

CD4: Structure and Interactions of an Immunoglobulin Superfamily Adhesion Molecule

STEPHEN C. HARRISON

Howard Hughes Medical Institute, Department of Biochemistry and Molecular Biology, Harvard University, Cambridge, Massachusetts 02138

Received March 19, 1993

CD4 is a glycoprotein present on the surface of certain T-lymphocytes. It was originally identified in the late 1970s as the target for a particular T-cell-specific monoclonal antibody.^{1,2} Subsequently, it has become clear that mature T-cells bear either CD4 or a different glycoprotein known as CD8 and that these two classes of lymphocytes interact with antigen presented by class II and class I MHC proteins, respectively.³ Thus, the presence of CD4 on a T-cell means that its antigen receptor (the "T-cell receptor" or "TCR") recognizes a peptide in the groove of a class II MHC molecule.⁴ It is believed that, on a mature T-cell, the TCR and CD4 jointly contact an antigen-bearing class II protein, forming a complex at the interface between the antigen-presenting cell and the T-cell^{5,6} (see Figure 1). Proper formation of this complex brings into proximity cytoplasmic elements associated with CD4 and with the TCR, and intracellular signals are generated that ultimately lead to T-cell activation.⁷

CD4 is also the HIV receptor.⁸⁻¹⁰ Its presence on a cell is sufficient for viral attachment, but entry (by fusion of viral and cellular membranes) requires additional, as yet unidentified cellular components.^{10,11} The surface glycoprotein of HIV is responsible for CD4 binding.¹² It is an oligomer—probably a tetramer¹³—of a species known as gp160, which is cleaved in mature virions into gp120 and gp41. The amino-terminal gp120 fragment carries the attachment function; the carboxy-terminal gp41 fragment carries the fusion and membrane-spanning functions.¹⁴

Domain Organization of CD4

The cloned cDNA for human,¹⁵ mouse,¹⁶ and rat¹⁷ CD4 revealed an amino acid sequence for the encoded protein with the organization characteristic of many cell-surface proteins (Figure 2a). A stretch of hydrophobic residues can be identified as a transmembrane segment, connecting a large extracellular portion with a short cytoplasmic tail. The extracellular part is composed of four regions with immunoglobulin (Ig) domain homology.¹⁷ Molecules containing such regions are said to belong to the "Ig superfamily".¹⁸ The amino-terminal domain of CD4 (D1) has many characteristics of an Ig variable (V) domain. The other three domains are less readily classified. The crystallographically determined structures¹⁹⁻²¹ show that D3 is also of the IgV design, despite its lack of a disulfide, and that D2

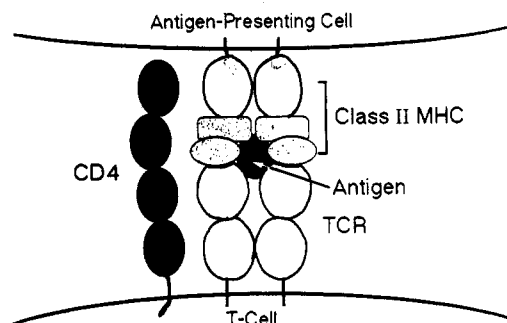


Figure 1. Diagram showing some of the molecules involved in the interaction between a T-cell and an antigen-presenting cell. The antigen, a peptide, is bound in a groove on the outer surface of a class II MHC glycoprotein. It is recognized by the T-cell receptor (TCR), together with its MHC surround. CD4 contacts one or more class II molecules, and it may also make lateral contacts with the TCR.

and D4 belong to a class variously designated H²² or C2.¹⁸ In human CD4, D3 and D4 each contain a glycosylation site. The cytoplasmic tail is probably too short to fold into a globular structure. Its function is to bind the Src-like tyrosine kinase known as p56^{lck} (Lck).²³ The interaction requires two cysteines in CD4 and two cysteines in the amino-terminal segment of Lck.²⁴ It has therefore been suggested that a metal ion,

