

# STRUCTURE, FUNCTION, AND DIVERSITY OF CLASS I MAJOR HISTOCOMPATIBILITY COMPLEX MOLECULES

*Pamela J. Bjorkman*

Division of Biology and Howard Hughes Medical Institute, California Institute of Technology, Pasadena, California 91125

*Peter Parham*

Department of Cell Biology, Stanford University, Stanford, California 94305

KEY WORDS: MHC, HLA molecules, peptide binding, polymorphism, T cell recognition.

---

## CONTENTS

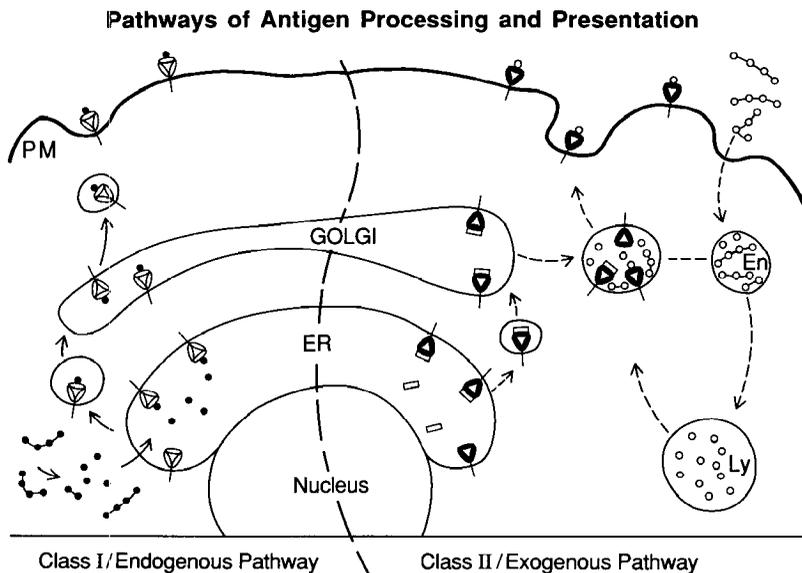
INTRODUCTION.....	254
<i>Class I MHC Polypeptides</i> .....	256
THE STRUCTURE OF HLA-A2.....	257
<i>Structural Overview</i> .....	257
<i>Structure of the Ig-like Domains</i> .....	259
<i>Structure of the <math>\alpha_1</math> and <math>\alpha_2</math> Domains</i> .....	262
<i>Size of Peptides that Fit in the Site</i> .....	267
SEQUENCE DIVERSITY IN CLASS I HLA HEAVY CHAINS.....	268
<i>The <math>\alpha_3</math> Domain</i> .....	275
<i>The <math>\alpha_1</math> and <math>\alpha_2</math> Domains</i> .....	275
ANTIGENIC PEPTIDE MOTIFS AND THE CONFORMATION OF BOUND PEPTIDES.....	280
WHAT IS SEEN BY A TcR?.....	281
<i>A Model for How TcRs Bind to the Peptide-MHC Complex</i> .....	282

## INTRODUCTION

Lymphocytes are the agents of antigenic specificity in the immune response. They divide into two groups: B cells that make immunoglobulins (Igs), and T cells that perform a heterogeneous set of functions including help for B cells, production of delayed-type hypersensitivity reactions, and the specific killing of virus-infected cells. Igs are the sole source of B cell specificity, whereas molecules from two diverse families of cell-surface glycoproteins, the T cell receptors (TcRs) and the major histocompatibility complex (MHC) glycoproteins, are the key elements of specificity in the T cell response to foreign antigens.

TcRs, in common with the Ig receptors of B cells, are products of somatically rearranging genes and are clonally expressed (reviewed in Refs. 1, 2 and M. M. Davis, this volume). There are, however, two important differences in antigen recognition by these related groups of lymphocyte receptors. The first is a difference in the type of antigenic determinant (epitope) recognized: TcRs recognize short, linear, peptide determinants of 10–20 amino acids (3–5), the generation of which usually requires unfolding and proteolytic fragmentation (“processing”) of the antigenic protein (6). By contrast, Igs commonly interact with epitopes formed by the three-dimensional structure of native proteins, although they can also be raised against peptides. The second difference in antigen recognition by Igs and TcRs is the involvement of a third molecule that performs the role of “presenting” the antigen to the receptor. For B cells such molecules do not exist; the antibody receptor forms a stable bimolecular complex with the antigenic protein. For T cells, the antigenic peptide must be bound by an MHC glycoprotein, and it is this complex of MHC molecule plus peptide that forms the structure recognized by the TcR. MHC glycoproteins are thus peptide-binding proteins, and can be considered as antigen-presenting molecules.

There are two structurally distinct, but related, families of MHC molecules that present antigens to two subsets of T cells: Class I MHC molecules present antigens to T cells that express the CD8 cell-surface glycoprotein, and class II MHC molecules present antigens to T cells that express the CD4 cell-surface glycoprotein (7, 8). Other less rigid distinctions between class I and II MHC molecules are in the categories of antigen presented and in the functional activities of the responding T cells. Class I molecules commonly present peptides derived from endogenously synthesized proteins, such as viral components produced upon virus infection, and result in the stimulation of CD8-bearing cytotoxic (killer) T cells (9). On the other hand, class II molecules generally present peptides derived from exogenously synthesized proteins and stimulate CD4-bearing helper T cells (10). This specificity in the antigens presented by class I and class II molecules is thought to result from cellular segregation of the compartments or sites at which class I and class II MHC molecules encounter peptide (11, 12) (Figure 1).



*Figure 1* Cartoon showing the cellular processing and presentation of antigens. Endogenous proteins (●-●-●) are degraded into peptides (●), which are bound by class I MHC molecules (open triangles) in the endoplasmic reticulum (ER), and thus transported through the Golgi to the plasma membrane (PM). Exogenous antigens (○-○-○) enter the cell by endocytosis and are degraded in endosomal (En) and lysosomal compartments (Ly). Peptides (○) encounter class II MHC molecules (filled-in triangles) in a poorly defined cellular compartment before transport to the plasma membrane. The invariant chain (open rectangle) transiently associates with class II MHC molecules during exocytosis.

That T cells have a dual requirement for recognition of both foreign antigen and a self-MHC molecule was first recognized by cellular immunologists in the early 1970s (13, 14): T cell recognition of antigen became known as "MHC restricted," with considerable debate focussing upon its molecular interpretation and whether one or two receptor molecules were involved (15). Transfection experiments have now demonstrated that a single TcR recognizes both foreign antigen and MHC molecule (16). Discovery of the T cell requirement for fragmented protein antigens led to investigation of the antigen processing pathway, and revealed that incubation of antigen-presenting cells with short peptides circumvented the need for processing and provided the necessary target for the TcR (3, 17). Combining the concepts of MHC restriction and antigen processing led to the hypothesis that MHC molecules actually bind processed antigen fragments, thereby forming ligands for TcRs. The first direct evidence for this model came with the demonstration of binding of specific antigen peptides from hen egg lysozyme to a murine class II MHC molecule (18).

The experiments that defined and characterized MHC restriction were only

possible because MHC genes are polymorphic, with the products of different alleles being distinguishable by both antibodies and TcRs. Indeed, alloreactive responses in which the immune system of one individual makes antibodies and T cells specific for the MHC molecules of another, are the basis for the immunological rejection of transplanted tissues that led to the discovery of the Major Histocompatibility Complex (19). The number of alleles at MHC loci and the complexity of their patterns of substitutions are unique when compared to other mammalian genes, and the basis for this diversity and its functional significance are long-standing questions (20). Analyses of MHC restriction (14), of T cell epitopes (5), and of peptide binding (21) all showed that polymorphic differences between MHC proteins affected the peptide antigens presented and the T cells that could respond. Characterization of MHC alleles revealed specific regions of the encoded protein at which sequence diversity was concentrated (22), and the marked predominance of coding over silent substitutions within those regions strongly suggested that diversification was the result of natural selection (23, 24).

A model for how MHC restriction might work was eventually developed (15, 25): a TcR was postulated to interact with a complex of a peptide bound to a particular class I or class II MHC molecule, and a T cell selected by one combination of peptide plus MHC would only be restimulated by that same or a closely related peptide-MHC complex. The specificity of restriction would then result from sequence variability in the MHC molecule, which determines both the peptides capable of binding and the TcR selected. Determination of the three-dimensional structure of HLA-A2, a human class I MHC molecule (26, 27), has confirmed the general principles of this model and revealed considerable insight into the details of the molecular interactions involved and of the role played by MHC polymorphism. In addition, the structure has helped to rationalize the kinetics of peptide binding to MHC molecules and the inaccessibility of the site in all but a fraction of purified MHC molecules (28, 29). More importantly, it has provided a mechanistic understanding of immunological tolerance—the discrimination between self and nonself—which is arguably the most fundamental property of the immune system.

This review is focussed on the three-dimensional structure of HLA-A2 and its interpretation in terms of the diversity of human class I MHC molecules, and their interactions with peptides, TcRs, and the CD8 glycoprotein. Some of the conclusions, but undoubtedly not all, will extend to class II MHC molecules.

### *Class I MHC Polypeptides*

Class I MHC molecules are noncovalently linked heterodimers (30). The heavy chain (44 kd) is encoded within the MHC and spans the membrane bilayer, oriented with its N-terminal major portion on the outside of the cell. The extracellular portion of the heavy chain has been divided into three

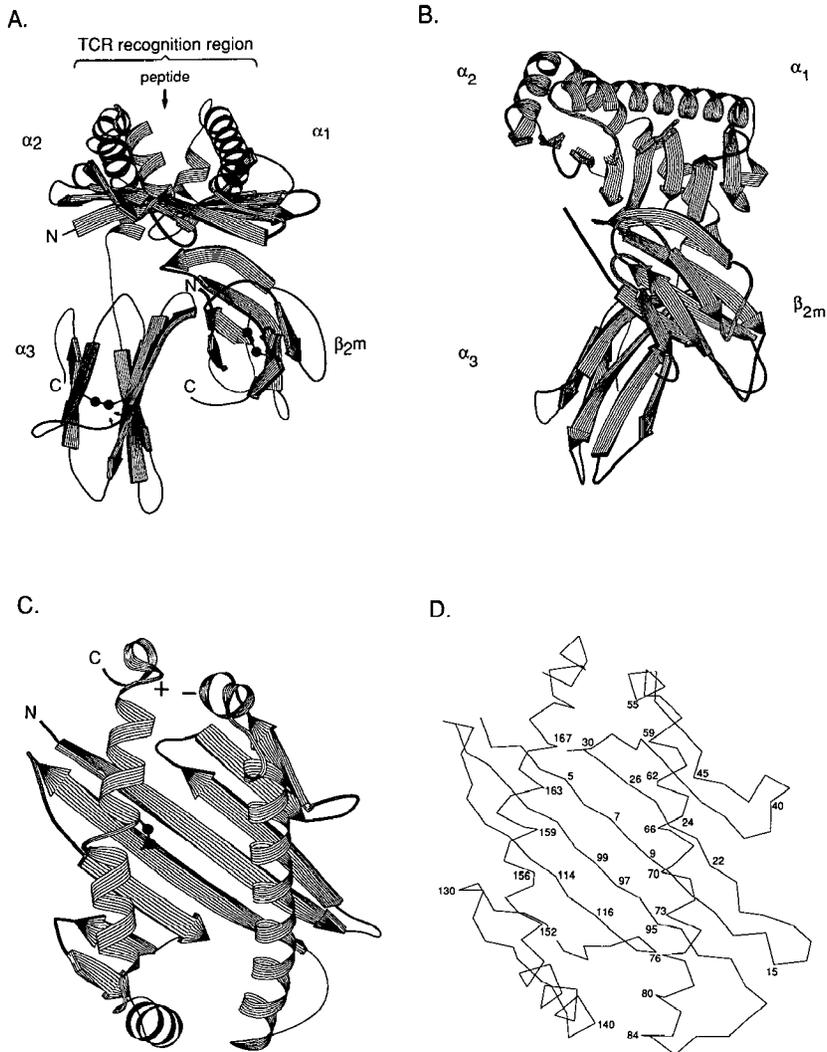
domains called  $\alpha_1$ ,  $\alpha_2$ , and  $\alpha_3$  (31), composed of 90, 92, and 92 amino acids, respectively, and encoded by separate exons (32). Following the  $\alpha_3$  domain is a membrane-spanning region of  $\sim 25$  amino acids and a cytoplasmic tail of  $\sim 30$  amino acids. The light chain is a non-MHC-encoded protein called  $\beta_2$ -microglobulin ( $\beta_2m$ , 12 kd) (33, 34). Both the  $\alpha_3$  and  $\beta_2m$  domains have sequence similarity to Ig constant regions (31, 35, 36, 37), and are therefore members of the Ig gene superfamily, which includes TcRs and class II histocompatibility antigens (38, 39). Most amino acid substitutions between the products of different class I alleles are localized to the  $\alpha_1$  and  $\alpha_2$  domains with the  $\alpha_3$  domain being relatively conserved, and  $\beta_2m$  invariant (22, 40). Since the sequence differences that result in restricted recognition by T cells and the differential ability to bind peptides mainly cluster in the  $\alpha_1$  and  $\alpha_2$  domains, it was expected that these domains would comprise the part of the molecule that binds peptides and is contacted by a TcR.

