Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

A reassessment of the MAdCAM-1 structure and its role in integrin recognition

J. Dando,[a] K. W. Wilkinson,[a] S. Ortlepp,[b] D. J. King[b]† and R. L. Brady[a]*

[a]Department of Biochemistry, University of Bristol, Bristol BS8 1TD, England, and [b]Celltech Ltd, 216 Bath Road, Slough SL1 4EN, England

† Current address: Corixa Corporation, 600 Gateway Boulevard, South San Francisco, CA 94080, USA.

Correspondence e-mail: l.brady@bris.ac.uk

Received 2 August 2001
Accepted 29 November 2001

PDB Reference: MAdCAM-1, 1gsm, r1gsmsf.

Mucosal addressin cell-adhesion molecule (MAdCAM-1) is a membrane-bound leukocyte receptor regulating both the passage and retention of leukocytes in mucosal tissues. A crystal structure for the two extracellular amino-terminal domains of human MAdCAM-1 has previously been reported, confirming their expected immunoglobulin superfamily topology. In this study, a second crystal structure of this fragment is described. Although the overall structure is similar to that previously reported, one edge strand in the amino-terminal domain is instead located on the opposite sheet. This alters the arrangement and conformation of amino acids in this region that have previously been shown to be crucial for ligand binding. MAdCAM-1 is also seen to form dimers within the crystal lattice, raising the possibility that oligomerization may influence the biological role of this adhesion molecule.

1. Introduction

Lymphocytes can be targeted to specific tissues *via* interactions between cell-surface receptor–ligand pairs. One example is mucosal addressin cell adhesion molecule (MAdCAM-1), which is present in the endothelium of mucosa. MAdCAM-1 binds both the integrin $\alpha4\beta7$ and L-selectin, regulating both the passage and retention of leukocytes in mucosal tissues. The extracellular portion of hMAdCAM-1 comprises two immunoglobulin superfamily (IgSF) domains, separated from the cell surface by an extended mucin-like region. A recent study (Tan *et al.*, 1998) reported the crystal structure of the integrin-binding IgSF domains of hMAdCAM-1, assigning both the NH_2-terminal domain (D1) and its neighbouring membrane-proximal domain (D2) to the I1 subset of IgSF folds. Domains of this subset comprise two β-sheets formed by strands *ABED* and *A′GFC* (Chothia & Jones, 1997; Harpaz & Chothia, 1994) and are believed to represent an intermediate fold between the variable (V) and constant (C) sets conventionally found in antibodies. Tan *et al.* (1998) also described a specific conformation for the *CD* loop in D1, a region believed to be of key importance for integrin binding. A further distinguishing feature in the crystal structure of hMAdCAM-1 is the presence of an extended loop region between strands *D* and *E* of D2. This region is reported to occupy the same face of the molecule as the protruding *CD* loop in D1 and these features have been proposed to represent two separate integrin-recognition motifs (Tan *et al.*, 1998; Wang & Springer, 1998).

However, inspection of the deposited coordinates for MAdCAM-1 (PDB code 1bqs) reveals several unusual features.

(i) Residues in both of the loop regions identified as important for integrin recognition exhibit relatively high temperature (*B*) factors. Main-chain atoms in the *CD* loop of D1 have temperature factors averaging 90 Å^2, compared with an overall average of 42 Å^2 for the remainder of the structure. This climbs to an average of 100 Å^2 for the extended *DE* loop (residues 149–157) in D2. One explanation may be that these regions are significantly mobile in the absence of contact with their integrin ligand.

(ii) The φ/ψ angles of four residues (42–45) in the proposed integrin-binding *CD* loop of D1 are in the disallowed regions of the Ramachandran plot (Fig. 1*a*). Tan *et al.* (1998) suggest this may reflect steric strain imposed in part by the neighbouring residue Arg70. This concept of steric strain, however, is not consistent with the high mobility of this region as evidenced by the local temperature factors. Further, although functional significance is often associated with Ramachandran outliers, unusual conformations are usually restricted to very short polypeptide segments (often a single residue), invariably restrained by hydrogen-bond networks, and commonly in or close to enclosed enzyme active sites (Herzberg & Moult, 1991). The D1 *CD* loop region of MAdCAM-1 does not share any of these characteristics.

(iii) The subsequent loop region (*D*/*E*) of D1 also exhibits unusual stereochemistry. Arg54 in this loop is again in a disallowed region of the Ramachandran plot and the average temperature factors for the loop residues are higher than the average for the remaining structure (Thr52, 57.9 Å^2; Gly53, 47.4 Å^2). The *D* strand connecting these loops, however, is consistent with the overall structure.

These unusual qualities in regions of the MAdCAM-1 structure believed to be central to its biological function have lead us to re-evaluate the crystal structure of this protein. In this study, we report an independent determination of the hMAdCAM-1 crystal structure. Our results are consistent with an alternative tracing of D1 that places it in the I2 subset of IgSF folds, with a new conformation for the integrin-binding loop of D1. In this form MAdCAM-1 forms a dimer, raising the possibility that MAdCAM-1 may be biologically active as an oligomer.

2. Experimental

2.1. Expression, purification and crystallization

The extracellular domains of hMAdCAM-1 were expressed as an Fc fusion protein using an inducible glutamine synthetase gene in murine NS0 cells (Barnes *et al.*, 2000). Residues 1–202 of the hMAdCAM-1 sequence (hMAdCAM-1) were included in the expressed protein, immediately followed at the C-terminus by the FXa cleavage site IEGR and then the sequence for the IgG Fc immunoglobulin fragment. The secreted fusion protein was purified directly from the cell supernatant using Protein A Sepharose and the Fc fragment was liberated by digestion with a 1:100 molar ratio of FXa (New England Biolabs) for 36 h under reducing conditions. Purified hMAdCAM-1 was obtained from the digestion mixture by ion-exchange chromatography [Mono S (Pharmacia) at pH 5, eluted with 0–0.5 *M* NaCl gradient]. hMAdCAM-1 includes a single N-linked glycosylation site