- (1) Reinherz, E. L.; Kung, P. C.; Goldstein, G.; Schlossman, S. F. *Proc. Natl. Acad. Sci. U.S.A.* 1979, 76, 4061-4065.
- (2) Reinherz, E. L.; Kung, P. C.; Goldstein, G.; Schlossman, S. F. *J. Immunol.* 1979, 234, 2894-2896.
- (3) Reviewed in: Parnes, J. R. *Adv. Immunol.* 1989, 44, 265-311.
- (4) Reviewed in: Janeway, C. A., Jr.; Rojo, J.; Saizawa, K.; Dianzani, U.; Portoles, P.; Tite, J.; Haque, S.; Jones, B. *Immunol. Rev.* 1989, 109, 77-92.
- (5) Doyle, C.; Strominger, J. L. *Nature* 1987, 330, 256-289.
- (6) Anderson, P.; Blue, M.-L.; Schlossman, S. *J. Immunol.* 1988, 140, 1732-1734.
- (7) Reviewed in: Klausner, R. D.; Samuelson, L. E. *Cell* 1991, 64, 875-878.
- (8) Sattentau, Q.; Dalgleish, A.; Weiss, R.; Beverley, P. C. L. *Science* 1986, 234, 1120-1123.
- (9) Dalgleish, A., et al. *Nature* 1984, 312, 763-766.
- (10) Sodroski, J.; Goh, W. C.; Rosen, C.; Campbell, K.; Haseltine, W. A. *Nature* 1986, 322, 470-474.
- (11) Ashorn, P. A., et al. *J. Virol.* 1990, 64, 2149-2156.
- (12) McDougal, J.; Kennedy, M.; Sligh, J.; Cort, S.; Mawle, A.; Nicholson, J. *Science* 1986, 231, 382-385.
- (13) Schwallier, M., et al. *Virology* 1989, 172, 367-369.
- (14) Leis, J. et al. *J. Virol.* 1988, 62, 1808-1809.
- (15) Maddon, P. et al. *Cell* 1985, 42, 93-104.
- (16) Tourville, B., et al. *Science* 1986, 234, 610-614.
- (17) Clark, S. J., et al. *Proc. Natl. Acad. Sci. U.S.A.* 1987, 84, 1649-1653.
- (18) Williams, A. F., et al. *Cold Spring Harbor Symp. Quant. Biol.* 1989, 54, 637-647.
- (19) Wang, J., et al. *Nature* 1990, 348, 411-418.
- (20) Ryu, S., et al. *Nature* 1990, 348, 419-426.
- (21) Brady, R. L., et al. *Science* 1993, 260, 979-982.
- (22) Hunkapiller, T.; Hood, L. *Adv. Immunol.* 1989, 44, 1-63.
- (23) Shaw, A. S., et al. *Cell* 1989, 59, 627-636.

Professor Harrison received his A.B. in chemistry and physics from Harvard College and his Ph.D. in Biophysics from Harvard University and is currently a Professor of Biochemistry and Molecular Biology and an Investigator in the Howard Hughes Medical Institute at Harvard University. He is also a Research Associate in Medicine at the Children's Hospital, Boston.

