UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

COLEGIO DE CIENCIAS Y HUMANIDADES
UNIDAD ACADEMICA DE LOS CICLOS
PROFESIONAL Y DE POSTGRADO

PROYECTO DE INVESTIGACION BIOMEDICA BASICA
INSTITUTO DE INVESTIGACIONES BIOMEDICAS

Reconocimiento Molecular: Relación estructura función en los receptores de células-T.

TESIS

Que para obtener el Grado de
DOCTOR EN INVESTIGACION BIOMEDICA BASICA
PRESENTA:

Juan Carlos Almagro Domínguez

México, D. F. Marzo de 1995

UNAM – Dirección General de Bibliotecas Tesis Digitales Restricciones de uso

PAGINACION VARIA

COMPLETA LA INFORMACION

a Carlos Larralde,
una de las personalidades más nobles
e interesantes que haya conocido,
inspirador además, de este trabajo.

a Ligia y Ana Sofía,
alegría de mi vida.

Agradecimientos

Quisiera agradecer en primer lugar, al Dr. Francisco Lara Ochoa,
por sus consejos oportunos,
y por el apoyo brindado en todo momento.

Después, al Dr. Eduardo Horjales Reboredo,
por sus enseñanzas sobre la estructura de las proteínas.

A los miembros del Comité tutorial:
Dr. Francisco Lara Ochoa,
Dr. Carlos Larralde,
Dr. Lino Días de León,
Dra. Carmen Gómez,
por las discusiones académicas sostenidas durante este periodo de mi vida,
que sin duda, fueron un elemento esencial en la maduración de mi
formación profesional.

Al jurado de tesis:
Dr. Francisco Lara Ochoa,
Dr. Eduardo Horjales Reboredo,
Dr. Xavier Soberón Mainero,
Dr. Enrique Ortega Soto,
Dr. Mario Calcagno Montans,
Dr. Edmundo Lamoyi Velázquez
Dr. Jaime Lagunez Otero,
por sus sugerencias y recomendaciones para mejorar el trabajo que aquí se
presenta.

A María Elena Ortega,
por su ayuda.

Al Profesor Jan Hermans y el Dr. Renán Cárdenas Marín,
por las discusiones sobre dinámica molecular.

Al Dr. Mario Amzel,
por las conversaciones sobre la estructura de las inmunoglobulinas.

A la Dirección General de Asuntos del Personal Académico,
por su apoyo a través del proyecto IN-206093.

Finalmente, y no por ello menos agradecido,
a mi amigo Enrique Vargas Madrazo.

Resumen

El objetivo central de la presente tesis es estudiar la relación estructura-función en el receptor de células-T (TCR), uno de los paradigmas de reconocimiento molecular. Esta familia de proteínas, junto con la familia de las inmunoglobulinas (Igs), son hasta el presente, las únicas fuentes conocidas de especificidad del sistema inmunológico. Los TCRs reconocen en combinación con las moléculas del complejo mayor de histocompatibilidad (MHC), péptidos relativamente sencillos producto de la fragmentación proteolítica de las proteínas antigénicas. Desde el punto de vista estructural, en contraste con el rápido avance en la caracterización del complejo péptido/MHC, para el TCR, no se dispone de una estructura tridimensional experimental. La falta de información estructural para esta familia de proteínas, ha conducido a que la relación estructura-función en el TCR sea poco comprendida.

Por analogía con las Igs de estructura conocida, se ha propuesto que el sitio de reconocimiento péptido/MHC en el TCR, se forma por seis regiones determinantes de la complementaridad (CDR), tres en el dominio Vα y tres en el dominio Vβ. Para identificar los CDRs en las secuencias de TCR, se utilizan aquí, tres índices de variabilidad simultáneamente. Los resultados muestran que a diferencia de las Igs, la región correspondiente al CDR-1 de las Igs, no es hipervariable en el TCR. La explicación del comportamiento relativamente conservado de los CDR-1, pudiera justificare en términos de que esta región, interactua con una región poco variable del MHC.

Para estudiar esta proposición se obtuvo, en ausencia de datos estructurales experimentales, un modelo atómico de interacción TCR/péptido/MHC. Los resultados de este modelo permiten asignar un papel para cada CDR en la interacción con el complejo péptido/MHC. El modelo obtenido muestra que, el CDR-1 de Vβ no interactua con la molécula de MHC, mientras que el CDR-1 de Vα, contacta con residuos relativamente conservados del MHC. Respecto al CDR-2 de Vα , el modelo indica que este, tiene un papel importante en el reconocimiento de las diferentes clases de MHC. El CDR-2 de Vβ por otra parte, pudiera determinar tanto el reconocimiento del MHC como el del péptido antigénico. Los CDR-3 de ambos dominios reconocen fundamentalmente al péptido antigénico, aunque también interactuan con el MHC.

Estas predicciones difieren cualitativamente de lo propuesto en la literatura. Así, la proposición del papel que juega cada CDR del TCR en la interacción con el complejo péptido/MHC sugiere una buena hipótesis de trabajo para diseñar experimentos precisos, que contribuyan a entender en términos estructurales, de que depende la especificidad de esta familia de proteínas.

Abstract

The main purpose of the present thesis is to study the structure-function relationship of the T-cell receptor (TCR), one of the molecular recognition paradigms. This family of proteins, together with the immunoglobulins (Igs), is at present, the only known source of the specificity of the immune system. TCRs recognise, in combination with the major histocompatibility complex molecules (MHC), relatively simple peptides, it product of the proteolytic fragmentation of the antigenic proteins. From a structural point of view, no three-dimensional structures are known for the TCR. The absence of structural information for this family of proteins, has been conduced to a poor understanding of the structure-function relationships in the TCRs.

By analogy with Igs of known structure, has been proposed that the peptide/MHC binding site of the TCRs, is formed by six complementarity determining regions (CDRs). Three of them placed at the Vα domain, and the other three in the Vβ domain. In order to identify the CDRs in the TCR sequences, three variability indexes are used here. Different to the Igs, the results indicate that the CDR-1 is a non hypervariable region in the TCR. One can suggest that the explication of such behaviour is inherently related to the interaction of this region with relatively conserved regions on the MHC molecules.

In order to study this proposition, in absence of experimental structural data, a three-dimensional atomic-model for the TCR/peptide/MHC interaction is obtained. The results allow us assign a role for each CDR in the interaction with the peptide/MHC complex. The model shows that the CDR-1 of Vβ does no interact with the MHC molecule, while the CDR-1 of Vα, interact with relatively conserved residues of the MHC. Concerning to the CDR-2 of Vα, the model indicates that this region has a fundamental role in the recognition of different class of MHCs. The CDR-2 of Vβ on the order hand, determine the recognition of the MHC molecule, as well as the antigenic peptide. The CDR-3 of both domains recognise mainly the antigenic peptide, although also interact with MHC.

The above predictions qualitatively differ to the propositions reported in the literature. Thus, the proposition of the role played by each CDR of the TCR, suggest a refined working hypothesis that can be tested by mutational experiments in order to understand the specificity mediated by TCRs.

Índice

Introducción

Capítulo I: Modelo de interacción TCR/péptido/MHC.

Capítulo II: VIR: A computational tool for the analysis of immunoglobulin sequences.

Capítulo III: Variability analysis of T-cell receptors using three variability indexes.

Capítulo IV: Molecular modeling of a T-cell receptor bound to a major histocompatibility complex molecule: Implications for the T-cell recognition.

Conclusiones

Bibliografía

Introducción

Las proteínas son uno de los componentes esenciales de los sistemas vivos, interviniendo de manera fundamental en todos los procesos vitales. Para realizar su función adecuadamente, las proteínas necesitan invariablemente interactuar físicamente con su ligando. Esta interacción tiene que ser además, lo suficientemente específica, como para que cada proteína reconozca a su ligando dentro de la compleja mezcla de moléculas que se encuentran en el exterior o interior de una célula.

Estas frases formuladas de la manera más simple posible, sugieren en principio, el proceso de reconocimiento molecular como elemento esencial de la función de las proteínas.

Uno de los paradigmas de reconocimiento molecular lo constituyen sin duda, los receptores del sistema inmunológico; inmunoglobulinas y receptores de células-T (Igs y TCRs, respectivamente). Estas dos familias de proteínas están dotadas de una gran capacidad para discernir, con alto grado de precisión, las moléculas que conforman un organismo de las que no forman parte de él.

Históricamente el estudio de los receptores del sistema inmunológico ha girado alrededor de dos preguntas. La primera se refiere a entender en que consiste la versatilidad de las Igs y los TCRs, capaces de reconocer de manera específica a cerca de 10^{15} moléculas diferentes.[1,2] La segunda pregunta es, si a pesar de esta gran versatilidad, existen reglas generales del reconocimiento molecular mediado por estas familias de proteínas.[3-5]

1.Berek C, Milstein C.1988. *Immunol. Rev.* **105**:5.
2.Davis MM.1990. *Ann. Rev. Biochem.* **59**:475.
3.Mian IS, Bradwell AR, Olson AJ. 1991 *J. Mol. Biol.* **217**:133.
4.Almagro JC, Vargas-Madrazo E, Zenteno-Cuevas R, Hernández-Mendiola V, Lara-Ochoa F.1995. *BioSystems.* **en prensa**
5.Vargas-Madrazo E, Horjales E, Lara-Ochoa F, Almagro JC.1995. *J.Mol. Biol.* **enviado**

Estas dos preguntas aunque son complementarias y tienen como objetivo común entender las bases del reconocimiento molecular mediado por los receptores del sistema inmunológico, son metodológicamente diferentes. En el primer caso se analizan las diferencias entre los miembros de las familias de las Igs o de los TCRs. En el segundo caso se buscan las similitudes. Las respuestas a la primera pregunta ayudan a entender en que consiste la especificidad de una Ig o de un TCR por su ligando particular. Las soluciones de la segunda pregunta, conducen a establecer principios generales del reconocimiento molecular mediado por las Igs y los TCRs.

Los resultados que se exponen en esta tesis son parte de un conjunto de trabajos desarrollados dentro del período de doctorado que han tenido como objetivo central la segunda pregunta acerca de los receptores del sistema inmunológico.[4-12]

Desde el punto de vista práctico, abordar el estudio de las Igs y los TCRs, como modelo de reconocimiento molecular, se ha visto facilitado por la gran cantidad de información disponible para estas dos familias de proteínas. Así por ejemplo, las Igs se consideran las moléculas mejor estudiadas de todas las proteínas.[13] Lo anterior se ilustra por el número de secuencias y estructuras tridimensionales que actualmente se depositan en los bancos de secuencias y estructuras cristalográficas. En octubre del pasado año, por ejemplo, en el

---

6.Lara-Ochoa F, Vargas-Madrazo E, Jiménez-Montaño MA, Almagro JC. 1994. *Biosystems.* **32**:1

7.Lara-Ochoa F, Vargas-Madrazo E, Almagro JC. 1995. *J. Mol. Evol.* **en prensa**

8.Vargas-Madrazo E, Almagro JC, Lara-Ochoa F, Jiménez-Montaño MA.1993. *ECAL'93.* p.1070.

9.Vargas-Madrazo E, Almagro JC, Lara-Ochoa F.1995. *J. Mol. Biol.* **246**: 74.

10.Almagro JC, Zenteno-Cuevas R, Vargas-Madrazo E, Lara-Ochoa F. 1995. *Int. J. Pep. & Prot. Res.* **en prensa**.

11.Almagro JC, Vargas-Madrazo E, Lara-Ochoa F, Horjales E.1995. *Prot. Sci.* **enviado**.

12.Almagro JC, Ceceña HA, Vargas-Madrazo E, Lara-Ochoa F.1995. *POP'95.* **en prensa**

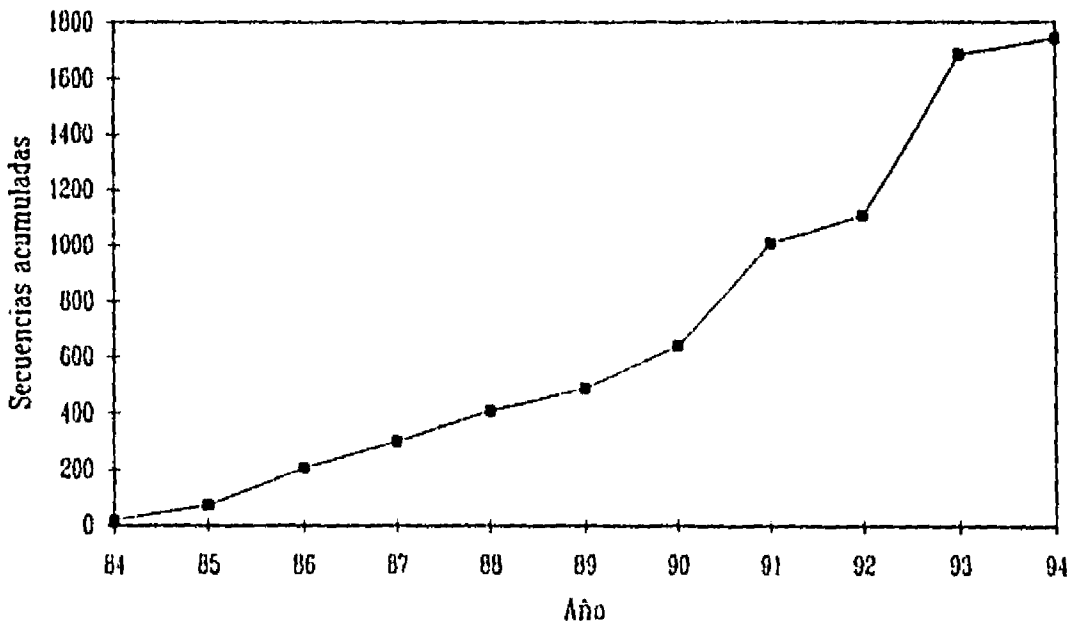
13.Padlan, 1994. *Mol. Immunol.* **31**:169.

14. Kabat EA, Wu TT, Perry HM, Gottesman KS, Foeller C. 1991. *NIH Publication No. 91-3242.*

banco de proteínas dè interés inmunológico[14] se compilaron aproximadamente 8,000 secuencias de Igs.[4] Mientras que en el banco de estructuras cristalográficas[15] (PDB) en fechas similares, el número de Igs era aproximadamente de 140 estructuras.

Similar a las Igs, en el presente se cuenta con un gran número de secuencias para los TCRs. En la Figura 1 se ofrece una idea de esta información. La figura muestra que, hasta la fecha, se han obtenido casi 1,800 secuencias de TCRs. Además, el creciente interés en esta familia de proteínas queda demostrado por el incremento casi exponencial en el número de secuencias de TCRs depositadas en el banco de proteínas de interés inmunológico[14], desde el reporte de la primera secuencia de un TCR en 1984.[16]

Figura.1. *Número de secuencias de TCRs depositadas en el banco de proteínas de interés inmunológico[14] hasta octubre de 1994 como función del año en que se reportaron. Los datos se obtuvieron con el paquete de programas VIR.*[4]



---

15. Bernstein FC, Koetzle TF, Williams GJB, Meyer EF Jr, Brice MD, Rodgers JR, Kennard O, Shimandouchi T, Tasumi M. 1977. *J. Mol. Biol.* **112**:535.

16. Hedrick SM, Cohen DI, Nielsen EA, Davis MM. 1984. *Nature.* **308**:149.

A diferencia de las Igs, para los TCRs sin embargo, no se dispone de estructuras tridimensionales experimentales. La falta de información estructural experimental para esta familia de proteínas, ha conducido a que la interpretación de la interacción mediada por los TCRs, este en franca desventaja con respecto al estudio de la interacción mediada por las Igs.

A pesar de lo anterior, los TCRs ofrecen una ventaja con respecto a las Igs en cuanto a establecer principios generales del reconocimiento molecular. Esto es, los TCRs son anteriores a las Igs desde el punto de vista evolutivo,[2] lo que implica que su mecanismo de reconocimiento molecular es más primitivo y en consecuencia, probablemente más simple. En efecto, las Igs reconocen directamente la estructura nativa de las proteínas antigénicas en solución.[17] Los TCRs en contraste, reconocen en combinación con las moléculas del complejo mayor de histocompatibilidad (MHC), péptidos relativamente sencillos producto de la fragmentación proteolítica de las proteínas antigénicas.[2] Este mecanismo de interacción por ser más restrictivo respecto al de las Igs, pudiera en principio, resultar más susceptible de ser descrito en términos generales.

Es por lo tanto deseable acercarse tanto como sea posible, en ausencia de datos estructurales experimentales para el TCR, a la descripción del mecanismo de reconocimiento molecular mediado por este receptor del sistema inmunológico. Los artículos que se seleccionaron para esta presentación [4,10,11] tienen este objetivo en común. En la sección que sigue se describen brevemente los artículos antes citados (4,10,11), y se explica el aporte de cada uno de ellos, a la descripción del mecanismo de reconocimiento molecular mediado por los TCRs; objetivo central de la tesis.

---

17. Davies DR, Padlan EA, Sheriff S.1990. *Ann. Rev. Biochem.* **59**:439.

Plan de la tesis

El cuerpo de la tesis esta compuesto de cuatro capítulos que se resumen a continuación:

Capítulo I: ***Modelo de interacción TCR/péptido/MHC***. En este capítulo se revisa el estado actual del conocimiento sobre la interacción TCR/péptido/MHC, sin comentar los resultados reportados en los artículos que conforman la tesis, para evitar repeticiones. Se incluye este capítulo siguiendo dos objetivos. El primero es brindar una tesis lo más autocontenida posible, es decir, facilitarle al lector las definiciones y los elementos necesarios para comprender los capítulos que siguen. El segundo objetivo, es definir los antecedentes de la tesis.

Capítulo II: ***VIR: A computational tool for the analysis of immunoglobulin sequences***. Este capítulo se conforma por un primer artículo[4] que describe una herramienta de cómputo para estudiar las secuencias de los receptores inmunológicos. VIR (**V**ariable domains of the **I**mmune **R**eceptors) fue originalmente diseñado con el propósito de manejar el gran volumen de información de que actualmente se dispone para los receptores inmunológicos. En la sección *Applications* de este capítulo, se delinean algunos resultados parciales del uso de esta herramienta en las Igs. Una versión ampliada de estos resultados puede verse en las referencias 7, 9, y en la tesis de doctorado de E. Vargas-Madrazo. La aplicación de este paquete de programas, extendido para las secuencias de TCRs, es la base metodológica del capítulo que sigue.

Capítulo III: ***Variability analysis of the T-cell receptor using three variability indexes***. Se conforma por el segundo artículo que se incluye en la tesis.[10] En este capítulo se utilizan simultáneamente tres índices de variabilidad para caracterizar las regiones de interacción con el complejo péptido/MHC en

los TCRs. Por analogía con las Igs de estructura conocida se ha propuesto que el sitio de reconocimiento péptido/MHC en el TCR se forma por seis regiones determinantes de la complementaridad (CDR), tres en el dominio V_α y tres en el dominio V_β. Los resultados que se presentan en este capítulo indican que las regiones hipervariables dentro de los dominios variables de los TCRs coinciden con el CDR-2 y CDR-3 definidos a partir de la homología de las secuencias de TCRs e Igs. Sin embargo, el CDR-1 de ambos dominios del TCR resultó ser no hipervariable. Estos resultados sugieren que el CDR-1, debido a su comportamiento relativamente conservado, puede tener un papel diferente al resto de los CDRs en el reconocimiento del complejo péptido/MHC.

Capítulo IV: ***Molecular modeling of a T-cell receptor bound to a major histocompatibility complex molecule: Implications for the T-cell recognition***. Se conforma por el tercer artículo que se incluye en la tesis.[11] En ausencia de datos de cristalografía de rayos-x para el TCR o para el complejo TCR/péptido/MHC, en este capítulo se obtiene un modelo tridimensional computacional para un complejo TCR/péptido/MHC. El análisis de la estructura es congruente con datos experimentales y con el análisis de variabilidad de las secuencias de TCRs, presentado en el capítulo anterior. En la discusión de este capítulo, partiendo del modelo tridimensional obtenido, se propone un papel para cada CDR en el reconocimiento del complejo péptido/MHC.

Finalmente, a manera de conclusiones, se resume en una sección aparte (**Conclusiones**) los principales resultados de los capítulos anteriores. Además, debido a que cada capítulo incluye sus referencias bibliográficas, al final de la tesis se lista una bibliografía completa con el objeto de facilitar su consulta.

Capítulo I:

Modelo de interacción TCR/péptido/MHC

Introducción

A mediados de la década de los 70's se demostró que las células-T requieren para su activación, del reconocimiento simultáneo de una proteína antigénica y una molécula del complejo mayor de histocompatibilidad (MHC) (Zinkernagel & Doherty, 1974). Años después, a principios de los 80's, se identificó con anticuerpos monoclonales dirigidos contra proteínas específicas de células-T, una glicoproteína de aproximadamente 90 Kd, compuesta de dos de cadenas amino ácidos (Allison et al., 1982). Un año más tarde, se demostró que la unión de anticuerpos específicos contra esta glicoproteína - llamada receptor de células-T (TCR) - activaban las células-T, mimetizando así, la interacción con el complejo antígeno/MHC (Acuto et al., 1983; Haskins et al., 1984). Estos hechos establecieron hacia mediados de los 80's que el TCR es el responsable del reconocimiento específico del complejo antígeno/MHC (Kronenberg et al., 1986).

El TCR se asocia no covalentemente con al menos otras cinco moléculas en la membrana de las células-T *'killer'* o *'helper'* [complejo del CD3 (Davis, 1990)]. La estrecha relación con otras proteínas en la membrana celular ha impedido obtener TCRs solubles y funcionales; materia prima de los estudios de cristalografía de rayos-x. Consecuentemente, en el presente no se dispone de una estructura experimental para el TCR o para el complejo TCR/péptido/MHC.

En ausencia de datos experimentales de cristalografía de rayos-x para el TCR, solo partiendo de sus secuencias de amino ácidos, se han utilizado dos estrategias para estudiar bases estructurales del reconocimiento molecular mediado por este receptor del sistema inmunológico. Estas estrategias se basan en el análisis de variabilidad de las secuencias conocidas de TCR (Kabat et al., 1991), por una parte, y por la otra, en el estudio de los TCRs a la luz de las estructuras tridimensionales conocidas de las Igs (Novotny, 1986, Chothia et al., 1988).

Los resultados derivados de tales estudios junto a la resolución de la primera estructura cristalográfica de una molécula MHC-I, hacia finales del año 1987 (Bjorkman et al., 1987a), brindaron las primeras líneas de evidencias para proponer un modelo hipotético de interacción TCR/péptido/MHC (Davis & Bjorkman, 1988; Chothia et al., 1988; Claverie et al., 1989, ver Figura 6 en el texto). Hasta la fecha, este modelo se sigue utilizando para estudiar las bases estructurales del reconocimiento del complejo péptido/MHC por el TCR (Jorgensen et al., 1992a; Patten et al., 1993; DiGiusto & Palmer, 1994). Este modelo, por su generalidad, ha resultado de gran utilidad para formular hipótesis sobre la evolución de los receptores de sistema inmunológico (Davis, 1990), e interpretar en términos globales, los mecanismos de reconocimiento molecular mediados por el TCR (Jorgensen et al., 1992a). Sin embargo, en su formulación actual, resulta insuficiente para estudiar los detalles finos de la interacción TCR/péptido/MHC. Así por ejemplo, en las conclusiones de un artículo reciente se dice: *"... the current models of the TCR structure* [se refiere al modelo citado] *while being a reasonable approximation of the general configuration of T cell receptors, may not serve to precisely predict the effect of specific residue variations, and their present form may not be sufficiently detailed to genetically engineer T cell receptors with predictable specificities"* (DiGiusto & Palmer, 1994).

En esta década se han obtenido avances tanto metodológicos como experimentales que permiten refinar el estudio estructural de la interacción TCR/péptido/MHC. En la primera parte de este capítulo se analizará el modelo de interacción TCR/péptido/MHC en su formulación actual, junto a una discusión de las evidencias experimentales que le dieron origen. En la segunda parte, se revisarán los avances metodológicos y experimentales que permiten proponer nuevos enfoques para estudiar el problema del reconocimiento molecular mediado por los TCRs.

Modelo de interacción TCR/péptido/MHC

La molécula de TCR se forma a partir de dos cadenas polipeptídicas, α y β, con una longitud que oscila en cada cadena, entre 230 y 304 amino ácidos (Davis, 1990). Estas dos cadenas de amino ácidos se ensamblan cada una, por un dominio amino terminal variable (V) de aproximadamente 110 amino ácidos, un dominio constante (C) de igual longitud, una región transmembrana de 12 a 15 amino ácidos, y una región citoplasmática carboxilo terminal de 20 a 24 amino ácidos (Figura 1a).

