPTGESY	Q16875	9	*	*	3.3
SD	ARLQYGLLIF	Q96BI3	10	0.2	*	*
AF	ARNKEIVHTF	Q8NI36	10	*	0.1	*
SL	ARYGKSPYLY	P31150	10	2.3	0.3	8.3
EC	GRAFSRKSLL	Q9BS31	10	0.2	*	*
RG	GRFGRKGVAI	Q14240	10	*	0.5	*

ER	GRFHGGNLFF	P55789	10	0.8	0.9	5.4
VT	GRFNGQFKTY	P63220	10	1.1	*	5.0
ST	GRIAHIPLYF	Q9NRG9	10	0.8	*	*
YT	GRIKAIQLEY	O43242	10	0.7	*	7.0
LF	GRKTGQAPGY	P99999	10	0.3	*	2.9
VR	GRWPGSSLYY	Q14739	10	4.5	*	9.4
TQ	GRWPKKSAEF	P18621	10	*	*	5.7
RY	GRYGGETKVY	Q16629	10	*	0.6	*
RS	GRYGRKGVAI	P38919	10	0.4	*	*
AG	GRYLGKKVQF	P55199	10	0.1	*	*
ML	HRFYGKNSSY	Q13283	10	0.3	0.9	0.6
TG	HRWLKGGVVL	P35613	10	*	*	3.6
SD	HRYGDGGSTF	P31943	10	8.8	2.1	1.3
KP	HRYRPGTVAL	Q6NXT2	10	0.8	1.2	0.5
EA	IQRTPKIQVY	P61769	10	*	0.1	*
II	KRFEQKGFRL	O60361	10	0.1	*	*
RL	KRFGPNVPAL	Q96SB8	10	0.2	*	0.8
TW	KRFSVPVQHF	Q99829	10	0.1	*	1.0
SN	KRHRPIGIGV	P23921	10	0.5	0.2	*
RY	KRNENQLIIF	Q15393	10	0.6	*	*
TD	KRNGVIIAGY	Q9UKF6	10	10.3	*	1.4
FL	KRNPGVKEGY	P22234	10	2.6	*	11.2
AL	KRQGRTLYGF	P62805	10	*	*	3.9
	MRLGSPGLLF	O75144	10	*	*	5.6
TI	RRDAPAGRKV	P62917	10	19.0	*	*
YN	RRFVNVVPTF	P62861	10	0.1	*	*
VL	RRHALIVQGF	Q9Y5Q9	10	*	0.5	*
RD	RRISGVDRYY	O15239	10	1.1	*	*
TA	RRKDAKSVKI	P63173	10	1.2	*	8.3
SY	RRKDGVFLYF	P62829	10	0.7	0.4	10.4
EK	RRKDHSFLGF	O75563	10	*	*	3.1
RL	RRLALFPGVA	P30101	10	1.7	1.7	3.4
ML	RRLGVQPSKY	Q9HAY2	10	0.5	*	1.6
AW	RRLSRDSGGY	Q14761	10	3.0	*	5.1
TL	RRNGLGQLGF	Q9P2F8	10	*	1.3	*
TW	RRNGPLPLKY	P35226	10	1.0	*	14.1
QT	RRRAAKNVSY	O14647	10	*	0.1	*
QE	RRRHYNGEAY	P31689	10	*	0.1	0.3
AL	RRSINQPVAF	Q9GZT3	10	1.2	*	2.2
DN	RRTGNKGLRL	Q14807	10	*	*	5.5
GR	RRYDRKQSGY	Q969Q0	10	1.3	0.5	1.0
IM	RRYGEQPASY	Q53H80	10	*	1.8	*
IM	RRYGTRPTSY	Q9H9L7	10	*	*	0.2
MM	RRYQDAIRVF	Q9Y262	10	0.9	*	4.9

EI	RRYQKSTELL	Q16695	10	2.1	*	6.1
ET	RRYRPTKNYL	O75934	10	0.3	0.1	0.8
SN	SRLEDQVIGV	Q2KHR3	10	*	*	11.0
AK	SRSSRAGLQF	Q6FI13	10	2.8	0.3	4.0
AA	WRRLPQAFRP	P13598	10	4.1	*	*
NR	GRFSSTTGLFY	P51817	11	0.5	0.4	10.8
EL	KRFNADNKLLL	Q9NRZ9	11	0.3	*	*
EK	KRYDREFLLGF	Q04637	11	0.9	0.1	1.7
DS	LRFGPNGGLVF	Q9UHW5	11	6.5	*	6.7
NE	LRVAPEEHPVL	P60709	11	*	0.2	*
KE	RRRTESINSAF	O96004	11	*	*	0.1
WL	RRYLENGKETL	P10319	11	3.4	*	13.8
EI	RRYQKSTELLI	Q71DI3	11	0.7	*	3.1
ET	SRFRQQRPLEF	Q15773	11	0.0	1.1	*
M	SRSVALAVLAL	P61769	11	2.5	*	20.3
RS	GRYGRKGVAINF	P38919	12	0.0	*	3.2
YN	RRFVNVVPTFGK	P62861	12	0.3	*	*
MC	RRKSSGGKGGSY	P01889	12	0.0	0.2	0.2
EI	RRYQKSTELLIR	Q16695	12	*	0.1	23.2
ES	RRFSPIPFPPLSY	Q6P2Q9	13	*	0.3	*
TF	YNELRVAPEEHPVL	P60709	15	*	0.0	*

[a]The amino acid sequence of B*27:04 ligands, their N-terminal flanking sequences, the accession number (A.N.) of the corresponding parental proteins, the peptide length and the intensity ratio in the cell line pairs where they were detected are indicated. Asterisks indicate that the corresponding peptide was not detected in that comparison. In the few cases where two similar peptide sequences involving isobaric residues were compatible with a same MS/MS spectrum, both possibilities were taken into account. All the sequences were determined by MALDI-TOF/TOF MS/MS form $[M+H]^+$ parental ions.

Table S4. Influence of internal sequence positions of B*27:04-bound nonamers on ERAP1-mediated trimming [a]

Subset [b]									
WE-I/JSL	**N[c]**	**Peptide Position**							
IR>3		**P3**	**P4**	**P5**	**P6**	**P7**	**P8**	**P9**	**Mean P3-P9**
WE- I	26	3.2	1.4	1.0	3.5	0.7	1.0	5.1	2.3
JSL	23	2.6	0.6	0.5	5.1	0.7	1.4	6.0	2.4
Ratio		**1.2**	**2.4**	**1.9**	**0.7**	**1.0**	**0.7**	**0.9**	**1.0**
IR≥1-3									
WE- I	34	3.3	0.9	0.8	3.6	0.5	1.1	6.0	2.3
JSL	34	2.7	0.9	0.9	5.0	0.4	1.5	5.7	2.4
Ratio		**1.2**	**1.0**	**1.0**	**0.7**	**1.5**	**0.8**	**1.1**	**1.0**
WE-I/C1R04									
IR>3		**P3**	**P4**	**P5**	**P6**	**P7**	**P8**	**P9**	**Mean P3-P9**
WE-I	42	3.4	1.9	1.1	4.1	0.7	2.4	5.4	2.7
C1R04	37	2.6	1.0	0.7	5.0	0.9	1.1	5.6	2.4
Ratio		**1.3**	**2.0**	**1.5**	**0.8**	**0.8**	2.2	1.0	**1.1**
IR≥1-3									
WE-I	35	3.2	1.2	0.9	4.2	0.8	1.8	5.9	2.6
C1R04	32	3.0	1.2	0.6	3.9	1.2	1.9	6.3	2.6
Ratio		**1.1**	**1.0**	**1.4**	**1.1**	**0.7**	**0.9**	**0.9**	**1.0**
KNE/C1R04									
IR>3		**P3**	**P4**	**P5**	**P6**	**P7**	**P8**	**P9**	**Mean P3-P9**
KNE	58	2.9	1.0	0.7	3.6	0.5	1.1	5.9	2.2
C1R04	64	2.8	0.7	0.3	3.7	0.9	1.2	5.4	2.1
Ratio		**1.0**	**1.6**	**2.2**	**1.0**	**0.6**	**0.9**	**1.1**	**1.0**
IR≥1-3									
KNE	32	3.1	1.6	1.0	3.6	0.3	1.6	6.0	2.5
C1R04	33	3.0	0.7	0.9	5.0	1.2	0.9	6.0	2.5
Ratio		**1.0**	**2.3**	**1.2**	**0.7**	**0.3**	**1.8**	**1.0**	**1.0**

[a]Each residue at a given position was assigned a score related to its effect on ERAP1-mediated trimming (**supplemental Table S1**). The mean score of each position for each peptide subset, as well as the corresponding ratios are given. [b]Peptide subsets whose ion peaks show the indicated intensity ratio (IR) relative to the other cell line. [c]Number of nonamers sequenced.

Apéndice III

Otros trabajos publicados durante el desarrollo de esta Tesis.

Endogenous Processing and Presentation of T-cell Epitopes from *Chlamydia trachomatis* with Relevance in HLA-B27-associated Reactive Arthritis*[S]

Juan J. Cragnolini, Noel García-Medel, and José A. López de Castro‡

Chlamydia trachomatis* triggers reactive arthritis, a spondyloarthropathy linked to the human major histocompatibility complex molecule HLA-B27, through an unknown mechanism that might involve molecular mimicry between chlamydial and self-derived HLA-B27 ligands. *Chlamydia*-specific $CD8^+$ T-cells are found in reactive arthritis patients, but the immunogenic epitopes are unknown. A previous screening of the chlamydial genome for putative HLA-B27 ligands predicted multiple peptides that were recognized *in vitro* by $CD8^+$ T-lymphocytes from patients. Here stable transfectants expressing bacterial fusion proteins in human cells were generated to investigate the endogenous processing and presentation by HLA-B27 of two such epitopes through comparative immunoproteomics of HLA-B27-bound peptide repertoires. A predicted T-cell epitope, from the CT610 gene product, was presented by HLA-B27. This is, to our knowledge, the first endogenously processed epitope involved in HLA-B27-restricted responses against *C. trachomatis* in reactive arthritis. A second predicted epitope, from the CT634 gene product, was not detected. Instead a non-predicted nonamer from the same protein was identified. Both bacterial peptides showed very high homology with human sequences containing the HLA-B27 binding motif. Thus, expression and intracellular processing of chlamydial proteins into human cells allowed us to identify two bacterial HLA-B27 ligands, including the first endogenous T-cell epitope from *C. trachomatis* involved in spondyloarthropathy. That human proteins contain sequences mimicking chlamydial T-cell epitopes suggests a basis for an autoimmune component of *Chlamydia*-induced HLA-B27-associated disease. *Molecular & Cellular Proteomics 8:1850–1859, 2009.

Chlamydia trachomatis is an obligate intracellular parasite that infects the urogenital epithelium. It is a very common pathogen and one frequently inducing reactive arthritis (ReA)[1] (1). Multiple strategies, including down-regulation of major histocompatibility complex (MHC) class I and class II expression (2–4) and persistence, have been developed by the bacteria to evade the immune system. Yet both $CD4^+$ and $CD8^+$ T-cell responses are activated upon infection (5). In particular, HLA-B27-restricted $CD8^+$ T-lymphocytes are found in patients with *Chlamydia*-induced ReA (6, 7). The role of these cells in the pathogenesis and evolution of ReA to chronic disease is probably mediated by IFN-γ. Secretion of this cytokine is critical for the protective immunity against *Chlamydia* (8) because it inhibits the bacterial growth (9). However, this is often insufficient to promote complete clearance of *C. trachomatis*, and actually IFN-γ-induced depletion of the tryptophan pool induces the differentiation of the metabolically active reticular bodies to persistent forms (10), which sustain chronic infection and ReA. The high prevalence of HLA-B27 among patients with *Chlamydia*-induced ReA (11), especially in its chronic form, suggests a pathogenetic mechanism based on interactive effects of the bacteria and HLA-B27 that seems unrelated to the capacity of *C. trachomatis* to infect or replicate into HLA-B27-positive cells (12). One such mechanism could be T-cell-mediated autoimmunity elicited by molecular/antigenic mimicry between chlamydial and self-derived HLA-B27 ligands. Antigenic mimicry between chlamydial and homologous α-myosin-derived peptides is crucial to inducing autoimmune myocarditis in mice (13). Breakdown of cytotoxic T-lymphocyte (CTL) tolerance to HLA-B27 was observed in transgenic rats upon exposure to *C. trachomatis* (14). Cross-reactivity between HLA-B27-restricted *Chlamydia*-specific CTL with self-derived HLA-B27 epitopes has not been reported. However, a biochemical basis for it was provided by the finding of an endogenously processed and presented peptide from the DNA primase of

From the Centro de Biología Molecular Severo Ochoa (Consejo Superior de Investigaciones Científicas and Universidad Autónoma de Madrid), Universidad Autónoma, 28049 Madrid, Spain

Received, February 25, 2009, and in revised form, May 6, 2009
Published, MCP Papers in Press, May 13, 2009, DOI 10.1074/mcp.M900107-MCP200

[1] The abbreviations used are: ReA, reactive arthritis; MHC, major histocompatibility complex; CTL, cytotoxic T-lymphocyte; GFP, green fluorescent protein; NQRA, Na^+-translocating NADH-quinone reductase subunit A; PqqC, pyrroloquinoline-quinone synthase-like protein; htMDM, high throughput mass data manager; s/n, signal-to-noise ratio; IFN, interferon; fr., fraction; E3, ubiquitin-protein isopeptide ligase; TCR, T-cell antigen receptor.

C. trachomatis with high homology to a self-derived HLA-B27 ligand (15, 16).

Because of the likely involvement of HLA-B27 in the pathogenesis of chronically evolving ReA, the role of $CD8^+$ T-cell responses in the protective immunity against *C. trachomatis* and the presence of HLA-B27-restricted T-cells in patients with *Chlamydia*-induced ReA, the identification of relevant chlamydial epitopes becomes crucial to establish the pathogenetic mechanism of this disease. Unfortunately a direct analysis of chlamydial HLA-B27 ligands expressed on infected cells is exceedingly difficult because of their extremely low amounts, which challenge even the most sensitive techniques of MS. In the case of *Chlamydia*, the situation is further complicated by the down-regulation of MHC class I expression shortly after infection (3, 4). To our knowledge, only one MHC class I ligand was recently identified, in the mouse system, from *Chlamydia muridarum*-infected cells using state-of-the-art MS techniques (17). Due in part to this difficulty, alternative approaches, such as expression cloning and synthetic peptide epitope mapping (18, 19) or MHC class I tetramer arrays (20), have been used to identify MHC class I-restricted chlamydial T-cell epitopes in mice. In a previous study (6) predictive algorithms were used to screen the whole genome of *C. trachomatis* for nonamer peptide sequences containing the HLA-B*2705 binding motif and a high probability of being generated by proteasomal cleavage. This led to identifying multiple sequences that, when used as synthetic peptides *in vitro*, stimulated $CD8^+$ T-cells from patients with *Chlamydia*-induced ReA. Such cells could also be detected in the synovial fluid of these patients using HLA-B27 tetramers complexed to some of these peptides (7).

Although these strategies identify chlamydial sequences that are recognized by $CD8^+$ T-cells they do not prove that these peptides are the endogenously processed epitopes that activated the natural T-cell responses to the bacteria *in vivo*. Because of the intrinsic cross-reactivity of T-cells (21, 22), it is conceivable that synthetic peptides recognized *in vitro* may be different from the natural epitopes generated by endogenous processing of the chlamydial proteins that elicit the HLA-B27-restricted T-cell responses in ReA patients. To investigate this issue we focused on two predicted epitopes (6). Stable transfectants expressing the corresponding chlamydial proteins fused to green fluorescent protein (GFP) were generated in a B*2705-positive cell line. The endogenous processing and presentation of the predicted epitopes or other peptides from the same bacterial protein were analyzed by comparative immunoproteomics analysis of the B*2705-bound peptide repertoires from transfected and untransfected cells and sequencing of peptides differentially presented on the bacterial protein transfectant.

EXPERIMENTAL PROCEDURES

Bacterial Gene Constructs and Transfectants—GFP bacterial fusion proteins were produced fusing the cDNA coding for the Na^+-translocating NADH-quinone reductase subunit A (NQRA) or pyrroloquinoline-quinone synthase-like protein (PqqC) (CT634 and CT610 gene products, respectively) of *C. trachomatis* serovar L2 (Advanced Biotechnologies, Columbia, MD) to the 3′-end of the GFP gene. Both cDNAs were amplified by PCR using the following primers: 5′-TCTCTCTCGAATTCTATGAAAATAGTTGTTTCTCGCGGA and 3′-TCTCTCTCGGATCCTAACGAGGAGGTTACCACATT for NQRA and 5′-CTCTCTCTAGATCTATGATGGAGGTGTTTATGAATTTT and 3′-CTCTCTCTGTCGACATAAGATTGATGACAACTACAACA for PqqC. The complete sequences were cloned into the pGFP-C1 vector (BD Biosciences Clontech) in the EcoRI and BamHI sites and in the BglII and SalI sites, respectively.

Stable transfectants were generated and characterized as previously described (16). Briefly the human HLA-A,B-deficient lymphoid cell line C1R (23) was electroporated with the GFP or GFP-chlamydial gene constructs and the RSV5 vector carrying the hygromycin resistance gene. Fluorescent hygromycin-resistant cells were selected by flow cytometry, and the expression of the fusion protein was analyzed by Western blot.

Isolation of HLA-B27-bound Peptides—B*2705-bound peptides were isolated from 1–2 $\times$ 10^{10} C1R-B*2705 cells and C1R-B*2705 transfectants expressing chlamydial proteins as described previously (24). Briefly cells were lysed in 1% Igepal CA-630 (Sigma), 20 mM Tris/HCl buffer, 150 mM NaCl, pH 7.5 in the presence of a mixture of protease inhibitors. HLA-B27-peptide complexes were isolated by affinity chromatography of the soluble fraction with the W6/32 monoclonal antibody (IgG2a, specific for a monomorphic HLA-A,B,C determinant) (25). HLA-B27-bound peptides were eluted at room temperature with 0.1% aqueous TFA, concentrated with Centricon 3 (Amicon, Beverly, MA), and fractionated by HPLC as described previously (26).

MALDI-TOF MS Analysis—HPLC fractions were analyzed in a MALDI-TOF/TOF mass spectrometer (4800 Proteomics Analyzer, Applied Biosystems, Foster City, CA). The samples were dried down with a SpeedVac system (Savant Global Medical Instrumentation, Ramsey, MN), reconstituted in 0.6 μl of TA buffer (33% aqueous acetonitrile, 0.1% TFA), loaded onto an Opti-TOF™ 384-well MALDI insert (Applied Biosystems), and allowed to dry at room temperature. Then 0.6 μl of matrix solution (α-cyano-4-hydroxycinnamic acid (Bruker Daltonics, Bremen, Germany)) at 3 mg/ml was added and allowed to dry at room temperature. MS data were acquired in the mass range 800–2000 Da in reflector positive mode at 25 kV and analyzed using the Data Explorer software, version 4.9 (Applied Biosystems). Each spectrum was externally calibrated with the Peptide Calibration Standard Mixture (Bruker Daltonics, product number 206195) to reach a typical mass measurement accuracy of $<$25 ppm.

Automated Comparative Analysis of MALDI-TOF MS Data—A systematic comparison of the MALDI-TOF MS spectra of the HPLC fractions from HLA-B27-bound peptide pools was carried out with an algorithm programmed in Visual Basic 6.0 and implemented as a macro within Microsoft© Excel, designated as high throughput mass data manager (htMDM). This tool recognizes the raw numeric data list from each individual mass spectrum from the MALDI-TOF/TOF mass spectrometer software. First, for comparison of distinct peptide pools, the ion peaks with a signal-to-noise ratio (s/n) smaller than 3-fold, relative to the average s/n of the MS spectrum, were removed. Second, ion peaks with a s/n below 5% of the maximal s/n observed in the MS spectrum were also ignored. Third, an exclusion list of matrix-related ion peaks was generated from direct MS analysis of the matrix alone. The ion peaks whose *m/z* was identical ($\pm$0.4) to those in this list were excluded. For comparison of MS spectra corresponding to correlative ($\pm$1) HPLC fractions from two peptide pools, ion peaks were considered to reflect identical peptides when their observed *m/z* were within a user-defined range, which depends on the

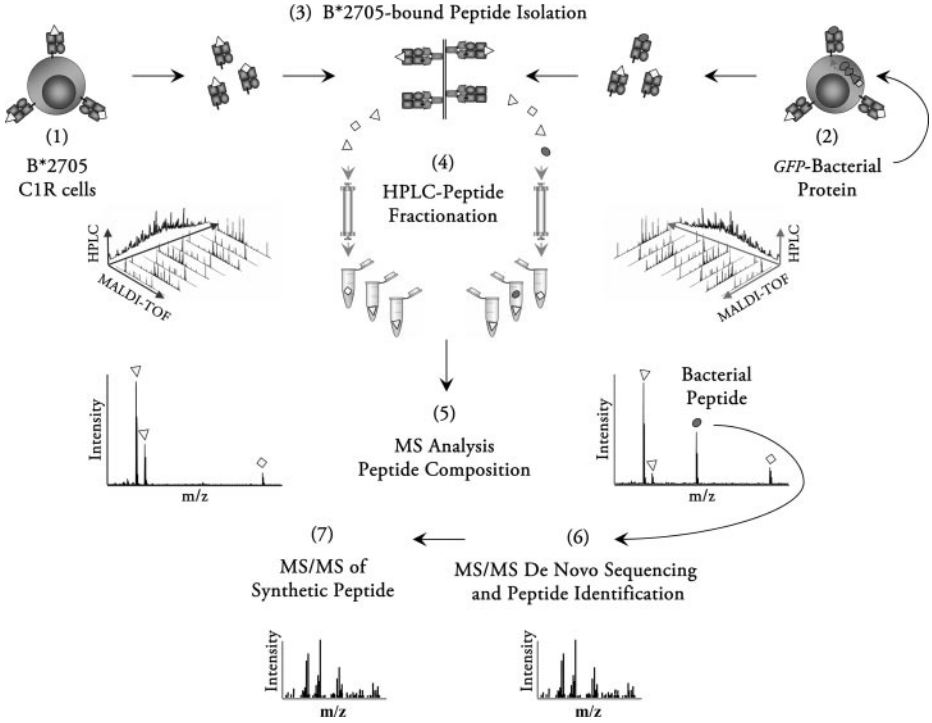


FIG. 1. **Summary of the strategy used to identify endogenously processed HLA-B27 ligands from *C. trachomatis*.** The B*2705-bound peptide repertoires were isolated in parallel experiments from about 1–2 $\times$ 10^{10} C1R transfectant cells expressing (*2*) or not expressing (*1*) a GFP bacterial fusion protein by immunopurification of HLA-B27 and acid extraction (*3*). The peptide pools were fractionated by HPLC under identical conditions and in consecutive runs (*4*). Each individual HPLC fraction was analyzed by MALDI-TOF MS, and the corresponding MS spectra of correlative fractions were systematically compared to identify peptides found only in the bacterial protein transfectant (*5*). The amino acid sequence of the differentially expressed peptides was determined by MS/MS fragmentation analysis (*6*). The corresponding peptide was chemically synthesized, and the MS/MS spectrum of the synthetic peptide was compared with that of the endogenous ligand (*7*).

accuracy of the instrument used. In our study this range was set at $\pm$0.4.

The htMDM tool also allows searching for specific ion peaks, such as candidate peptide epitopes or peptides of a particular size. A user-defined inclusion list with the *m/z* values of interest can be stored and used to automatically screen all the MS spectra for the presence of ion peaks with identical ($\pm$0.4) *m/z* for their quick identification. The htMDM output consists of a listing of filtered *m/z* values with the corresponding retention time or HPLC fraction number. Matches with the exclusion and inclusion directories are marked as well as shared and specific *m/z* signals from each of the peptide repertoires compared. To verify the performance of the htMDM tool, both automatic and manual comparisons were carried out.

Peptide Sequencing—MALDI-TOF-TOF MS/MS fragmentation spectra were acquired with Data Explorer, version 4.9, at 1 kV using collision-induced dissociation with atmospheric air and a precursor mass window of $\pm$2.5 Da. A s/n of 10 was used for processing data. Interpretation of the MS/MS spectra was done manually but was assisted by various tools as follows. Manual inspection of the spectrum usually allowed us to derive a tentative sequence. This was used to screen the chlamydial protein sequences NQRA and PqqC (UniProtKB/Swiss-Prot accession numbers O84639 and O84616, respectively) for a possible match using the Mascot software, version 2.2 (Matrix Science Inc, Boston, MA). The MS-product tool (version 5.1.8) (University of California San Francisco), which generates a list of theoretical fragment ions from a given peptide sequence, was then used to match the candidate sequences from the chlamydial proteins to our experimental MS/MS spectra. The assigned sequences were formally confirmed by comparing the experimental MS/MS spectra with those of the corresponding synthetic peptides.

Homology Searches—The sequence homology between chlamydial peptides and human proteins was analyzed in the non-redundant protein database using the Blastp software (National Center for Biotechnology Information, Bethesda, MD). The most recent versions of the database (Release 33, January 22, 2009) and search algorithm available were used.