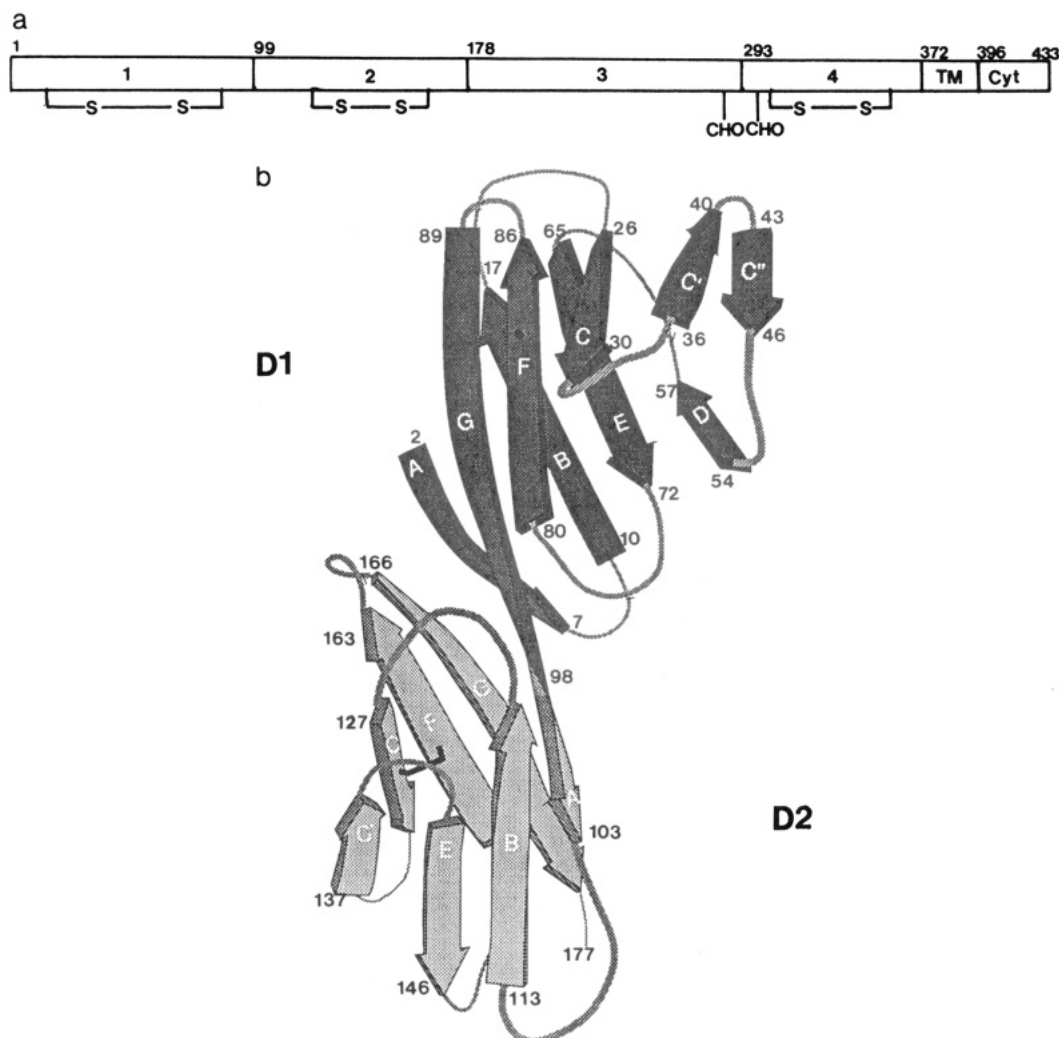


Figure 2. (a) Amino acid sequence organization of the human CD4 molecule.¹⁵ The four Ig-like domains are numbered; the disulfide bonds are shown; the glycosylation sites are marked. CHO: glycosylation site. TM: transmembrane domain. Cyt: cytoplasmic tail. (b) Ribbon diagram, showing organization of CD4 D1D2. D1 is more darkly hatched than D2. Residue numbers at the beginning and end of each secondary structure element are shown.

such as Zn^{2+} , could link the two proteins, but there is yet no direct evidence for such a ligand.

X-ray crystallographic studies of CD4 fragments containing domains 1 and 2 (D1D2)^{19,20} and domains 3 and 4 (D3D4)²¹ have revealed their structures. Crystals of D1–D4 are poorly ordered,²⁵ and the exact relationship between the two halves of the extracellular moiety is still not known. D1 contains the surface recognized by the HIV gp120, and mutational analysis based on the structure leads to a very precise definition of the amino acid residues involved. A similar sort of analysis shows that the class II MHC contact includes a large surface on D1 as well as a significant part of D2. Possible lateral contacts between D3 or D4 and the TCR have not yet been defined.