## THE STRUCTURE OF HLA-A2

The three-dimensional structure of HLA-A2 was determined using X-ray crystallographic techniques (26, 27). Prior to crystallization, the molecule was purified after papain cleavage from the cell membranes of the homozygous human lymphoblastoid cell line JY (41, 42). Papain cleaves the HLA heavy chain 13 residues before the transmembrane region (43), yielding a soluble heterodimeric fragment that retains all the serologically detected epitopes of the native cell-surface molecule. The soluble form of HLA-A2 crystallized in two closely related space groups, and isomorphous crystals of a second human class I MHC molecule, HLA-Aw68, were also obtained (44, 45). Both crystal forms grew as very thin ( $\sim 20$  to  $100 \mu\text{m}$ ) plates. Data from crystals of one space group were collected using high-intensity X-rays from synchrotron sources (46), which allowed the chain tracing of  $\sim 90\%$  of the backbone of HLA-A2. The remainder of the structure was traced after averaging electron density maps derived from the two HLA-A2 space groups (26; M. A. Saper et al, manuscript in preparation). The structure of HLA-A2 has now been refined to 2.6 Å resolution (M. A. Saper et al, manuscript in preparation; 47), and the structure of HLA-Aw68 has been solved and refined to 2.6 Å resolution (48). HLA-A2 and HLA-Aw68 differ by 13 amino acids; 6 in the  $\alpha_1$  domain, 6 in the  $\alpha_2$  domain, and 1 in the  $\alpha_3$  domain (49). The  $\alpha$ -carbon backbones of the two molecules are very similar (48); thus, the following description of the structure of HLA-A2 applies to HLA-Aw68, and presumably to other class I MHC molecules as well.

### *Structural Overview*

The HLA-A2 molecule is composed of two sets of structurally homologous domains; one set being the membrane-proximal  $\alpha_3$  and  $\beta_2m$  domains, and the other set being the membrane-distal  $\alpha_1$  and  $\alpha_2$  domains (Figure 2). As



**Figure 2** Schematic illustrations of HLA-A2 structure.  $\beta$ -strands shown as thick arrows,  $\alpha$ -helices represented as helical ribbons. Amino and carboxy termini of the heavy and light chains are indicated as "N" and "C." Disulfide bonds indicated as two connecting spheres. Figures modified from Ref. 26 and prepared with the aid of a program described in Ref. 126. (a) The four domains viewed from the side. The approximate location of residues 227 and 245 (implicated in CD8 binding) indicated with small arrows. (b) The molecule has been rotated 90° from the orientation in Figure 2a. (c) The  $\alpha_1$  and  $\alpha_2$  domains as viewed from the top of the molecule showing the peptide-binding site and surface presumably contacted by a TcR. (d) Carbon- $\alpha$  atoms of the  $\alpha_1$  and  $\alpha_2$  domains with residue numbers.

anticipated from their primary sequences (31, 35, 37), the  $\alpha_3$  and  $\beta_2m$  domains are folded into  $\beta$ -sandwich structures resembling Ig constant regions. However, these Ig domains interact with each other in a novel arrangement that has not been found in any of the known antibody structures including Fcs, Fabs, or light chain dimers (50, 51). The common fold of the  $\alpha_1$  and  $\alpha_2$  domains does not resemble an Ig variable or constant region, although  $\alpha_2$  contains a disulfide bond reminiscent of those found in typical Ig domains. The  $\alpha_1$  and  $\alpha_2$  domains sit on top of the Ig-like domains forming a "platform" consisting of an eight-stranded  $\beta$ -pleated sheet topped by two long  $\alpha$ -helices that lie diagonally across the sheet. The extracellular portion of the molecule is about 70 Å high (as oriented in Figure 2a), and cross-sectional dimensions are about 50 Å and 40 Å. Several lines of evidence suggest that the binding site for antigenic peptides is a deep cleft located between the two long  $\alpha$ -helices (27).

### *Structure of the Ig-like Domains*

The  $\beta$ -sandwich structure of the  $\alpha_3$  and  $\beta_2m$  domains are each composed of two antiparallel  $\beta$ -pleated sheets connected by an internal disulfide bond, as shown schematically in Figure 3. One sheet has four  $\beta$ -strands, while the other sheet has three. Both MHC Ig-like domains closely resemble antibody constant domains, with the regions of greatest similarity being located in the seven  $\beta$ -strands. The loops connecting these strands show some structural differences from those in antibody constant domains, an observation also made from a comparison of the structure of bovine  $\beta_2m$  with an Ig constant domain (52). As is common in  $\beta$ -structure, extensive hydrogen bonding is found between the main chain atoms of adjacent strands within each sheet.

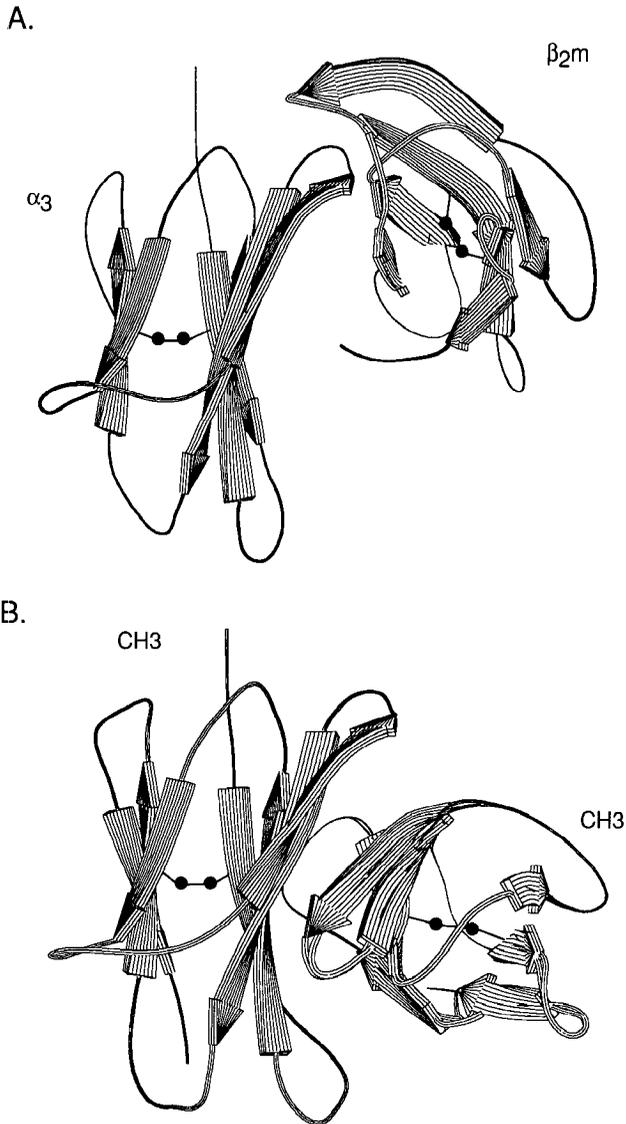
HLA-A2, Aw68, and  $\beta_2m$  represent the first members of the Ig gene superfamily, other than antibodies, for which three-dimensional structural information is available. Because MHC molecules contain two Ig-like domains, prior to determination of the three-dimensional structure it was generally assumed that the mode of subunit interaction in MHC molecules would be the same as found in antibody structures. Antibody constant domains contact each other with their four-stranded  $\beta$ -pleated sheets together, and one domain is related to the other by a dyad axis (a 180° rotation) of symmetry (50, 51). This same subunit interaction is preserved in the CH-CL pair of an Fab, the CH3 dimer in Fc, and the CL-CL dimer in a light chains dimer. It was therefore a surprise to find that the two Ig-like domains in the HLA-A2 structure interact in a novel way not previously found relating pairs of antibody domains. In the class I structure,  $\alpha_3$  is related to  $\beta_2m$  by a 146° rotation followed by a 13 Å translation along the rotation axis direction (26), instead of the 180° rotation without a translation that relates pairs of antibody domains. This interaction is unlikely to be an artifact caused by the packing of

HLA molecules within the crystal, since it is found in both crystal forms of HLA-A2, for which the packing of adjacent molecules differs. A schematic view of a pair of antibody constant domains can be compared with the  $\alpha_3$ - $\beta_2m$  pair in Figures 3a and 3b.

An examination of Figure 3a reveals that the  $\alpha_3$  and  $\beta_2m$  residues that are involved in the contact interface are within their respective four-stranded sheets, but the contact pairs are not symmetrical as would be found if the two domains were related by a two-fold symmetry axis. There are more extensive contacts between  $\beta_2m$  and the  $\alpha_1$ - $\alpha_2$  platform than between  $\alpha_3$  and the platform (26), because the effect of the 13 Å translation that relates  $\alpha_3$  and  $\beta_2m$  is to move  $\beta_2m$  up, so that it supports the  $\alpha_1$ - $\alpha_2$  platform. Thus  $\alpha_3$  interacts extensively only with  $\beta_2m$ , whereas  $\beta_2m$  interacts with all three domains of the heavy chain. Circular dichroism experiments suggest that the heavy chain of HLA loses its structure when separated from  $\beta_2m$  (53), and analysis of antigenic reactivity suggests that association with  $\beta_2m$  is accompanied by substantial changes in the conformation of the heavy chain (54–56). Association with  $\beta_2m$  also appears to be required for cell-surface expression of HLA class I heavy chains (57, 58), although one murine class I heavy chain (H-2D<sup>d</sup>) can be detected in small quantities at the surface of cells in the absence of  $\beta_2m$  (56). These studies indicated that association with  $\beta_2m$  facilitates or stabilizes the folding of class I heavy chains during biosynthesis (59), an idea substantiated by the way in which  $\beta_2m$  contacts all three heavy chain domains in the HLA-2 structure. Furthermore, studies of mutant cell lines defective in association of  $\beta_2m$  and class I heavy chains have shown that binding of an appropriate peptide may be critical for correct folding of the heavy chain, association with  $\beta_2m$ , and transport of the complex to the cell surface (60).

The HLA-A2 structure shows that the presence of a pair of Ig-like domains in a molecule need not necessarily imply an association between those domains that is similar to that found in antibodies. As crystal structures of other members of the Ig superfamily are determined, additional ways for pairs of antibody domains to interact may be found. It will be especially interesting to see if the novel mode of association of the Ig-like domains in class I molecules is preserved in a class II MHC structure, in which there are two membrane-associated polypeptides ( $\alpha$  and  $\beta$ ), each having one Ig-like domain and one domain similar to either  $\alpha_1$  or  $\alpha_2$  of the class I heavy chains (61).

**CD8 INTERACTS WITH THE  $\alpha_3$  DOMAIN** The mature T lymphocytes that have receptors recognizing class I MHC molecules invariably express the CD8 glycoprotein (7). This molecule exists as a homodimer of  $\alpha$  chains, or a heterodimer of  $\alpha$  and  $\beta$  chains. The striking correlation between CD8 expression and class I MHC restriction of a T cell's receptor led to the



*Figure 3* Comparison of the interaction of the  $\alpha_3$  and  $\beta_{2m}$  domains of class I MHC molecules (3a) with that of two Ig constant domains (3b). Constant domains are the two CH3 domains from Fc (127). The two Ig domains in HLA-A2 are paired in an interaction not found relating pairs of antibody constant domains. Prepared with a program described in Ref. 126.

hypothesis that there is specific biological interaction between class I MHC molecules and CD8 (62). Many indirect observations provided support for such interaction, which has been more directly demonstrated with an *in vitro* cell-binding assay (63).

This assay also revealed that HLA-Aw68 is unlike other class I MHC molecules, including HLA-A2, in that it does not bind, or has a highly reduced affinity for, CD8 (64). *In vitro* mutagenesis showed that, of the 13 amino acid differences between HLA-A2 and HLA-Aw68, a single substitution of the  $\alpha_3$  domain at position 245 is solely responsible for the change in CD8 binding. The structures of the  $\alpha$  carbon backbones of A2 and Aw68 are very similar, and it is likely that the change of side chain—alanine in A2 to valine in Aw68—is responsible for the difference, and that this region of the  $\alpha_3$  domain is directly involved in the interaction with CD8 (48).

Independent evidence for this model comes from study of a mutant mouse class I MHC molecule that was selected for loss of a monoclonal antibody—defined epitope of the  $\alpha_3$  domain (65). In this mutant H-2D<sup>d</sup> molecule, glutamic acid at position 227 is replaced by lysine. Extensive analyses of the capacity of T cells to recognize this mutant molecule indicate that it has lost the capacity to interact with CD8 (66). Residue 227 is an exposed residue found on the loop between strands 3 and 4 of the Ig fold of  $\alpha_3$ , whereas residue 245 is on strand 5 of the fold and is only partially exposed. These two residues are only 10.5 Å apart and could easily form part of a single CD8-binding site. Comparison of other point mutations within the  $\alpha_3$  domain has identified further positions within the loop between strands 3 and 4 and in strand 4 that affect CD8 binding. The loop (residues 220 to 229) contains five acidic and no basic residues. Mutational analysis shows acidic groups are important for CD8 binding, suggesting that the negatively charged loop may be a central site of contact with CD8 (67).