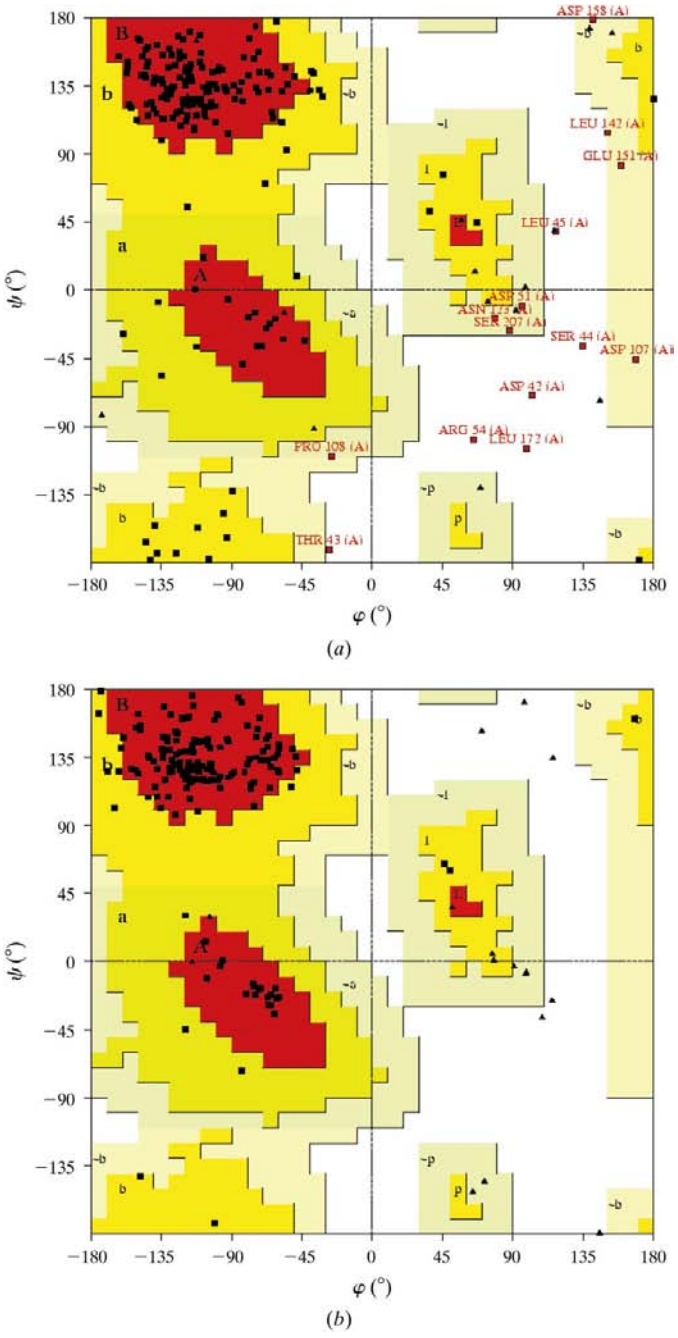

Figure 1
Ramachandran plots for the two MAdCAM-1 models. Ramachandran plots from the coordinates for (*a*) the current model for human MAdCAM-1 and (*b*) 1bqs. (The figure was prepared using *PROCHECK*; *et al.*, 1993.) The φ and ψ angles for individual amino acids are represented by black squares, except for glycine residues which are shown as triangles. The most favoured regions are shaded in red, with further allowed and generously allowed regions in orange and yellow, respectively. In the new model for MAdCAM-1, 91% of non-glycine residues are in the most favoured region, with the remaining 9% in the additional allowed regions. For the 1bqs model, 76% of non-glycine residues are in the most favoured regions, 16% in the additional allowed regions, 4% in the generously allowed regions and about 4% in disallowed regions.

Table 1
Diffraction data and refinement statistics for hMAdCAM-1.

Values in parentheses are for the outermost resolution shell.

Diffraction data	
Unit-cell parameters (Å,°)	a = 64.0, b = 99.2, c = 70.5, $\alpha = \beta = \gamma$ = 90.0
Resolution range (Å)	30–1.9
Completeness (%)	99.0 (98.5)
Total No. of observed reflections	102574
Average $I/\sigma(I)$	43.3 (4.7)
$R_{\rm merge}$ (%)	3.3 (23.0)
Refined model	
$R_{\rm free}$ (all data, 20–1.9 Å) (%)	25.9
$R_{\rm cryst}$ (all data, 20–1.9 Å) (%)	22.3
R.m.s.d. bond lengths (Å)	0.008
R.m.s.d. bond angles (Å)	0.032
Average B factor, main-chain atoms (Å^2)	29.4
Average B factor, side-chain atoms (Å^2)	32.7

(Asn61). Carbohydrate was removed from the purified protein by digestion with PNGase-F (New England Biolabs) for 24 h at 303 K, after which the deglycosylated form of the protein (DG-hMAdCAM-1) was isolated by further passage over the Mono-S column as described above.

Crystals of DG-hMAdCAM-1 were grown using the vapour-diffusion technique. 2 µl hanging drops of the protein at 10 mg ml^{-1} in 20 m*M* Tris buffer pH 8.0 were equilibrated against 1 ml well solution containing 16%(*w*/*v*) PEG 4000, 0.5 *M* Li_2SO_4 and 0.1 *M* Tris pH 7.75. These crystallization conditions are very similar to those described by Tan *et al.* (1998).

2.2. Data collection, structure solution and refinement

Crystals were transferred incrementally to solutions containing increasing amounts of glycerol then flash-frozen at 100 K in the crystallization buffer with 15% glycerol. Diffraction data were collected on beamline X11 at the EMBL Outstation of the DESY synchrotron at Hamburg and processed using *DENZO* (Otwinowski, 1993). The data are summarized in Table 1.

Structure solution was straightforward using the phases from the published structure of hMAdCAM-1 (Tan *et al.*, 1998) determined from isomorphous crystals with near-identical unit-cell parameters (see legend to Table 1). The structure of DG-hMAdCAM-1 was refined using iterative cycles of *REFMAC* (Murshudov *et al.*, 1999) and manual rebuilding with the X-BUILD option in *QUANTA* (MSI). Statistics for the final model are summarized in Table 1 and a Ramachandran plot for the final model coordinates is shown in Fig. 1(*b*). The corresponding plot from the published coordinates for hMAdCAM-1 is shown in Fig. 1(*a*).