RESULTS

Identification of Endogenously Processed HLA-B27-restricted T-cell Epitopes from C. trachomatis in Human Cells—The following strategy (Fig. 1) was originally developed in a previous study from our laboratory (16) and was here adapted to investigate the endogenous processing and presentation of HLA-B27 ligands from *C. trachomatis* recognized by $CD8^+$ T-cells from ReA patients (6). 1) B*2705-C1R cells were used as recipients to generate stable transfectants expressing bacterial fusion proteins. 2) To achieve efficient expression of chlamydial proteins in human cells, fusion constructs were generated in which the GFP coding sequence was placed at the 5′-end of the corresponding chlamydial gene. 3) B*2705-bound peptide repertoires were isolated from large amounts of cells expressing or not expressing the bacterial protein. 4) Both peptide pools were fractionated by HPLC in consecutive runs and identical conditions to minimize alterations in the peptide elution patterns. 5) The MALDI-TOF MS spectrum of every HPLC fraction from each peptide pool was obtained and compared with the correlative (*i*), previous ($i - 1$), and following ($i + 1$) fractions of the other pool to look for peptides found only in the bacterial transfectant. 6) The amino acid sequences of the differential peptides were determined from the corresponding MS/MS spectra. 7) The assigned sequences were validated by comparison with the MS/MS spectra of the corresponding synthetic peptides. If this analysis failed to reveal the presence of the predicted T-cell epitope, a blind MS/MS fragmentation analysis was carried out in the HPLC fraction corresponding to the ex-

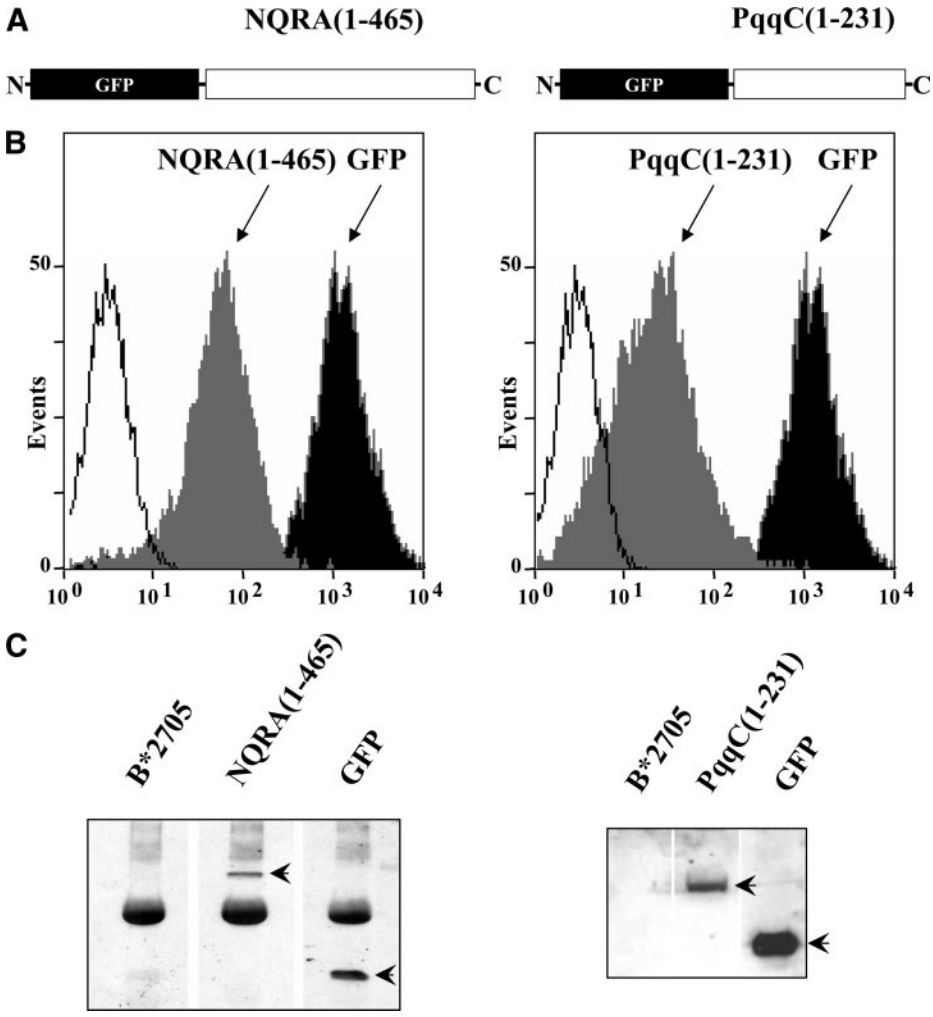


FIG. 2. **Expression of chlamydial fusion proteins in B*2705-C1R cells.** *A*, schematic structure of the GFP-NQRA and GFP-PqqC fusion proteins. *B*, flow cytometry analysis showing the GFP-associated fluorescence of the NQRA (*left panel*) and PqqC (*right panel*) fusion protein transfectants. Untransfected B*2705-C1R cells (*white* histogram) or cells transfected with GFP alone (*black* histogram) were included as controls. *C*, Western blot analysis showing the expression of the NQRA (*left*) and PqqC (*right*) fusion proteins in the respective transfectant cells. The immunoblot was done on the material immunoprecipitated with anti-GFP antibody (NQRA) or on whole lysates (PqqC). In the former case, the prominent nonspecific band corresponds to the heavy chain of the anti-GFP antibody used for immunoprecipitation. In both cases the fusion protein and GFP are noted with *upper* and *lower arrows*, respectively.

pected retention time of the peptide epitope and in neighbor fractions using a narrow gate corresponding to the *m/z* of the peptide.

*Expression of Chlamydial Fusion Proteins in B*2705-C1R Cells*—We focused on the CT634 and CT610 genes, coding for the NQRA and PqqC proteins, respectively (27), which contain sequences recognized *in vitro* as exogenously added synthetic peptides by $CD8^+$ T-cells from ReA patients (6). In addition, PqqC stimulated IFN-γ secretion by peripheral blood mononuclear cells from *Chlamydia*-infected patients with urogenital infection (28). NQRA is a polypeptide of 465 residues involved in Na^+ transport from the cytoplasm to the periplasm. PqqC is a polypeptide of 231 residues that is expressed in the native state as a homodimer. It is a *Chlamydia*-specific toxin that is secreted into the host cytoplasm and induces apoptosis in mammalian cell lines (29, 30). Both proteins are natural components of elementary bodies (31). GFP-chlamydial gene constructs (Fig. 2*A*) were transfected into B*2705-C1R cells. The stable expression of the fusion proteins was established by flow cytometry (Fig. 2*B*) and Western blot analyses (Fig. 2*C*).

A Single HLA-B27 Chlamydial Ligand Differentially Detected in NQRA Transfectants Is Unrelated to the Predicted T-cell Epitope—HLA-B27-bound peptide pools were isolated from B*2705-C1R cells, B*2705-C1R transfected with GFP alone, and B*2705-C1R expressing the NQRA fusion protein. A total of 4895 ion peaks from the NQRA transfectant were compared with 4958 and 4526 ion peaks from untransfected B*2705-C1R and GFP transfectants, respectively. Only one ion peak from the NQRA transfectant, with *m/z* 1054.7 in HPLC fr. 208, was absent from B*2705-C1R cells (Fig. 3*A*) and from the transfectant expressing GFP alone (not shown). The amino acid sequence of this peptide was determined by MALDI-TOF/TOF MS/MS and assigned as a nonamer, KRALLEIVI, spanning residues 86–94 of the chlamydial NQRA protein. The assignment was confirmed by comparison of the MS/MS spectrum with that of the corresponding synthetic peptide (Fig. 3*B*). The NQRA-(86–94) peptide was unrelated to the predicted T-cell epitope from this protein spanning residues 330–338 (MRDHTITLL) (6). The endogenously processed peptide was not predicted in that report.

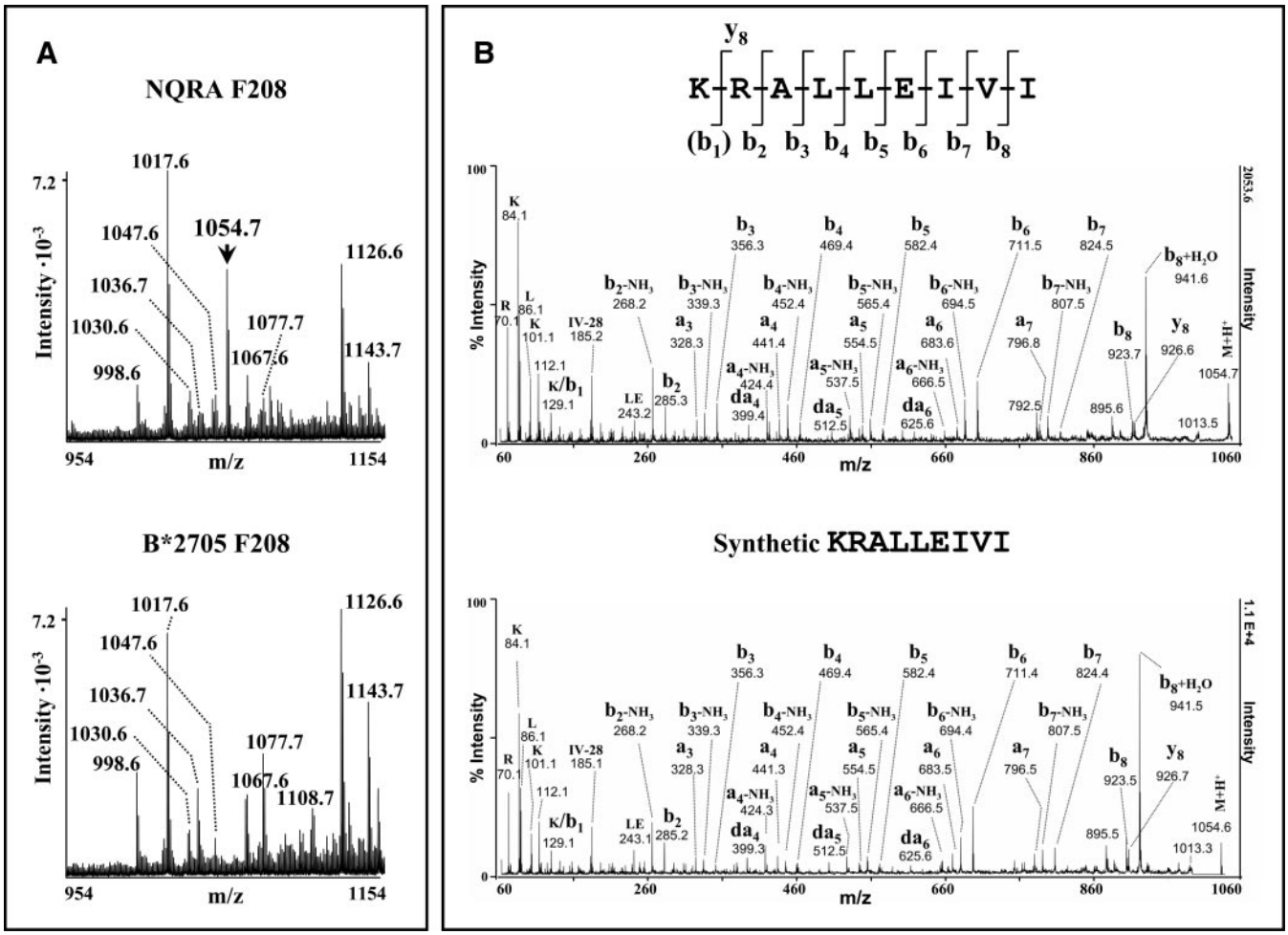


FIG. 3. **A single endogenous HLA-B27 ligand identified from the chlamydial NQRA protein is unrelated to the predicted T-cell epitope.** *A*, MALDI-TOF MS spectra of fr. 208 from the HLA-B27-bound peptide pools isolated from the NQRA transfectant (*above*) and untransfected B*2705-C1R cells (*below*) showing a single ion peak, at *m/z* 1054.7, differentially present in the bacterial transfectant. Only the relevant sections of the spectra are shown. This ion peak was also not detected in the adjacent fractions 207 and 209 of the untransfected cells (not shown). *B*, MALDI-TOF/TOF MS/MS spectrum of the ion peak at *m/z* 1054.7 from the bacterial NQRA transfectant showing the assigned sequence (*above*) and MS/MS spectrum of the corresponding synthetic peptide (*below*). The assignment of the ion peak with *m/z* 129.1 as b1 is ambiguous because it is also consistent with a Lys immonium ion.

The possibility that low amounts of NQRA-(330–338) might have escaped detection in our comparison was addressed by determining the elution position of the synthetic peptide in the same chromatographic conditions used for fractionation of the HLA-B27-bound peptide pools and looking for the corresponding ion peak in the MALDI-TOF MS spectra of that and neighbor (±6) HPLC fractions from the bacterial transfectant. No evidence for an ion peak consistent with NQRA-(330–338) was found within the detection limits of the instrument used (data not shown). Of note, this peptide was very inefficient in inducing HLA-B27 refolding *in vitro* (7), suggesting that it is a low affinity ligand.

The Predicted T-cell Epitope from the Chlamydial PqqC Protein Is Endogenously Processed and Presented by HLA-B27—The same strategy was used for determining the endogenous processing of HLA-B27 ligands from PqqC. Upon systematic comparison of the MALDI-TOF MS spectra from the HPLC-fractionated HLA-B27-bound peptide pools involving 4905 ion peaks from the PqqC transfectant, two ion peaks not found in the corresponding HPLC fractions from B*2705-C1R cells (Figs. 4*A* and 5*A*) or B*2705-C1R expressing GFP alone (data not shown) were detected. One of these peaks, with *m/z* 1055.7 in HPLC fr. 184 from the PqqC transfectant (Fig. 4*A*), had the same molecular mass and chromatographic retention time as the predicted T-cell epitope from this protein spanning residues 70–78, ARKLLLDNL (6). The amino acid sequence of the differentially expressed B*2705 ligand was determined by MALDI-TOF/TOF MS/MS, and its assignment as the chlamydial T-cell epitope was formally confirmed by comparison of the MS/MS spectrum with that of the corresponding synthetic peptide (Fig. 4*B*).

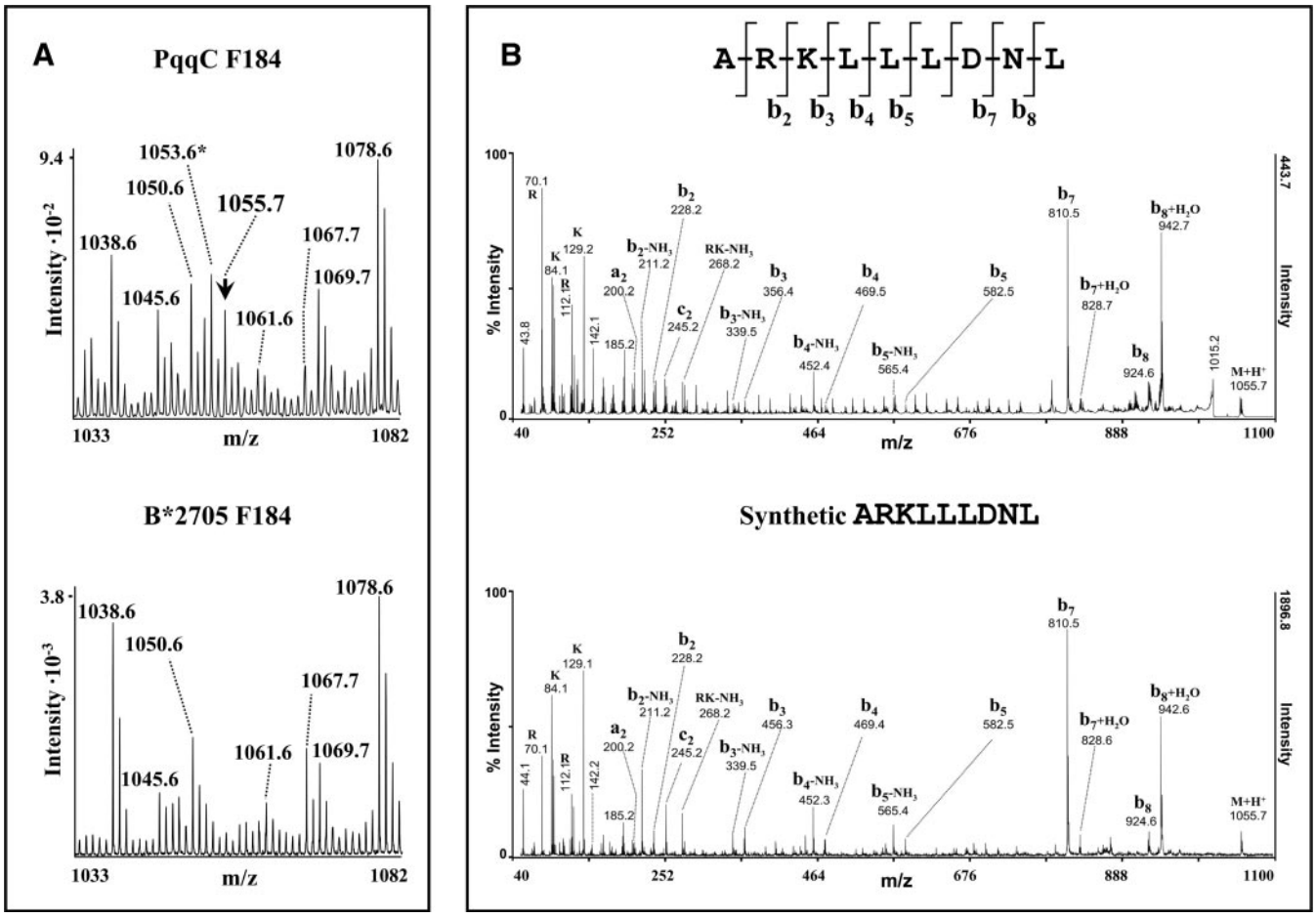


FIG. 4. **A single endogenous HLA-B27 ligand identified from the chlamydial PqqC protein shows identity with the predicted T-cell epitope.** *A*, MALDI-TOF MS spectra of fr. 184 from the HLA-B27-bound peptide pools isolated from the PqqC transfectant (*above*) and untransfected B*2705-C1R cells (*below*) showing a single detected ion peak, at *m/z* 1055.7, differentially present in the bacterial transfectant. Only the relevant sections of the spectra are shown. This ion peak was also not detected in the adjacent fractions 183 and 185 of the untransfected cells (not shown). The ion peak at *m/z* 1053.6, labeled with an *asterisk*, was a shared ligand found in the adjacent fr. 185 from B*2705-C1R. *B*, MALDI-TOF/TOF MS/MS spectrum of the ion peak at *m/z* 1055.7 from the bacterial PqqC transfectant showing the assigned sequence (*above*) and MS/MS spectrum of the corresponding synthetic peptide (*below*). The periodically spaced ion peaks in the *upper* MS/MS spectrum are related to each other by 44 *m/z* units with a related series differing by 16 *m/z* units. They arise from traces of the Igepal CA-630 detergent (octylphenyl polyethylene glycol, $(C_2H_4O)nC_{14}H_{22}O$) used for cell lysis.

The second differential ion peak, with *m/z* 1183.5, was found in HPLC fr. 182 from the PqqC transfectant (Fig. 5*A*). The amino acid sequence of this peptide was determined to be TRFSYAEYF (Fig. 5*B*). It was unrelated to the chlamydial PqqC protein but matched exactly the N-terminal region of the human ankyrin repeat and SOCS box protein 2 (ASB2; UniProtKB/Swiss-Prot accession number Q96Q27). This is a likely substrate recognition component of E3 ubiquitin-protein ligase complexes (32).

NQRA-(86–94) and PqqC-(70–78) Are Highly Homologous to Human Protein Sequences with the HLA-B27 Binding Motif—The chlamydial B*2705 ligand NQRA-(86–94) was screened for homologous sequences against the human proteome database. This search identified a highly homologous sequence (Table I) from the DnaJ homolog DJB14 (UniProtKB/Swiss-Prot accession number Q8TBM8). The following features confer special significance to this homology. First, the human counterpart contains the binding motif of HLA-B27, including Arg-2, a nonpolar C-terminal residue (Met-9), and a small N-terminal residue (Ala-1) that is frequent among B*2705 ligands (33). Second, six of nine consecutive residues in the human homolog are identical to the chlamydial peptide. These include Arg-2, the secondary anchor residue Ala-3, and the central region (residues 4–7), which is recognized by the T-cell antigen receptor (TCR) (34). Third, this stretch of identical residues is followed by two conservative differences at positions 8 and 9 involving aliphatic residues.

An analogous search was performed for the chlamydial PqqC-(70–78) epitope. Again this peptide showed a very high homology with residues 172–180 of the exportin 6 (UniProtKB/Swiss-Prot accession number Q96QU8). This human sequence also shows six of nine identical residues, including the

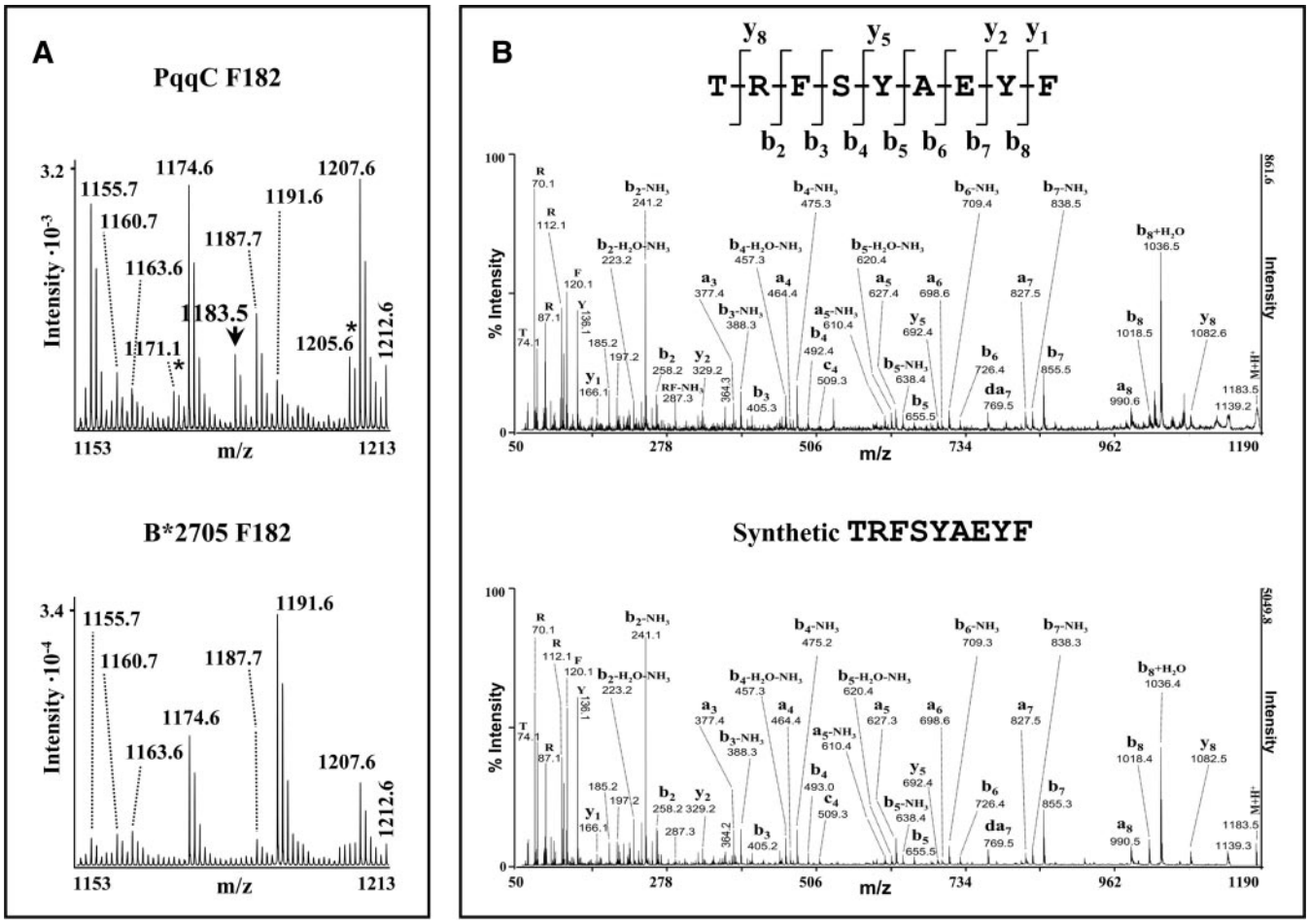


FIG. 5. **Identification of a self-derived HLA-B27 ligand differentially detected in PqqC transfectant cells.** *A*, MALDI-TOF MS spectra of fr. 182 from the HLA-B27-bound peptide pools isolated from the PqqC transfectant (*above*) and untransfected B*2705-C1R cells (*below*) showing an ion peak, at *m/z* 1183.5, differentially present in the bacterial transfectant. Only the relevant sections of the spectra are shown. This ion peak was also not detected in the adjacent fractions 181 and 183 of the untransfected cells (not shown). The ion peaks at *m/z* 1171.1 and 1205.6 of the bacterial transfectant, labeled with an *asterisk*, were shared ligands found in an independent preparation and in the adjacent fraction 183 from B*2705-C1R, respectively (not shown). *B*, MALDI-TOF/TOF MS/MS spectrum of the ion peak at *m/z* 1183.5 from the bacterial PqqC transfectant showing the assigned sequence (*above*) and MS/MS spectrum of the corresponding synthetic peptide (*below*).

TABLE I
Homology of naturally processed HLA-B27 ligands from C. trachomatis with human protein sequences

Sequence[a]	Identity	Organism	Parental protein (residues)	UniProtKB/Swiss-Prot accession no.
	%			
KRALLEIVI		*C. trachomatis*	Na^+-translocating NADH-quinone reductase subunit A, NQRA-(86–94)	O84639
ARALLEIIM	66.7	*Homo sapiens*	DnaJ homolog protein, DNAJB14-(43–51)	Q8TBM8
ARKLLLDNL		*C. trachomatis*	Pyrroloquinoline-quinone synthase protein, PqqC-(70–78)	O84616
LRKLLLDQV	66.7	*H. sapiens*	Exportin-6 protein, Exp6-(172–180)	Q96QU8

[a] Residues in the human sequences identical to the bacterial peptides are underlined.

Arg-2 motif of HLA-B27 ligands, the secondary anchor residue Lys-3, and identity in the whole central region. The two sequences also show chemically similar aliphatic residues at the anchor positions 1 and 9 and polar residues at the accessible position 8 (Table I). Thus, both endogenously processed HLA-B27 ligands from *C. trachomatis* are highly homologous to human sequences containing the HLA-B27 binding motif and many features that might favor TCR cross-recognition.

DISCUSSION

The recognition of HLA-B27-restricted chlamydial antigens by CTL is potentially important in the pathogenesis of spon-

dyloarthropathies at least for two reasons. First because it may favor the establishment of bacterial persistence, a crucial event in chronic infection, through the production of IFN-γ (10). Second because it might induce autoimmune cross-reactivity and tissue injury through molecular mimicry with self-derived HLA-B27 ligands (13, 35, 36). Thus, the report of a series of chlamydial HLA-B27 ligands, selected from the screening of the bacterial genome with predictive algorithms, that were recognized by $CD8^+$ T-cells from patients with ReA (6, 7) was a significant step toward defining the nature of the CTL responses in this disease. However, this approach has several limitations. The most important one is that, because of the intrinsic T-cell cross-reactivity (21, 22), a synthetic peptide recognized by CTL *in vitro* could be unrelated to the natural epitope that stimulated that CTL *in vivo*. Moreover the predictive algorithms were based on parameters that may not be highly accurate because of insufficient refinement of MHC class I binding motifs (37) or to lack of correspondence between proteasomal cleavage *in vivo* and *in vitro* (38). Finally the screening of the chlamydial genome with these predictive algorithms was limited to nonameric sequences, although a significant percentage of HLA-B27 ligands are longer (33). Thus, the relevance of the predicted T-cell epitopes must be validated by demonstrating their endogenous processing and presentation *in vivo*.