It is of note that strands 4 and 5 of the  $\alpha_3$  domain comprise the major site of interaction with  $\beta_2$ -microglobulin (26). That substitutions within this region also affect CD8 binding suggests that  $\beta_2$ m may directly contribute to interaction with CD8, and raises the possibility that the unique juxtaposition of the Ig-like domains in class I MHC molecules is an adaptation to binding CD8.

### *Structure of the $\alpha_1$ and $\alpha_2$ Domains*

The  $\alpha_1$  and  $\alpha_2$  domains have been the focus of most interest, because this part of the HLA structure is responsible for peptide binding and MHC-restricted recognition by TcRs. Although there is only minimal, if any, sequence similarity between  $\alpha_1$  and  $\alpha_2$  (31), the two domains share a common tertiary structure and probably a common origin. Each is composed of a four-stranded antiparallel  $\beta$ -pleated sheet that forms the amino-terminal half of the domain, followed by a long  $\alpha$ -helical region (Figures 2c, 2d, and 4). When  $\alpha_1$  and  $\alpha_2$

are paired in the HLA structure, they are related by a nearly exact two-fold rotation axis, so that a single eight-stranded antiparallel  $\beta$ -pleated sheet is formed from the four-stranded sheets in each domain. The  $\alpha$ -helices then lie on top of the sheet running antiparallel to each other, their helical axes making an angle of approximately  $40^\circ$  with respect to the  $\beta$ -strands in the sheet.

The presence of approximate dyad symmetry relating two homologous domains of a single polypeptide chain has been observed in other proteins in which repetitive domains are suggested to have arisen by gene duplication, resulting in the formation of a so-called "intramolecular dimer" (68). These types of proteins are suggested to have evolved from two identical polypeptides that formed a dimer with exact two-fold symmetry. If a gene duplication occurred so that the two polypeptides were in series in a single polypeptide chain, the resulting intramolecular dimer would be an acceptable structure if the carboxyl end of one subunit was close to the amino end of the other in the original intermolecular dimer. The interface between domains in the intramolecular dimer should require few changes, since the two halves had evolved to fit together in the original dimer. Examples of proteins that may have evolved this way include ferredoxin, parvalbumin, and pepsin (68), as well as the  $\alpha_1$  and  $\alpha_2$  domains of class I MHC molecules. The N-terminal domains of class II MHC  $\alpha$  and  $\beta$  subunits are hypothesized to be structurally similar to class I  $\alpha_1$  and  $\alpha_2$  (69), which as a pseudo intermolecular dimer, may represent an intermediate in the evolution of the class I MHC peptide-binding region. It is also interesting to note that in almost all symmetric multidomain enzymes or binding proteins as well as in the MHC  $\alpha_1$  and  $\alpha_2$  domains, the active center is located at the interface between domains, and close to the symmetry axis relating them (68).

The helical regions of both the  $\alpha_1$  and  $\alpha_2$  domains are composed of two  $\alpha$ -helices. The first helix in  $\alpha_1$  is a short helix of  $\sim 1\frac{1}{2}$  turns (residues 50–55) that meets a longer helix of  $7\frac{1}{2}$  to 8 turns (residues 57–85) at an angle of approximately  $110^\circ$ . The  $\alpha_2$  domain's helical region also starts with a short helix (residues 138–148;  $\sim 3$  helical turns) that meets a longer helix (residues 150–175,  $\sim 7\frac{1}{2}$  helical turns) at an approximate angle of  $130^\circ$ . The short helices at the amino-termini of the helical regions in  $\alpha_1$  and  $\alpha_2$  form the means by which the floor of the domain (i.e. the  $\beta$ -pleated sheet) is linked with the top of the domain (the long helices). The long helices of  $\alpha_1$  and  $\alpha_2$  are respectively in length about 41 Å and 37 Å, and are located about 18 Å apart (as measured between the  $\alpha$ -carbon atoms of residues 67 and 160). The mean interaxial distance for two helices placed next to each other is 9.4 Å (70); thus, there is an obvious gap between the two MHC  $\alpha$ -helices, which has been hypothesized to be the site where antigenic peptides bind (27). The axes of the two long helices make an angle of approximately  $20^\circ$  with respect to each other, so that in both cases, the amino-terminal residues of these helices

project to the highest points of the molecule, and may therefore be the most accessible residues to a TcR recognizing the top surface of peptide-MHC complex. A view of the molecule that illustrates this point is shown in Figure 2*b*.

**LOCATION OF THE BINDING SITE FOR ANTIGENIC PEPTIDES** The large cleft between the  $\alpha$ -helices is the most likely location of the single peptide-binding site proposed on the basis of peptide competition experiments (21). In addition to the obvious fact that the cleft is on the top surface of the molecule where it would be easily accessible to TcRs, there are other supporting lines of evidence:

1. The HLA-A2 crystal structure shows an unknown molecule or mixture of molecules to be occupying this site, presumably representing peptide(s) that copurified and cocrystallized with the HLA protein. The fact that molecules in the site remained bound during the entire purification procedure is consistent with the slow kinetic off-rate observed for peptides bound to class II MHC molecules (71).
2. The site has a size and shape consistent with accommodation of peptides of the size known either to bind to MHC molecules, or to be recognized by T cells after incubation with target cells (5, 72). [A peptide of 12–20 residues in an  $\alpha$ -helical conformation could fit into the groove although, depending upon the amino acid side chains, the fit could be quite tight. Alternatively, the site could accommodate about eight residues of peptide in an extended conformation (27). Longer peptides or even extended loops of proteins could bind to the site if they protruded over one or both ends of the site.]
3. Much of the sequence variation between HLA-A,B,C molecules (Figure 4) is found in residues that line the sides and bottom of the groove (27). This finding correlates with and provides a molecular mechanism for the observed specificity in peptide binding to different allelic products. In particular, 18 of the 20 positions with high variability in amino acids have side chains that point into the groove and potentially alter peptide binding, and 10 of the 13 substitutions that distinguish HLA-A2 and HLA-A28 are in such positions.
4. Residues known to affect recognition of class I MHC molecules by T cell clones fall in the site in positions where they could make direct contacts with bound peptides or with a TcR recognizing the peptide-MHC complex (27).

The floor of the peptide-binding site is formed by residues on the first and second  $\beta$ -strands of the  $\alpha_1$  and  $\alpha_2$  domains whose side chains point “up”

(direction defined by Figure 2a), and its sides are formed by residues on the two  $\alpha$ -helices that point "in," toward the site. The residues on the floor and sides of the site with inward-pointing side chains have been classified as potentially interacting with peptides (downward-facing triangles in Figure 4), and candidates for direct recognition by TcRs if they are located on one of the two  $\alpha$ -helices and their side chains point "up" (upward-facing triangles in Figure 4). Residues with side chains in an intermediate position between "up" and "in" could interact with either peptides or TcRs (diamonds in Figure 4) (Ref. 27).

**IS THE MHC PEPTIDE-BINDING SITE ALWAYS OCCUPIED?** An important implication of the occupation of the binding site in the HLA-A2 crystal structure is that cell-surface MHC molecules may always be filled with endogenous peptides derived from "self"-proteins. The presentation of "self"-peptides in addition to "foreign" peptides had been postulated by various authors (73, 74). The studies of Townsend et al (60), suggesting that peptide binding is essential for correct folding of the heavy chain and transport to the cell surface, would support this hypothesis, as does the finding of a heterogeneous mixture of peptide material associated with purified class II MHC molecules that can be released by treatment at pH 2.5 (Ref. 28). It therefore seems likely that any preparation of purified class I or class II MHC molecules will contain bound peptides, and that in vitro peptide-binding studies have been largely measuring the binding of exogenous peptide only after displacement of endogenous material. This may explain the small fraction of MHC molecules—5–10% for purified class II MHC molecules (28) and 0.3% for class I MHC (29)—with sites accessible to an experimental peptide. The slow on-rate observed for peptide binding to class II MHC molecules (71) could also be due to the necessity for displacement of endogenous material before binding.

It is probably not necessary that endogenous peptide be displaced from every surface MHC molecule on a cell before T cell recognition occurs, but it is not known what percentage of surface MHC molecules must be occupied by a foreign peptide to trigger a T cell response.

*Presentation of self-peptides and immunological tolerance* A current view is that assembly and cell-surface expression of class I MHC molecules requires binding of peptides, a process that occurs in the endoplasmic reticulum (60). The peptides presumably derive from endogenous proteins through routine processes of intracellular degradation and turnover. Although little is known of the parameters determining which peptides are presented, they are likely to include the concentrations of peptides, the specificity of systems that transport peptides into the endoplasmic reticulum, and the residues in the combining

site of the class I heavy chains. In a healthy cell, the class I MHC molecules at the plasma membrane will display a selection of peptides derived from self-proteins that do not provoke an aggressive T cell response, because the immune system develops a state of nonresponsiveness or tolerance to self during ontogeny. When that same cell is infected with a virus, viral peptides will compete with endogenous peptides for binding to class I MHC molecules. When presented at the cell surface, they will be recognized as foreign and stimulate an aggressive T cell response (9).

In vivo, class II molecules do not bind the same spectrum of peptides as class I molecules (75). This difference is not just the result of differences in combining site specificity, but of distinct intracellular pathways for peptide binding (Figure 1). Although class II  $\alpha$  and  $\beta$  polypeptides are synthesized and assembled in the endoplasmic reticulum, they do not appear to bind the peptides they will present until reaching a site where peptides from exogenous, endocytosed proteins predominate over those from endogenous proteins. The mechanistic basis for this difference is not understood, but probably involves the invariant chain protein, which shows a specific but temporary association with class II molecules during their protracted transit from the endoplasmic reticulum to the plasma membrane (76). One possibility is that part of the invariant chain fills the groove of the class II molecule, acting as a keeper peptide, during passage from the endoplasmic reticulum to the peptide-binding compartment.

That self-peptides are bound by MHC molecules and brought to cell surfaces is indicated by studies of Lorentz & Allen (77, 78). They isolated T cell hybridomas from mice immunized with an allelic form of hemoglobin, and showed that self-hemoglobin peptides were presented by class II MHC-bearing cells in the strain of mice that was immunized. Thus, there appear to be no specific mechanisms for distinguishing self from nonself at the level of peptides binding to MHC molecules and being expressed at the cell surface.

Evidence points to the discrimination between self and nonself being exercised at the level of the T cell, with MHC and peptide presentation contributing to the selection and control of the TcR repertoire. T cells mature in the thymus gland; it is there that rearrangement and expression of TcR genes occurs and where a small fraction of the receptor-bearing cells are selected for differentiation into mature, circulating peripheral T cells (4, 79). In essence, selection is a matching of TcRs against the class I and II MHC molecules of the individual, and evidence has accumulated to invoke both positive and negative elements. Within the thymus, the TcRs of immature T cells are exposed to class I and II molecules complexed with peptides derived from endogenous proteins from thymic cells, and from proteins derived from other cells that are taken up by thymic cells (4, 79, 80). Potentially autoreactive cells with receptors that interact too strongly with any of these

complexes are eliminated, and such clonal deletion is a major contributor to self-tolerance (81–83). On the other hand, cells with a certain strength of interaction with a self-MHC molecule, such that are not autoreactive but would be stimulated by the complex of that MHC molecule with a foreign peptide, are positively selected (8, 84).

Negative selection in the thymus explains tolerance to all thymic proteins, which will include all housekeeping proteins, and to proteins such as those in the blood that can be taken up by the thymic cells. For example, Lorentz & Allen (78) found that thymic cortical epithelial cells presented the hemoglobin self-antigen in fetal thymus during the period of time that selection against self-reactive T cells is known to occur. There are, however, likely to be specialized self-proteins that are processed and presented in certain cells and tissues but not in the thymus. T cells with receptors specific for these MHC peptide complexes will leave the thymus, and studies with transgenic mice show that alternative mechanisms of peripheral tolerance exist to prevent them from turning autoreactive (85). These mechanisms do not appear to involve elimination but an inactivation of the T cells that requires continued presentation of the antigen and may, on removal of antigens, be reversed and lead to an autoimmune reaction.

Thus it is possible to explain both the selection and activation of T cells in terms of the presentation of self- and foreign peptides at critical points in the development of T cells. A consequence of this scheme is that tolerance will not be induced against self-peptides that are not bound by self-MHC molecules.

### *Size of Peptides That Fit in the Site*

From the size of the HLA-A2 peptide-binding site, it appears that for bound peptides longer than 20 residues, one or both ends would have to extend beyond the site. It is known that lysozyme (86), ribonuclease (87), and myoglobin (88) can bind to class II MHC molecules and be presented as unfolded, but intact polypeptide chains, suggesting that in these cases, the peptide epitope is binding with the rest of the polypeptide chain of the protein protruding out of one or both ends of the site. These observations were extended by a study in which a 34-residue lysozyme peptide was found to bind to a class II MHC molecule without further proteolytic processing (89). This peptide was synthesized with D-amino acid residues flanking the known T cell epitope region, thus ensuring that the extra residues would not be cleaved by any protease. The core residues involved in the T cell determinant were protease sensitive unless bound to the class II molecule, demonstrating that an MHC molecule can protect a determinant from proteolytic degradation, a finding that is consistent with the proposal that MHC molecules may protect antigens from a total degradation during antigen processing (10).

