Acta Cryst. (1996). D**52**, 369–379

Structure of a Functional Fragment of VCAM-1 Refined at 1.9 Å Resolution

JIA-HUAI WANG,[a]* THILO STEHLE,[a,b] BLAKE PEPINSKY,[c] JIN-HUAN LIU,[a] MICHAEL KARPUSAS[c] AND LAURELEE OSBORN[c]

[a]*Department of Molecular and Cellular Biology, Harvard University, Cambridge, Massachusetts 02138, USA,* [b]*Howard Hughes Medical Institute, Harvard University, Cambridge, Massachusetts 02138, USA, and* [c]*Biogen, Inc., 14 Cambridge Center, Cambridge, Massachusetts 02142, USA*

(*Received 20 June 1995; accepted 4 September 1995*)

Abstract

The crystal structure of the functional amino-terminal two-domain fragment of human vascular cell adhesion molecule 1 (VCAM-1) has been determined at 1.9 Å resolution. The crystals contain two copies of the molecule in the asymmetric unit. The structure was solved by multiple isomorphous replacement, using lead and selenium derivatives. Anomalous scattering had to be used to resolve the phase ambiguity of a lead derivative. Since the selenium derivative has very small isomorphous differences, the local scaling algorithm had to be used to obtain an interpretable difference Patterson map. The initial phases were improved by non-crystallographic averaging, solvent flattening and histogram matching. The structure has been refined to a crystallographic R factor of 20.4% (15–1.9 Å, $F \geq 3\sigma$) and consists of two Ig domains (D1 and D2). The angle between these domains differs by 12° between the two copies of the molecule in the crystallographic asymmetric unit, demonstrating that some movement is possible at the interface. In the amino-terminal domain D1 there is an 'extra' disulfide bond, in addition to the conserved cross-sheet disulfide bond, at the top of the molecule. This bond, a hallmark of the integrin-binding subclass of Ig superfamily proteins, makes the top of this domain very compact. The feature that projects most prominently from D1 is the *CD* loop, near the base of the domain. The key residue for integrin binding, Asp40, is located in this loop and is easily accessible.

1. Introduction

VCAM-1 (vascular cell adhesion molecule 1) is a versatile cell-surface adhesion molecule, originally discovered on endothelial cells stimulated by inflammatory cytokines. It belongs to a subclass of the Ig superfamily which also includes the ICAM's (intercellular cell adhesion molecules 1, 2 and 3) and MAdCAM-1 (mucosal addressing cell adhesion molecule 1) (Briskin, McEvoy & Butcher, 1993; Fawcett *et al.*, 1992; Staunton, Dustin & Springer, 1989; Staunton, Marlin, Stratowa, Dustin & Springer, 1988; Vazeux *et al.*, 1992), and which is characterized by unique structural and functional features. Structurally, each member of this subclass has an 'extra' pair of cysteine residues, compared to classic Ig-like modules, in the domains directly involved in counter-receptor binding. Functionally, all members of this subclass bind members of the integrin family, a large group of heterodimeric metal-ion-dependent proteins best known for their roles in development and in leukocyte and platelet adhesion in the vasculature (Hynes, 1992). Both ICAM-1 and VCAM-1 have been subverted to serve as viral receptors by members of the picornavirus family. ICAM-1 is a receptor for human rhinovirus, the cause of the common cold (Greve *et al.*, 1989), and VCAM-1 is recognized by murine encephalomyocarditis virus (Huber, 1994). In addition to its role in leukocyte recruitment, VCAM-1 has been implicated in a growing list of developmental processes, including hematopoiesis (Miyake *et al.*, 1991), muscle and nerve development (Rosen *et al.*, 1992; Sheppard, McGuillan, Iademarco & Dean, 1995) and cardiac and placental formation (Gurtner *et al.*, 1995; Kwee *et al.*, 1995; Yang, Rayburn & Hynes, 1995).

The extracellular part of the major form of VCAM-1 consists of seven Ig domains (Hession *et al.*, 1991; Polte, Newman & Gopal, 1990). Of these, both domains 1 and 4 bind to the α4-integrins $\alpha4\beta1$ (VLA-4) and $\alpha4\beta7$ (Elices *et al.*, 1990; Kilger *et al.*, 1995; Osborn, Vassallo & Benjamin, 1992; Vonderheide & Springer, 1992), which are present on most types of circulating leukocytes, with the exception of the abundant neutrophils active in acute inflammation. α4-bearing leukocytes (lymphocytes, monocytes, natural killer cells, eosinophils and basophils) are recruited from the bloodstream to sites of infection or inflammation by interaction with VCAM-1. This step usually occurs following a preliminary weak interaction with one of the carbohydrate-recognizing molecules known as selectins, resulting in rolling on the wall of the blood vessel. It has recently been shown that VCAM-1 itself can mediate rolling (Alon *et al.*, 1995; Berlin *et al.*, 1995), as well as the tight adhesion previously hypothesized to be the sole function of CAM/integrin interactions (reviewed by Springer, 1994).

This versatility is presumably related to the integrin's ability to change its affinity (or avidity) for its receptor quickly, in response to signals from inside or outside of the leukocyte (reviewed by Hemler & Lobb, 1995).

The crystal structure of the ligand-binding N-terminal two-domain fragment of VCAM-1 (VCAM-D1D2) is the first to be determined within its subclass (Jones *et al.*, 1995; Wang *et al.*, 1995). The structure shows the location of a key region involved in integrin binding, and offers a good explanation as to why a cyclic peptide can inhibit this interaction (Wang *et al.*, 1995). In this report, we give details on the structure determination and refinement of the molecule at 1.9 Å. Based on the refined model, we discuss the architecture of the domains and compare them with Ig domains. We also give reasons why the rather large domain interface can accomodate substantial movement. Finally, we discuss the structure of the integrin-binding loop and the way it is stabilized by internal interactions as well as the interaction with a neighboring loop.