Our approach to this issue involved the expression of chlamydial proteins in human cells and the identification of endogenously processed bacterial HLA-B27 ligands by MS. Several methodological aspects deserve consideration. First, is this a sound alternative to the identification of chlamydial epitopes from infected cells? Second, are the proteins analyzed likely to reach the MHC class I antigen processing pathway under conditions of natural infection? Third, is an MS-based approach sensitive enough to detect natural T-cell epitopes, which might be expressed in very low amounts and still recognized by CTL (39, 40)? The two first questions can only be addressed by considering the mechanism of antigen cross-presentation (41). Chlamydial proteins should reach the MHC class I pathway of the professional antigen-presenting dendritic cells after uptake of antigen-loaded apoptotic bodies. Bacterial proteins in the endosomal compartment of dendritic cells can be transferred to the cytosol by incompletely characterized mechanisms (42, 43) and incorporated into the proteasome pathway. In principle, any bacterial protein might be processed in this way, although many unknown factors in the endosome-to-cytosol transfer mechanism may impose unforeseen limitations on the chlamydial proteins amenable to proteasomal degradation. Other mechanisms that bypass protein entry into the cytosol may also play a role, such as endoplasmic reticulum-phagosome fusion or antigen processing in the endocytic compartment followed by peptide interchange by the MHC class I molecule upon recycling between the endosomal compartment and the cell surface. However, the actual contribution of these pathways to cross-presentation remains ill defined (41). Thus, the generation of HLA-B27 epitopes upon processing of endogenously synthesized chlamydial proteins may be similar to antigen processing through the cytosolic pathway as it occurs on antigen-presenting cells upon natural infection. However, we cannot rule out that processing of a fusion protein in a B cell line, as analyzed in this study, might differ from processing of a natural bacterial protein in dendritic cells. In *Chlamydia*-infected host cells the most obvious candidates to reach the cytosolic pathway for MHC class I-mediated antigen presentation are the proteins secreted into the cytoplasm (44) or those in the vacuolar membrane (18, 19). However, other chlamydial proteins or protein fragments might reach the cytosol through cross-presentation mechanisms that remain largely undefined in non-professional antigen-presenting cells.

Another question relates to the sensitivity of MS relative to T-cell recognition. We grossly estimated that, in our experimental conditions, the detection limit for HLA-B27 ligands would be around 100–200 copies per cell. This might be 1–2 orders of magnitude below the sensitivity of CTL, which might recognize only a few epitope copies per cell (39, 40). Yet it is reasonable to assume that a transfected protein will yield much more peptide copies presented by HLA-B27 than on infected cells because a much greater protein amount will reach the cytosolic degradation pathway. Thus, it is likely that our system can detect peptides that might be presented upon natural infection in extremely low amounts. For this reason, a candidate chlamydial peptide not found in the B27-bound peptide pool in our analysis is unlikely to be presented on infected cells. Yet this possibility cannot be ruled out.

With the above considerations in mind, our results demonstrate both the reach and the limitations of the epitope prediction algorithms used on the chlamydial genome (6). Although they identified at least one peptide, PqqC-(70–78), they failed to predict another, NQRA-(86–94), and predicted an epitope that is apparently not processed *in vivo*, NQRA-(330–338). Thus, predictive algorithms for putative T-cell epitopes can only be considered as a first step to orient an experimental search for naturally processed antigens. To our knowledge, PqqC-(70–78) is the first HLA-B27-restricted antigen from *C. trachomatis* recognized by $CD8^+$ T-cells from ReA patients that has been shown to be endogenously processed and presented *in vivo*. Yet this does not definitely prove that PqqC-(70–78) is a target of the immune response to *C. trachomatis*, which would require showing that the same peptide is presented on *Chlamydia*-infected cells or that a peptide-specific T-cell clone recognizes such cells.

We have previously shown that a peptide derived from the chlamydial DNA primase with high homology to a natural self-derived ligand of HLA-B27 is endogenously processed and presented by this molecule, providing a sound basis for molecular mimicry between chlamydial and self-derived HLA-

B27 ligands (16). The two chlamydial peptides endogenously processed and presented by HLA-B27 in the present study have closely related human sequences, including the same or similar anchor residues and identical TCR-interacting regions. This level of homology between short bacterial peptides and human proteins may be relatively frequent and not restricted to these examples (45). Whether the human homologs are actually processed as natural HLA-B27 ligands is a question of obvious relevance that was not addressed in this study. In the absence of a complete examination of this issue the significance of these homologies is unknown. However, they show that there are multiple possibilities for the generation of self-derived HLA-B27 ligands with obvious mimicry of naturally processed chlamydial epitopes.

We do not know the reason for the differential expression of a self-derived peptide in the PqqC transfectant. There are three possibilities. First, the peptide could be a shared ligand that was detected by chance only in this transfectant but not in the other cells analyzed. Second, the parental ASB2 protein of this self-derived ligand was up-regulated by chance in the particular PqqC transfectant used in our study. Third, up-regulation of ASB2 might have occurred as a specific consequence of the expression of the GFP-PqqC fusion protein. We consider the first possibility unlikely because the differential ion peak was observed in the same transfectant but not in any of the others in two distinct experiments involving independently obtained peptide pools. We cannot rule out the second possibility as it would require repeating this analysis in a second, independently obtained PqqC transfectant, which was not attempted. The third possibility would be consistent with the speculation that GFP-PqqC expression might up-regulate certain components of the ubiquitin/proteasome pathway of protein degradation to which ASB2 belongs. Because this issue is unrelated to the main point of this study, it was not further pursued.

In conclusion, expression of chlamydial proteins in HLA-B27-positive human cells is a reliable and sensitive way to identify endogenously processed bacterial ligands presented by HLA-B27 *in vivo*. Using this approach we demonstrated, for the first time to our knowledge, the endogenous processing and presentation of an HLA-B27-restricted epitope recognized by $CD8^+$ T-cells from ReA patients but failed to confirm the endogenous generation of a second T-cell epitope identified by predictive algorithms. In contrast, we identified a non-predicted nonamer from the same parental protein that was endogenously processed and presented by HLA-B27, revealing the limitations of predictive methods for identifying natural MHC class I ligands and T-cell epitopes. The very close similarity of the two bacterial HLA-B27 ligands identified in this study to amino acid sequences containing the HLA-B27 binding motif in the human proteome suggests that this is a possible source of self-derived peptides mimicking chlamydial epitopes. This would be consistent with an autoimmune mechanism of ReA in which T-cell cross-reactivity between *Chlamydia*-specific and self-derived HLA-B27 ligands may play a crucial role.

Acknowledgment—We thank the staff of the proteomics facilities at the Centro Nacional de Biotecnología, Madrid (member of the ProteoRed network) for help in MS.

* This work was supported by Ministry of Science and Innovation Grants SAF2005/03188 and SAF2008/00461 and an institutional grant from the Fundación Ramón Areces (to the Centro de Biología Molecular Severo Ochoa).

S The on-line version of this article (available at http://www.mcponline.org) contains supplemental material.

‡ To whom correspondence should be addressed: Centro de Biología Molecular Severo Ochoa, c/ Nicolás Cabrera N. 1, Universidad Autónoma, 28049 Madrid, Spain. Tel.: 34-91-196-4554; Fax: 34-91-196-4420; E-mail: aldecastro@cbm.uam.es.

REFERENCES

1. Zeidler, H., Kuipers, J., and Köhler, L. (2004) Chlamydia-induced arthritis. *Curr. Opin. Rheumatol.* **16,** 380–392
2. Zhong, G., Fan, T., and Liu, L. (1999) Chlamydia inhibits interferon gamma-inducible major histocompatibility complex class II expression by degradation of upstream stimulatory factor 1. *J. Exp. Med.* **189,** 1931–1938
3. Zhong, G., Liu, L., Fan, T., Fan, P., and Ji, H. (2000) Degradation of transcription factor RFX5 during the inhibition of both constitutive and interferon gamma-inducible major histocompatibility complex class I expression in chlamydia-infected cells. *J. Exp. Med.* **191,** 1525–1534
4. Zhong, G., Fan, P., Ji, H., Dong, F., and Huang, Y. (2001) Identification of a chlamydial protease-like activity factor responsible for the degradation of host transcription factors. *J. Exp. Med.* **193,** 935–942
5. Loomis, W. P., and Starnbach, M. N. (2002) T cell responses to Chlamydia trachomatis. *Curr. Opin. Microbiol.* **5,** 87–91
6. Kuon, W., Holzhütter, H. G., Appel, H., Grolms, M., Kollnberger, S., Traeder, A., Henklein, P., Weiss, E., Thiel, A., Lauster, R., Bowness, P., Radbruch, A., Kloetzel, P. M., and Sieper, J. (2001) Identification of HLA-B27-restricted peptides from the *Chlamydia trachomatis* proteome with possible relevance to HLA-B27-associated diseases. *J. Immunol.* **167,** 4738–4746
7. Appel, H., Kuon, W., Kuhne, M., Wu, P., Kuhlmann, S., Kollnberger, S., Thiel, A., Bowness, P., and Sieper, J. (2004) Use of HLA-B27 tetramers to identify low-frequency antigen-specific T cells in Chlamydia-triggered reactive arthritis. *Arthritis Res. Ther.* **6,** R521–534
8. Lampe, M. F., Wilson, C. B., Bevan, M. J., and Starnbach, M. N. (1998) Gamma interferon production by cytotoxic T lymphocytes is required for resolution of Chlamydia trachomatis infection. *Infect. Immun.* **66,** 5457–5461
9. Shemer, Y., and Sarov, I. (1985) Inhibition of growth of Chlamydia trachomatis by human gamma interferon. *Infect. Immun.* **48,** 592–596
10. Brunham, R. C., and Rey-Ladino, J. (2005) Immunology of Chlamydia infection: implications for a Chlamydia trachomatis vaccine. *Nat. Rev. Immunol.* **5,** 149–161
11. Keat, A. C., Maini, R. N., Nkwazi, G. C., Pegrum, G. D., Ridgway, G. L., and Scott, J. T. (1978) Role of Chlamydia trachomatis and HLA-B27 in sexually acquired reactive arthritis. *Br. Med. J.* **1,** 605–607
12. Young, J. L., Smith, L., Matyszak, M. K., and Gaston, J. S. (2001) HLA-B27 expression does not modulate intracellular Chlamydia trachomatis infection of cell lines. *Infect. Immun.* **69,** 6670–6675
13. Bachmaier, K., Neu, N., de la Maza, L. M., Pal, S., Hessel, A., and Penninger, J. M. (1999) Chlamydia infections and heart disease linked through antigenic mimicry. *Science* **283,** 1335–1339
14. Popov, I., Dela Cruz, C. S., Barber, B. H., Chiu, B., and Inman, R. D. (2002) Breakdown of CTL tolerance to self HLA-B*2705 induced by exposure to Chlamydia trachomatis. *J. Immunol.* **169,** 4033–4038
15. Ramos, M., Alvarez, I., Sesma, L., Logean, A., Rognan, D., and López de Castro, J. A. (2002) Molecular mimicry of an HLA-B27-derived ligand of arthritis-linked subtypes with chlamydial proteins. *J. Biol. Chem.* **277,** 37573–37581
16. Cragnolini, J. J., and de Castro, J. A. (2008) Identification of endogenously

presented peptides from Chlamydia trachomatis with high homology to human proteins and to a natural self-ligand of HLA-B27. *Mol. Cell. Proteomics* **7,** 170–180
17. Karunakaran, K. P., Rey-Ladino, J., Stoynov, N., Berg, K., Shen, C., Jiang, X., Gabel, B. R., Yu, H., Foster, L. J., and Brunham, R. C. (2008) Immunoproteomic discovery of novel T cell antigens from the obligate intracellular pathogen Chlamydia. *J. Immunol.* **180,** 2459–2465
18. Fling, S. P., Sutherland, R. A., Steele, L. N., Hess, B., D'Orazio, S. E., Maisonneuve, J., Lampe, M. F., Probst, P., and Starnbach, M. N. (2001) CD8+ T cells recognize an inclusion membrane-associated protein from the vacuolar pathogen Chlamydia trachomatis. *Proc. Natl. Acad. Sci. U.S.A.* **98,** 1160–1165
19. Starnbach, M. N., Loomis, W. P., Ovendale, P., Regan, D., Hess, B., Alderson, M. R., and Fling, S. P. (2003) An inclusion membrane protein from Chlamydia trachomatis enters the MHC class I pathway and stimulates a CD8+ T cell response. *J. Immunol.* **171,** 4742–4749
20. Grotenbreg, G. M., Roan, N. R., Guillen, E., Meijers, R., Wang, J. H., Bell, G. W., Starnbach, M. N., and Ploegh, H. L. (2008) Discovery of CD8+ T cell epitopes in Chlamydia trachomatis infection through use of caged class I MHC tetramers. *Proc. Natl. Acad. Sci. U.S.A.* **105,** 3831–3836
21. Wucherpfennig, K. W., and Strominger, J. L. (1995) Molecular mimicry in T cell-mediated autoimmunity: viral peptides activate human T cell clones specific for myelin basic protein. *Cell* **80,** 695–705
22. Evavold, B. D., Sloan-Lancaster, J., Wilson, K. J., Rothbard, J. B., and Allen, P. M. (1995) Specific T cell recognition of minimally homologous peptides: evidence for multiple endogenous ligands. *Immunity* **2,** 655–663
23. Zemmour, J., Little, A. M., Schendel, D. J., and Parham, P. (1992) The HLA-A,B "negative" mutant cell line C1R expresses a novel HLA-B35 allele, which also has a point mutation in the translation initiation codon. *J. Immunol.* **148,** 1941–1948
24. Paradela, A., García-Peydró, M., Vázquez, J., Rognan, D., and López de Castro, J. A. (1998) The same natural ligand is involved in allorecognition of multiple HLA-B27 subtypes by a single T cell clone: role of peptide and the MHC molecule in alloreactivity. *J. Immunol.* **161,** 5481–5490
25. Barnstable, C. J., Bodmer, W. F., Brown, G., Galfre, G., Milstein, C., Williams, A. F., and Ziegler, A. (1978) Production of monoclonal antibodies to group A erythrocytes, HLA and other human cell surface antigens. New tools for genetic analysis. *Cell* **14,** 9–20
26. Paradela, A., Alvarez, I., García-Peydró, M., Sesma, L., Ramos, M., Vázquez, J., and López De Castro, J. A. (2000) Limited diversity of peptides related to an alloreactive T cell epitope in the HLA-B27-bound peptide repertoire results from restrictions at multiple steps along the processing-loading pathway. *J. Immunol.* **164,** 329–337
27. Stephens, R. S., Kalman, S., Lammel, C., Fan, J., Marathe, R., Aravind, L., Mitchell, W., Olinger, L., Tatusov, R. L., Zhao, Q., Koonin, E. V., and Davis, R. W. (1998) Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trachomatis. *Science* **282,** 754–759
28. Olsen, A. W., Follmann, F., Højrup, P., Leah, R., Sand, C., Andersen, P., and Theisen, M. (2007) Identification of human T cell targets recognized during Chlamydia trachomatis genital infection. *J. Infect. Dis.* **196,** 1546–1552
29. Stenner-Liewen, F., Liewen, H., Zapata, J. M., Pawlowski, K., Godzik, A., and Reed, J. C. (2002) CADD, a Chlamydia protein that interacts with death receptors. *J. Biol. Chem.* **277,** 9633–9636
30. Schwarzenbacher, R., Stenner-Liewen, F., Liewen, H., Robinson, H., Yuan, H., Bossy-Wetzel, E., Reed, J. C., and Liddington, R. C. (2004) Structure of the Chlamydia protein CADD reveals a redox enzyme that modulates host cell apoptosis. *J. Biol. Chem.* **279,** 29320–29324
31. Skipp, P., Robinson, J., O'Connor, C. D., and Clarke, I. N. (2005) Shotgun proteomic analysis of Chlamydia trachomatis. *Proteomics* **5,** 1558–1573
32. Heuzé, M. L., Guibal, F. C., Banks, C. A., Conaway, J. W., Conaway, R. C., Cayre, Y. E., Benecke, A., and Lutz, P. G. (2005) ASB2 is an Elongin BC-interacting protein that can assemble with Cullin 5 and Rbx1 to reconstitute an E3 ubiquitin ligase complex. *J. Biol. Chem.* **280,** 5468–5474
33. Lopez de Castro, J. A., Alvarez, I., Marcilla, M., Paradela, A., Ramos, M., Sesma, L., and Vázquez, M. (2004) HLA-B27: a registry of constitutive peptide ligands. *Tissue Antigens* **63,** 424–445
34. Garboczi, D. N., Ghosh, P., Utz, U., Fan, Q. R., Biddison, W. E., and Wiley, D. C. (1996) Structure of the complex between human T-cell receptor, viral peptide and HLA-A2. *Nature* **384,** 134–141
35. Benjamin, R., and Parham, P. (1990) Guilt by association: HLA-B27 and ankylosing spondylitis. *Immunol. Today* **11,** 137–142
36. Bachmaier, K., and Penninger, J. M. (2005) Chlamydia and antigenic mimicry. *Curr. Top. Microbiol. Immunol.* **296,** 153–163
37. Ruppert, J., Sidney, J., Celis, E., Kubo, R. T., Grey, H. M., and Sette, A. (1993) Prominent role of secondary anchor residues in peptide binding to HLA-A2.1 molecules. *Cell* **74,** 929–937
38. Rock, K. L., and Goldberg, A. L. (1999) Degradation of cell proteins and the generation of MHC class I-presented peptides. *Annu. Rev. Immunol.* **17,** 739–779
39. Sykulev, Y., Joo, M., Vturina, I., Tsomides, T. J., and Eisen, H. N. (1996) Evidence that a single peptide-MHC complex on a target cell can elicit a cytolytic T cell response. *Immunity* **4,** 565–571
40. Purbhoo, M. A., Irvine, D. J., Huppa, J. B., and Davis, M. M. (2004) T cell killing does not require the formation of a stable mature immunological synapse. *Nat. Immunol.* **5,** 524–530
41. Lin, M. L., Zhan, Y., Villadangos, J. A., and Lew, A. M. (2008) The cell biology of cross-presentation and the role of dendritic cell subsets. *Immunol. Cell Biol.* **86,** 353–362
42. Ackerman, A. L., Giodini, A., and Cresswell, P. (2006) A role for the endoplasmic reticulum protein retrotranslocation machinery during crosspresentation by dendritic cells. *Immunity* **25,** 607–617
43. Rock, K. L. (2006) Exiting the outside world for cross-presentation. *Immunity* **25,** 523–525
44. Kleba, B., and Stephens, R. S. (2008) Chlamydial effector proteins localized to the host cell cytoplasmic compartment. *Infect. Immun.* **76,** 4842–4850
45. Davies, J. M. (2000) Introduction: Epitope mimicry as a component cause of autoimmune disease. *Cell Mol. Life Sci.* **57,** 523–526

Presentation of Cytosolically Stable Peptides by HLA-B27 Is Not Dependent on the Canonic Interactions of N-Terminal Basic Residues in the A Pocket[1]

Patricia Gómez, Carla Mavian, Begoña Galocha, Noel García-Medel, and José A. López de Castro[2]

HLA-B27 binds peptides with R at position 2. Additionally, a substantial fraction of the HLA-B27-bound peptide repertoire has basic residues at position 1. It is unclear whether this is determined by structural complementarity with the A pocket of the peptide-binding site, by the increased availability of peptides with dibasic N-terminal sequences resulting from their cytosolic stability, or both. To distinguish between these possibilities two B*2705 mutants were generated in which one or two A pocket surface residues stabilizing the peptidic R1 side chain were changed: E163T and E163T-W167S. Both mutants bound a large fraction of the constitutive peptide repertoire of B*2705. Moreover, 90 B*2705 ligands of known sequence were examined for their endogenous presentation by the mutants. The E163T mutation alone had a limited effect on binding of peptides with R1 or K1 and on the relative frequencies of N-terminal residues. However, it decreased the overall stability of the molecule. The E163T-W167S mutant also bound many of the B*2705 ligands with N-terminal basic residues, but its preference for G1 was significantly decreased. The results indicate that the capacity of HLA-B27 to bind peptides with N-terminal basic residues is largely independent of the canonic interactions that stabilize at least the R1 side chain. Thus, the prevalence of HLA-B27 ligands with dibasic N-terminal sequences may be significantly influenced by the increased availability of these peptides resulting from their cytosolic stability. This confers to HLA-B27 a unique capacity to present Ags generated in low amounts, but resistant to intracellular degradation. *The Journal of Immunology,* **2009, 182: 446–455.**

HLA-B27 is a major susceptibility factor for ankylosing spondylitis and other spondyloarthropathies (1, 2). Although the mechanism of this association is not known, the peptide-binding properties of the molecule are thought to be critical, either through presentation of foreign peptide Ags showing molecular mimicry with self-ligands (3), or through a more general influence of peptide ligands in determining folding and stability of the native molecule (4). Additionally, HLA-B27 is a very efficient Ag-presenting molecule (5), determining good protective responses against a variety of infections, including HIV (6–9) and hepatitis C (10). These observations justify the interest of characterizing the peptide-binding specificity of HLA-B27 and its molecular basis.

A major feature of HLA-B27 ligands is their almost absolute restriction for R at position (P)[3] 2 (11, 12), which is determined by the structure of the B pocket, of which a critical feature is the E45 residue (13, 14). HLA-B27 ligands show higher heterogeneity at the N-terminal (P1) position, although basic (R, K) and small residues (G, A, S) are predominant; for example, in a published registry of 108 constitutive B*2705 nonamer ligands (12), the frequency of basic and small residues was 27.8% (R1, 20.4%; K1, 7.4%) and 41.6% (G1, 23.1%; A1, 10.2%; S1, 8.3%), respectively. P1 residues bind in the A pocket. Their interactions include a conserved network of hydrogen bonds established with the peptidic N terminus within the pocket itself (13), as well as contacts of residues at the surface of the pocket with the P1 side chains. The influence of P1 residues on binding to HLA-B*2705 was assessed in one study with poly-A analogs containing the R2 motif and changes at P1 (15). In this study R1 was the residue that made the highest contribution to binding.

X-ray analyses have determined the binding mode of the peptidic R1 to HLA-B27 and provided an explanation for its contribution to peptide affinity (16). Three residues at the surface of the A pocket interact with the R1 side chain of HLA-B27 ligands: R62, E163, and W167. E163 forms a salt bridge with the guanidinium group of the peptidic R1. Additionally, this residue interacts with R62 through π-π stacking of their guanidinium groups and with the indole ring of W167 through van der Waals interactions of the aliphatic R1 side chain. Thus, R62 and W167 form a tightly packed sandwich with the peptidic R1, allowing for its stable anchoring to HLA-B27.

Although structural considerations might explain the prevalence of R1 among HLA-B27 ligands, an alternative, not mutually exclusive, explanation was recently proposed on the basis that peptides with N-terminal dibasic sequences (i.e., RR, KR) were particularly resistant to amino-peptidase-mediated cytosolic

Centro de Biología Molecular "Severo Ochoa" (Consejo Superior de Investigaciones Científicas and Universidad Autónoma de Madrid), Universidad Autónoma, Madrid, Spain

Received for publication September 5, 2008. Accepted for publication October 30, 2008.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[1] This work was supported by Grant SAF2005/03188 from the Ministry of Science and Technology and an institutional grant from the Fundación Ramón Areces to the Centro de Biología Molecular "Severo Ochoa".

[2] Address correspondence and reprint requests to Dr. José A. López de Castro, Centro de Biología Molecular "Severo Ochoa", C/Nicolás Cabrera, N.1, Universidad Autónoma, 28049 Madrid, Spain. E-mail address: aldecastro@cbm.uam.es

[3] Abbreviations used in this paper: P, position: P1, N terminal position; T163, E163T; T163S167, E163T-W167S; C1R, HMy2.C1R; HC, heavy chain; TFA, trifluoroacetic acid; MS, mass spectrometry; *m/z,* mass-to-charge.

degradation (17). Thus, such peptides, even if generated in relatively low amounts, would advantageously compete for survival in the cytosol and reach the MHC class I loading pathway. HLA allotypes, such as HLA-B27, with specificity for peptides with R2 would then preferentially bind peptides with N-terminal basic residues due to their increased availability, relative to other ligands.

The present study was undertaken to assess the role of the A pocket and, in particular, of the structural determinants of the peptidic R1 binding, in the preference of HLA-B27 for peptides with dibasic N-terminal sequences. It was reasoned that if the preference of HLA-B27 for R1 was essentially determined by A pocket interactions, presentation of peptides with N-terminal basic residues would be severely compromised by mutating the contact residues at the surface of the pocket. Conversely, if the contribution of P1 anchoring to the overall peptide affinity were not critical, even severe disruption of the interactions that stabilize the peptidic R1 side chain would not affect the preferential binding of peptides with basic P1 residues to HLA-B27, due to their relatively high availability. Thus, we generated two HLA-B*2705 mutants lacking only E163 (E163T, herein designated as T163) or both this residue and W167 (E163T-W167S, herein designated as T163S167) and compared their peptide repertoires, the P1 residue frequencies among their constitutive peptide ligands, and the molecular stability of these mutants, relative to B*2705.

Materials and Methods

Cell lines, Abs, and flow cytometry

HMy2.C1R (C1R) is an HLA-A-negative human lymphoid cell line with low expression of its endogenous HLA-B35 and -Cw4 molecules (18). C1R transfectants expressing B*2705 and the T163 mutant were obtained using genomic DNA and were previously described (19, 20). The T163S167 mutant was obtained from the full-length cDNA of T163 cloned into pcDNA3 (Invitrogen) by PCR-mediated site-directed mutagenesis using a QuickChange site-directed mutagenesis kit (Stratagene) with the following primers: 5′-CACGTGCGTGGAGTCGCTCCGCAGATACC3-′ and 3′-GGTATCTGCGGAGCGACTCCACGCACGTG5-′. After confirming the correct sequence of the double mutant, it was transfected into 10^7 C1R cells by electroporation of 10 μg DNA at 260 V and 960 μF. The transfectants were selected with increasing concentrations of G418 (Invitrogen), ranging from 0.5 to 1 mg/ml. All the cell lines were cultured in RPMI 1640 medium supplemented with 2 mM L-Gln and 10% FBS (Invitrogen). The mAbs used were W6/32 (IgG2a, specific for a monomorphic HLA class I determinant of the native heterodimer) (21), ME1 (IgG1, specific for HLA-B27, B7, B22) (22), HC10 (IgG2a), which recognizes free class I H chain (HC) (23), and the anti-γ-tubulin mAb (IgG1) GTU88 (Sigma-Aldrich). Flow cytometry was performed as previously described (24).

Western blot analysis

Cells (~10^6) were lysed at 4°C for 20 min in 0.5% Igepal CA-630 (Sigma-Aldrich), 50 mM Tris-HCl (pH 7.4), and 5 mM $MgCl_2$ containing a cocktail of protease inhibitors (Complete Mini, Roche). After centrifugation of whole lysates and SDS-PAGE, HLA class I HC and γ-tubulin were revealed with HC10 or the anti-γ-tubulin mAb, respectively, using human-absorbed HRP-conjugated goat anti-mouse Ig (SouthernBiotech) as secondary Ab.