Human fibrinogen represents an example of a protein that does not require processing (i.e. conversion into a peptide form that associates with an MHC molecule), but will stimulate T cells when associated with class II MHC molecules as an intact, and presumably native, protein (90). The fibrinogen epitope that is recognized has been localized to a 27-amino-acid region near the C-terminus of one of the chains, and is hypothesized to be a conformationally flexible portion capable of inserting into the peptide-binding site.

Although we have primarily discussed the binding of peptide antigens to MHC molecules, it is possible that nonpeptide molecules could bind to and be presented by MHC molecules. The heme moiety of cytochrome *c* has recently been shown to be a dominant MHC-restricted T cell epitope (91). It has also been suggested that the binding of nonpeptide organic compounds to MHC molecules might play a role in the ability of mice or rats to distinguish differences in MHC haplotypes of potential mates. Class I molecules are excreted into urine, and rats can distinguish the smell of urine samples from donors that differ genetically only at the class I MHC locus (92). It has been proposed that different allelic class I products might associate with a unique mixture of volatile, endogenous metabolites, leading to the ability to distinguish MHC haplotype by olfaction (92).

## SEQUENCE DIVERSITY IN CLASS I HLA HEAVY CHAINS

Diversity in the class I MHC molecules expressed by an individual is formed by compounding factors. The first is that multiple, divergent heavy chain loci are codominantly expressed. Their number varies between species and it can also vary between individuals within a species (20, 93). Certain "classical" class I MHC loci are expressed on a majority of cells types: in mice they are the *H-2L*, *K*, and *D* loci, and in humans they are the *HLA-A,B,C* loci. Products of other "nonclassical" class I genes exhibit restricted tissue distributions and can be either secreted, retained in the cell, or expressed at the cell surface. Examples of "nonclassical" human class I genes are *HLA-E*, *F*, and *G* (94). With few exceptions (95), class I molecules that present antigens to T cells are products of classical class I genes, and physiological functions for the products of nonclassical class I genes have yet to be established. Even the functionality of classical class I loci can pose problems; for example, *HLA-C* shows a significantly reduced cell-surface expression compared to *HLA-A* and *B*, and its capacity to present antigens to human T cells is uncertain.

The considerable variation in the numbers of class I genes and of expressed class I genes provides evidence for evolutionary plasticity of the MHC and the continuing production of new antigen-presenting genes coupled with elimina-

tion or inactivation of older genes. The *HLA-AR* pseudogene provides an example of a product of the latter process (32, 96). *HLA-AR* is closely linked to *HLA-A*, and similarities to that gene suggest that *HLA-A* and *HLA-AR* were the products of an ancestral gene duplication. Although *HLA-AR* sequences are very similar to *HLA-A,B,C*, they have one or two nucleotide changes that ensure this gene does not encode functional antigen-presenting heavy chains. One point substitution changes the cysteine at position 164 that forms the disulfide bond of the  $\alpha_2$  domain to phenylalanine, and a single nucleotide deletion that produces a frameshift in the middle of exon 4 results in a non-Ig-like sequence for  $\alpha_3$  accompanied by premature termination (Figure 4a). *HLA-AR* amino acid sequences are consistent with the idea that *HLA-AR* proteins were used in the past for antigen presentation, but the gene was subsequently inactivated and became a pseudogene. Accompanying this absence of contemporary function is a limited polymorphism and sequence diversity, especially at positions that show high variability in *HLA-A, B, or C* molecules.

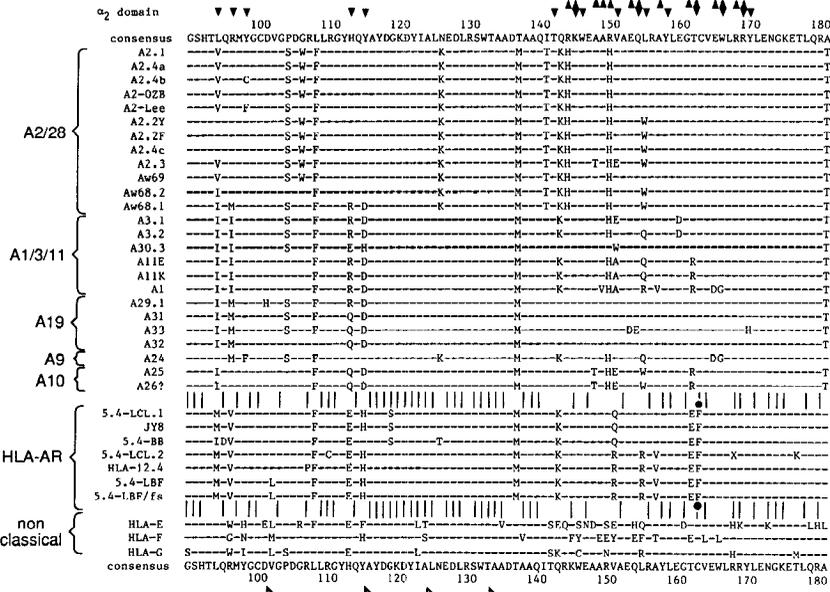
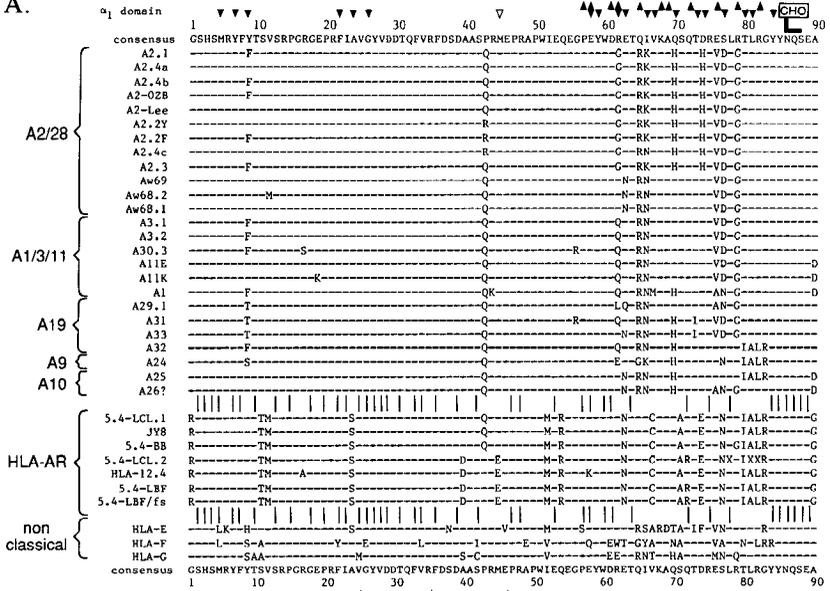
The second factor that creates diversity in the classical class I MHC is their

---

**Figure 4** Amino acid sequences of human class I HLA heavy chains. The consensus sequence is from Ref. 22 and identities with the consensus are shown by a dash. Downward-facing triangles designate residues postulated to interact with peptides, upward-facing triangles indicate residues postulated to interact with TcRs, and diamonds indicate residues that are postulated to interact with peptides and/or TcRs. These assignments are modifications of those described in Ref. 27, taking into account the possibility that TcRs may interact with residues that point "up" and away from the peptide-binding groove, in addition to those that point "up," or "up" and "in." The single invariant site of N-linked carbohydrate addition at position 86 is shown by CHO. Residues of  $\alpha_3$  that are implicated in binding CD8 from *in vitro* mutagenesis experiments are shown with an asterisk. The secondary structure is given below each domain sequence with  $\beta$  strands being represented by arrows, helices by squiggles, and loops by straight lines. (a)  $\alpha_1$  and  $\alpha_2$  domains of *HLA-A, AR, E, F, and G*. The *HLA-A* sequences are divided into five families as described (128). Vertical lines show residues that are conserved in all *HLA-A,B,C* sequences. With the exception of *HLA-12.4*, all *HLA-AR* alleles have a single nucleotide deletion in codon 227 (indicated by an X). For one sequence, 5.4-LBF, the consequences of the frameshift, including termination in codon 272, are shown on line 5.4-LBF/fs; otherwise the *HLA-AR* sequences are translated in the same reading frame as *HLA-A,B,C*. References for sequences are: *HLA-E,F,G* (94); *HLA-AR* (22; P. D. Ennis et al, manuscript submitted; 96); A29, A30, A31, A33 (128, 129); A11K (130); A2 subtypes (98, 131, 132). (b)  $\alpha_1$  and  $\alpha_2$  domains of *HLA-B*. References for sequences are: B37, B57 (22, 101; P. D. Ennis et al, manuscript submitted); B27 subtypes (132); B51(W) (133); B51, Bw52 (134); B13.1 (135); B35 (136); B38, B39 (137); BSNA, (M. Takiguchi, personal communication). The new and old designation for B27 subtypes are given (97). B? is an as-yet-unassigned molecule that is related to Bw58. (c)  $\alpha_3$  domains of *HLA-A,AR,E,F,G, and B*. (d)  $\alpha_1, \alpha_2,$  and  $\alpha_3$  sequences of *HLA-C* molecules. References for sequences are: Cw1, Cw2.1 (138); Cw2.2, Cw11 (101); Cw5, CwBL18 (139); Cw6 (D) (140); Cw6 (W), Cw7 (133); JY328 (141); C1.9, C1.10 (129); Cb-1, Cb-2 (142); BeWoC1 (143); Cx52 (144); pC1 (145).

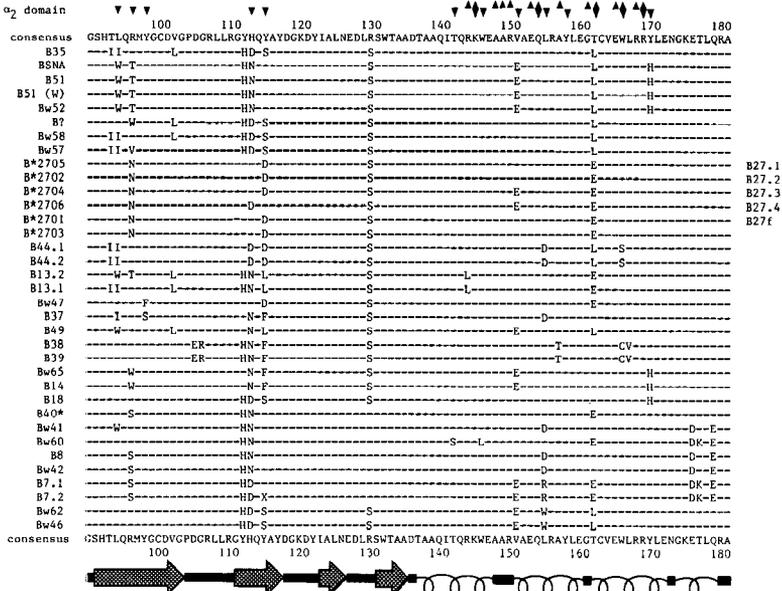
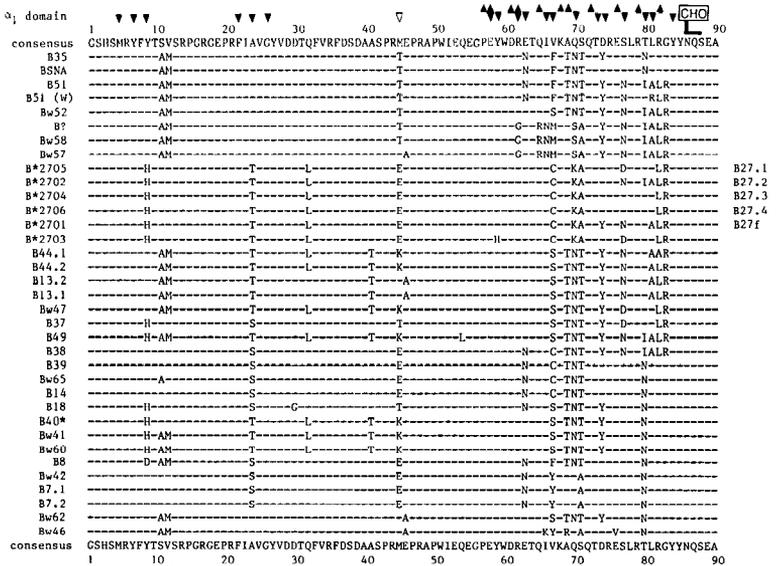
270 BJORKMAN & PARHAM

A.



Annual Rev. Biochem. 1990.59:253-288. Downloaded from arjournals.annualreviews.org by CALIFORNIA INSTITUTE OF TECHNOLOGY on 09/08/05. For personal use only.

B.



