Received 6 June; accepted 12 October 1990.

1. Davis, R. Jr *et al. Proc. 21st Int. Conf. Cosmic Rays* Vol. 12 (ed. Protheroe, R. J.) 143–151 (Physics Publications, University of Adelaide, 1990).
2. Rowley, J. K., Cleveland, B. T. & Davis, R. Jr in *Proc. Conf. Solar Neutrinos and Neutrino Astronomy* Vol. 126 (eds Cherry, M. L., Lande, K. & Fowler, W. A.) 1–21 (Am. Inst. Phys., 1985).
3. Bahcall, J. N., Davis, R. Jr & Wolfenstein, L. *Nature* **334,** 487–493 (1988).
4. Coffey, H. E. (ed.) *Solar-Geophysical Data, Rep. No. 540-Part 1* (National Geophysical Data Center, Boulder, 1989).
5. *Cosmic Ray Nucleonic Intensity Data* (Bartol Research Institute, Newark, Delaware, 1970–1989).
6. Bevington, P. R. *Data Reduction and Error Analysis for the Physical Sciences* (McGraw Hill, New York, 1969).
7. Bahcall, J. N., Field, G. B. & Press, W. H. *Astrophys. J.* **320,** L69–L73 (1987).
8. Gaisser, T. K. & Stanev, T. in *Proc. Conf. Solar Neutrinos and Neutrino Astronomy* Vol. 126 (eds Cherry, M. L., Lande, K. & Fowler, W. A.) 277–281 (Am. Inst. Phys., 1985).
9. Fireman, E. L., Cleveland, B. T., Davis, R. Jr & Rowley, J. K. in *Proc. Conf. Solar Neutrinos and Neutrino Astronomy* Vol. 126 (eds Cherry, M. L., Lande, K. & Fowler, W. A.) 22–31 (Am. Inst. Phys., 1985).
10. Cisneros, A. *Astrophys. Space Sci.* **10,** 87–91 (1971).
11. Voloshin, M. B., Vysotskii, M. I. & Okun, L. B. *Soviet J. nucl. Phys.* **44,** 440–441 (1986).
12. Voloshin, M. B. & Vysotskii, M. I. *Soviet J. nucl. Phys.* **44,** 544–545 (1986).
13. King, J. H. *Interplanetary Medium Data Book* Suppl. 4 (1985–1988) (National Space Science Data Center, Greenbelt, 1989).
14. Howard, R. *A. Rev. Astr. Astrophys.* **22,** 131–155 (1984).
15. Mikheyev, S. P. & Smirnov, S. *Nuovo Cim.* **C9,** 17–26 (1986).
16. Cribier, M., Hampel, W., Rich, J. & Vignaud, D. *Phys. Lett.* **B182,** 89–94 (1986).
17. Bahcall, J. N. & Press, W. H. *Astrophys. J.* (submitted).
18. Filippone, B. & Vogel, P. *Phys. Lett.* **B246,** 546–550 (1990).
19. Bahcall, J. N. & Ulrich, R. K. *Rev. mod. Phys.* **60,** 297–372 (1988).
20. Turck-Chieze, S., Cahen, S., Casse M & Doom, C. *Astrophys. J.* **335,** 415–424 (1988).
21. Hirata, K. S. *et al. Phys. Rev. Lett.* **63,** 16–19 (1989).
22. Hirata, K. S. *et al. Phys. Rev. Lett.* **65,** 1297–1300 (1990).
23. Suzuki, A. KEK Preprint 90-31 (1990).
24. Raffelt, G. Max Planck Institute preprint MPI-PAE/PTH 6/90 (1990).
25. Lim, C-S. & Marciano, W. *J. Phys. Rev.* **D37,** 1368–1373 (1988).

ACKNOWLEDGEMENTS. We thank K. Lande and R. Davis Jr for discussions and for unpublished data, and B. Cleveland for simulation data. We thank B. Filippone for pointing out the problem with the mean ^{37}Ar production rate. This work is supported by the NSF and the Department of Energy.

Atomic structure of a fragment of human CD4 containing two immunoglobulin-like domains

Jiahuai Wang*, Youwei Yan*†, Thomas P. J. Garrett*†, Jinhuan Liu*, David W. Rodgers*, Robert L. Garlick‡, George E. Tarr§, Yasmin Husain§, Ellis L. Reinherz§‖ & Stephen C. Harrison*†

* Harvard University, Department of Biochemistry and Molecular Biology, and † Howard Hughes Medical Institute, 7 Divinity Avenue, Cambridge, Massachusetts 02138, USA
‡ Upjohn Company, Kalamazoo, Minnesota 49007, USA
§ Laboratory of Immunobiology, Dana Farber Cancer Institute and ‖ Department of Medicine, Harvard Medical School, 44 Binney Street, Boston, Massachusetts 02115, USA

The structure of an N-terminal fragment of CD4 has been determined to 2.4 Å resolution. It has two tightly abutting domains connected by a continuous β strand. Both have the immunoglobulin fold, but domain 2 has a truncated β barrel and a non-standard disulphide bond. The binding sites for monoclonal antibodies, class II major histocompatibility complex molecules, and human immunodeficiency virus gp120 can be mapped on the molecular surface.

THE cell-surface glycoprotein CD4 is expressed on most thymocytes and on the subset of peripheral T lymphocytes that includes helper T cells[1,2] and class II major histocompatibility complex (MHC) specific cytotoxic T cells[3–5]. Several experiments suggest that CD4 contacts nonpolymorphic regions of class II MHC molecules and that this interaction is important for triggering through the T-cell receptor by class II MHC-presented antigen[3–8]. The CD4/class II MHC interaction is also critical for a key stage in thymocyte differentiation[9]. CD4 may play a direct part in T-cell activation[10], and a specific tyrosine kinase associated with the cytoplasmic domain of CD4 provides a possible route for ligand-dependent signal transduction[11,12].

Human CD4 is the receptor for HIV[13–18]. Binding of the viral glycoprotein, gp120, to CD4 is the initial step in viral entry, leading to fusion of viral and cell membranes[19,20]. Blocking viral attachment by inhibiting gp120–CD4 binding is a potentially valuable antiviral strategy in AIDS therapy.

CD4 (relative molecular mass 55,000) consists of an extracellular portion (residues 1–371), a transmembrane segment (372–395), and a cytoplasmic tail (396–433) (refs 21, 22). The N-terminal 100 residues have clear sequence similarity to immunoglobulin variable domains (particularly Vκ) (ref. 22), and mapping of sites for anti-CD4 monoclonal antibodies indicates that this part indeed folds like an immunoglobulin variable region[23,24]. Less striking similarities suggest that residues 100–180, 180–290 and 290–370 (roughly) may also be homologous to immunoglobulin domains[22,25]. The extracellular portion has therefore been described as a concatenation of four immunoglobulin-like units. We use the domain terminology here, because our structure bears out predictions about domains 1 and 2, but evidence for distinct third and fourth domains is only indirect. Recombinant products corresponding to the intact extracellular region ('domains 1–4') are monomeric[41], and the CD4 glycoprotein is believed to be monomeric on the cell surface as well[26]. The two *N*-glycosylation sites on human CD4 are in domains 3 and 4[21]. HIV gp120 binds to various soluble CD4 derivatives[27–33]. The known critical contacts are confined to domain 1 (refs 23, 33–37). Contacts to class II MHC molecules probably extend over domains 1 and 2 (refs 38, 39).