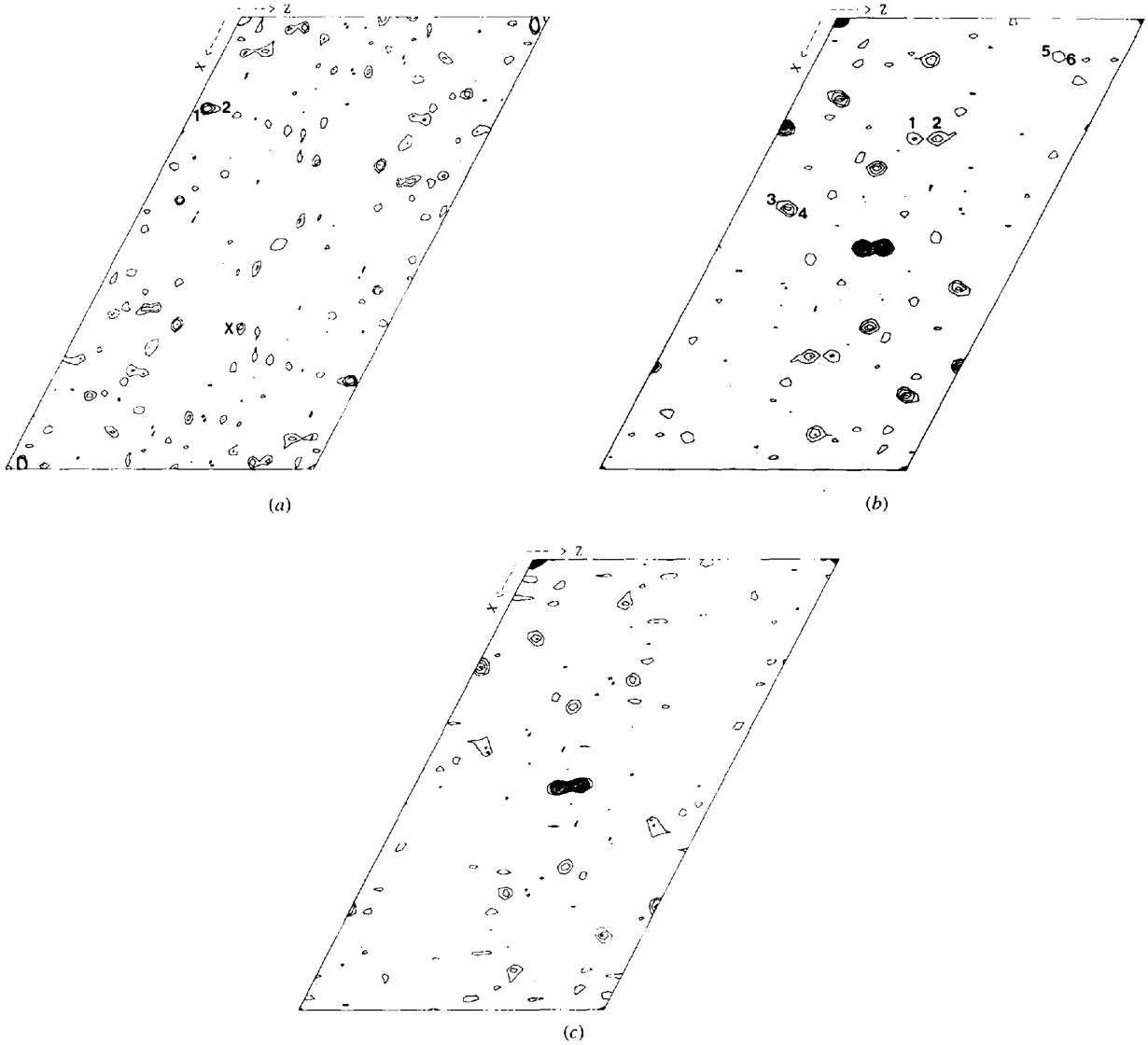

Fig. 1. Harker sections ($y = 0$) of difference Patterson maps. (*a*) Anomalous difference Patterson map for the lead derivative (5.0–3.5 Å). Contours are given in 1σ intervals, starting at 2σ. The sites are labeled 1 and 2, the cross peak is marked with X. (*b*) Difference Patterson map for the selenomethionyl derivative (12–3.5 Å) with local scaling applied to the derivative data set. The six Harker peaks are labeled. Contours are given in 2σ intervals, starting at 2σ. The remaining peaks above 4σ are double-weighted non-Harker peaks. The huge peaks near the center of the unit cell represent the translation vector relating atoms in two independent molecules in the asymmetric unit. Because sites 1, 2, 5 and 6 have similar y coordinates, many non-Harker peaks fall onto the ($y = 0$) section. (*c*) The same section of the difference Patterson map for the selenomethionyl derivative, produced with least-squares scaling (program *RSTATS*, *CCP4* program package). Contours are at the same level as in Fig. 1(*b*). The map is much noisier, and several peaks seen in Fig. 1(*b*) are absent.

2. Experimental

2.1. *Crystallization, data collection and data processing*

The crystals used in this structure analysis were grown as previously described and belong to space group $C2$, with unit-cell dimensions $a = 122.09$, $b = 48.86$, $c = 73.44$ and $\beta = 117.40°$ (Wang, Pepinsky, Karpusas, Liu & Osborn, 1994). Two molecules are present in the crystallographic asymmetric unit. Data were collected from flash-frozen crystals at 118 K on a Nicolet/Siemens multiwire area detector (Siemens, Inc.), and integrated and reduced using *BUDDHA* (Blum, Metcalf, Harrison

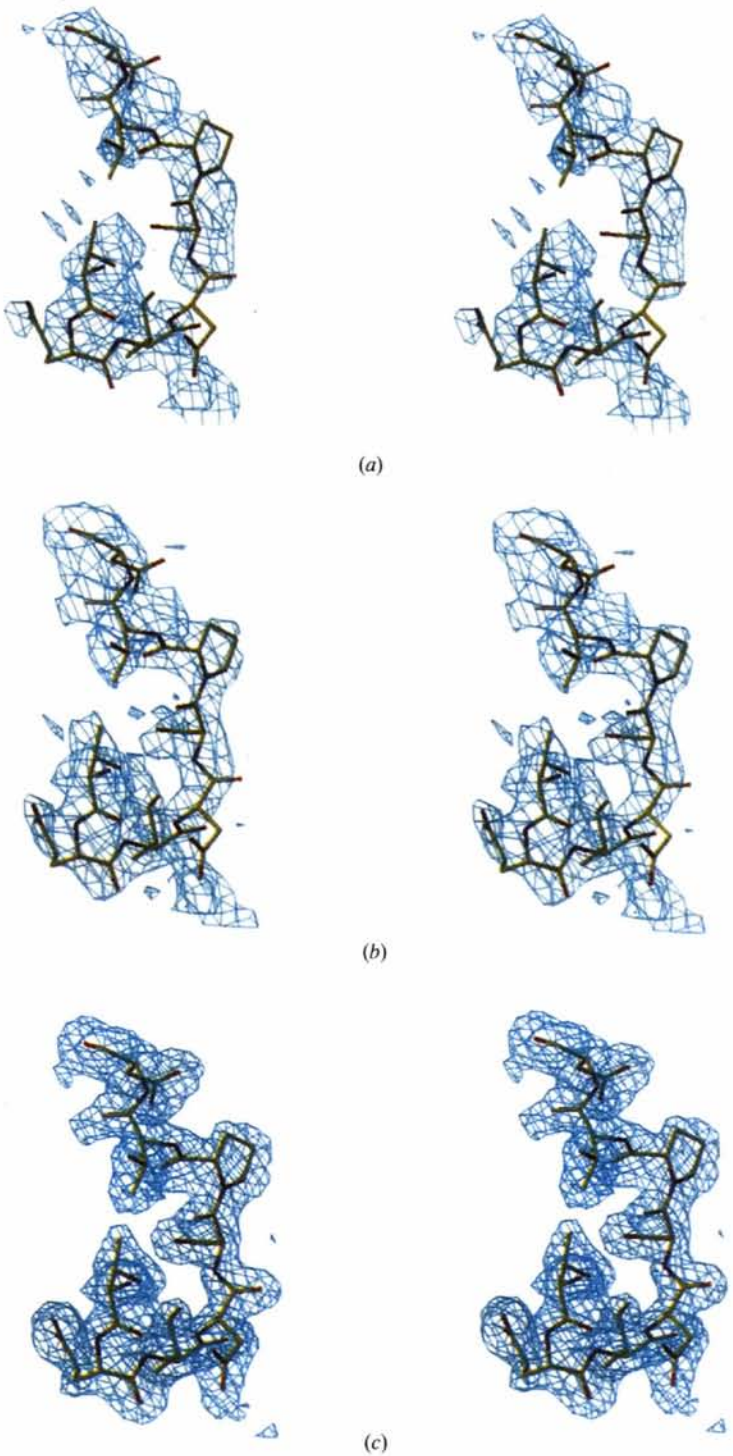

Fig. 2. Stereoviews of the electron-density maps used at different stages of the structure determination. All contours shown are at 1σ. The coordinates (residues 37–44) show parts of the *CD* loop of the refined model. (*a*) MIR-map at 2.5 Å resolution. (*b*) $F_o \exp(i\alpha_{comb})$ map at 2.5 Å resolution, after a final round of density modification with program *DM* and using the final envelope. (*c*) Final $(2F_o - F_c)\exp(i\alpha_{calc})$ map phased at 1.9 Å resolution. The views have been generated with *O* (Jones, Zhou, Cowan & Kjeldgaard, 1991).