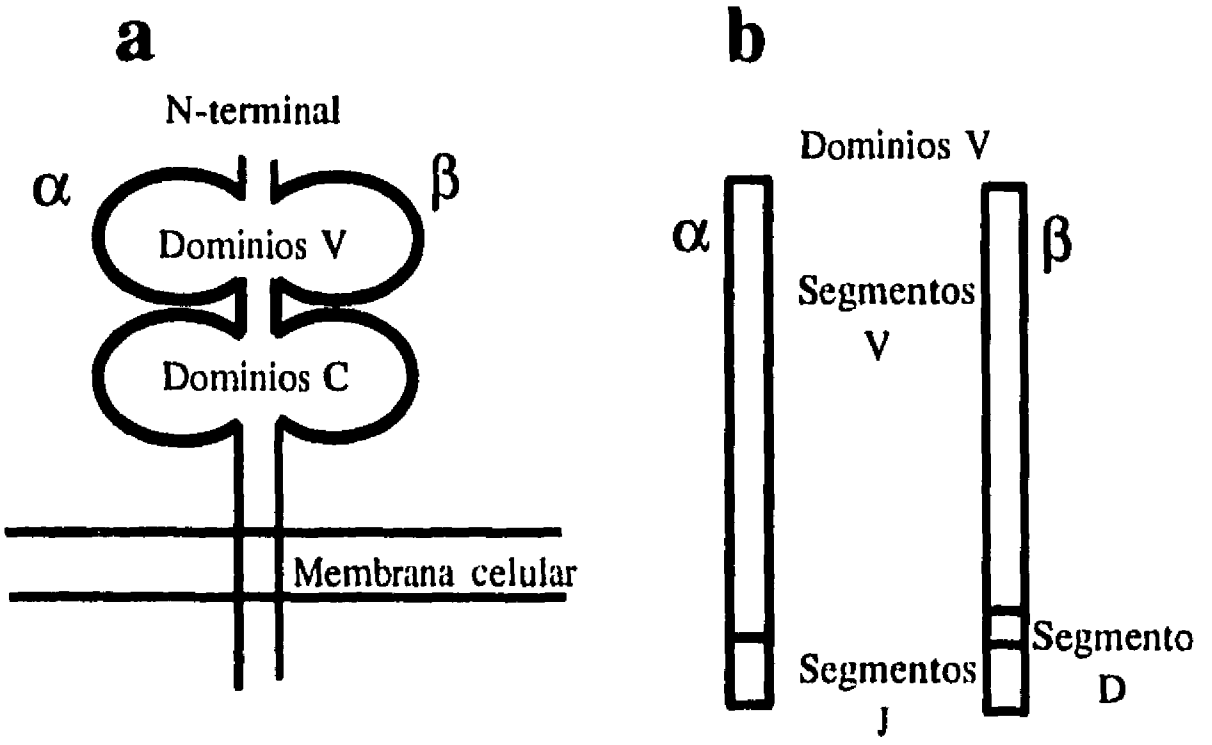


Figura 1a. *Estructura primaria de las cadenas de amino ácidos α y β de los TCRs.*
b. *Codificación genética de los dominios funcionales del TCR (Davis, 1990).*

Los dominios amino terminal, Vα y Vβ, son los responsables del reconocimiento específico del complejo péptido/MHC (Davis, 1990). Genéticamente, estos dominios se codifican por la unión de varias familias de segmentos génicos (Davis, 1990). El dominio Vα es codificado por los segmentos génicos V_α y J_α, mientras que el dominio Vβ, es producto de los segmentos génicos V_β, D y J_β (Figura 1b). El análisis de variabilidad de las secuencias de amino ácidos de estos productos génicos, fue el primer elemento para establecer como los dominios Vα y Vβ de los TCRs interactuan con el complejo péptido/MHC.

El análisis de variabilidad se desarrolló originalmente por Wu & Kabat (1970) para estudiar las secuencias de Igs. Este tipo de análisis parte de establecer que las familias de proteínas que reconocen a ligandos variables - como es el caso de las Igs o de los TCRs - deben tener en sus regiones funcionales alta variación de los amino ácidos (Wu & Kabat, 1970). Para identificar estas regiones, se construye un alineamiento múltiple de secuencias y se analiza cada posición del alineamiento con un índice que estima numéricamente la variabilidad de los amino ácidos (Wu & Kabat , 1970).

Con el índice definido por Wu & Kabat (1970) se estudiaron las primeras secuencias conocidas de Igs. Los resultados indicaron que el sitio de reconocimiento antigénico de esta familia de proteínas, se forma por seis regiones hipervariables o regiones determinantes de la complementaridad (CDR); tres en el dominio V_L y tres en el dominio V_H (dominio variable de la cadena ligera y pesada, respectivamente) (Wu & Kabat 1970, Kabat & Wu, 1971). El valor predictivo de este análisis quedó demostrado, al comprobar en las primeras estructuras tridimensionales de Igs (Poljak et al., 1973; Segal et al., 1974), que estas seis regiones hipervariables se localizaban en los *'loops'* que interactuan con el antígeno. Así, el éxito de esta metodología para identificar las regiones responsables del reconocimiento antigénico en las Igs,

condujo a utilizar, en las primeras secuencias conocidas de TCRs, este tipo de análisis.

Con el uso del índice de Kabat (Wu & Kabat, 1970) en el TCR, se identificó una región de máxima variabilidad en la unión de los segmentos génicos $V_\alpha J_\alpha$ o $V_\beta DJ_\beta$ (Kabat et al., 1991). Esta concentración de variabilidad en las regiones $V_\alpha J_\alpha$ o $V_\beta DJ_\beta$ del TCR, indicó que éstas pudieran tener un papel relevante en el reconocimiento del complejo péptido/MHC (Davis & Bjorkman et al., 1988).

El análisis de las secuencias de TCRs mostró por otra parte, que desde el punto de vista estructural, los dominios Vα y Vβ, pudieran compartir el mismo plegamiento que la porción Fv de las Igs (Novotny et al., 1986; Chothia et al., 1988) (Figura 2). La porción Fv de las Igs se forma por la unión no covalente de los dominios V_H y V_L (Amzel & Poljak, 1979). Cada uno de estos dominios se conforma a su vez por: (1) dos láminas β antiparalelas, (2) seis *'loops'* que interactuan con el antígeno, tres en cada dominio, (3) los *'loops'* restantes que no forman parte del sitio de reconocimiento antigénico, y (4) la interface entre estos dominios que unen espacialmente a los *'loops'* que forman el sitio de reconocimiento antigénico (Amzel & Poljak, 1979).

Puesto que el plegamiento de las Igs y de los TCRs pudiera resultar similar, se propuso que el sitio de interacción con el antígeno en los TCRs, es geométricamente equivalente al de las Igs (ver Figura 2). Esta proposición se reafirmó por el hecho de que los *'loops'* que forman el sitio de interacción con el antígeno en las Igs, son similares en longitud, a los *'loops'* equivalentes en el TCR (Chothia et al., 1988). De esta manera, las Igs de estructura conocida se han utilizado como base para simular la interacción con el complejo péptido/MHC (Davis & Bjorkman 1988; Chothia et al., 1988; Claverie et al., 1989). Esta

posibilidad fue el segundo elemento en la inferencia de un mecanismo general de interacción TCR/péptido/MHC (Davis & Bjorkman 1988; Claverie et al., 1989).

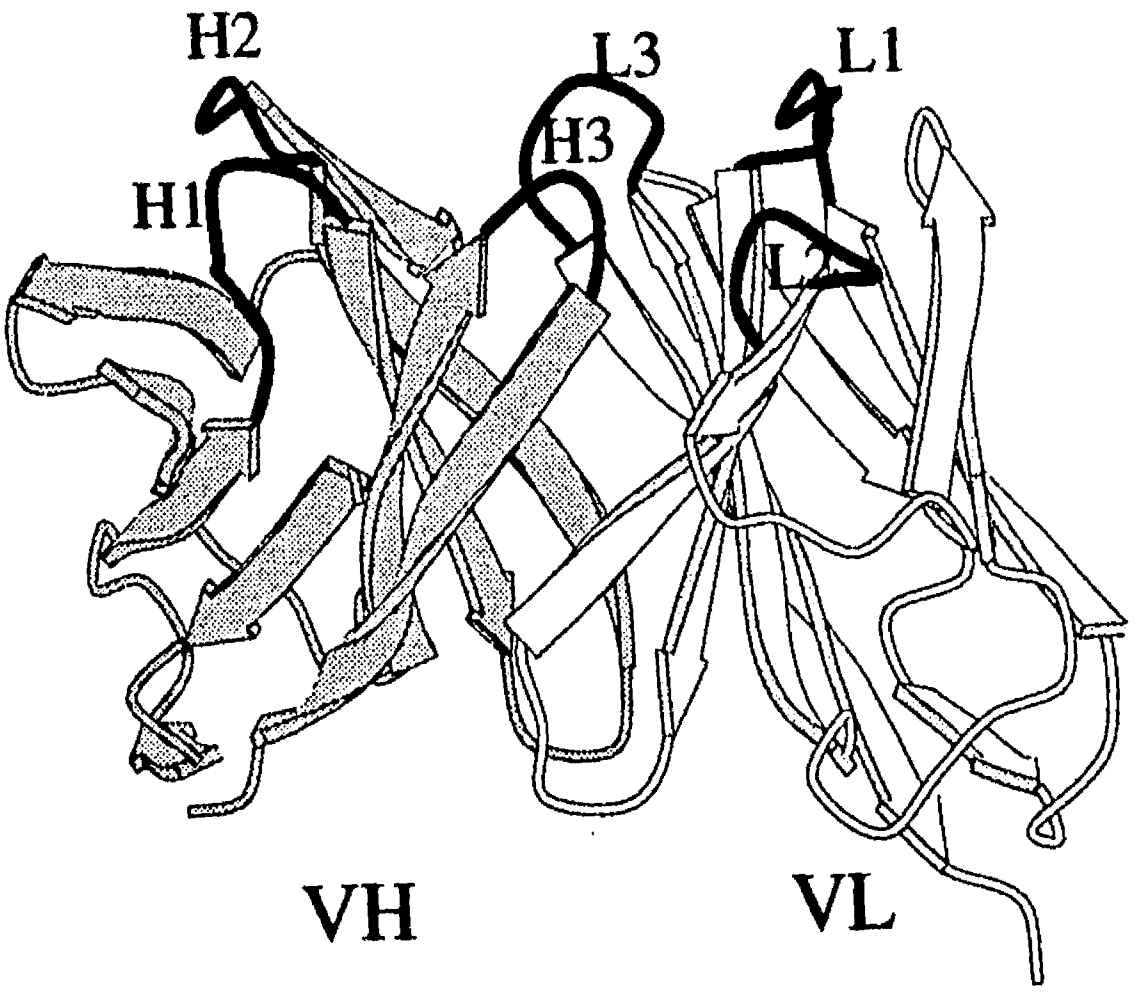


Figura 2. *Estructura de un dimero de dominios V_L y V_H de una Ig. V_H se representa más oscuro que dominio V_L , para facilitar la observación de la interface entre los dominios. Los seis 'loops' que forman el sitio de reconocimiento antigénico se colorean en gris oscuro; para V_L se denotan: L1, L2 y L3, para V_H: H1, H2 y H3. La figura se obtuvo con el paquete de representación molecular en PostScript* MOLSCRIPT *(Kraulis, 1991) utilizando las coordenadas cristalográficas del fragmento Fv D1.3 (Bhat et al., 1990) entrada 1FVA del Banco de datos de estructuras cristalográficas (PDB) (Bernstein et al., 1977).*

El tercer elemento para proponer el modelo de interacción TCR/péptido/MHC, fue la resolución por cristalografía de rayos-x de una molécula de MHC-I (Bjorkman et al., 1987a). La estructura de la molécula de HLA-A2, reveló que el MHC-I se forma por la asociación de cuatro dominios estructurales, α1, α2, α3 y β2m (Figura 3). Los dominios próximos a la membrana, α3 y β2m, son similares a los dominio constantes de las Igs (Bjorkman et al., 1987a). Los dominios distales de la membrana, α1 y α2, difieren substancialmente del plegamiento típico de las Igs, conformados cada uno, por una lámina β antiparalela de cuatro hebras β y una hélice α (ver Figura 3).

En las moléculas funcionales de MHC-I, la asociación de los dominios α1 y α2 forman una plataforma de una lámina β de ocho hebras β antiparalelas, por encima de la cual se ubican las hélices α (Figura 4). Entre las dos hélices α y la lámina β queda una cavidad de aproximadamente 25 Å de longitud, 10 Å de ancho y 11 Å de profundidad (Bjorkman et al., 1987a). Esta cavidad aloja a los péptidos antigénicos en la molécula de MHC hacia la porción más distal de la membrana celular (Bjorkman et al., 1987b). Así, la superficie que se genera por el péptido antigénico junto con la superficie de las hélices α de la molécula de MHC, sugieren una región ideal para la interacción con la molécula de TCR (Bjorkman et al., 1987b).

Si se toma un fragmento Fv de una Ig para simular la estructura de los dominios funcionales de un TCR, al observar el sitio de reconocimiento antigénico desde la orientación que 've' el complejo péptido/MHC, el CDR-3 de ambos dominios (L3 y H3) queda hacia el centro del sitio de reconocimiento, mientras que los CDR-1 y 2 (L1, L2, H1 y H2) quedan hacia los flancos (Figura 5). El área que forman los *'loops'* hipervariables de la Ig es aproximadamente de 500 $Å^2$ (25 Å de

largo por 20 Å de ancho). Esta área es similar a la que forman las hélices α de la molécula de MHC, en la parte superior de la cavidad donde se aloja el péptido antigénico (ver Figura 4).

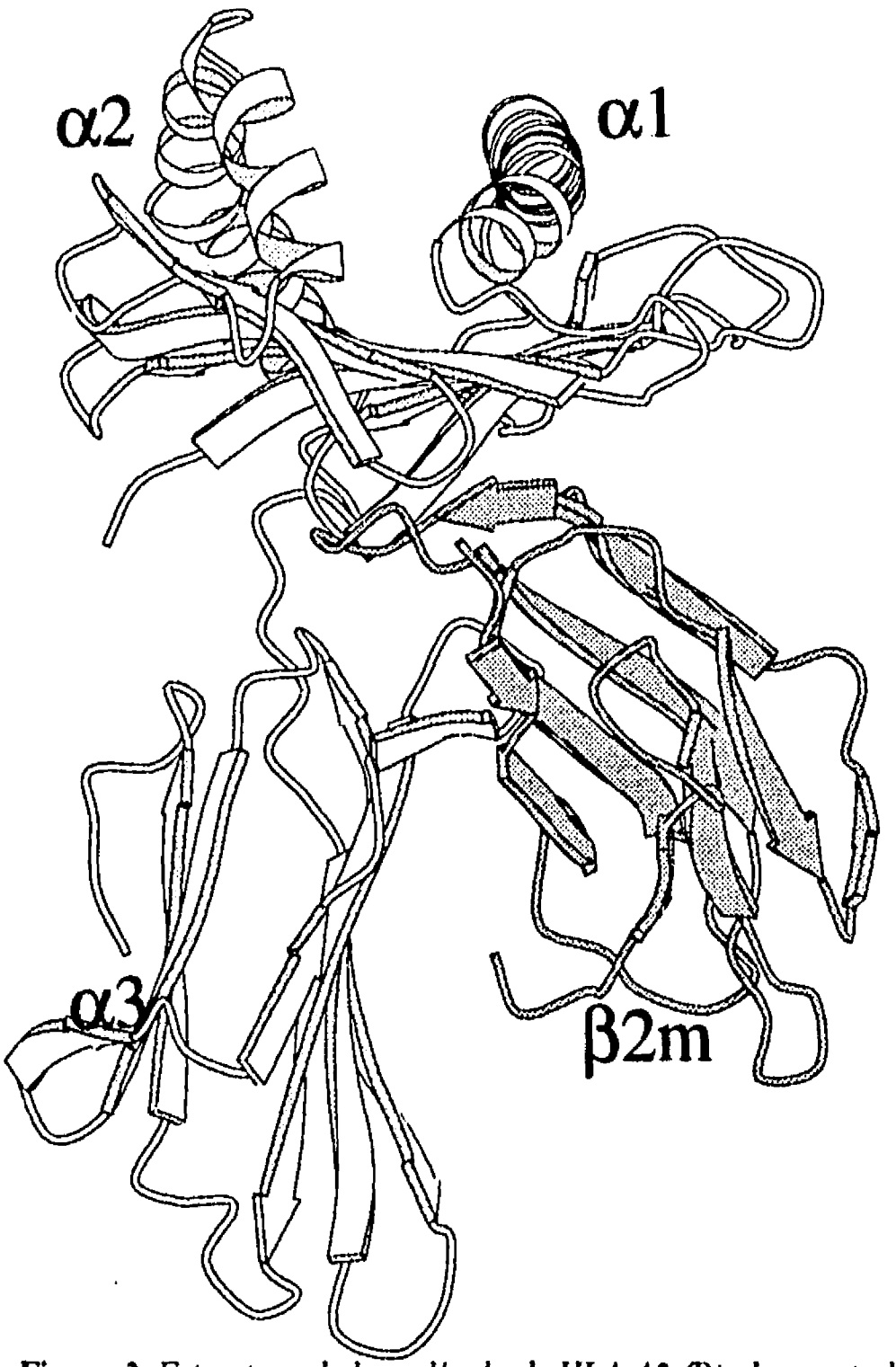


Figura 3. *Estructura de la molécula de HLA-A2 (Bjorkman et al., 1987a), entrada del PDB (1HLA). La β2m se representa más oscuro para diferenciarla del resto de la estructura. La Figura se obtuvo con el paquete de representación molecular en PostScript* MOLSCRIPT *(Kraulis, 1991).*

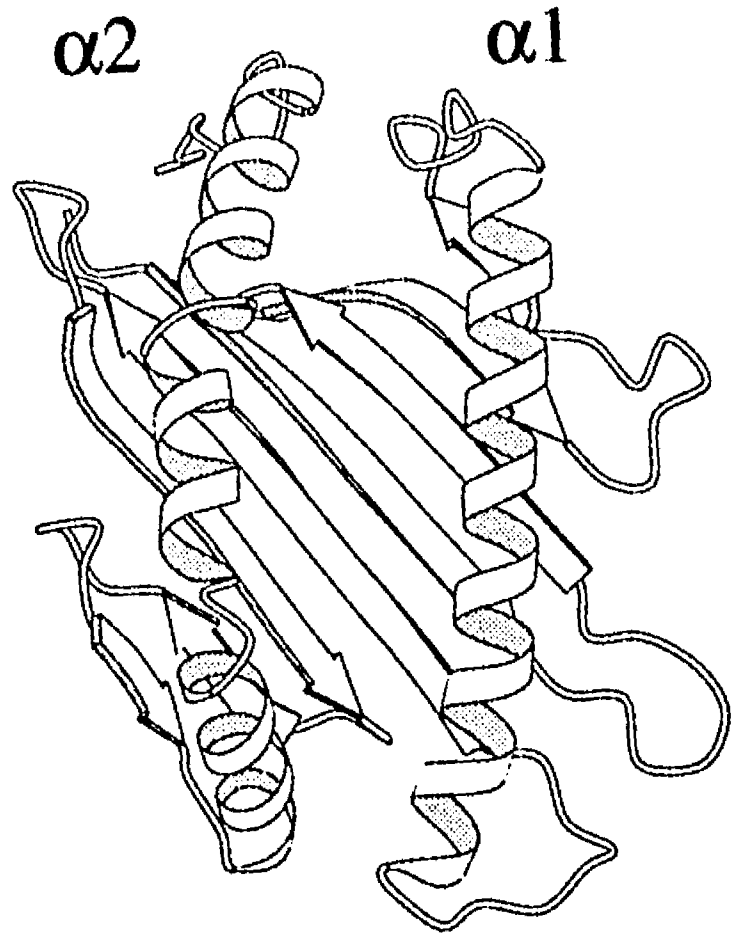


Figura 4. *Estructura los dominios α1 y α2 de la molécula HLA-A2 rotada 90° respecto a la Figura 3.*

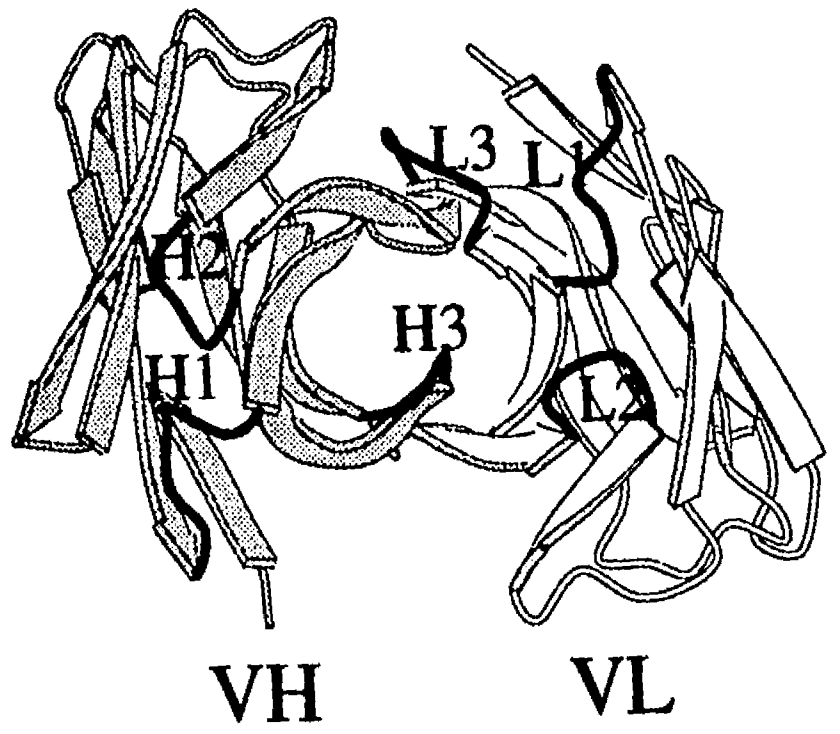


Figura 5. *Sitio de reconocimiento de una Ig visto desde el antigeno. La Figura se obtuvo rotando 90° la estructura de la Figura 2 (Fv D1.3).*

Por analogía con las Igs, la concentración de la variabilidad en los TCRs corresponde estructuralmente con el CDR-3 (Davis & Bjorkman, 1988; Chothia et al., 1988, Claverie et al., 1989). Entonces, si se coloca la estructura del Fv de la Ig sobre la molécula de MHC, de manera tal que los CDR-3 de ambos dominios queden alineados en la dirección de péptido antigénico, el CDR-1 y CDR-2 de uno de los dominios del Fv, reconocen la hélice α de un dominio de la molécula de MHC, mientras que el CDR-1 y CDR-2 del otro dominio reconocen la otra hélice α del MHC (Figura 6).

Debe notarse que partiendo de la información anterior no se puede decidir cual es la orientación correcta de la molécula de Ig sobre el MHC (Davis & Bjorkman, 1988; Claverie et al., 1989). De aquí que se consideren dos modelos alternativos. Los modelos alternativos se obtienen por rotaciones de 90° o de 180° en la molécula de Ig respecto a la molécula de MHC. Una rotación de 90° en la molécula de Ig, implica que los CDR-3 interactuan directamente con las hélices del MHC, mientras que una rotación 180° en la molécula de Ig, implicaría solo un cambio en la orientación relativa de la Ig con respecto a la molécula de MHC (Davis & Bjorkman, 1988; Claverie et al., 1989).

Este modelo por su generalidad ha sido de gran utilidad para formular hipótesis sobre la evolución de los receptores del sistema inmunológico (Davis, 1990), diseñar experimentos (Ajitkumar et al., 1988; Nalefski et al., 1990; Patten et al., 1993), y explicar los rasgos generales de interacción TCR/péptido/MHC (Jorgensen et al., 1992a). Sin embargo, como se mencionó antes, se admiten varias orientaciones en la molécula de TCR sobre el complejo péptido/MHC. Además, si se traslada la molécula de Ig sobre la superficie del MHC, diferentes residuos del TCR pudieran interactuar con el péptido y/o con el MHC. Esta falta de precisión en el

modelo descrito ha impedido avanzar en el estudio estructural de los detalles finos del reconocimiento mediado por los TCRs (Prochnicka-Chalufour et al., 1991).

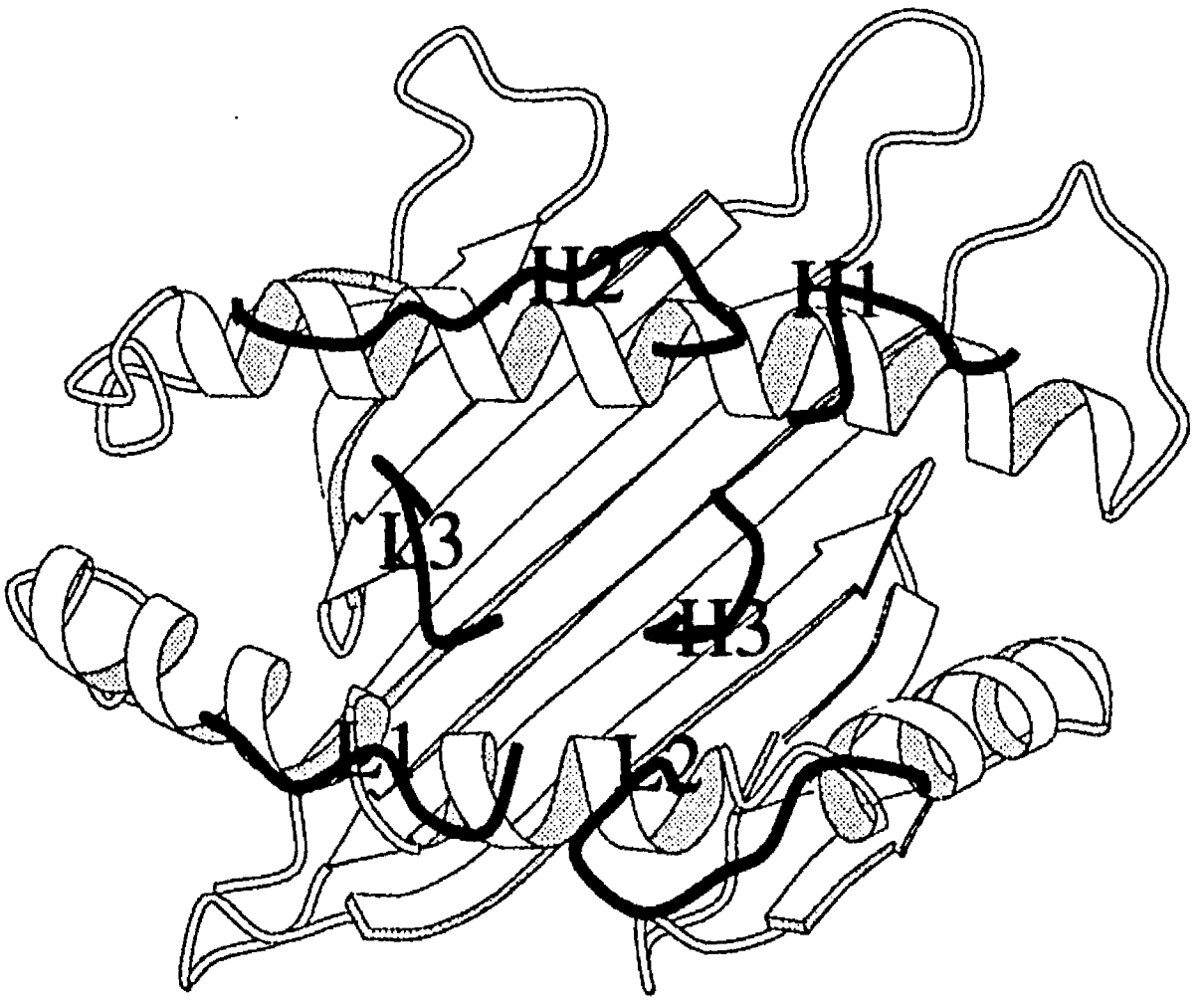


***Figura** 6. Modelo de interacción TCR/péptido/MHC según Davis & Bjorkman (1988), Chothia et al. (1988) y Claverie et al., (1989) obtenido a partir de una Ig simulando una molécula de TCR y la estructura cristalográfica del HLA-A2. Orientación de la molécula de Ig donde L1 y L2 interactuan con el dominio α1 del MHC-I y H1 y H2 interactuan con el dominio α2 del MHC-I. La figura se obtuvo colocando la estructura de la Figura 2 (Ig) sobre la molécula de HLA-A2 (Figura 4) como sugiere el modelo de interacción TCR/péptido/MHC. En la Ig solo se muestran los 'loops' hipervariables para facilitar la observación del modelo.*

En las secciones que continúan se revisarán los avances metodológicos y experimentales que permiten refinar el estudio estructural de la interacción TCR/péptido/MHC, su aplicación, constituye los capítulos que siguen de la tesis.