Several groups have obtained crystals of recombinant soluble CD4 derivatives[40–42]. Those from four-domain species have large unit cells, and they diffract relatively poorly. By contrast, crystals of N-terminal, two-domain products diffract to high resolution. We report here on a structure obtained from two forms of the two-domain species. Independent results on similar crystals are reported in the accompanying paper by Ryu *et al.*[43].

Structure determination

We studied two forms of truncated CD4, referred to here as CD4(1–182) and CD4(1–183). CD4(1–182) was produced in secreted form by transfected Chinese hamster ovary (CHO) cells; it contains the first 182 residues of the intact molecule. CD4(1–183) was produced as an inclusion-body protein in *Escherichia coli* using the vector pTrp2-CD4/183; it contains an N-terminal methionine and 183 residues of authentic CD4. It was refolded as described in ref. 44. Both products were purified by immunoaffinity chromatography and gel filtration[29,44].

Three crystal forms were identified, all in space group C2, and called A, B and C. Both products seem to give all three

forms. We focused on the C form of CD4 (1-182), which diffracts to spacings of at least 2.2 Å. This report is based primarily on a determination of the 2.4 Å resolution structure from those crystals, using multiple isomorphous replacement (MIR), followed by iterative cycles of refinement and model building. We also solved the structure of the A form of CD4 (1-183) to 3 Å resolution by molecular replacement, and we used averaging of density in the two unit cells to resolve ambiguities at late stages in the refinement of the C-form structure.

Aspects of the C-form structure determination are summarized in Table 1*a*. We identified three heavy-atom derivatives (K_3OsCl_6, K_2PtBr_4 and UO_2SO_4) by soaking under varied conditions of pH and ionic strength, to minimize non-isomorphism. The Os and Pt derivatives were solved by examination of isomorphous and anomalous-difference Patterson functions; the uranyl derivative, by a difference Fourier synthesis. Introduction of the heavy atoms greatly reduced the attainable resolution, as shown in Table 1*a*. Moreover, non-isomorphism, especially of the uranyl derivative, seemed to limit the accuracy of the phase determinations, and the MIR map was of marginal quality (Fig. 1). Nonetheless, clear density for two lobes, each with β sheets, could be seen. Solvent flattening[45], using an envelope determined by visual inspection and using the Bricogne program suite for density modification[46], improved the clarity of the map. The expectation that one part of the molecule should resemble an immunoglobulin variable domain enabled us to fit most of domain 1, including many side chains, using the program FRODO[47], and we could then build segments of a polyalanine model into parts of domain 2, for which the density was weaker. About 60% of the polypeptide chain, with about half the atoms, was fitted at this stage and preliminarily refined at 3.5 Å resolution using simulated annealing in the program XPLOR[48,49]. MIR and partial structure phases were combined[46], Fourier coefficients were computed with modified Sim weights[50], and solvent regions in the phase-combined map were flattened, as described above. Three rounds of model building, refinement, and phase combination, increasing the resolution at each stage, led to an essentially complete structure at 2.8 Å resolution. This model revealed the immunoglobulin topology of domain 2. Four rounds of refinement and rebuilding, using omit-refine and difference maps to adjust the model and to add some remaining residues, extended the resolution to 2.6 Å. Individual thermal parameters were then introduced, and the resolution was extended to 2.4 Å in several more rounds of refinement and local rebuilding, with the aid of maps produced by density-averaging between A-form and C-form structures, as described below. This averaging not only resolved ambiguities in certain surface loops, but also reduced bias in those regions for which the MIR maps had not provided a clear guide for initial model building. Refinement parameters appear in Table 1*c*. The R factor (using $F > \sigma$) is 22% for data between 6 and 2.4 Å. Ordered solvent has not so far been included.

The r.m.s. error in atomic positions of 0.2 Å was estimated by comparing coordinates before and after a last round of molecular dynamics at 300 K. The errors are not evenly distributed throughout the structure, however. The least well determined residues have main chain atoms with large thermal parameters (>50 Å^2); these are residues 20-23, 103-107, 124-126,

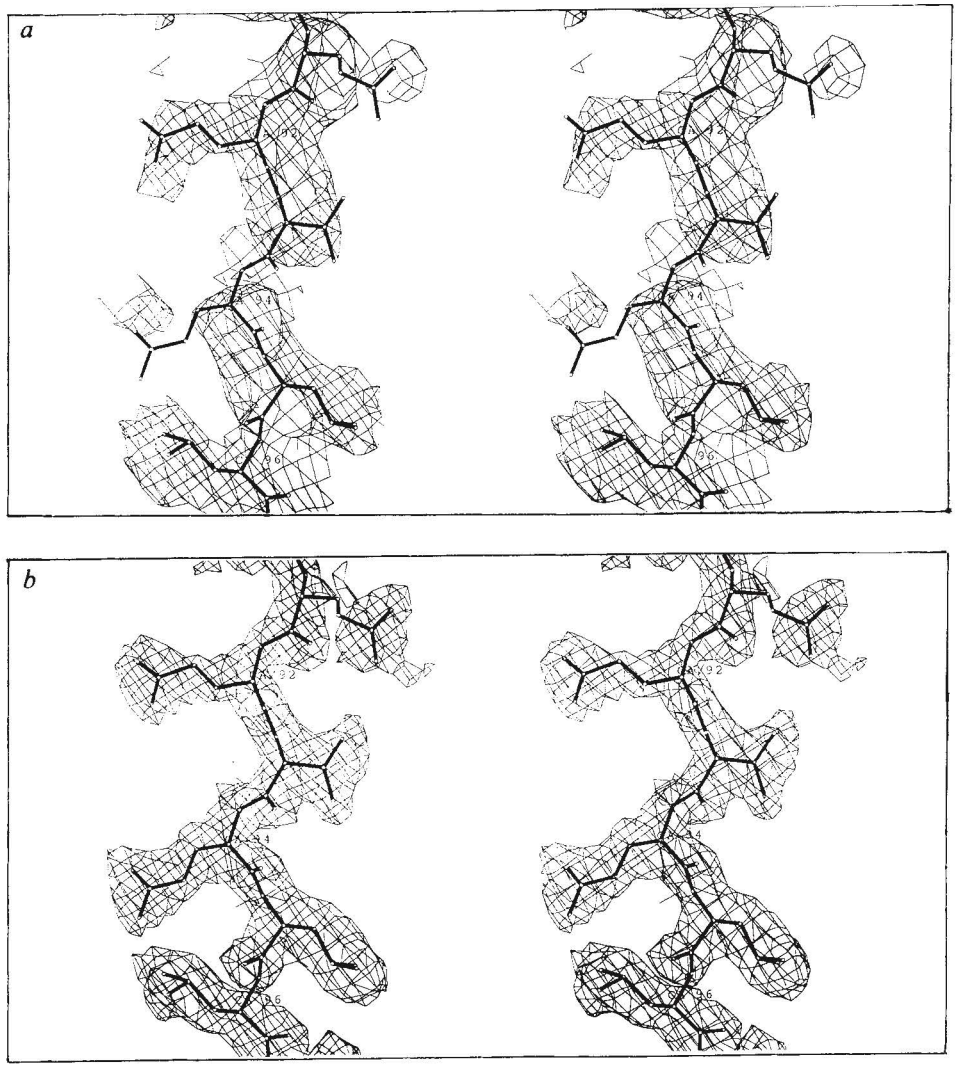

FIG. 1 Portion of the C-form electron density map, showing density for residues 92-96 at two stages of the structure determination. *a*, MIR map, 3.5 Å resolution. *b*, $(2F_o - F_c)$ map after the last round of refinement reported here, 2.4 Å resolution.