& Wiley, 1987) and the *CCP*4 program package (Collaborative Computing Project, Number 4, 1994). An extensive search for heavy-atom derivatives was carried out by collecting portions of data to 3.5 Å resolution. Several of these data sets showed significant isomorphous differences, ranging from 11 to 35%. However, the difference Patterson maps calculated with a complete data set were featureless, except in one case: trimethyllead acetate at 10 m*M*, where a single elongated peak at 4σ was seen on the Harker section. The native Patterson map had shown earlier that the two independent molecules are related by a translation vector (0.49, 0, 0.51) perpendicular to the crystallographic dyad (Wang *et al.*, 1994). The elongated peak in the difference Patterson map was, therefore, a superposition of the two Harker peaks at low resolution. Because the two heavy-atom sites have roughly the same y coordinates, use of anomalous scattering was required to resolve the phase ambiguity in the monoclinic space group. To obtain good anomalous differences to 3.5 Å using a conventional X-ray source and area detector, the following measures proved to be critical. First, a single frozen crystal allowed about seven measurements on average for each anomalous difference without noticeable intensity decay, which significantly increased the signal-to-noise ratio. Second, the raw data were grouped in batches in a way that odd numbered batches are at a detector setting angle of ($\Phi \rightarrow \Phi + \Delta\Phi$) and even numbered batches are at ($180° + \Phi \rightarrow 180° + \Phi + \Delta\Phi$). The program *AGROVATA* was modified to calculate anomalous differences from Bijvoet mates by taking the differences from adjacent batches only. This procedure substantially reduced the intensity difference due to absorption. The Harker section of the anomalous difference Patterson map at 5–3.5 Å is shown in Fig 1(*a*). Even at this resolution range, the peaks are clearly seen.

A second derivative data set was collected from a crystal grown from selenomethionine-substituted VCAM-D1D2. Special attention was paid when collecting and processing the 2.5 Å selenomethionyl data set, in order to obtain small isomorphous differences. (i) The unit-cell dimensions of several flash-frozen crystals were tested, and the crystal selected for data collection had the smallest difference (0.2% or less) in unit-cell parameters compared to the native crystal. (ii) The program *DSCALEAD* (Rould, 1996) was used to apply local scaling between the native and derivative data. Taken together, the two procedures significantly improved the signal-to-noise ratio in the difference Patterson map. The map was strikingly clear and showed essentially all of the Harker and cross peaks for the six selenium sites. The Harker section of this map is shown in Fig. 1(*b*). By comparison, Fig 1(*c*) shows the same section of a difference Patterson map produced with conventional least-squares scaling of the selenomethionyl derivative (*CCP*4 program *RSTATS*). The local scaling algorithm is especially effective for derivatives like selenomethione, which have small but isomorphous differences. The isomorphous difference for the selenomethionyl derivative is only 8.7% to 2.5 Å, whereas the R_{sym} values for the native and the derivative data sets in the same resolution range are 5.0 and 6.7%, respectively. A difference-Fourier map of the selenomethionyl derivative, phased using single isomorphous replacement plus anomalous scattering from the lead derivative, fixed the heavy-atom configuration and the relative origin of the two heavy-atom derivatives. Heavy-atom coordinates were initially refined with *HEAVY* (Terwilliger & Eisenberg, 1983) at 12–3.5 Å. Later, the maximum-likelihood phase refinement as implemented in *MLPHARE* (*CCP*4 program package) was applied to compute a multiple isomorphous replacement (MIR) map at 12–2.5 Å (single isomorphous replacement between 3.5 and 2.5 Å). The MIR statistics have been reported previously (Wang *et al.*, 1995).

2.2. *Phase improvement*

Despite the efforts described above, the MIR map was of marginal quality. The translation vector, which relates the two independent molecules and is perpendicular to the crystal dyad, generates a pseudo-dyad parallel to the crystal dyad at roughly ($x = 1/4$, $z = 1/4$). We, therefore, averaged the map using the local twofold symmetry (program *SKEWPLANE*, *CCP*4 program package). In the averaged map, a molecular boundary encompassing two abutting domains was clearly detectable and allowed us to draw a molecular envelope using *ENVEDIT* (Blum, 1990). Density modification (program *DM*; Cowtan, 1994) was then used to apply solvent flattening, histogram matching and twofold averaging of the MIR map. The initial resolution range was 15–4.5 Å; the resolution was gradually extended to 2.5 Å in 15 cycles. We allowed the transform to be refined and tightened the envelope to match the 44% solvent content of the crystals. The map calculated with the phases from *DM* was, after further averaging, readily interpretable for the major part of the structure, and in particular for D1.

2.3. *Model building and crystallographic refinement*

An initial model was built for D1 and the framework of D2 using *QUANTA* (Molecular Simulations, Inc.) and *O* (Jones, Zhou, Cowan & Kjeldgaard, 1991). The model was then transformed back to the crystal frame to generate the two independent molecules, *A* and *B*. At this stage, conformational differences between the loops in D2 of the two molecules were apparent. The model was then subjected to rigid-body refinement using *X-PLOR* (Brünger, Kuriyan & Karplus, 1987). At the outset, 10% of the data were set aside to monitor the 'free' R factor (Brünger, 1992). The bulk-solvent correction method as implemented in *X-PLOR* (Version 3.1) was used for refinement. Rigid-body refinement was performed at

Table 1. *Crystallographic refinement*

Round	Resolution range (Å)	No. of water molecules	No. of reflections	Simulated annealing	Starting R factors R_{free} (%)	Starting R factors R_{work} (%)	Final R factors R_{free} (%)	Final R factors R_{work} (%)
1	15–2.7	0	10473	2000–300 K	49.3	47.1	39.0	27.8
2	15–2.3	0	16179	3000–300 K	46.7	45.6	37.4	26.6
3	15–2.3	24	16179	3000–300 K	37.9	31.0	36.4	26.0
4	15–2.3	67	16179	3000–300 K	38.1	29.9	34.7	24.4
5	15–2.2	178	18112	500–300 K	36.5	28.1	35.2	24.3
6	15–2.0	263	22083		34.9	25.2	32.6	22.0
7	15–2.0	277	22083		32.1	23.1	29.4	19.9
8	15–1.9	303	24026		32.3	26.0	28.9	20.4

Refinement was carried out with *X-PLOR* (Brünger, Kuriyan & Karplus, 1987). A typical round consisted of 100–150 cycles of energy minimization, 15 steps of temperature-factor refinement, a simulated-annealing run ('slowcool' protocol), which was omitted in the later stages, followed by another 100 cycles of minimization and 15 steps of temperature-factor refinement. The solvent correction method and a ($F > 3\sigma$) cut off were used throughout the refinement. Temperature-factor refinement was carried out using a lower resolution cut off of 6.0 Å. In rounds 1–6, non-crystallographic symmetry restraints were used.