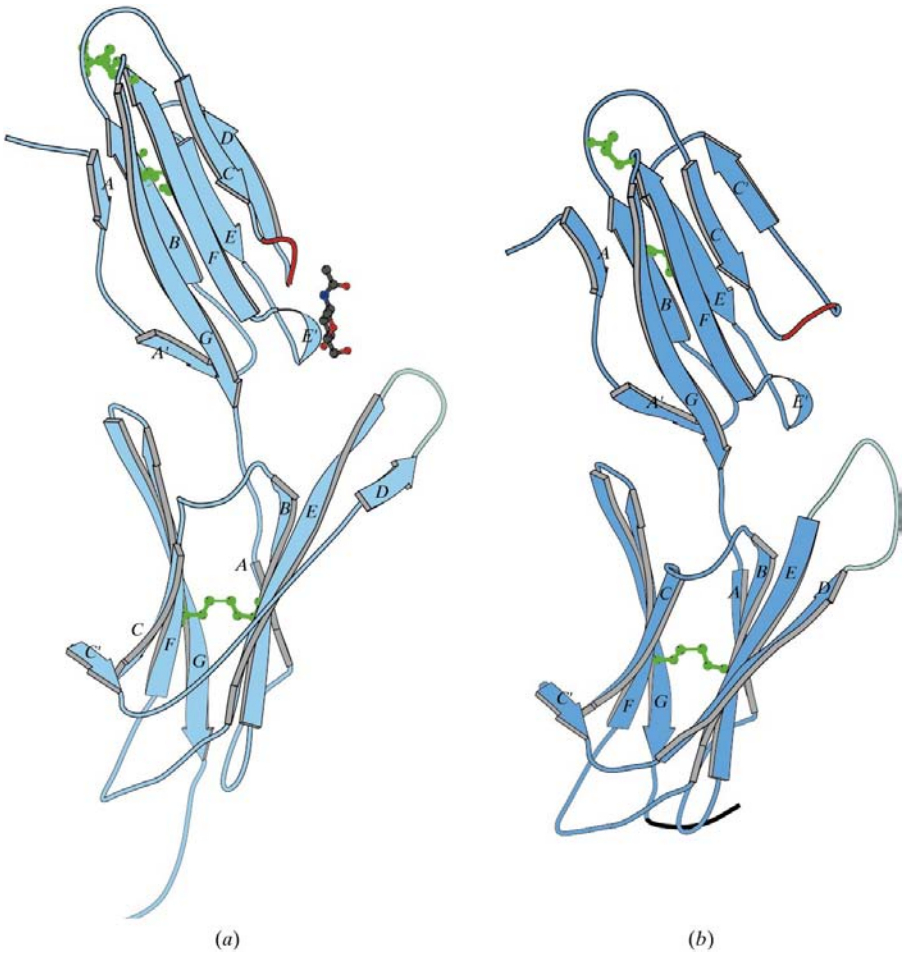

Figure 2
Structure of the amino-terminal domains of human MAdCAM-1. (*a*) Ribbon drawing showing the previous N-terminal two-domain structure of human MAdCAM-1 (PDB code 1bqs). (*b*) In the same orientation, ribbon drawing showing the structure of the N-terminal two domains of human MAdCAM-1 as described in this study. Residues forming disulfide bonds are shown in green in ball-and-stick representation. Residues in the *C*/*D* (*C*/*C*′ in the previous structure) loop of domain 1 are shown in red and residues in the *D*/*E* loop of domain 2 are shown in cyan. The *N*-acetyl glucosamine group attached to Asn61 in the 1bqs structure is shown in ball-and-stick representation with C atoms black, O atoms red and N atoms blue. (The figure was prepared using *MOLSCRIPT*; Kraulis, 1991.)

3. Results and discussion

The sequences of constructs used for this and the previous hMAdCAM-1 study (Tan *et al.*, 1998) are essentially identical except at the C-terminus. Unsurprisingly, the 5–6 amino acids at the C-terminus, which do not form part of the second domain, differ in their conformation. In the earlier structure, these residues protrude away from the domain close to residues in the *C*/*D* loop of an adjacent molecule in the crystal lattice, although no contacts are

made with these residues. This potential disruption is avoided in the current structure, where the shorter C-terminus remains close to the domain. This may contribute to the clarified electron density of regions at the base of domain 1 in the current structure (see later).

MAdCAM-1 is extensively post-translationally modified *in vivo*, primarily through O-linked glycosylation on the membrane stalk extending beyond domain 2. hMAdCAM-1 also has one N-linked glycosylation site within D1 (Asn61), which is not essential for activity (Tan *et al.*, 1998) and is not present in the murine form. One minor difference between the construct used for the earlier structure determination of hMAdCAM-1 and in the present study relates to this modification. Tan and coworkers used recombinant protein expressed in lectin-resistant CHO cells, which secrete predominantly high-mannose structures that can readily be removed using mannosidase. Only the first *N*-acetyl glucosamine, directly attached to Asn61, is visible in their electron-density maps. In the present study, we have expressed hMAdCAM-1 as an Fc-fusion protein in NS0 cells – which retain authentic glycosylation – then removed the glycosylation with PNGase-F. This glycosidase hydrolyses the asparagine–glucosamine link and hence the protein used in this study is fully deglycosylated. Both proteins crystallize isomorphously, with only a small change in the unit-cell parameters (in this study, the unit-cell parameters are $a = 64.0$, $b = 99.2$, $c = 70.5$ Å, $\alpha = \beta = \gamma = 90.0^\circ$. In the Harvard study, the unit-cell parameters were $a = 65.8$, $b = 101.1$, $c = 70.0$ Å, $\alpha = \beta = \gamma = 90.0^\circ$).

As expected, the overall structures of the two forms of hMAdCAM-1 are largely similar. In both cases, two IgSF domains related by a similar angle are observed. Both structures have been refined to comparable *R* factors ($R_{\rm cryst} = 22.3\%$, $R_{\rm free} = 25.9\%$ using all data in the resolution range 20–2.1 Å in this study; $R_{\rm cryst} = 22.3\%$, $R_{\rm free} = 28.0\%$ using data with $|F| > 2\sigma$ in the resolution range 15–2.2 Å in the previous study), although the Ramachandran plots for each vary considerably in the number of outliers (Fig. 1). The overall structural similarity is reflected in the root-mean-square differences calculated for 177 equivalent C^α positions of 0.93 Å. Nonetheless, the interpretation of the electron density between the two structures at the *C*/*C*′ (*C*/*D* in previous study) and *C*′/*E* (*D*/*E* in previous study) loops of domain 1 differs markedly. These alternative interpretations lead to different connectivities in the final structures, with significant implications for the positioning of residues proposed to be crucial for integrin binding.