131–138 and 151–154. Little or no electron density is present for residues beyond 176, and the side-chain positions for residues 1, 24, 43, 129, 139 and 150 are also uncertain. As indicated by early electron density maps, the average thermal parameters for domains 1 and 2 differ substantially (20 Å^2 and 43 Å^2, respectively), but this difference seems less dramatic in the A form.

The structure of CD4(1–183) in the A form was solved by molecular replacement using data to 3.3 Å resolution (Table 1*b*) and a model from CD4(1–182) in the C form at an intermediate refinement stage. The fast rotation function[51] and a Patterson-space rotation search[52] gave identical solutions. A translation search, followed by rigid-body refinement with either CORELS[53] or XPLOR[49], gave $R = 39\%$ to 3.3 Å. Computations in which the model was split into two domains gave no relative reorientation, showing absence of hinge-bending at the junction. One round of simulated annealing refinement with XPLOR reduced the R factor to 25% for data 7–3 Å. Density averaging between the two space groups, using data to 3 Å resolution and carried out as described for HLA[54], gave significant phase improvement, and maps calculated with those phases substantially accelerated completion of the C-form refinement. The C-form model at a late stage was then used as a starting point for a round of simulated annealing refinement in the A form (to 3 Å), resulting in an R factor of 20% for a partially refined structure with good geometry (see Table 1*c*). Successful refinement and averaging show that CD4(1–183) is essentially identical in structure to CD4(1–182).

Molecular structure

The polypeptide chain of CD4(1–182) folds into two abutting domains (Fig. 2). Each domain is an antiparallel β barrel with the connectivity characteristic of an immunoglobulin fold. The two domains pack so closely, one against the other, that there is a large common hydrophobic interface. Moreover, the strand from residue 89–104 passes continuously from one domain to the next, so that residue 98 forms a hydrogen bond in a sheet of domain 1 and residue 99 forms a hydrogen bond in a sheet of domain 2. The overall dimensions are roughly $65 \times 35 \times 25$ Å, giving the molecule a rod-like shape. The N terminus is near one end and the C terminus near the other. Evidence from electron microscopy[42] and hydrodynamic measurements[41] and

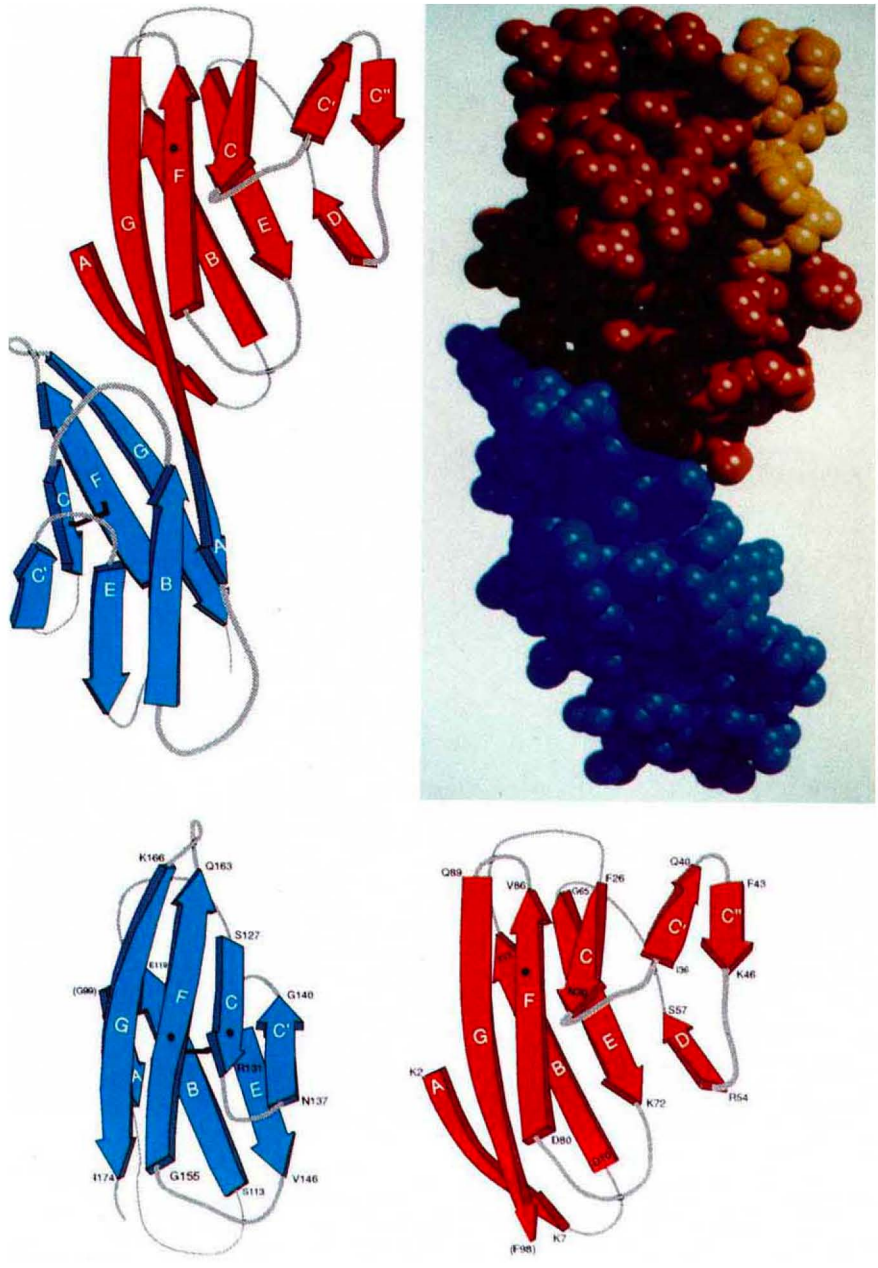

FIG. 2 *a*, Backbone representation of CD4(1–182). Domain 1 is in red, domain 2 in blue; β strands are indicated by letters, separately in each domain. Strand A of domain 2 is continuous with strand G of domain 1. Note that domains 1 and 2 are related by a rotation of ~160° and a translation along the axis of the molecule. Disulphide bonds are shown as solid lines; only the trace is visible of the disulphide bond between strands B and F in domain 1. *b*, Solid representation of CD4(1–182), in an orientation similar to *a*. The C″ ridge of domain 1, implicated in the binding of HIV gp120, is highlighted. *c*, Representations of domains 1 and 2 oriented to show the similarity of their folded structures. First and last residues in each strand are indicated by single-letter code and sequence numbers.

inferences from crystal-packing analyses of CD4(1-375) (refs 41, 42) suggest that the full extracellular segment is a rod-like molecule about 125 Å long and 25-30 Å wide. The shape of CD4(1-182) is clearly appropriate for half of such a rod, with the C terminus of its polypeptide chain at the membrane-proximal end. The orientation chosen for the illustrations therefore places the membrane below the model. The parts of the molecule at the top of the illustrations can be assumed to be the most accessible on the cell surface.