10–4.5 Å, treating each domain as a single rigid body and refining only one domain at a time while keeping the others fixed. This procedure significantly improved the agreement between model and map, in particular for molecule *B*. The rigid-body refinement revealed that D1 and D2 of molecules *A* and *B* are related by different transformation matrices. The model coordinates were then used to generate two separate envelopes for the two domains using *MAMA* (Kleywegt & Jones, 1994). Further density modification using these envelopes and the new transforms improved the density map for D2 considerably. Three rounds of iterative runs of density modification and model building yielded about 95% of the structure.

Coordinates were refined using the energy-minimization and simulated-annealing options in *X-PLOR* (Brünger *et al.*, 1987). The resolution range for the first refinement round was 15–2.7 Å; the resolution was extended stepwise to 15–1.9 Å for the final round. The model was checked with $(2F_o - F_c)\exp(i\alpha_c)$ and $(F_o - F_c)\exp(i\alpha_c)$ maps. At several stages, the quality of the model was also checked with omit maps. Water molecules were added with ARP (Lamzin & Wilson, 1993), using a minimum density of 3.5σ in the $(F_o - F_c)\exp(i\alpha_c)$ map and possible hydrogen-bond formation as acceptance criteria. After each round of refinement, the water molecules were checked against $(2F_o - F_c)\exp(i\alpha_c)$ maps: water molecules with electron density below 1.0σ were deleted from the coordinate list. All water molecules included in the final model are hydrogen bonded (donor–acceptor distance < 3.5 Å) to protein atoms or to other water molecules. The waters are numbered consecutively, starting at 400. The stereochemical quality of the model was monitored using *PROCHECK* (Laskowski, MacArthur, Moss & Thornton, 1993). Temperature factors were only restrained in the first seven refinement rounds. In the first six rounds, non-crystallographic symmetry restraints were used. Earlier rounds included a simulated-annealing run (typically 3000–300 K 'slow-cooling' protocol), which was omitted in the later stages.

Table 2. *Statistics on crystallographic data*

Resolution range (Å)	R_{sym}* (%)	Redundancy	Completness (%)	R_{free}† (%)	R_{work}† (%)
15.0–6.0	3.0	4.6	99.2	25.1	16.5
6.0–4.3	3.0	4.7	99.9	22.9	14.3
4.3–3.5	4.2	5.0	99.5	25.5	14.7
3.5–3.0	6.0	5.0	98.8	29.5	18.6
3.0–2.7	8.9	4.8	95.8	30.6	23.1
2.7–2.5	12.7	4.4	91.4	33.9	25.0
2.5–2.3	18.7	4.1	88.4	33.4	26.5
2.3–2.1	23.8	3.6	82.4	32.8	29.1
2.1–2.0	22.3	1.6	61.9	32.9	27.9
2.0–1.9	36.3	1.4	54.9	32.1	28.9
15.0–1.9	5.9	3.9	82.9	28.9	20.4

*$R_{sym} = \sum_h \sum_i |I_{hi} - I_h| / \sum_{hi} I_{hi}$. †$R_{free} = R_{work} = \sum_i |F_{obs} - F_{calc}| / \sum_i F_{obs}$. 'Free' and 'work' refer to the free and working data sets, respectively.

The refinement reached convergence after eight rounds. The course of the refinement is shown in Table 1. Data statistics are given in Table 2. Fig. 2 shows examples for the MIR map, the density-modified, solvent-flattened and averaged map produced with *DM*, and the final $(2F_o - F_c)\exp(i\alpha_c)$ map at 1.9 Å resolution.*

3. Results and discussion

3.1. *Model accuracy*

The final model consists of all 195 amino acids for each of the two copies of VCAM-D1D2 in the asymmetric unit and 303 water molecules. The final crystallographic R factors for the free and the working data sets are 28.9 and 20.4%, respectively (15–1.9 Å; $F > 3\sigma$, only 3% reflections were removed by this criterion). The model has good geometry; the r.m.s. devi-

* Atomic coordinates and structure factors have been deposited with the Protein Data Bank, Brookhaven National Laboratory (Reference: 1VSC, R1VSCSF). Free copies may be obtained through The Managing Editor, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England (Reference: LI0212). At the request of the authors, the atomic coordinates and structure factors will remain privileged until 1 June 1996.

ations for bond lengths and bond angles are 0.016 Å and 2.1°, respectively. Fig 3(*a*) shows the real-space density correlation (Jones *et al.*, 1991) for the main chain of both molecules. Most main-chain and side-chain atoms are well defined by electron density in the $(2F_o - F_c)\exp(i\alpha_c)$ map. Exceptions are residues 142–148 of molecule *B* (*C'E* loop), residues 103–109 (*AB* loop) and 160–167 (*EF* loop) of molecule *A* and the last two C-terminal residues of each molecule. With the exception of Glu76, the main-chain torsion angles of all residues fall into favorable regions of the Ramachandran plot (Ramachandran & Sasisekharan, 1968). Glu76 is located at the $(i + 1)$ position of a β-turn. The coordinate error of the model is around 0.25 Å, as judged from a Luzzati plot (Luzzati, 1952).

A temperature-factor plot is shown for both molecules in Fig. 3(*b*). The temperature factors agree very well with the density correlation coefficients in Fig. 3(*a*): regions with low temperature factors have a high density correlation coefficients and *vice versa*. The average temperature factors for D1 are 23.0 Å^2 (molecule *A*) and 22.3 Å^2 (molecule *B*). For D2, the values are 43.4 Å^2 (molecule *A*) and 33.6 Å^2 (molecule *B*). Thus, D2 has a higher mobility than D1 in both molecules. In addition, D2 is more mobile in molecule *A* than in molecule *B*.

The nature of the crystal contacts enables us to understand the differences in mobility among the domains. The molecules pack in a layer-like fashion, with their longest dimension lying in the *ac* plane (perpendicular to the crystallographic dyads along axis *b*) of the unit cell. A view of a part of such a layer is given in Fig. 4. The generally higher mobility of D2 can be rationalized by its overall shape and its interactions with neighboring molecules: the domain has several protruding loops which mainly participate in crystal lattice formation. D1, on the other hand, has a very compact structure and the crystal contacts involve residues which are less mobile. A good example is the crystal contact which involves the main chain (strand *A*) of D1 and the *FG* loop of D2 (see Fig. 4). The higher mobility of D2 in molecule *A*, as compared to molecule *B*, appears to be a result of fewer crystal contacts involving that domain.

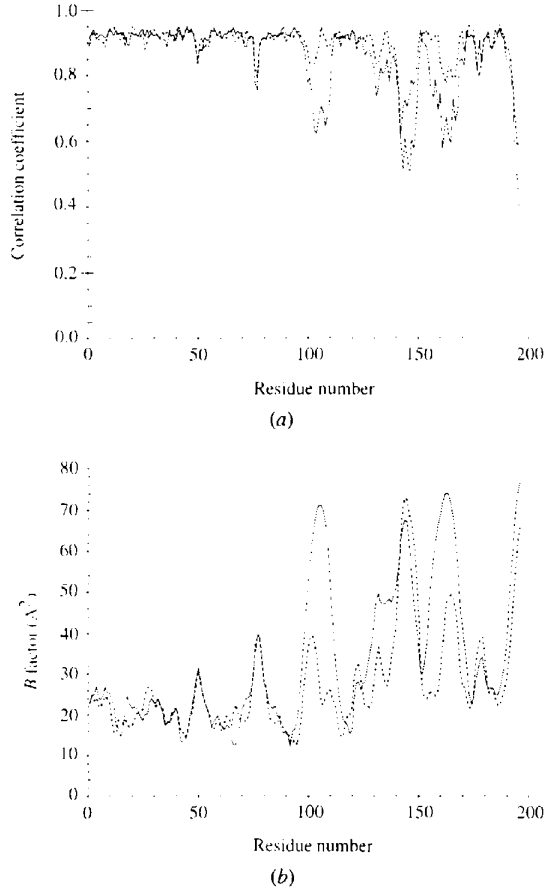

Fig. 3. Real-space density correlation coefficents (*a*) and temperature factors (*b*) of the main-chain atoms of refined VCAM-D1D2 as functions of the residue number. Molecule *A*, solid lines; molecule *B*, broken lines. The real-space correlation coefficient has been calculated with *O* (Jones, Zhou, Cowan & Kjeldgaard, 1991) using the final $(2F_o - F_c)\ \exp(i\alpha_c)$ map.