In the model from the current study, at the base of the *C* strand the polypeptide chain loops to form a further *C*′ strand on the edge of the *GFC* sheet, before crossing to the adjacent sheet at the top of the domain to form the *E* strand. This topology (*ABE*|*A*′*GFCC*′; see Fig. 2*b*) mirrors that observed in domain 2 of hMAdCAM-1 and has been described as an I2 set IgSF domain. This differs from the model previously published for this domain of hMAdCAM-1, in which at the base of the *C* strand the peptide chain crosses to the adjacent *ABED* sheet to form the *D* strand. The resulting topology (*ABED*|*A*′*GFC*; see Fig. 2*a*) has been classified by Wang & Springer (1998) as an I1 IgSF domain and is also observed in other structures such as the N-terminal domain of VCAM-1 (Jones *et al.* 1995). The source of these differing interpretations can be seen in the assembly of adjacent molecules within the crystal lattice to form a dimer (Fig. 3). Residues 45–50 from the symmetry-related molecule in the current model occupy the same density as residues 45–50 from the first molecule in the previous structure. Residues in the connecting loops (*C*/*C*′ and *C*′/*E* in the current structure and *C*/*D* and *D*/*E* in the earlier structure) occupy unique positions in each structure.

Carbohydrate structures extending from the site close to the base of D1 are likely to contact the mobile *DE* loop in D2, which is disordered within both crystal lattices. In addition, the *C*/*C*′ loop from the adjacent molecule of the dimer formed within the crystal lattice is in close proximity to this glycosylation site. This region, which contains Asp42 from the putative integrin-binding site, is poorly ordered in both crystal structures. It is not clear whether the relative mobility of these peptide regions is an artefact resulting from removal of glycosylation or is replicated in the presence of the heterogeneous and mobile carbohydrate structures. We note, however, that in both models for hMAdCAM-1 carbohydrate structures at the Asn61 position could at least partially obscure access to *C*/*C*′ loop residues. There is no glycosylation motif at this or adjacent regions in the sequence for murine MAdCAM-1. An examination of the expected location of carbohydrate at this site does not suggest that its presence would be inhibitory for formation of the dimer observed in the crystals.

3.1. The *C*/*C*′ and *C*′/*E* loops of domain 1

Retrieval of the deposited structure factors from the PDB for the previous structure (entry 1bqs) has allowed us to compare the observed electron density and its interpretation for each structure. Coordinates for the 1bqs structure were first subjected to five cycles of maximum-likelihood refinement using *REFMAC* to ensure the maps subsequently generated were comparable to those used in the current interpretation. This led to minimal changes in the coordinates ($R_{\rm cryst} = 22.8\%$, $R_{\rm free} = 28.8\%$ for all data in the resolution range 20–2.2 Å) and the subsequent electron-density maps in the loop regions for both structures are shown in Fig. 4.

Figs. 4(*a*) and 4(*b*) show the density and corresponding interpretation in the region of residues 40–45 (*C*/*C*′ loop in hMAdCAM-1, *C*/*D* in 1bqs). Neither structure shows unambiguously interpretable density in this region. This region appears to be mobile within the crystals, perhaps attributable to its proximity to the glycosylation site (Asn61) in the adjacent strand. We note, however, that residues 42–45 inclusive in the 1bqs model all fall within disallowed regions of the Ramachandran plot (Fig. 1). Tan and coworkers argue that this unusual conformation is likely to be associated with function. A buried un-neutralized arginine (Arg70) is suggested as the source of this conformation, making hydrogen bonds to the carbonyl groups of Asp42 and Thr43.

However, the poor quality of the electron density in this region is not consistent with these residues being intentionally constrained in an unfavourable conformation. When we attempted to refine our current model with these residues in similar positions, no improvement in the Ramachandran plot could be obtained. We note that although Tan and coworkers state that 92.3% of residues in the 1bqs model are in the favoured region of the Ramachandran plot, *PROCHECK* analysis of the 1bqs structure (Fig. 1*a*) shows only 76.3% in most favoured regions, with 16.0% in the additional allowed regions, a further 4.1% in the generously allowed regions and 3.6% in disallowed regions. By comparison, in the current model with these residues in the *C*/*C*′ location there are no Ramachandran outliers. The Ramachandran plot (Fig. 1*b*) shows a tighter distribution, with 92.2% of residues in the most favoured regions, 7.8% in additional allowed regions and no residues in either generously allowed or disallowed regions. Hence, although the electron density in this region is poor in both cases, the current study structure of this region is more plausible in view of our general understanding of preferred polypeptide conformations.

Ambiguity in the course of the chain trace in the region of residues 42–45 can be clarified by reference to the density at the opposite end of the connected strand. If – as in the current model – the connectivity at the base of the *C* strand is modelled to a *C*′ strand on the same sheet, then necessarily at the carboxy-terminal end of this *C*′ strand the chain must cross to the opposite sheet to begin the *E* strand. This situation is seen unambiguously in the density from the present study (Fig. 4*c*), supporting our assignment of the connectivity in the poorly ordered *C*/*C*′ region. On the other hand, if the polypeptide chain crosses to the other sheet at the base of the *C* strand (*C*/*D*) as in the 1bqs structure, at the carboxy end of the *D* strand a hairpin-type turn is required to continue to the *E* strand on the same sheet. The observed density in the current study does not support a hairpin turn in this region as seen in the 1bqs model (Fig. 4*d*) in which the main chain follows density assigned to the Arg54 side chain in the current structure. The temperature factors for main-chain atoms in the corresponding residues are all relatively high. A further concern is that Arg54 falls in a disallowed region of the Ramachandran plot in the 1bqs model and Thr52 is in the generously allowed region for left-handed helical turns. The model from our study provides a more plausible interpretation of the density in this region, hence confirming the topology of domain 1 as *ABE*|*A*′*GFCC*′ as illustrated in Fig. 2(*b*).