Análisis de variabilidad de los TCRs

En las Igs, los dos primeros CDRs se codifican por los segmentos génicos V_H y V_L, mientras que el tercer CDR se localiza en la unión de los segmentos génicos V_LJ y V_HDJ_H (Wu & Kabat 1970, Kabat & Wu, 1971). Similar a las Igs, en los TCRs, el análisis de variabilidad con el índice de Wu & Kabat (1970) identifica una región hipervariable en la unión de los segmentos genéticos $V_\alpha J_\alpha$ o $V_\beta DJ_\beta$ (Kabat et al., 1991). Sin embargo, en los primeros ~ 90 amino ácidos, codificados por los segmentos génicos, V_α o V_β, no se han obtenido de manera concluyente, regiones hipervariables (Kronenberg et al., 1986; Davis, 1990, Kabat et al., 1991).

La falta de definición de regiones hipervariables en los segmentos génicos Vα y Vβ responde a que los TCRs son mucho más variables a lo largo de su secuencia de amino ácidos con respecto a las Igs. Si se considera que los TCRs son anteriores a las Igs desde el punto de vista evolutivo (Davis et al., 1990), la explicación de tal comportamiento resulta de que los TCRs han tenido mayor tiempo de evolución y consecuentemente mayor acumulación de mutaciones neutrales en sus secuencias de amino ácidos (Jores et al., 1990). La mayor dispersión de mutaciones a lo largo de las secuencias de TCRs con respecto a las Igs, impide delimitar regiones de concentración de variabilidad (Kabat et al., 1991). Desde el punto de vista funcional, la falta de definición de los CDR-1 y 2 en los TCRs limita el estudio de estas regiones en el reconocimiento del complejo péptido/MHC.

Para resolver esta dificultad se propuso a principios de los 90's una modificación del índice de Kabat junto con un método de filtrado de los perfiles de variabilidad por series de Fourier (Jores et al., 1990). Con la aplicación de esta nueva metodología a 159 secuencias de la cadena amino ácidos Vβ se delinearon cuatro CDRs, tres de los cuales corresponden con los CDRs de las Igs (Jores et al., 1990). En el caso de las secuencias de la cadena de amino ácidos Vα, la aplicación de esta metodología resultó infructuosa, puesto que no se delimitaron regiones de hipervariabilidad en este dominio de los TCRs (Jores et al., 1990).

Recientemente, se propuso una nueva medida de variabilidad (Shenkin et al., 1991). Este índice se basa en la fórmula de entropía de Shannon (Shenkin et al., 1991). Según los autores (Shenkin et al., 1991), este índice de variabilidad tiene mejores propiedades matemáticas que los índices anteriores, como por ejemplo, continuidad en todo su dominio (Shenkin et al., 1991). Esta medida de variabilidad aún no se ha aplicado a los TCRs.

Es deseable por lo tanto, aplicar los tres índices de variabilidad simultáneamente a las secuencias de los dominios funcionales de los TCRs, seguidos del filtrado con series de Fourier de los perfiles de variabilidad. El objetivo de aplicar esta metodología a los TCRs es identificar las regiones equivalentes al CDR-1 y 2 de las Igs en los TCRs. El uso simultáneo de los tres índices, pudiera compensar las deficiencias de cada uno de ellos por separado en estimar la variabilidad. El filtrado con series de Fourier daría un criterio objetivo de comparación de cuales son las regiones que se identifican con cada índice. Esta proposición es el capítulo III de la tesis.

Datos experimentales sobre la interacción TCR/péptido/MHC

Los estudios experimentales para definir los componentes moleculares de la interacción TCR/péptido/MHC, hasta hace tres años, eran escasos e indirectos (para una revisión del tema ver Jorgensen et al., 1992a). Esto impedía, junto a la ausencia de estructuras tridimensionales para el TCR o para el complejo TCR/péptido/MHC, profundizar en el estudio del reconocimiento molecular mediado este receptor del sistema inmunológico.

En los últimos tres años, sin embargo, estudios basados en el análisis de las secuencias de TCR en respuestas a péptidos naturales o sintéticos, en el contexto de diferentes alelos de MHC, evidenciaron que deben existir contactos físicos entre el TCR y el complejo péptido/MHC (Jorgensen et al., 1992a). Uno de los sistemas mejor estudiados es el complejo 5C.C7/MCC (83-103)/I-E^k (TCR/péptido/MHC, respectivamente) (Hedrick et al., 1988; Jorgensen et al., 1992b). En este sistema utilizando un conjunto de péptidos del MCC con mutaciones puntuales, se determinó que cargas reversas (+ → - o - → +) en los residuos 99 y 102 del péptido, seleccionaban diferentes TCR funcionales (Jorgensen et al., 1992b). Las mutaciones compensatorias que se expresan en el TCR se localizan en las regiones de unión $V_\alpha J_\alpha$ y $V_\beta D J_\beta$. Estos resultados indican que estas regiones interactuan directamente con el péptido antigénico, como predice el modelo general de interacción TCR/péptido/MHC antes descrito (Jorgensen et al. 1992b).

Los contactos entre residuos del TCR y el péptido antigénico permiten además seleccionar la orientación en que el TCR reconoce al complejo péptido/MHC (Jorgensen et al., 1992b). Así, esta información experimental, aporta un elemento adicional que restringe los grados de libertad (rotaciones y traslaciones)

de la molécula de TCR con respecto al complejo péptido/MHC. El modelo tridimensional del complejo 5C.C7/MCC(93-103)/I-E^k es el capítulo IV de la tesis.

Referencias

Acuto O, Hussey R, Fitzgeral K, Protentis J, Meuer S, Reinherz E. 1983. The human T-cell receptor: Appearance in ontogeny and biochemical relationship of α and β subunits on IL-2 dependent clones and T-cell tumors. *Cell. 34:*717-726.

Ajitkumar P, Geier SS, Kesari KV, Borriello F, Nakagawa M, Bluestone JA, Saper MA, Wiley DC, Nathenson SG. 1988. Evidence that multiple residues on both the α-helices of the class I MHC molecule are simultaneously recognized by the T cell receptor. *Cell 54:*47-56.

Allinson JP, McIntyre BM, Bloch D. 1982. Tumor-specific antigen of murine T-lymphoma defined with monoclonal antibody. *J. Immunol.129:*2293-3000.

Amzel, LM, Poljak RJ. 1979. Three-dimensional structure of immunoglobulins. *Annu. Rev. Biochem.* **48**, 961-997.

Bernstein FC, Koetzle TF, Williams GJB, Meyer EF Jr, Brice MD, Rodgers JR, Kennard O, Shimandouchi T, Tasumi M. 1977. The Protein Data Bank: A computer-based archival file for macromolecular structures. *J. Mol. Biol. 112*: 535-542.

Bhat TN, Bentley GA, Fischmann TO, Boulot G, Poljak RJ. 1990. Small . rearragements in structures of Fv and Fab fragments of antibody D1.3 on antigen binding. *Nature 347:*483-485.

Bjorkman PJ, Saper MA, Samraoui B, Bennett WS, Strominger JL, Wiley DC. 1987a. Structure of the human class I histocompatibility antigen, HLA- A2. *Nature 329:*507-512.

Bjorkman PJ, Saper MA, Samraoui B, Bennet WS, Strominger JL, Wiley DC. 1987b. The foreign antigen binding site and T cell recognition regions of class I histocompatibility antigens. *Nature. 329:*512-519.

Chothia C, Boswell DR, Lesk, AM. 1988. The outline structure of the T-cell α:ß receptor. *EMBO J. 7*:3745-3755.

Claverie J-M, Prochnicka-Chalufour A, Bougueleret L. 1989. Implications of a Fab-like structure for T-cell receptor. *Immunol. Today. 10*:10-14.

Davis MM, Bjorkman PJ. 1988. T-cell antigen receptor genes and T-cell recognition. *Nature 334*:395-402.

Davis MM. 1990. T cell receptor gene diversity and selection. *Annu. Rev. Biochem. 59*: 475- 496.

DiGiusto DL, Palmer, E. 1994. An analysis of sequence variation in the β chain framework and complementarity determining regions of an allo-reactive T cell receptor. *Mol. Immunol. 31*: 693-699.

Haskins K, Kappler J, Marrack P. 1984. The major histocompatibilty complex-restricted antigen receptor on T-cells. *Annu. Rev. Immunol. 2*:51-66.

Hedrick SM, Engel I, McElligot DL, Fink, PJ, Hsu ML, Hansburg D, Matis A. 1988. Selection of amino acid sequences in the β chain of T cell antigen receptor. *Science 239:*1541-1544.

Jores R, Alzari PM, Meo T. 1990. Resolution of hypervariable regions in T-cell receptor ß chains by a modified Wu-Kabat index of amino acids diversity. *Proc. Natl. Acad. Sci. USA. 87:*9138-9142.

Jorgensen JL, Reay PA, Ehrich EW, Davis MM. 1992a. Molecular components of T-cell recognition. *Annu. Rev. Immunol. 10*:835-873.

Jorgensen JL, Esser U, Fazekas de St. Groth B, Reay PA, Davis MM. 1992b. Mapping T-cell receptor-peptide contacts by variant peptide immunization of single-chain transgenics. *Nature 355*:224-230.

Kabat EA, Wu TT. 1971. Attempts to locate complementarity determining residues in the variable positions of light and heavy chains. *Ann. NY. Acad. Sci. 190*:382-383.

Kabat EA, Wu TT, Perry HM, Gottesman KS, Foeller C. 1991. Sequences of Proteins of Immunological Interest. 5th ed. Bethesda, MD: U.S. Department of Health and Human Services, Public Health Service, National Institutes of Health, NIH Publication No. 91-3242.

Kraulis PJ. 1991. MOLSCRPT: A program to produce both detailed and schematic plots of protein structures. *J. Appl. Cryst. 24:*946-950.

Kronenberg M, Siu G, Hood LE, Sastri N. 1986. The molecular genetics of the T-cell antigen receptor and T-cell antigen recognition. *Annu. Rev. Immunol. 4:*529-591.

Nalefski EA, Wong JGP, Rao A 1990. Amino acids substitution in the first complementary-determining region of a murine T-cell receptor α chain affect antigen-major histocompatibility complex recognition. *J. Biol. Chem. 265:*8842-8846.

Novotny J, Tonegawa S, Saito H, Kranz DM, Eisen H. 1986. Secondary, tertiary, and quaternary structure of T-cell-specific immunoglobulin-like polypeptide chains. *Proc. Natl. Acad. Sci. USA. 83*:742-746.

Patten, P.A., Rock, EP, Sonoda, T, Fazekas de St.Groth, B, Jorgensen, JL, Davis, MM. 1993. Transfer of putative complementary-determining regions loops of T-cell receptors V domains confer toxin reactivity but not peptide/MHC specificity. *J. Immunol. 150*: 2281-2294.

Poljak RJ, Amzel LM, Avey H P, Chen B L, Phizacherley RP, Saul F. 1973. Three-dimensional structure of the Fab' fragment of a human immunoglobulin at 2.8-Å resolution. *Proc. Natl. Acad. Sci. U.S.A. 70*:3305-3310.

Prochnicka-Chalufour A, Casanova J-L, Avrameas S, Claverie J-M, Kourilsky P. 1991. Biased amino acids distribution in regions of the T cell receptors and MHC molecules potentially involved in their association. *Int. Immunol. 3*:853-864.

Segal DM, Padlan AE, Cohen GH, Rudikoff S, Potter M, Davies DR. 1974. The three-dimensional structure of a phosphorylcholine-binding mouse immunoglobulin Fab and the nature of the antigen binding site. *Proc. Natl. Acad. Sci. USA. 71*:4298-4302.

Shenkin PS, Erman B, Mastrandrea LD. 1991. Information-theoretical entropy as a measure of sequence variability. *Proteins. 11*:297-313.

Wu TT, Kabat EA. 1970. An analysis of the sequences of the variable regions of Bence Jones proteins and myeloma light chains and their implications for antibody complementarity. *J. Exp. Med. 132*: 211-250.

Zinkernagel RM, Doherty PC. 1974. Immunological surveillance against altered self components by sensitised T lymphocytes in lymphocytic choriomeningitis. *Nature 251*:547-548.

Capítulo II:

VIR: A computational tool for analysis of immunoglobulin sequences

BIO1479
/RLB

BioSystems VOL (1994) PP–PP

VIR: A computational tool for analysis of immunoglobulin sequences

J.C. Almagro*[a], E. Vargas-Madrazo[b], R. Zenteno-Cuevas[a], V. Hernández-Mendiola[b], F. Lara-Ochoa[a]

[a]*Instituto de Química, UNAM Circuito Exterior, Ciudad Universitaria, C.P. 04510, México, D.F., México*
[b]*Instituto de Investigaciones Biológicas, Universidad Veracruzana, Xalapa, Veracruz, México*

Received 27 May 1994; accepted 12 August 1994

Abstract

In this paper a microcomputer software named VIR (Variable domains of the Immune Receptors) is reported. This package can be used in sequence studies of immunoglobulin variable domains. The main features of the VIR software in the sequences management are: (1) ease of information recovery/extraction from amino acid sequences; and (2) its capability to obtain multiple sequence alignments with predefined characteristics (i.e. specie and/or specificity). As an analytical tool, the VIR package employs such multiple sequence alignments to compute: (1) tables showing amino acid frequencies; (2) three variability indexes; (3) identity matrices; (4) random samples; and (5) sequences with possible canonical structures. Thus the software reported here here is proposed as a useful tool to carry out detailed studies of immunoglobulin variable domains.

Keywords: Data bases; Variability analysis; Sequence analysis; Pattern recognition; Canonical structures

1. Introduction

One of the core problems in molecular biology concerns the constraints determining structure-function relationships in proteins. In principle, it can be safely assumed that such constraints can be found by analyzing amino acid patterns in key positions of a multiple sequence alignment (Zuckerkandl, 1976). In the particular case of the immune system, where the only source of specificity at molecular level seems to come from two proteins: T-cell receptors (TCR) and immunoglobulins (Ig), a long standing question prevails: is it possible to find amino acid patterns to reveal the structure-function relationships in the antigen binding sites of TCR and Ig?.

Prediction of the antigen binding site or complementarity determining regions (CDRs) in Ig by variability analysis, prior to the resolution of Ig structures (Wu and Kabat, 1970), made possible the beginning of understanding specificity (function) mediated by these molecules. Subsequently to the resolution of several Ig structures, analysis of its three-dimensional conformation have revealed canonical structures in five of its six CDRs (Chothia and Lesk, 1987; Chohia et al., 1989). Existence of these canonical structures in the CDRs of Ig implies that just a few main-chain conformations are present in a large set of Ig molecules with

* Corresponding author, Tel.: +525 616 2576; Fax: +525 616 2217; E-mail address: salazar@redvax1.dgsca.unam.mx.

0303-2647/94/$07.00
SSDI 0303-2647(94)01479-Q

Dear Author,
Please return corrections as soon as possible by fax or courier. Originals by Express Mail
Thank You

different loop amino acid sequences and specificities (Chothia and Lesk, 1987; Chothia et al., 1989). Recent studies have enhanced the knowledge of Ig structure-function relationships by finding definite amino acid bias when analyzing positions that interact with antigens in antigen-antibody complexes (Mian et al., 1991; Lara-Ochoa et al., 1994). This suggests a general mechanism underlying the molecular recognition mediated by Ig (Lara-Ochoa et al., 1994).

A next step is to correlate functional and structural information. In the opinion of the authors, the main limitation to achieve this, has been the absence of user-friendly computational tools to perform such analysis. To circumvent this difficulty in this paper a microcomputer software named VIR (Variable domains of the Immune Receptors), developed to easily retrieve and analyze Ig amino acid sequences, is reported. VIR interfaces with Kabat's data base, which is known to possess all currently updated information on Ig sequences (see for example Kabat et al., 1991). To do so, this program makes use of SEQHUNT internet service which, in turn, has the necessary tools to access this information. The data are translated to VIR data base format. This data base structure optimizes the information management and its recovery form Ig sequences, and allows use of the different analytical tools available with the VIR package.

On its analytical side, VIR uses two different approaches: statistical analysis and pattern recognition. Statistical analysis provides computation of: (1) tables of amino acid usage by position; and (2) three different variability indexes. Correspondingly, the pattern recognition approach provides: (1) patterns compatible with key residues responsible for determining the Ig-fold; and (2) canonical structures in the CDRs of Ig sequences.

In order to properly outline the work in the following sections, a complete — but not exhaustive — description of the VIR data bases format and the management and analytical tools are given in the first place. Secondly, examples of the two different approaches in Ig sequence analysis are shown and discussed. Finally, the significance of the VIR software is discussed in the light shed by-current trends in data bases management and sequence analysis.

2. Description of the VIR package

All programs were written in the Turbo Pascal version 5.5 language, and were compiled to run on an IBM-compatible PC microcomputer. A user-friendly menu-driven graphical interface system was developed with on-line help available to explain each and every item appearing in the main menu or submenus.

In order to readily start using the package two data bases are available with it. One corresponds to the variable domains of light (V_L) chains while the other does so with the heavy (V_H) chains. Each data base contains a multiple sequence alignment as well as the information associated with each sequence. In the V_H case, the data base contains all the sequences compiled in the Kabat's data base up to April 1994. For V_L the data base contains the sequences from the same source of V_H sequences. Both of them are ASCII files. Therefore, data bases available with the program, updated ones, or those created by the user, can be displayed, edited or printed using a text processor capable of importing ASCII files (WordPerfect, Word, Word for Windows, and the like).

2.1. VIR data base structure

Functional, technical and bibliographic information in the VIR data bases is codified on a label associated with each sequence in the multiple alignment according to the following description:

Describer	Position on label
Technique	1
Year	2–3
Species	4–5
Journal	6–8
Volume	9–11
First page of the paper	12–16
Name given by Kabat et al. (1991)	17–30
Particular specificity	31–35
General specificity	36
Binding constant (if known)	37
Subclass (kappa or lambda)	38
Subgroups of Kabat et al. (1991)	39–40
Space	41–50
Sequence	51

Thus, each line of the VIR data base is composed of a label (left side) and the sequence codified in the one letter amino acid code (right side), for example:

```
position:  1         10        20        30        40        50
           |    |    |    |    |    |    |    |    |    |    |
           N89RNPNA86 2341 3.14.9          LEVANUNK5*********DIQMF...AS
           |----------Label-----------------------------------|sequence|
```

The label is used by the management tools of the package, while the sequence is processed by the analytical tools (see below). In order to codify the describers of the label, a commitment was made between classifying the maximum information and avoidance redundancies or ambiguities. In the case of specificity, judged by the authors as the most important functional describer, it should be emphasized that two describers, particular and general, are used. General specificity refers to the groups of specificities (for example anti-protein Ig), and it can thus be used to directly obtain data bases with global specificities. Particular specificities refers to the specificities (for example anti-lysozyme Ig), allowing to stratify the search.

Concerning the multiple sequence alignments, two conventions have been proposed in the literature regarding insertions numbering and placement inside CDRs of Ig: (i) Kabat et al. (1991); and (ii) Chothia et al. (1987). We chose the convention proposed by Chothia et al. (1987) because it has been developed on the basis of comparisons among Ig of known three-dimensional structures.

2.2. Tools of management and analysis in the VIR package

The package consists of a master program that controls four main modules: Open, Data, Analysis and Up-to-date (Fig. 1). (Words in bold italics are those shown to be chosen in all menus and submenus below described).

1. The Open module is the one in charge of loading data bases, either those available with the package *Ig* or any *Other* crated by the user in VIR format.

2. The Data module is the data base manager of the package. It is divided in the two functions below indicated by the first indentation level:

(i) To provide a complete characterization of the data base currently being used, the Information submenu (see Fig. 1) is used. This submenu report:

Total number of sequences.

Number of sequences by *Technique* (amino acid sequences or nucleotide translation).

Number of sequences according to the *Year* of publication.

Number of sequences by *Specie*.

Number of sequences by *Specificity*.

Number of sequences with reported *Affinity*.

Number of kappa or lambda (*k* or *l*) Sequences. (Only applicable to V_L.)

Number of sequences by Kabat's *Subgroup* (Kabat et al., 1991).

Matrix: this item allows self-comparison of the sequence data base (that is to say, with

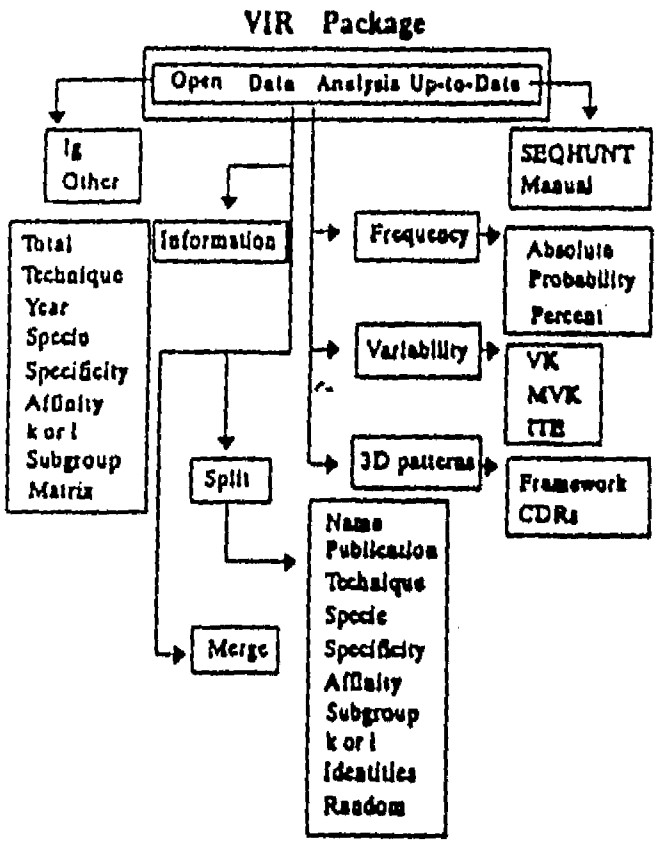


Fig. 1. VIR package menu system.

itself) making possible to identify redundancies in it. (ii) To create new data bases with predefined features (sequences from a single specificity, for example). This is achieved using the Split submenu and the Merge option (see Fig. 1). The latter merges two data bases, while the former creates a new data base given one pre-established criterion. Below, the different criteria concerning the Split submenu are explained. Generation of a data base with sequences given its *Name(s)*.

Generation of a data base according to bibliographic information (*Publication*)

Generation of a data base according to the *Technique* used to obtain the amino acid sequences (amino acid sequences or nucleotide translation). Generation of a data base with sequences of a single *Specie*.

Generation of a data base with sequences of a single *Specificity* (General and particular).

Generation of a data base selecting those sequences with known *Affinity* constant.

Generation of a data base selecting those sequences within a Kabat *Subgroup* (Kabat et al., 1991).

Generation of a data base containing only kappa or lambda (*k* or *l*) sequences.

Generation of a data base obtained by analyzing *Identities* among amino acids with a query sequence within a range of maximum and minimum Identity threshold.

Generation of a *Random* data base with a given number of sequences to be taken from the data base in use.

For the Species, Specificity, Publication and Subgroup options, submenus listing the corresponding different classifications are displayed by the package.

Output files from the Split submenu are in VIR data base format (see VIR data bases structure above). Hence, combining the Split submenu and the Merge option, new data bases performing any combination of the described items can be obtained, for example, a not-before-hand included data base containing human anti-lysozyme with known Affinity constant.

3. The Analysis module is the third one of the package and was designed bearing two conceptually different techniques in mind to analyze retrieved data bases: statistical techniques and sequence pattern analysis. Concerning statistical analysis two methods are offered, the most simple one computes of amino acid frequencies via the Frequency submenu composed of three options: *Absolute*, *Probability* and *Percent* (see Fig. 1). Output coming from this submenu are tables in ASCII format listing the positions (rows) and the numerical values associated with each amino acid (columns). The second statistical technique is Variability analysis (Variability submenu, see Fig. 1). Three variability indexes are provided: Kabat's variability index (*VK*) (Wu and Kabat, 1970), modified Kabat's variability index (*MVK*) (Jores et al., 1990), and informational-theoretical entropy index (*ITE*) (Shenkin et al., 1991). Output coming from this submenu consists of tables in ASCII format listing the positions in the amino acid sequence and the numerical values given to each position in the multiple sequence alignment.