Isolation of HLA-B27-bound peptides

HLA-B27-bound peptides were isolated from ~10^{10} C1R transfectant cells as previously described (25). Briefly, cells were lysed in 1% Igepal CA-630 (Sigma-Aldrich), 20 mM Tris-HCl buffer, and 150 mM NaCl (pH 7.5) in the presence of a cocktail of protease inhibitors. After ultracentrifugation, cell lysates were subjected to affinity chromatography using the W6/32 mAb. HLA-B27-bound peptide pools were eluted with 0.1% aqueous trifluoroacetic acid (TFA) at room temperature, filtered through Centricon 3 (Millipore), concentrated, and subjected to reverse-phase HPLC fractionation. This was conducted in a Waters Alliance system using a Vydac 218TP52 column at a flow rate of 100 μl/min, as previously described (26). Fractions of 50 μl were collected.

Mass spectrometry (MS) analysis and peptide sequencing

The peptide composition of HPLC fractions was analyzed by MALDI-TOF MS using an Autoflex MALDI-TOF mass spectrometer (Bruker Daltonik) in positive ion reflector mode. Dried fractions were resuspended in 0.5 μl of 33% aqueous acetonitrile and 0.1% TFA. They were deposited onto an AnchorChip 600/384 target plate (Bruker Daltonik) and allowed to dry at room temperature. Then, 0.5 μl of matrix solution (α-cyano-4-hydroxycinnamic acid in 33% aqueous acetonitrile and 0.1% TFA) at 2 mg/ml were added and allowed to dry again at room temperature. MS spectra were processed using the MoverZ software (version 2001.02.03: www.bioinformatics.genomicsolutions.com/moverZDL.html).

In some instances individual HPLC fractions were reanalyzed using a MALDI-TOF/TOF instrument (4800 Proteomics Analyzer, Applied Biosystems). In these cases, the dried HPLC fractions were reconstituted with 0.6 μl of 33% aqueous acetonitrile and 0.1% TFA, loaded onto an Opti-TOF 384-well MALDI insert (Applied Biosystems) and allowed to dry at room temperature. Then, 0.6 μl of the matrix solution (α-cyano-4-hydroxycinnamic acid in 33% aqueous acetonitrile and 0.1% TFA) was added at 3 mg/ml. These mass spectra were acquired in reflector positive mode and processed using the 4000 Series Explorer software version 3.5.

Peptide sequencing was conducted by quadrupole ion trap nanoelectrospray MS/MS in a Thermo Finnigan LCQ DECA-XP instrument, as previously described (27). Alternatively, a Surveyor HPLC system coupled online to an LTQ LIT instrument (Thermo Finnigan) was used. Peptides were concentrated in a reverse-phase precolumn (0.32 × 30 mm, BioBasic 18, Thermo Electron) and eluted on an analytical column (0.18 × 150 mm, BioBasic 18, Thermo Fisher Scientific) using a 72- to 86-min gradient from 5% to 40% solvent B (solvent A: 0.1% formic acid; solvent B: 80% acetonitrile, 0.1% formic acid) at a flow rate of 2 μl/min. Normalized collision energy was set to 35% and a 3-Da mass window was used to fragment selected parental ions.

Some peptide sequences were obtained using an esquire3000^{plus} ion trap mass spectrometer (Brucker) after online chromatographic separation of samples as follows. Samples were dissolved in buffer A (0.5% acetic acid in water) and loaded onto a column (100 mm × 100 μm internal diameter) packed with 5-μm Kromasil C18 beads (EKA Chemicals) and fractionated in a Famos-Switchos-Ultimate chromatographic system (LCPackings) with a 45-min linear gradient of 5–30% buffer B (90% acetonitrile, 0.5% acetic acid in water) at 500 nl/min. MS/MS spectra were acquired by automatic switching between MS and MS/MS mode using dynamic exclusion. Some samples were analyzed using the multiple reaction monitoring mode, which allows specific peptide masses to be selected.

Interpretation of the mass spectra was done manually but assisted by various software tools as follows. Manual inspection of the spectrum usually allowed us to determine a partial sequence. This information, together with the mass-to-charge (m/z) ratio of the parent ion (in all cases, charge 2 parent ions were selected for sequencing in this study), was used as input data for a Mascot (version 2.2) search (www.matrixscience.com) in the human protein entries of the MSDB database (release 2006/31/08, Imperial College, London), using a window of 0.8 m/z units. Of the 20 output sequences showing the highest scores in this preliminary search, those few showing the canonical R2 motif of HLA-B27 ligands and absence of "prohibited" residues for HLA-B27 binding, such as N-terminal or C-terminal Pro, were selected. From each of these sequences, a list of theoretical fragment ions was generated using the MS product tool of the ProteinProspector package (version 4.27.2 basic: prospector.ucsf.edu/) as an assistance to match the putative candidate sequences to our experimental MS/MS spectrum. When one single proper match was obtained, the peptide sequence was assigned.

Residue frequency and statistical analysis

To facilitate the analysis of residue frequencies, a software tool was developed, designated as MSearcher, which recognizes peptide series written as unformatted text databases. These can be directly generated in a text editor with data transferred from conventional word processors, such as Microsoft Office applications. The software calculates several parameters from single or multiple (up to four) peptide positions. These parameters include the residue, residue type, or residue size frequencies, and the deviation from the mean values of these parameters in the proteome. The software was implemented in a user-friendly graphical interface and is freely available upon request. For statistical analysis the Fisher's exact test was used.

Thermostability assay

The thermostability assay was performed as detailed elsewhere (28). Briefly, C1R transfectants were pulse-labeled for 15 min and chased at 0,

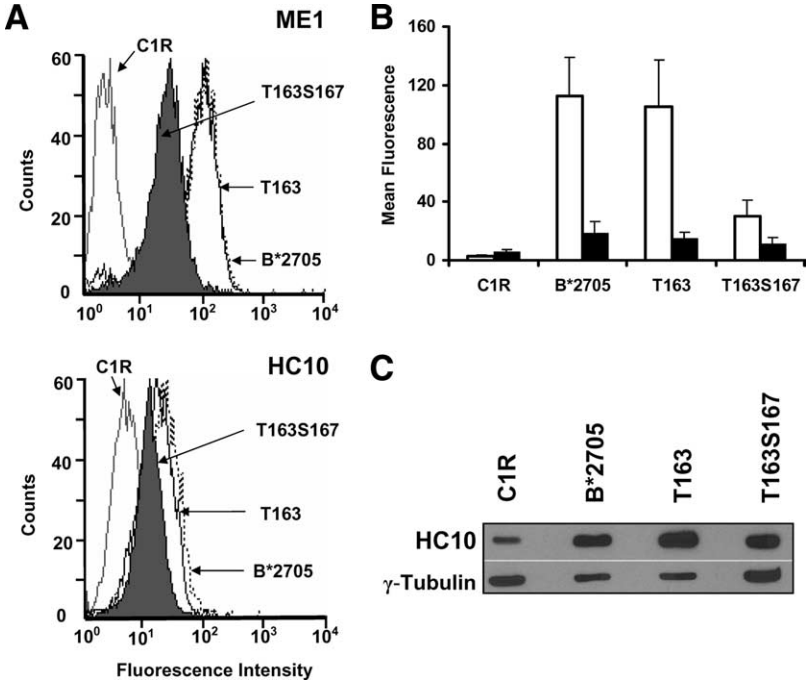


FIGURE 1. Expression of HLA-B*2705 and pocket A mutants. *A*, Flow cytometry analysis of B*2705, T163, and T163S167 on C1R transfectants, stained with ME1 and HC10. Untransfected C1R cells were used as a control. A representative experiment is shown. *B*, Comparison of the surface expression of B*2705, T163, and T163S167 heterodimers (open bars) and free HC (filled bars). The data are expressed as mean fluorescence ± SD of 12–15 experiments. The mean HC10/ME1 fluorescence ratio for B*2705, T163, and T163S167, after correcting for the background values in untransfected C1R cells, was 0.12, 0.09, and 0.21, respectively. *C*, HLA class I protein expression was analyzed by Western blot of whole lysates from untransfected C1R, B*2705, T163, and T163S167 transfectants stained with HC10. In this experiment equal amounts of material were analyzed from untransfected C1R and T163S167. Four-fold less amounts were used for B*2705 and T163. Relative class I HC expression, as estimated from the respective HC/tubulin ratios from three experiments, was: B*2705:1; T163: 1.7 ± 0.2; T163S167: 0.33 ± 0.15; untransfected C1R: 0.15.

2, and 4 h. The lysates were incubated for 1 h at various temperatures, ranging from 4°C to 50°C, immunoprecipitated with W6/32, and analyzed by SDS-PAGE. The amount of HLA-B27 heterodimer precipitated at each temperature at any given chase time was expressed as a percentage of the amount precipitated at 4°C, and it was plotted, for each time point, as a function of the temperature.

Results

*Surface expression of HLA-B*2705 and pocket A mutants*

The expression of HLA-B*2705, T163, and T163S167 on the surface of C1R transfectant cells was analyzed by flow cytometry with the ME1 and HC10 mAb, which recognize the folded heterodimers and the unfolded HC, respectively (Fig. 1, *A* and *B*). The surface expression of B*2705 and T163 heterodimers was very similar and ~3-fold higher than for the double mutant. Similar results were obtained with W6/32 (data not shown), indicating that the lower expression of the T163S167 mutant was not due to modification of the ME1 epitope. Indeed, the HC protein expression of the T163S167 mutant, as estimated by Western blot, was significantly lower than for the other two variants (Fig. 1*C*). The fluorescence ratio obtained with HC10 and ME1 (HC10/ME1 ratio) was also similar for B*2705 and T163, but was ~2-fold higher for T163S167 (Fig. 1*B*). This might be consistent with higher cell surface dissociation of the double mutant, compared with B*2705 or T163.

*Limited effect of the T163 and T163S167 mutations on the B*2705-bound peptide repertoire*

The B*2705-bound peptide pool was compared with those of T163 and T163S167 on the basis of identity of molecular mass and chromatographic retention time of the corresponding peptide species (Table I). The strategy used was the same as in previous studies from our laboratory (29–32). The peptide pools to be compared were isolated from the corresponding C1R transfectant cells after immunoaffinity purification with W6/32 and acid extraction. They were fractionated by HPLC in consecutive runs under identical conditions, and each of the peptide-containing fractions was analyzed by MALDI-TOF MS. The MS spectrum of each HPLC fraction from one variant was compared with the correlative previous and following fractions from the other variant to take into account small shifts in the retention time of peptides that may occur even in consecutive chromatographic runs. Ion peaks with the same (±0.8) *m/z* in this comparison were considered to reflect identical ligands of both variants. This assumption was validated for significantly overlapping peptide pools in previous analyses of HLA-B27 subtypes (29–31) and, in this study, by sequencing the relevant peptides from multiple variants. Ion peaks found in only one peptide pool were considered to reflect variant-specific ligands.

With this criterion, T163 shared with B*2705 ~85% of its peptide repertoire. It is likely that some of the differences may actually consist of peptides that are found in lower amounts, and below the detection limit of the MALDI-TOF instrument, in one of the variants (see below). Thus, the observed peptide sharing is probably a minimum estimation.

In the comparison with T163S167, twice as many ion peaks were detected for B*2705 than for the double mutant. This is probably due to the lower expression level of the corresponding transfectant. Thus, only the peptide sharing of the mutant relative to the wild type was taken into account, since differential ion peaks in this comparison cannot be explained by lower protein expression. T163S167 shared with B*2705 44.5% of its peptide repertoire.

Table I. *Comparison of the peptide repertoires from B*2705 and pocket A mutants*[a]

	Expt. 1		Expt. 2	
	B*2705	T163	B*2705	T163S167
Peptides compared	755	644	1356	450
Shared ligands	548 (72.6%)	548 (85.1%)	334 (24.6%)	334 (44.5%)
Differential ligands	207 (27.4%)	96 (14.9%)	1022 (75.4%)	416 (55.5%)

[a] B*2705 was compared with each mutant in separate experiments. Of the 755 B*2705 ligands compared in experiment 1, 514 (68.1%) were also in the larger set of B*2705 ligands compared in experiment 2.

Table II. *Presence of the set 1 of B*2705 ligands in the T163 and T163S167 mutants*[a]

Peptide Set 1 (n = 54)	B2705	T163	T163S167	Ref.[b]
8-mers (n = 1)	8-mers (n = 1)	8-mers (n = 1)	8-mers (n = 0)	
RRFFPYYV	**	*	Not found	(12)
9-mers (n = 29)	9-mers (n = 29)	9-mers (n = 29)	9-mers (n = 27)	
ARLFGIRAK	**	*	**	(12)
ARLKEVLEY	**	*	**	(12)
ARLQTALLV	**	*	*	(12)
FRYNGLIHR	**	*	*	(12)
GRIDKPILK	**	**	**	(12)
GRFSGLLGR	**	*	**	(12)
GRIGQAIAR	**	**	Not found	This study
GRIGVITNR	**	**	**	(12)
GRIPGIYGR	**	*	*	(12)
GRLTKHTKF	**	**	*	(12)
IRAAPPPLF	**	**	Not found	(12)
IRLPSQYNF	**	**	*	(12)
KRFDDKYTL	**	*	**	(33)
KRFEGLTQR	**	*	**	(12)
KRLVVFDAR	**	*	*	(12)
KRYKSIVKY	**	*	**	(33)
LRFPGQLNA	**	**	*	(32)
LRNQSVFNF	**	*	*	(12)
LRVTPFILK	**	*	*	(33)
QRFGPPVSR	**	**	**	This study
QRKKAYADF	**	*	*	(12)
QRNVNVFKF	**	*	*	(33)
RRDFNHINV	**	**	**	(33)
RRFGDKLNF	**	*	**	(12)
RRFFPYYVY	**	**	**	(12)
RRYQKSTEL	**	**	**	(12)
SRFPEALRL	**	*	*	(12)
SRLAIRNEF	**	**	**	(33)
SRTPYHVNL	**	*	**	(12)
10-mers (n = 12)	10-mers (n = 12)	10-mers (n = 12)	10-mers (n = 10)	
ARYGKSPYLY	**	*	*	(31)
GRFNGQFKTY	**	**	*	(12)
GRIKAIQLEY	**	**	Not found	(12)
HRFYGKNSSY	**	**	*	(12)
HRFEQAFYTY	**	*	*	(12)
KRFSVPVQHF	**	*	*	(33)
NRFAGFGIGL	**	*	**	(12)
RRFVNVVPTF	**	**	*	(12)
RRISGVDRYY	**	**	*	(12)
RRKDGVFLYF	**	*	*	(33)
RRLALFPGVA	**	**	**	(12)
RRLQIEDFEA	**	*	Not found	(31)
11-mers (n = 10)	11-mers (n = 10)	11-mers (n = 10)	11-mers (n = 7)	
ARFSPDDKYSR	**	**	*	(33)
ARNPSLKQQLF	**	*	*	(31)
RRFVNVVPTFG	**	**	*	(33)
RRLQIEDFEAR	**	*	Not found	(12)
RRYLENGKETL	**	**	*	(12)
SRAGLQFPVGR	**	**	Not found	(12)
SRAGPLSGKKF	**	*	**	(12)
SRSQTSSFFTR	**	*	*	This study
VRLLLPGELAK	**	**	**	(12)
YRVTLNPPGTF	**	*	Not found	(33)
12-mers (n = 1)	12-mers (n = 1)	12-mers (n = 1)	12-mers (n = 1)	
RRFVNVVPTFGK	**	**	*	(12)
13-mers (n = 1)	13-mers (n = 1)	13-mers (n = 1)	13-mers (n = 0)	
RRYLENGKETLQR	**	*	Not found	(12)

[a] Peptides labeled with double asterisks (**) were directly sequenced from the corresponding HLA-B27 variant. Peptides labeled with a single asterisk (*) were assigned to the corresponding variant on the basis of the presence of an ion peak with the same chromatographic retention time and m/z ratio in the MALDI-TOF MS spectrum.

[b] Previously reported and new sequences of B*2705 ligands are specified. All the sequences from the mutants were obtained in this study.

Again, the peptide sharing observed is probably a minimum estimation determined by the sensitivity of the MS instrument used. For instance, of 30 ion peaks found in B*2705, but not in the double mutant, in this comparison using a conventional MALDI-TOF instrument, 17 (56.7%) were detected using the much more sensitive (~20- to 100-fold) MALDI-TOF/TOF analyzer. If the differences found in the comparison of the peptide repertoires in Table I were corrected by this percentage, the peptide overlap of T163 and T163S167 with B*2705 would be 93.5% and 76%, respectively.

These results indicate that the single T163 mutation has a limited effect on peptide specificity, allowing for binding of a large

Table III. *Presence of the peptide set 2 in B*2705, T163, and T163S167*[a]

Peptide Set 2 (n = 41)	B2705	T163	T163S167	Ref.[b]
Shared (n = 32)				
9-mers (n = 27)	9-mers (n = 27)	9-mers (n = 27)	9-mers (n = 21)	
ARAALQELL	*	**	*	This study
ARDETEFYL	**	**	*	This study
ARFPETPAF	*	**	*	This study
ARLPSLNKL	**	**	Not found	(31)
ARNPPGFAF	*	**	**	This study
ARSEQFINL	**	**	*	This study
ARSKEVINR	**	**	*	This study
FRYQGHVGA	**	**	Not found	This study
GRAGPSYSM	**	*	Not found	This study
GRGSFKTVY	**	**	Not found	This study
GRYPGVSNY	**	**	*	(12)
IRHPNIITL	**	**	Not found	(12)
LRLEAGLNR	**	*	*	This study
LRYPMAVGL	**	**	*	(12)
NRYDGIYKV	**	*	*	This study
QRDDILINR	(s)[c]	(s)[c]	(s)[c]	(31)
QRDGYQQNF	*	**	*	This study
QRSSGLVQR	*	**	*	This study
RRARGKVKV	**	**	*	(31)
RRKDGVFLY	*	**	*	This study
SRAGIATQF	**	*	*	This study
SRFSLENNF	**	**	Not done	(31)
SRTPVLMNF	**	*	**	(12)
TRFGIAAKY	*	*	**	This study
TRFLAEEGF	**	**	*	This study
TRNNFAVGY	*	**	*	This study
VRLPPITKF	*	*	(s)[d]	This study
10-mers (n = 5)	10-mers (n = 5)	10-mers (n = 5)	10-mers (n = 4)	
GRFPDGTNGL	**	**	Not found	This study
GRYSDSLVQK	**	**	*	This study
HRYGDGGSTF	**	**	*	(12)
KRNPGVKEGY	**	**	*	(12)
KRYAVPSAGL	**	*	*	This study
B*2705-specific ligands				
9-mers (n = 4)	9-mers (n = 4)	9-mers (n = 0)	9-mers (n = 0)	
ARDNTINLI	**	ND[e]	ND	(31)
GRQGQAITL	**	ND	Not found	This study
SRIGLLTRI	**	Not found	Not found	This study
SRLKESFLV	**	ND	Not found	This study
Mutant-specific ligands				
9-mers (n = 4)	9-mers (n = 0)	9-mers (n = 1)	9-mers (n = 4)	
DRYDGMVGF	Not found	Not found	**	
ERFNVPVSL	Not found	*	**	
ERIPDQLGY	Not found	Not found	**	
GRVNLNVLR	Not found	ND	(s)[d]	
11-mers (n = 1)	11-mers (n = 0)	11-mers (n = 0)	11-mers (n = 1)	
NRFSTPEQAAK	Not found	Not found	**	

[a] Peptides labeled with double asterisks (**) were directly sequenced from the corresponding HLA-B27 variant. Peptides labeled with a single asterisk (*) were assigned to the corresponding variant on the basis of the presence of an ion peak with the same chromatographic retention time and m/z ratio in the MALDI-TOF MS spectrum.

[b] Previously reported and new sequences of B*2705 ligands are specified. All the sequences from the mutants were obtained in this study.

[c] This peptide was directly sequenced by nanoelectrospray MS/MS from B*2705, T163, and T163S167, but it was not detected by MALDI-TOF in any of the three variants.

[d] This peptide was directly sequenced by nanoelectrospray MS/MS from T163S167, but it was not detected by MALDI-TOF in this variant.

[e] ND indicates not found by MALDI-TOF and not reanalyzed by MALDI-TOF/TOF.

majority of B*2705 ligands to this mutant. The double mutation has a larger effect, but it is still compatible with binding of a substantial fraction of the B*2705-bound peptide repertoire.

*E163 in B*2705 is dispensable for ligands with basic P1 residues*

To determine the effect of the T163 mutation on residue usage at P1, the presence of two peptide sets in the T163-bound peptide pool was analyzed. Set 1 (Table II) consisted of 54 B*2705 ligands of known sequence that give significant signals in the MALDI-TOF MS spectrum (33) and included some prominent components of the B*2705-bound peptide pool. Set 2 (Table III) consisted of 41 ion peaks of diverse intensity, or even not detected by MALDI-TOF MS, which were sequenced in this study from B*2705, the mutants, or both. Some of these peptides were already known B*2705 ligands (12, 31). The predominant P1 residues among set 1 peptides were basic (R, 26.4%; K, 9.4%) and small (G, 15.1%; A, 11.3%; S, 11.3%). These frequencies reflected approximately those of a wider series of B*2705 ligands (12). Among the peptides in set 2, the most notorious difference was the almost 5-fold lower frequency of R1 among the B*2705 ligands, relative to those in set 1 (Fig. 2*A*).

Of a total of 90 B*2705 ligands in both peptide sets, including all 54 from set 1 and 36 from set 2, 26.7% had R or K and

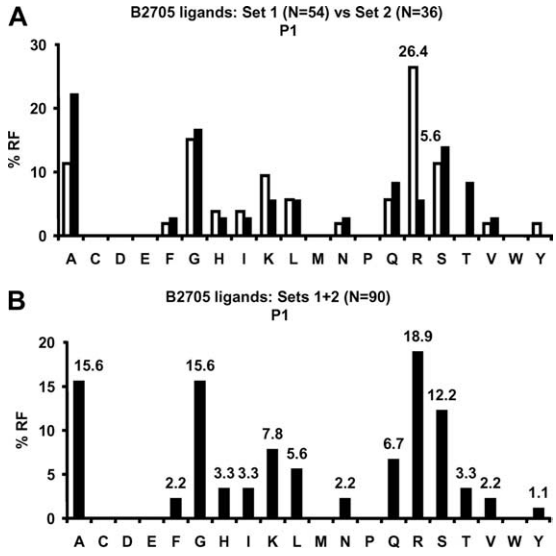


FIGURE 2. P1 residue frequencies among HLA-B*2705 ligands. *A*, Comparison of the percentage frequency of each amino acid residue (% RF) at P1 between the 54 B*2705 ligands from set 1 (open bars) and the 36 B*2705 ligands from set 2 (filled bars). *B*, Percentage RF at P1 among all 90 B*2705 ligands from sets 1 and 2.

43.4% had small residues (G, A, S) at P1 (Fig. 2*B*). Only four (4.4%) of these B*2705 ligands, all from set 2, were not found in T163, and none of them had R1 or K1 (Table III). Thus, the E163 residue in B*2705 is dispensable for binding peptides with basic P1 residues.

To examine the possibility that the T163 mutation could have a quantitative effect on the peptides with basic P1 residues bound to the mutant, the amounts of 15 shared ligands with R1 recovered from B*2705 and T163 were estimated from the corresponding ion peak intensities in the MALDI-TOF spectra. This is only an approximate estimation since MALDI-TOF MS is not a quantitative technique and peak intensities may be influenced by multiple factors, including the composition of the sample. With this limitation in mind, a pattern of relative peptide yields consistent with a variant-related bias was not observed (Table IV). The intensity of individual ion peaks from T163, relative to those from B*2705, ranged from ~25% to 259%, suggesting that peptides with R1 bound the mutant with similar efficiency as did the wild type, with some peptide-to-peptide differences. The only exception was RRYLENGKETLQR, whose yield from the mutant was <1% relative to the wild type. This peptide is derived from residues 169–181 of the HLA-B27 HC. The possibility that the T163 mutation may influence the generation of this ligand seems unlikely, since the related peptide RRYLENGKETL was recovered from the mutant and the wild type with similar yields (Table IV).

*Most of the B*2705 ligands with basic P1 residues are presented by T163S167*

The presence of the set 1 and set 2 peptides in the T163S167-bound peptide pool was analyzed by MALDI-TOF MS and their sequences were determined by nano-ESI MS/MS as above. The lower expression of the T163S167 mutant resulted in lower peptide yields. To account for the possibility that the lack of detection of B*2705 ligands in the double mutant was due simply to low recovery, those ion peaks found in B*2705 but not in the double mutant upon MALDI-TOF analysis were reanalyzed using a significantly more sensitive MALDI-TOF/TOF instrument. Only when a given ion peak from the double mutant was not detected after this second analysis was it assigned as a differentially bound peptide.

Table IV. *Ion peak intensities of peptides with R1 in T163 and B*2705*

Peptide	Fraction No.	M+H	Total Intensity	Percentage T163[a]
RRFFPYYV				
B*2705	183–185	1146.41	5,958,270,976	49.1
T163	183–184	1146.41	2,927,624,192	
RRDFNHINV				
B*2705	146–148	1170.54	6,150,422,528	54.5
T163	146–147	1170.41	3,352,428,544	
RRFFPYYVY				
B*2705	188–190	1310.61	7,997,227,008	103.3
T163	188–189	1310.43	8,262,516,736	
RRFGDKLNF				
B*2705	161–163	1152.62	7,589,593,088	50.2
T163	161–162	1152.59	3,806,855,168	
RRKDGVFLY				
B*2705	164–165	1153.43	3,750,232,064	109.7
T163	164–165	1153.29	4,114,350,080	
RRYQKSTEL				
B*2705	119–122	1180.67	9,489,350,656	99.5
T163	120–122	1180.58	9,442,426,880	
RRFVNVVPTF				
B*2705	185–188	1234.57	7,074,611,200	49.5
T163	185	1234.54	3,501,195,264	
RRISGVDRYY				
B*2705	138–143	1284.34	5,627,740,160	259.0
T163	138–143	1284.31	14,577,631,232	
RRLALFPGVA				
B*2705	188–189	1099.51	4,111,073,280	172.7
T163	187–189	1099.68	7,101,218,816	
RRLQIEDFEA				
B*2705	167–168	1276.55	2,434,269,184	46.3
T163	167	1276.49	1,126,170,624	
RRFVNVVPTFG				
B*2705	180–181	1291.51	4,186,308,608	118.0
T163	179–180	1291.88	4,940,365,824	
RRLQIEDFEAR				
B*2705	159–160	1432.58	4,241,489,920	25.4
T163	159	1432.58	1,078,984,704	
RRYLENGKETL				
B*2705	139–143	1378.57	6,870,892,544	72.0
T163	138–140	1378.38	4,948,426,752	
RRFVNVVPTFGK				
B*2705	170–173	1419.94	10,785,914,880	55.0
T163	169–172	1419.64	5,930,352,640	
RRYLENGKETLQR				
B*2705	129–130	1662.90	4,489,609,216	0.8
T163	129	1662.82	36,175,872	

[a] Percentage intensity of individual ion peaks from T163, relative to B*2705.