Domain 1. As predicted from sequence alignments, the structure of the first domain resembles very closely the variable domain of an immunoglobulin light chain. Fig. 3*a* shows a comparison with the Vκ domain of the Bence-Jones protein REI[55], and Fig. 4 presents the sequence alignment corresponding most closely to this three-dimensional superposition. Immunoglobulin variable domains have two β sheets, with strands conventionally labelled ABCC′C″DEFG[56,57]. Strands ABED form one sheet; strands GFCC′C″, the other. Strand A actually begins in the

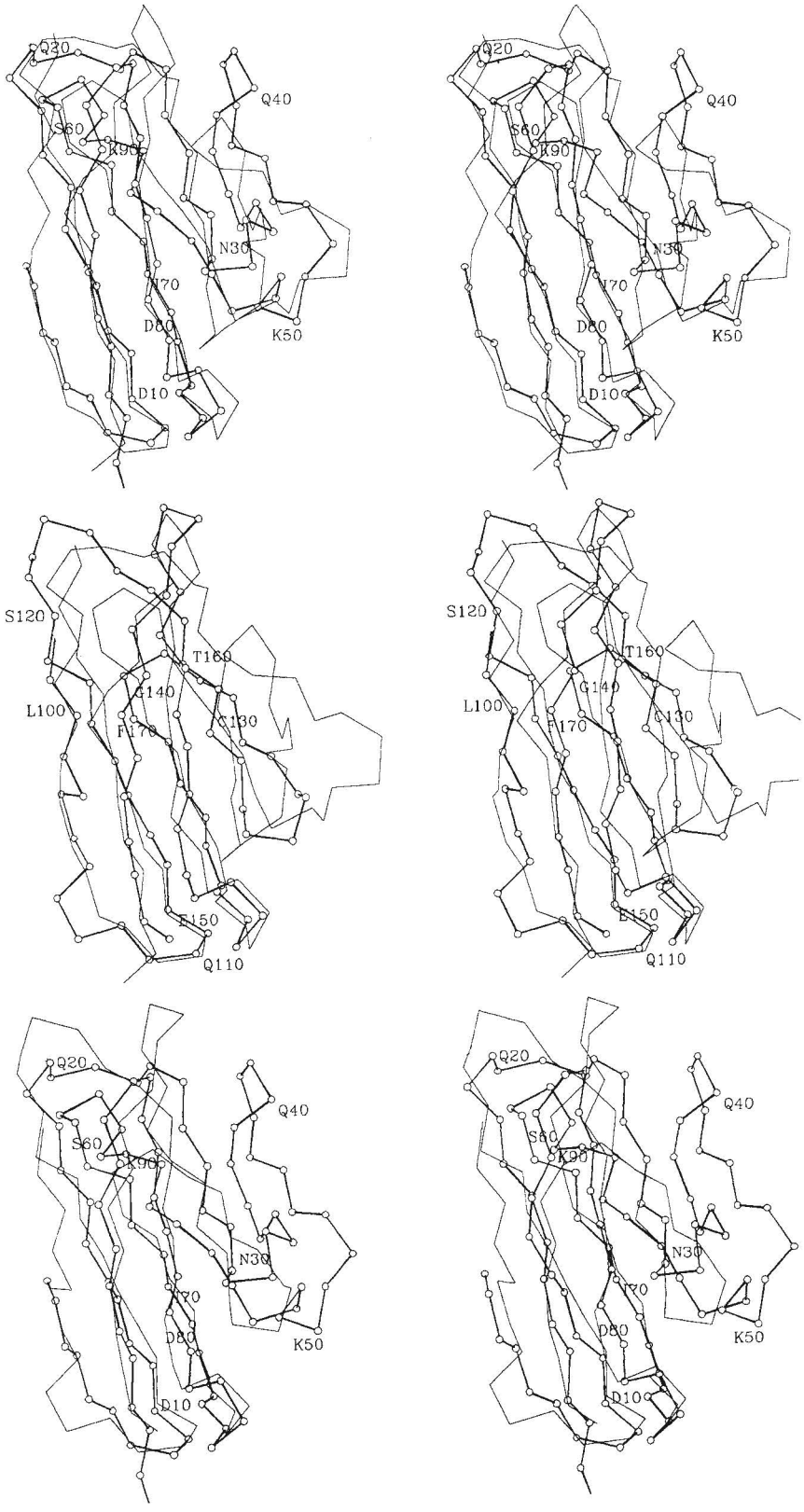

FIG. 3 Stereo drawings of pairwise α-carbon backbone superpositions of CD4 domain 1 (residues 1-98), CD4 domain 2 (residues 99-176), and immunoglobulin Vκ domain of Bence-Jones protein REI[55]. *a*, CD4 domain 1 (thick lines and residue numbers) versus REI (thin lines). The drawing is based on fitting the 53 common atoms in the so-called Pin region (ref. 56). The r.m.s. fit is 0.83 Å, comparable to pairwise comparisons among various immunoglobulin domains[56]. *b*, CD4 domain 2 (thick lines and residue numbers) versus REI (thin lines). The fit is based on 42 Cα atoms from strands A, B, E, F and G, selected by visual inspection. The r.m.s. agreement is 1.45 Å. *c*, CD4 domain 1 (thick lines and residue numbers) versus domain 2 (thin lines). The fit is based on 38 corresponding Cα atoms in strands A, B, C, E, F and G. The r.m.s. agreement is 1.28 Å.

first sheet, alongside B, and passes across to the second, alongside G. The first residue of CD4 corresponds to the residue of REI immediately after this switch, and the two sheets of CD4 domain 1 therefore contain strands BED and AGFCC′C″ respectively.

As in all standard immunoglobulin domains[58,59], strand B contains the first cysteine (residue 16) of a disulphide, and strand C contains a conserved tryptophan (residue 28) packed against it. The turn from C to C′ is more abrupt than in REI and the C′C″ corner correspondingly more extended. This extension of the C′C″ β ribbon is essentially a β turn. Its high thermal parameters imply some flexibility. It packs against the side chain of Trp 62, which projects from an insertion (relative to REI) forming a helical turn in the DE loop. In heavy-chain variable domains, the size of the framework residue at position 71 in the conventional numbering scheme has been shown to be an important determinant of the conformation of the C′C″ loop in CDR2[60]. In CD4, the side chain of Trp 62 occupies the space corresponding to heavy-chain residue 71, and the large tryptophan ring clearly holds the C′C″ turn in its outward-projecting position. With Leu 44, in the C″ strand, and Phe 67, in the DE loop, Trp 62 forms an extension of the immunoglobulin-like hydrophobic core. Through its effects on the shape of the C′C″ side of the domain, this extended hydrophobic core seems to be important for interactions with class II MHC molecules and HIV gp120 (see below).

The first residue of strand D is an arginine, conserved in immunoglobulin variable domains. It forms a salt bridge to a conserved aspartic acid in the EF corner. These residues in the CD4 sequence have been noted previously and cited as evidence for a variable-domain-like folded structure. Another conserved residue, Tyr 82, lies in strand F, with its side chain in the hydrophobic core, forming a hydrogen bond to the main-chain carbonyl of Asp 78. Strand F also contains the second cysteine (residue 84) of the characteristic immunoglobulin disulphide. The FG corner is five residues shorter than in REI. This loop corresponds to the CDR3 region of an immunoglobulin variable domain, where it is especially prominent and contributes critically to antigen binding. Since antigen binding is not a function of CD4, it is not surprising that the polypeptide chain makes an abrupt β turn at residues 87–88, rather than maintaining a large surface projection.