The second analytical (non-statistical) technique before mentioned, namely pattern recognition (3D patterns submenu in Fig. 1), was developed to estimate the relationship among Ig of known three-dimensional structures and amino acid sequences. The *Framework* option compares amino acid sequences with a pattern of residues identified as mainly responsible for determining the Ig-fold (Chothia et al., 1988), while the *CDRs* option searches for canonical structure patterns in the data base currently being used.

4. The Up-to-date module is an interface with the SEQHUNT internet service to update data bases already created. As in previous examples, it is further subdivided in two items: *SEQHUNT* and *Manual*. SEQHUNT converts VIR data base format into Kabat's data base format and vice versa. This is the interface with the Kabat's data bases containing currently updated sequences of Ig variable domains. Manual allows the user to actualize the data bases or to analyze a sequence not reported by the SEQHUNT service simply by typing its sequence.

3. Applications

In order to show the potential use of the package, two applications of the VIR software are

described in this section. Application (a) shows an example of a statistical analysis, while application (b) performs an analysis of a typical immunologic problem.

3.1. Example of predefined data bases and amino acid usage

It has recently been shown that several positions within CDRs of Ig — historically considered as randomly hypervariable — have preference for certain amino acids (Vargas-Madrazo et al., 1994a; Lara-Ochoa et al., 1994). The bias found in the amino acid usage of these positions could be due to: (i) a general feature of the Ig molecular recognition; or (ii) some artefact introduced by certain over-represented specificities resulting in predominance of some sequences in the data base (e.g. anti-hapten Ig). The second possibility could be discarded by building balanced samples with equal number of sequences of different specificities and then, comparing their amino acid usage with that of the total sample (total data base).

Thus, two samples were built as follows:

Sample 1. Using the Split submenu in the Data module a data base was generated by randomly choosing one sequence from each specificity (210 different specificities) in the total data base.

Sample 2. Following the already mentioned procedure another data base was built choosing five sequences from specificities having at least five sequences (52 different specificities) in the total data base.

Each data base was loaded with the Other option in the Open module, and using the Percent option in the Frequency submenu, the corresponding tables of amino-acids percent by positions were computed. The results for one position in the V_L are shown in Fig. 2. It was observed that the patterns of amino acid usage found in the three samples are rather similar. Statistical calculations confirm this result, and the same behavior was obtained for other positions in CDRs (Vargas-Madrazo et al., in preparation). Therefore, this analysis discards the existence of any bias as a consequence of a predominance of certain specificities in the total sample analyzed, suggesting that the preferential use of certain amino acids in CDRs is a general feature of the Ig molecular recognition.

3.2. Example of canonical structures

Genetically, the first two CDRs of V_H (H1 and H2) of Ig are produced by the V_H germline genes (Tonegawa, 1983). Additionally to V_H germline genes, most species retain a pool of V_H pseudogenes estimated to be 30% of the total number of V_H genes (Kodeira et al., 1986). Besides, polymorphism studies of human V_H pseudogenes report high sequence conservation among unrelated individuals (Pascual and Capra, 1991). These facts, suggest a possible functional role for pseudogenes in the human immune response diversity (Wysocki and Gefter, 1989; Pascual and Capra, 1991). In this application additional arguments in favor of this hypothesis are given by the combined use of Framework and CDRs items, within 3D patterns submenu (Vargas-Madrazo et al., 1994b).

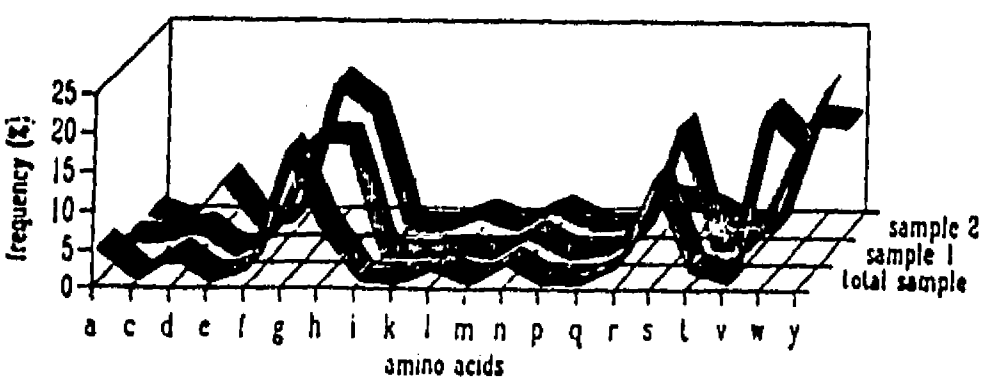


Fig. 2. Amino acid usage for position 91 of V_L. This position is selected as an example because it does not determine the canonical conformation (Chothia and Lesk, 1987) and contacts the antigen in 6 of the 8 antigen–antibody complexes (Padlan and Kabat, 1991). The number of sequences in each sample is: total sample: 1934 sequences; sample 1: 210 sequences; sample 2: 260 sequences.

Table 1
Structural divergence for V_H human germline genes and pseudogenes. We consider as defective or mutated, those positions not having amino acids codified in the Framework item. Defective substitutions on each sample were averaged dividing the total number of defective positions by the number of sequences in each sample

Pseudogenes	Germline genes
6.3	0.5

It has been proposed that the number of mutations accumulated in pseudogenes with respect to germline genes is principally a function of either: (i) the time elapsed since inactivation of the gene has occurred; or (ii) possible functional restrictions acting on the new state of the gene (Pascual and Capra, 1991). In the first case, random mutation distributions should be expected while in the latter a pattern — a canonical structure, that is to say — should emerge. Since human V_H germline genes show a large percentage of canonical structures within the first two CDRs (Chothia et al., 1992; Vargas-Madrazo et al., 1994b), then it might be possible that such patterns could also be present in these sequences.

In order to test the above mentioned hypothesis comparisons of mutations distributed in framework and CDR positions were carried out. So, the framework option was used on human V_H germline genes and pseudogenes to determine the degree of structural divergence (non-conservative amino acid substitutions) accumulated at framework positions (Table 1). As can be seen, mutations that may alter the framework stability of the human V_H pseudogenes are about 12 times as divergent as those of human V_H germline genes. A high divergence degree in pseudogenes, with respect to the germline genes, suggests that the production of a structurally stable variable domain is not the function of selection for the V_H pseudogenes. Therefore, if producing a structurally stable domain is not the function of selection, then pseudogenes could be a source of additional diversity for the CDRs or, otherwise, be irrelevant for the immune response.

Table 2
Sequences with canonical structures patterns in human V_H germline genes and pseudogenes

CDRs	Pseudogenes (%)	Germlines genes (%)
H1	74	100
H2	43	96

Contrasting with the fact that pseudogenes have an average of six destructive mutations per sequences in frame work positions (see Table 1), it was found using the CDRs item that 3/4 of the sequences have canonical structures for H1, while half of sequences do so for H2 (Table 2). The case of H2 having less sequences with canonical structures might be closely related to the fact that this CDR is, in general, less conserved (Chothia et al., 1992). Taken together, these results suggest that functional restrictions are directly responsible of canonical structure conservation (Vargas-Madrazo et al., 1994).

4. Discussion

In the last 13 years, since the first sequence compilation (Dayhoff and Eck, 1966), accumulation of quantitative information about nucleotide and amino acid sequences at an ever increasing rate makes it very difficult to assimilate and/or to analyze its possible meanings. Projects to map and sequence complex genomes including the human one are underway (Bell, 1990). Sequences analysis, in the molecular biology field can be divided into two main branches: (i) DNA or amino acid sequences collection in data bases; and (ii) tools development for sequence analysis research.

Regarding the first point, in the Ig case, the functional, technical, bibliographic data and sequences, have been collected from the Kabat's data base. In April 1994, the number of Ig sequences or fragments in this data base roughly amounted to 8000 (4700 for V_H and 3300 for V_L), classified under 460 different specificities. All this information made it possible for careful evolutionary and functional analysis in order to begin to understand the structural basis of the immune recognition mediated by Ig.

In parallel with this large amount of experimental data, in the last 20 years the Ig-fold has been the subject of many structural investigations, from which, antibodies are considered the best known structure of all proteins to date (Padlan, 1994). This structural knowledge permits starting from a more robust working hypothesis to study the relation between sequences and their three-dimensional structure. Thus, combining this kind of analysis with functional information (specificities and/or affinity constant), structure-function relationship studies can be improved in Ig to account for the second branch of the sequences analysis. The main limitation to achieve this, however, has been the absence of user-friendly 'immunologist-language-like' tools to manage and analyze Ig sequences. The software described here is a first attempt to carry out this demand.

In the Analysis module of the VIR package three main analytical tools are given: amino acids frequency; variability analysis; and pattern recognition analysis. Calculation of amino acid frequencies allows analysis of amino acid propensities in positions of interest in Ig sequences. Application (a) shows the combined use of functional information from Ig sequences and this kind of analysis. Such analysis allows to study the general and particular properties of CDRs throughout the construction of samples considering the functional information relevant to the process (i.e specificity).

In relation to the variability indexes, variability analysis by Kabat index, at the time of its development, played a major role in the structural analysis of Ig, successfully predicting the regions responsible for the antigen specificity, before X-ray data could be obtained (Wu and Kabat, 1970). Other two variability indexes have been developed; the modified Kabat index (Jores et al., 1990), and the informational-theoretical entropy index (Shenkin et al., 1991). Simultaneous use of the Kabat index and the modified Kabat index has been employed as an analytical tool to investigate the residues responsible for the different detailed specificities of various lectins (Young and Oomen, 1992). Recently, variability analysis was extended to simultaneously use the three currently known variability indexes to understand the interaction between the TCR and peptide-MHC complex (Almagro et al., 1994). This shows the potential uses of this analysis in other protein families or, in our case, in the detailed evolutionary or functional analysis of Ig.

Finally, the last implementation, pattern recognition, deals with the proposition of supplying an analytical tool to perform studies of the relation between sequences and their corresponding three-dimensional structures. Application (b) describes a structural analysis of Ig sequences with an unknown functional role. Thus, it is expected that this analysis will allow a more rational view of the molecular recognition process mediated by antibodies (Chothia et al., 1992).

In summary, the software described is an attempt to provide a package of tools with a unified format that brings together the most relevant methods to analyze the Ig variable domains. Of course, this proposition can be applied to other protein families. For the immune system it is currently being planned to extend it to TCR and MHC molecules.

4.1. Availability of the program

The executable programs and data bases can be obtained by sending an e-mail message to the following address: salazar@redvax 1.dgsca.unam.mx.

Acknowledgements

The authors wish to thank Hector Ceceña for very helpful comments and carefully reading the manuscript. This work was partially supported by DGAPA, grant No. IN-206093. EVM was supported by a grant of CONACyT.

References

Almagro, J.C., Zenteno-Cuevas, R., Vargas-Madrazo, E. and Lara-Ochoa, F., 1994, Variability analysis of the T-cell receptors using three variability indexes. (Submitted).

Bell, G.I, 1990, The human genome: An introduction, in: Computers and DNA, Bell, G.I. and Marr T.G. (eds.) Santa Fe Institute studies in the Sciences of Complexity, Vol. 7 (Addison Wesley) pp. 3–11.

Chothia, C. and Lesk, A.M., 1987, Canonical structures for the hypervariable regions of immunoglobulin. J. Mol. Biol. 196, 901–917.

Chothia, C., Boswell, D.R. and Lesk, A.M., 1988, The outline structure of the T cell α:β receptor. EMBO J. 7, 3745–3755.

Chothia, C., Lesk, A.M., Tramontano, A., Levitt, M., Smith-Gill, S. J., Air, G., Sheriff, S., Padlan, E.A., Davies, D., Tulip, W.R., Colman, P.M., Spinelli, S., Alzari, P.M. and Poljak, R.J., 1989, Conformations of immunoglobulin hypervariable regions. Nature (Lond.) 342, 877–883.

Chothia, C., Lesk, A.M., Gherardi, E., Tomlinson, I.M., Walter, G., Marks, J.D., Llewelyn, M.B. and Winter, G., 1992. Structural repertoire of the human VH segments. J. Mol. Biol. 227, 799–817.

Dayhoff, M.O. and Eck, R.V., 1966, Atlas of Protein Sequence and Structure, Vol.2 (NBRF Press, Silver Spring).

Kodeira, M., Kinashi, T., Umemura, I., Matsuda, F., Noma, T., Ono, Y. and Honjo, T., 1986, Organization and evolution of the variable region genes of the human immunoglobulin heavy chain. J. Mol. Biol. 190, 529–541.

Jores, R., Alzari, P.M. and Meo, T., 1990, Resolution of hypervariable regions in T-cell receptor β chains by a modified Wu-Kabat index of amino acids diversity. Proc. Natl. Acad. Sci. USA. 87, 9138–9142.

Kabat, E.A., Wu, T.T., Perry, H.M., Gottesman, K.S. and Foeller, C., 1991, Sequences of Proteins of Immunological Interest. 5th edn. (U.S. Department of Health and Human Services, Public Health Service, National Institutes of Health, NIH, Bethesda, MD) Publication No. 91–3242.

Lara-Ochoa, F., Vargas-Madrazo, E., Jiménez-Montaño, M.A. and Almagro, J.C., 1994, Patterns in the complementarity determining regions of immunoglobulin. BioSystem 32, 1–9.

Mian, I.S., Bradwell, A.R. and Olson, A.J., 1991, Structure, function and properties of antibody binding sites. J. Mol. Biol. 217, 133–151.

Padlan, E.A. and Kabat, E.A., 1991, Modeling of antibody combining sites. Methods Enzymol. 202-B, 3–21.

Padlan E.A., 1994, The antibody molecules. Mol. Immunol. (in press).

Pascual, V. and Capra, D., 1991, Human immunoglobulin heavy-chain variable region genes: Organization, polymorphism, and expression. Adv. Immunol. 49, 1–74.

Shenkin, P.S., Erman, B. and Mastrandrea, L.D., 1991, Information-theoretical entropy as a measure of sequence variability. Proteins 11, 297–313.

Tonegawa, S., 1983, Somatic generation of antibody diversity. Nature (Lond.) 302, 575–581.

Vargas-Madrazo, E., Ochoa-Lara, F. and Jiménez-Montaño, M., 1994a, A skew distribution of amino acids at recognition sites of the hypervariable regions of immunoglobulin. J. Mol. Evol. 38, 100–104.

Vargas-Madrazo, E., Almagro, J.C. and Lara-Ochoa, F., 1994b, Structural repertoire in VH pseudogenes of immunoglobulin: Comparison with human germline genes and human functional sequences (submitted).

Wu, T.T. and Kabat, E.A., 1970, An analysis of the sequences of the variable regions of Bence Jones proteins and myeloma light chains and their implications for antibody complementarity. J. Exp. Med. 132, 211–250.

Wysocki, L.J. and Gefter, M.L., 1989, Gene conversion and the generation of antibody diversity. Annu. Rev. Biochem. 58, 509–531.

Young, M.N. and Oomen, R.P., 1992, Analysis of sequence variation among legume lectins A ring of hypervariable residues forms the perimeter of the carbohydrate binding site. J. Mol. Biol. 228, 924–934.

Zuckerkandl, E., 1976, Evolutionary processes and evolutionary noise at molecular level. J. Mol. Evol. 7, 167–183.

Capítulo III:

Variability analysis of the T-cell receptors using three variability indexes

JOBNAME: PEPTID PAGE: 1 SESS: 4 OUTPUT: Wed Oct 26 10:33:51 1994
/gamma/dric/munksgaard/310/005/545 - 2 -

Int. J. Peptide Protein Res. 0, 1994, 1–7
Printed in Belgium – all rights reserved

INTERNATIONAL JOURNAL OF
PEPTIDE & PROTEIN RESEARCH
ISSN 0367-8377

Variability analysis of the T-cell receptors using three variability indexes

JUAN C. ALMAGRO[1], ROBERTO ZENTENO[1], ENRIQUE VARGAS-MADRAZO[2] and FRANCISCO LARA-OCHOA[1]

[1] *Institute of Chemistry, National University of Mexico, University City, Mexico and* [2] *Institute of Biological Investigations, Veracruzana University, Xalapa, Veracruz, Mexico*

Received 29 February, accepted for publication 13 August 1994

In absence of a three-dimensional structure for TCR molecules, several attempts to identify their hypervariable regions by variability methods have been made; this subject is still troublesome. In this paper three different variability indexes were used: (i) the Kabat index, which is the classical measure of sequence variability, (ii) the modified Kabat index, successfully used in the β-chain of T-cell receptors and (iii) an information-theoretical entropy concept, recently proposed as an improved measure of the variability. In order to identify the hypervariable regions in the TCR sequences, a Fourier filtering was applied on each variability profile.

Results show that the three variability indexes have distinct resolutions for different levels of variability. Thus, the simultaneous use of these indexes compensates for the deficiency of any one of them in estimating variability. Applying the Fourier filtering, it is found that the hypervariable regions here identified, roughly coincide with the defined CDR-2 and CDR-3 in TCR by analogy with Ig. However, no hypervariable in the CDR-1 of α- and β-chains was found. The study on the influence of sample size in variability analysis, indicates that results are independent of the sample size. Considering current structural models of TCR–peptide–MHC interaction, one can suggest that the low-variability characteristics of these regions is inherently related to the interaction with relatively conserved region on the α-helices of MHC.

Key words: Fourier filtering; hypervariable regions; information-theoretical entropy; Kabat variability index; modified Kabat variability index; TCR–peptide–MHC interaction

The TCR is a heterodimer of two polypeptide chains α and β (or and δ). The α- and β-chains are composed of an amino terminal V domain (V_α and V_β) and a carboxyl terminal C domain (C_α and C_β) (1). The V_α domain is encoded by the gene segment V_α and by the joining segment J_α, whereas the V_β domain is encoded by three gene segments: V_β, the diversity segment D, and J_β (1). The main functions of this heterodimer involve its specific interaction (via its V domains) with short and linear antigenic peptides bounded to MHC molecules (1, 2). Although rapid advances in the structural characterization of peptide–MHC complexes (3–9) have been made, the nature of the interaction between the TCR molecule and peptide–MHC complex is poorly understood (10). This is partly due to the lack of experimental three-dimensional structures of TCR and TCR–peptide–MHC complex (10).

In absence of experimental three-dimensional structures of TCR, it has been proposed that its most variable regions in the V domains must correspond to the peptide–MHC binding sites (11–15). However, the precise identification of the hypervariable regions in the V_α and V_β domains of TCR using the classical variability measure [Kabat variability index (16)] is still troublesome, except for the $V_\alpha J_\alpha$ and $V_\beta DJ_\beta$ junctions (17). Failure to identify the hypervariable regions in the V_α and V_β gene segments by the Kabat index is a consequence of the absence of dominant peaks of variability (17, 18). In order to solve this difficulty, a modification of the Kabat index has been proposed (18). Application of this modified Kabat index to 159 sequences of TCR V_β domain, followed by an objective criterion based on a Fourier series expansion, outlines hypervariability regions unambiguously (18). However, application of this method to 82 sequences of the V_α domain do not outline such regions properly (18).

Abbreviations: TCR, T-cell receptor; MHC, major histocompatibility complex; Ig, immunoglobulin; CDRs, complementarity determining regions; V, variable domain; C, constant domain.

Recently, another measure of amino acid variability based on an informational-theoretical entropy was proposed (19). This measurement method has more desirable mathematical properties than the Kabat index, as well as an intuitive and statistical meaning which agrees well with the notion of variability (19). However, this index has not been applied to TCR.

In this paper we apply the three abovementioned variability indexes on the most sequences from the V_α and V_β domains reported in the literature (17). The hypervariable regions on each variability profile were identify by the Fourier filtering proposed by Jores *et al.* (18).

The hypervariable regions outlined with this methodology show good correspondence with the CDR-2 and CDR-3 proposed for the TCR by comparison with the V domain of Ig (20). However, the regions corresponding to the CDR-1 in both V_α and V_β domains were found to be relatively conserved. We suggest that the low variability in the CDR-1 of TCR is due to the interaction between this region and relatively conserved regions on the α-helices of MHC molecules.

MATERIAL AND METHODS

Data bases

Two computer-readable data bases, one for the sequences from V_α domain and the other for the sequences from the V_β domain, were built by digitalizing the multiple alignment reported by Kabat *et al.* (17). All sets of TCR sequences in both data bases consisted of 221 amino acid sequences for the V_α domain (123 complete sequences and 98 sequence fragments) on the one hand, and of 319 amino acid sequences for the V_β domain (141 complete sequences and 178 sequence fragments) on the other.

Variability indexes

Calculations of variability indexes were performed on each data base using the following definitions:

(i) Kabat index (V_K) (16):

$$V_K = k/p_1,$$

where k is the number of different amino acid types that appear at a given position and p_1 is the frequency of the most common amino acid at that position.

(ii) Modified Kabat index (V_J) (18):

$$V_J = j/np_1,$$

where j is the number of distinguishable amino acid pairs that appear at a given position and np_1 is the frequency of the most common amino acid pair at that position.

(iii) Information-theoretical entropy index (V_S) (19):

$$V_S = 2^S$$

where

$$S = -\sum_{i=1}^{m} p_i \log_2 p_i$$

and p_i is the frequency of amino acids at position i, and m is the number of different amino acids at position i.

Hence, six variability data sets were obtained, three from each data base.

Fourier filtering

In order to obtain comparable results from the three variability indexes, the original data sets were normalized by extracting the data set average from each data set. The hypervariable regions on each normalized data set were identified using the procedure proposed by Jores *et al.* (18). This procedure consists of the following: (i) the forward Fourier transform is applied to a variability data set (in our case to the normalized data); (ii) the first and last 10 coefficients (higher coefficients) are selected, and their inverse Fourier transform is obtained (procedure referred as Fourier filtering); (iii) the filtered data are divided by their root mean-square (r.m.s.) deviation. Those regions scoring values strictly greater than one are considered hypervariable (18).

Sample size dependence

Since the sample size might have an influence on the variability analysis (18), we looked for any possible dependence between the number of sequences used in the analysis and the shape of each Fourier filtered variability profile. Then, three samples with increasing number of sequences (20, 120 and 220 sequences) were randomly taken from the α- or β-data base, and each sample was processed by the above described Fourier filtering method.

RESULTS

Variability plots

The variability plots for the V_α and V_β domains of TCR using the three indexes are shown in Fig. 1. Along the V gene segments, the three indexes measure different degrees of variability. As expected (1, 17), one dominant peak of variability in the $V_\alpha J_\alpha$ and $V_\beta DJ_\beta$ junction, is clearly identified with the three of them.

Unlike the case of Ig (16), use of the Kabat index, V_K, along the V gene yields an ambiguous distribution of variability, making exact identification of hypervariable regions difficult to accomplish (Fig. 1). The modified Kabat index V_J shows a more heterogeneous variability in the V gene, although an exact delimitation of the boundaries of the hypervariable regions is not possible. Finally, the distribution of variabilities obtained by the information-theoretical entropy index V_S gives indistinguishable peaks of variability.

In order to explain the above indexes behavior, the values of V_K versus V_S and V_K versus V_J are plotted for each position of the V_α and V_β domains (Fig. 2A and

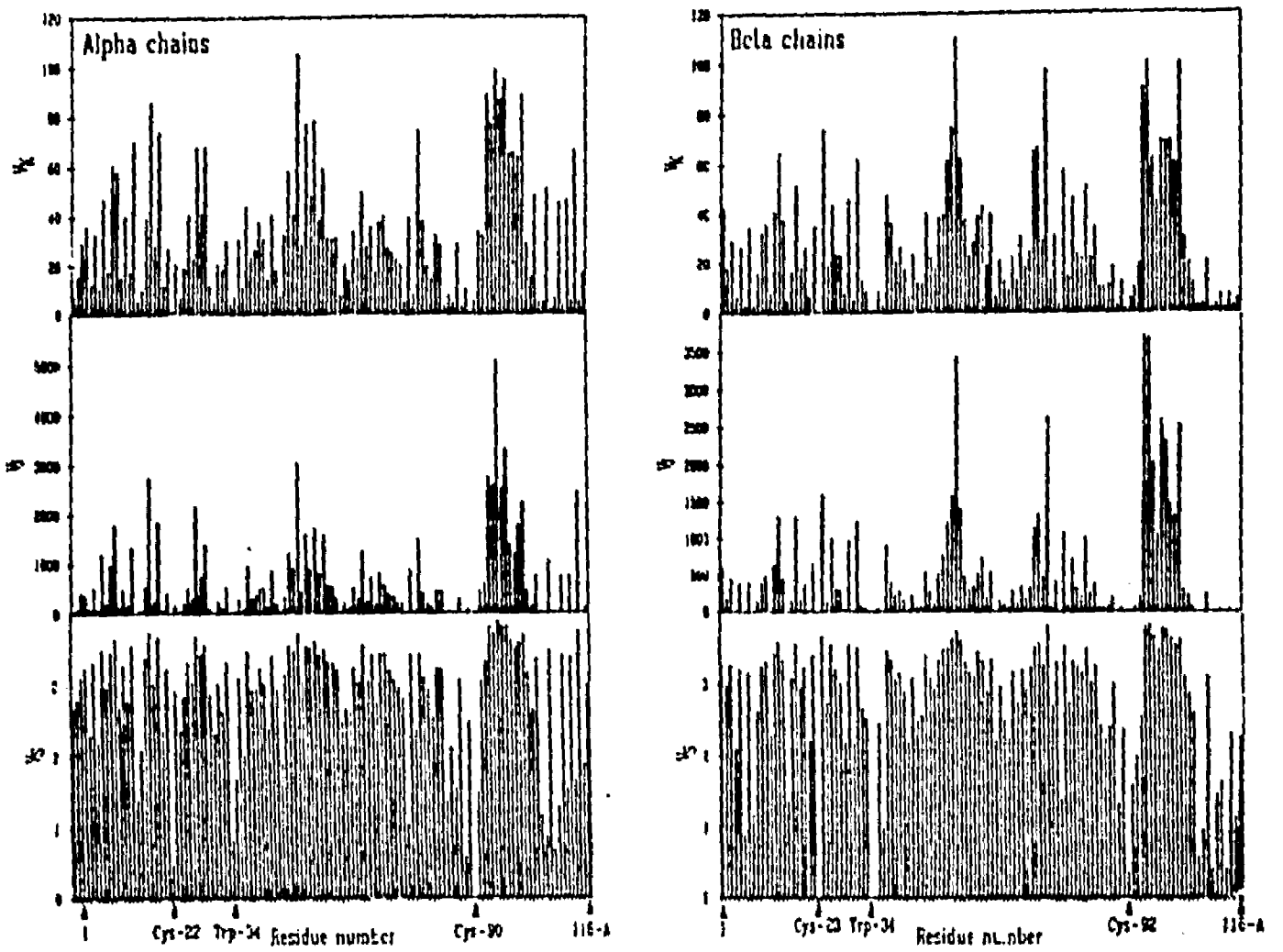


FIGURE 1

Variability plots for V_α and V_β domains using V_K, V_J and V_S. Arrowheads indicate the positions of invariant residues according to Kabat's numbering (17). Variability plot for V_α domain include the insertions 30A, B and 116A and for V_β domain include the insertions 30A, 62A and 116A.