Of the 54 B*2705 ligands in set 1, 45 (83%) were found in the double mutant (Table II). These shared ligands included all 5 peptides with K1 and 11 of the 15 peptides (73.3%) with R1. These results indicate that, despite the severe disruption of the A pocket environment in T163S167, this mutant still retains a significant capacity to bind peptides with basic P1 residues.

Of the 41 peptides in set 2 (Table III), 25 (61%) were found in both peptide pools, 11 (26.8%) were found in B*2705 but not in the double mutant, and 5 (12.2%) were found in the double mutant but not in B*2705. All four peptides with R1 or K1 were found among the shared ligands. Small P1 residues (G, A, S) showed a lower joint frequency among shared, relative to B*2705-specific ligands in this set (40% and 81.9%, respectively). This was due mainly to the higher frequency of G1 and S1 among B*2705-specific ligands compared with the shared subset (36.4% vs 8% and 27.3% vs 8%, respectively). These differences did not reach statistical significance, although the increased frequency of G1 among B*2705-specific ligands approached this limit (Fig. 3*A*).

Of the five peptides sequenced from the double mutant but not found in B*2705, four showed acidic or N residues at P1, which are absent or found with low frequency among B*2705 ligands

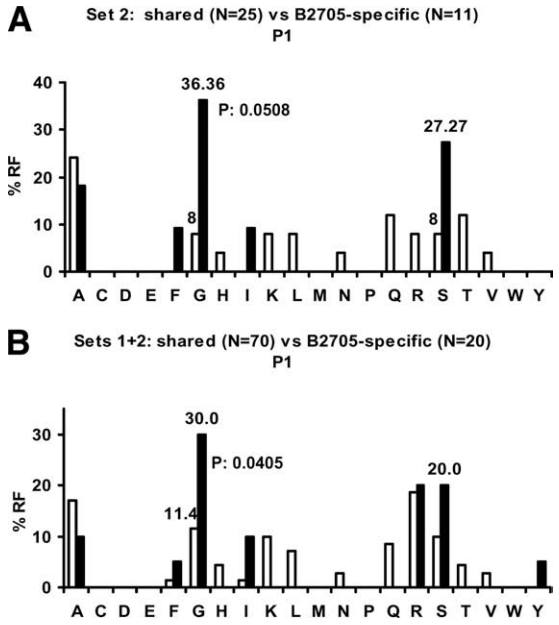


FIGURE 3. Comparison of P1 residue frequencies among HLA-B*2705 ligands shared or not with the T163S167 mutant. *A*, Comparison of the percentage frequency of each amino acid residue (% RF) at P1 between the 25 B*2705 ligands from set 2 shared with T163S167 (open bars) and the 11 peptides from the same set found only in the wild type (filled bars). *B*, Comparison of the percentage RF at P1 between the 70 B*2705 ligands from sets 1 and 2 shared with T163S167 (open bars) and the 20 peptides from both sets found in the wild type but not in the double mutant (filled bars).

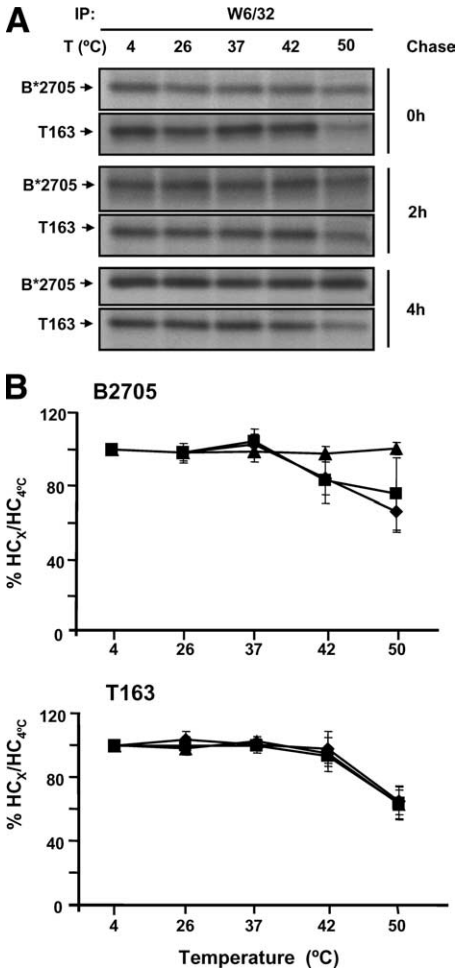


FIGURE 4. Thermostability of B*2705 and T163. *A*, C1R transfectants were labeled for 15 min and chased for the indicated times. Equal aliquots of the lysates were kept at 4°C or heated at the indicated temperatures for 1 h before immunoprecipitation with W6/32, separated by SDS-PAGE, and analyzed by fluorography. The left arrows indicate the HC. A representative experiment is shown. *B*, The percentage of W6/32-reactive HLA/peptide complexes recovered after heating at 0 (♦), 2 (■), or 4 h (▲) chase was plotted as the intensity value of the class I HC at any given temperature (HC_X) relative to that at 4°C ($HC_{4°C}$). Data are means ± SD of four (B*2705) or five (T163) experiments.

(12). The fifth peptide, GRVNLNVLR, was not detected by MALDI-TOF MS in any of the variants. Thus, although it was not found in the wild type by nanoelectrospray MS/MS, it is difficult to rule out its presence in the B*2705-bound pool.

Overall, of 90 B*2705 ligands in both sets 1 and 2, 70 (77.8%) were also found in the double mutant and 20 (22.2%) were not. The frequency of peptides with basic P1 residues was similar or higher among the subset of shared ligands (R1, 18.6%; K1, 10%) than among the B2705-specific ones (R1, 20%; K1, 0%), confirming that binding of peptides with these motifs is largely maintained in the absence of both E163 and W167. As noted in the analysis of set 2, the overall frequency of small P1 residues (G, A, S) among shared ligands was smaller than among differentially bound ones (38.5% and 60%, respectively). The most notorious difference was the increased frequency of G1 among B*2705-specific ligands, relative to shared ones (30% vs 11.4%), which reached statistical significance (Fig. 3*B*). This suggests that the absence of W167 may disfavor binding of peptides with G1.

The T163 mutation decreases the thermostability of HLA-B27

The effect of the T163 mutation on the stability of HLA-B27 was analyzed by thermostability analyses in pulse-chase experiments. Lysates from pulse-labeled cells were incubated at various temperatures and immunoprecipitated with the conformation-sensitive mAb W6/32 at different chase times. The thermostability of the molecule was measured as the percentage of HLA-B27 HC immunoprecipitated at a given temperature, relative to the amount recovered at 4°C (Fig. 4). The double mutant could not be analyzed in these experiments because its low expression level precluded a clear distinction of the T163S167 HC from that of the endogenous class I molecules of C1R cells. Both B*2705 and T163 showed a significant stability in these experiments, with ~70% recovery of W6/32-precipitated material at 50°C at 0 h chase time. B*2705 increased its stability with time, reaching 100% thermostability at 50°C at 4 h chase time. The T163 mutant failed to do so, suggesting that this mutation decreased the overall stability of HLA-B27/peptide complexes. The time-dependent increase in the stability of B*2705 occurred equally in the presence of brefeldin A (data not shown), indicating that it takes place in the endoplasmic reticulum.

Discussion

The specificity of Ag presentation by HLA class I molecules is determined essentially by the structural complementarity of peptide ligands with the Ag-binding groove and, in particular, with the side chain-binding pockets (34). For HLA-B27 ligands, the primary anchor positions are P2, which is restricted almost exclusively to R, and the C-terminal position, which for B*2705 is predominantly restricted to basic, aliphatic, and aromatic residues. Other positions, particularly P1, P3, and P7, make a significant

contribution to peptide binding, but they show a larger allowance for different residues. They are called secondary anchor positions. Basic and small P1 residues jointly account for 69.4% of the nonamers and 82% of the decamers in a published registry of B*2705 ligands (12). This restriction, together with substantial interactions of the peptidic R1 side chain in the A pocket of HLA-B27 (16), seemed to suggest that the frequency of basic P1 residues, particularly R, among HLA-B27 ligands is determined by structural complementarity, and that the peptidic R1 is a significant anchor residue, as also suggested by in vitro binding assays with poly-A analogs (15).

In addition to the binding features, Ag presentation depends on the availability of appropriate ligands. This, in turn, is determined by the efficiency of their generation along the Ag-processing pathway, and by the resistance of the peptide to cytosolic degradation. It has been estimated that ~2000 protein molecules must be processed for each peptide copy that binds to an MHC class I molecule (35, 36). This is due to the extraordinary amino-peptidase activity in the cytosol to which peptides are exposed before binding to TAP, which results in a very short half-life of peptides in vivo (37). Thus, peptides resistant to amino-peptidase-mediated degradation would need lower thresholds for their generation to reach TAP, and they would compete advantageously for binding to MHC class I molecules. A consequence of this would be that such ligands might be presented and act as Ags even if they are produced in relatively low amounts or from nonabundant proteins in the cell. For this reason, the demonstration that peptides with N-terminal dibasic sequences were particularly resistant to cytosolic amino peptidases (17) was potentially very relevant for the immunological properties of HLA-B27 and a few other allotypes with specificity for R2 residues. On the basis of these observations it was proposed that the high frequency of HLA-B27 ligands with basic P1 residues would be significantly determined by their increased cytosolic stability and availability, rather than being a consequence of structural complementarity with the A pocket environment. A problem with this interpretation was that the frequency of R2 among ligands of non-B27 HLA class I molecules is not necessarily associated with a correspondingly high frequency of basic P1 residues, suggesting a critical role of the structural complementarity between the MHC protein and the peptide ligand in determining P1 preferences. For instance, of 17 HLA-B14 (B*1402 or B*1403) ligands with R2 from one study (32), none showed R or K at P1. Of 39 B*3901 ligands with R2 or K2 listed in the SYFPEITHI database (www.syfpeithi.de), none showed R or K at P1. Similarly, none of nine B*3909 reported ligands with R2 (38) showed a basic P1 residue. Finally, of five natural ligands of Gogo*0101 with R2, only one had K1 (39). All three HLA-B14, HLA-B39, and Gogo*0101 have T instead of E at position 163, but they have R62 and W167. Thus, in the absence of an appropriate structure of the peptide binding site, it seemed that the increased cytosolic stability of peptides with N-terminal dibasic sequences was not sufficient for their prevalent presentation by HLA class I molecules.

Our study sought to determine the role of the A pocket in HLA-B27 in the remarkable restriction of this molecule for basic and small P1 residues. In particular, the importance of residues E163 and W167 in determining the preference of HLA-B27 for R1 was investigated by mutating either E163 alone or both E163 and W167 to other naturally occurring polymorphisms at these positions. Contrary to our expectations, the T163 change had a limited effect on P1 residue usage, including R1. It is possible that, in the presence of this mutation, R1 residues can still bind with moderate loss of peptide stability. Support for this explanation can be found, for instance, in the crystal structure of H-K^b-VSV8 peptide complex (40), in which the peptidic R1 adopts an alternative binding mode with R62 and W167, possibly imposed by the presence of T163 in H-K^b (16). Thus, the lack of basic P1 residues among the peptide ligands of HLA-B14 or HLA-B39 cannot be simply due to their lack of E163.

Although the cumulative effect of the T163 and S167 mutations had a more significant effect on peptide specificity, the impact of the presumably severe disruption of the A pocket surface induced by these mutations on the frequency of basic P1 residues was much smaller than might have been anticipated. These results strongly suggest that the native structure of the A pocket surface in HLA-B27 is not a major determinant of the preference of this molecule for basic P1 residues. It is conceivable that peptide anchoring to HLA-B27 mediated by positions other than P1 may provide sufficient stability even if A pocket interactions are severely disrupted, obviating the requirement for an optimal anchor residue at P1. It is also possible that, even in the double mutant, basic P1 residues might adopt an alternative binding mode that provides significant stabilization of the P1 side chain. Although this is difficult to rule out in the absence of an x-ray diffraction model of the mutant, it would appear less likely, due to the drastic alteration of the A pocket surface induced by the double mutation. Note that the overall contribution of the P1 side chain to binding is not only a consequence of the alteration of nonbinding interaction patterns, but also of the changes induced by the mutations in the conformational entropy of binding, a component whose importance has been recently highlighted both in the process of molecular recognition by proteins (41) and for P1 and other anchor residues of MHC class I molecules (42). Since loss of conformational entropy counteracts the favorable contribution of nonbonding interactions, the mutations introduced in the A pocket surface of HLA-B27 may not only disrupt such contacts, which contribute to the enthalpic component of the Gibbs free energy of P1 interaction, but they also decrease the loss of conformational entropy by allowing a larger conformational space of the P1 side chain, thus partially compensating for the disrupting effect on the enthalpic component. We cannot formally rule out this possibility, but it seems unlikely that entropic effects alone might explain binding of peptides with basic P1 residues to pocket A surface mutants with significantly altered nonbonding interactions.

If HLA-B27 ligands can bypass the requirement of the canonic pocket A interactions to bind peptides with R1, this would support the claims by Herberts et al. (17) that the predominance of HLA-B27 ligands with R or K at P1 is favored by their resistance to amino-peptidase-mediated degradation in the cytosol, even if this feature is not reflected in a similar predominance among other HLA class I molecules, for which pocket A interactions might possibly be more determinant.

The only statistically significant effect induced by the double mutation on P1 residue frequencies concerned the increased number of peptides with G1 among those bound by B*2705 but not by T163S167. The high frequency of HLA-B27 ligands with G1 has not yet been satisfactorily explained, since this residue is thermodynamically disfavored compared with R1 (16), and the susceptibility of peptides with N-terminal GR sequences to cytosolic degradation has not been determined. However x-ray diffraction analyses showed that, in the presence of a peptidic G1 residue, the side chain of W167 in HLA-B27 folds down onto the peptide N terminus, which binds in the A pocket through a conserved network of hydrogen bonds, shielding it from the solvent (16, 43). To the extent to which this conformation may provide some stability, compensating in part for the loss of interactions with the peptidic R1 side chain, the absence of W167 in the double mutant may disfavor binding of peptides with G1, which would explain our

results. Alternatively, an increased allowance of the double mutant for P1 residues less favored in B*2705 might also result in a decreased frequency of G1 among T163S167-bound peptides.

Thermostability analyses showed that both the wild type and the T163 mutant bind highly stable peptide repertoires. Some time-dependent increase of the global thermostability was observed for B*2705, which occurred equally in the presence of brefeldin A (data not shown). Therefore, the effect is probably due to the tapasin-mediated optimization of the B*2705-bound peptide repertoire in the endoplasmic reticulum, as reported in a previous study (44). This optimization was not observed in the T163 mutant. Thus, although peptide specificity was little affected by this mutation, it induced a certain loss of stability on its bound peptide repertoire. Although the low expression level of the T163S167 mutant precluded a reliable estimation of its thermostability, the larger cell surface dissociation observed by flow cytometry suggests that the overall stability of the T163S167/peptide complexes might be lower than for the wild type or the T163 mutant. If so, although the double mutation has a limited effect on the peptide repertoire and, in particular, on the frequency of peptides with basic P1 residues, the molecular stability of the complexes might be affected.

In conclusion, our results indicate that the native structure of the A pocket surface in HLA-B27 is largely dispensable for binding peptides with N-terminal basic residues to this molecule, although W167 may be significant in determining the high frequency of G1 among HLA-B27 ligands. That peptides with R or K at P1 bind to B27 mutants with a severely altered A pocket surface strongly suggests that HLA-B27 ligands are well stabilized in the peptide binding groove by other contacts. Therefore, in agreement with the suggestion by Herberts et al. (17), the high frequency of HLA-B27 ligands with N-terminal basic residues may be more determined by their cytosolic stability than by the canonic interactions of these residues with those at the A pocket surface. The potential consequences of this fact for the physiological function of HLA-B27 as an Ag-presenting molecule were properly noted in that previous study: HLA-B27 would present peptides generated in lower amounts than required for presentation by other MHC molecules, because such peptides would be less destroyed in the cytosol. The relevance of this feature for the pathogenetic role of HLA-B27 in spondyloarthropathies is less clear, since both disease-associated (i.e., B*2705) and non-disease-associated subtypes (i.e., B*2706) have identical structures in the A pocket and its vicinity, as well as similar high frequencies of ligands with basic P1 residues (12). However, individual peptides may be pathogenetically relevant in the context of only some subtypes, due to subtype-specific binding or antigenicity.

Acknowledgments

We thank the staff of the Proteomics facilities at the Centro de Biología Molecular "Severo Ochoa" and Centro Nacional de Biotecnología, Madrid (both members of the ProteoRed network) for help in mass spectrometry. Special thanks are given to our colleagues Miguel Marcilla for help in the anaysis of MS/MS data and to Andreas Ziegler and Barbara Uchanska-Ziegler (Institut fur Immungenetik, Charite-Universitatsmedizin Berlin) for critical reading of the manuscript and valuable insights.

Disclosures

The authors have no financial conflicts of interest.

References

1. Brewerton, D. A., F. D. Hart, A. Nicholls, M. Caffrey, D. C. James, and R. D. Sturrock. 1973. Ankylosing spondylitis and HL-A 27. *Lancet* 1: 904–907.
2. Brewerton, D. A., M. Caffrey, A. Nicholls, D. Walters, J. K. Oates, and D. C. James. 1973. Reiter's disease and HL-A 27. *Lancet* 2: 996–998.
3. Benjamin, R., and P. Parham. 1990. Guilt by association: HLA-B27 and ankylosing spondylitis. *Immunol. Today* 11: 137–142.
4. Marcilla, M., and J. A. Lopez de Castro. 2008. Peptides: the cornerstone of HLA-B27 biology and pathogenetic role in spondyloarthritis. *Tissue Antigens* 71: 495–506.
5. Gomard, E., M. Sitbon, A. Toubert, B. Begue, and J. P. Levy. 1984. HLA-B27, a dominant restricting element in antiviral responses? *Immunogenetics* 20: 197–204.
6. McNeil, A. J., P. L. Yap, S. M. Gore, R. P. Brettle, M. McColl, R. Wyld, S. Davidson, R. Weightman, A. M. Richardson, and J. R. Robertson. 1996. Association of HLA types A1–B8-DR3 and B27 with rapid and slow progression of HIV disease. *Q. J. Med.* 89: 177–185.
7. Kelleher, A. D., C. Long, E. C. Holmes, R. L. Allen, J. Wilson, C. Conlon, C. Workman, S. Shaunak, K. Olson, P. Goulder, et al. 2001. Clustered mutations in HIV-1 gag are consistently required for escape from HLA-B27-restricted cytotoxic T lymphocyte responses. *J. Exp. Med.* 193: 375–386.
8. den Uyl, D., I. E. van der Horst-Bruinsma, and M. van Agtmael. 2004. Progression of HIV to AIDS: a protective role for HLA-B27? *AIDS Rev.* 6: 89–96.
9. Schneidewind, A., M. A. Brockman, J. Sidney, Y. E. Wang, H. Chen, T. J. Suscovich, B. Li, R. I. Adam, R. L. Allgaier, B. R. Mothe, et al. 2008. Structural and functional constraints limit options for cytotoxic T-lymphocyte escape in the immunodominant HLA-B27-restricted epitope in human immunodeficiency virus type 1 capsid. *J. Virol.* 82: 5594–5605.
10. Neumann-Haefelin, C., S. McKiernan, S. Ward, S. Viazov, H. C. Spangenberg, T. Killinger, T. F. Baumert, N. Nazarova, I. Sheridan, O. Pybus, et al. 2006. Dominant influence of an HLA-B27 restricted $CD8^+$ T cell response in mediating HCV clearance and evolution. *Hepatology* 43: 563–572.
11. Jardetzky, T. S., W. S. Lane, R. A. Robinson, D. R. Madden, and D. C. Wiley. 1991. Identification of self peptides bound to purified HLA-B27. *Nature* 353: 326–329.
12. Lopez de Castro, J. A., I. Alvarez, M. Marcilla, A. Paradela, M. Ramos, L. Sesma, and M. Vazquez. 2004. HLA-B27: a registry of constitutive peptide ligands. *Tissue Antigens* 63: 424–445.
13. Madden, D. R., J. C. Gorga, J. L. Strominger, and D. C. Wiley. 1992. The three-dimensional structure of HLA-B27 at 2.1 Å resolution suggests a general mechanism for tight peptide binding to MHC. *Cell* 70: 1035–1048.
14. Villadangos, J. A., B. Galocha, F. Garcia, J. P. Albar, and J. A. Lopez de Castro. 1995. Modulation of peptide binding by HLA-B27 polymorphism in pockets A and B, and peptide specificity of B*2703. *Eur. J. Immunol.* 25: 2370–2377.
15. Lamas, J. R., A. Paradela, F. Roncal, and J. A. Lopez de Castro. 1999. Modulation at multiple anchor positions of the peptide specificity of HLA-B27 subtypes differentially associated with ankylosing spondylitis. *Arthritis Rheum.* 42: 1975–1985.
16. Hillig, R. C., M. Huelsmeyer, W. Saenger, K. Welfle, R. Misselwitz, H. Welfle, C. Kozerski, A. Volz, B. Uchanska-Ziegler, and A. Ziegler. 2004. Thermodynamic and structural analysis of peptide-and allele-dependent properties of two HLA-B27 subtypes exhibiting differential disease association. *J. Biol. Chem.* 279: 652–663.
17. Herberts, C. A., J. J. Neijssen, J. de Haan, L. Janssen, J. W. Drijfhout, E. A. Reits, and J. J. Neefjes. 2006. Cutting edge: HLA-B27 acquires many N-terminal dibasic peptides: coupling cytosolic peptide stability to antigen presentation. *J. Immunol.* 176: 2697–2701.
18. Zemmour, J., A. M. Little, D. J. Schendel, and P. Parham. 1992. The HLA-A,B "negative" mutant cell line C1R expresses a novel HLA-B35 allele, which also has a point mutation in the translation initiation codon. *J. Immunol.* 148: 1941–1948.
19. Calvo, V., S. Rojo, D. Lopez, B. Galocha, and J. A. Lopez de Castro. 1990. Structure and diversity of HLA-B27-specific T cell epitopes: analysis with site-directed mutants mimicking HLA-B27 subtype polymorphism. *J. Immunol.* 144: 4038–4045.
20. Villadangos, J. A., B. Galocha, D. Lopez, V. Calvo, and J. A. Lopez de Castro. 1992. Role of binding pockets for amino-terminal peptide residues in HLA-B27 allorecognition. *J. Immunol.* 149: 505–510.
21. Barnstable, C. J., W. F. Bodmer, G. Brown, G. Galfre, C. Milstein, A. F. Williams, and A. Ziegler. 1978. Production of monoclonal antibodies to group A erythrocytes, HLA and other human cell surface antigens: new tools for genetic analysis. *Cell* 14: 9–20.
22. Ellis, S. A., C. Taylor, and A. McMichael. 1982. Recognition of HLA-B27 and related antigens by a monoclonal antibody. *Hum. Immunol.* 5: 49–59.
23. Stam, N. J., H. Spits, and H. L. Ploegh. 1986. Monoclonal antibodies raised against denatured HLA-B locus heavy chains permit biochemical characterization of certain HLA-C locus products. *J. Immunol.* 137: 2299–2306.
24. Vázquez, M. N., and Lopez de Castro, J. A. 2005. Similar cell surface expression of β_2-microglobulin-free heavy chains by HLA-B27 subtypes differentially associated with ankylosing spondylitis. *Arthritis Rheum.* 52: 3290–3299.
25. Paradela, A., M. Garcia-Peydro, J. Vazquez, D. Rognan, and J. A. Lopez de Castro. 1998. The same natural ligand is involved in allorecognition of multiple HLA-B27 subtypes by a single T cell clone: role of peptide and the MHC molecule in alloreactivity. *J. Immunol.* 161: 5481–5490.
26. Paradela, A., I. Alvarez, M. Garcia-Peydro, L. Sesma, M. Ramos, J. Vazquez, and J. A. Lopez de Castro. 2000. Limited diversity of peptides related to an alloreactive T cell epitope in the HLA-B27-bound peptide repertoire results from restrictions at multiple steps along the processing-loading pathway. *J. Immunol.* 164: 329–337.
27. Marina, A., M. A. Garcia, J. P. Albar, J. Yague, J. A. Lopez de Castro, and J. Vazquez. 1999. High-sensitivity analysis and sequencing of peptides and proteins by quadrupole ion trap mass spectrometry. *J. Mass Spectrom.* 34: 17–27.