Domain 2. The second domain proceeds directly from the first. There is no equivalent of an elbow or hinge, and the A strand in domain 2 is a direct continuation of the G strand in domain 1. This continuous β-strand connector shows how a stiff, rod-like molecule can be constructed of sequential immunoglobulin-like domains. The structure of domain 2 deviates from a conventional immunoglobulin domain in two important respects: the disulphide links cysteines within one sheet rather than across the barrel, and there is a short connection between the end of strand C and the beginning of strand E. We have called the central part of this connection C′, because it participates in four antiparallel β-sheet hydrogen bonds with strand C and therefore corresponds roughly to strand C′ of a variable domain rather than to strand D of a constant domain. Indeed, much of the three-dimensional structure of domain 2 superposes well on the first domain and on the Vκ domain of REI. The comparisons are shown in Fig. 3*b*, *c*, and the sequence alignments derived from these superpositions are given in Fig. 4. Segments of domain 2 that correspond closely to segments of REI are the beginning of the A strand where it runs alongside B, the second part of the AB loop, the B strand and part of the C strand, the entire segment from the beginning of the E strand to the end of the F strand, and the second part of the G strand. The AB and EF loops superpose well on REI, including a conserved glycine in AB and a conserved aspartic acid in EF, both characteristic of immunoglobulin variable domains[59]. One model for domain 2 based on sequence alignments accurately assigned residues to strands and correctly predicted the placement of the disulphide[25].

The first cysteine of the disulphide (residue 130) is in strand C, next to its partner (residue 159) in strand F. It occupies the position taken by the nearly invariant tryptophan in immunoglobulin structures. Disulphides between hydrogen-bonded residues in antiparallel β strands are rare, because the positioning is unfavourable for the required geometry of the S–S bridge[61], but sequence alignments do predict side-by-side disulphide bonds in other members of the immunoglobulin superfamily[25]. The cysteine residue in strand F corresponds to the 'usual' position for the second member of the disulphide. In effect, domain 2 of CD4 has lost the cysteine in strand B and acquired one in strand C. A different disulphide migration seems to have occurred in the α chain of CD8[62]. Three cysteines are found in its immunoglobulin-like domain, two in the 'usual' locations and a third two residues before the conserved tryptophan in strand C. It has been reported that the disulphide in this domain links this 'extra' cysteine in strand C with cysteine in strand B, leaving the one in strand F unpaired[62]. Thus CD4 and CD8 both seem to have immunoglobulin-like domains with unconventional disulphide linkages. The third domain of CD4, not present in the 1–182 fragment, is believed to have an immunoglobulin-like structure with no disulphide bond[25]. Precedent for such a structure is found in the two immunoglobulin-like domains of the *E. coli* chaperone protein, pap D[63].

The unusual placement of the disulphide bond produces little distortion in domain 2 with respect to REI or domain 1, and despite extensive differences in the identity of particular residues, the overall shape of the hydrophobic core is maintained. The placement of the cysteine in strand C eliminates the normally conserved tryptophan, and a leucine occupies the position of the standard cysteine in strand B. There is a tryptophan instead of a tyrosine at position 157, two residues before the second cysteine in strand F. Its ring-NH group forms a hydrogen bond with the main-chain carbonyl of residue 153; this interaction corresponds precisely to the bond formed by the tyrosine-OH group in REI and CD4 domain 1. Another hallmark of immunoglobulin variable domains, the salt bridge between an arginine at the beginning of strand D and an aspartic acid in the EF loop, is partially modified in domain 2. The

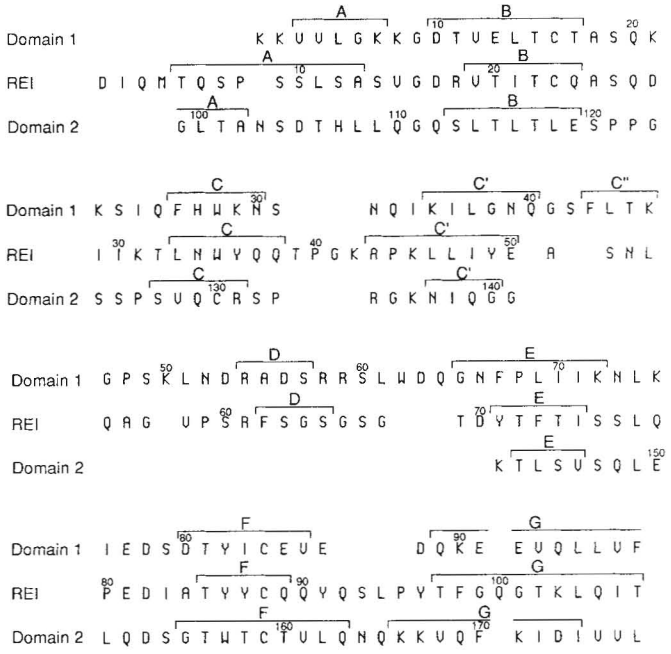

FIG. 4 Sequence alignments implied by the superpositions in Fig. 3. Strands were assigned in CD4 by the method of Kabsch and Sander[68], using our best-refined coordinate set. Strand assignments in REI are based on the Brookhaven database entry[55].

TABLE 1 Statistics for data collection, phase determination and refinement

(*a*) C form

Crystallization: CD4(1–182), 16–18 mg-ml^{-1} by hanging-drop method. Well solution: 20% PEG 4K, 0.05 M HEPES (pH 7.85), 0.05 M KCl, 0.1% azide, 20 °C
Unit cell: Space group *C*2, a=84.23 Å, b=30.65 Å, c=88.94 Å, β=118.43°

		Resolution shells (Å)					
	Overall	15–6	6–4	4–3.3	3.3–3	3–2.6	2.6–2.4
Native							
Number of measurements	31,554	2,547	7,485	6,543	4,283	7,423	3,142
Unique reflections	7,829	508	1,249	1,369	1,018	2,180	1,505
Expected unique reflections	8,048	521	1,254	1,372	1,018	2,199	1,684
Completeness (%)	97.7	97.5	99.6	99.8	100	99.6	89.4
Completeness (%) with $I/\sigma>1$	85.7	97.1	99.2	98.5	95.3	83.9	58.2
R_{merge} (%)*	8.6	6.2	8.7	9.0	11.2	15.1	21.1
Mean figure of merit	0.59	0.70	0.68	0.36			
Derivatives							
K_3OsCl_6							
Number of unique reflections	3,046	515	1,240	1,291			
R_{merge} (%)*	8.0	7.8	7.5	14.3			
R_c (%)†	71	58	83	81			
Phasing power‡	1.54	1.19	1.73	1.41			
K_2PtBr_4							
Number of unique reflections	2,913	509	1,202	1,202			
R_{merge} (%)*	9.0	8.1	8.7	14.9			
R_c (%)†	71	64	69	92			
Phasing power‡	1.88	1.48	1.90	1.70			
UO_2SO_4							
Number of unique reflections	1,741	500	1,241				
R_{merge} (%)*	7.1	6.1	7.6				
R_c (%)†	66	68	62				
Phasing power‡	1.10	1.08	1.05				
Refinement:							
R(%)§ using $F/\sigma>1$	22.1		16.6	19.9	24.5	28.9	30.5
using $F/\sigma>3$	21.1		16.6	19.7	23.5	27.1	26.5

(*b*) A form

Crystallization: CD4(1–183), 16–18 mg ml^{-1} by hanging-drop method. Well solution: 20% PEG 4K, 0.05 M HEPES (pH 7.85), 0.1% azide, 4 °C
Unit cell: Space group *C*2, a=80.15 Å, b=32.14 Å, c=75.98 Å, β=103.08°

	Overall	15–6	6–4	4–3.3	3.3–3
Native					
Number of measurements	7,692	1,729	2,491	2,244	1,228
Unique reflections	3,400	459	1,039	1,128	774
Expected unique reflections	3,930	489	1,167	1,290	984
Completeness (%)	86.5	93.8	89.0	87.4	78.7
Completeness (%) with $I/\sigma>1$	72.9	91.6	85.4	73.1	48.5
R_{merge} (%)*	12.5	9.2	9.7	18.3	38.4
Refinement					
R (%)§ using $F/\sigma>1$	19.1		17.1	19.8	23.8
using $F/\sigma>3$	18.0		16.7	18.4	21.4