2B). V_K was used as reference because it is the most applied variability measure in current literature.

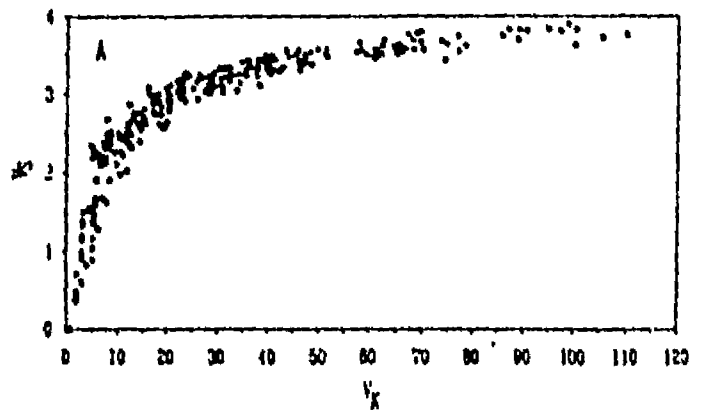
Correlation between V_K and V_S (Fig. 2A) shows a hyperbolic relationship. In the range from $V_S = 0$ to $V_S = 2.5$ ($V_K < 10$) the slope of the curve is very large. The sensitivity decreases from 2.5 up to values of 3.5 (from 10 to 30 in V_K), where asymptotical behavior is observed. This relationship corresponds to a distorted variability plot of V_S (Fig. 1), where the values of low variability are overestimated while medium and high variability values are underestimated with respect to V_K.

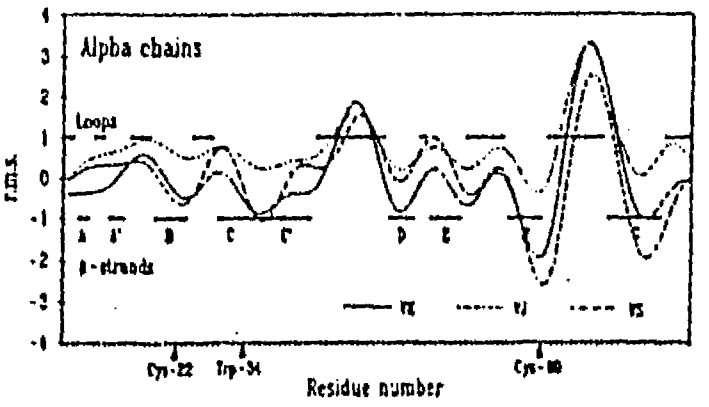
In the range from $V_J = 500$ to $V_J = 6000$ ($V_K > 30$) correlation of V_K with respect to V_J shows (Fig. 2B) a high dispersion of data points, as previously observed by Jores *et al.* (18). In the low variability range ($V_J < 500$ and $V_K < 30$) a smaller dispersion is found; however, the sensitivity (slope of the curve) of V_J with respect to V_K decreases. As a result, in contrast with the range of medium and high values, differences among low variability values ($V_J < 500$) for the variability plots of V_J (Fig. 1) are practically indistinguishable.

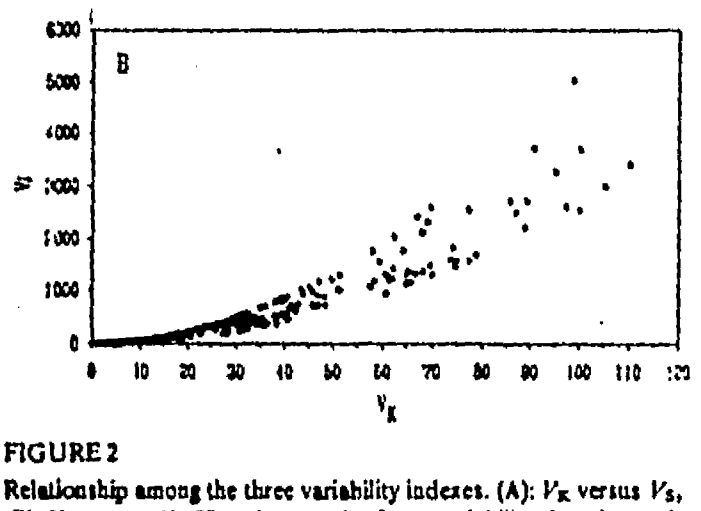
The above comparison shows that V_K, V_J and V_S estimate variability differently. While V_J operates well for high and medium variability values, V_S increases the resolution for low variability values. This suggests that the simultaneous use of the three indexes compensates for the deficiency of any one of them in estimating variability.

Fourier filtering

Fourier filtering was applied to objectively define hypervariable regions in the variability plots (Fig. 3). The two hypervariable regions identified as peak of variability directly on the variability profile in the $V_\alpha J_\alpha$ and $V_\beta DJ_\beta$ junction, were better distinguished after the filtering procedure was applied. This hypervariable region correspond to the loop linking of β-strands F-G in the Fourier smoothed plots of variability. Another hypervariable region was found in both V_α and V_β domains







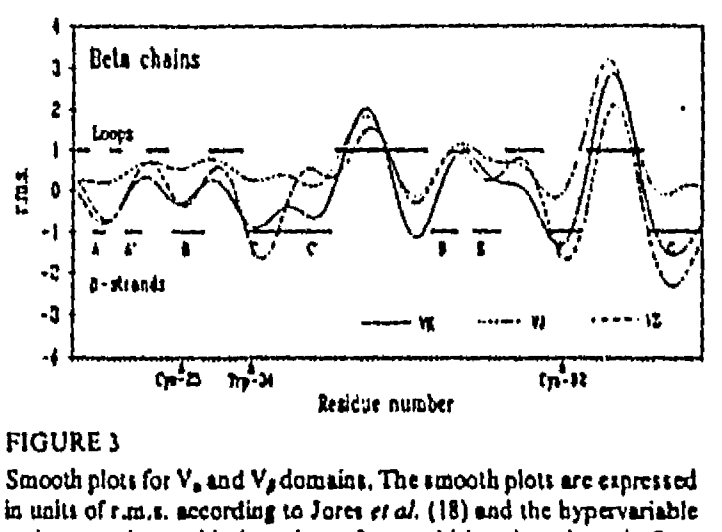


FIGURE 2

Relationship among the three variability indexes. (A): V_K versus V_S, (B): V_K versus V_J. The plots results from variability data for each position of V_α and V_β domains (121 scatter points from V_α domain and 119 scatter points from V_β domain).

FIGURE 3

Smooth plots for V_α and V_β domains. The smooth plots are expressed in units of r.m.s. according to Jores *et al.* (18) and the hypervariable regions are those with the values of r.m.s. higher than the unit. Secondary structure (β-strands and loops) of V domains of TCR are those predicted by Chothia *et al.* (20). Arrowheads indicate the positions of invariant residues that determine the V domain fold according to Kabat's numbering (17).

using all variability indexes, which corresponds to the loop linking of β-strands C'-D.

In addition, when the modified Kabat index V_J was used, a hypervariable region was found only in the V_β domain with slightly, but strictly, greater than one r.m.s. value. This hypervariable region, corresponding to the loop linking β-strands D-E, was reported previously by Jores *et al.* (18), using this same index.

Sample size

Since variability analysis is a statistical procedure, the number of sequences analyzed may have an influence in the hypervariable regions identified (18). The study on the influence of sample size in variability analysis (Fig. 4 shows that the shape of the smoothed plots changes only when a small number of sequences (20) are used independently of the index applied. For values over 120 sequences (approximately half the number of sequences in the data bases) the shape of the already mentioned plots does not change with sample size. Therefore, the identified hypervariable regions are independent of sample size, at least for the set of sequences in the data bases.

Because data bases used in the present study do not only contain complete sequences (see data bases), it would seem reasonable to assume that for the sequence fragments (especially for the $V_\alpha J_\alpha$ and $V_\beta DJ_\beta$ junctions) results in the definition of hypervariable regions would be different. Following the assumption, two additional data bases containing only complete sequences were built and Fourier filtering was applied. The results are similar to those using the original data bases (data not shown). This additional test demonstrates that occurrence of fragments of sequences in data bases is irrelevant when identifying hypervariable regions.

DISCUSSION

In the absence of three-dimensional structures of TCR, another approach has been proposed to identify the peptide–MHC binding site of this molecule (20, 21). This approach, based on the homology of TCR sequences with Ig of known structures, suggest that the six CDRs of TCR must be similar to the six CDRs of Ig (20). Then, a comparison of the hypervariable regions found here for the TCR with the CDRs proposed

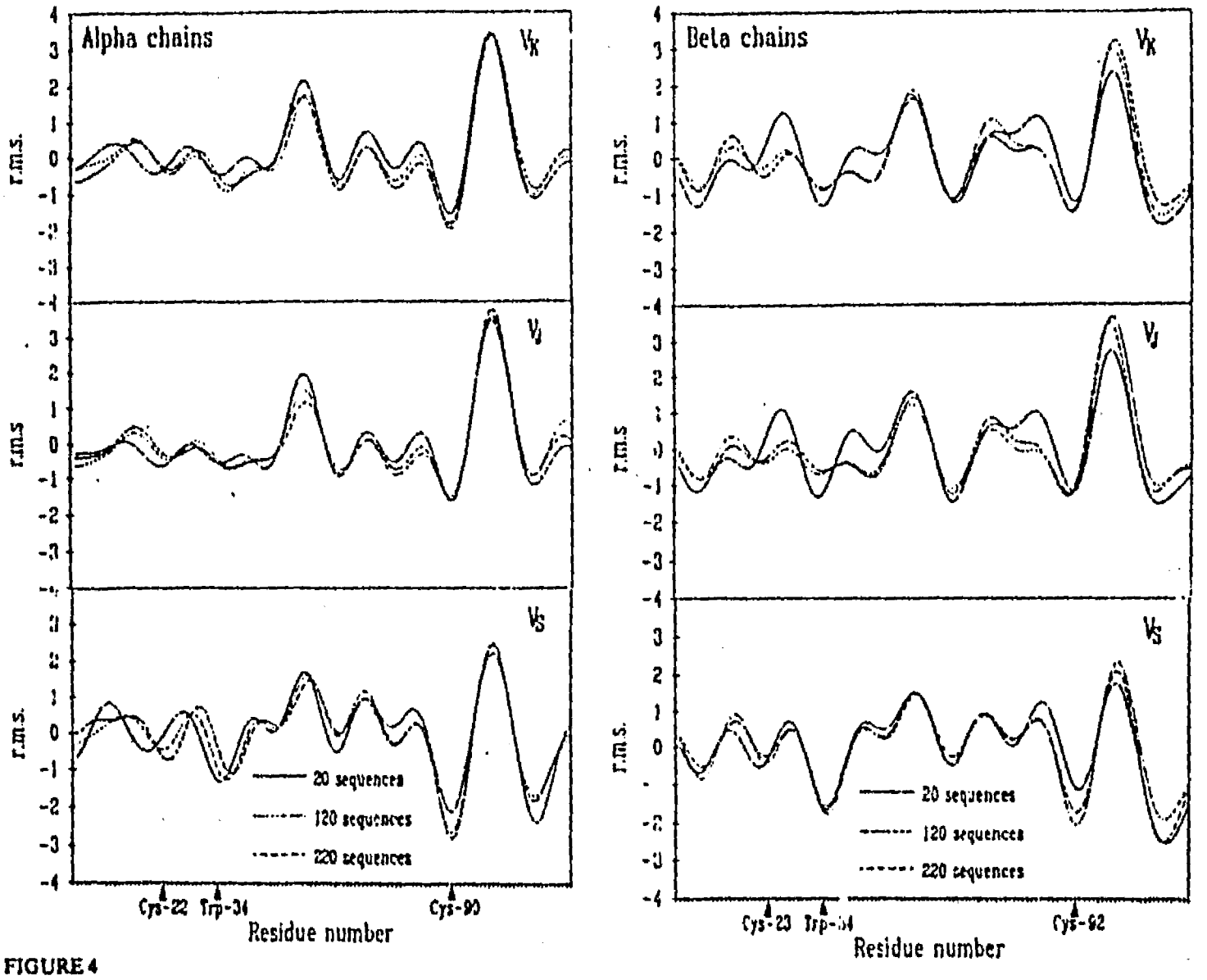


FIGURE 4

Sample size dependence on the variability indexes. The smooth plots are expressed in unit of r.m.s. deviations of the smoothing variability plot according to Jores *et al.* (18). Arrowheads indicate the positions of invariant residues determining folds in the V domain according to Kabat's numbering (17).

by homology with Ig (20) could be valuable to contrast our results.

The comparison shows good correspondence in the hypervariable region located as the loop linking of β-strands C'-D with CDR-2, and the hypervariable region corresponding to the loop linking of β-strands F-G with CDR-3 (Table I). Nevertheless, the CDR-1, placed at the loop linking β-strands B-C, was not identified as a hypervariable region by the variability analysis (Fig. 3). This result, however, is consistent with the observation that only limited sequence variations occur in CDR-1 in both V_α and V_β domains of TCR (20).

To explain the discrepancy found in the region corresponding to the CDR-1, two hypotheses can be made: (1) these regions do not interact with the peptide–MHC complex, or (2) these regions interact with low variability regions of peptide–MHC complex, and consequently do not require a high level of amino acid replacements. In principle, the first hypothesis should be discarded because recent experimental data suggest the involvement of at least the CDR-1 of V_α domain in the peptide–MHC complex recognition (22–24). To test the second hypothesis the conservation of amino acids in the regions of peptide–MHC complex which interact with the CDR-1 of TCR was analyzed by plotting their variability.

The TCR–peptide–MHC interaction models proposed in literature (2, 25, 26) were considered tentatively to localize interacting regions between TCR and peptide–MHC complex. According to these models, in the case of MHC class-I molecules, the CDR-1 and CDR-2 of V_α domain interact with helix-2, while CDR-1 and CDR-2 of V_β domain interact with helix-1 (2, 25, 26). The CDR-3 of both domains of TCR recognize the

TABLE I

Localitation of the hypervariable regions in the V_α and the V_β domains of TCR and their comparison with the CDRs proposed by analogy with Ig

Index	Hypervariable regions	
	V_α	V_β
V_K	49–55	51–58
	93–102	96–104
V_l	48–56	50–58
		70–73
	92–102	95–104
V_s	51–56	53–58
	95–101	98–103
Consensus[a]	?	?
	51–56	53–58
	95–101	98–103
CDRs[b]	25–30A	26–30A
	48–56	47–57
	93–104	93–104

[a] Consensus refer to the hypervariable regions simultaneously identified by the three indexes.

[b] CDRs defined by Chothia *et al.* (20) translate to the Kabat *et al.* (17) numbering to facilitate the comparison with the hypervariable regions.

antigenic peptide buried in the cleft shaped by the two α-helices of MHC (2, 25, 26). Recent studies have acknowledge the interaction of CDR-3 with the peptide suggesting that orientation of the TCR with respect to the MHC molecule is the predicted one (27). These hypothetical models present a general view of the interaction between TCR and peptide–MHC complex, however, the precise location of regions on the α-helices of MHC molecules which interact with CDR-1 and CDR-2 of TCR is not defined.

In order to obtain a precise location of regions of α-helices of MHC which interact with the CDR-1 and CDR-2 of TCR three suppositions were made: (1) The CDRs of TCR and Ig are geometrically equivalent. (2) The TCR-MHC interaction is similar in class-I and class-II MHC molecules. (3) The CDR-1 and CDR-2 of TCR should be interacting with the region limited by the first and last conserved residues of the α-helices contacting the peptides in class I MHC molecules [residues 59–84 at helix-1, and residues 143–171 at helix-2 (4–8)].

The first supposition agrees with Ganju *et al.* (23) and allows to delimit the interaction regions of CDR-1 and CDR-2 of TCR on the α-helices of MHC using Ig of known structure. The second supposition it is supported by the features of degeneracy that appear to have the TCR–MHC interaction (10). In fact, has been shown that TCR with identical V gene have cross-reactive idiotypes (28). This proposition is useful to interchange information between the class-I and class-II MHC molecules. The third supposition is consistent with mutagenesis studies showing that the area encompassed by these residues coincides with the possible interface between TCR and MHC molecules (29). This last supposition delimitates the boundaries of the interaction regions of CDRs of TCR on the α-helices of MHC molecules.

Thus, using an X-ray crystallographic structure of peptide–MHC class-I complex (4) because is best characterized than the peptide–MHC class-II complex, the regions of α-helices of MHC which interact with the CDR-1 and CDR-2 of the TCR can be outlined as follows. The CDR-1 of the V_β domain interacts with the second half of helix-1 (residues 71–84), and the CDR-2 interact with the first half (residues 59–72). The region in contact with CDR-1 of V_α domains in the helix-2 is from residue 157 to residue 171, while CDR-2 interacts with the region running from residue 143 up to residue 156.

The variability of the regions defined above were analyzed using the three variability indexes on the sequences of MHC class I compiled by Kabat *et al.* (18), 151 sequences from the helix-1 and 168 sequences from the helix-2. Average variabilities for the three indexes are given in Table 2. In spite of that the three indexes estimate the variability differently; the average variabilities for any one of them in the regions which interact with CDR-1 in both domains of TCR is less than the average variabilities in the interacting MHC-CDR-2 regions. These analysis suggest that CDR-1 of V_α and V_β domains are relatively low-variability regions because they interact with low-variability regions of MHC molecules and consequently do not require a high level of amino acid replacements.

TABLE 2

Average of the variabilities of α-helices of the MHC class I molecules

Index	Helix-1		Helix-2	
	Positions 59–70	Positions 71–84	Positions 143–156	Positions 157–171
V_K	16.7	13.4	14.6	7.2
V_l	132.7	57.9	35.4	21.4
V_s	1.5	1.4	1.3	0.7

The suggestion explains the discrepancy found between the present variability analysis and the CDR-1 proposed for the TCR by homology with the Ig (20). In addition, this proposition has interesting implications for the interaction between TCR molecule and peptide–MHC complex. That is, if the CDR-1 of V_α and V_β domains is relatively conserved, as has been shown by the present study, then it might be that these regions only contribute to the total energy in the ternary complex and do not have a critical role in the determination of the fine specificity of the TCR–MHC interaction.

In summary, we have shown that the simultaneous use of the three variability indexes, followed by an objective criterion, based on Fourier filtering, delineate unambiguously hypervariable regions in the V_α and V_β domains of TCR. The agreement between the hypervariable regions identified by this methodology, and CDR-2 and CDR-3 predicted by comparison of TCR sequences and Ig of known structure, has enable us to analyze in detail the discrepancy found in the CDR-1. The suggestion that the CDR-1 of V_α and V_β domains are relatively conserved because they interact with low-variability regions of MHC molecules add an interesting hypothesis to the interaction model between the TCR and peptide–MHC complex.

Finally, the variability analysis at the time of its development played a major role in the structural analysis of Ig, successfully predicting the regions responsible for the antigen specificity before X-ray data were obtained (17). Actually, other two variability indexes has been developed [V_J and V_S (18, 19)]. Recently, an example of the simultaneous use of V_K and V_J as analytical tools in the absence of X-ray data has been employed to characterize the binding site of legume lectins (30). In this paper we extend the variability analysis to the simultaneous use of the three variability indexes (V_K, V_J and V_S) in the absence of TCR X-ray structures.

ACKNOWLEDGMENTS

We thank Dr. E. Lamoyi, Dr. J. Lagunez-Otero and B.E. Lara for helpful comments, and H. Cecefia for revision of the final version of the manuscript. J.C.A. was supported by DGAPA grant IN-206093. E.V.M. was supported by CONACyt.

REFERENCES

1. Davis, M.M. (1990) *Annu. Rev. Biochem.* 59, 475–496
2. Bjorkman, P.J. & Parham, P. (1990) *Annu. Rev. Biochem.* 59, 253–288
3. Bjorkman, P.J., Saper, M.A., Samraoui, B., Bennett, W.S., Strominger, J.L. & Wiley, D.C. (1987) *Nature* 329, 507–512
4. Fremont, D.H., Matsumura, M., Stura, E.A., Peterson, P.A. & Wilson, I.A. (1992) *Science* 257, 919–927
5. Matsumura, M., Fremont, D.H., Peterson, P.A. & Wilson, I.A. (1992) *Science* 257, 927–934
6. Latron, F., Pazmany, L., Morrison, J., Moots, R., Saper, M.A., McMichael, A. & Strominger, J.L. (1992) *Science* 257, 964–967
7. Guo, H.-C., Jardetzky, T.S., Garrett, T.P., Lane, W.S., Strominger, J.L. & Wiley, D.C. (1992) *Nature* 360, 364–366
8. Silver, M.L., Guo, H.-C., Strominger, J.L. & Wiley, D.C. (1992) *Nature* 360, 367–369
9. Brown, J.H., Jardetzky, T.S., Gorga, J.C., Stern, L.J., Urban, R.G., Strominger, J.L. & Wiley, D.C. (1993) *Nature* 364, 33–39
10. Jorgensen, J.L., Reay, P.A., Ehrich, E.W. & Davis, M.M. (1992) *Annu. Rev. Immunol.* 10, 835–873
11. Patten, P., Yokota, T., Rothbard, J., Chien, Y., Arai, K. & Davis, M.M. (1984) *Nature* 312, 40–46
12. Becker, D.M., Patten, P., Chien, Y., Yokota, T., Eshhar, Z., Gledlin, M., Gascoigne, N.R.J., Goodnow, C., Wolf, R., Arai, K. & Davis, M.M. (1985) *Nature* 317, 430–434
13. Yoshikai, Y., Kimura, N., Toyonaga, B. & Mak, T.M. (1986) *J. Exp. Med.* 164, 90–103
14. Leiden, J.M. & Strominger, J.L. (1986) *Proc. Natl. Acad. Sci. USA* 83, 4456–4460
15. Bougueleret, L. & Claverie, J.-M. (1987) *Immunogenetics* 26, 304–308
16. Wu, T.T. & Kabat, E.A. (1970) *J. Exp. Med.* 132, 211–250
17. Kabat, E.A., Wu, T.T., Perry, H.M., Gottesman, K.S. & Foeller, C. (1991) *Sequences of Proteins of Immunological Interest*. 5th edn., Bethesda, MD. US Department of Health and Human Services, Public Health Service, National Institutes of Health, NIH Publication No. 91-3242
18. Jores, R., Alzari, P.M. & Meo, T. (1990) *Proc. Natl. Acad. Sci. USA* 87, 9138–9142
19. Shenkin, P.S., Erman, B. & Mastrandrea, L.D. (1991) *Proteins* 11, 297–313
20. Chothia, C., Boswell, D.R. & Lesk, A.M. (1988) *EMBO J.* 7, 3745–3755
21. Novotny, J., Tonegawa, S., Saito, H., Kranz, D.M. & Eisen, H. (1986) *Proc. Natl. Acad. Sci. USA* 83, 742–746
22. Nalefski, E.A., Wong, J.G.P. & Rao, A. (1990) *J. Biol. Chem.* 265, 8842–8846
23. Ganju, R.K., Smiley, S.T., Bajorath, J., Novotny, J. & Reinherz, E.L. (1992) *Proc. Natl. Acad. Sci. USA* 89, 11552–11556
24. Kobori, J.A., Hood, L. & Shastri, N. (1992) *Proc. Natl. Acad. Sci. USA* 89, 2940–2944
25. Davis, M.M. & Bjorkman, P.J. (1988) *Nature* 334, 395–402
26. Claverie, J.-M., Prochnicka-Chalufour, A. & Bougueleret, L. (1989) *Immunol. Today* 10, 10–14
27. Jorgensen, J.L., Esser, U., Fazekas de St. Groth, B., Reay, P.A. & Davis, M.M. (1992) *Nature* 355, 224–230
28. Rupp, F., Brecher, J., Giellin, M.A., Mosman, T., Zinkernagel, R., Hengartner, H. & John, R.H. (1987) *Proc. Natl. Acad. Sci. USA* 84, 219–222
29. Ajitkumar, P., Geier, S.S., Kesari, K.V., Borriello, F., Nakagawa, M., Bluestone, J.A., Saper, M.A., Wiley, D.C. & Nathenson, S.G. (1988) *Cell* 54, 47–56
30. Young, M.N. & Oomen, R.P. (1992) *J. Mol. Biol.* 228, 924–934

Address:

Juan Carlos Almagro
Instituto de Química
Universidad Nacional Autónoma de México
Circuito Exterior, Ciudad Universitaria
C.P. 04510, México, D.F.
Tel: (525)616-25-76
Fax: (525)616-22-17
E-mail: salazar@redvax1.dgsca.unam.mx

PEPTID/94-3230/7

Capítulo IV:

Molecular modeling of a T-cell receptor bound to a mayor histocompatibility complex molecule: Implications for the T-cell recognition

Title page:

MOLECULAR MODELING OF A T-CELL RECEPTOR BOUND TO A MAYOR HISTOCOMPATIBILITY COMPLEX MOLECULE: IMPLICATIONS FOR THE T-CELL RECOGNITION

Juan C. Almagro[+], Enrique Vargas-Madrazo[•], Francisco Lara-Ochoa and Eduardo Horjales

Instituto de Química,
Universidad Nacional Autónoma de México,
Circuito Exterior, Ciudad Universitaria,
C.P. 04510, México, D.F.