In the new model for hMAdCAM-1, rather than being buried within the fold and supporting the loop structure of residues 42–45 through a network of hydrogen bonds as seen in the 1bqs model, the guanidinium group of Arg70 is now located at the base of a surface cleft. In the monomeric form, this leaves the arginine surface accessible and potentially free to interact with ligands such as $\alpha4\beta7$. However, within the crystals hMAdCAM-1 forms dimers (see below and Fig. 3). In the oligomer the arginine side chain forms a hydrogen bond with the main-chain carbonyl group of Ala66 from the same molecule and the main-chain carbonyl group of Leu45 from the crystallographically related monomer. Residues on the same face of the molecule as Arg70 have been shown by mutational studies (Green *et al.*, 1999) to be necessary for ligand binding (see later).

hMAdCAM-1 binds to the integrin $\alpha4\beta7$ (Berlin *et al.*, 1993) and hVCAM-1 binds to both $\alpha4\beta7$ and $\alpha4\beta1$ (Ruegg *et al.*, 1992). Previous studies have identified a key integrin-binding residue in domain 1 of both hVCAM-1 and hMAdCAM-1. This residue – Asp42 in hMAdCAM-1 (Shyjan *et al.*, 1996) and Asp40 in hVCAM-1 (Wang *et al.*, 1996) – is located in the *CC*′/*CD* loop in each structure. This is distinct from the hICAM-1 and hICAM-2 structures, where the key residue has been identified as a glutamate residue lying at the end of the *C* strand on a flat surface (Casasnovas *et al.*, 1997, 1998; Bella *et al.*, 1998). However, the hICAM molecules bind the integrin ligands $\alpha L\beta2$ and $\alpha M\beta2$ (Springer, 1990) which contain an I domain that the $\alpha4$ ligands lack and this may explain the differing topology around the acidic residues. As hMAdCAM-1 and hVCAM-1 both bind $\alpha4$ integrins, a degree of similarity might be expected around the binding site for these ligands. However, when the previous model for D1 of hMAdCAM-1 was superimposed on D1 from hVCAM-1, the key aspartate residues were found to be displaced by about 8 Å. Additionally, Asp42 in the hMAdCAM-1 structure is largely buried away from solvent, whereas the aspartate in hVCAM-1 protrudes in a manner consistent with binding to ligand (Jones *et al.*, 1995). Overlaying the new model for hMAdCAM-1 with hVCAM-1 shows that the positions of the two aspartate C^{α} atoms are only 2 Å apart (r.m.s.d. = 0.93 Å for all equivalent C^{α} atoms in the domains), leading to their hydrophilic heads adopting almost identical positions (Fig. 5). These conformations would appear to be more consistent with both of these molecules being capable of binding $\alpha4$ ligands.

3.2. The extended *DE* loop in domain 2

When compared with many other IgSF integrin receptors, hMAdCAM-1 has an extra long insert in the *DE* loop of domain 2. This region contains a very high proportion of negatively charged residues, with seven of the 11 residues of this region (149–159) being either glutamate or aspartate. Substitution of six of the nine acidic residues in the *DE* loop of domain 2 with alanine results in significant loss of $\alpha_4\beta_7$ binding (Green *et al.*, 1999). Unfortunately, this region in the electron-density maps of both 1bqs and of the current study is poorly ordered. Those residues that have been included in this loop have very high temperature factors, indicating high mobility. It is therefore only possible to speculate as to how this *DE* loop may function structurally in $\alpha_4\beta_7$ binding.

It has been proposed that this negatively charged loop may function as an antenna orienting the hMAdCAM-1 molecule above the cell membrane (Tan *et al.*, 1998). While this is possible, mutational studies point to this loop being more directly involved in ligand binding (Green *et al.*, 1999; Newham *et al.*, 1997). If both the *DE* and *C*/*C*′ (*CD*) loops are involved in ligand binding, it seems reasonable that both will adopt a more rigid conformation when bound to the integrin

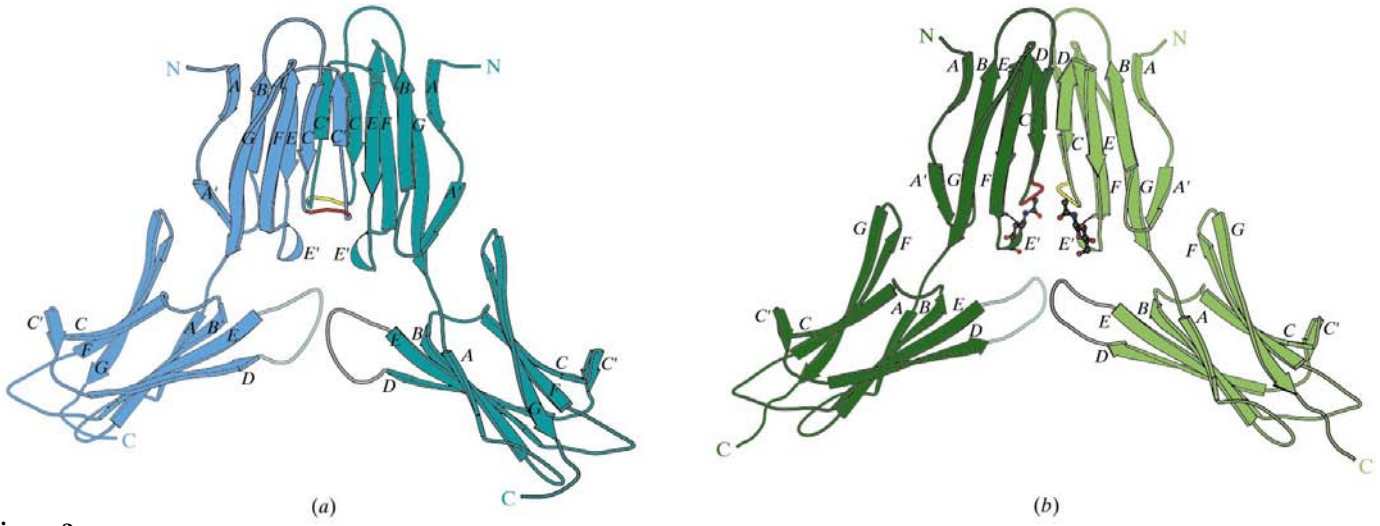

Figure 3
Possible dimers of hMAdCAM-1. Ribbon diagrams showing the crystal packing of monomers of human MAdCAM-1 to form dimers in the structures of (*a*) the current model for human MAdCAM-1 (blue) and (*b*) 1bqs (green). Individual monomers within each dimer are distinguished by dark and light shading. Residues in the *C*/*D* (*C*/*C*′) loop of domain 1 are shown in red and yellow. Residues in the *D*/*E* loop of domain 2 are shown in cyan and grey. The glycosylation site in the 1bqs structure is shown in ball-and-stick representation, with C atoms black, O atoms red and N atoms blue. (The figure was prepared using *MOLSCRIPT*; Kraulis, 1991.)