Annual Rev. Biochem. 1990.59:253-288. Downloaded from arjournals.annualreviews.org by CALIFORNIA INSTITUTE OF TECHNOLOGY on 09/08/05. For personal use only.





unprecedented genetic polymorphism. Currently 20, 40, and 10 antigens of the *HLA-A*, *B*, and *C* loci have been found that are expressed at significant frequency within human populations (97). However, these numbers clearly represent underestimates of the total, since novel alleles continue to be found as more precise and sensitive methods are applied to HLA typing. As far as we can tell, diversity in HLA-A,B,C molecules is due solely to amino acid substitutions in the MHC-encoded heavy chains, with no polymorphism being contributed by  $\beta_2m$ . There are different alleles of  $\beta_2m$  in mice, but they only differ by single amino acids from the common form. By contrast, the amino acid sequences of class I MHC heavy chains exhibit complex patterns of substitution involving many positions throughout the polypeptide chain (Figure 4).

HLA-A, B, and C heavy chains are homologous structures, showing only minor differences in the size of the transmembrane and cytoplasmic domains. Within the 274 residues of the  $\alpha_1$ ,  $\alpha_2$ , and  $\alpha_3$  domains, a total of 165 positions or 61% of the total are invariant between the 25 HLA-A, 35 HLA-B, and 18 HLA-C sequences that have been determined (Figure 4). These positions include the invariant glycosylation site at position 86 in  $\alpha_1$ , the cysteines at positions 101, 164, 203, and 259 that form the disulfide bonds of  $\alpha_2$  and  $\alpha_3$ , and many of the residues involved in contacting  $\beta_2m$  (Figure 4).

Nonclassical class I heavy chains encoded by the *HLA-E*, *F*, and *G* genes share many of these features. For example, consideration of the 165 positions that are invariant in all 76 HLA-A,B,C sequences shows that 155, 150, and 153 are preserved in HLA-E, F, and G respectively, and that the majority of differences involve conservative substitutions (Figure 4*a, c*). Such comparisons show that the HLA-E, F, and G gene products have only slightly diverged from antigen-presenting class I MHC molecules, and that any functions that these nonclassical class I molecules perform are likely to be related to the binding of peptides and/or interaction with the TcR complex.

Certain features in the classical class I sequences can be used to distinguish the products of the *HLA-A*, *B*, and *C* loci. At 31 positions in the amino acid sequence, products of a single locus are conserved, but differ from products of the other loci (one position in the leader peptide, two in  $\alpha_1$ , one in  $\alpha_2$ , four in  $\alpha_3$ , 13 in the transmembrane sequence, and nine in the cytoplasmic domains). On the basis of the substitutions at these positions, one can readily assign sequences to a particular locus. It is of interest that 22 of the 31 "locus-specific" positions are in the transmembrane and cytoplasmic domains. They may represent neutral substitutions acquired since gene duplication led to the creation of separate HLA-A, B, and C loci, or reflect as-yet-undiscovered functional differences in the carboxy-terminal regions of their products.

### *The $\alpha_3$ Domain*

Within the  $\alpha_1$ ,  $\alpha_2$ , and  $\alpha_3$  domains, products of a locus can differ by a single amino acid or by as many as 30 substitutions. Although the differences are predominantly found in the  $\alpha_1$  and  $\alpha_2$  domains, variation in the  $\alpha_3$  domain may be of importance, especially with respect to CD8 interactions (64, 66). HLA-B molecules are noticeably homogeneous in the  $\alpha_3$  domain, with 25 of the 34 sequences being identical (Figure 4c). Only three different substitutions are found, one shared by seven sequences and the other two being each represented in a single sequence. There is greater variation in HLA-A locus molecules, substitutions being found at nine positions and allelic products differing by up to seven amino acids. Two observations show functional effects due to natural substitutions in the  $\alpha_3$  domain: first, T cells can distinguish a substitution of alanine for glutamic acid at position 236 in the HLA-A2-0ZB subtype (98); second (as discussed above), an alanine to valine substitution at position 245 in HLA-Aw68.1 and HLA-Aw68.2 reduces the affinity for CD8 (64). HLA-C molecules show a pattern closer to HLA-A than to HLA-B in  $\alpha_3$ , variation occurring at 12 positions and sometimes involving nonconservative substitutions. A significant locus-specific difference in  $\alpha_3$  is the presence of arginine at position 239 in all HLA-B molecules and of glycine at this position in all HLA-A and C molecules (Figure 4c,d).

### *The $\alpha_1$ and $\alpha_2$ Domains*

From the first comparisons of allelic sequences (23, 99, 100), it was apparent that exons 2 and 3 encoding the  $\alpha_1$  and  $\alpha_2$  domains had a higher ratio of coding to noncoding substitutions than would be expected from random mutation. These observations argued for diversity being the result of positive natural selection, which has been confirmed and extended by a subsequent analysis that has made use of the three-dimensional structure (24). In this study, Hughes & Nei showed that the ratio of replacement to silent substitutions at positions involved in the antigen-combining site was significantly higher than at positions distant from the site. They concluded that class I MHC polymorphism is the result of natural selection and can be explained in terms of heterozygote advantage. While this idea may be criticized as overly simplistic (101), there is general agreement that MHC polymorphism results from positive selection due presumably to advantage from making T cell responses against particular pathogens. A new ability to respond to a particular antigen could derive from improvement either in presentation of peptides or in thymic selection of appropriate TcRs.

Analysis of diversity in MHC molecules has been aided by application of the parameter of variability originally used by Wu & Kabat (102) in comparison of Ig sequences. Variability at each position in a set of homologous

sequences is defined as the number of different amino acids found at a particular position divided by the frequency of the most common. Calculation of variability for all positions in the  $\alpha_1$ ,  $\alpha_2$ , and  $\alpha_3$  domains of HLA-A, B, and C molecules shows that positions of substitution can be divided into two groups having variability that is low ( $<4.0$ ) and high ( $\geq 4.0$ ) (Table 1). The division at a variability of 4.0 is somewhat arbitrarily set: it is the value obtained if two amino acids are present at equal frequency and is also the maximum variability found amongst the products of six alleles of HLA-AR, an inactivated gene for which T cell-related selection for diversity no longer exists and that can serve as a negative control.

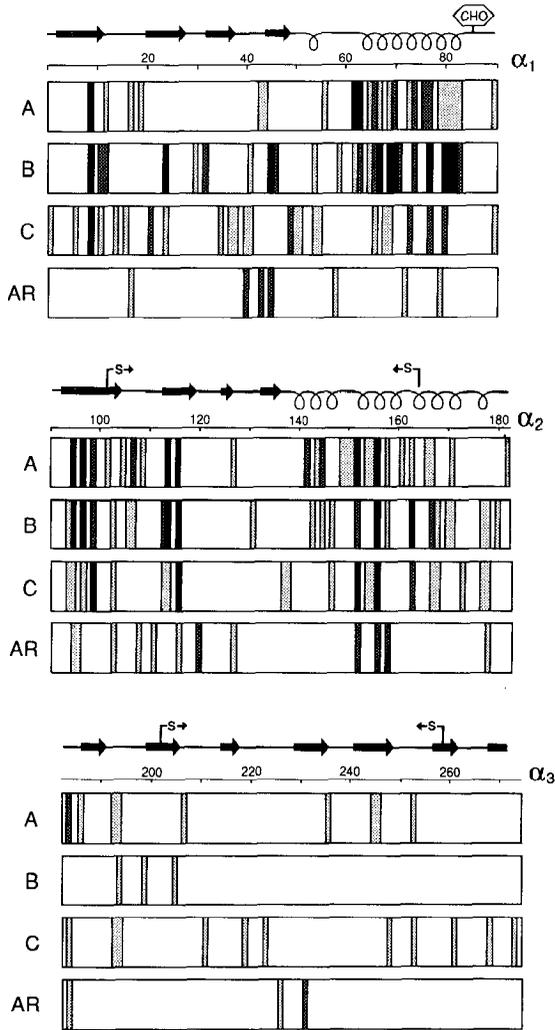
Positions of high variability are found in the sequences of all three loci; however, their number and distribution are very distinct (Table 1, Figure 5). Nine, fifteen, and five positions of high variability are found respectively for

**Table 1** Positions with high variability ( $> 4.0$ ) in HLA-A,B,C molecules

Position	Variability <sup>a</sup>			Potential contact <sup>b</sup>
	HLA-A	HLA-B	HLA-C	
$\alpha_1$				
9	9.1	4.8	10.0	Peptide
24		7.0		Peptide
45		10.0		Peptide
62	13.9			TcR + Peptide
63	4.2			Peptide
67		13.5		Peptide
69		4.8		TcR
70		6.4		Peptide
77		6.2		Peptide
80		7.0		Peptide
81		5.0		Peptide
82		5.5		TcR
$\alpha_2$				
95	5.4	5.0		Peptide
97	5.8	14.0		Peptide
99			5.2	Peptide
114	8.3	6.2		Peptide
116	6.3	15.9	10.2	Peptide
152	5.9		5.5	Peptide
156	8.3	5.6	10.3	Peptide
163		8.1		TcR + Peptide

<sup>a</sup> Variability defined according to Wu & Kabat (102) index.

<sup>b</sup> See Figure 4 and text for details. The possibility that residue 45 contacts a bound peptide is discussed in the text.



**Figure 5** Comparison of variability ( $V$ ) in amino acid sequence in the extracellular domains of human class I MHC heavy chains. The variability parameter as defined by Wu & Kabat (102) was calculated for 25 HLA-A, 35 HLA-B, 18 HLA-C, and 6 HLA-AR sequences; calculations were made within each group of sequences. Positions with "high variability" ( $V \geq 4.0$ ) are shown in solid black; positions with intermediate variability ( $3.0 \leq V < 4.0$ ) are shown in heavy stipple; positions with low variability ( $1.0 < V < 3.0$ ) are shown in light stipple, and positions with no variability ( $V = 1$ ) are shown in white. The predicted protein sequences of the HLA-AR alleles were analyzed ignoring the frameshift induced by a single nucleotide deletion. The secondary structure of each domain is indicated over the top of each diagram;  $\beta$  strands are indicated by arrows, helices by squiggles, and loops by straight lines. The cysteines involved in the conserved disulfide bond formation are given by S and the conserved glycosylation site by CHO.

HLA-A, B, and C with only three positions—9, 116, and 156—being held in common. The distribution of positions of high variability in HLA-C molecules is more similar to HLA-A, though in other more conserved parts of the gene, HLA-C sequences are more closely related to HLA-B.

Positions of high variability are those that exhibit the greatest number of amino acids (three to six), and they tend to have more radical substitution than positions of low variability, which commonly have just two chemically similar amino acids. A striking correlation is found between these positions of high variability and residues proposed to contact peptides and/or T cells on the basis of the three-dimensional structure (Ref. 27; see Figure 4a and 4b and Table 1). Of the 20 positions of high variability, 16 can be classified as peptide-interacting, two fall into the TcR-interacting category, and two are in intermediate positions where they could interact with peptides and/or TcRs. Position 45 has high variability, but is located on the fourth  $\beta$ -strand at the back of the  $\alpha_1$  helix, rather than on either the first and second  $\beta$ -strands or the  $\alpha$ -helices that form the peptide-binding site. However, this residue was suggested to be capable of influencing the site since its side chain pointed toward it (27). Further analysis shows that it forms the base of a hydrophobic cavity extending from the peptide-binding groove (48). Residues at positions 24, 26, 34, 45, and 67 are important in formation of this cavity. Of these, glycine 26 and valine 34 are invariant in HLA-A,B,C sequences. The other three positions exhibit variation that differs dramatically between products of the three loci. Only two combinations of amino acids are found in HLA-A and HLA-C molecules, whereas 13 distinct combinations exist in HLA-B molecules (Table 2). Thus there is both conservation and divergence of this pocket within the products of different class I loci.

Almost all of the peptide- or TcR-interacting residues within the peptide-binding site show some sequence variability. Notable exceptions are residues within the  $\alpha_1$  domain that form the floor of the site (methionine 5, phenylalanine 22, and glycine 26 are completely conserved, and tyrosine 7 and alanine 24 are conserved in all HLA-A molecules). By contrast, the residues on the  $\alpha_2$  domain side of the floor are some of the most highly variable within class I MHC sequences (Table 1). Significant differences in the diversity of the two helices are seen in comparison of the products of the three loci. The  $\alpha_1$  helix of HLA-C molecules is highly conserved, with nine peptide and TcR-binding positions that show polymorphism in HLA-A or -B being invariant. By contrast, HLA-B molecules have the greatest diversity in their  $\alpha_1$  helices, as evidenced by the presence of seven positions of high variability, compared to none in HLA-C and two in HLA-A (Table 1). In the  $\alpha_1$  helix, HLA-A molecules are intermediate between HLA-B and HLA-C. The situation in the  $\alpha_2$  helix is, however, different: the sequences of HLA-B molecules show less diversity than HLA-A molecules. For example, 11 of the 20 TcR- and peptide-binding positions of this helix are conserved amongst 35 HLA-B

**Table 2** Diversity in residues<sup>a</sup> of the 45 pocket

	Position in sequence					HLA subtype
	24	26	34	45	67	
A	G	V	M	V		all A except A1
A	G	V	M	M		A1
A	G	V	T	F		B35, SNA, 51
A	G	V	T	S		B52
A	G	V	T	M		B58
A	G	V	M	M		B57
T	G	V	E	C		B27
T	G	V	K	S		B44, 47, 40, 41
T	G	V	M	S		B13, 49
S	G	V	E	C		B16, 14
S	G	V	T	S		B18, 34
S	G	V	E	F		B8
S	G	V	E	Y		B7, 42
A	G	V	M	S		B62
A	G	V	M	Y		B46
A	G	V	G	Y		all other C
S	G	V	G	Y		Cw1, Cw6, JY328

<sup>a</sup> Amino acids indicated by single-letter code.

molecules, compared to 7 of the 20 for 25 HLA-A molecules. In particular, the amino-terminal part of the helical region, residues 139–155, is more conserved in HLA-B compared to HLA-A sequences. The pattern in the  $\alpha_2$  helix of HLA-C molecules is closer to that of HLA-B than HLA-A molecules. In combining the two  $\alpha$  helices, there are three distinctive patterns of diversity, providing further evidence for the functional divergence of HLA-A, B, and C gene products in their interactions with peptides and TcRs.