Structure of D1D2

The first two domains (Figure 2b) form a tightly linked, rod-like unit.^{19,20} D1 closely resembles an Ig V κ module. It has nine strands in two β -sheets, one with strands AGFCC'C'' and the other with strands BED. The residues at the core of the domain correspond

precisely to these found in Ig structures: a pair of disulfide-bridged cysteines, a tryptophan packed against them, and several other conserved hydrophobic residues. There are in addition some specific hallmarks of light-chain variable domains:²² a glycine in the AB loop, an arginine–aspartic acid salt bridge linking the beginning of strand D with the EF loop, and a buried tyrosine in the F strand hydrogen bonded to the carbonyl of the same aspartic acid. D1 differs from an IgV domain in two key ways: it lacks the features responsible for dimerization of variable modules in an Ig molecule,²⁶ and it contains elaborations of the C'C'' and DE loops. The absence of the typical IgV dimerization structures (an extended CC' loop and a bulge in the G strand) is consistent with the observation that soluble CD4 is monomeric.^{25,27} The elaborations of the C'C'' and DE loops probably have a role in binding to class II MHC molecules²⁸—the specific function of CD4. The two loops are spatially adjacent. The DE loop contains about two turns of irregular helix in the segment from 58 to residue 64. The side chain of Trp 62, in the middle of this segment, lies beneath the C'C''

(24) Turner, J. M., et al. *Cell* 1990, 60, 755–765.

(25) Kwong, P. D., et al. *Proc. Natl. Acad. Sci. U.S.A.* 1990, 87, 6423–6427.

(26) Colman, P. M. *Adv. Immunol.* 1988, 43, 99–132.

(27) Kirchhausen, T.; Harrison, S. C. In preparation.

(28) Moebius et al. *Proc. Natl. Acad. Sci. U.S.A.* 1992, 89, 2008–2102.

turn, making it project away from the body of the domain. The C'C'' loop, which corresponds to the CDR2 of light-chain IgV domains, is also longer than in most antibody molecules. Thus, the C'C'' edge of the domain has a curious salience, with the side chain of Phe 43 projecting directly outward.

D2 resembles a truncated IgV module, in which strands C' and D have been replaced by an abrupt connection between C' and E. Strands GFCC' lie in one sheet; ABE, in the other. The disulfide bond is in a "nonstandard" location. It connects strands C and F within a sheet, rather than B and F across the core of the domain. The C-strand cysteine occupies the sequence position usually held by the conserved tryptophan, and the two bridged cysteines are therefore in register on adjacent strands, with their main-chain hydrogen bonds directed away from each other. The resulting disulfide geometry is likely to be slightly strained.²⁹ The packing of the two sheets upon each other also deviates from the conserved relationship seen in the Ig molecules.³⁰ Nevertheless, the structures of the AB and EF loops closely resemble those of corresponding IgV segments.

The two domains are linked by a continuous β -strand. That is, the G-strand D1 proceeds without interruption into the A-strand D2. The amino acid sequence contains six consecutive nonpolar residues, and their side chains point alternately into the cores of the two domains. There is substantial buried surface (550 Å²) at the domain interface, stabilizing the rod-like configuration of the fragment. Large-scale hinge motion at this "lap joint" is unlikely.

Structure of D3D4

The structure of a rat D3D4 fragment, expressed as a recombinant protein in CHO cells, has recently been determined.²¹ The two domains form a unit joined by a continuous β -strand, just like D1D2. Indeed, the relationship between D3 and D4 closely resembles the D1/D2 contact. D3 has an IgV-like fold, but the two sheets are significantly farther apart. There is no disulfide, and the residues that would correspond to the two cysteines are a phenylalanine and a leucine. The tryptophanyl side chain shifts in the C strand to fill the gap left by the presence of Gly, rather than Tyr or Trp, in the F strand. Thus, the fundamental IgV design is maintained, despite lack of conservation in the region sometimes called the "pin" of the domain.³⁰ D4 resembles D2, with the same seven-strand topology, but the AB loop is substantially truncated, giving the membrane proximal tip of the domain a more compact character. The disulfide is in the "normal" location.