[•]Instituto de Investigaciones Biológicas,
Universidad Veracruzana,
Xalapa, Veracruz, México.

[+] corresponding author

Short title: Molecular modeling of a TCR/peptide/MHC complex.

Total number of pages: 37 (including one table and six figures).

Abbreviations: Ig: immunoglobulin, TCR: T-cell receptor, MHC: major histocompatibility complex, MCC: moth cytochrome *c*, PDB: Protein data bank; V: variable domain; V_L: variable domain of light chain; V_H: variable domain of heavy chain; Vα: variable domain of α chain; Vβ: variable domain of β chain.

Key words: TCR/peptide/MHC interaction complex, Variability analysis

ABSTRACT

The main functions of the T-cell receptor (TCR) involve its specific interaction with short and linear antigenic peptides bounded to the major histocompatibility complex (MHC) molecules. In absence of a three-dimensional structure for the TCR and for the TCR/peptide/MHC complex, several attempt to characterize the structural components of the TCR/peptide/MHC interaction has been made; this subject is still troublesome.

In this paper a computer-based three-dimensional model for a TCR/peptide/MHC complex (5C.C7/moth cytochrome *c* peptide 93-103/I-E^k) was obtained. The complex surface shows a high complementarity between the 5C.C7 structure and the peptide/I-E^k molecule. The mapping of residues involved in the TCR/peptide/MHC interaction shows close agreement with mutational experiments (Jorgensen et al., *Annu. Rev. Immunol.* **10**:835,1992). Moreover, the results are consistent with a recent variability analysis of TCR sequences using three variability indexes (Almagro et al., *Int. J. Pep & Prot Res.*, 1995, in press). Accordingly, the three-dimensional model of 5C.C7/MCC peptide 93-103/I-E^k complex provides a framework to generate testable hypotheses about TCR recognition. Starting from this three-dimensional model the role played by each loop that form the peptide/MHC binding site of the TCR molecule is proposed.

INTRODUCTION

The specificity of the immune system is related to two protein families: Igs and TCRs. Igs directly recognize, via their V_L and V_H domains, the three-dimensional structure of native antigenic proteins in solution (Davies et al., 1990), whereas TCRs recognize, via their Vα and Vß domains, short antigenic peptides bound to the MHC molecules (Davis, 1990). At present, significant advances have been made in the structural characterization of antigenic recognition mediated by Igs (Davies et al., 1990; Wilson & Stanfield, 1993; Padlan, 1994). In contrast, the nature of the interaction between TCRs and peptide-MHC complexes is poorly understood (Jorgensen et al., 1992a). This is partly due to the lack of an experimental three-dimensional structure for TCRs and for TCR/peptide/MHC complexes.

In absence of three-dimensional structures for TCRs, sequence analysis of their V domains show that the Vα/Vß dimer has a framework structure which is very close to the Fv fragment of Igs (Novotny et al., 1986; Chothia et al., 1988). Further comparison has structurally assigned the six hypervariable loops that form the antigen binding site of Igs to the binding site of TCRs (Chothia et al., 1988), termed in TCRs: α1, α2 and α3 for the Vα, and ß1, ß2 and ß3 in the Vß.

Genetically, α1 and α2 are encoded by the Vα gene, whereas α3 is produced by the recombination of an additional gene segment: Jα (Davis, 1990). In a similar way ß1 and ß2 are produced by the Vß gene, while ß3 is a result of the recombination of two additional gene segments: D and Jß (Davis, 1990). Based on the fact that α3 and ß3 concentrate the genetic diversity of TCRs on the one hand, and the similitude

between TCRs and Igs on the other, several hypothetical models for the TCR/peptide/MHC interaction have been proposed (Davis & Bjorkman, 1988; Chothia et al., 1988; Claverie et al. 1989). These models predict that α1, α2, ß1 and ß2 recognize the α-helices of MHC molecules, while α3 and ß3 interact with the antigenic peptide (Davis & Bjorkman, 1988; Chothia et al., 1988; Claverie et al. 1989).

The above models, for their generality, have been very useful to think about the evolution of the immune receptors (Davis & Bjorkman, 1988; Claverie et al., 1989; Davis, 1990), to explain overall features of the TCR/peptide/MHC interaction (Jorgensen et al., 1992b), and to suggest experiments (Ajitkumar et al., 1988; Nalefski et al., 1990; Patten et al., 1993; Lone et al., 1994). However, in the current form they may not be sufficiently detailed to delimitate which segments of the helical regions of MHC molecules interact with the TCR (Prochnicka-Chalufour et al., 1991; Almagro et al., 1995a). Consequently, they do not assign a role for the α1, α2, ß1 and ß2 in the TCR-MHC interaction.

In Igs as well as in TCRs, it has been proposed that the most variable regions of their V domains, must be placed in those loops that structurally form the binding site (Kabat et al., 1991). However, the precise identification of the hypervariable regions within the Vα and Vß domains using the classical variability measure [Kabat variability index (Wu & Kabat, 1970)] have been troublesome in all the TCR loops, except for α3 and ß3 (Kabat et al., 1991). Failure to identify the hypervariable regions in α1, α2, β1 and β2 of TCRs by the Kabat index, is a consequence of the absence

of dominant peaks of variability in the amino acids sequences encoded by the Vα and Vβ gen segments (Kabat et al., 1991). In order to solve this difficulty three variability indexes followed by a Fourier filtering have been recently used (Almagro et al., 1995a). This methodology unambiguously delineates hypervariable regions at the α2 and β2 loops together with the known hypervariable region placed at the α3 and ß3. However the regions corresponding to α1 and ß1 have been found relatively conserved (Almagro et al., 1995a). These results suggest that α1 and ß1 could have a different role with respect to α2 and β2 in the interaction with the MHC molecules (Almagro et al., 1995a).

In the last few years increasing experimental evidence has been accumulated for the interaction between TCR and peptide/MHC complexes (Jorgensen, et al., 1992b; Nalefski et al., 1992; Lone et al., 1994, DiGiusto & Palmer, 1994). One of the best studied experimental systems is the T-cell response to different MCC peptides bound to the I-E^k molecules (Hedrick et al., 1988; Jorgensen et al., 1992b). In this system, correlation of TCR gene segment usage with variant of the C-terminal peptide from MCC, demonstrated that charge substitutions on different residues of the peptide elicit opposite charges in residues of α3 or ß3 of TCRs (Jorgensen et al., 1992b). This result implies a particular positioning on the peptide/MHC structure for the TCR. Therefore using this experimental information allows us to refine the TCR/peptide/MHC interaction models proposed at present in order to understand the role played by α1, α2, β1 and β2 in the MHC interaction.

Thus, in this paper a computer-based three-dimensional model for the 5C.C7

sequence of TCRs in complex with peptide 93-103 of MCC [MCC (93-103)] and the I-E^k molecule was obtained. In the light shed by this atomic model, the structural role played by each loops that form the binding site of the TCR molecule in the interaction with peptide/MHC complex is discussed.

RESULTS

The atomic model of the TCR molecule: The three-dimensional structure for the 5C.C7 sequence was obtained by starting from the assumption that the Vα/Vβ dimer of TCRs has the typical fold of an Fv fragment of Ig (Novotny et al., 1986; Chothia et al, 1988). In Igs, the Fv fragment is composed of a large conserved framework of β-sheets and six hypervariable loops denoted L1, L2 and L3 for V_L and H1, H2 and H3 for V_H (Wu & Kabat, 1970, Poljak et al., 1973) that conform the antigen binding site. Structural analysis of the binding site of Igs revealed that there are a small number of main chain conformations found for five of the six hypervariable loops, termed canonical structures (Chothia & Lesk, 1987). A canonical structure is determined (1) by the loop size and (2) by the presence of certain residues at key sites in both the hypervariable loop and the framework region (Chothia & Lesk, 1987; Chothia et al., 1989; Tramontano et al., 1990). Molecular modeling procedures based on these rules have been demonstrated to be capable of correctly predict the antigen binding sites of different Fv fragments in advance of their experimental determination (Chothia et al., 1989).

Therefore, the framework of 5C.C7 sequence was modeled by using as

reference the coordinates of the D1.3 Fv fragment (Bhat et al., 1990), one of the best resolved and best refined Fv structures (1.8 Å resolution and an R-value of 0.158). The alignment between the reference and 5C.C7 sequence was made taking into account the amino acid pattern at those 40 sites proposed as crucial to maintain the conserved features of the Fv framework (Chothia et al., 1988) (Figure 1).

Having modeled the framework of the 5C.C7 sequence, the possibility of assembling the antigen biding site based on coordinates from canonical structures was investigated. To achieve this, the size of canonical structures described for Igs was compared to the size of those loops that form the binding site of the 5C.C7 sequence. Three of the six loops, α1, ß1 and ß2 were found to have the proper size of canonical structures, and consequently were built using this modeling procedure. The rest of the loops (α2, α3 and ß3) do not have the appropriate size for a canonical structure as described at present. Thus, α3 and β3 were modeled starting from coordinates of Ig loops found in a data base of Ig structures by a loop search procedure. No loops were found in the case of α2 of the appropriate size in Igs data base. This loop was modeled as a deletion of two residues of L2 in the reference structure. In Figure 2 the resulting loops that comprise the binding site of 5C.C7 after molecular refinement of the complete structure (see Material and Methods section) are shown.

Among the loops modeled with canonical structures, the coordinates of L1 type 1 were assigned to α1. This loop has residues different from those that stabilize the canonical conformation: Asp instead of Ile at position 2, Phe instead of Ala at position 25 and Val instead of Leu at position 33, and Ser instead of Tyr at position

71. However, after the molecular refinement of the structure the results show that these residues are mutually compensated to fit the characteristic packing of the canonical structure L1 type 1 in α1 (see Figure 2). This finding agrees with Ganju et al. (1992) modeling the binding site of the TCR sequence RFL8.3 based on the canonical structure motif.

The other loop modeled based on a canonical structure was ß1. This loop is two residues shorter than the canonical structures described for H1 (Chothia & Lesk, 1987; Chothia et al., 1992). However, this loop is equal in size to α1. Given this observation, along with the fact that Vα and Vβ possess features that are similar to both Igs domains, V_L and V_H (Chothia et al., 1988), β1 was built with the canonical structure L1 type 1 also. However, at position 29 within the loop, ß1 has a large polar residue (Histidine), differing from those residues (small hydrophobic) described as key to maintain this canonical structure (Chothia & Lesk, 1987; Chothia et al., 1989). An analysis of the refined structure (see Figure 2) shows that, like in α1, mutations in residues of the framework can pack this canonical structure in Vß. As can be seen from Figure 2, His at position 29 is stabilized by a hydrogen bond with Ser at framework position 94 of Vß. A study of the Vß sequences has acknowledged this finding. In the compilation of Vß sequences made by Chothia et al. (1988) this loop size is the most abundant within the sequences (roughly 90%). Except for four, all sequences having this size, have histidine at position 29. Serine also is highly conserved at position 94 among the sequences with this loop size. The conservation of serine at position 94 brings forth a structural context that fits histidine as the anchor

residue of the canonical structure L1 type 1 in Vß.

The third 5C.C7 loop built with a canonical structure was ß2. This loop has the same size as the canonical structure H2 type 3 (Chothia & Lesk, 1987). This canonical structure has residues with positive values for Φ and Ψ and usually, but not in all cases, Gly, Asn, or Asp, at positions 54 or 55 (Chothia et al. 1989) (position 52 and 53 of D1.3 V_H domain in Figure 1). Residue 53 is a Glu in the 5C.C7 loop with Φ having a positive value and Ψ is equal to -25°. Asn at position 51 also presents a positive Φ. These values of the dihedral angles produce a slight distortion of the conformation, but still fill the requirements for such canonical structure.

For those loops with modeling not based on canonical structures (α3 and ß3), the r.m.s. difference between the Cα atoms of the x-ray coordinates used as reference structure and the final modeled structure was for α3 0.57 Å and 1.36 Å for ß3. The differences show that these loops can be accommodated in the 5C.C7 framework. This result is consistent with the assessment of the final 5C.C7 structure quality by the 3D-profile method (Lüthy et al., 1992). The structure has no regions scoring less than 0 (Figure 3), and the global score of each V domain of the TCR is comparable with global scores of x-ray structures of ~ 100 amino acids lengths (Lüthy et al., 1992).

Finally, in order to assure the model stability, the 5C.C7 structure was subjected to 30 ps of molecular dynamic simulation at 300 K in a 4 Å shell water. The r.m.s. deviations over the Cα atoms between the framework of the starting structure and an average structure of the last 20 ps of simulation was of 1.41 Å. This r.m.s.

deviation is comparable with the correspondent value obtained among known x-ray structures of Igs (Chothia & Lesk, 1987), as well as differences between an x-ray structure and an average structure of 40 ps of dynamical simulation for a Fv fragment in similar simulation conditions (Almagro et al., 1995b). In the loops that form the binding site, the r.m.s. differences from the Cα atoms with respect to 5C.C7 model structure are similar to the differences from the framework when compared the initial model and the average structure. Thus, together with the 3D-profile method this test indicates that the 5C.C7 model provides a good starting structure to simulate the interaction with the MCC(93-103)/I-E^k.

The atomic model of the TCR/peptide/MHC complex: The three-dimensional model of the MCC peptide 93-103 in complex with the I-E^k molecule was obtained by taking advantage of the recently resolved x-ray crystallographic coordinates of HLA-DR1 molecule in complex with influenza hemagglutinin peptide 306-318 at 2.75 Å (Stern et al., 1994). In Figure 4 the sequence alignment of the α1 and ß1 domains of DR1 and I-E^k is shown.

The alignment has no insertions or deletions. In addition, the sequences have identities of 78% for α1 domain and 67% for ß1 domain. Consequently, the direct mutations in side chains would produce a good starting model, further refined by molecular dynamics. Starting from this structure and the 5C.C7 structure described in the above section, the atomic model for the 5C.C7/MCC(93-103)/I-E^k complex was obtained.

The complex was built by manual orientation of the 5C.C7 molecule upon the

MCC(93-103)/I-E^k complex, until the distance among the residues identified being in contact by mutational studies was optimal [Glu-97 of α3 and Asn-100 of ß3 with Lys-99 and Thr-102 in MCC respectively (Jorgensen et al., 1992b)]. Furthermore, the total energy upon the complex formation in each orientation was evaluated by calculating the van der Waals and electrostatic intermolecular interactions. The solvent accessible surface of the loops that form the binding site of TCR and the peptide/α-helices of MHC was visually analyzed during graphic manual changes. This procedure gives an additional criterion concerning surface complementarity.

Within the minimal interaction energy and the best complementarity between 5C.C7 and MCC(93-103)/I-E^k, the complex was subjected to molecular dynamic refinement in a water shell. No major conformational changes were observed between the single molecules and the molecules of the complex, which indicate that the complex refinement only introduced local optimization of the side chain interactions. The resulting structure is show in Figure 5 and kinemage 1.

Analysis of the complex interface shows that it is extended over a large region with maximum dimensions of 25 Å x 20 Å, similar to those predicted by mutagenesis studies (Ajitkumar et al., 1988). The solvent-accessible surface buried in complex formation was estimated in ~ 900 $Å^2$ for MCC/I-E^k and ~ 900 $Å^2$ for 5C.C7. This surface is higher in size than the lysozyme-antibody interfaces of 690-746 $Å^2$ (Davies et al., 1990). However, this is very similar to the solvent-accessible surface involved in the neuraminidase and the NC41 antibody complex of 886 A^2 (Davies et al., 1990). The similitude with the neuraminidase-antibody complex can be related to a larger

curvature radius of the interaction region of the MHC and neuraminidase molecules when compared to the corresponding regions of the lysozyme molecule.

Topologically the 5C.C7 interface is very irregular; it has a deep depression on one side of the surface and an essentially plane region at the other side (Figure 6a). The depression builds by α1, α2 and β3 (Figure 6b) is highly complementarity with a prominent kink at the helical region of the ß-1 domain of I-E^k. The loop ß2 of 5C.C7 on the other side of the interface is also highly complementary to the MHC molecule running along the helical region of the α1 domain of I-E^k (Figure 6b). It should be noted that this striking complementarity makes it difficult for a different positioning of the TCR on the peptide-MHC complex.

The estimate of the peptide/MHC residues in contact with the 5C.C7 molecule shows that seven residues of the MCC peptide and nineteen residues of the I-E^k molecule are involved in the interaction. Twenty-six residues of the 5C.C7 interact with the MCC(93-103)/I-E^k complex (Table 1). Twenty-four residues are located in five of six loops that form the binding site among them, and two residues are placed in the loop linking the ß-strands D-E of V_α. β1 loop of 5C.C7 does not interact directly either with the MHC molecule or with the peptide (see Table 1). In α1 and α2 only residues contacting the MHC molecule were found. In contrast, the rest of the loops (ß2, α3 and ß3) interact with both the MCC peptide and the MHC molecule.

As predicted by the general TCR/peptide/MHC models (Davis & Bjorkman, 1988; Chothia et al., 1988; Claverie et al. 1989), and experimental results (Jorgensen

et al., 1992b), α3 and ß3 contact the MCC peptide. However, it is interesting to note that residues of α3 and ß3 also contact the helices of I-E^k molecule (see Table 1). These results, taken together, allow us to propose in structural terms, the way the 5C.C7 molecule recognizes the MCC(93-103)/I-E^k complex.

DISCUSSION

Analysis of the TCR/peptide/MHC interface: The mapping of the 5C.C7 residues which interacts with the MCC (93-103)/I-E^k structure shows that ß1 does not contact the peptide/MHC complex. In addition the location of ß1 in the structure suggests that any contact of this loop with the MHC molecule (or the peptide) should imply disruption of some contacts among the residues used to orient the TCR molecule upon the peptide/MHC complex (see kinemage 1). This result suggests that β1, at least in this particular complex, do not have a direct role in MHC recognition.

The residues that interact with the peptide/MHC complex show that α1 and α2 loops are implied only in the recognition of the MHC molecule (see Table 1), while β2, α3 and β3, contact both the MCC peptide and the I-E^k molecule. In T-cell responses to the MCC peptide presented by I-E^k molecules it has been shown that the TCRs use the same V_α gen segment [Vα 11.1 (Hedrick et al., 1988; Jorgensen et al., 1992b)], whereas the use of V_β does not so. In V_β, changes in a single residue in the MCC peptide induce changes in very different patterns of V_β usage (Jorgensen, et al., 1992b). If it is considered that α1 and α2 are encoded by the Vα gene segment and β2 is encoded by the Vβ gen segment (Davis, 1990), the fact that α1

and α2 contact only the MHC molecule explains why the TCRs use the same Vα. In addition, the usage of different Vβ genes is explained by the fact that, in contrast to the Vα gen segment, several residues from ß2 contact the α-helix of α1 domain of I-E^k, as well as several residues of the MCC peptide. Thus, mutation of the peptide would imply changes in the pattern Vβ gen usage in order to accommodate mutations in the peptide.

Although ß2 contacts both the MCC peptide and the I-E^k molecule, there are some residues (from residue 52 to residue 55, and residue 57) that only interact with the peptide and do not contact the MHC molecule (see Table 1). The region of the ß2 that interacts with the I-E^k molecule is placed mainly at the ß-strand located from residue 53 to residue 58 (see kinemage 1). In this interacting region the positions 54 and 55 of 5C.C7 have two hydrophobic residues, Val and Leu, respectively. These residues form a hydrophobic patch flanked by hydrophilic residues. This hydrophobic patch is in contact with a hydrophobic stretch at the α-helix of the MHC molecule, it formed by Leu-64, Ala-65, Ala-68 and Val-69. In solution, these two hydrophobic regions would be exposed to the solvent, which is energetically unfavorable. After complex formation, the hydrophobic regions in both TCR and I-E^k are stabilized with one another. This observation suggests that the hydrophobic stretch in the α1 domain of I-E^k molecule is important for the recognition of the MHC molecule. Supporting this observation, mutations in the I-E^k sequence within this region avoid the TCR-MHC interaction (Jorgensen et al., 1992a).

Thus, the region placed at the β-strand of β2 an one side of the TCR

interface, and α1 and α2 on the other side, would be responsible for the recognition of the MHC molecules. In α3 and β3, several residues interact with the peptide and also contact the helical regions of the I-E^k molecule. Therefore α3 and β3, together with the residues β2 which interact with the peptide, would determine the fine recognition of the antigenic peptides in combination with a certain particular MHC molecule.

Implications for the T-cell recognition: In absence of structural experimental information for TCRs, variability analysis of TCR sequences shows that α1 and ß1 are relatively conserved (Almagro et al., 1995a). This result suggests that α1 and ß1 would have a different role with respect to α2 and β2 in the interaction of the MHC molecules (Almagro et al., 1995a). The general TCR/peptide/MHC interaction models (Davis & Bjorkman, 1988; Chothia et al., 1988; Claverie et al. 1989), based on the similitude between TCRs and Igs, have not been sufficiently detailed to study the role played by α1, α2, ß1 and ß2 in the TCR-MHC interaction (Prochnicka-Chalufour et al., 1991).

In the preceding section, several features for the interaction of α1, α2, β1 and β2 were highlighted for the particular case of the 5C.C7/MCC (93-103)/I-E^k complex. The results are consistent with the experimental data available for this complex and explain general characteristic of T-cell responses to MCC peptides presented by the I-E^k molecule. Thus, this TCR/peptide/MHC interaction model can be useful in providing a structural framework for generating testable hypotheses about the contribution of α1, α2, ß1 and ß2 to the TCR-MHC interaction.

Overall the experimental structure for functional domains of MHC class II, α1 and ß1, are similar to that of class I (Brown et al., 1994). Additionally, comparison of the peptides bonded to the MHC class I and class II shows that the peptide conformation in both classes of MHCs is extended. In fact, a comparison of Cα atoms between peptides bound to class I molecules (PDB entry; 1HHK [Madden et al., 1993]) and class II (HLA-DR1 molecule [Stern et al., 1994]) at region which interact with the 5C.C7 in our model (residues 97-103, see Table 1) gives r.m.s. differences of 1.0 Å. The similitude between class I and II MHC molecules, and the likeness between Cα conformations of the antigenic peptides, indicate that the surface of peptide/MHC molecules presented to the TCRs is similar. If it is considered that the TCR molecules bind the peptide/MHC class I complexes in the same orientation as peptide/class II complexes, as been suggested by the interaction between superantigen and MHC class I molecules (Bellio et al., 1994) and superantigen class II molecules (Jardetzki et al., 1994), then several general suggestions can be made for the TCR/peptide/MHC interaction.

Analysis of 5C.C7/MCC(93-103)/I-E^k structure showed that β1 loop of 5C.C7 does not contact the MCC(93-103)/I-E^k complex. Consistent with our model, experimental data indicates that substitutions in the residues within this region of TCRs do not affect the response to the allo-antigens (DiGiusto and Palmer, 1994). This result agrees also with the low variability found for this loop (Almagro et al., 1995a). In spite of that, experiments with single alanine substitutions on residues of β1 indicate that this region may be involved in the peptide-MHC interaction (Lone

et al., 1994; Kang et al., 1994; Bellio et al., 1994).

The discrepancy found among the experimental results at the β1 loop can be explained by: (1) In other TCR molecules, different of 5C.C7, loops shorter in size than β2 and α3 of 5C.C7 sequence (those loops that block the interaction of β1 with the MHC molecule, see kinemage 1) may allow to the β1 loop interact with the helical region of MHC; and (2) Analysis of the binding site of 5C.C7 shows that the anchor residue of this loop, His-29, is packed with Ser-94 placed at the α3 loop. This tight interaction among residues of β1 and α3 might to modify the proper positioning of the α3 loop to recognize both, the peptide or the MHC molecules. Consequently, amino acids substitutions in His-29 may abolish the T-cell response. These two suggestions are not mutually exclusive and will have different weight in each particular TCR/peptide/MHC system.

In α1 loop, only three residues interact with the I-E^k molecule (see Table 1). Experimental data shows that the α1 loop of TCR is implied in the recognition of other TCR/peptide/MHC complexes (Nalefski et al., 1992). An analysis of those residues which interacts with the MHC molecule in our model, shows that all these residues interact only with non polymorphic residues of MHC (Brown et al., 1992). This also is consistent with the variability analysis, and suggests that α1 is not a hypervariable region because this interact with residues relatively conserved of MHC molecules.

Analysis of the residues in α2 that interact with the MHC, indicate that its contacts are with the kink region of β1 domain of MHC. Comparison between class

I and class II molecules show structural differences at this region (Brown et al., 1993). The differences comes from a more protuberant kink in the helical region of β1 domain of class II molecules when compared with the correspondent region of MHC class I molecules (Brown et al., 1993). This region differing in structure also differ in its pattern of polymorphism, and even more, insertions and deletions are localized at this place within the MHC sequences (Brown et al., 1988). In a compilation of TCR sequences it was shown that five different sizes can be found in α2 loop of TCRs (Chothia et al., 1988). This fact contrasts with the conserved size of its equivalent loop in Igs, L2 (Chothia, et al., 1988). The variability in size of the α2 loop indicates that several conformations are possible for this loop. together with the structural differences of MHCs molecule at the kink region, this suggest that the interaction between α2 loop and the kink region of MHCs is critical for distinguishing among different classes of MHCs.

In β2 loop the residues which interact with the MHC molecule are placed in a structurally conserved region, because it is located on a ß-strand (see kinemage 1). The region in the MHC which interacts with β2 loop is also structurally conserved; it is placed in the middle of the α-helical region of α1 domain of MHC. Therefore one can suggest that, at this region only side-chain mutations that conserved the match of the TCR/MHC interaction are allowed. Additionally, the interaction TCR-MHC in this region can be modulated by the peptide residues bounded to the MHC molecule, as observed by the mapping of the residues at the 5C.C7/MCC(93-103)/I-E^k complex.