Several interesting features of MHC peptide-binding sites are revealed by an examination of the HLA-A2 structure and the sequences of other class I MHC molecules (Figure 4). First, there are many large aromatic residues in the binding site, most notably a number of tyrosine side chains in positions where they could potentially interact with bound peptides. Since many of the tyrosines are conserved, this is a general feature of class I peptide-binding grooves, but the significance of the tyrosine side chain for peptide binding is unknown. One could speculate that a tyrosine side chain is able to interact with many of the different types of side chains that would be found on the high number of peptides capable of binding to any given class I molecule, by virtue of its aromatic moiety that could be involved in hydrophobic interactions, and the polar hydroxyl group that could be involved in hydrogen bonding. A high percentage of tyrosine side chains has also been noted in many antibody-combining sites (102a).

Another feature conserved in class I peptide-binding sites is a salt bridge that links the N-terminal end of the  $\alpha_1$  helix to the C-terminal end of the  $\alpha_2$  helix effectively to hold together one end of the site. The residues involved in the salt bridge are glutamic acid 55 and arginine 170 (conserved in all human class I sequences with the exception of HLA-Cw3 and the nonclassical HLA-E and -G molecules), and its location on the HLA-A2 structure is highlighted in Figure 2c. The end of the site that contains the salt bridge also has many large aromatic residues which, together with the residues forming the salt bridge, effectively make this end of the site more occluded than the other end. The lack of space at this end of the site is a general feature of class I MHC peptide-binding sites, since the large residues contributing to it are mostly conserved (in HLA-A2 these residues are phenylalanine 22, tyrosine 7, and tyrosine 99 on the floor of the site, and tyrosine 59, tyrosine 159, and tryptophan 167 on the sides of the binding site). Although the helices at the other end of a class I site are not linked by a salt bridge, this end of the site is not open, owing to the presence of the conserved tyrosine at position 84, which points in to block this end of the site.

It is interesting to note that most of the sequence variability within the peptide binding site falls in positions where side chains point towards the site (potentially peptide interacting) rather than up into solution (potentially TcR interacting). In fact, an examination of the upward-pointing residues in Figure 4 (designated by upward pointing triangles or by diamonds) reveals that, for many different alleles of HLA molecules, all of the residues pointing up are the same. Thus, the MHC side chains contacted directly by a TcR recognizing the peptide-MHC complex would be identical in these cases. The means by which TcRs distinguish these MHC alleles must be through the indirect effects that different inward-pointing side chains have on the conformation of bound peptides and/or the upward pointing MHC residues.

## ANTIGENIC PEPTIDE MOTIFS AND THE CONFORMATION OF BOUND PEPTIDES

The conserved features of class I peptide-binding sites must play some role in determining what types of peptides will bind to class I MHC molecules, and their secondary structure conformation. It is not yet possible to predict accurately from a protein sequence which fragments will be antigenic or what conformation they would adopt when bound to an MHC molecule. Several predictive methods have been proposed for the identification of T cell epitopes within protein sequences. One method supposes that antigenic peptides form amphipathic  $\alpha$ -helical structures (103), and assumes that the nonpolar side of the helix contacts the MHC molecule and the polar side faces the solvent or contacts a TcR. However, an examination of the HLA-A2 structure clearly suggests that because of the depth of the binding site, a peptide will

contact the MHC molecule on three of its sides, leaving the fourth side exposed to solvent or in a position to interact with a T cell receptor (27). There is also no structural reason that peptides need to be amphipathic structures with a nonpolar side interacting with MHC residues, because the HLA-A2-binding site contains a number of polar and charged residues, as well as nonpolar amino acids. Examination of other class I sequences (Figure 4) and sequences of class II MHC molecules aligned to HLA-A2 (69), suggests that other MHC-binding sites are not particularly hydrophobic either. Another predictive method to locate which peptides can serve as T cell immunogens searches for a motif within the primary sequence of a protein (5). The motif that these authors found by analysis of 57 T cell epitopes is charged residue or glycine, followed by two or three hydrophobic residues, ending with a polar residue or glycine. Although this consensus sequence is quite degenerate, a number of peptides known to be T cell epitopes do not contain it (104), so it cannot account for all peptides capable of binding to MHC molecules.

Currently, there are no physical-chemical data regarding the secondary structures of peptides bound to MHC molecules. Studies to address this question have been done for peptides that interact with class II molecules. These studies involved assaying the ability of T cells to recognize target cells pulsed with an altered peptide, and determining the binding ability of the altered peptide using an *in vitro* peptide-binding assay. Based on these types of experiments, an  $\alpha$ -helical conformation was suggested for one peptide bound to a class II molecule (105), and an extended strand conformation was suggested for another (106). A study of the HLA-A2 site does not immediately suggest what conformation a specific peptide would adopt when bound to a class I molecule. However, using computer graphics to dock peptides into the HLA-A2 site suggests that a peptide in a standard  $\alpha$ -helical conformation is a very tight fit in the binding site, and thus minor movements of side chains or an even opening up of the site by displacing the  $\alpha$ -helices might be necessary before binding or releasing a peptide in a helical conformation. Such a movement need not result in a large structural change, but could be caused by a flexing or straightening of the  $\alpha_1$ - $\alpha_2$   $\beta$ -sheet (which has the standard right-handed twist) resulting in a very slight moving of the helices. The movement of the MHC helices would have to be slight since the  $\alpha_2$  helix is tethered to the  $\beta$ -sheet floor of the site by a disulfide bond. Another change that might facilitate binding or release of a peptide is breaking the previously discussed salt bridge linking positions 55 and 170.

## WHAT IS SEEN BY A TcR?

Although analysis of the specificity of T cells clearly demonstrated they were MHC restricted, i.e. they had a dual specificity for both antigen and an MHC

molecule (13, 14), the mechanism for this dual recognition remained unclear. Combining knowledge of the molecular nature of TcRs gained through biochemical and later molecular cloning studies (107–110) with the structure of MHC molecules, it is possible to make some predictions about the nature of the interaction between TcRs and the MHC-peptide complex.

A simple molecular interpretation of MHC-restricted recognition by T cells would be that TcRs recognize MHC residues as well as peptide residues in the MHC-peptide complex. Alternatively, it has been proposed that MHC-restricted TcRs recognize determinants only on the peptide antigen, and that the apparent dependence upon the correct MHC haplotype for T cell activation is a consequence of cells with different MHC genotypes selecting, protecting, and presenting a different set of peptides from any given antigenic molecule (74, 111). However, since the top surface of a peptide-MHC complex would probably be relatively flat, in the sense that the side chains from the peptide and the side chains on the MHC helices would protrude to approximately the same height (Figure 6), it seems likely that a TcR that contacts peptide will inevitably interact with residues on the MHC molecule. (Of course, the top surface of a peptide-MHC complex would not be completely flat, since the HLA-A2 helices show some irregularities such as a kink in the  $\alpha_2$  helix at residue 163, and the helices are not perfectly parallel to each other.) Nevertheless, it would appear that a TcR recognizing the top surface of the complex cannot avoid making direct contact with side chain atoms from the MHC molecule as well as the peptide.

### *A Model for How TcRs Bind to the Peptide-MHC Complex*

Models proposing how TcRs might bind to their ligands (2, 112, 113) have used what is known structurally about antibodies to provide a first order model for TcR structure. Amino acid sequence comparisons suggest that TcR variable (V) regions are folded into  $\beta$ -sandwich structures resembling Ig V regions (109, 110, 114). Antibody V domains have a conserved mode of interaction so that the hypervariable loops forming the antigen-binding site are found in the same relative positions with respect to each other (115). A study of Ig sequences and structures has identified conserved amino acids that are critical for maintaining the V region's contact surface, and most of these amino acids are found in analogous positions in TcR V region sequences (112, 116). It is therefore likely that TcR V domains will not only fold into tertiary structures similar to antibody V regions, but that the chain pairing and resulting combining sites of TcRs will also be similar to those described for antibodies (112). From studies of the crystal structures of protein antigen-antibody complexes (117–119), it has been learned that the antibody-ligand interface is an area of about  $25 \times 30 \text{ \AA}$ . The antigenic determinant is made up of amino acid main chain and side chain atoms from 16–17 residues that are

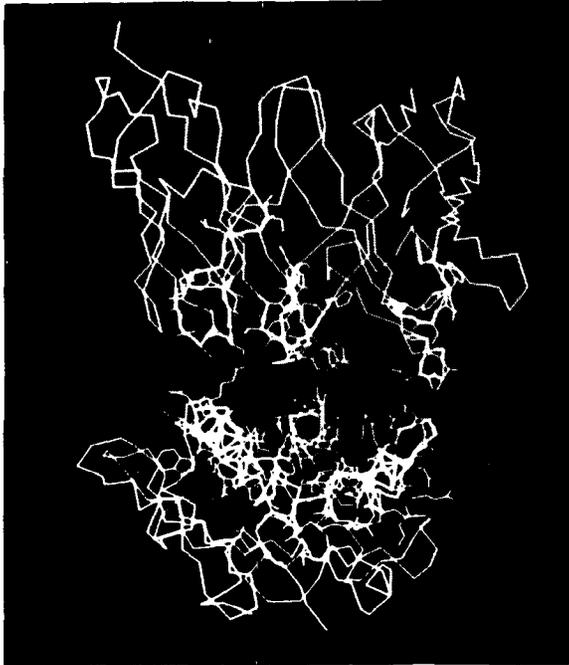
not linear in the primary sequence, but brought together by the tertiary structure of the protein antigen. No major conformational changes take place in either the antibody or the antigen after complex formation (117–119). These same principles have been proposed to apply to TcR recognition of peptide-MHC complexes (120).

Although variability in TcR regions is less localized than in Ig V regions, hypervariability is found in the locations corresponding to the three classic Ig hypervariable regions of complementarity-determining regions (CDRs) (121). TcR chains, like those of Igs, are assembled by the relatively random joining of different coding segments (V, D, and J) to constant region genes. The first and second CDR are encoded within the V gene segment itself, and the third CDR is formed by the junction of the V gene segment with smaller gene segments such as the D (diversity) and J (joining) gene segments. A comparison of the pattern of diversity in antibodies and TcRs shows that the receptors have much less variability in their first and second CDRs than antibodies, primarily because there are far fewer TcR V gene segments than Ig V gene segments. By contrast, considerably more diversity is concentrated in the TcR CDR3 or junctional region as compared to Igs (discussed in more detail in Ref. 2 and in M. M. Davis, this volume). In order to rationalize the striking concentration of TcR diversity within the junctional region and relative lack of diversity elsewhere, it should be relevant to consider what TcRs are known to do: namely to recognize a large number of small molecules (i.e. peptides) embedded in physically larger and much less diverse MHC molecules. The simplest interpretation of the skewing of diversity in the TcR case toward the CDR3-equivalent region is that the amino acids in this region are mainly interacting with peptide determinants, and that residues within the much less diverse CDR1- and CDR2-equivalent regions are primarily involved in contacts to MHC determinants.

In antibodies, the first and second CDRs of one V domain are separated from their counterparts on the other domain, and the space between them is occupied by the CDR3 regions from each chain (115). Structural studies of antibodies complexed with protein antigens have shown that side chains from all six CDRs contact the antigen, and that the entire interface is a rather flat surface with protrusions and depressions in the antibody being complementary to the antigen surface (117–119). When the (hypothetical) TcR-combining site is compared to the top surface of the MHC-peptide complex that is its ligand, it is interesting to note that the  $\alpha$ -helices that make up the sides of the MHC peptide-binding site are separated by about the same distance as separates the CDR1 and CDR2 regions of one V domain from another (18 Å). Thus the top surface of an MHC-peptide complex could interact with the combining site of an Ig-like TcR such that the limited diversity in the CDR1- and CDR2-equivalent regions on TcR  $V_\alpha$  and  $V_\beta$

contacts the side chains of the MHC  $\alpha$ -helices, leaving the centrally located and very diverse CDR3-equivalent regions to interact with the peptide (Figure 6). The alignment of CDR1 and CDR2 with MHC determinants and CDR3 with peptide determinants agrees with studies correlating TcR sequences with known antigen and MHC specificities (122–124), as well as site-directed mutagenesis experiments to test antigen fine specificity (125).