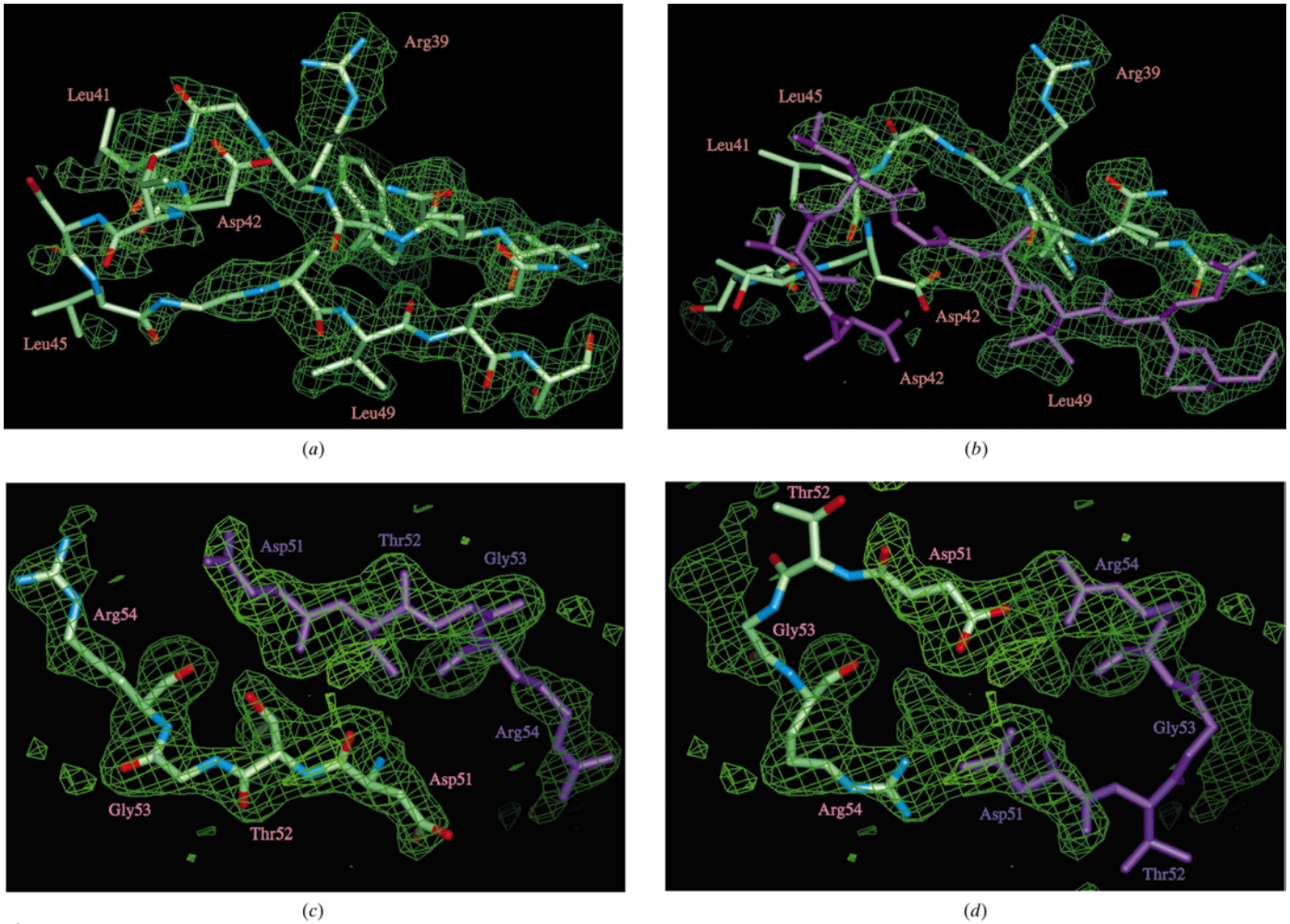

Figure 4
Electron density of changeover region in domain 1. Electron density with overlaid models for residues 36–51 in (*a*) the current structure of human MAdCAM-1 and (*b*) 1bqs and for residues 51–54 in (*c*) the current structure of human MAdCAM-1 and (*d*) 1bqs. In each case, the electron density is from an omit map $|F_{obs}| - |F_{calc}|$ calculated using observed structure factors from the current study and calculated structure factors from models in which residues 36–54 were given zero occupancy during refinement. Atoms from symmetry-related molecules are shown in magenta.

ligand. Both these loops occupy the same face of the hMAdCAM-1 molecule and further mutations by Green and coworkers also map residues crucial for ligand binding to this face of the molecule, for example Arg70. hVCAM-1 has an equivalent negatively charged loop, the $C'E$ loop in domain 2. This loop is shorter and less negatively charged than the equivalent loop in hMAdCAM-1 (Jones *et al.*, 1995) and its function is also not known, but mutational studies have suggested it is important for integrin binding (Newham *et al.*, 1997).