Structure of the Complete Extracellular Fragment (D1-D4)

Electron microscopy²⁷ and hydrodynamic measurements²⁵ show that the D1-D4 fragment—sometimes referred to a sCD4 ("soluble CD4")—is monomeric and rod-like. An end-to-end packing of the D1D2 and D3D4 structures can therefore be modeled. There is a sequence of six nonpolar residues at the D2/D3 boundary, very similar to the one joining D1 and D2 and D3

(29) Richardson, J.; Richardson, D. C. In *Prediction of Protein Structure and the Principles of Protein Conformation*; Fasman, G., Ed.; Plenum: New York, 1989.

(30) Lesk, A. M.; Chothia, C. *J. Mol. Biol.* 1982, 160, 325.

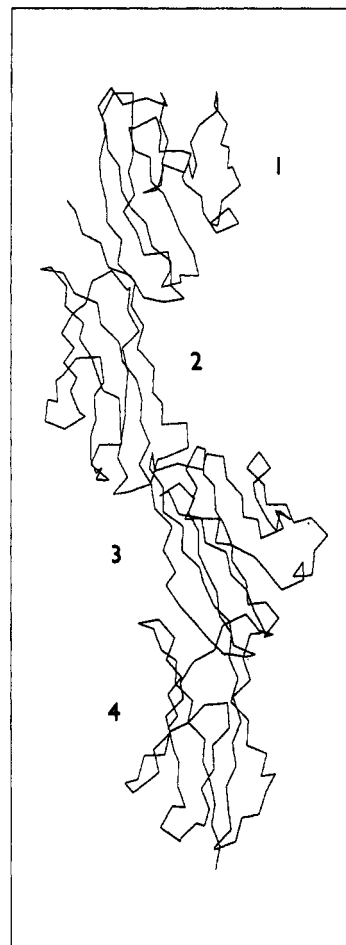


Figure 3. Carbon- α trace of a model for CD4, D1-D4, based on the assumptions outlined in the text. D3D4 has been represented by D1D2, because coordinates for the latter are not yet available. The correct orientation of D4 is slightly different, and the various loops of D3D4 are actually somewhat differently configured than the corresponding loops of D1D2.¹⁹⁻²¹

and D4.³¹ Other characteristic features of the D1/D2 interaction appear to be present at the D2/D3 contact as well. We have proposed that the interface between D2 and D3 resembles the one between D1 and D2.^{31,32} The diagram in Figure 3 shows the α trace of a model based on this idea. The modeled four-domain structure has a zigzag appearance. It is about 120 Å long, in agreement with measurements from electron microscopy.²⁷

CD4 as HIV Receptor

The surface on CD4 for binding HIV gp120 has been mapped, by studying the effects of CD4 mutations on gp120 affinity.³³⁻³⁶ The association is a tight one ($K^d \sim 1-10$ nM), and it is essentially irreversible at low temperature. Results from a number of laboratories are summarized in Figure 4. The C'C'' ridge is at the center of the site, with Phe 43 in the C'C'' turn having a critical role. There are graded effects of different

(31) Garrett et al. *J. Mol. Biol.* In press.

(32) Harrison et al. *Cold Spring Harbor Symp. Quant. Biol.* 1992, 57, 541-548.

(33) Peterson, A.; Seed, B. *Cell* 1988, 54, 65-72.

(34) Sattentau, Q., et al. *J. Exp. Med.* 1989, 170, 1319-1334.

(35) Brodsky, M. H.; Warton, M.; Myers, R. M.; Littman, D. R. *J. Immunol.* 1990, 144, 3078-3086.

(36) Moebius et al. *J. Exp. Med.* 1992, 176, 507-517.