The use of a hypothetical model for describing the TcR-peptide-MHC trimolecular complex points out the obvious need for continued structural work in order to arrive at a complete description of the precise molecular nature of this interaction. Determination of the crystal structure of a human class I MHC molecule has allowed us a glimpse of how MHC molecules use their diversity to bind peptides and present them to T cells. However, the conformation that antigenic peptides adopt when bound to an MHC molecule



*Figure 6* Proposed mode of interaction of TcR (represented at top of figure) with peptide-MHC complex (bottom of figure). TcRs are proposed to bind to their ligands with the low diversity encoded within their CDR1 and 2 regions contacting side chains from the MHC  $\alpha$ -helices, and the high diversity encoded within their CDR3 regions contacting side chains from a bound peptide (here shown as a hypothetical 12-residue poly-valine  $\alpha$ -helix). FabJ539 (147) is used as a model for a TcR. Molecules are depicted as carbon- $\alpha$  backbones, with side chains included in the regions proposed to interact. Figure modified from Ref. 2 and prepared with the FRODO graphics program (146).

is still an open question, which can best be answered by the crystal structures of one or more specific peptide-MHC complexes. Our present knowledge would also be greatly enhanced by the determination of a structure of a TcR alone or as a complex with its peptide-MHC ligand. As methods improve for production of protein in the quantities required for crystallization, we can look forward to new crystal structures of and new insight into the molecules involved in the cellular immune response.

#### ACKNOWLEDGMENTS

PJB is a PEW Scholar and PP is a scholar of the Leukemia Society. The authors' research is supported by grant from USPHS (PP) and the Howard Hughes Medical Institute (PJB). We thank Patricia Massard for preparation of the manuscript.

#### Literature Cited

1. Hedrick, S. M. 1988. *Adv. Immunol.* 43:193-234
2. Davis, M. M., Bjorkman, P. J. 1988. *Nature* 334:395-402
3. Shimonkevitz, R., Kappler, J. W., Marrack, P., Grey, H. M. 1983. *J. Exp. Med.* 158:303-16
4. Marrack, P., Kappler, J. 1987. *Science* 238:1073-79
5. Rothbard, J. B., Taylor, W. R. 1988. *EMBO J.* 7:93-100
6. Allen, P. M. 1987. *Immunol. Today* 8: 270-73
7. Parnes, J. R. 1989. *Adv. Immunol.* 44: 265-311
8. Teh, H. S., Kisielow, P., Scott, B., Kishi, H., Uematsu, Y., et al. 1988. *Nature* 335:229-35
9. Townsend, A., Bodmer, H. 1989. *Annu. Rev. Immunol.* 7:601-24
10. Unanue, E. R., Allen, P. M. 1987. *Science* 236:551-57
11. Germain, R. N. 1986. *Nature* 322:687-89
12. Bevan, M. J. 1987. *Nature* 325:192-94
13. Rosenthal, A. S., Shevach, E. M. 1973. *J. Exp. Med.* 138:1194-212
14. Zinkernagel, R. M., Doherty, P. C. 1974. *Nature* 248:701-2
15. Matzinger, P. 1981. *Nature* 292:497-501
16. Dembic, Z., Haas, W., Weiss, S., McCubrey, J., Kiefer, H., et al. 1986. *Nature* 320:232-38
17. Townsend, A. R. M., Rothbard, J., Gotch, F. M., Behador, G., Wraith, D., et al. 1986. *Cell* 44:959-68
18. Babbitt, B. P., Allen, P. M., Matsuuda, G., Haber, E., Unanue, E. R. 1985. *Nature* 317:359-61
19. Snell, G. D. 1944. *Science* 100:272-73
20. Klein, J., Figueroa, F. 1986. *CRC Crit. Rev. Immunol.* 6:295-386
21. Guillet, J. G., Lai, M. Z., Briner, T. J., Smith, J. A., Gefter, M. L. 1986. *Science* 324:260-63
22. Parham, P., Lomen, C. E., Lawlor, D. A., Ways, J. P., Holmes, N., et al. 1988. *Proc. Natl. Acad. Sci. USA* 85: 4005-9
23. Jaulin, C., Perrin, A., Abastado, J.-P., Dumas, B., Papamatheakis, J., et al. 1985. *Immunogenetics* 22:453-70
24. Hughes, A. L., Nei, M. 1988. *Nature* 335:167-70
25. Schwartz, R. H. 1985. *Annu. Rev. Immunol.* 3:237-61
26. Bjorkman, P. J., Saper, M. A., Samraoui, B., Bennett, W. S., Strominger, J. L., et al. 1987. *Nature* 329:506-12
27. Bjorkman, P. J., Saper, M. A., Samraoui, B., Bennett, W. S., Strominger, J. L., et al. 1987. *Nature* 329:512-18
28. Buis, S., Sette, A., Colon, S. M., Grey, H. M. 1988. *Science* 242:1045-47
29. Chen, B. P., Parham, P. 1989. *Nature* 337:743-45
30. Cresswell, P., Turner, M. J., Strominger, J. L. 1973. *Proc. Natl. Acad. Sci. USA* 70:1503-7
31. Orr, H. T., Lopez de Castro, J. A., Lancet, D., Strominger, J. L. 1979. *Proc. Natl. Acad. Sci. USA* 76:5839-42
32. Malissen, M., Malissen, B., Jordan, B. R. 1982. *Proc. Natl. Acad. Sci. USA* 79:893-97
33. Grey, H. M., et al. 1973. *J. Exp. Med.* 138:1608-12
34. Peterson, P. A., Rask, L., Lindblom, J.

- B. 1974. *Proc. Natl. Acad. Sci. USA* 71:35-39
35. Tragardh, L., Rask, L., Wiman, K., Fohlman, J., Peterson, P. A. 1979. *Proc. Natl. Acad. Sci. USA* 76:5839-42
36. Peterson, P. A., Cunningham, B. A., Berggard, I., Edelman, G. M. 1972. *Proc. Natl. Acad. Sci. USA* 69:1697-701
37. Smithies, O., Poulik, M. D. 1972. *Science* 175:187-89
38. Williams, A. 1987. *Immunol. Today* 8:298-303
39. Hunkapiller, T., Hood, L. 1989. *Adv. Immunol.* 44:1-63
40. Michaelson, J. 1983. *Immunogenetics* 17:219-60
41. Turner, M. J., Cresswell, P., Parham, P., Strominger, J. L., Mann, D. L., et al. 1975. *J. Biol. Chem.* 250:4512-19
42. Parham, P., Alpert, B. N., Orr, H. T., Strominger, J. L. 1977. *J. Biol. Chem.* 252:7555-67
43. Springer, T. A., Strominger, J. L. 1976. *Proc. Natl. Acad. Sci. USA* 73:2481-85
44. Bjorkman, P. J. 1984. *Crystallographic studies of HLA*. PhD thesis. Harvard Univ.
45. Bjorkman, P. J., Strominger, J. L., Wiley, D. C. 1985. *J. Mol. Biol.* 186:205-10
46. Bjorkman, P. J., Bennett, W. S., Wiley, D. C. 1989. In *Symposium on Synchrotron Radiation in Biology (Proceedings)*. New York: Brookhaven Natl. Lab. In press
47. Maddon, D., Saper, M. A., Garrett, T. P. J., Bjorkman, P. J., Strominger, J. L., et al. 1989. In *Symposium on Immunological Recognition (Proceedings)*, pp. 353-59. Cold Spring Harbor, NY: Cold Spring Harbor Lab. Press
48. Garrett, T. P. J., Saper, M. A., Bjorkman, P. J., Strominger, J. L., Wiley, D. C. 1989. *Nature* 342:692-96
49. Holmes, N., Parham, P. 1985. *EMBO J.* 4:2849-54
50. Amzel, L. M., Poljak, R. J. 1979. *Annu. Rev. Biochem.* 48:961-77
51. Davies, D. R., Metzger, H. 1983. *Annu. Rev. Immunol.* 4:167-204
52. Becker, J. W., Reeke, G. N. 1985. *Proc. Natl. Acad. Sci. USA* 82:4225-29
53. Lancel, D., Parham, P., Strominger, J. L. 1979. *Proc. Natl. Acad. Sci. USA* 76:3844-48
54. Krangel, M. S., Biddison, W. E., Strominger, J. L. 1979. *J. Immunol.* 130:1856-62
55. Yokoyama, K., Geier, S., Uehara, H., Nathenson, S. G. 1985. *Biochemistry* 24:3002-10
56. Allen, H., Fraser, J., Flyer, D., Clavin, S., Flavell, R. 1986. *Proc. Natl. Acad. Sci. USA* 83:7447-51
57. Arce-Gomez, B., Jones, E. A., Barnstable, J., Solomon, E., Bodmer, W. F., 1978. *Tissue Antigens* 11:96-101
58. Seong, R. H., Clayberger, C. A., Krensky, A. M., Parnes, J. R. 1988. *J. Exp. Med.* 167:288-99
59. Hansen, T. H., Myer, N. B., Lee, D. R. 1988. *J. Immunol.* 140:3522-27
60. Townsend, A., Ohlen, C., Bastin, J., Ljunggren, H.-G., Foster, C., et al. 1989. *Nature* 340:443-48
61. Kaufman, J. F., Auffray, C., Korman, A. J., Shackelford, D. A., Strominger, J. L. 1984. *Cell* 36:1-13
62. Swain, S. L. 1983. *Immunol. Rev.* 74:129-42
63. Norment, A. M., Salter, R. D., Parham, P., Engelhard, V. H., Litman, D. R. 1988. *Nature* 336:79-81
64. Salter, R. D., Norment, A. M., Chen, B. P., Clayberger, C., Krensky, A. M., et al. 1989. *Nature* 338:345-47
65. Potter, T. A., Bluestone, J. A., Rajan, T. V. 1987. *J. Exp. Med.* 166:956-66
66. Connolly, J. M., Potter, T. A., Wormstall, E.-M., Hansen, T. H. 1988. *J. Exp. Med.* 168:325-41
67. Parham, P., Benjamin, R. J., Chen, B. P., Clayberger, C., Ennis, P. D., et al. 1989. See Ref. 47, pp. 529-43
68. McLachlan, A. D. 1980. In *Protein Folding*, ed. R. Jaenicke, pp. 79-99. North Holland: Elsevier
69. Brown, J. H., Jardetzky, T., Saper, M. A., Samraoui, B., Bjorkman, P. J., et al. 1988. *Nature* 332:845-50
70. Chothia, C. 1984. *Annu. Rev. Biochem.* 53:537-72
71. Buus, S., Sette, A., Colon, S. M., Jenis, D. M., Grey, H. M. 1986. *Cell* 47:1071-77
72. Livingstone, A. M., Fathman, C. G. 1987. *Annu. Rev. Immunol.* 5:477-501
73. Benacerraf, B. 1978. *J. Immunol.* 120:1809-12
74. Claverie, J.-M., Kourilsky, P. 1986. *Annu. Inst. Pasteur Immunol.* D137 (3):3-21
75. Morrison, L. A., Lukacher, A. E., Braciale, V. L., Fan, D. P., Braciale, T. J. 1986. *J. Exp. Med.* 164:1478-89
76. Stockinger, B., Pessara, U., Lin, R. H., Habicht, J., Grez, M., et al. 1989. *Cell* 56:683-89
77. Lorentz, R. G., Allen, P. M. 1988. *Proc. Natl. Acad. Sci. USA* 85:5220-23
78. Lorentz, R. G., Allen, P. M. 1989. *Nature* 337:560-62
79. Sprent, J., Lo, D., Gao, E.-K., Ron, Y. 1988. *Immunol. Rev.* 101:173-90
80. Zinkernagel, R. M., Callahan, G. N.,