3.3. Dimer formation

In the present interpretation of hMAdCAM-1, two symmetry-related molecules form a dimer within the crystal lattice through an extensive interface formed by residues from the edge of the β-sandwich in domain 1 (Fig. 3). Domain 1 exhibits a pronounced V-shaped groove at this edge, partially exposing some of the hydrophobic residues in the domain core. Main-chain atoms from the edge C' and E strands form antiparallel hydrogen bonds with, respectively, the E and C' strands of the adjacent molecule in the dimer. This forms two continuous β-sheets extending across both faces of the dimer. This arrangement creates a quasi-continuous and common buried hydrophobic core for both D1 domains. In the new model the solvent-accessible surface area buried between the domains is 2236 Å^2, dominated by 1251 Å^2 of hydrophobic contributions compared with 985 Å^2 from hydrophilic atoms (all calculated in the absence of H atoms using *QUANTA*). In the previous model, adjacent molecules within the crystal lattice could also be interpreted as a dimer, although the buried surface area (1022 Å^2 hydrophilic contribution, 800 Å^2

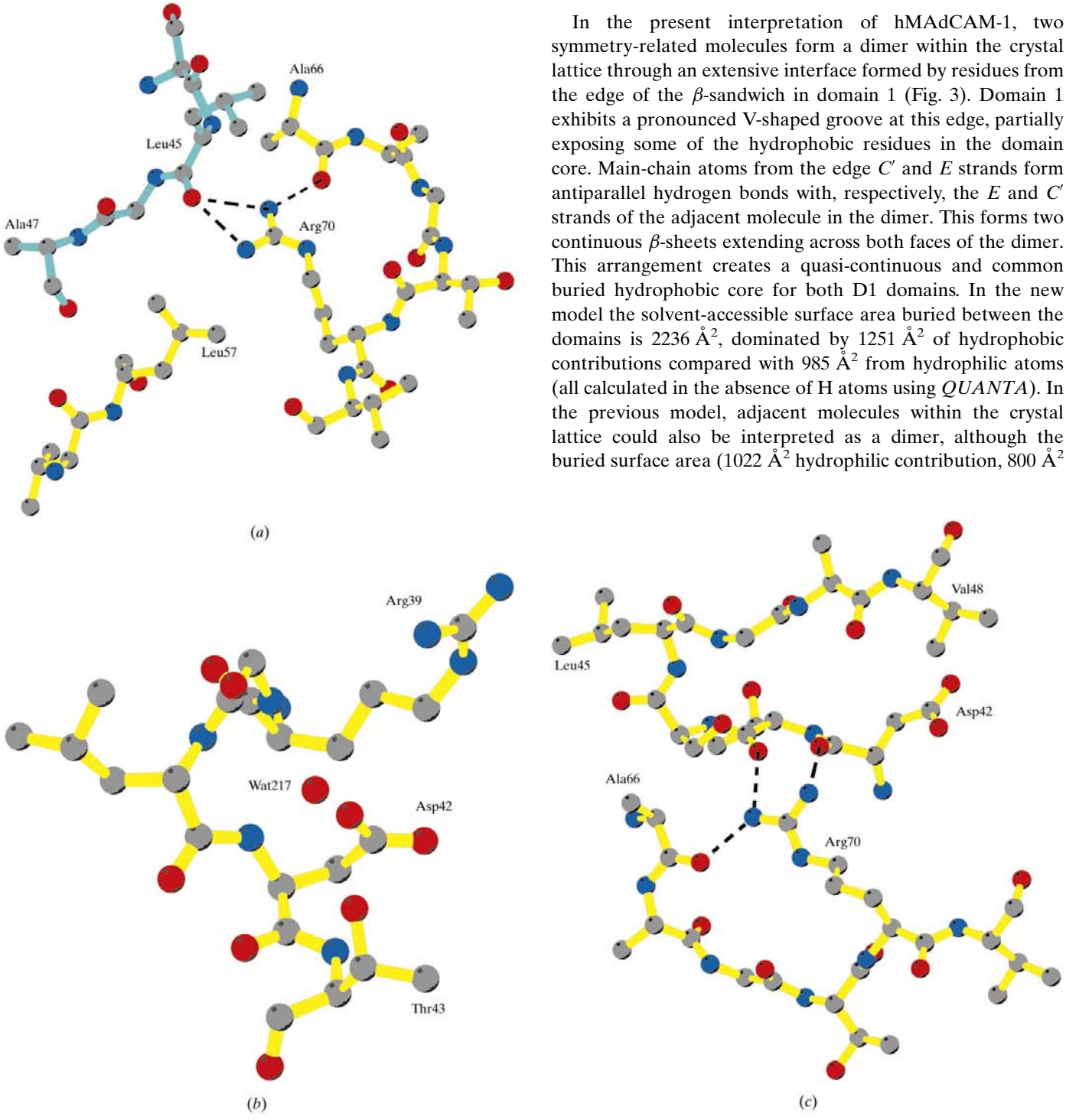

Figure 5
Interactions of Asp42 and Arg70. Diagram to illustrate the local conformation and interactions involving residues (*a*) Arg70 and (*b*) Asp42 in the current structure of human MAdCAM-1. (*c*) shows the very different interactions involving residues Arg70 and Asp42 in the 1bqs structure of human MAdCAM-1. Symmetry-related molecules are shown in cyan and hydrogen bonds are indicated by dashed lines. (The figure was prepared using *MOLSCRIPT*; Kraulis, 1991.)

hydrophobic contribution, 1823 Å^2 in total) is less pronounced. Edge-to-edge dimerization of IgSF domains has previously been observed in other proteins, including human ICAM-1 (Reilly *et al.*, 1995; Casasnovas *et al.*, 1998). As the Ig domains of hMAdCAM-1 are believed to project approximately 22 nm away from the cell surface by an extended mucin-like region (Shyjan *et al.*, 1996), there appears to be no steric restriction to forming a dimer of the type observed in the crystal.

One residue that makes an important contribution at the dimer interface is Arg70, forming a bifurcated charged hydrogen bond across the dimer interface with the carbonyl group of Leu45 from the adjacent molecule. Mutational studies indicate Arg70 is essential for integrin binding (Green *et al.*, 1999). This side chain is buried in the 1bqs structure, forming hydrogen bonds with the carbonyl groups of residues 42 and 43 that lie in disallowed regions of the Ramachandran plot. In the new model, the charged guanadinium group would be solvent exposed in a single monomer. Disruption of the intermolecular bond made by Arg70 in the dimer would be expected to be detrimental for dimer formation. This raises the possibility that the reduction in integrin-binding activity reported for mutants of this residue might result from disruption of a dimeric structure essential for activity.