3.2. *Overall structure*

The VCAM-D1D2 molecule consists of two concatenated Ig folds. Fig. 5 gives a stereoview of the Cα tracing of the molecule. Each of the two domains has seven β-strands, forming a β-barrel, which run approximately perpendicular to the crystallographic dyad. This orientation probably accounts for strong (4–5σ) features

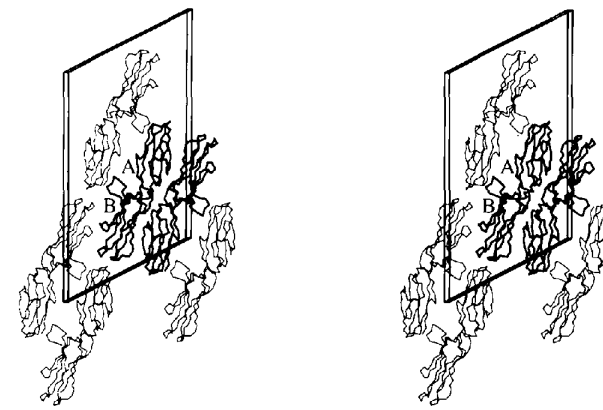

Fig. 4. Stereoview of the layer-like crystal packing of VCAM-D1D2. The reference molecules *A* and *B* are shown in thick lines, the crystallographically related molecules in thin lines. The unit cell is displayed as well. The layer is mainly generated by the extensive contact between the reference molecules which are related by non-crystallographic symmetry, and by a crystal contact involving strand *A* of D1 and the *FG* loop of D2. In this contact, residues 1–3 of strand *A* and residues 178–180 of the *FG* loop form a short antiparallel β-sheet. Note how the *FG* loop is pulled away from D2 to accommodate this interaction. The view was prepared using *MOLSCRIPT* (Kraulis, 1991).

on the Harker section of the native Patterson map, which we believe to be the accumulated vectors between β-strands related by a twofold symmetry axis. In addition to the extensive β-sheet structure, there are three short segments of 3_{10} helix: residues 62–65, 121–124 and 161–164. The first such helix is located in the *EF* loop of D1, the second and third are in the *BC* and *EF* loops of D2. There are three *cis*-prolines: Pro7, Pro60 and Pro120. The residue preceding Pro7 is the last in strand *A*, and its carbonyl is hydrogen bonded to strand *B*. The chain then kinks over, joining the opposite sheet. Pro60 is the second residue on the *EF* loop of D1, where the chain also kinks sharply. Pro120 is located at the interdomain interface.

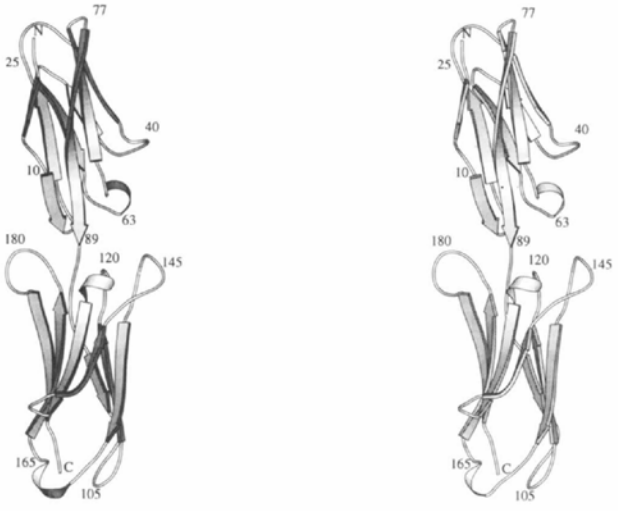

Fig. 5. Stereo drawing of VCAM-D1D2 (molecule *A*). Some residues are labeled. This figure was prepared using *MOLSCRIPT* (Kraulis, 1991).

3.3. *Domain structure*

Ig folds have been classified into 'sets', based on the specific arrangements of their β-strands. VCAM-D1 (residues 1–89), falls into a recently identified intermediate or I-set (Harpaz & Chothia, 1994). The domain has the strand sequence characteristic of Ig constant domains (C1-set), but it also has a number of detailed features characteristic of Ig variable domains (V-set). A V-set domain contains strands *ABED* in one sheet and $A'GFCC'C''$ in the other, while VCAM-D1, like a C1-set domain, lacks strands C' and C''. The conformation of the *AB* and *EF* loops in VCAM-D1, however, follows the conserved V-set pattern, including a particular main-chain hydrogen bond between these two loops. VCAM-D1 also has a *cis*-proline in strand *A* (Pro7), where the chain kinks before entering strand A'. This is another V-set characteristic. Of the 89 Cα atoms in D1, 77 can be superimposed onto the corresponding atoms of a representative V-set domain, Ig-REI, (Epp, Lattman, Schiffer, Huber & Palm, 1975) with a remarkably small r.m.s. deviation of 1.3 Å; by contrast, only 56 Cα atoms correspond closely (r.m.s. deviation of 1.2 Å) to those of a typical C1-set structure, the light-chain Fab fragment KOL (Marquart, Deisenhofer, Huber & Palm, 1980). Therefore, D1 is probably classified best as a truncated V-set or I-set as defined by Harpaz & Chothia (1994). Based on their three-dimensional structures, an alignment of VCAM-D1 with Ig-REI and the N-terminal domains of CD4 and CD2 is given in Fig. 6.

The top end of a V-set domain corresponds to the region of the hypervariable loops in an antibody mol-

```
                A       A'      B       C
VCAM-D1   FKIETTPESRYLAQIGDSVSLTCSTTG--CESPFFSWRTQ----ID----------------
REI       DIQMTQSP-SSLSASVGDRVTITCQASQ--DIIKYLNWYQQTPGKAPKLLIYE-A---SNLQAG
CD4-D1          KKVVLGKKGDTVELTCTASQ--KKSIQFHWKNS----NQIKILGNQGSF-LTKGPS
CD2-D1        RDSGTVWGALGHGINLNIPNFQMTDDIDEVRWER-----GSTLVAEFKRKMKPFLK--

                D       E       F       G
VCAM-D1   -SPLNGKVTNEG---TTSTLTMNPVSFGNEHSYLCTATC-ESRKL-EK-GIQVEIY
REI       -VPSRFSGSGSG---TDYTFTISSLQPEDIATYYCQQYQ-SLPYT-FGQGTKLQIT
CD4-D1    KLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVE-----D-QKE-EVQLLVF
CD2-D1    --SGAFEILA------NGDLKIKNLTRDDSGTYNVTVYSTNGTRILDK-ALDLRIL
```

ig. 6. Sequence alignment, based on three-dimensional structures. Shown are the sequences of the N-terminal domain of VCAM-1, the variable domain of Ig-REI (Epp, Lattman, Schiffer, Huber & Palm, 1975), and the N-terminal domains of CD4 (Wang *et al.*, 1990) and CD2 (Jones, Davis, Williams, Harlos & Stuart, 1992). The β-strands of VCAM-D1 are marked with solid lines. The Cα atoms of 77 residues in VCAM-1 and Ig-REI can be superimposed on each other with an r.m.s. deviation of 1.3 Å. These residues are highlighted. Every tenth VCAM-1 residue is marked by a dot.

ecule. In the case of VCAM-D1, this 'hypervariable' end is closed by a disulfide bond between Cys28 at the beginning of the *BC* loop and Cys75 at the beginning of the *FG* loop; this bond, a hallmark of the VCAM/ICAM Ig subclass, makes the tip of D1 very compact. It has a left-handed spiral conformation as opposed to the right-handed conformation of the central disulfide bond between Cys23 and Cys71. With the distal end of D1 closed, the most protruding feature is the *CD* loop near the base of the domain.