(*c*) Refined structure

R.m.s. deviations of restrained parameters from ideal values‖

	Bonds	Angles	Dihedral angles	Chiral and planar restraint	Thermal parameter restraints Bond	Angles
C form	0.027 Å	4.19°	29.5°	1.97°	0.94 Å^2	1.24 Å^2
A form	0.025 Å	4.86°	29.3°	1.86°	0.74 Å^2	1.00 Å^2

Data were collected with CuKα radiation, generated by an Elliot GX-13 rotating anode and collimated by Franks double-mirror optics, using a Siemens–Nicolet area detector controlled by the programs described in ref. 66. Integrated intensities were scaled and merged using the CCP4 program suite, which was also used for many of the subsequent steps in the structure determination. Heavy-atom parameters were refined and MIR phases calculated using the program HEAVY[67]. The three derivatives were prepared in somewhat different ways: (1) K_3OsCl_6, 1 mM in 25% PEG 4K, 0.05 M HEPES, (pH 7.85) 0.05 M KCl, 3-day soak; (2) K_2PtBr_4, 1 mM in 25% PEG 4K, 0.05 M HEPES (pH 7.85), with no salt to prevent cell change, 3-day soak and 4–6-hour washing; (3) UO_2SO_4 saturated in 25% PEG 4K, 0.05 M HEPES (pH 7.0), 0.2 M KCl, 2-day soak.

* $R_{merge}=\Sigma_h\Sigma_i|I_{hi}-I_h|/\Sigma_{hi}I_{hi}$, where h are unique reflection indices, I_{hi} are intensities of redundant, symmetry-related reflections of index h, and I_h is the mean intensity for reflections of index h.

† R_C, Cullis R factor for centric reflections.

‡ Phasing power, mean value of the heavy-atom structure factor amplitudes divided by the residual lack-of-closure error.

§ $R=\Sigma|F_o-F_c|/\Sigma F_o$, where F_o and F_c are observed and calculated structure factor amplitudes, respectively. The refinement does not include reflections at spacings larger than 6 Å.

‖ For definitions of restraints as applied, see the program manual of XPLOR[49].

absence of strand D eliminates the usual arginine, but its place seems to be taken by Lys 136 at the CC′ corner.

The last few residues of domain 2 appear to be disordered in CD4(1–182). Val 176 is the last residue placed in our most recent model. It lies just at the position where the G strand emerges from the β sheet in the middle of a cluster of hydrophobic residues. The disordered residues 177–182 are likely to form the connection to domain 3. The alignments described above show that they correspond to residues 98–103 (the last of domain 1 and the first five of domain 2). We note that fragments 95–100 and 174–179 are both sequences of hydrophobic residues, suggesting that if the next part of CD4 is indeed immunoglobulin-like in its fold, then the link between domains 2 and 3 may be similar to the one between domains 1 and 2. Many cell-surface proteins are composed of concatenated immunoglobulin-like elements[25,59], and the interdomain connection seen in this first view of such a protein is also likely to be found in other structures.

Antibody binding sites

Sites for binding monoclonal antibodies to CD4 have been mapped by several groups by testing their reactivity with CD4 altered at one or more residues[24,64,65], by generating 'escape mutants' of CD4[23], or by analysing patterns of cross competition between pairs of antibodies[15]. A summary of the results for several monoclonals that map to domain 1 is shown in Fig. 5 by using a diagram of the CD4 structure. Only changes that selectively affect certain monoclonals are shown, as changes that affect a broad spectrum of antibodies probably destabilize the entire domain. As expected, the residues that appear to interact with a given antibody cluster in a single patch on the CD4 surface. Crystallographic analyses of complexes between antibodies and their protein antigens show that the interface is approximately 25–30 Å in diameter[57]. The sites defined by observable changes in binding to mutated CD4 are likely to represent a minimum estimate for the actual antigen-antibody interface, because available variants do not necessarily saturate the surface and because changes in residues near the periphery of the interface may have a less pronounced effect than changes in centrally located side chains. Indeed, most of the patches in Fig. 5 are about 15 Å in diameter. There are clearly two sorts of mutations that affect antibody binding without destroying overall stability: alterations of an exposed residue, defining a contact at the interface, and alterations of a buried residue, probably creating local conformational distortion. Only exposed residues are shown in Fig. 5 An example of a change in a buried residue is A55F, which selectively diminishes binding of OKT4d. The bulky phenylalanyl side chain would be expected to distort the C″D corner (residues 47–54), precisely where surface changes affect OKT4d affinity.

The binding sites defined by the summary in Fig. 5 are generally consistent with patterns of cross-blocking between the corresponding antibodies[15]. The size of an immunoglobulin antigen-combining region (two domains each the size of domain 1) implies substantial excluded volume, and two antibodies binding anywhere on the same side of domain 1 might be expected to interfere. Thus VIT4 and MT 151, which both 'see' the AGFCC′C″ face of domain 1, compete effectively with each other. Neither competes with Leu3a, which seems to bind to the right-hand side of the molecule as depicted in Fig. 5*a*, nor with OKT4a, which binds to the BED face. T4/18T3A9 competes with both VIT4 and Leu3a, as the mapping of their sites requires. One noninterfering pair consists of Leu3a and OKT4d. The diagram in Fig. 5*a* suggests that both these antibodies interact with the C′C″ edge of domain 1, but that Leu3a approaches from 'above' and OKT4d from 'below'. Monoclonal MT151 selects for escape mutants on both domain 1 (residue 94) and domain 2 (residue 165) (ref. 23). These residues are about 12 Å apart. Mutation of lysine 1, which lies between them, also diminishes MT151 binding[65].

Class II MHC interaction

Amino-acid substitutions that interfere with functional interaction of CD4 and class II MHC molecules on a target cell occur on both domains 1 and 2 (ref. 38). Mutations in the CC′ and C″D corners of domain 1 and along the A strand of domain 2 have the strongest effects (arrows in Fig. 5). Monoclonal OKT4a, which binds to the BED face of domain 1 (Fig. 5), competes with the MHC interaction, as do Leu3a and HIV gp120[38,39]. Examination of the model shows that there must be a very extended region of lateral contact to account for these interferences. The relative orientation of the two domains produces a noticeable concavity in the overall shape of the two-domain fragment, and this surface might provide a notch for interlocking CD4 with a class II MHC molecule projecting 'downwards' from the opposite cell (in the orientation of Fig. 5). If the shape of the MHC molecule were roughly complementary to this notch, most of the widespread contacts implied by the mutational data could occur. In any case, the distribution of the implied contacts is consistent with the notion that CD4 must project beyond the outer, antigen-presenting domain of the MHC molecule to interact with nonpolymorphic regions near its base. Moreover, the total length of the CD4 molecule is sufficient for CD4 to line up with a T-cell receptor bound to an antigen-presenting class II molecule and to form a specific ternary complex.