28. Merino, E., B. Galocha, M. Vázquez, and J. A. López de Castro. 2008. Disparate folding and stability of the ankylosing spondylitis-associated HLA-B*1403 and B*2705 proteins. *Arthritis. Rheum*. 58: 3693–3703.
29. Sesma, L., V. Montserrat, J. R. Lamas, A. Marina, J. Vazquez, and J. A. Lopez de Castro. 2002. The peptide repertoires of HLA-B27 subtypes differentially associated to spondyloarthropathy (B*2704 and B*2706) differ by specific changes at three anchor positions. *J. Biol. Chem.* 277: 16744–16749.
30. Ramos, M., A. Paradela, M. Vazquez, A. Marina, J. Vazquez, and J. A. Lopez de Castro. 2002. Differential association of HLA-B*2705 and B*2709 to ankylosing spondylitis correlates with limited peptide subsets but not with altered cell surface stability. *J. Biol. Chem.* 277: 28749–28756.
31. Gomez, P., V. Montserrat, M. Marcilla, A. Paradela, and J. A. López de Castro. 2006. B*2707 differs in peptide specificity from B*2705 and B*2704 as much as from HLA-B27 subtypes not associated to spondyloarthritis. *Eur. J. Immunol.* 36: 1867–1881.
32. Merino, E., V. Montserrat, A. Paradela, and J. A. Lopez de Castro. 2005. Two HLA-B14 subtypes (B*1402 and B*1403) differentially associated with ankylosing spondylitis differ substantially in peptide specificity, but have limited peptide and T-cell epitope sharing with HLA-B27. *J. Biol. Chem.* 280: 35868–35880.
33. Marcilla, M., J. J. Cragnolini, and J. A. Lopez de Castro. 2007. Proteasome-independent HLA-B27 ligands arise mainly from small basic proteins. *Mol. Cell. Proteomics* 6: 923–938.
34. Garrett, T. P., M. A. Saper, P. J. Bjorkman, J. L. Strominger, and D. C. Wiley. 1989. Specificity pockets for the side chains of peptide antigens in HLA-Aw68. *Nature* 342: 692–696.
35. Yewdell, J. W., E. Reits, and J. Neefjes. 2003. Making sense of mass destruction: quantitating MHC class I antigen presentation. *Nat. Rev. Immunol.* 3: 952–961.
36. Princiotta, M. F., D. Finzi, S. B. Qian, J. Gibbs, S. Schuchmann, F. Buttgereit, J. R. Bennink, and J. W. Yewdell. 2003. Quantitating protein synthesis, degradation, and endogenous antigen processing. *Immunity* 18: 343–354.
37. Reits, E., A. Griekspoor, J. Neijssen, T. Groothuis, K. Jalink, P. van Veelen, H. Janssen, J. Calafat, J. W. Drijfhout, and J. Neefjes. 2003. Peptide diffusion, protection, and degradation in nuclear and cytoplasmic compartments before antigen presentation by MHC class I. *Immunity* 18: 97–108.
38. Yague, J., M. Ramos, J. Vazquez, A. Marina, J. P. Albar, and J. A. Lopez de Castro. 1999. The South Amerindian allotype HLA-B*3909 has the largest known similarity in peptide specificity and common natural ligands with HLA- B27. *Tissue Antigens* 53: 227–236.
39. Urvater, J. A., H. Hickman, J. L. Dzuris, K. Prilliman, T. M. Allen, K. J. Schwartz, D. Lorentzen, C. Shufflebotham, E. J. Collins, D. L. Neiffer, et al. 2001. Gorillas with spondyloarthropathies express an MHC class I molecule with only limited sequence similarity to HLA-B27 that binds peptides with arginine at P2. *J. Immunol* 166: 3334–3344.
40. Fremont, D. H., M. Matsumura, E. A. Stura, P. A. Peterson, and I. A. Wilson. 1992. Crystal structures of two viral peptides in complex with murine MHC class I H-2Kb. *Science* 257: 919–927.
41. Frederick, K. K., M. S. Marlow, K. G. Valentine, and A. J. Wand. 2007. Conformational entropy in molecular recognition by proteins. *Nature* 448: 325–329.
42. Zhou, P., X. Chen, and Z. Shang. 2009. Side-chain conformational space analysis (SCSA): a multi conformation-based QSAR approach for modeling and prediction of protein-peptide binding affinities. *J. Comput. Aided Mol. Des. In press.*
43. Hülsmeyer, M., R. C. Hilling, A. Volz, M. Rühl, W. Schröder, W. Saenger, A. Ziegler, and B. Uchanska-Ziegler. 2002. HLA-B27 subtypes differentially associated with disease exhibit subtle structural alterations. *J. Biol. Chem.* 277: 47844–47853.
44. Williams, A. P., C. A. Peh, A. W. Purcell, J. McCluskey, and T. Elliott. 2002. Optimization of the MHC class I peptide cargo is dependent on tapasin. *Immunity* 16: 509–520.

DIVERSITY OF NATURAL SELF-DERIVED LIGANDS PRESENTED BY DIFFERENT HLA CLASS I MOLECULES IN TRANSPORTER ANTIGEN PROCESSING-DEFICIENT CELLS

Elena Lorente [1], Susana Infantes [1], Eilon Barnea [2], Ilan Beer [3], Alejandro Barriga [1], Noel García-Medel [4], Fátima Lasala [5], Mercedes Jiménez [5], Arie Admon [2], Daniel López [1,5,*]

[1] Unidad de Procesamiento Antigénico and [5] Unidad de Proteómica, Centro Nacional de Microbiología, Instituto de Salud Carlos III, 28220 Majadahonda (Madrid), Spain. [2] Department of Biology, Technion-Israel Institute of Technology, 32000 Haifa, Israel. [3] IBM Haifa Research Lab, 31905 Haifa, Israel. [4] Centro de Biología Molecular Severo Ochoa, CSIC/Universidad Autónoma de Madrid, 28049 Madrid, Spain.

Running Title: Natural HLA peptidome in TAP-deficient cells

* Correspondence to: Dr. Daniel López. Unidad de Procesamiento Antigénico. Centro Nacional de Microbiología. Instituto de Salud Carlos III. 28220 Majadahonda (Madrid), Spain. Tel: +34 91 822 37 08, FAX: +34 91 509 79 19, E-mail address: dlopez@isciii.es.

The authors declare no conflicts of interest.

ABBREVIATIONS

Ab, Antibody;

CTL, Cytotoxic T lymphocyte;

EBV, Epstein-Barr virus;

ERAAP, Endoplasmic Reticulum aminopeptidase associated with antigen processing;

GRAVY, Grand average of hydropathicity;

HLA, Human leukocyte antigen;

MHC, Major histocompatibility complex;

TAP, Transporter associated with antigen processing;

VACV, Vaccinia virus;

SUMMARY

The transporter associated with antigen processing (TAP) translocates the cytosol-derived proteolytic peptides to the endoplasmic reticulum lumen where they complex with nascent human leukocyte antigen (HLA) class I molecules. Non-functional TAP complexes and viral or tumoral blocking of these transporters leads to reduced HLA class I surface expression and a drastic change in the available peptide repertoire. Using mass spectrometry to analyze complex human leukocyte antigen HLA-bound peptide pools isolated from large numbers of TAP-deficient cells, we identified 337 TAP-independent ligands naturally presented by four different HLA-A, -B, and -C class I molecules with very different TAP dependency from the same cell line. The repertoire of TAP-independent peptides examined favored increased peptide lengths and a lack of strict binding motifs for all four HLA class I molecules studied. The TAP-independent peptidome arose from 183 parental proteins, the majority of which yielded one HLA ligand. In contrast, TAP-independent antigen processing of very few cellular proteins generated multiple HLA ligands. In addition, the secretory vesicle-like organelles were the main source of parental proteins for TAP-independent HLA ligands. Finally, a predominant endoproteolytic peptidase specificity for Arg/Lys or Leu/Phe residues in the P_1 position of the scissile bond was found for the TAP-independent ligands. These data draw a new and intricate picture of TAP-independent pathways.

INTRODUCTION

Proteolysis, by the proteasome and other cytosolic proteases, of both newly synthesized proteins and the mature cell proteome continuously generates short peptides that are transported into the endoplasmic reticulum (ER) by the transporter associated with antigen processing (TAP) (1). These peptides are assembled with a nascent HLA class I heavy chain and β2-microglobulin to generate stable HLA/peptide complexes that are exported to the cell membrane and subjected to cytotoxic $CD8^+$ T lymphocyte recognition (reviewed in (2)).

Non-functional TAP complexes, which can be produced by mutations in the TAP gene, have been described in both humans (3) and mice (4). Patients with an HLA class I deficiency have a reduced functional $CD8^+$ population but may appear asymptomatic for long periods of time with only a limited susceptibility to chronic respiratory bacterial infections. Thus, their immune systems must be reasonably efficient, and in addition to different unaltered layered defenses, it remains possible that the reduced cytolytic $CD8^+$ αβ T subpopulation that is specific for TAP-independent antigens may contribute to immune defenses that protect against severe infections in these individuals.

Although TAP-independent viral epitopes are known (reviewed in (5-7)), few studies have analyzed the cellular TAP-independent HLA class I peptide repertoire. TAP-deficient cells have been described as having very limited antigen processing capacity, with predominant proteolytic ER signal peptidase (SPase) activity(8,9). Therefore, is the TAP-independent HLA peptidome so limited in TAP-deficient cells, as suggested by these studies? Of major interest is the identification of self-derived ligands presented in the same cells by several common HLA antigens, which was described previously with very different TAP dependency. Therefore, using a high-throughput immunopeptidomics analysis, we analyzed the TAP-independent HLA peptidome isolated from large numbers of TAP-deficient cells and bound to different

HLA alleles. In this study, we identified more than three hundred TAP-independent ligands bound to different HLA-A, -B, and –C class I molecules.

EXPERIMENTAL PROCEDURES

Cell lines.

T2 is a human cell line that has a large homozygous deletion within the MHC, including both TAP genes and all of the functional class II genes (10-12). In addition, this cell line expresses low levels of HLA class I molecules on the cell surface (13). T2 cells transfected with B*2705 have been previously described (14). Transfected RMA-S cells expressing HLA-B*2705 have also been previously described (15). All cell lines were cultured in RPMI 1640 supplemented with 10% fetal bovine serum (FBS) and 5 μM β-mercaptoethanol.

Synthetic peptides.

Peptides were synthesized in a peptide synthesizer (model 433A; Applied Biosystems, Foster City, CA) and purified by reversed-phase HPLC. The correct molecular mass of each peptide was established with a Reflex IV MALDI-TOF instrument (Brucker-Franzen Analytik, Bremen, Germany), and their correct composition was determined with a Deca XP LCQ mass spectrometer (Thermo Fisher, San Jose, CA).

Isolation of HLA-bound peptides.

HLA-bound peptides were isolated from 4 x 10^{10} healthy or vaccinia (VACV)-WR-infected T2-B27 transfected cells. Cells were lysed in 1% CHAPS (Sigma), 20 mM Tris/HCl buffer, and 150 mM NaCl, pH 7.5, in the presence of a protease inhibitor cocktail (16,17). HLA-peptide complexes were isolated via affinity chromatography of the soluble fraction of cell extracts with the following mAbs, used sequentially: PA2.1 (anti-HLA-A2) (18), ME1 (anti-HLA-B27) (19), and W6/32 (specific for a monomorphic HLA class I determinant) (20) (Supplemental Figure 1). HLA-bound peptides were eluted at room temperature with 0.1% aqueous trifluoroacetic acid (TFA), separated

from the large subunits and concentrated with a Centricon 3 ultrafiltration device (Amicon, Beverly, MA) exactly as previously described (16,17).

Electrospray-Orbitrap mass spectrometry analysis.

Peptide mixtures recovered after the ultra-filtration step were concentrated with Micro-Tip reverse-phase columns (C_{18}, 200 µl, Harvard Apparatus, Holliston, MA) (16). Each C_{18} tip was equilibrated with 80% acetonitrile in 0.1% TFA, washed with 0.1% TFA, and then loaded with the peptide mixture. The tip was then washed with an additional volume of 0.1% TFA and the peptides were eluted with 80% acetonitrile in 0.1% TFA. Peptide samples were then concentrated to approximately 18 µl using vacuum centrifugation (16,17).

HLA class I peptides that had been immunoprecipitated with each HLA-specific mAb were analyzed by µLC-MS/MS using an Orbitrap XL mass spectrometer (Thermo Fisher) fitted with a capillary HPLC (Eksigent, Dublin, CA) (16,17). The peptides were resolved on homemade Reprosil C18 capillary columns (75 micron ID) (21) with a 7%-40% acetonitrile gradient for 2 h in the presence of 0.1% formic acid. The seven most intense masses that exhibited single-, double-, and triple-charge states were selected for fragmentation from each full mass spectrum by collision-induced disintegration.

Database searches.

Sequest 3.31 (Thermo-Fisher) (22) was used for peak-list generation of the µLC-MS/MS data. The peaks were identified using various search engines: Proteome Discoverer 1.0 SP1 (Thermo-Fisher) combining the results of Sequest 3.31 and Bioworks Browser 3.3.1 SP1 (Thermo-Fisher) (22), using the human part of the NCBI database (Jun 2009) including 574,003 proteins. The search was not limited by enzymatic specificity, the peptide tolerance was set to 0.01 Da, and the fragment ion

tolerance was set to 0.5 Da (16,17). Oxidized methionine was searched as a variable modification. Other search criteria were set such that the search was not limited by any methodological bias (selection of individual protein, use of HLA consensus scoring algorithms, etc.).

Identified peptides were selected if the following criteria were met: mass accuracy ≤ 0.005 Da; Sequest Xcorr > 1.5 for singly, > 2.5 for doubly, and > 3.5 for triply charged peptides; ΔCn > 0.1; Proteome Discovered P score > 20; and P(pep) < $1x10^{-2}$ with Bioworks Browser (16,17). When the MS/MS spectra fitted more than one peptide, only the highest scoring peptide was selected. The false-positive rate for peptide identification was set to 2% based on a search of a reversed database. In addition, some synthetic peptides were made, and their MS/MS spectra were used to confirm the assigned sequences.

MHC/peptide stability assays.

The following synthetic peptides were used as controls in complex stability assays: Flu NP (SRYWAIRTR, HLA-B27-restricted) (23), RSV M_{76-84} (SRSALLAQM, HLA-B27-restricted) (16), and C4CON (QYDDAVYLK, HLA-Cw4-restricted) (24). RMA-S B*2705 transfectant cells, a cell line deficient in TAP that expresses low levels of cell surface MHC class I, were incubated at 26°C for 16 h in RPMI 1640 medium supplemented with 10% heat-inactivated FBS. This allows the expression of empty MHC class I molecules (without antigenic peptide) at the cellular membrane that are stable at 26°C but not at 37°C. The cells were washed and incubated for 2 h at 26°C with various concentrations of peptide in the same medium. The cells were maintained at 37°C for an additional 4 h and then collected for flow cytometry. This method allows empty MHC class I molecules to become internalized and can thus discriminate between bound and unbound peptides. MHC expression was measured using 100 µl of hybridoma culture supernatant containing ME1 (anti-HLA-B27) mAb as previously

described (25). Samples were acquired on a FACSCanto flow cytometer (BD Biosciences, San Jose, CA, USA) and analyzed using CellQuest Pro 2.0 software (BD Bioscience). Cells incubated without peptide had peak fluorescence intensities close to background staining with secondary Ab alone. The fluorescence index was calculated at each time point as the ratio of the mean channel fluorescence of the sample to that of the control incubated without peptide. The data are mean values of the three experiments.

RESULTS

Physiological processing generates multiple cellular ligands bound to HLA alleles with distinct TAP dependency in the same human TAP-deficient cell line.

To date, approximately 70 human TAP-independent MHC class I ligands are known (7,9), and are mostly restricted to HLA-A2 and derived by cleavage of signal sequences generated by the SPase complex. Thus, the comparison of TAP-independent peptide pools derived from HLA-A2, HLA-B27, which was previously described as an allele with high TAP dependency (26), and other HLA class molecules, such as HLA-B51 or- Cw1, with no data about their TAP dependency could be relevant in the study of alternative antigen processing pathways.

In a previous study, HLA-A2, -B27, -B51, and -Cw1-bound peptide pools were isolated from large amounts of either healthy or the VACV-infected human TAP-deficient cell line (Supplemental Figure 1) (17). These peptide mixtures were subsequently separated by reverse-phase HPLC and analyzed by mass spectrometry. Using several software technologies, eleven fragmentation spectra present in the VACV-infected HLA-bound peptidic pools, but absent in the respective control uninfected pools, were resolved as peptides of VACV viral proteins (17). These theoretical assignments were confirmed by their identity with the MS/MS spectrum of the corresponding synthetic peptide (17), thereby validating the software technology used for the identification of such HLA-bound ligands. Moreover, the use of identical software tools in these same samples resolved 112, 78, and 196 fragmentation spectra with high confidence parameters as peptidic sequences of different human cellular proteins bound to HLA-A2 (Supplemental Table 1), -B27 (Supplemental Table 2), and -B51 or -Cw1 (Supplemental Table 3), respectively.

As a control, the reverse database search of the same HLA-bound peptide pools yielded a 2% false positive rate. Two different peptide sequences were selected as additional controls for assignment because they did not have the canonical anchor motif for HLA-presenting molecules and thus could be candidates for erroneous assignments. The first of these control peptides has ∆mass and Sequest scores very close to the criterion limits (see peptide INKGKGFGF in Supplemental Table 2). The second peptide is longer (14 residues, TPEGGPAPPPYSEV) and is included in a nested set of N-extended peptides of the same protein (see below). Both the experimentally detected and the corresponding synthetic peptide MS/MS spectra were identical (Supplemental Figures 2 and 3). To confirm that HLA-B27 is the MHC class I molecule that presents these ligands, MHC/peptide complex stability assays were performed using TAP-deficient RMA-S cells transfected with the HLA-B27 molecule. The two ligands bound to HLA-B*2705 class I molecules with an affinity substantially lower than other natural high-affinity ligands (Supplemental Fig. 4), as would be expected due to the absence of HLA-B27 anchor motifs. As such, these peptides could be considered low-affinity ligands. Poor binding peptides lacking P2 anchor motifs have been eluted from both murine (8) and human (17) TAP-deficient cells and were described previously.

A comparison of the HLA peptidome from untreated and vaccinia virus-infected cells showed that approximately one-third of the cellular HLA ligands detected were common to both conditions (Supplemental Tables 1, 2, and 3). Approximately one-half and one-sixth of the peptides were found to be specifically associated with either untreated or vaccinia virus-infected cells, respectively (Supplemental Tables 1, 2, and 3). To identify associations among vaccinia-mediated changes in the infected cell, host proteins encoding HLA-A2 and -B27 unique self-peptides were mapped using the Ingenuity Pathway Analysis software. This analysis did not reveal significant differences in pathways associated with the infection.

Collectively, these results indicate that a similar broad range of TAP-independent ligands was endogenously processed and presented by different HLA class I molecules in the same infected cells, despite their differences in TAP dependency (26).

Structural features of TAP-independent HLA ligands.

HLA-A2, -B51, and Cw1 class I molecules usually bind peptides approximately 9-11 residues long (SYFPEITHI database: http://www.syfpeithi.de (27)), whereas HLA-B27 could accommodate peptides up to 13-14 residues in a bulged conformation (SYFPEITHI database, (28)). The analysis of size indicated that approximately 60% of the TAP-independent HLA-A2, -B51, or –Cw1 ligands and 40% of the -B27 ligands are longer than those identified in TAP-sufficient cells (Figure 1).

We next studied the anchor motif requirements of these ligands. The HLA-A2, -B51, and –Cw1 alleles present peptides with partially similar anchor motifs (SYFPEITHI database). The classical position 2 anchor motifs are as follows: for HLA-A2 binding, Leu or Met; for HLA-B51, Pro and Ala; and for HLA-Cw1, Ala and Leu. Aliphatic CΩ residues (SYFPEITHI database) are common between these three HLA alleles. For HLA-B27, the anchor motif consists of Arg or Gln at P2 and basic or aliphatic CΩ residues (SYFPEITHI database, (28)). However, the respective anchor motifs were absent in 60-70% of the TAP-independent HLA ligands (Table 1). Likewise, the amino acid preference at the CΩ position was studied and revealed major discrepancies among the HLA-A2, -B51 and –Cw1 ligands of TAP-sufficient versus TAP-deficient cells, although no differences in HLA-B27 ligands were found (Table 2). These data suggest that the relative contribution of the CΩ pocket to the stabilization of unusual peptides differs among these HLA class I alleles.

It is well documented that HLA-A2 binds signal sequence-derived peptides generated by the cleavage of signal sequences from the parental polypeptide by the

signal peptidase (SPase) complex (9). We found that 11% of peptides that bound to the HLA-A2 molecule were derived from the signal sequence of various proteins (Table 3). In addition, 3% of HLA-B27 and -B51 or -Cw1 ligands are located in the region generated by SPase activity (Table 3).). Most importantly, a significant fraction of bound ligands identified from HLA-A2 (17%)-, -B27 (23%)- and -B51 or -Cw1 (24%)-associated repertoires are located at the C-terminal position of their respective proteins. Thus, only one endoproteolytic cleavage was needed to release these particular ligands. In contrast, the remainder of peptides required two endoproteolytic cleavages for their generation. Unexpectedly, a similar fraction of the double cleaved HLA ligands were nested set peptides with an identical core but with N- or C-extended residues from the same protein (Table 3 and Supplemental Figures 1, 2, and 3). Indeed, one ligand found in the HLA-A2 and -B51 or -Cw1 peptide pools was both N- and C-terminally extended (Table 3 and Supplemental Figure 1 and 3). Representative nested set peptides are depicted in Figure 2.

TAP-independent processing generates multiple HLA ligands from the same cellular proteins.

In addition to N- and C-extended peptides (Supplemental Tables 1, 2 and 3, and Figure 2), some proteins contributing different clustered HLA ligands were identified by mass spectrometry (Supplemental Tables 1, 2, 3, and 4). Figure 3 depicts a representative example. In the first 80 residues of myosin heavy chain 9 protein, different adjacent or superimposed HLA ligands were processed. In addition, other N- and C-terminal extended peptides that were identified as binding to different HLA class I molecules were identified from the same protein (Figure 3, panel B). Thus, these data indicate that extensive processing of N- and C-terminal regions of some proteins occurs via TAP-independent antigen processing pathways. These recurrent

endoproteolytic activities were not restricted based on their location within the polypeptide because several HLA ligands were identified that constitute internal regions of some proteins (Figure 4 and Supplemental Table 4). In total, one-fifth of the identified proteins were processed, and their ligands were later presented by two different HLA class I molecules. Approximately 6% of the identified proteins generated multiple TAP-independent ligands that associated with HLA-A2, -B27, and -B51 or –Cw1 class I molecules (Table 4 and Supplemental Table 4).

Collectively, these data indicate that some proteins could be widely processed by proteases via TAP-independent pathways, yielding multiple peptides that could be presented by different HLA class I molecules. Thus, these data suggest that greater proteolytic processing occurs than previously supposed in TAP-independent antigen processing pathways.

Several cytoskeleton proteins are widely processed in TAP-deficient cells.

As shown in Table 4, some proteins were more extensively processed by the TAP-independent pathways (see also Supplemental Tables 4 and 5). The TAP-independent processing of these ten proteins generated one-third of the ligands associated with HLA-A2, -B27, and -B51 or –Cw1 class I molecules (Tables 4 and 5 and Supplemental Tables 4 and 5). In addition to two proteins of the secretory pathway (the HLA class I molecules and the CD74a protein), most of the extensively processed proteins were associated with the cytoskeleton (Table 5), which suggests that these proteins may have special access to regions of the secretory pathway with high cleavage activities.

Cellular location of parental proteins for TAP-independent HLA ligands.

The 337 TAP-independent ligands identified by mass spectrometry arose from 183 parental proteins. Table 6 shows the predicted organelle location of these proteins based on the Gene Ontology database (http://www.geneontology.org) (29). The majority of these proteins were from the cytoplasm, nucleus, and plasma membrane, whereas much lower percentages were from other subcellular compartments. One-third of the parental proteins were from the ER (8%), Golgi (4%), plasma membrane (14%), and secretory granules (6%), which likely represents their direct processing from transported proteins by resident proteases in the HLA-loading compartments. Strikingly, approximately two-thirds of the parental proteins were from non-secretory subcellular compartments, such as the nucleus (32%), cytoplasm (22%), cytoskeleton (9%), and mitochondria (5%). As such, their respective proteins and/or processed ligands must be transported into the ER lumen by a yet unknown pathway to yield TAP-independent antigen presentation.

The secretory vesicle-like organelle is the main source of parental proteins for HLA TAP-independent ligands.

Various TAP-independent ligands from several lysosomal proteins (e.g., lysosomal multispanning membrane protein 5, Figure 2) were identified. Thus, an attractive hypothesis arose that pointed to protein digestion within this degradative organelle as the source of most of the parental proteins yielding the TAP-independent ligands identified by mass spectrometry. Thus, the distribution of both the source proteins of the TAP-independent peptidome and that of the lysosomes (30,31) was compared. Table 6 shows a very different pattern of organelle location of constitutive proteins between this subcellular organelle and the TAP-independent peptidome. Thus, the lysosomes are not likely to be the primary source of peptides in the TAP-independent peptidome. As some lysosomal proteins are expressed in other secretory organelles, and because diverse vesicular organelles harbor different organelle-specific

constitutive proteins, the composition of several organelles was analyzed. Only secretory vesicles from human neutrophils (32) have shown a similar predominance of cytoplasm, nucleus and plasma membrane in the ratios observed for the TAP-independent peptidome (Table 6). Thus, secretory-like vesicles could be a relevant source of the parental proteins observed during TAP-independent antigen processing. Further, 51% of the parental proteins identified as part of the TAP-independent peptidome (Supplemental Table 5) were previously found in the proteome of this secretory vesicle (32). Thus, secretory vesicles and their related organelles were the main source of proteins processed to generate TAP-independent HLA ligands.

Cleavage specificity of peptidases over TAP-independent ligands.

Next, to study the specificity of peptidases involved in the generation of ligands in TAP-defective cells, an analysis was carried out that used mass spectrometry to detect residues on both sides of the hydrolyzed bonds of HLA ligands. When several nested peptides were found, as indicated in Figure 2, amino and/or carboxyl peptidase activities could be assumed. Accordingly, only the higher possible HLA ligand was examined. Figure 5 shows the distribution of amino acids found in the immediate flanking positions of scissile bonds (P_1 or P'_1 residues). The relative abundance of most amino acids was similar when P'_1, but not P_1, residues of HLA ligands were analyzed (panel B versus C). This was also true when the analysis was performed on each individual HLA class I molecule studied (data not shown). Additionally, no correlation was found when similar analyses of P_2, P_3, P'_2, and P'_3 positions were performed (data not shown). In summary, these results indicate that proteases with specificity to some residues in P_1, but not other positions, make the endoproteolytic cleavages to generate TAP-independent ligands. Only four amino acids (Arg, Leu, Lys, and Phe) are increased in the P_1 positions of scissile bonds, which accounted for up to 10% of the total cleavages detected (Fig. 5, panel B). These four major P_1 residues

could be processed by different proteases or may represent the specificities of one or a few proteases that generate both P_1 cleavage positions. To resolve these questions, an analysis of the correlation between these four specific amino acid residues and the opposite P_1 N- or C-end residues of HLA ligands was carried out. When Arg was the P_1 C-end of an HLA ligand identified by mass spectrometry, only two amino acid residues (Arg and Lys) were predominantly located in the corresponding P_1 N-end position (Fig. 6B, white boxes). This was true also for the reverse situation; when Arg was the P_1 N-end position, the amino acid residues predominantly located in the P_1 C-end position were also Arg and Lys (Fig. 6B, black boxes). Further, an identical correspondence in the analysis of Lys residue was found with both the P_1 N- and C-end positions (Fig. 6C). Similar analyses with Phe P_1 cleavages have shown that only Phe and Leu residues were mainly located in the equivalent P_1 N- or C-end positions of scissile bonds (Fig. 6D). A minor correlation with the Leu cleavage analysis was found, although Phe and Leu remained as the most abundantly detected residues (Fig. 6E). As several endoproteolytic peptidases have specificity for Arg/Lys or Leu/Phe in the P_1 position of the scissile bond (MEROPS database: http://merops.sanger.ac.uk (33)), the enzymatic activity of only two of these types of peptidases could explain more than half of the identified cleavages derived from the detected TAP-independent ligands (Table 7).

Low hydrophobicity in TAP-independent virus ligands.