VCAM-D2 (residues 90–195) can best be described as a C2-set Ig fold (Williams & Barclay, 1988). Strands *ABE* form one β-sheet and strands GFCC′ form the other. The domain is significantly larger than the average C2 set, and the residues that account for its increased size are found mostly in the *C′E* and the *FG* loops. These loops are 11–12 residues and 6–8 residues longer, respectively, than the corresponding loops in the structures of CD4-D2, CD4-D4 or CD2-D2. Loop *C′E* is rather flexible; loop *FG* is well ordered but it has, in our structure, a conformation that is stabilized by a crystallographically related molecule (see Fig. 4). We expect the *FG* loop to be more mobile in solution. Both loops extend beyond the body of D2 and flank the base of D1 (Fig. 5).

Strand *A* of D2 begins at residue 96, following a segment which kinks at Pro92 and Pro95. Despite the lack of secondary structure of the peptide segment 90–95, it is nevertheless structurally well integrated into D2. There are two main-chain hydrogen bonds between Ser90 and Tyr119, and as a result of these interactions, Pro120 is constrained in a *cis*-configuration. We note that the interdomain link in VCAM-D1D2 is quite different to the continuous β-strand connector between D1 and D2 in CD4-D1D2 (Ryu *et al.*, 1990; Wang *et al.*, 1990) or D3 and D4 in CD4-D3D4 (Brady *et al.*, 1993).

3.4. *Domain interface*

We have previously shown that, when the two copies of the VCAM-D1D2 molecules are superimposed based only on the atoms of D1, the orientation of D2 of the two molecules differs by 12°, and we have designated residue 89 as the 'pivot' that mainly accommodates this difference (Wang *et al.*, 1995). Jones *et al.* (1995) reported a 6° difference for a similar calculation using the two independent VCAM-D1D2 molecules present in their crystals. Table 3 lists the main-chain conformation angles of residues 88–91 for molecules *A* and *B*, showing that the largest differences are at residue 89. The difference appears to be related to the molecular packing in the crystal, which creates a significantly different environment for Tyr89. In molecule *B*, the side chain of Tyr89 forms a hydrogen bond to a crystallographically related molecule, whereas, in molecule *A*, it does not.

In spite of the difference in the relative orientation of D1 and D2 in the two molecules *A* and *B* present in the crystal, the domain interfaces are rather similar. The interdomain contact is largely defined by the *AB* and *EF* loops of D1 and the *BC* and *C′E* loops of D2. The extensive *FG* loop of D2 is only weakly involved in the interface, but since this loop is pulled

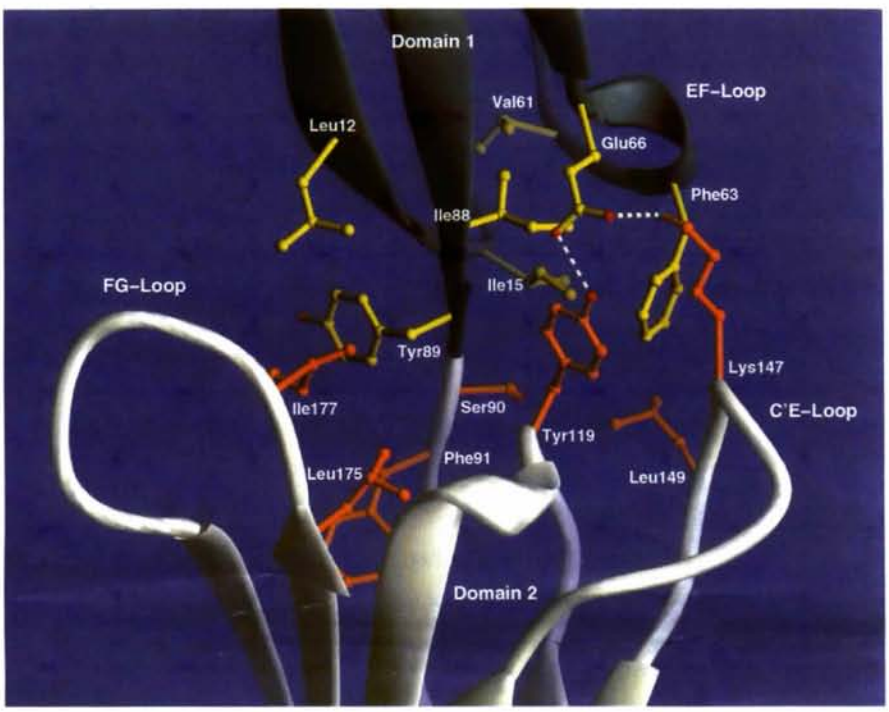

Fig. 7. View of the interdomain interface of VCAM-D1D2. Residues which primarily form the interface are shown in atomic detail. The observed 12° angle difference between D1 and D2 in the two molecules *A* and *B* demonstrates that one domain can move to a certain degree with respect to the other. In the view shown here, that movement would be perpendicular to the plane of the paper, along the length of the cradle generated by the *C′E* and *FG* loops of D2. These two loops seem to guide the movement in one direction. The mainly hydrophobic nature of the domain interface accommodates the movement without causing significant structural changes. The representation has been produced with *RIBBONS* (Carson, 1987).

Table 3. *Torsion angles of residues connecting D1 and D2*

Residue	Molecule *A* φ (°)	Molecule *A* ψ (°)	Molecule *B* φ (°)	Molecule *B* ψ (°)
Ile88	−130	155	−128	150
Tyr89*	−151	163	−131	154
Ser90	−146	128	−151	121
Phe91	−158	59	−154	63

*Hydrogen bond (Tyr89—OH···Thr159—OG1) to crystallographically related molecule in *B* but not in *A*.

away from the domain by a crystal contact (see Fig. 4), it could interact more closely with the base of D1 in solution. The interface can be subdivided roughly into two hydrophobic patches, which are separated by the residues linking the two domains. On one side, the hydrophobic patch is formed by the side chains of Ile15, Val61, Phe63 and Ile88 from D1 and Tyr119, Leu149 and the Cβ atom of Ser90 from D2. On the other side, a second patch is formed by the side chains of Leu12 and Tyr89 from D1 and Phe91, Leu175 and Ile177 from D2. There is only one interdomain hydrogen bond, between Tyr119 OH and Glu66 OE1, present in both copies of the interface. The interface of molecule *A* is further stabilized by a salt bridge between Lys147 and Glu66, not present in molecule *B*. A view into the interface is given in Fig. 7.