In conclusion, on the basis of the 5C.C7/MCC(93-103)/I-E^k model it is possible to suggest a structural role for α1, α2, β1 and β2 in the MHC interaction. These suggestions can be tested by mutational experiments in absence of structural experimental results for the TCR/peptide/MHC complex in order to understand the structural basis of the recognition mediated by TCR.

MATERIAL AND METHODS

Molecular modeling of 5C.C7: All modeling procedures were performed using the program INSIGHT-II (Biosym Technologies, San Diego) running on a Silicon Graphic 4D/35. The coordinates of the reported X-ray structure of the D1.3 Fv fragment (entry: 1FVA [Bath et al., 1990]) obtained from PDB (Bernstein et al. 1977) served as reference in the framework modeling procedure of the 5C.C7 Fv fragment. The hypervariable loops α1, ß1 and ß2 were modeled with canonical structures. For α1 and ß1 the coordinates of canonical structure L1 type 1 of the Fab J539 (PDB entry: 2FBJ [Bhat T.N., Padlan E.A, Davies D.R, in preparation]) were assigned. For ß2, the canonical structure type 3 of H2 of the reference structure (D1.3) was used. Two other hypervariable loops (α3 and β3) were modeled starting from the best choice of Igs loops found by using the loop search procedure of the INSIGHT-II package. The data base of Igs structures was built with structures collected from the PDB version of October-93. The x-ray structures found by the loop search procedure and then used as a template to built α3 and β3 were the coordinates corresponding to the residues L90-L97 of KOL (PDB entry: 2FB4

[Marquart et al., 1980]) in α3, and the coordinates corresponding to the residues J95-J101 of B13I2 (PDB entry: 1IGF [Stanfield et al., 1992]) for β3. The hypervariable loops α2 were modeled as a deletion of two amino acids in the atomic coordinates of the reference (D1.3 Fv fragment).

Molecular dynamic refinement: All molecular calculations were carried out using the Discover simulation software package (version 2.7) from Biosym Technologies running on a CRAY Y-MP 4/464 computer. The parameters used were those of the Consistent Valence Force Field (Hagler, 1985). No cross-terms were used in the energy expression and a simple harmonic potential was chosen for the bond stretching terms. Calculations were conducted with a group based non-bonded cut-off of 10.5 Å imposed over a switching distance of 1.5 Å and updating the non-bonded pair list every 20 time steps using the leap-frog algorithm with a 1 fs time step. All of the hydrogen atoms in the proteins were explicitly modeled. The refinement protocol consisted of successive steps of 5 ps of molecular dynamics at 300 K, as follows: 1) refinement of each modeled loop independently; 2) a global refinement of all loops together; 3) a global side chain refinement of the structures fixing the backbone atoms; 4) a global refinement of the structure tethering Cα atoms with a force constant of 60 Kcal/Å; 5) a global refinement of all structure. Before each step, the beginning structure was minimized for 250 steps using the steepest descent method followed by 250 steps using the conjugate gradient method.

Steps 1-3 were archived in vacuo with a constant dielectric of one and the protein having a net total neutral charge. Step 4-5 were performed with the residues

charged to simulate pH=7.0 and a 4 Å thick water shell surrounding the protein to stabilize the charges at the surface of the molecule. In this later case a linear distance-dependent dielectric was used to compensate for the water-vacuo interface. The distance-dependent model together with a water shell surrounding the protein has been shown to achieve a trajectory which better agrees with the crystal structure when compared with a constant dielectric model (Guenot & Kollman, 1992).

The 5C.C7 model refinement consisted of the above-mentioned five steps. The MCC(93-103)/I-E^k structure was only subjected to steps 3-5 because of the quality of the initial model. After obtaining a good orientation for the 5C.C7/MCC(93-103)/I-E^k complex, the structure was finally subjected to steps 3-5.

Acknowledgments: The authors wish to thank D. Eisenberg for 3D profile program, Don C. Wiley and coworkers for kindly providing atomic coordinates for peptide-MHC-II complex, M. Amzel and P. Alzari for valuable comments in early developments of this work, to the UNAM supercomputing center for the facilities in the use of the CRAY Y-MP 4/464 computer and to P. Reidy for technical assistance. JCA was supported by DGAPA, grant No. IN-206093. EVM was supported by FOMES, SNI and CONACyT.

REFERENCES

Ajitkumar P, Geier SS, Kesari KV, Borriello F, Nakagawa M, Bluestone JA, Saper MA, Wiley DC, Nathenson SG. 1988. Evidence that multiple residues on both the α-helices of the class I MHC molecule are simultaneously recognized by the T cell receptor. *Cell 54*:47-56.

Almagro JC, Zenteno-Cuevas R. Vargas-Madrazo E, Lara-Ochoa, F. 1995a. Variability analysis of the T-cell receptors using three variability indexes. *Int. J. Pep. & Prot. Res.* in press.

Almagro JC, Ceceña HA, Vargas-Madrazo E, Lara-Ochoa F. 1995b. Molecular dynamics simulation of an immunoglobulin Fv fragment In: *POP'95*. in press.

Bhat TN, Bentley GA, Fischmann TO, Boulot G, Poljak RJ. 1990. Small rearrangements in structures of Fv and Fab fragments of antibody D1.3 on antigen binding. *Nature 347*:483-485.

Bellio M, Lone Y-C, Calle-Martin O, Malissen B, Abastado J-P, Kourilsky P. 1994. The Vβ complementary determining region 1 of major histocompatibility complex (MHC) class I-restricted T cell receptor is involved in the recognition of peptide/MHC I and superantigen/MHCII complex. *J. Exp. Med. 179*:1087-1097.

Bernstein FC, Koetzle TF, Williams GJB, Meyer EF Jr, Brice MD, Rodgers JR, Kennard O, Shimandouchi T, Tasumi M. 1977. The Protein Data Bank. A computer-based archival file for macromolecular structures. *J. Mol. Biol. 112*:535-542.

Brown JH, Jardetzky T, Saper MA, Samraoui B, Bjorkman PJ, Wiley Don C. 1988. A hypothetical model of the foreign antigen binding site of class II histocompatibility molecules. *Nature 332*: 845-850.

Brown JH, Jardetzky TS, Gorga JC, Stern LJ, Urban RG, Strominger JL, Wiley DC. 1993. Three-dimensional structure of the human class II histocompatibility antigen HLA-DR1. *Nature 364*:33-39.

Chothia C, Lesk AM. 1987. Canonical structures for the hypervariable regions of immunoglobulin. *J. Mol. Biol. 196*:901-917.

Chothia C, Boswell DR, Lesk, AM. 1988. The outline structure of the T-cell α:ß receptor. *EMBO J. 7*:3745-3755.

Chothia C, Lesk AM, Tramontano A, Levitt M, Smith-Gill SJ, Air G, Sheriff S, Padlan EA, Davies D, Tulip WR, Colman PM, Spinelli S, Alzari PM, Poljak RJ. 1989. Conformations of immunoglobulin hypervariable regions. *Nature (London) 342*:877-883.

Chothia C, Lesk AM, Gherardi E, Tomlinson IM, Walter G, Marks JD, Llewelyn MB, Winter G. 1992. Structural repertoire of the human VH segments. *J. Mol. Biol. 227*:799-817.

Claverie J-M, Prochnicka-Chalufour A, Bougueleret L. 1989. Implications of a Fab-like structure for T-cell receptor. *Immunol. Today. 10*:10-14.

Connolly ML. 1983. Solvent-accessible surfaces of proteins and nucleic acids. *Science 221*:709-713.

Davies DR, Padlan EA, Sheriff S. 1990. Antibody-Antigen Complexes. *Annu. Rev. Biochem. 59*:439-473.

Davis MM, Bjorkman PJ. 1988. T-cell antigen receptor genes and T-cell recognition. *Nature 334*:395-402.

Davis MM. 1990. T cell receptor gene diversity and selection. *Annu. Rev. Biochem. 59*:475- 496.

DiGiusto DL, Palmer, E. 1994. An analysis of sequence variation in the β chain framework and complementarity determining regions of an allo-reactive T cell receptor. *Mol. Immunol. 31*:693-699.

Ganju RK, Smiley ST, Bajorath J, Novotny J, Reinherz EL. 1992. Similarity between fluorescein-specific T-cell receptor and antibody in chemical details of antigen recognition. *Proc. Natl. Acad. Sci. USA. 89*:11552-11556.

Genot J & Kollman PA. 1992. Molecular dynamics studies of a DNA-binding protein: 2. An evaluation of implicit and explicit solvent models for the molecular dynamics simulation of the *Escherichia coli trp* represor. *Protein Sci. 1*:1185-1205.

Hagler AT. 1985. Theoretical simulation of conformation, energetics, and dynamics of peptides. In: *Conformation in Biology & Drug Design, The peptides*. Meienhofer J. Ed. Academic Press, New York 7:213-228.

Hedrick SM, Engel I, McElligot DL, Fink, PJ, Hsu ML, Hansburg D, Matis A. 1988. Selection of amino acid sequences in the β chain of T cell antigen receptor. *Science 239:*1541-1544.

Jardetzky TS, Brown JH, Gorga JC, Stern LJ, Urban RG, Chi Y-i, Stauffacher C, Strominger JL, Wiley DC. 1994. Three-dimensional structure of a human class II histocompatibility molecule complexed with superantigen. *Nature 368*:711-718.

Jorgensen JL, Reay PA, Ehrich EW, Davis MM. 1992a. Molecular components of T-cell recognition. *Annu. Rev. Immunol. 10*:835-873.

Jorgensen JL, Esser U, Fazekas de St. Groth B, Reay PA, Davis MM. 1992b. Mapping T-cell receptor-peptide contacts by variant peptide immunization of single-chain transgenics. *Nature 355*:224-230.

Kabat EA, Wu TT, Perry HM, Gottesman KS, Foeller C. 1991. Sequences of Proteins of Immunological Interest. 5th ed. Bethesda, MD: U.S. Department of Health and Human Services, Public Health Service, National Institutes of Health, NIH Publication No. 91-3242.

Kabesh W, Sander C. 1983. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers. 22:*2577-2637.

Kang J, Chambers, C, Pawling J., Scott C, Hozumi N. 1994. Conserved amino acids residues in the complementarity determining region 1 of the TCR β-chain are involved in the recognition of conventional Ag and Mls-1 superantigen. *J. Mol. Immunol. 152*:5305-5317.

Kobori JA, Hood L, Shastri N. 1992. Structure-function relationship among T-cell receptors specific for lysozyme peptides bound to A^b or A^{bm12} molecules. *Proc. Natl. Acad. Sci. USA. 89*:2940-2944.

Lone Y-C, Bellio M, Prochnicka-Chalufour A, Ojcius DM, Boissel N, Ottenhoff TMH, Klausner RD, Abasto J-P, Kourilsky P. 1994. Role of the CDR1 region of the TCR β chain in the binding to purified MHC-peptide complex. *Int. Immunol. 6*:1561-1565.

Lüthy R, Bowie JV, Eisenberg D. 1992. Assessment of protein models with three-dimensional profiles. *Nature. 356*:83-85.

Madden DR, Garboczi DN, Wiley DC. 1993. The antigenic identity of peptide-MHC complexes: A comparison of conformation of five viral peptides presented by HLA-A2. *Cell. 75*:693-708.

Marquart M, Deisenhofer J, Huber R. 1980. Crystallographic refinement and atomic models of the intact immunoglobulin molecule Kol and its antigen-binding fragment at 3.0 Å and 1.9 Å resolution. *J. Mol. Biol. 141*:369-391.

Nalefski EA, Kasibhatla S, Rao A 1992.Funtional analysis of the binding site on the T cell receptor α chain. *J. Exp. Med. 175*:1553-1563.

Novotny J, Tonegawa S, Saito H, Kranz DM, Eisen H. 1986. Secondary, tertiary, and quaternary structure of T-cell-specific immunoglobulin-like polypeptide chains. *Proc. Natl. Acad. Sci. USA.* 83:742-746.

Padlan EA. 1994. The anatomy of the antibody molecule. *Mol. Immunol. 31*:169-193.

Patten PA, Rock EP, Sonoda T, Fazekas de St.Groth B, Jorgensen JL, Davis MM. 1993. Transfer of putative complementary-determining regions loops of T-cell receptors V domains confer toxin reactivity but not peptide/MHC specificity. *J. Immunol. 150*: 2281-2294.

Poljak RJ, Amzel LM, Avey H P, Chen B L, Phizacherley RP, Saul F. 1973. Three-dimensional structure of the Fab' fragment of a human immunoglobulin at 2.8-Å resolution. *Proc. Natl. Acad. Sci. U.S.A. 70*:3305-3310.

Prochnicka-Chalufour A, Casanova J-L, Avrameas S, Claverie J-M, Kourilsky P. Biased amino acids distribution in regions of the T cell receptors and MHC molecules potentially involved in their association. *Int. Immunol. 3*:853-864.

Stanfield RL, Fieser TM, Lerner RA, Wilson IA. 1990. Crystal structures of an antibody to a peptide and its complex with peptide antigen at 2.8 Å. *Science. 248*: 712-719.

Stern LJ, Brown JH, Jardetzky TS, Gorga JC, Urban RG, Strominger JL, Wiley DC. 1994. Crystal structure of the human class II MHC protein HLA-DR1 complexed with an antigenic peptide from influenza virus. *Nature*. in press

Tramontano, A., Chothia, C. & Lesk, A. M. (1990). Framework residue 71 is a major determinant of the position and conformation of the second hypervariable region in the V_H domains of immunoglobulins. *J. Mol. Biol. 215*:175-182.

Wilson IA, Stanfield RL. 1993. Antibody-antigen interactions. *Curr. Op. Struct. Biol. 3*: 113-118.

Wu TT, Kabat EA. 1970. An analysis of the sequences of the variable regions of Bence Jones proteins and myeloma light chains and their implications for antibody complementarity. *J. Exp. Med. 132*:211-250.

FIGURE AND TABLE CAPTIONS

Figure 1: Alignment between the reference (D1.3) and the TCR (5C.C7) sequence. Upper part of the figure: V_L/V_α domains, lower part: V_H/V_β. Within residue numbering those residues that determine the Ig-fold (Chothia et al., 1998); **b**: buried in ß-sheets, **t**: residues in turns, **i**: at the interface domains. In the sequences, letters in bold show the conserved ß-sheets framework, letters in italics represent the hypervariable loops. Discontinuous lines under TCR sequence (---) indicate framework residues with backbone coordinates assigned using the reference structure.

Figure 2. Comparison of the loops that form the antigen binding site in the 5C.C7 molecule (light blue) after molecular refinement and the x-ray structures using as reference in the molecular modeling procedure (orange). A red sphere highlight the Cα atoms of those residues that determine the canonical structures in the Ig hypervariable loops. In α1 and β1 loop the residues at position 2 are not represented in the figure. In the β1 loop is shows the distance of the possible hydrogen bond between the His-29 and Ser-94.

Figure 3. 3D-profile of the V domains of 5C.C7 atomic model.

Figure 4. Alignment between HLA-DR1 and I-E^k sequences. Upper part of the figure: α1 domains, lower part: β1 domains. In the sequences, letters in bold

show the β-sheets, letters in italics shows α-helices as were defined for the HLA-DR1 structure by Kabcsh & Sander algorithm (Kabcsh & Sander, 1983).

Figure 5. Ribbons drawing of the 5C.C7/MCC(93-103)/I-E^k atomic model. Color code for the molecules: I-E^k: light blue; MCC(93-103): blue; 5C.C7 framework: white; for the loops that form the antigen binding site, α1: violet, α2: red, α3: brown, β1: light green, β2: yellow, β3: green.

Figure 6. Solvent accessible surface generated by the residues which interact in the 5C.C7/MCC(93-103)/I-E^k complex. The surface was determined by the Connolly algorithm (Connolly, 1993) with a sphere of 1.7 Å diameter. **a**. View of the complex with the molecules separated by 10 Å to show the surface complementarity. **b**. Surface view obtained by rotating 90° the 5C.C7 and MCC(93-103)/I-E^k structures with respect to the figure **a**. Color code as in figure 5.

Table 1. Residues which interact in the 5C.C7/MCC(93-103)/I-E^k complex. An asterisk represent a possible hydrogen bond, two asterisk a possible salt bridge, no asterisk van der Waals contacts. Definition of atoms in contact as in Padlan (1994).

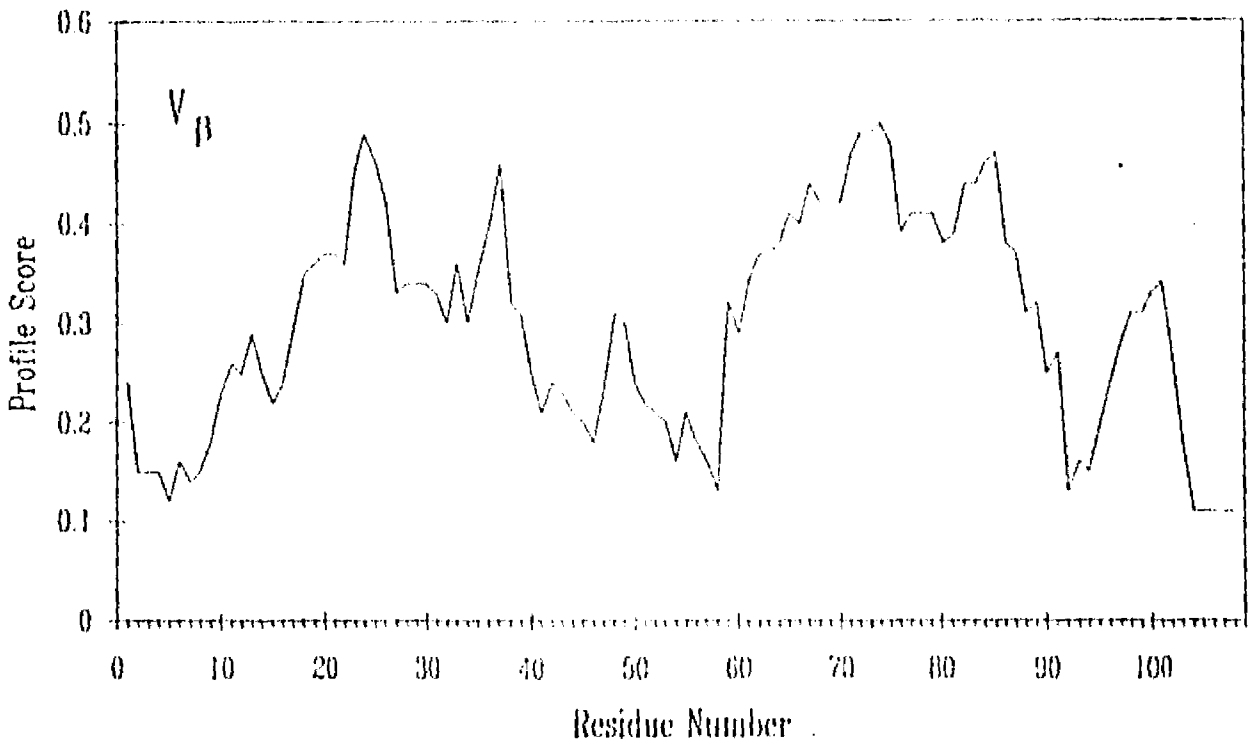
V_α/V_L:

```
          1         10        20        30        40        50
          .|.b.b....|.b..t..b.b.b.b.....|.b.bibi..t..i.ibb..|....|..
ALPHA:  gdqVeQspsalsLheGtgSaLrCnFtttmra-VqWFRKnsRgsLinLFy --lasgtk
           . |||  ||.  |  . .  | .        . |..*  * |. *|*|
D1.3L:  divLtQspaslsAsvGetVtItCrAsgnihnyLaWYQQkqGksPqLLvy ytttladg
           A     A'        B       α1     C        C'       α2
          ------------------------        ----------------   ------

           60        70        80        90        100       110
          ..|tt.b.....|b.b.b..t.|.t.b|bibi|.........|.....ib.bb.b.b.
ALPHA:  engRLkSafdskerySt LhIrdAqleDsGtYFCAa     easntnkvvFGtGTrLqVl
             |. .           |.|    | || | |.|.              *|| || | .
D1.3L:  vpsRFsGsgsg--tqYsLkInslqpeDfGsYYCQh     --fwstprtFGgGTkLeIk
           D         E                   F       α3      G
          ----------  ---------------------              ----------
```

V_β/V_H:

```
          1         10        20        30abc     40        50
          |..b|b...|..b..t..b|b.b.b....|....b.bibi..t..i.ibb..|.......
BETA :  nskViQtprylvKgqGqkAkMrCiPekghpv---VfWYQQnkNneFkFlInfgngevlgg
           . .     .*  *| . . | .          . |..|  .  . . .|*
D1.3H:  qvqLQES gpglVapSQsLsItCtVsgfsltgygVnWVRQppGkgLeWLGmiwgdgntdy
           A    A'        B      ß1       C        C'    ß2
          ------ ----------------          ------------------------

           60        70        80        90        100       110
          ..|...tt.b..| ....b.b.b|.t...t.b|bibi.....|.....ib.bb.b.b...
BETA :  idmtekRFsAecp-snspCsLeIqsSeaqDsAlYLCAs  slnnansdytFGsGTrLlVi
             |.|.    | | . | .  .    | | |.||              || || | |
D1.3H:  nsalksRLsIskdnsksqVfLkMnsLhtdDtArYYCAr  --erdyrldyWGsGTtLtVs
           D          E                   F       ß3      G
          ------------------------------------              ----------
```


V_α
Profile Score
0.6
0.5
0.4
0.3
0.2
0.1
0
0
10
20
30
40
50
60
70
80
90
100
Residue Number

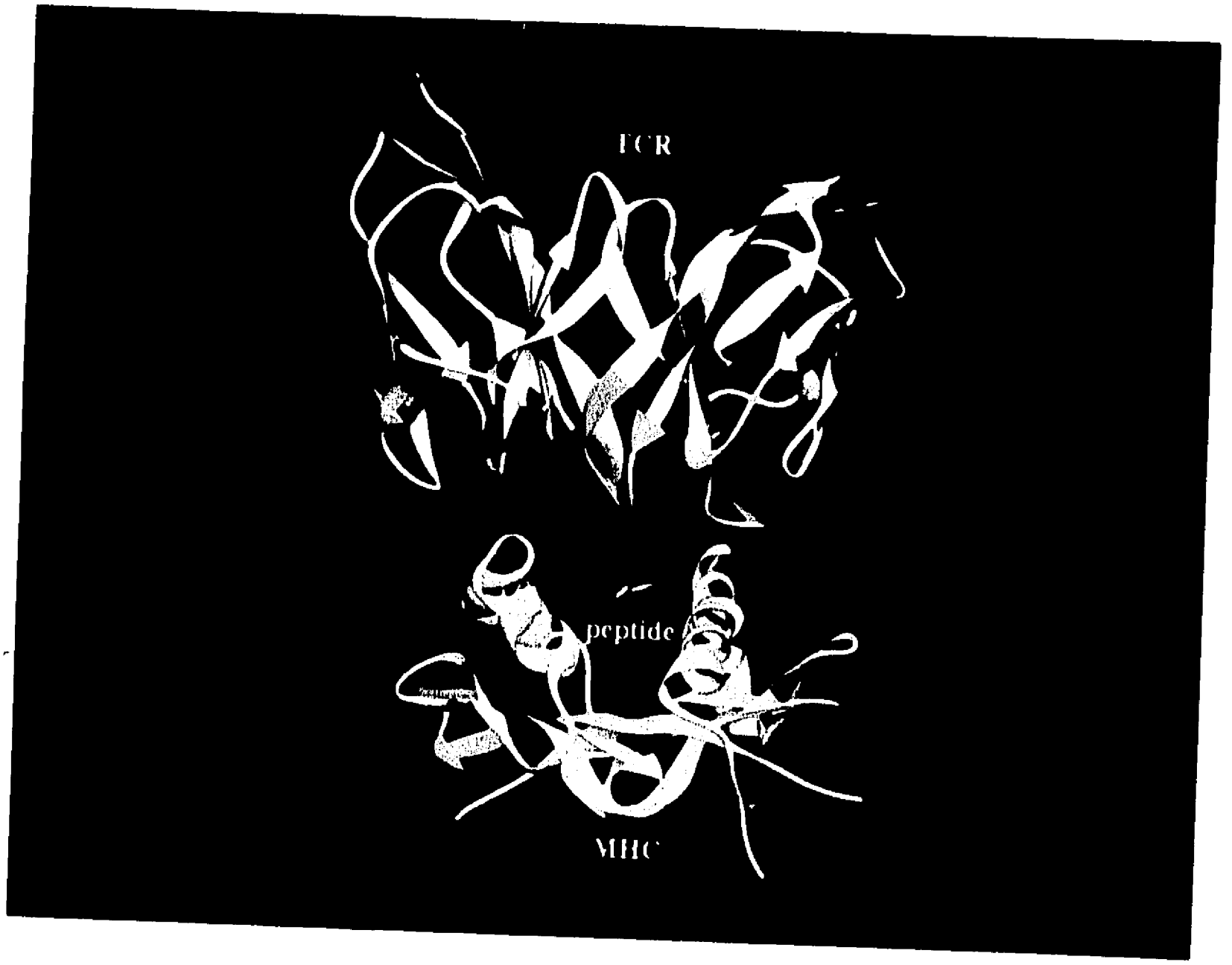

V_β
Profile Score
0.6
0.5
0.4
0.3
0.2
0.1
0
0
10
20
30
40
50
60
70
80
90
100
Residue Number

α1 domain:

```
                10        20        30        40        50
          ..|....|....|....|....|....|....|....|....|....|....|
I-E^k_α : eehtiiqaefyllpdkrgefmfdfdgdeifhvdieksetiwrleefakfaspe
          ||| |||||||| ||  ||||||||||||||||  | || ||||||  ||| |
 DR1_α: EEHVIIQAEFYLNPDQSGEFMFDFDGDEIFHVDMAKKETVWRLEEFGRFASFE
          S1             S2        S3        S4

              60        70        80
          ....|....|....|....|....|....
I-E^k_α:  aggalaniavdkanldymkersnntpdan
          ||||||||||||||  |  ||| ||  |
 DR1_α: AQGALANIAVDKANLEIMTKRSNYTPITN
          H1
```

ß1 domain:

```
                  10        20        30        40        50
            ....|....|....|....|....|....|....|....|....|....|
I-E^k_α : frpwpleycksechfyngtqrvrllvryfynleenlrfdsdvgefravte
            ||   |  |||||||||| ||||| |  || ||  |||||||| |||||
 DR1_α: TRPRFLWQLKFECHFFNGTERVRLLERCIYNQEESVRFDSDVGEYRAVTE
              S1'             S2'         S3'         S4'

                  60        70        80        90
            ....|....|....|....|....|....|....|..
I-E^k_α : ltelgrpdaenwnsqpeilegkraevdtvcrhnyeifdnflvpr
            |||||||||| ||||   ||| || ||| |||||      | | |
 DR1_α: LTELGRPDAEYWNSQKDLLEQRRAAVDTYCRHNYGVGQSFTVQR
              H1'       H2'           H3'
```


TCR
peptide
MHC

V_α	TCR residues in contact	peptide/MHC residues in contact
α1	T-27	T-77(ß)
	R-29	A-73(ß), E-74(ß)**,T-77(ß)*
	A-31	E-66(ß)
α2	Y-48	Q-64(ß)
	L-52	P-65(ß), E-69(ß)
	A-53	P-65(ß)
α3	E-97	K-99(p)**
	S-99	T-77(ß)*, V-78(ß)*
	N-100	T-77(ß), V-78(ß), H-81(ß), A-96(p)
	T-101	E-59(α)*
	N-102	Q-70(ß)*, Y-97(p)*, K-99(p),E(74(ß)*
Other	S-68	E-69(ß)*
	K-69	D-76(ß)
$V_ß$		
ß2	Q-50	V-69(α), A-65(α), Q-100(p)*
	N-51	A-72(α), T-102(p)
	Q-52	K-71(α), A-72(α)
	E-53	K-71(α)**
	V-54	L-64(α), A-68(α), V-69(α)
	L-55	Q-61(α), L-64(α)
	Q-56	Q-61(α), Y-97(p)*
	Q-57	Q-61(α)
ß3	N-100	T-102(p)*, K-103(p)*
	N-101	Q-64(ß)*, A-101(p), T-102(p)*,E-66(ß)
	A-102	E-66(ß), K-99(p)
	N-103	E-66(ß)*
	S-104	Q-64(ß),E-66 (ß)

Conclusiones

El objetivo central de la presente tesis fue estudiar la relación estructura-función en los TCRs. Para esto, en el capitulo I se delineó el modelo de interacción TCR/péptido/MHC propuesto hasta el presente y se analizaron sus limitaciones. En el mismo capítulo, se discutieron los avances metodológicos y experimentales que permiten refinar el estudio estructural del reconocimiento molecular mediado por este receptor del sistema inmunológico.