Binding of gp120

Mutations that influence gp120 binding cluster near the edge of domain 1 that contains strand C″ (refs 23, 24, 34–36, 64, 65). A concordance of such changes is presented in Fig. 6. As in the case of changes that alter antibody binding, these sites include both exposed and buried residues. Mutations that alter affinity of several monoclonal antibodies with nonoverlapping binding sites are not included, because their effect can be interpreted as a general destabilization of domain 1. The mutations in buried

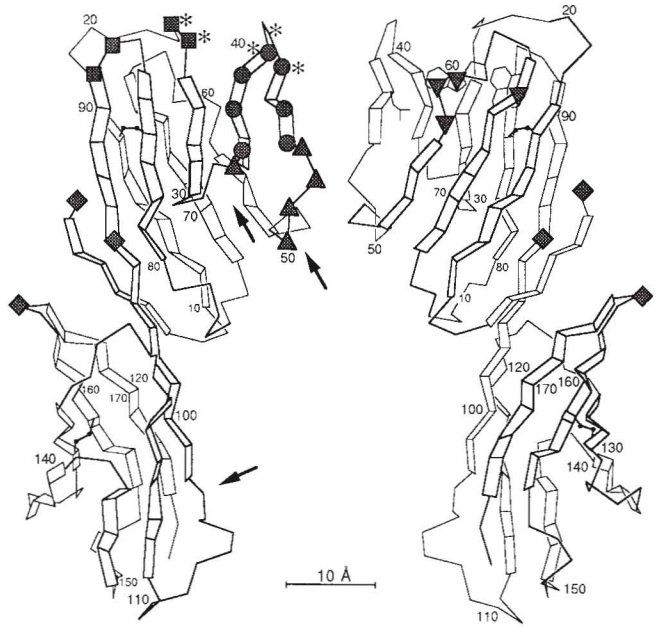

FIG. 5 Monoclonal antibody-binding sites and regions of MHC interaction on domains 1 and 2 of CD4. The directions of view are from the AGFCC′C″ face (left) and BED face (right) of domain 1. The left-hand view corresponds roughly to Fig. 2*a*. The pleated ribbons are β strands; the line segments define connecting loops. Each segment joins successive α-carbon positions. Small numbers label every tenth residue. The symbols represent locations of exposed residue that affect binding of a particular antibody: ▨, Leu3a (refs 23, 24, 64, 65); ▼, OKT4a (refs 23, 24, 64); ▲, OKT4d (refs 23, 24, 64); ▨, VIT4 (refs 23, 24); *, T4/18T3A9 (refs 23, 24); ◆, MT151 (refs 23, 24, 65). Arrows show loops where substitutions affect MHC binding[38]. In the right-hand view, side chains of Leu 44, Trp 62 and Phe 67 have been included to illustrate the extended hydrophobic core supporting the C′C″ loop.

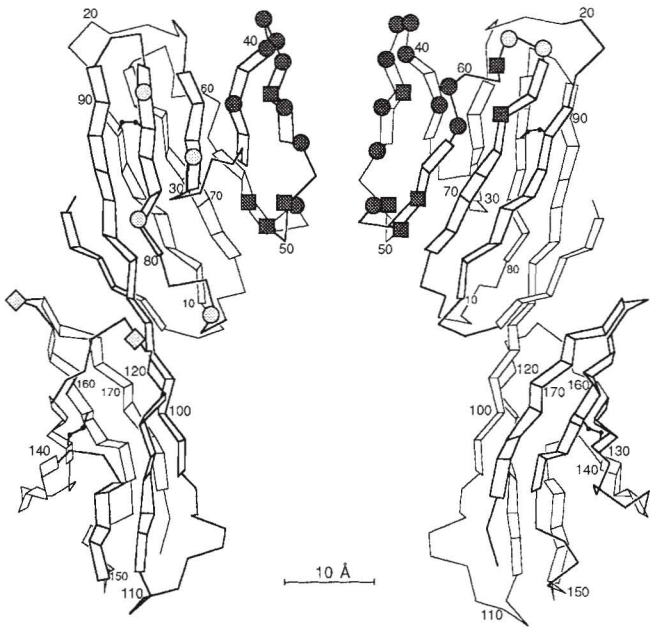

FIG. 6 Residues that affect gp120 interaction with CD4. The views are described in the caption to Fig. 5. The symbols represent classes of residues as follows: ●, exposed (refs 23, 24, 36, 64, 65); ■, buried (refs 24, 34, 36, 64); ◍, some exposed residues that are displaced from the C'C" ridge (ref. 65); ◈, positions in domain 2 reported to have some effect on gp120 interaction (refs 34, 35).

residues shown in Fig. 6 can therefore be presumed to influence binding by generating local, rather than global, perturbations in the folded structure. Many of the exposed residues that can be inferred to contact gp120 lie between positions 38 and 52. Their side chains decorate the C" edge, forming a ridge about 25 Å long (see Fig. 2*b*). They include an exposed phenylalanine at position 43 in the C'C" turn. Several mutations from the adjacent β-sheet surfaces also affect the gp120 interaction. Several of them involve side chains, such as Arg 59, that project toward the ridge, but results of a recent survey by alanine-scanning mutagenesis[65] implicate some residues displaced farther from the C" edge (see Fig. 6). Because changes in the A and B strands at the opposite edge of the domain and in the BC loop at the 'top' do not affect gp120 binding, we propose that the viral glycoprotein approaches domain 1 from the C" side and that it may fit over the two faces to some extent. It appears not to contact the surface of the CC' corner (residues 30–35), which projects strongly from the AGFCC'C" face. The C" ridge is supported at either end by hydrophobic pockets, in which mutations of buried residues reduce gp120 affinity. Examples of such mutations are W62C and F67L, which would perturb packing of side chains in the extended hydrophobic core beneath the C'C" turn, and S49Y and A55F, which would perturb the C"D corner.

What can we infer about the complementary site on gp120? Definition of the contact surface on CD4 by mutagenesis is clearly still incomplete, but the summary in Fig. 6 shows that it is likely to require a region of interaction on gp120 at least 25 Å long and 12 Å wide. The data suggest that the gp120 site is a groove rather than a flat surface, but specifically designed mutations are needed to confirm this and to estimate a depth. Because gp120 is a large protein, there may be additional contacts. These might be less critical for binding, and they might therefore not be detected by competition experiments with truncated forms of CD4. This possibility could explain reports of effects from changes near residues 122 and 164 in domain 2[34,35] and from a substitution at position 77 (ref. 65). If these observations do reflect local interactions, some part of gp120 could be imagined to extend across the middle of CD4 as seen in Fig. 6. The C" ridge seems nonetheless to be the most important surface from the perspective of inhibitor design. □

Received 16 October; accepted 23 October 1990.