In cells infected with the Epstein-Barr virus (EBV), only peptides with high hydrophobicity from the BRLF1 and LMP2 proteins were processed by TAP-independent pathways (34,35). In contrast, hydrophobicity is not a necessary condition for the TAP-independent presentation of vaccinia virus ligands (17). Because the total number of viral ligands in both studies was low (eight and eleven ligands from EBV or vaccinia virus, respectively), the study of the hydrophobicity of more abundant cellular TAP-independent ligands could be important. The GRAVY mean of approximately

1,000 HLA-A2, -B27, -B51, and -Cw1 TAP-dependent ligands previously described (Table 1) was only 0.0 ± 1.0, and significant differences between the different HLA alleles studied were not found (Table 8), indicating that no hydrophobicity of these HLA ligands exists. A very similar GRAVY measurement (-0.3 ± 0.8) was found for the 286 TAP-independent ligands described in this report (Table 7). These results show that hydrophobicity is not a necessary condition for the overall TAP-independent presentation of cellular ligands.

DISCUSSION

Identification of self-derived HLA ligands by mass spectrometry analysis, contributes to a better understanding of the mechanisms of antigen presentation that are associated with the cellular immune response. While several hundred peptides bound to specific MHC class I alleles are identified in any given immunoproteomics analysis from TAP-sufficient cells, only some tens have been described for TAP-independent HLA ligands from the still very limited number of immunoproteomics studies from TAP-deficient cells. In the first study, 22 and 27 cellular peptides bound to the murine H-2K^{d} class I molecule were immunoprecipitated from human and mouse TAP-deficient cells respectively (8). Later, 50 HLA-A2 or -B51 ligands were presented TAP-independently in a human TAP-deficient cell line (9). Currently, using a sequential immunoprecipitation of several HLA class I molecules, we identified several hundred TAP-independent ligands that were processed and presented by the four class I molecules expressed in the same cell population**.** Thus, the HLA peptidome generated by TAP-independent antigen processing pathways is more diverse than previously assumed. In agreement with previous studies (8,9), the major features of the peptide repertoire bound to classical MHC class I from TAP-deficient cells revealed increased peptide lengths and a lack of strict binding motifs in all HLA class I molecules studied in the current report. In addition, in a previous study using the same HLA peptidome, several vaccinia low affinity ligands were identified that did not conform to the normal anchor motifs used by HLA-A2, -B27 and –Cw1 class I molecules (17). Nonetheless, the identified HLA-A2 low-affinity ligand generated long-term CTL memory responses against vaccinia virus, even in an HLA-A2 transgenic TAP^{+} mouse model (17). These data indicate that low-affinity ligands, such as the self-derived peptides identified in the current report, have functional relevance.

Immunoproteomics analysis of the MHC class II peptidome usually shows ligands represented by several length variants of the same core, thereby forming sets of nested peptides varying by several residues at the N- or C-terminal ends (summarized in SYFPEITHI database). Further, in the analysis of the TAP-independent peptide repertoire that is associated with murine non-classical MHC class I molecules, some H-2Qa-1^{b}-bound peptides were identified as length variants of the same central core (36). In the current report, multiple sets of up to ten nested peptides were endogenously presented by different HLA class I molecules (Supplemental Table 4), suggesting that amino acid trimming at both the N- and C- terminus might be a general attribute of TAP-independent peptides. The similarities between the classical and non-classical peptide repertoires of HLA class II and TAP-independent HLA class I suggest at least a partial contribution of TAP-independent peptide loading in the post invariant chain-degrading compartments.

Unlike previous studies (8,9), the higher number of ligands identified in the current report allows the identification of extensive TAP-independent antigen processing mechanisms, yielding multiple HLA ligands in a reduced subgroup representing a small number of cellular proteins. Most of the parental proteins identified herein were cytoskeletal proteins, suggesting that a specific but uncharacterized mechanism of protein and/or peptide transport to secretory subcellular organelles must be hypothesized. Previously, autophagic processes present at low levels in different cell lines have been suggested to explain TAP-independent antigen processing (37). In addition, as peptides derived from cytosolic proteins are found bound to class II MHC molecules (38-40), the mechanisms involved in peptides and/or protein transport from the cytosol to secretory pathways could be similar for both TAP-independent class I and class II MHC ligands. This possibility is supported by our identification of nested sets of TAP-independent peptides that are very similar to those previously reported to be associated with HLA class II molecules. The similar composition of the secretory

vesicle proteome that includes both HLA class I and class II molecules (32), as well as the TAP-independent ligands identified in the current report, suggest that the antigen processing of proteins from these or similar organelles could be a relevant source of the TAP-independent peptidome. This vision of MHC class I molecules, from cell surface protein recycling, to entering into classical MHC class II compartments and later being transported back to the plasma membrane associated with endocytic/secretory vesicle peptides, might be the case for the peptides identified in the current report and is supported by two previous studies. At first, a measles virus F protein epitope was previously presented by class I molecules in TAP-independent, acidic-sensitive manner (41). Second, uncharacterized TAP-independent peptide-HLA-B27 complexes were generated by a proteasome-independent, but chloroquine-sensitive, pathway (42). Thus, this endocytic/secretory pathway may exist under normal conditions where it may contribute to a minor fraction of presented ligands. However, when the highly predominant TAP-dependent ligands are absent, as in TAP-deficient cells, this process may predominate.

Individual HLA class I alleles have shown different TAP dependencies. The prevalent HLA-A2 allele is considered to be the least TAP-dependent (43). HLA-B7 and –B8 alleles can also bind some ligands that are processed via mechanisms other than TAP transport. In contrast, other MHC class I molecules, including HLA-A3, -A24, and –B27, have been described as mainly TAP-dependent (26). In the present report, a similarly broad TAP-independent peptidome was identified by MS for these HLA alleles with differing TAP requirements. Thus, quantitative rather than qualitative differences (probably associated with the high efficiency rate of the SPase, accounting for the majority of signal sequence cleavages (44)), are responsible for the diverse overall expression of various MHC class I molecules in TAP-deficient cells (42).

The global picture emerging from the current report is consistent with the model depicted in Figure 7. Some ligands, mostly HLA-A2-restricted, were processed by the SPase, in accordance with previous studies (9,42,45,46). Endoproteolytic peptidases, exhibiting specificity for Arg/Lys or Phe/Leu in the P_1 position of the scissile bond, play an important role in the generation of many ligands associated with the four HLA class I alleles studied herein. However, some TAP-independent ligands must be produced by other, yet uncharacterized protease activities. Finally, a fraction of the longest variants from the same core protein could be folded within the HLA molecules, thereby protecting them from the trimming activity. The remaining molecules might be accessed by amino and/or carboxypeptidases and trimmed to shorter peptides. Several rounds of trimming and HLA/peptide stabilization of fractions of these peptides could form the sets of nested peptides identified by mass spectrometry. Recurrent and sequential amino-terminal trimming and MHC protection have been demonstrated for a nested set of abundant and equally antigenic murine HLA class I epitopes from TAP-sufficient cells that ranged between 9-15 residues (47,48). Whereas the role of the ER-resident peptidase ERAP in the N-terminal end trimming of different MHC class I ligands was previously well defined (49-51), to date, very few studies have implied a role for carboxypeptidases in antigen processing in the vesicular pathway. Indirect evidence has been reported in two cases: First, the proteolytic action of furin in the secretory pathway is required to generate an antigenic viral epitope (52). After cleavage by furin, several C-terminal residues must be trimmed from the precursor peptide to generate the optimal epitope, suggesting that carboxypeptidases are involved. Second, several signal sequence-derived peptides generated by SPase complexes have C-terminal-extended residues when compared to the optimal HLA-bound epitope (8,9,36), indicating that carboxypeptidases may be involved in antigen processing of these ligands. A direct role for the carboxypeptidase ACE was described for the processing of peptides for MHC class I (53). Recently, undefined carboxypeptidases were involved in the antigen processing of a vaccinia-derived TAP-independent epitope (54).Finally, the

present report indirectly implicates carboxyproteases in the generation of nested sets of peptides bound to several HLA class I alleles.

In summary, different and complex processing pathways are required to generate the HLA class I peptide repertoire in TAP-deficient cells.

REFERENCES

1. York, I. A., Goldberg, A. L., Mo, X. Y., and Rock, K. L. (1999) Proteolysis and class I major histocompatibility complex antigen presentation. *Immunol.Rev.* **172,** 49-66

2. Jensen, P. E. (2007) Recent advances in antigen processing and presentation. *Nat.Immunol.* **8,** 1041-1048

3. Cerundolo, V. and de la Salle, H. (2006) Description of HLA class I- and CD8-deficient patients: Insights into the function of cytotoxic T lymphocytes and NK cells in host defense. *Semin.Immunol.* **18,** 330-336

4. van Kaer, L., Ashton Rickardt, P. G., Ploegh, H. L., and Tonegawa, S. (1992) TAP1 mutant mice are deficient in antigen presentation, surface class I molecules, and $CD4^{-}8^{+}$ T cells. *Cell* **71,** 1205-1214

5. Del Val, M. and López, D. (2002) Multiple proteases process viral antigens for presentation by MHC class I molecules to $CD8^{+}$ T lymphocytes. *Mol.Immunol.* **39,** 235-247

6. Johnstone, C. and Del Val, M. (2007) Traffic of proteins and peptides across membranes for immunosurveillance by $CD8^{+}$ T Lymphocytes: A topological challenge. *Traffic* **8,** 1486-1494

7. Larsen, M. V., Nielsen, M., Weinzierl, A., and , L. O. (2006) TAP-Independent MHC Class I Presentation. *Current Immunology Reviews* **2,** 233-245

8. Suri, A., Walters, J. J., Levisetti, M. G., Gross, M. L., and Unanue, E. R. (2006) Identification of naturally processed peptides bound to the class I MHC molecule H-$2K^{d}$ of normal and TAP-deficient cells. *Eur.J.Immunol.* **36,** 544-557

9. Weinzierl, A. O., Rudolf, D., Hillen, N., Tenzer, S., van Endert, P., Schild, H., Rammensee, H. G., and Stevanovic, S. (2008) Features of TAP-independent MHC class I ligands revealed by quantitative mass spectrometry. *Eur J Immunol.* **38,** 1503-1510

10. Salter, R. D., Howell, D. N., and Cresswell, P. (1985) Genes regulating HLA class I antigen expression in T-B lymphoblast hybrids. *Immunogenetics* **21,** 235-246

11. Erlich, H., Lee, J. S., Petersen, J. W., Bugawan, T., and DeMars, R. (1986) Molecular analysis of HLA class I and class II antigen loss mutants reveals a homozygous deletion of the DR, DQ, and part of the DP region: implications for class II gene order. *Hum.Immunol.* **16,** 205-219

12. DeMars, R., Chang, C. C., Shaw, S., Reitnauer, P. J., and Sondel, P. M. (1984) Homozygous deletions that simultaneously eliminate expressions of class I and class II antigens of EBV-transformed B-lymphoblastoid cells. I. Reduced proliferative responses of autologous and allogeneic T cells to mutant cells that have decreased expression of class II antigens. *Hum.Immunol.* **11,** 77-97

13. Salter, R. D. and Cresswell, P. (1986) Impaired assembly and transport of HLA-A and -B antigens in a mutant TxB cell hybrid. *EMBO J.* **5,** 943-949

14. Kuipers, J. G., Raybourne, R., Williams, K. M., Zeidler, H., and Yu, D. T. (1996) Requirements for HLA-B*2705-binding peptides with special regard to the transporter associated with antigen processing (TAP). *Clin Exp.Rheumatol.* **14,** 523-529

15. Villadangos, J. A., Galocha, B., and Lopez de Castro, J. A. (1994) Unusual topology of an HLA-B27 allospecific T cell epitope lacking peptide specificity. *J.Immunol.* **152,** 2317-2323

16. Infantes, S., Lorente, E., Barnea, E., Beer, I., Cragnolini, J. J., García, R., Lasala, F., Jiménez, M., Admon, A., and López, D. (2010) Multiple, non-conserved, internal viral ligands naturally presented by HLA-B27 in human respiratory syncytial virus-infected cells. *Mol.Cell.Proteomics* **9,** 1533-1539

17. Lorente, E., Infantes, S., Barnea, E., Beer, I., Garcia, R., Lasala, F., Jimenez, M., Vilches, C., Lemonnier, F. A., Admon, A., and Lopez, D. (2011) Multiple viral ligands naturally presented by different class I molecules in transporter antigen processing-deficient vaccinia virus-infected cells. *J Virol* **86,** 527-541

18. Parham, P. and Bodmer, W. F. (1978) Monoclonal antibody to a human histocompatibility alloantigen, HLA-A2. *Nature* **276,** 397-399

19. Ellis, S. A., Taylor, C., and McMichael, A. (1982) Recognition of HLA-B27 and related antigen by a monoclonal antibody. *Hum.Immunol.* **5,** 49-59

20. Barnstable, C. J., Bodmer, W. F., Brown, G., Galfre, G., Milstein, C., Williams, A. F., and Ziegler, A. (1978) Production of monoclonal antibodies to group A erythrocytes, HLA and other human cell surface antigens-new tools for genetic analysis. *Cell* **14,** 9-20

21. Ishihama, Y., Rappsilber, J., Andersen, J. S., and Mann, M. (2002) Microcolumns with self-assembled particle frits for proteomics. *J Chromatogr.A* **979,** 233-239

22. Eng, J., McCormack, A., and Yates, J. (2009) An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J.Amer.Soc.Mass.Spect.* **5,** 976-989

23. Wang, M., Lamberth, K., Harndahl, M., Roder, G., Stryhn, A., Larsen, M. V., Nielsen, M., Lundegaard, C., Tang, S. T., Dziegiel, M. H., Rosenkvist, J., Pedersen, A. E., Buus, S., Claesson, M. H., and Lund, O. (2007) CTL epitopes for influenza A including the H5N1 bird flu; genome-, pathogen-, and HLA-wide screening. *Vaccine* **25,** 2823-2831

24. Fan, Q. R., Garboczi, D. N., Winter, C. C., Wagtmann, N., Long, E. O., and Wiley, D. C. (1996) Direct binding of a soluble natural killer cell inhibitory receptor to a soluble human leukocyte antigen-CW4 class I Major Histocompatibility Complex molecule. *Proc.Natl.Acad.Sci.U.S.A.* **93,** 7178-7183

25. López, D., Samino, Y., Koszinowski, U. H., and Del Val, M. (2001) HIV envelope protein inhibits MHC class I presentation of a cytomegalovirus protective epitope. *J.Immunol.* **167,** 4238-4244

26. Anderson, K. S., Alexander, J., Wei, M., and Cresswell, P. (1993) Intracellular transport of class I MHC molecules in antigen processing mutant cell lines. *J Immunol.* **151,** 3407-3419

27. Rammensee, H. G., Bachmann, J., Emmerich, N. P. N., Bachor, O. A., and Stevanovic, S. (1999) SYFPEITHI: database for MHC ligands and peptide motifs. *Immunogenetics* **50,** 213-219

28. Ben Dror, L., Barnea, E., Beer, I., Mann, M., and Admon, A. (2010) The HLA-B*2705 peptidome. *Arthritis Rheum.* **62,** 420-429

29. Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., and Sherlock, G. (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat.Genet.* **25,** 25-29

30. Schmidt, H., Gelhaus, C., Nebendahl, M., Lettau, M., Lucius, R., Leippe, M., Kabelitz, D., and Janssen, O. (2011) Effector granules in human T lymphocytes: the luminal proteome of secretory lysosomes from human T cells. *Cell Commun.Signal.* **9,** 4

31. Schmidt, H., Gelhaus, C., Nebendahl, M., Lettau, M., Lucius, R., Leippe, M., Kabelitz, D., and Janssen, O. (2011) Effector granules in human T lymphocytes: proteomic evidence for two distinct species of cytotoxic effector vesicles. *J Proteome.Res* **10,** 1603-1620

32. Uriarte, S. M., Powell, D. W., Luerman, G. C., Merchant, M. L., Cummins, T. D., Jog, N. R., Ward, R. A., and McLeish, K. R. (2008) Comparison of proteins expressed on secretory vesicle membranes and plasma membranes of human neutrophils. *J Immunol.* **180,** 5575-5581

33. Rawlings, N. D., Barrett, A. J., and Bateman, A. (2010) MEROPS: the peptidase database. *Nucleic Acids Res* **38,** D227-D233

34. Lautscham, G., Mayrhofer, S., Taylor, G., Haigh, T., Leese, A., Rickinson, A., and Blake, N. (2001) Processing of a multiple membrane spanning Epstein-Barr virus protein for $CD8^+$ T cell recognition reveals a proteasome-dependent, transporter associated with antigen processing-independent pathway. *J.Exp.Med.* **194,** 1053-1068

35. Lautscham, G., Rickinson, A., and Blake, N. (2003) TAP-independent antigen presentation on MHC class I molecules: lessons from Epstein-Barr virus. *Microbes.Infect.* **5,** 291-299

36. Oliveira, C. C., van Veelen, P. A., Querido, B., de Ru, A., Sluijter, M., Laban, S., van der Burg, S. H., Offringa, R., and van Hall, T. (2009) The nonpolymorphic MHC Qa-1b mediates CD8+ T cell surveillance of antigen-processing defects. *J Exp.Med* **207,** 207-221

37. Monastyrska, I., Rieter, E., Klionsky, D. J., and Reggiori, F. (2009) Multiple roles of the cytoskeleton in autophagy. *Biol.Rev Camb.Philos.Soc* **84,** 431-448

38. Suri, A., Walters, J. J., Kanagawa, O., Gross, M. L., and Unanue, E. R. (2003) Specificity of peptide selection by antigen-presenting cells homozygous or heterozygous for expression of class II MHC molecules: The lack of competition. *Proc.Natl Acad.Sci U.S.A* **100,** 5330-5335

39. Nelson, C. A., Roof, R. W., McCourt, D. W., and Unanue, E. R. (1992) Identification of the naturally processed form of hen egg white lysozyme bound

to the murine major histocompatibility complex class II molecule I-Ak. *Proc.Natl Acad.Sci U.S.A* **89,** 7380-7383

40. Mukherjee, P., Dani, A., Bhatia, S., Singh, N., Rudensky, A. Y., George, A., Bal, V., Mayor, S., and Rath, S. (2001) Efficient presentation of both cytosolic and endogenous transmembrane protein antigens on MHC class II is dependent on cytoplasmic proteolysis. *J Immunol.* **167,** 2632-2641

41. Gromme, M., UytdeHaag, F. G., Janssen, H., Calafat, J., van Binnendijk, R. S., Kenter, M. J., Tulp, A., Verwoerd, D., and Neefjes, J. (1999) Recycling MHC class I molecules and endosomal peptide loading. *Proc.Natl Acad.Sci U.S.A* **96,** 10326-10331

42. Lorente, E., Garcia, R., and Lopez, D. (2011) Allele-dependent processing pathways generate the endogenous human leukocyte antigen (HLA) class I peptide repertoire in TAP-deficient cells. *J Biol.Chem.* **286,** 38054-38059

43. Spies, T. and DeMars, R. (1991) Restored expression of major histocompatibility class I molecules by gene transfer of a putative peptide transporter. *Nature* **351,** 323-324

44. Suciu, D., Chatterjee, S., and Inouye, M. (1997) Catalytic efficiency of signal peptidase I of Escherichia coli is comparable to that of members of the serine protease family. *Protein Eng* **10,** 1057-1060

45. Wei, M. L. and Cresswell, P. (1992) HLA-A2 molecules in an antigen-processing mutant cell contain signal sequence-derived peptides. *Nature* **356,** 443-446

46. Henderson, R. A., Michel, H., Sakaguchi, K., Shabanowitz, J., Appella, E., Hunt, D. F., and Engelhard, V. H. (1992) HLA-A2.1-associated peptides from a mutant cell line: a second pathway of antigen presentation. *Science* **255,** 1264-1266

47. Samino, Y., López, D., Guil, S., Saveanu, L., van Endert, P. M., and Del Val, M. (2006) A long N-terminal-extended nested set of abundant and antigenic major histocompatibility complex class I natural ligands from HIV envelope protein. *J.Biol.Chem.* **281,** 6358-6365

48. Infantes, S., Samino, Y., Lorente, E., Jiménez, M., García, R., Del Val, M., and López, D. (2010) H-2Ld class I molecule protects an HIV N-extended epitope from in vitro trimming by endoplasmic reticulum aminopeptidase associated with antigen processing. *J Immunol.* **184,** 3351-3355

49. Saric, T., Chang, S. C., Hattori, A., York, I. A., Markant, S., Rock, K. L., Tsujimoto, M., and Goldberg, A. L. (2002) An IFN-g-induced aminopeptidase in the ER, ERAP1, trims precursors to MHC class I-presented peptides. *Nat.Immunol.* **3,** 1169-1176

50. Serwold, T., González, F., Kim, J., Jacob, R., and Shastri, N. (2002) ERAAP customizes peptides for MHC class I molecules in the endoplasmic reticulum. *Nature* **419,** 480-483

51. Saveanu, L., Carroll, O., Lindo, V., Del Val, M., López, D., Lepelletier, Y., Greer, F., Schomburg, L., Fruci, D., Niedermann, G., and van Endert, P. M. (2005)

Concerted peptide trimming by human ERAP1 and ERAP2 aminopeptidase complexes in the endoplasmic reticulum. *Nat.Immunol.* **6,** 689-697

52. Gil-Torregrosa, B. C., Castaño, A. R., López, D., and Del Val, M. (2000) Generation of MHC class I peptide antigens by protein processing in the secretory route by furin. *Traffic* **1,** 641-651

53. Shen, X. Z., Billet, S., Lin, C., Okwan-Duodu, D., Chen, X., Lukacher, A. E., and Bernstein, K. E. (2011) The carboxypeptidase ACE shapes the MHC class I peptide repertoire. *Nat.Immunol.*

54. Lorente, E., Garcia, R., Mir, C., Barriga, A., Lemonnier, F. A., Ramos, M., and Lopez, D. (2012) The role of metalloproteases in vaccinia virus epitope processing for transporter associated with antigen processing (TAP)-independent human leukocyte antigen (HLA)-B7 class I antigen presentation. *J Biol.Chem.* **in press,** doi: 10.1074/jbc.M111.314856.

FOOTNOTES

We thank Dr. J. A. López de Castro (Centro de Biología Molecular Severo Ochoa, Madrid, Spain) and Dr. David Yu (University of California, Los Angeles, CA) for the cell lines.

This work was supported by grants to D. L. from the Ministerio de Ciencia e Innovación and the FIPSE Foundation and to A. A. from the ISF 916/05. The funding agencies had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. The authors have no conflicting financial interests.

FIGURE LEGENDS

Figure 1. Length distribution of naturally processed peptides presented by HLA class I molecules in a TAP-deficient T2 cell line.

Respective totals of 112, 78, and 196 TAP-independent ligands were co-immunoprecipitated with HLA-A2 (filled bars), -B27 (open bars), and –B51 or –Cw1 (dashed bars) class I molecules.

Figure 2. Representative nested set peptides of ligands identified by mass spectrometry.

Diagram of representative nested set ligands N-extended (upper panel), C-extended (lower panel) or doubly extended (middle panels). The name of the protein source is boxed in each respective panel.

Figure 3. Naturally processed peptides from myosin heavy chain 9 identified by mass spectrometry.

Diagram of identified ligands bound to HLA class I molecules in the first 80 (panel A) or last 40 (panel B) residues from myosin heavy chain 9 protein. Ligands specific for HLA-A2 (white boxes), -B27 (black boxes), and –B51 or –Cw1 (gray boxes) are depicted in the lower section of each panel.

Figure 4. Naturally processed peptides from MRCL2, β-actin, and glyceraldehyde 3-P dehydrogenase proteins identified by mass spectrometry.

Diagram of identified ligands bound to HLA class I molecules from MRCL2 (upper panel), β-actin (middle panel), and glyceraldehyde 3-P dehydrogenase (lower panel). Ligands specific for HLA-A2 (white boxes), -B27 (black boxes), and –B51 or –Cw1 (gray boxes) are depicted in each panel.

Figure 5. Analysis of N- and C-end cleavage specificity in HLA class I ligands.

A diagram of residues involved in the generation of naturally processed HLA class I ligands by peptidase cleavages is shown (panel A). Distribution of P_1 (panel B) and P'_1 (panel C) amino acid residues of the scissile bonds created by peptidase cleavage.

Figure 6. Analysis of correspondence of N- and C-end cleavage specificity in HLA class I ligands.

Panel A: A diagram of the residues involved in the generation of naturally processed HLA class I ligands by peptidase cleavage is shown. P1 N-end (white boxes) and P1 C-end (black boxes). The level of correspondence between specific amino acid residues is indicated at the top right corner of each panel and opposite P_1 N-end (white boxes) or C-end (black boxes) residue of HLA ligand (panels B-D).

Figure 7. A model of the diversity of proteases and parallel processing pathways involved in TAP-independent self-derived antigen presentation.

The model shows the components involved in each of the proposed pathways.

Table 1

Amino acid preference at the anchor motif P2 position in TAP-dependent versus TAP-independent ligands

Residue	HLA-A2		Residue	HLA-B27		Residue	HLA-B51, -Cw1	
	TAP $^{+}$ [a]	TAP $^{-}$ [b]		TAP $^{+}$ [c]	TAP $^{-}$ [d]		TAP $^{+}$ [e]	TAP $^{-}$ [f]
L	68 [g]	38	R	98	22	L	5	14
M	5	4	Q	2	6	P	34	18
						A	42	7
Σ	73	42	Σ	100	28	Σ	81	39

[a] 424 HLA-A2 ligands from the SYFPEITHI database.
[b] 112 HLA-A2 TAP-independent ligands, see Supplemental Table 1.
[c] 571 HLA-B27 ligands (28).
[d] 78 HLA-B27 TAP-independent ligands, see Supplemental Table 2.
[e] 68 HLA-B51 and 9 HLA-Cw1 ligands from the SYFPEITHI database.
[f] 196 HLA-B51 and -Cw1TAP-independent ligands, see Supplemental Table 3.
[g] Data are expressed in percentages.

Table 2

Amino acid preference at the CΩ position in TAP-dependent versus TAP-independent ligands

Residue	HLA-A2		Residue [a]	HLA-B27		Residue [a]	HLA-B51, -Cw1	
	TAP $^{+}$ [a]	TAP $^{-}$ [b]		TAP $^{+}$ [c]	TAP $^{-}$ [d]		TAP $^{+}$ [e]	TAP $^{-}$ [f]
L	38 [g]	17	R	22	20	V	21	11
V	35	11	F	20	17	I	40	6
I	11	3	K	20	23	L	22	14
A	8	10	L	18	14			
			Y	11	5			
Σ	92	41	Σ	91	79	Σ	83	31

[a] 424 HLA-A2 ligands from the SYFPEITHI database.
[b] 112 HLA-A2 TAP-independent ligands, see Supplemental Table 1.
[c] 571 HLA-B27 ligands (28).
[d] 78 HLA-B27 TAP-independent ligands, see Supplemental Table 2.
[e] 68 HLA-B51 and 9 HLA-Cw1 ligands from the SYFPEITHI database.
[f] 196 HLA-B51 and Cw1 TAP-independent ligands, see Supplemental Table 3.
[g] Data are expressed in percentages.