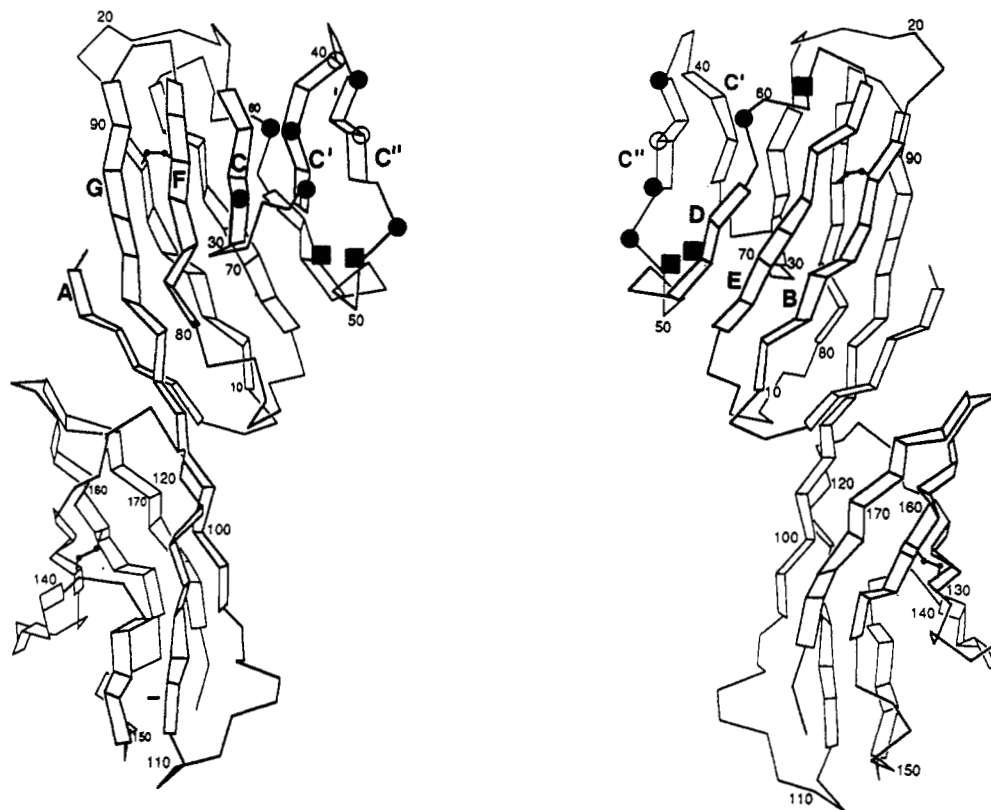


Figure 4. Residues in D1 that influence gp120 binding.³⁵ Solid circles represent positions where changes in surface side chains significantly diminish affinity; solid squares represent positions where changes in inward-facing side chains decrease affinity; open circles show surface positions where certain changes lead to small increases in affinity. The left- and right-hand diagrams are "front" and "back" views, respectively; the front view corresponds to Figure 2b.

substitutions at position 43: Leu, Trp, and Tyr perturb the interaction to some degree, changing the dissociation constant from 1 to about 20 nM, while Ile and Ala render the binding undetectable.³⁶ It is therefore likely that gp120 contains a specificity pocket to receive the exposed phenylalanine side chain. The importance of the projecting conformation of the C'C'' turn is clear from the effects of changes in Trp 62, which underlies the C'C'' turn. Phe 43 is surrounded on the molecular surface by several positively charged residues that also contribute to the contact: Lys 35, Lys 46, and Arg 59.³⁶ The total area of the surface defined by residues identified in this way is about 900 Å². There is presumably a groove in gp120 complementary to this region.

The interaction between CD4 and HIV involves more than simple docking. Some flexibility in the CD4 C'C'' ridge is implied by comparison of coordinates from two different crystal forms.^{19,37} There are shifts of up to 2 Å at the top (Phe 43) and bottom (Pro 48) of the C'' strand. Even larger conformational changes are believed to occur in the native HIV envelope glycoprotein.³⁸ HIV from many primary patient isolates has a low affinity for CD4, but gp120 dissociated from such a virus binds tightly.³⁹⁻⁴¹ Moreover, as expected from this result, CD4 induces shedding of gp120 from the virus particles.⁴² Thus, the gp120/gp41 oligomer be-