1. Reinherz, E. L., Kung, P. C., Goldstein, G. & Schlossman, S. F. *Proc. natn. Acad. Sci. U.S.A.* **76,** 4061–4065 (1979).
2. Reinherz, E. L., Kung, P. C., Goldstein, G. & Schlossman, S. F. *J. Immun.* **234,** 2894–2896 (1979).
3. Meuer, C., Schlossman, S. F. & Reinherz, E. L. *Proc. natn. Acad. Sci. U.S.A.* **79,** 4395–4399 (1982).
4. Biddison, W., Rao, P., Thalle, M., Goldstein, G. & Shaw, S. *J. exp. Med.* **156,** 1065–1076 (1982).
5. Krensky, A. M., Reiss, C. S., Mier, J. W., Strominger, J. L. & Burakoff, S. J. *Proc. natn. Acad. Sci. U.S.A.* **79,** 2365–2369 (1982).
6. Gay, D. *et al. Nature* **328,** 626–629 (1987).
7. Engleman, E. C., Benike, C. J., Grumet, C. & Evans, R. L. *J. Immun.* **127,** 2124–2129 (1981).
8. Doyle, C. & Strominger, J. L. *Nature* **330,** 256–289 (1987).
9. Kruisbeek, A. M. *et al. J. exp. Med.* **161,** 1029–1047 (1985).
10. Eichmann, K., Jonsson, J. I., Falk, I. & Emmrich, F. *Eur. J. Immun.* **17,** 643–650 (1987).
11. Rudd, C. E., Trevillyan, J. M., Dasgupta, J. D., Wong, L. L. & Schlossman, S. F. *Proc. natn. Acad. Sci. U.S.A.* **85,** 5190–5194 (1988).
12. Veillette, A., Bookman, M. A., Horak, E. M. & Bolen, J. B. *Cell* **56,** 301–308 (1988).
13. Klatzmann, D. *et al. Science* **225,** 59–63 (1984).
14. Dalgleish, A. *et al. Nature* **312,** 763–766 (1984).
15. Sattentau, Q., Dalgleish, A., Weiss, R. & Beverley, P. C. L. *Science* **234,** 1120–1123 (1986).
16. McDougal, J. S. *et al. J. Immun.* **137,** 2937–2944 (1986).
17. McDougal, J. S. *et al. Science* **231,** 382–385 (1986).
18. Maddon, P. *et al. Cell* **47,** 333–348 (1986).
19. Sodroski, J., Goh, W. C., Rosen, C., Campbell, K. & Haseltine, W. A. *Nature* **322,** 470–474 (1986).
20. Lifson, J. *et al. Nature* **323,** 725–728 (1986).
21. Maddon, P. *et al. Cell* **42,** 93–104 (1985).
22. Clark, S. J. *et al. Proc. natn. Acad. Sci. U.S.A.* **84,** 1649–1653 (1987).
23. Peterson, A. & Seed, B. *Cell* **54,** 65–72 (1988).
24. Sattentau, Q. *et al. J. exp. Med.* **170,** 1319–1334 (1989).
25. Williams, A. F., Davis, S. J., He, Q. & Barclay, A. N. *Cold Spring Harb. Symp. quant. Biol.* **54,** 637–647 (1989).
26. Sayre, P. & Reinherz, E. L. *Eur. J. Immun.* **15,** 291–295 (1985).
27. Smith, D. H. *et al. Science* **238,** 1704–1707 (1987).
28. Fisher, R. A. *et al. Nature* **331,** 76–78 (1988).
29. Hussey, R. E. *et al. Nature* **331,** 78–81 (1988).
30. Deen, K. C. *et al. Nature* **331,** 82–84 (1988).
31. Traunecker, A., Luke, W. & Karjalainen, K. *Nature* **331,** 84–86 (1988).
32. Berger, E. A., Fuerst, T. R. & Moss, B. *Proc. natn. Acad. Sci. U.S.A.* **85,** 2357–2361 (1988).
33. Richardson, N. E. *et al. Proc. natn. Acad. Sci. U.S.A.* **85,** 6102–6106 (1988).
34. Clayton, L. K. *et al. Nature* **335,** 363–366 (1988).
35. Mizukami, T., Fuerst, T. R., Berger, E. A. & Moss, B. *Proc. natn. Acad. Sci. U.S.A.* **85,** 9273–9277 (1988).
36. Arthos, J. *et al. Cell* **57,** 469–481 (1989).
37. Landau, N. R., Warton, M. & Littman, D. R. *Nature* **334,** 159–162 (1988).
38. Clayton, L. K., Sieh, M., Pious, D. A. & Reinherz, E. L. *Nature* **339,** 548–551 (1989).
39. Lamarre, D. *et al. Science* **245,** 743–746 (1989).
40. Davis, S. J. *et al. J. molec. Biol.* **213,** 7–10 (1990).
41. Kwong, P. D. *et al. Proc. natn. Acad. Sci. U.S.A.* **87,** 6423–6427 (1990).
42. Harrison, S. C. *et al.* manuscript in preparation.
43. Ryu, S.-E. *et al. Nature* **348,** 419–426 (1990).
44. Garlick, R. L. *et al. AIDS Res. hum. Retrovirus* **6,** 465–479 (1990).
45. Wang, B. C. *Meth. Enzym.* **115,** 90–111 (1985).
46. Bricogne, G. *Acta crystallogr. A* **32,** 832–847 (1976).
47. Jones, T. A. *J. Appl. crystallogr.* **11,** 268–272 (1978).
48. Brünger, A., Kuriyan, J. & Karplus, M. *Science* **235,** 458–460 (1987).
49. Brünger, A. T. *XPLOR Manual, Version 2.1.* (Yale University, New Haven, Connecticut, USA).
50. Stuart, D. & Artymiuk, P. *Acta crystallogr. A* **40,** 713–716 (1985).
51. Crowther, R. A. in *The Molecular Replacement Method* 173–183 (ed. Rossmann, M. G.) (Gordon and Breach, New York, 1972).
52. Huber, R. in *Molecular Replacement. Proceedings of the Daresbury Study Weekend,* 58–61 (SERC, Daresbury, UK, 1985).
53. Sussman, J. L., Holbrook, S. R., Church, G. M. & King, S. H. *Acta crystallogr.* **33,** 800–804 (1977).
54. Bjorkman, P. J. *et al. Nature* **329,** 506–511 (1987).
55. Epp, O. *et al. Biochemistry* **14,** 4943–4952 (1975).
56. Lesk, A. M. & Chothia, C. *J. molec. Biol.* **160,** 325–342 (1982).
57. Colman, P. M. *Adv. Immun.* **43,** 99–132 (1988).
58. Kabat, E. *et al. Sequences of Proteins of Immunological Interest* (US Dept. of Health & Human Services, Bethesda, Maryland, 1987).
59. Hunkapiller, T. & Hood, L. *Adv. Immun.* **44,** 1–63 (1989).
60. Tramontano, A., Chothia, C. & Lesk, A. M. *J. molec. Biol.* **215,** 175–182 (1990).
61. Richardson, J. & Richardson, D. C. in *Prediction of Protein Structure and the Principles of Protein Conformation* (ed. Fasman, G.) (Plenum, New York, 1989).
62. Kirszbaum, L. *et al. J. Immun.* **142,** 3931–3936 (1989).
63. Holmgren, A. & Bränden, C. I. *Nature* **342,** 248–251 (1989).
64. Brodsky, M. H., Warton, M., Myers, R. M. & Littman, D. R. *J. Immun.* **144,** 3078–3086 (1990).
65. Ashkenazi, A. *et al. Proc. natn. Acad. Sci. U.S.A.* **87,** 7150–7154 (1990).
66. Blum, M., Metcalf, P., Harrison, S. C. & Wiley, D. C. *J. appl. Crystallogr.* **20,** 235–242 (1987).
67. Terwilliger, T. C. & Eisenberg, D. *Acta crystallogr. A* **39,** 813–817 (1983).
68. Kabsch, W. & Sander, C. *Biopolymers* **22,** 2577–2637 (1983).

ACKNOWLEDGEMENTS. We thank D. C. Wiley and many members of our laboratories for discussion and advice; N. B. Jansen, H. F. Meyer, D. P. Palermo, P. A. Wells and others at the Upjohn Co. for preparation of CD4 fragments; W. Weirenga, L. Post and H. Einspahr, also of Upjohn, for support and encouragement; W. Kabsch and K. C. Holmes, for advice on phase combination; W. Hendrickson and colleagues for exchanging manuscripts before submission; and R. Helimiss for help with illustrations. The work was supported by the NIH (grants to S.C.H., E.L.R. and D. C. Wiley). Atomic coordinates have been deposited in the Brookhaven database, accession number 2CD4.