A comparison of the interfaces in the two molecules shows that the total buried surface areas are 816 and 775 Å^2 for molecules *A* and *B*, respectively (calculated with *SURFACE*; Collaborative Computational Project, Number 4, 1994). Since the angles between the two domains differ in the two molecules, the interface must accommodate to the shift. The primarily hydrophobic nature of the interface may account for this flexibility. The two hydrophobic patches act as 'cushions' on which the domains seem to be able to rock back and forth. The interface is flanked on two sides by the *C'E* and *FG* loops of D2, giving it a cradle-like appearance, similar to that seen in CD4-D1D2 (Garrett, Wang, Yan, Liu & Harrison, 1993). In the case of CD4, it is the *BC* and *FG* loop of D2 that form a cradle into which D1 inserts. No significant movement is possible there. For VCAM-D1D2, the observed movement is along the length of the cradle, rather than perpendicular to it, and the loops on each side of the cradle seem to guide the direction of this movement. The axis that connects the two 'cushions' is roughly perpendicular to the movement.

3.5. *Integrin binding*

We have previously discussed the importance of the N-terminal corner of the *CD* loop of VCAM-D1 and provided evidence that the cyclic peptide Cys-Gln-Ile-Asp-Ser-Pro-Cys can effectively block the adhesion between VCAM-D1D2 and integrin $\alpha 4\beta 1$ (Wang *et al.*, 1995), presumably by mimicking the structure of the peptide segment Thr37-Gln-Ile-Asp-Ser-Pro-Leu. Our refined structure shows that this loop is very well ordered. Its conformation is stabilized by a number of interactions. There are four internal hydrogen bonds: a main-chain hydrogen bond between Thr37 O and Asp40 N in the β-turn, and a network involving the side chains of Thr37 and Ser41 and main-chain atoms of Ile39 and Ser41. The limited conformational freedom of Pro42 adds to the rigidity of the loop. The side chain of Leu43 packs against a hydrophobic patch generated by the side chains of Trp35, Val47, Leu56 and Tyr69, which supports the loop structure as well. There are also a number of polar interactions between this *CD* loop and the spatially adjacent *EF* loop of the same domain. Some hydrogen bonds involving Asn65, His67, Ser68 and Tyr69 of the *EF* loop link the two loops directly, and some interactions are mediated by water molecules (to Glu66 and His67). The interactions involving the *CD* loop are shown in Fig. 8.

The side chain of Asp40 has been shown to be a key residue in integrin binding (Osborn *et al.*, 1994; Vonderheide, Tedder, Springer & Staunton, 1994). It lies at the N-terminal edge of the *CD* loop and points away from it. The side chain is not involved in any intramolecular interaction, and it is, therefore, fully accessible for the counter-receptor. The peptide segment Thr37-Gln-Ile-Asp-Ser-Pro-Leu is conserved in VCAM-D1 of

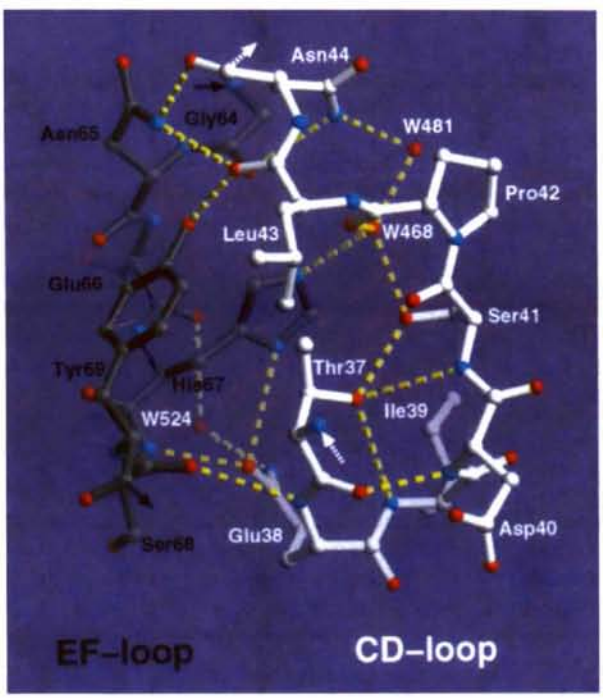

Fig. 8. Interactions involving the *CD* loop of VCAM-D1. Hydrogen bonds within the *CD* loop are shown. Also shown are interactions between the *CD* and *EF* loops, including the hydrogen bonds mediated by water molecules. Hydrogen bonds are represented with broken yellow lines. View generated with *RIBBONS* (Carson, 1987).

all species characterized to date (Hession *et al.*, 1992) and also in the functional human VCAM-D4 domain (Osborn *et al.*, 1992; Vonderheide & Springer, 1992) The conformation of this loop is, therefore, characteristic of VCAM-1 as an α4-integrin-binding molecule. There is evidence that the *EF* loop of VCAM-D1 may also be important for integrin binding (Osborn *et al.*, 1994). The close proximity of the *CD* and *EF* loops and the extent of the interactions between them suggests that the *EF* loop plays a role in stabilizing and orienting the *CD* loop. Strand *F* is several residues longer than in the typical V-set domains, which points to the significance of the *EF* loop.

The *C'E* loop of D2 is also close to the *CD* loop of D1 and hence to the integrin binding site. Jones *et al.* have discussed the possible contribution of this loop to integrin binding (Jones *et al.*, 1995). There is no evidence that this long and highly charged loop plays a role in binding, but its proximity, partial inter-species conservation and the fact that it is even longer and more charged in MAdCAM-1, which shares the ability to bind $\alpha4\beta7$-integrins with VCAM-1, suggests that it may participate. The rocking movement of domains along the cradle formed by the *C'E* and *FG* loops of D2 will change the orientations of the *CD* loop of D1 and the *C'E* loop of D2 with respect to each other, *i.e.* bringing them closer together or farther apart. This motion may allow a well defined loop from one domain and a flexible loop from the other domain to work together on ligand binding. It is possible, for example, that the *C'E* loop may become ordered upon binding.

Recently, the crystal structure of the A domain (also named I domain) from the α-subunit of the integrin Mac-1 ($\alpha_m\alpha_2$) has been reported (Lee, Rieu, Arnaout & Liddington, 1995). The authors have suggested a general metal-ion-dependent adhesion site for ligand (*e.g.* ICAM-1) binding and further proposed that a similar motif may exist in the β-subunit of these integrins. It is conceivable that the interaction between α4-integrin and VCAM-1, which involves Asp40 of VCAM-D1, is mediated by a metal-ion-dependent site as well. A conclusive answer must await the structure determination of such a complex.

We thank Stephen C. Harrison for generous support, advice and critical reading. We also thank Tomas Kirchhausen for encouragement, and Michael G. Rossmann and members of the Harrison group for helpful discussions. JHW and JHL were supported by NIH grant AI 30361 to Stephen C. Harrison.

References

Alon, R., Kassner, P. D., Carr, M. W., Finger, E. B., Hemler, M. E. & Springer, T. A. (1995). *J. Cell Biol.* **128**, 1243-1253.

Berlin, C., Bargatze, R. F., Campbell, J. J., von Andrian, U. H., Szabo, M. C., Hasslen, S. R., Nelson, R. D., Berg, E. L., Erlandsen, S. L. & Butcher, E. C. (1995). *Cell*, **80**, 413-422.

Blum, M. (1990). PhD thesis, Harvard University, Massachussetts, USA.

Blum, M., Metcalf, P., Harrison, S. C. & Wiley, D. C. (1987). *J. Appl. Cryst.* **20**, 235-242.

Brady, R. L., Dodson, E. J., Dodson, G. G., Lange, G., Davis, S. J., Williams, A. F. & Barclay, A. N. (1993). *Science*, **260**, 979-983.