It is also interesting to note that in the dimer the two loops thought to be important for ligand binding are brought closer together than in the monomer alone. In the dimer the *C*/*C*′ (*C*/*D*) loop from domain 1 and the *D*/*E* loop from domain 2 lie in an exposed position on the dimer surface, only 15 Å apart. Mutations in either of these loops could disrupt assembly of a functional dimer. We note that mutations in both of these regions have been reported to decrease integrin binding (Green *et al.*, 1999).

Nonetheless, the existence of this dimeric form of MAdCAM-1 outside of the crystal lattice remains speculative. Analytical ultracentrifugation sedimentation equilibrium analysis of the recombinant MAdCAM-1 used in this study has shown the vast majority of the protein exists in monomeric form within solution (data not shown).

4. Conclusions

Previous analyses of the hMAdCAM-1 structure have relied heavily on expected homology with VCAM-1 for interpretation of both the chain topology and biologically active surfaces. In this study, we have identified a small but significant variation in the MAdCAM-1 fold, in a region believed to be central to the integrin ligand-binding site. This reinterpretation of the hMAdCAM-1 domain 1 topology is supported by significant improvements in the quality of the Ramachandran plot for this new model when compared with the earlier model and improved refinement statistics. Nonetheless, we cannot discount the possibility that these differences exist as the recombinant proteins used in each study are from different sources. In particular, removal of the N-linked glycosylation from Asn61 might lead to small structural perturbations. However, we note that this glycosylation site is not well conserved across species, arguing against it being a stringent determinant of a structure required for biological action. Further, the close isomorphism of both crystalline forms of the protein suggests a consistency of overall topology not evident when the two models are compared.

The difficulties in assigning an unambiguous chain trace for hMAdCAM-1 domain 1 are clear evidence of inherent flexibility in the edge strand of this IgSF domain. This is consistent with the notion that this strand may rearrange or 'strand-swap' as an accompaniment to oligomer assembly. The large extended hydrophobic core formed within the dimer in the crystals described in this study would be expected to stabilize dimer formation, although there is currently little evidence for significant populations of this dimer when these proteins are in solution. Although it is possible this strand swapping is induced by the close packing of molecules within the crystal lattice, the dimer produced is consistent with some of the ligand-binding data. Weak molecular associations are believed to dominate many molecular interactions at the cell surface and are not uncommon between proteins tethered to the membrane. In the absence of a crystal structure of hMAdCAM-1 bound to its integrin ligand, it is not possible to ascertain whether the dimer observed in this study is biologically relevant. Nonetheless, the alternative model presented in this study provides a reasonable interpretation of the mutational analyses and raises the possibility that oligomeric assembly should be considered as a possible regulator of MAdCAM-1 function. Further studies are required to examine the relevance of the MAdCAM-1 dimer either at the cell surface or in complex with its integrin ligand.

We are grateful to the BBSRC for the award of a postgraduate studentship to JD, and to the staff at Daresbury Laboratory for access to the synchrotron for data-collection facilities.

References

Barnes, L. M., Bentley, C. M. & Dickson, A. J. (2000). *Cytotechnology*, **32**, 109–123.
Bella, J., Kolatkar, P. R., Marlor, C. W., Greve, J. M. & Rossmann, M. G. (1998). *Proc. Natl Acad. Sci. USA*, **95**, 4140–4145.
Berlin, C., Berg, E. L., Briskin, M. J., Andrew, D. P., Kilshaw, P. J., Holzmann, B., Weissmann, I. L. Hamann, A. & Butcher, E. C. (1993). *Cell*, **74**, 185–189.
Casasnovas, J. M., Springer, T. A., Liu, J., Harrison, S. C. & Wang, J. (1997). *Nature (London)*, **387**, 312–315.
Casasnovas, J. M., Stehle, T., Liu, J.-H., Wang, J.-H. & Springer, T. A. (1998). *Proc. Natl Acad. Sci. USA*, **95**, 4134–4139.
Chothia, C. & Jones, E. Y. (1997). *Annu. Rev. Biochem.* **66**, 823–862.
Green, N., Rosebrook, J., Cochran, N., Tan, K., Wang, J., Springer, T. A. & Briskin, M. J. (1999). *Cell Adhes. Commun.* **7**, 167–181.
Harpaz, Y. & Chothia, C. (1994). *J. Mol. Biol.* **238**, 528–539.
Herzberg, O. & Moult, J. (1991). *Proteins*, **11**, 223–229.
Jones, E., 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**, 539–544.
Kraulis, P. (1991). *J. Appl. Cryst.* **24**, 946–950.
Laskowski, R. A, MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.

Murshudov, G. N., Lebedev, A., Vagin, A. A., Wilson, K. S. & Dodson, E. J. (1999). *Acta Cryst.* D**55**, 247–255.
Newham, P., Graig, S. E., Seedon, G. N., Schofield, N. R., Ress, A., Edwards, M., Jones, Y. & Humphries, M. J. (1997). *J. Biol. Chem.* **272**, 19429–19440.
Otwinowski, Z. (1993). *Proceedings of the CCP4 Study Weekend. Data Collection and Processing*, edited by L. Saywer, N. Isaacs & S. Bailey, pp. 56–62. Warrington: Daresbury Laboratory.
Reilly, P. L., Woska, J. R., Jeanfavre, D. D., McNally, E., Rotlein, R. & Bormann, B. J. (1995). *J. Immunol.* **155**, 529–532.
Ruegg, C. A., Postigo, A., Sikroski, E. E., Butcher, E. C., Pytela, R. & Erle, D. J. (1992). *J. Cell. Biol.* **117**, 179–189.
Shyjan, A. M., Bertagnolli, M., Kenney, C. J. & Briskin, M. J. (1996). *J. Immunol.* **156**, 2851–2857.
Springer, T. A. (1990). *Nature (London)*, **346**, 425–434.
Tan, K., Casasnovas, J. M., Liu, J., Briskin, M. J., Springer, T. A. & Wang, J. (1998). *Structure*, **6**, 793–801.
Wang, J. & Springer, T. A. (1998). *Immunol. Rev.* **163**, 197–215.
Wang, J., Stehle, T., Pepinsky, B., Liu, J., Kaprusas, M. & Osborn, L. (1996). *Acta Cryst.* D**52**, 369–379.