- Althage, A., Cooper, S., Klein, P. A., et al. 1978. *J. Exp. Med.* 147:882-96
81. Kappler, J. W., Roehm, N., Marrack, P. 1987. *Cell* 49:273-80
82. Kappler, J. W., Staerz, U., White, J., Marrack, P. 1988. *Nature* 332:35-40
83. Pullen, A. M., Marrack, P., Kappler, J. W. 1988. *Nature* 335:796-801
84. Sha, W. C., Nelson, C. A., Newberry, R. D., Kranz, D. M., Russell, J. H., et al. 1988. *Nature* 336:73-76
85. Morahan, G., Allison, J., Miller, J. F. A. P. 1989. *Nature* 339:622-24
86. Allen, P. M., Unanue, E. R. 1984. *J. Immunol.* 132:1077-79
87. Lorentz, R. G., Tyler, A. N., Allen, P. M. 1988. *J. Immunol.* 141:4124-28
88. Streicher, H. Z., Berkower, I. J., Busch, M., Gurd, F. R. N., Berzofsky, J. A. 1984. *Proc. Natl. Acad. Sci. USA* 81:6831-35
89. Donermeyer, D. L., Allen, P. M. 1989. *J. Immunol.* 142:1063-68
90. Lee, P., Matsuuda, G. R., Allen, P. M. 1988. *J. Immunol.* 140:1063-68
91. Cooper, H. M., Corradin, G., Paterson, Y. 1988. *J. Exp. Med.* 168:1127-43
92. Singh, P. B., Brown, R. E., Roser, B. 1987. *Nature* 327:161-64
93. Lawlor, D. A., Zemmour, J., Ennis, P. D., Parham, P. 1990. *Annu. Rev. Immunol.* 8:In press
94. Orr, H. T. 1989. In *Immunobiology of HLA*, ed. B. Dupont, 2:33-40. New York: Springer-Verlag
95. Zidovik, D., Rotlic, M., McKune, K., Guerder, S., MacKay, C., et al. 1989. *Nature* 340:646-49
96. Zemmour, J., Koller, B. H., Ennis, P. D., Geraghty, D. E., Lawlor, D. A., et al. 1989. Submitted
97. Bodmer, W. F., Albert, E., Bodmer, J. G., Dupont, B., Mach, B., et al. 1987. In *Immunobiology of HLA*. Vol I: *Histocompatibility Testing*, ed. B. Dupont, pp. 72-79. New York: Springer-Verlag
98. Castaño, R., Ezquerra, A., Domenech, N., López de Castro, J. A. 1988. *Immunogenetics* 27:345-55
99. N'Guyen, C., Sodoyer, R., Trucy, J., Strachan, T., Jordan, B. R. 1985. *Immunogenetics* 21:479-89
100. Holmes, N., Ennis, P., Wan, A. M., Denney, D. W., Parham, P. 1987. *J. Immunol.* 139:936-41
101. Parham, P., Lawlor, D. A., Lomen, C. E., Ennis, P. D. 1989. *J. Immunol.* 142:3937-50
102. Wu, T. T., Kabat, E. A. 1970. *J. Exp. Med.* 132:211-50
- 102a. Padlan, E. A., Silverton, E. W., Sheriff, S., Cohen, G. H., Smith-Gill, S. J., et al. 1989. *Proc. Natl. Acad. Sci. USA* 86:5938-42
103. DeLisi, C., Berzofsky, J. A. 1985. *Proc. Natl. Acad. Sci. USA* 82:7048-52
104. Claverie, J.-M., Kourilsky, P., Langlade-Demoyden, P., Prochnicka-Chalufour, A., Dadaglio, G., et al. 1988. *Eur. J. Immunol.* 18:1547-53
105. Allen, P. M., Matsuuda, G. R., Evans, R. J., Dunbar, J. B., Marshall, G. R., et al. 1987. *Nature* 327:713-15
106. Sette, A., Buus, S., Colon, S., Smith, J. A., Miles, C. 1987. *Nature* 328:395-99
107. Allison, J. P., McIntyre, B. W., Bloch, D. 1982. *J. Immunol.* 129:2293-300
108. Haskins, K., Kubo, R., White, J., Pigeon, M., Kappler, J., et al. 1983. *J. Exp. Med.* 157:1149-69
109. Hedrick, S. M., Cohen, D. I., Nielsen, E. A., Davis, M. M. 1984. *Nature* 308:149-53
110. Yanagi, Y., Yoshikai, Y., Leggett, K., Clark, S. P., Aleksander, I., et al. 1984. *Nature* 308:145-49
111. Werdelin, O. 1986. *Immunol. Today* 8:80-84
112. Chothia, C., Boswell, D. R., Lesk, A. M. 1988. *EMBO J.* 7:3745-55
113. Claverie, J.-M., Prochnicka-Chalufour, A., Bougueleret, L. 1989. *Immunol. Today* 10:10-14
114. Chien, Y.-H., Becker, D., Lindsten, T., Okamura, M., Cohen, D., et al. 1984. *Nature* 312:31-35
115. Chothia, C., Novotny, J., Bruccoleri, R., Karplus, M. 1985. *J. Mol. Biol.* 186:651-63
116. Novotny, J., Tonegawa, S., Saito, H., Kranz, D. M., Eisen, H. N. 1986. *Proc. Natl. Acad. Sci. USA* 83:742-46
117. Amit, A. G., Mariuzza, R. A., Phillips, S. E. V., Poljak, R. J. 1986. *Science* 233:747-53
118. Colman, P. M., Laver, W. G., Varghese, J. N., Baker, A. T., Tulloch, P. A., et al. 1987. *Nature* 326:358-63
119. Sheriff, S., Silverton, E. W., Padlan, E. A., Cohen, G. H., Smith-Gill, S. J., et al. 1987. *Proc. Natl. Acad. Sci. USA* 84:8075-79
120. Poljak, R. J. 1987. *Ann. Inst. Pasteur/Immunol.* 138:175-80
121. Patten, P., Yokota, T., Rothbard, J., Chien, Y.-H., Arai, K.-I., et al. 1984. *Nature* 312:40-46
122. Fink, P. J., Matis, L. A., McElligott, D. L., Bookman, M., Hedrick, S. M. 1986. *Nature* 321:219-26
123. Winoto, A., Urban, J. L., Lan, N. C., Goverman, H., Hood, L., et al. 1986. *Nature* 324:679-82
124. Hedrick, S. M., Engel, I., McElligott,

- D. L., Fink, P. J., Hsu, M.-L., et al. 1988. *Science* 239:1541-44
125. Engel, I., Hedrick, S. M. 1988. *Cell* 54:473-84
126. Priestle, J. P. 1988. *J. Appl. Crystallogr.* 21:572-76
127. Deisenhofer, J. 1981. *Biochemistry* 20:2361-71
128. Kato, K., Trapani, J. A., Alloppenna, J., Dupont, B., Yang, S. Y. 1989. *J. Immunol.* 143:3371-78
129. Cianetti, L., Testa, U., Scotto, L., La Valle, R., Simeone, A., et al. 1989. *Immunogenetics* 29:30-91
130. Mayer, W. E., Jonker, M., Klein, D., Ivanyi, P., van Seventer, G., et al. 1988. *EMBO J.* 7:2765-74
131. Holmes, N., Epstein, H., Kennedy, L. J. 1989. *Immunogenetics* 29:112-16
132. López de Castro, J. A. 1989. *Immunol. Today* 10:239-46
133. Pohla, H., Kuon, W., Tabaczewski, P., Doerner, C., Weiss, E. H. 1989. *Immunogenetics* 29:297-307
134. Hayashi, H., Ennis, P. D., Ariga, H., Salter, R. D., Parham, P., et al. 1988. *J. Immunol.* 142:306-11
135. Kato, K., Dupont, B., Yang, S. Y. 1989. *Immunogenetics* 29:117-20
136. Ooba, T., Hayashi, H., Karaki, S., Tanabe, M., Kano, K., et al. 1989. *Immunogenetics*. In press
137. Muller, C., Weiss, E., Schmidt, W. 1989. *Immunogenetics* 30:200-7
138. Gussow, D., Rein, R. S., Meijer, I., de Hoog, W., Seemann, G. H. A., et al. 1987. *Immunogenetics* 25:313-22
139. Tibensky, D., DeMars, R., Holowachuk, E. W., Delovitch, T. L. 1989. *J. Immunol.* 143:348-55
140. Mizuno, S., Kang, S. H., Lee, H. W., Trapani, J. A., Dupont, B., et al. *Immunogenetics* 29:323-30
141. Srivastava, R., Duceman, B. W., Biro, P. A., Sood, A. K., Weissman, S. M. 1985. *Immunol. Rev.* 84:93-113
142. Takiguchi, M., Nishimura, I., Hayashi, H., Karaki, S., Kariyone, A., et al. 1989. *J. Immunol.* 143:1372-78
143. Ellis, S. A., Strachan, T., Palmer, M. S., McMichael, A. J. 1989. *J. Immunol.* 142:3281-85
144. Takata, H., Inoko, H., Ando, A., Hara-naka, M., Watanabe, B., et al. 1989. *Immunogenetics* 28:265-70
145. Szöts, H., Reithmüller, G., Weiss, E., Mco, T. 1986. *Proc. Natl. Acad. Sci. USA* 83:1428-32
146. Jones, T. A. 1985. *Methods Enzymol.* 115:157-71
147. Suh, S. W., Bhat, T. N., Navia, M. A., Cohen, G. H., Rao, D. N., et al. 1986. *Proteins: Struct. Funct. Genet.* 1:74-80



## CONTENTS

HOW TO SUCCEED IN RESEARCH WITHOUT BEING A GENIUS, <i>Oliver H. Lowry</i>	1
PYRUVOYL-DEPENDENT ENZYMES, <i>Paul D. van Poelje and Esmond E. Snell</i>	29
PHYTOCHELATINS, <i>Wilfried E. Rauser</i>	61
RECENT TOPICS IN PYRIDOXAL 5'-PHOSPHATE ENZYME STUDIES, <i>Hideyuki Hayashi, Hiroshi Wada, Tohru Yoshimura, Nobuyoshi Esaki, and Kenji Soda</i>	87
SELENIUM BIOCHEMISTRY, <i>Thressa C. Stadtman</i>	111
BIOCHEMISTRY OF ENDOTOXINS, <i>Christian R. H. Raetz</i>	129
OCCLUDED CATIONS IN ACTIVE TRANSPORT, <i>Ian M. Glynn and S. J. D. Karlish</i>	171
CHEMICAL NUCLEASES: NEW REAGENTS IN MOLECULAR BIOLOGY, <i>David S. Sigman and Chi-hong B. Chen</i>	207
CADHERINS: A MOLECULAR FAMILY IMPORTANT IN SELECTIVE CELL-CELL ADHESION, <i>Masatoshi Takeichi</i>	237
STRUCTURE, FUNCTION, AND DIVERSITY OF CLASS I MAJOR HISTOCOMPATIBILITY COMPLEX MOLECULES, <i>Pamela J. Bjorkman and Peter Parham</i>	253
DNA HELICASES, <i>Steven W. Matson and Kathleen A. Kaiser-Rogers</i>	289
THE MITOCHONDRIAL PROTEIN IMPORT APPARATUS, <i>Nikolaus Pfanner and Walter Neupert</i>	331
UNUSUAL COENZYMES OF METHANOGENESIS, <i>Anthony A. DiMarco, Thomas A. Bobik, and Ralph S. Wolfe</i>	355
PEPTIDES FROM FROG SKIN, <i>Charles L. Bevins and Michael A. Zasloff</i>	395
CLATHRIN AND ASSOCIATED ASSEMBLY AND DISASSEMBLY PROTEINS, <i>James H. Keen</i>	415
ANTIBODY-ANTIGEN COMPLEXES, <i>David R. Davies, Eduardo A. Padlan, and Steven Sheriff</i>	439

T CELL RECEPTOR GENE DIVERSITY AND SELECTION, <i>Mark M. Davis</i>	475
THE BACTERIAL PHOSPHOENOLPYRUVATE:GLYCOSE PHOSPHOTRANSFERASE SYSTEM, <i>Norman D. Meadow, Donna K. Fox, and Saul Roseman</i>	497
SELF-SPlicing OF GROUP I INTRONS, <i>Thomas R. Cech</i>	543
STRUCTURE AND FUNCTION OF CYTOCHROME <i>c</i> OXIDASE, <i>Roderick A. Capaldi</i>	569
TRANSITION-STATE ANALOGUES IN PROTEIN CRYSTALLOGRAPHY: PROBES OF THE STRUCTURAL SOURCE OF ENZYME CATALYSIS, <i>Elias Lolis and Gregory A. Petsko</i>	597
INTERMEDIATES IN THE FOLDING REACTIONS OF SMALL PROTEINS, <i>Peter S. Kim and Robert L. Baldwin</i>	631
REGULATION OF VACCINIA VIRUS TRANSCRIPTION, <i>Bernard Moss</i>	661
BIOCHEMICAL ASPECTS OF OBESITY, <i>Henry Lardy and Earl Shrago</i>	689
RNA POLYMERASE B (II) AND GENERAL TRANSCRIPTION FACTORS, <i>Michèle Sawadogo and André Sentenac</i>	711
SEQUENCE-DIRECTED CURVATURE OF DNA, <i>Paul J. Hagerman</i>	755
CYTOKINES: COORDINATORS OF IMMUNE AND INFLAMMATORY RESPONSES, <i>Ken-ichi Arai, Frank Lee, Atsushi Miyajima, S. Miyatake, Naoko Arai, and Takashi Yokota</i>	783
THE FAMILY OF COLLAGEN GENES, <i>Eero Vuorio and Benoit de Crombrughe</i>	837
DEFENSE-RELATED PROTEINS IN HIGHER PLANTS, <i>Dianna J. Bowles</i>	873
MOTOR PROTEINS OF CYTOPLASMIC MICROTUBULES, <i>Richard B. Vallee and Howard S. Shpetner</i>	909
DNA RECOGNITION BY PROTEINS WITH THE HELIX-TURN-HELIX MOTIF, <i>Stephen C. Harrison and Aneel K. Aggarwal</i>	933
CAMP-DEPENDENT PROTEIN KINASE: FRAMEWORK FOR A DIVERSE FAMILY OF REGULATORY ENZYMES, <i>Susan S. Taylor, Joseph A. Buechler, and Wes Yonemoto</i>	971
THE CLASSIFICATION AND ORIGINS OF PROTEIN FOLDING PATTERNS, <i>Cyrus Chothia and Alexei V. Finkelstein</i>	1007
INDEXES	
Author Index	1041
Subject Index	1101
Cumulative Index of Contributing Authors, Volumes 55–59	1123
Cumulative Index of Chapter Titles, Volumes 55–59	1126