En el capítulo II se desarrolló un paquete de programas denominado VIR. Este paquete de programas permite manejar, de manera eficiente, el gran volumen de información generado hasta la fecha sobre los receptores del sistema inmunológico. La base para describir VIR fue la información disponible sobre las Igs. La aplicación de esta herramienta de cómputo al manejo de las secuencias de los TCRs, fue la base metodológica del capítulo III.

En el capítulo III, se utilizaron tres índices de variabilidad simultáneamente con el objetivo de estudiar las secuencias conocidas de los TCRs. Los perfiles de variabilidad obtenidos por cada uno de los índices, se filtraron con series de Fourier para delimitar objetivamente las regiones de máxima variabilidad o hipervariables. Estas regiones deben formar el sitio de reconocimiento antigénico del TCR.

Los resultados muestran que los tres índices tienen distinta resolución para diferentes niveles de variabilidad. Así, su uso simultáneo, compensa las deficiencias de cada uno de ellos por separado para estimar la variabilidad. Al aplicar el filtrado con series de Fourier, se encontró que en los dominios funcionales de los TCRs, a diferencia de las Igs, la región correspondiente al CDR-1, no es hipervariable en el TCR.

Si se considera el modelo general de interacción descrito en el capítulo I, la explicación del comportamiento relativamente conservado del CDR-1 en

los TCRs, se justifica en términos de que esta región, interactua con una región poco variable del MHC. Esta proposición tiene implicaciones para descripcion general de la interacción TCR/péptido/MHC. Esto es, el CDR-1 de los dominios Vα y Vβ del TCR pudiera no tener un papel importante en la definición de la especificidad fina del complejo péptido/MHC. Una derivación de lo anterior es que se reduce el estudio de la especificidad fina del TCR por el complejo péptido/MHC a los otros CDRs que forman el sitio de reconocimiento antigénico.

Para estudiar los detalles estructurales del reconocimiento TCR/péptido/MHC, en el capítulo IV, se obtuvo un modelo atómico de interacción TCR/péptido/MHC. Los resultados de este modelo permiten asignar un papel para cada CDR en la interacción con el complejo péptido/MHC.

El modelo obtenido muestra que el CDR-1 de Vβ no interactua con la molécula de MHC, mientras que el CDR-1 de Vα contacta con residuos relativamente conservados del MHC. Este resultado es congruente con la proposición del capítulo III. Además de este resultado, el análisis estructural del complejo TCR/péptido/MHC muestra que el CDR-2 de Vα pudiera tener un papel importante en el reconocimiento de las diferentes clases de MHC. El CDR-2 de Vβ por otra parte, pudiera determinar tanto el reconocimiento del MHC como el del péptido antigénico. Estas proposiciones tomadas en conjunto difieren cualitativamente de las predicciones del modelo general de interacción TCR/péptido/MHC discutido en el capitulo I, donde el papel de los CDR-1 y 2 de los TCRs en la interacción con el complejo péptido/MHC era incierto.

La proposición del papel que juega cada CDR del sugiere una buena hipótesis de trabajo para diseñar experimentos precisos que aporten nuevas luces a la descripción de la interacción TCR/péptido/MHC. Además, la

descripción detallada de la interacción TCR/péptido/MHC que se presenta en el capítulo IV, pudiera servir de referencia para explicar resultados experimentales ya realizados.

Perspectivas

Como se mencionó en la introducción de la tesis, nuestro trabajo fue orientado por una pregunta: ¿Existen reglas del reconocimiento molecular mediado por los receptores inmunológicos: Igs y TCRs? Esta pregunta se hizo explícita en el capítulo II para diseñar VIR.

Una respuesta afirmativa a esta pregunta implicaría poder comenzar a entender las bases moleculares del reconocimiento mediado por las Igs y los TCRs, hasta el presente la única fuente molecular de especificidad del sistema inmunológico. Desde el punto de vista aplicado, implicaría poder diseñar de manera racional Igs o TCRs de especificidad predefinida.

En los estudios realizados por nuestro grupo sobre las Igs (ver tesis de doctorado de E. Vargas-Madrazo, 1995) se llega a la conclusión general de que un primer reconocimiento general antígeno/Ig es determinado por la forma de sitio de reconocimiento de este receptor del sistema inmunológico. Esto sugiere que las Igs tienen restricciones geométricas en su mecanismo de reconocimiento molecular.

En la presente tesis se predice que cada CDR que forma el sitio de reconocimiento antigénico en el TCR puede tener una función particular en el reconocimiento de su ligando, el complejo péptido/MHC. Puesto que existen similitudes tanto a nivel genético como estructural entre las Igs y los TCRs, se pudiera combinar la experiencia de las Igs con estos resultados presentados para los TCRs. Esto permitirá en el futuro buscar similitudes y diferencias en los mecanismos básicos de reconocimiento molecular de los receptores del

sistema inmunológico. Así, la pregunta central de nuestro trabajo cobra una perspectiva aún más general. Este es el trabajo futuro.

Conclusiones/pág.4

Bibliografía

Acuto O, Hussey R, Fitzgeral K, Protentis J, Meuer S, Reinherz E. 1983. The human T-cell receptor: Appearance in ontogeny and biochemical relationship of α and β subunits on IL-2 dependent clones and T-cell tumors. *Cell. 34:*717-726.

Ajitkumar P, Geier SS, Kesari KV, Borriello F, Nakagawa M, Bluestone JA, Saper MA, Wiley DC, Nathenson SG. 1988. Evidence that multiple residues on both the α-helices of the class I MHC molecule are simultaneously recognized by the T cell receptor. *Cell. 54:*47-56.

Allinson JP, McIntyre BM, Bloch D. 1982. Tumor-specific antigen of murine T-lymphoma defined with monoclonal antibody. *J. Immunol.129:*2293-3000.

Almagro JC, Vargas-Madrazo E, Zenteno-Cuevas R, Hernández-Mendiola V, Lara-Ochoa F. 1995. VIR: A computational tool for the analysis of immunoglobulin sequences. *BioSystems*. in press.

Almagro JC, Zenteno-Cuevas R. Vargas-Madrazo E, Lara-Ochoa F. 1995. Variability analysis of the T-cell receptors using three variability indexes. *Int. J. Pep. & Prot. Res.* in press.

Almagro JC, Ceceña HA, Vargas-Madrazo E, Lara-Ochoa F. 1995. Molecular dynamics simulation of an immunoglobulin Fv fragment In: *POP'95*. Oxford University press. in press.

Almagro JC, Vargas-Madrazo E, Lara-Ochoa F, Horjales E. 1995. Molecular modeling of a T-cell receptor bound to a major histocompatibility complex: Implications for the T-cell recognition. enviado a *Protein Sci.*

Amzel LM, Poljak RJ. 1979. Three-dimensional structure of immunoglobulins. *Annu. Rev. Biochem. 48:*961-997.

Becker DM, Patten P, Chien Y, Yokota T, Eshhar Z, Giedlin M, Gascoigne NRJ, Goodnow C, Wolf R, Arai K, Davis MM. 1985. Variability and repertoire size of T-cell receptor V_α gene segments. *Nature 317:*430-434.

Bell GI. 1990. The human genome: An introduction. pp. 3-11. In: *Computers and DNA*, ed. Bell GI and Marr TG. Santa Fe Institute studies in the Sciences of Complexity Vol.7. Addison Wesley.

Bellio M, Lone Y-C, Calle-Martin O, Malissen B, Abastado J-P, Kourilsky P. 1994. The Vβ complementary determining region 1 of major histocompatibility complex (MHC) class I-restricted T cell receptor is involved in the recognition of peptide/MHC I and superantigen/MHC II complex. *J. Exp. Med.179*:1087-1097.

Berek C, Milstein C. 1988. The dynamic nature of the antibody repertoire. *Immunol. Rev. 105*:5-26.

Bhat TN, Bentley GA, Fischmann TO, Boulot G, Poljak RJ. 1990. Small rearrangements in structures of Fv and Fab fragments of antibody D1.3 on antigen binding. *Nature 347:*483-485.

Bernstein FC, Koetzle TF, Williams GJB, Meyer EF Jr, Brice MD, Rodgers JR, Kennard O, Shimandouchi T, Tasumi M. 1977. The Protein Data Bank. A computer-based archival file for macromolecular structures. *J. Mol. Biol. 112*:535-542.

Bjorkman PJ, Parham P.1990. Structure, function, and diversity of class I major histocompatibility complex. *Ann. Rev. Biochem. 59*:253-288.

Bjorkman PJ, Saper MA, Samraoui B, Bennett WS, Strominger JL, Wiley DC 1987. Structure of the human class I histocompatibility antigen, HLA-A2. *Nature 329:*507-512.

Bjorkman PJ, Saper MA, Samraoui B, Bennet WS, Strominger JL, Wiley DC. 1987. The foreign antigen binding site and T cell recognition regions of class I histocompatibility antigens. *Nature. 329:*512-519.

Bougueleret L, Claverie J-M. 1987. Variability analysis of the human and mouse T-cell receptor ß chains. *Immunogenetics 26*:304-308.

Brown JH, Jardetzky T, Saper MA, Samraoui B, Bjorkman PJ, Wiley Don C. 1988. A hypothetical model of the foreign antigen binding site of class II histocompatibility molecules. *Nature 332*:845-850.

Brown JH, Jardetzky TS, Gorga JC, Stern LJ, Urban RG, Strominger JL, Wiley 1993. Three-dimensional structure of the human class II histocompatibility antigen HLA-DR1. *Nature 304:*33-39.

Chothia C, Lesk AM. 1987. Canonical structures for the hypervariable regions of immunoglobulin. *J. Mol. Biol. 196:*901-917.

Chothia C, Boswell DR, Lesk, AM. 1988. The outline structure of the T-cell α:ß receptor. *EMBO J. 7*:3745-3755.

Chothia C, Lesk AM, Tramontano A, Levitt M, Smith-Gill SJ, Air G, Sheriff S, Padlan EA, Davies D, Tulip WR, Colman PM, Spinelli S, Alzari PM, Poljak RJ. 1989. Conformations of immunoglobulin hypervariable regions. *Nature (London) 342*:877-883.

Chothia C, Lesk AM, Gherardi E, Tomlinson IM, Walter G. Marks JD, Llewelyn MB, Winter G. 1992. Structural repertoire of the human VH segments. *J. Mol. Biol. 227*:799-817.

Claverie J-M, Prochnicka-Chalufour A, Bougueleret L. 1989. Implications of a Fab-like structure for T-cell receptor. *Immunol. Today. 10*:10-14.

Connolly ML. 1983. Solvent-accessible surfaces of proteins and nucleic acids. *Science 221:*709-713.

Davies DR, Padlan EA, Sheriff S. 1990. Antibody-Antigen Complexes. *Annu. Rev. Biochem. 59*: 439-473.

Davis MM, Bjorkman PJ. 1988. T-cell antigen receptor genes and T-cell recognition. *Nature 334*:395-402.

Davis MM. 1990. T cell receptor gene diversity and selection. *Ann. Rev. Biochem. 59*:475- 496.

Dayhoff MO, Eck RV. 1966. *Atlas of Protein Sequence and Structure*. Vol.2 NBRF Press, Silver Spring.

DiGiusto DL, Palmer, E. 1994. An analysis of sequence variation in the β chain framework and complementarity determining regions of an allo-reactive T cell receptor. *Mol. Immunol. 31*:693-699.

Fremont DH, Matsumura M, Stura EA, Peterson PA, Wilson IA. 1992. Crystal structures of two viral peptides in complex with murine MHC class I H-2K^b. *Science 257*:919-927.

Ganju RK, Smiley ST, Bajorath J, Novotny J, Reinherz EL. 1992. Similarity between fluorescein-specific T-cell receptor and antibody in chemical details of antigen recognition. *Proc. Natl. Acad. Sci. USA. 89*:11552-11556.

Genot J & Kollman PA. 1992. Molecular dynamics studies of a DNA-binding protein: 2. An evaluation of implicit and explicit solvent models for the molecular dynamics simulation of the *Escherichia coli trp* repressor. *Prot. Sci. 1*:1185-1205.

Guo H-C, Jardetzky TS, Garrett TP, Lane WS, Strominger JL, Wiley DC. 1992. Different length peptides bind to HLA-Aw68 similarly at their ends but bulge out in the middle. *Nature 360*:364-366.

Haskins K, Kappler J, Marrack P. 1984. The major histocompatibility complex-restricted antigen receptor on T-cells. *Ann. Rev. Immunol. 2*:51-66.

Hedrick SM, Cohen DI, Nielsen EA, Davis MM. 1984. Isolation of CDNA clones encoding T cell-specific membrane associated proteins. *Nature 308*:149-153.

Hedrick SM, Engel I, McElligot DL, Fink, PJ, Hsu ML, Hansburg D, Matis A. 1988. Selection of amino acid sequences in the β chain of T cell antigen receptor. *Science 239*:1541-1544.

Jores R, Alzari PM, Meo T. 1990. Resolution of hypervariable regions in T-cell receptor ß chains by a modified Wu-Kabat index of amino acids diversity. *Proc. Natl. Acad. Sci. USA. 87*:9138-9142.

Jorgensen JL, Reay PA, Ehrich EW, Davis MM. 1992. Molecular components of T-cell recognition. *Annu. Rev. Immunol. 10*:835-873.

Jorgensen JL, Esser U, Fazekas de St. Groth B, Reay PA, Davis MM. 1992. Mapping T-cell receptor-peptide contacts by variant peptide immunization of single-chain transgenics. *Nature 355*:224-230.

Kabat EA, Wu TT. 1971. Attempts to locate complementarity determinine residues in the variable positions of light and heavy chains. *Ann. NY. Acad. Sci. 190*:382-383.

Kabat EA, Wu TT, Perry HM, Gottesman KS, Foeller C. 1991. Sequences of Proteins of Immunological Interest. 5th ed. Bethesda, MD: U.S. Department of Health and Human Services, Public Health Service, National Institutes of Health, NIH Publication No. 91-3242.

Kabcsh W, Sander C. 1983. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers. 22:*2577-2637.

Kang J, Chambers, C, Pawling J., Scott C, Hozumi N. 1994. Conserved amino acids residues in the complementarity determining region 1 of the TCR β-chain are involved in the recognition of conventional Ag and Mls-1 superantigen. *J. Mol. Immunol. 152:*5305-5317.

Kodeira M, Kinashi T, Umemura I, Matsuda F, Noma T, Ono Y, Honjo T, 1986. Organization and evolution of the variable region genes of the human immunoglobulin heavy chain. *J. Mol. Biol. 190:*529-541.

Kobori JA, Hood L, Shastri N. 1992. Structure-function relationship among T-cell receptors specific for lysozyme peptides bound to A^b or $A^{bm\text{-}12}$ molecules. *Proc. Natl. Acad. Sci. USA. 89:*2940-2944.

Kraulis PJ. 1991. MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr. 24:*946-950.

Kronenberg M, Siu G, Hood LE, Sastri N. 1986. The molecular genetics of the T-cell antigen receptor and T-cell antigen recognition. *Ann. Rev. Immunol. 4:*529-591.

Lara-Ochoa F, Vargas-Madrazo E, Jiménez-Montaño MA, Almagro, JC. 1994. Patterns in the complementarity determining regions of immunoglobulin. *BioSystems 32:*1-9.

Lara-Ochoa F, Vargas-Madrazo E, Almagro JC. Distribution of the use of amino acids in the hypervariable regions of immunoglobulins. *J. Mol. Evol.* in press.

Latron F, Pazmany L, Morrison J, Moots R, Saper MA, McMichael A, Strominger JL. 1992. A critical role for conserved residues in the cleft of HLA-A2 in presentation of a nonapeptide to T cell. *Science 257:*964-967.

Leiden JM, Strominger JL. 1986. Generation of diversity of the ß chain of human T-lymphocyte receptor for antigen. *Proc. Natl. Acad. Sci. USA. 83:*4456-4460.

Lone Y-C, Bellio M, Prochnicka-Chalufour A, Ojcius DM, Boissel N, Ottenhoff TMH, Klausner RD, Abasto J-P, Kourilsky P. 1994. Role of the CDR1 region of the TCR β chain in the binding to purified MHC-peptide complex. *Int. Immunol. 6:*1561-1565.

Lüthy R, Bowie JV, Eisenberg D. 1992. Assessment of protein models with three-dimensional profiles. *Nature. 356:*83-85.

Madden DR, Garboczi DN, Wiley DC. 1993. The antigenic identity of peptide-MHC complexes: A comparison of conformation of five viral peptides presented by HLA-A2. *Cell. 75:*693-708.

Matsumura M, Fremont DH, Peterson PA, Wilson IA. 1992. Emerging principles for the recognition of peptide antigens by MHC class I molecules. *Science 257:*927-934.

Marquart M, Deisenhofer J, Huber R. 1980. Crystallographic refinement and atomic models of the intact immunoglobulin molecule Kol and its antigen-binding fragment at 3.0 Å and 1.9 Å resolution. *J. Mol. Biol. 141:*369-391.

Mian IS, Bradwell AR, Olson AJ. 1991. Structure, function and properties of antibody binding sites. *J. Mol. Biol. 217:*133-151.

Nalefski EA, Wong JGP, Rao A. 1990. Amino acids substitution in the first complementary-determining region of a murine T-cell receptor α chain affect antigen-major histocompatibility complex recognition. *J. Biol. Chem. 265:*8842-8846.

Nalefski EA, Kasibhatla S, Rao A. 1992. Functional analysis of the binding site on the T cell receptor α chain. *J. Exp. Med. 175:*1553-1563.

Novotny J, Tonegawa S, Saito H, Kranz DM, Eisen H. 1986. Secondary, tertiary, and quaternary structure of T-cell-specific immunoglobulin-like polypeptide chains. *Proc. Natl. Acad. Sci. USA. 83:*742-746.

Padlan EA, Kabat EA. 1991. Modeling of antibody combining sites. *Methods Enzymol. 202-B:*3-21.

Padlan EA. 1994. The anatomy of the antibody molecule. *Mol. Immunol. 31:*169-193.

Pascual V, Capra D. 1991. Human immunoglobulin heavy-chain variable region genes: Organization, polymorphism, and expression. *Adv. Immunol. 49:*1-74.

Patten P, Yokota T, Rothbard J, Chien Y, Arai K, Davis MM. 1984. Structure, expression and divergence of T-cell receptor ß-chain variable regions. *Nature 132:*40-46.

Patten PA, Rock EP, Sonoda T, Fazekas de St.Groth B, Jorgensen JL, Davis MM. 1993. Transfer of putative complementary-determining regions loops of T-cell receptors V domains confer toxin reactivity but not peptide/MHC specificity. *J. Immunol. 150:*2281-2294.

Poljak RJ, Amzel LM, Avey HP, Chen BL, Phizacherley RP, Saul F. 1973. Three-dimensional structure of the Fab' fragment of a human immunoglobulin at 2.8-Å resolution. *Proc. Natl. Acad. Sci. U.S.A. 70:*3305-3310.

Prochnicka-Chalufour A, Casanova J-L, Avrameas S, Claverie J-M, Kourilsky P. 1991. Biased amino acids distribution in regions of the T cell receptors and MHC molecules potentially involved in their association. *Int. Immunol. 3:*853-864.

Rupp F, Brecher J, Giellin MA, Mosman T, Zinkernagel R, Hengarthner H, Joho RH. 1987. T cell antigen receptors with identical variable regions but different diversity and joining regions gene segments have distintic specificities but cross-reactive idiotypes. *Proc. Natl. Acad. Sci. USA. 84:*219-222.

Segal DM, Padlan AE, Cohen GH, Rudikoff S, Potter M, Davies DR. 1974. The three-dimensional structure of a phosphorylcholine-binding mouse immunoglobulin Fab and the nature of the antigen binding site. *Proc. Natl. Acad. Sci. USA. 71:*4298-4302.

Shenkin PS, Erman B, Mastrandrea LD. 1991. Information-theoretical entropy as a measure of sequence variability. *Proteins. 11:*297-313.

Silver ML, Guo H-C, Strominger JL, Wiley DC 1992. Atomic structure of a human MHC molecule presenting an influenza virus peptide. *Nature 360:*367-369.

Stanfield RL, Fieser TM, Lerner RA, Wilson IA. 1990. Crystal structures of an antibody to a peptide and its complex with peptide antigen at 2.8 Å. *Science. 248:*712-719.

Stern LJ, Brown JH, Jardetzky TS, Gorga JC, Urban RG, Strominger JL, Wiley DC. 1994. Crystal structure of the human class II MHC protein HLA-DR1 complexed with an antigenic peptide from influenza virus. *Nature*. in press

Tonegawa S. 1983. Somatic generation of antibody diversity. *Nature (London) 302:*575-581.

Tramontano A, Chothia C., Lesk AM. 1990. Framework residue 71 is a major determinant of the position and conformation of the second hypervariable region in the V_H domains of immunoglobulins. *J. Mol. Biol. 215*:175-182.

Vargas-Madrazo E, Almagro JC, Lara-Ochoa F, Jiménez-Montaño MA. 1993. Amino acids pattern in the recognition sites of immunoglobulins. *ECAL'93.*

Vargas-Madrazo E, Almagro JC, Lara-Ochoa F. 1995. Structural repertoire in V_H pseudogenes of immunoglobulins: Comparison with human germline genes and human amino acid sequences. *J. Mol. Biol. 246:*74-81.

Vargas-Madrazo E, Horjales E, Lara-Ochoa F, Almagro JC. 1995. Canonical Structure repertoire of the antigen-binding site suggest strong geometrical restrictions associated to the mechanism of immune recognition. enviado a *J. Mol. Biol.*

Wilson IA, Stanfield RL. 1993. Antibody-antigen interactions. *Curr. Op. Struct. Biol. 3*:113-118.

Wu TT, Kabat EA. 1970. An analysis of the sequences of the variable regions of Bence Jones proteins and myeloma light chains and their implications for antibody complementarity. *J. Exp. Med. 132*:211-250.

Wysocki LJ, Gefter ML. 1989. Gene conversion and the generation of antibody diversity. *Annu. Rev. Biochem. 58:*509-531.

Yoshikai Y, Kimura N, Toyonaga B, Mak TM. 1986. Sequences and repertoire of human T cell receptor α chain variable region genes in mature T lymphocytes. *J. Exp. Med. 164:*90-103.

Young MN, Oomen RP (1992) Analysis of sequence variation among legume lectins: A ring of hypervariable residues forms the perimeter of the carbohydrate-binding site. *J. Mol. Biol. 228:*924-934.

Zinkernagel RM, Doherty PC. 1974. Immunological surveillance against altered self components by sensitized T lymphocytes in lymphocytic choriomeningitis. *Nature 251:*547-548.

Zucherkandl E. 1976. Evolutionary Processes and Evolutionary Noise at Molecular Level. *J. Mol. Evol. 7:*167-183.