Table 3

Major features of TAP-independent HLA ligands

Type of peptide	HLA-A2	HLA-B27	HLA-B51, -Cw1
Signal sequence [a]	11 [b]	3	3
C-terminal [c]	17	23	24
N-extended [d]	10	18	11
C-extended [e]	4	8	4
N- and C-extended	1	0	1
Total of extended	15	26	16

[a] Peptides located at the signal sequence of their respective proteins.
[b] Data are expressed as a percentage of the total TAP-independent ligands.
[c] Peptides located at the C-terminal position of their respective protein.
[d] N-extended peptides with respect to the minimal ligand identified with identical core.
[e] C-extended peptides with respect to the minimal ligand identified with identical core.

Table 4

HLA restriction and the number of proteins from TAP-independent ligands

Number of HLA alleles	% of Proteins [a]	% of TAP-independent ligands [a]
One	75	45
Two	19	22
HLA-A2 and –B27	2	2
HLA-A2 and –B51 or –Cw1	11	10
HLA-B27 and –B51 or –Cw1	6	10
Three	6	33

[a] Of the total shown in Supplemental Tables 1, 2, and 3.

Table 5

Subcellular location of proteins with TAP-independent ligands presented by HLA-A2, -B27, and -B51 or –Cw1

Protein	Gi accession	Subcellular location [a]	N^{er} of peptides
Beta Actin	4501885	cytoskeleton	24
CD74a	68448544	ER-Golgi-endosome [b]	7
Glyceraldehyde-3-phosphate dehydrogenase	7669492	cytosol	6
HLA-A2	717123	ER-Golgi-membrane	15
Heterogeneous nuclear ribonucleoprotein A1	133254	nucleus, cytoskeleton	5
Heterogeneous nuclear ribonucleoprotein U	126302554	nucleus, cytoskeleton	10
Myosin regulatory light chain MRCL2	15809016	cytoskeleton	11
Myosin, heavy polypeptide 9, non-muscle	12667788	cytoskeleton	30
TTD non-photosensitive 1 protein	20162566	nucleus	1
Tubulin alpha 6	14389309	cytoskeleton	3

[a] Cell location based on gene ontology analysis (http://www.geneontology.org) (29).
[b] ER, endoplasmic reticulum

Table 6

Distribution of proteins by cell location

Location	TAP$^-$ peptidome	Lysosomes [a]	Secretory vesicles [b]
Cytoplasm	22 [c]	16	23
Cytoskeleton	9	0	8
Endoplasmic reticulum	8	2	4
Extracellular	0	0	5
Golgi	4	1	2
Mitochondria	5	2	6
Nucleus	32	5	22
Plasma membrane	14	2	24
Secretory granule [d]	6	72	8

[a] From human T cells (30,31).
[b] From human neutrophils (32).
[c] Data are expressed as the percentage of proteins listed by cell location based on an analysis of gene ontology (http://www.geneontology.org) (29).
[d] Secretory granules refers to melanosomes, lysosomes, platelet granules, endosomes, synaptosomes, exosomes or cytolytic granules, as defined in references (30,31).

Table 7

Summary of the predominant peptidase-dependent cleavage residues
in TAP-independent ligands

Cleavaged residue	HLA-A2		HLA-B27		HLA-B51, -Cw1		Total HLA alleles	
	P_1 N-end [a]	P_1 C-end [a]	P_1 N-end [a]	P_1 C-end [a]	P_1 N-end [a]	P_1 C-end [a]	P_1 N-end [a]	P_1 C-end [a]
K/R	34 [e]	24	34	42	27	25	33	30
L/F	24	29	20	30	24	27	24	28
Σ	58	53	54	72	51	52	57	58

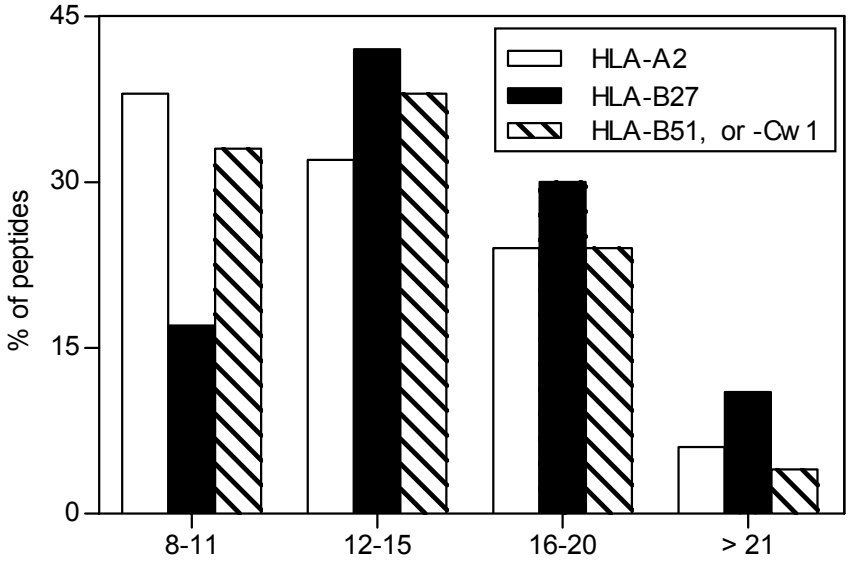
[a] From figures 6 and 7.
[b] Data are expressed as the percent of total cleavages.

Table 8

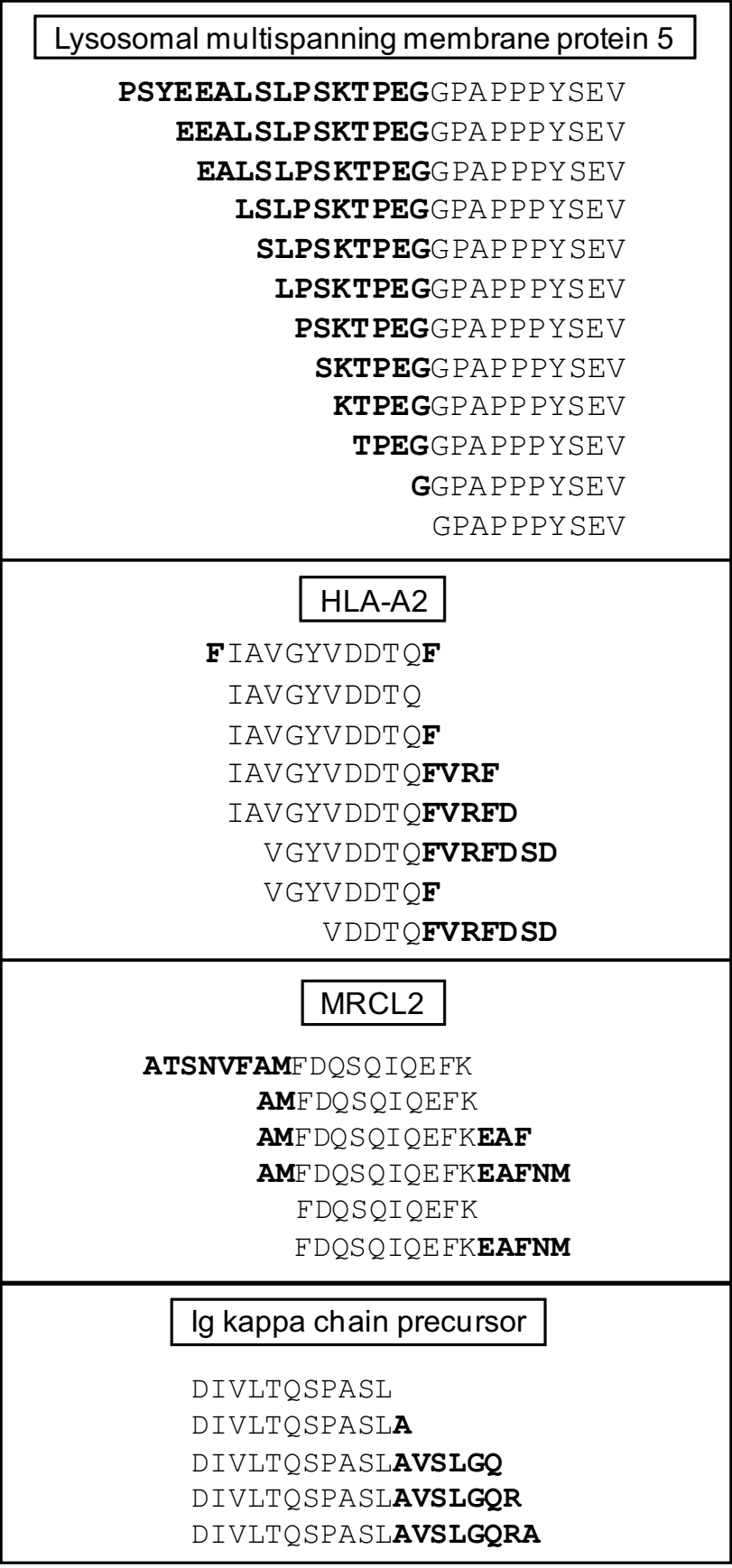
Hydrophobicity and TAP dependency of HLA class I ligands

HLA	TAP $^{+}$	TAP $^{-}$
-A2	0.7 ± 0.9	0.0 ± 0.8
-B27	-0.4 ± 0.8	-0.6 ± 0.7
-B51, and -Cw1	0.3 ± 0.9	-0.3 ± 0.8
Total	0.0 ± 1.0	-0.3 ± 0.8

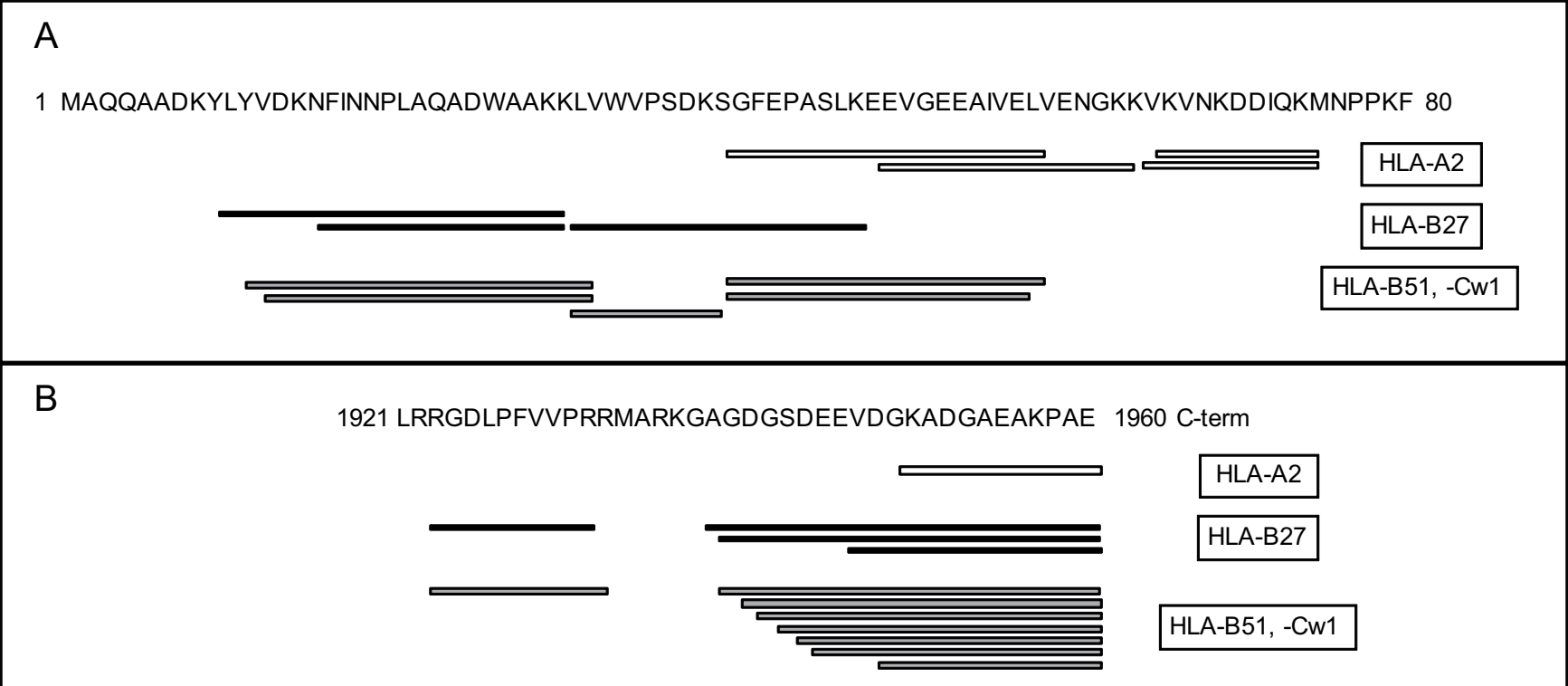
[a] The mean ± SD of the hydropathicity of the HLA class I ligands previously shown in Table 1 is shown, according to the grand average of hydropathicity (GRAVY) scale (ProtParam tool, ExPASy Proteomics Server, http://www.expasy.ch).



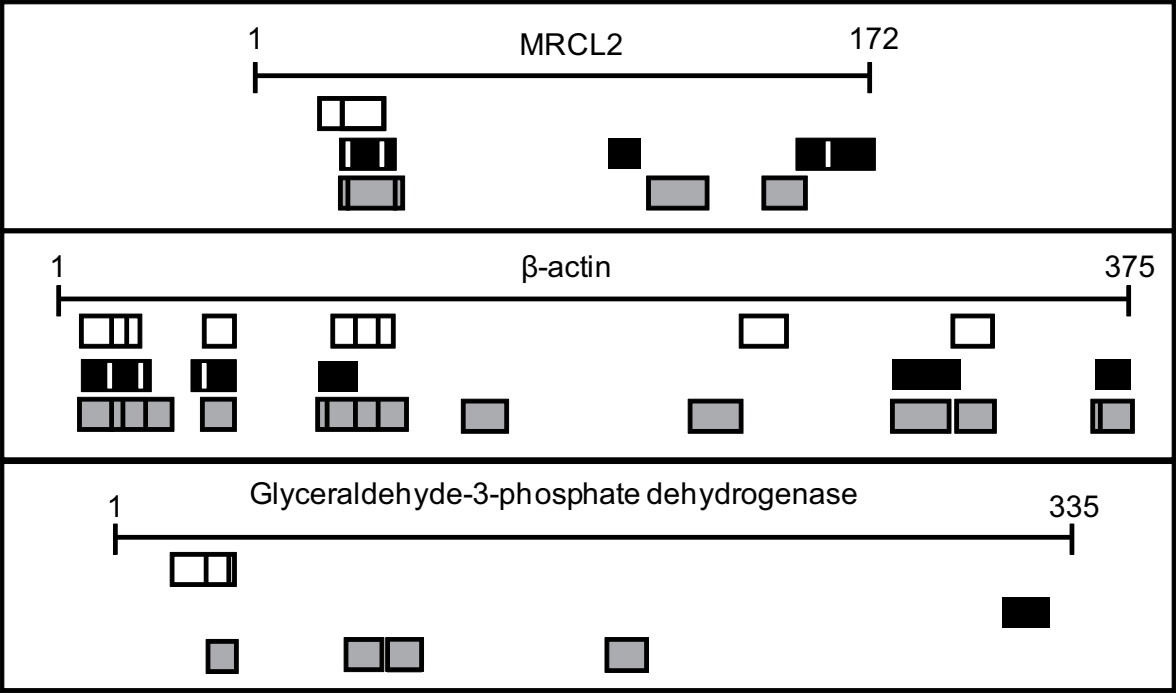
Lorente et al Figure 1



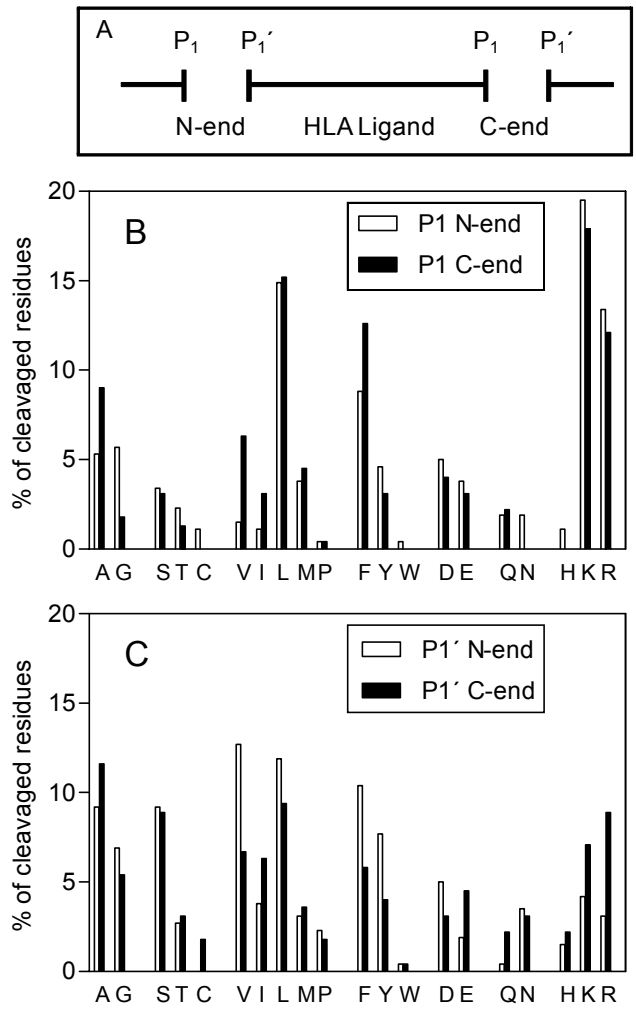
Lorente et al. Figure 2



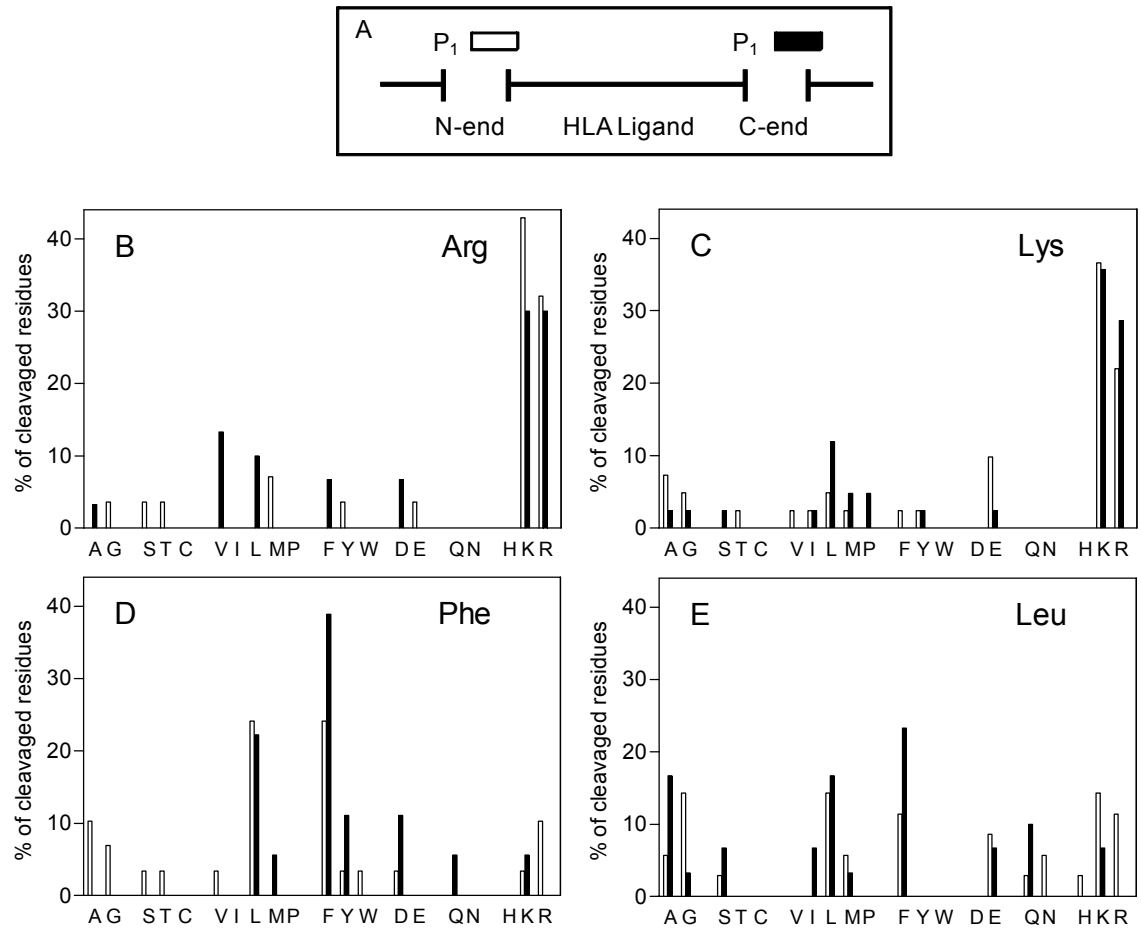
Lorente et al. Figure 3



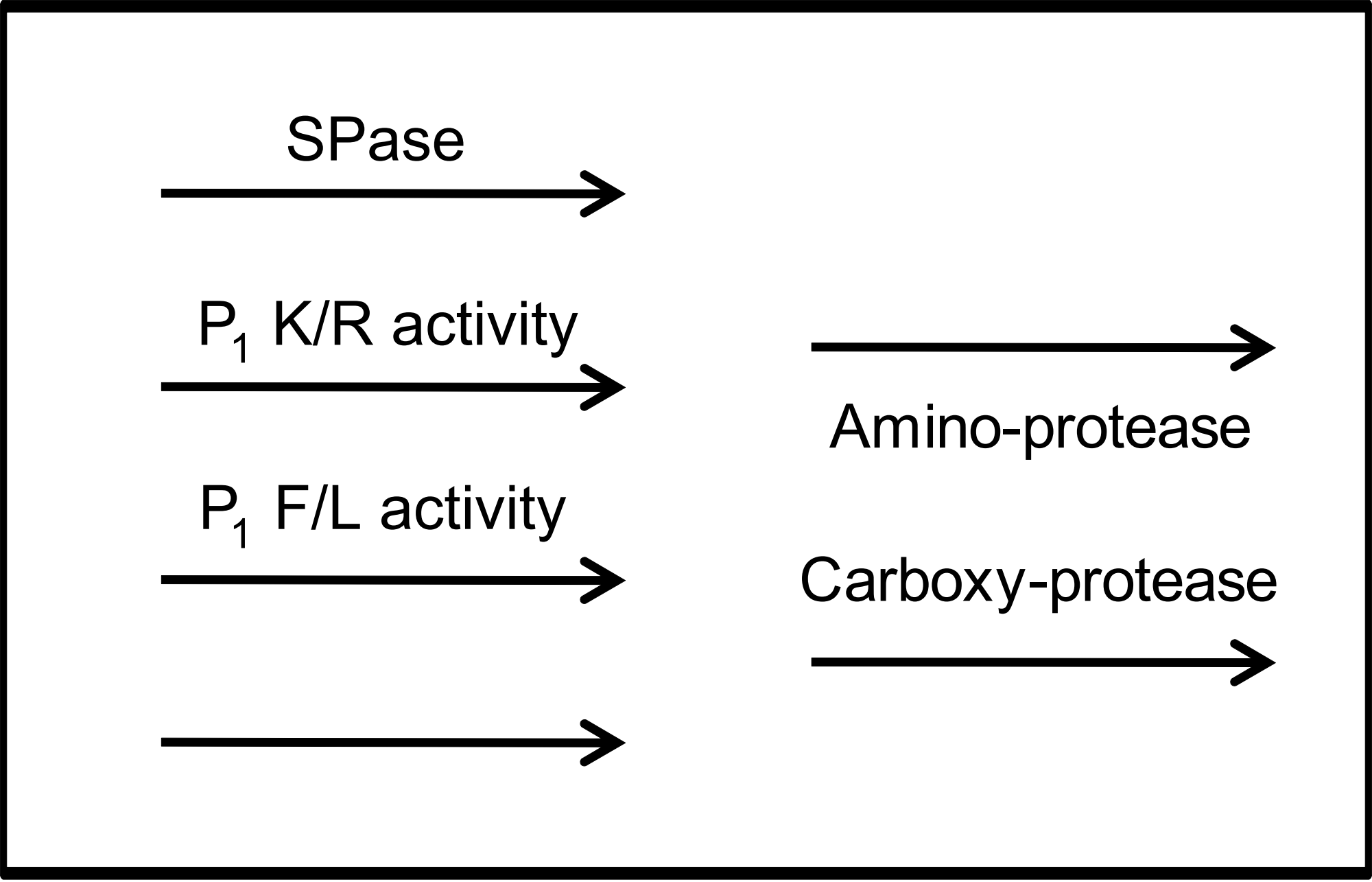
Lorente et al. Figure 4



Lorente et al. Figure 5



Lorente et al. Figure 6



Lorente et al Figure 7

Lista de trabajos publicados durante esta Tesis

Juan J. Cragnolini, Noel García-Medel, and José A. López de Castro. Endogenous Processing and Presentation of T-cell Epitopes from Chlamydia trachomatis with relevance in HLA-B27-associated Reactive Arthritis. Molecular and Cellular Proteomics, **8**: 1850-1859. 2009.

Patricia Gómez, Carla Mavian, Begoña Galocha, Noel García-Medel, and José A. López de Castro. Presentation of Cytosolically stable Peptides by HLA-B27 is not Dependent on the Canonic Interactions of N-Terminal Basic Residues in the A Pocket. The Journal of Immunology, **182**: 446-455. 2009.

Garcia-Medel N, Sanz A, Barnea E, Admon A, Lopez de Castro JA. The origin of proteasome-inhibitor resistant HLA class I peptidomes: a study with HLA-A*68:01. Molecular and Cellular Proteomics, **11**: M111. 2011.

Noel García-Medel, Alejandro Sanz-Bravo, Dung Van Nguyen, Begoña Galocha, Patricia Gómez-Molina, Adrián Martín-Esteban, Carlos Álvarez-Navarro, José Antonio López de Castro. Functional interaction of the ankylosing spondylitis associated ERAP1 polymorphism and HLA-B27 in vivo. Manuscript submitted.

Elena Lorente, Susana Infantes, Eilon Barnea, Ilan Beer, Ruth García, Noel García-Medel, Fátima Lasala, Mercedes Jiménez, Arie Admon, Daniel López. Diversity of natural self-derived ligands presented by different HLA class I molecules in transporter antigen processing-deficient cell. Manuscript submitted.