haves as if it undergoes a cooperative allosteric transition, with CD4 as the triggering ligand.³² There could be a significant conformational change in the gp120 fragment when CD4 binds: alternatively, the quaternary organization of the gp120/gp41 oligomer might simply interfere with proper docking of CD4 in the putative gp120 groove. In the former case, the conformational change in gp120 required for CD4 interaction would be accompanied by a loosening of its contact to gp41. In influenza virus hemagglutinin—the one case where the structure of a viral attachment glycoprotein is known⁴³—receptor binding as such does not trigger a conformational change, but proton binding, which accompanies transfer to the low-pH endosomal compartment of a cell, does. There is loosening of the contact between the receptor-binding domain and the base of the hemagglutinin, and a concomitant exposure of a fusion activity.⁴³ The properties of the influenza hemagglutinin thus provide a paradigm for thinking about the HIV envelope protein, with CD4 playing the combined role of receptor and conformational trigger. Drug development strategies could be aimed either at inhibiting the CD4/gp120 interaction in order to prevent binding or at inducing the conformational change in unbound HIV particles in order to disarm the virus. CD4 mimicry might well be a useful starting point for both strategies. Soluble CD4 itself has been unsuccessful as a drug, because of the high concentrations needed to bind most HIV isolates.⁴⁴ The structure of the gp120 site and the nature of its linkage to a

(37) Yan, Y., et al. In preparation.

(38) Sattentau, Q. J.; Moore, J. P. *J. Exp. Med.* 1991, 174, 407-415.

(39) Ashkenazi, A., et al. *Proc. Natl. Acad. Sci. U.S.A.* 1991, 88, 6056-6060.

(40) Brighty, D. W., et al. *Proc. Natl. Acad. Sci. U.S.A.* 1991, 88, 7802-7805.

(41) Turner, S., et al. *Proc. Natl. Acad. Sci. U.S.A.* 1992, 89, 1335-1339.

(42) Moore, J. P., et al. *Science* 1990, 250, 1139-1142.

(43) Wiley, D. C.; Skehel, J. J. *Annu. Rev. Biochem.* 1987, 56, 365-394.

(44) Moore, J. P., et al. *J. Virol.* 1992 66, 235-243.

conformational change will determine whether smaller molecules that resemble the C'C' ridge can circumvent this problem.

CD4 as MHC Ligand

The interaction of CD4 with class II MHC molecules has been more difficult to study than its gp120 binding, because the weak association can be detected only in a cell adhesion assay, where many contacts occur in parallel.^{5,45} The conclusions drawn by different laboratories are not entirely consistent, but the actual results are relatively similar.^{28,46-48} Mutation of various residues in D1 and D2 alters the observed adhesion or other measures of interaction. Those in D1 lie on all lateral faces of the domain, but not at the tip; those in D2 lie primarily in the D1-proximal part of the domain.⁴⁸ There is one report of an effect from D3,⁴⁶ but most D3 changes do not appear to influence MHC binding.⁴⁸ The simplest picture that can be derived from this

(45) Clayton, L. K., et al. *Nature* 1989, 339, 548-551.

(46) Fleury, S. et al. *Cell* 1991, 66, 1037-1049.

(47) Lamarre, D., et al. *Science* 1989, 245, 743-746.

(48) Moebius, U., et al. *Proc. Natl. Acad. Sci. U.S.A.* In press.

(49) D. C. Wiley, personal communication.

pattern of effects involves interaction with more than one class II molecule, so that D1 is "surrounded" on both sides by class II contacts. In a T-cell, it is likely that CD4 will associate laterally with a TCR complex. Domains 1 and 2 will then project beyond the TCR and interact with the MHC class II molecule bound to it. Oligomerization of this TCR/MHC/CD4 assembly—by any of a variety of mechanisms—could bring more than one MHC molecule in contact with D1 and D2 from a given CD4. Clustering of TCR assemblies is indeed believed to be required for signal transduction and T-cell activation. The crystallographically-determined structure of a class II MHC protein should soon be available,⁴⁹ and docking studies may suggest a more specific model for the MHC/CD4 interaction in such a complex.

Work in my laboratory that is included in this review was carried out by J. Wang, Y. Yan, T. Garrett, and J. Liu, in collaboration with U. Moebius, L. Clayton, and others in the laboratory of E. Reinherz (Dana-Farber Cancer Center, Boston). I thank Leo Brady (University of York) for helpful discussions and all the authors of ref 21 for communicating results prior to publication. Supported in part by NIH Grant AI-30361.