Briskin, M. J., McEvoy, L. M. & Butcher, E. C. (1993). *Nature (London)*, **363**, 461-467.

Brünger, A. T. (1992). *Nature (London)*, **355**, 472-475.

Brünger, A. T., Kuriyan, J. & Karplus, M. (1987). *Science*, **235**, 458-460.

Carson, M. (1987). *J. Mol. Graphics*, **5**, 103-106.

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.

Cowtan, K. (1994). *Jnt CCP4 and ESF-EACBM Newslett Protein Crystallogr.* **31**, 34-38.

Elices, M. J., Osborn, L., Takada, Y., Crouse, C., Luhowskyj, S., Hemler, M. E. & Lobb, R. R. (1990). *Cell*, **60**, 577-584.

Epp, O., Lattman, E. E., Schiffer, M., Huber, R. & Palm, W. (1975). *Biochemistry*, **14**, 4943-4952.

Fawcett, J., Holness, C. L. L., Needham, L. A., Turley, H., Gatter, K. C., Mason, D. Y. & Simmons, D. L. (1992). *Nature (London)*, **360**, 481-484.

Garrett, T. P., Wang, J., Yan, Y., Liu, J. & Harrison, S. C. (1993). *J. Mol. Biol.* **234**, 763-778.

Greve, J. M., Davis, G., Meyer, A. M., Forte, C. P., Yost, S. C., Marlor, C. W., Kamarck, M. E. & McClelland, A. (1989). *Cell*, **56**, 839-847.

Gurtner, G., Davis, V., Li, H., McCoy, M. J., Sharpe, A. & Cybulsky, M. I. (1995). *Genes Develop.* **9**, 1-14.

Harpaz, Y. & Chothia, C. (1994). *J. Mol. Biol.* **238**, 528-539.

Hemler, M. E. & Lobb, R. R. (1995). *Curr. Opin. Hematol.* **2**, 61-67.

Hession, C., Moy, P., Tizard, R., Chisholm, P., Williams, C., Wysk, M., Burkly, L., Miyake, K., Kincade, P. & Lobb, R. (1992). *Biochem. Biophys. Res. Commun.* **183**, 163-169.

Hession, C., Tizard, R., Vassallo, C., Schiffer, S. B., Goff, D., Moy, P., Chi-Rosso, G., Luhowskyj, S., Lobb, R. & Osborn, L. (1991). *J. Biol. Chem.* **266**, 6682-6685.

Huber, S. (1994). *J. Virol.* **68**, 3453-3458.

Hynes, R. O. (1992). *Cell*, **69**, 11-25.

Jones, E. Y., Davis, S. J., Williams, A. F., Harlos, K. & Stuart, D. I. (1992). *Nature (London)*, **360**, 232-239.

Jones, E. Y., Harlos, K., Bottomley, M. J., Robinson, R. C., Driscoll, P. C., Edwards, R. M., Clements, J. M., Dudgeon, T. J. & Stuart, D. I. (1995). *Nature (London)*, **373**, 537-544.

Jones, T. A., Zhou, J.-Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110-119.

Kilger, G., Needham, L. S., Nielsen, P. J., Clements, J., Vestweber, D. & Holzmann, B. (1995). *J. Biol. Chem.* **270**, 5979-5984.

Kleywegt, G. J. & Jones, T. A. (1994). *From First Map to Final Model*, edited by S. Bailey, R. Hubbard & D. Waller, pp. 59–66. Warrington: Daresbury Laboratory.

Kraulis, P. J. (1991). *J. Appl. Cryst.* **24**, 946-950.

Kwee, L., Baldwin, H. S., Shen, H. M., Stewart, C. L., Buck, C., Buck, C. A. & Labow, M. A. (1995). *Development*, **121**, 489-503.

Lamzin, V. S. & Wilson, K. S. (1993). *Acta Cryst.* D**49**, 129-147.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283-291.

Lee, J.-O., Rieu, P., Arnaout, M. A. & Liddington, R. (1995). *Cell*, **80**, 631-638.
Luzzati, V. (1952). *Acta Cryst.* **5**, 802-810.
Marquart, M., Deisenhofer, J., Huber, R. & Palm, W. (1980). *J. Mol. Biol.* **141**, 369-391.
Miyake, K., Medina, K., Ishihara, K., Kimoto, M., Auerbach, R. & Kincade, P. W. (1991). *J. Cell Biol.* **114**, 557-565.
Osborn, L., Vassallo, C. & Benjamin, C. D. (1992). *J. Exp. Med.* **176**, 99-107.
Osborn, L., Vassallo, C., Browning, B. G., Tizard, R., Haskard, D. O., Benjamin, C. D., Dougas, I. & Kirchhausen, T. (1994). *J. Cell Biol.* **124**, 601-608.
Polte, T., Newman, W. & Gopal, T. V. (1990). *Nucleic Acids Res.* **18**, 5901-5901.
Ramachandran, G. N. & Sasisekharan, V. (1968). *Adv. Prot. Chem.* **23**, 283-437.
Rosen, G. D., Sanes, J. R., LaChance, R., Cunningham, J. M., Roman, J. & Dean, D. C. (1992). *Cell*, **69**, 1107-1119.
Rould, M. (1996). *Methods Enzymol.* In the press.
Ryu, S.-E., Kwong, P. D., Truneh, A., Porter, T. G., Arthos, J., Rosenberm, M., Dai, X., Xuong, N., Axel, R., Sweet, R. W. & Hendrickson, W. A. (1990). *Nature (London)*, **348**, 419-423.
Shepparu, A. M., McGuillan, J. J., Iademarco, M. F. & Dean, D. C. (1995). *J. Biol. Chem.* **270**, 3710-3719.
Springer, T. A. (1994). *Cell*, **76**, 301-314.
Staunton, D. E., Dustin, M. L. & Springer, T. A. (1989). *Nature (London)*, **339**, 61-64.
Staunton, D. E., Marlin, S. D., Stratowa, C., Dustin, M. L. & Springer, T. A. (1988). *Cell*, **52**, 925-933.
Terwilliger, T. C. & Eisenberg, D. (1983). *Acta Cryst.* A**39**, 813-817.
Vazeux, R., Hoffman, P. A., Tomita, J. K., Dickinson, E. S., Jasman, R. L., St. John, T. & Gallatin, W. M. (1992). *Nature (London)*, **360**, 485-488.
Vonderheide, R. H. & Springer, T. A. (1992). *J. Exp. Med.* **175**, 1433-1442.
Vonderheide, R. H., Tedder, T. F., Springer, T. A. & Staunton, D. E. (1994). *J. Cell Biol.* **125**, 215-222.
Wang, J. H., Pepinsky, B., Karpusas, M., Liu, J. H. & Osborn, L. (1994). *Proteins Struct. Funct. Genet.* **20**, 287-290.
Wang, J. H., Pepinsky, B., Stehle, T., Liu, J. H., Karpusas, M., Browning, B. & Osborn, L. (1995). *Proc. Natl Acad. Sci.* **92**, 5714-5718.
Wang, J. H., Yan, Y., Garrett, T. P., Liu, J., Rodgers, D. W., Garlick, R. L., Tarr, G., Husain, E., Reinherz, E. L. & Harrison, S. C. (1990). *Nature (London)*, **348**, 411–418.
Williams, A. F. & Barclay, A. N. (1988). *Ann. Rev. Immunol.* **6**, 381-405.
Yang, J. T., Rayburn, H. & Hynes, R. O. (1995). *Development*, **121**, 549-560.