The structure of immunoglobulin superfamily domains 1 and 2 of MAdCAM-1 reveals novel features important for integrin recognition

Kemin Tan[1], Jose M Casasnovas[2†], Jin-huan Liu[1], Michael J Briskin[3], Timothy A Springer[2*] and Jia-huai Wang[1,4*]

Background: Mucosal addressin cell adhesion molecule 1 (MAdCAM-1) is a cell adhesion molecule that is expressed on the endothelium in mucosa, and guides the specific homing of lymphocytes into mucosal tissues. MAdCAM-1 belongs to a subclass of the immunoglobulin superfamily (IgSF), the members of which are ligands for integrins. Human MAdCAM-1 has a unique dual function compared to other members in the same subclass in that it binds both the integrin α4β7, through its two IgSF domains, and a selectin expressed on leukocytes, via carbohydrate sidechains. The structure determination of the two IgSF domains and comparison to the N-terminal two-domain structures of vascular cell adhesion molecule 1 (VCAM-1) and intercellular adhesion molecules (ICAM-1 and ICAM-2) allow us to assess the molecular basis of the interactions between integrins and their preferred ligands.

Results: The crystal structure of a fragment containing the two IgSF domains of human MAdCAM-1 has been determined to 2.2 Å resolution. The structure of MAdCAM-1 reveals two separate integrin-recognition motifs. The key integrin-binding residue, Asp42, resides in the CD loop of domain 1; a buried arginine residue (Arg70) plays a critical role in maintaining the conformation of this loop. The second binding site is associated with an unusual long D strand in domain 2. The D and E strands extend beyond the main body of the domain, forming a negatively charged β ribbon unique to MAdCAM-1. This ribbon is located on the same face as the key aspartate residue in domain 1, consistent with evidence that it is involved in integrin binding.

Conclusions: The structural comparison of MAdCAM-1 to other members of the same IgSF subclass reveals some interesting features. Firstly, MAdCAM-1, like VCAM-1, has the key integrin-binding residue located on the protruding CD loop of domain 1 and binds to an integrin that lacks an I domain. This is in contrast to ICAM-1 and ICAM-2 where the key residue is located at the end of the C strand on a flat surface and which bind to integrins that contain I domains. Secondly, architectural differences in the CD loops of MAdCAM-1 and VCAM-1 cause an 8 Å shift in position of the critical aspartate residue, and may partly determine their binding preference for different integrins. Finally, the unusual charge distribution of the two-domain fragment of MAdCAM-1 is predicted to orient the molecule optimally for integrin binding on the top of its long mucin-like stalk.

Addresses: [1]Laboratory of Immunobiology, Dana-Farber Cancer Institute, 44 Binney Street, Boston, MA 02115, USA, [2]The Center for Blood Research, and Department of Pathology, Harvard Medical School, 200 Longwood Avenue, Boston, MA 02115, USA, [3]LeukoSite Inc., Cambridge, MA 02142, USA and [4]Department of Pediatrics, Harvard Medical School, Boston, MA 02115, USA.

[†]Present address: Karolinska Institute, Department of Biosciences at NOVUM, Center for Biotechnology, S-141 57 Huddinge, Sweden.

*Corresponding authors.
E-mail: jwang@red.dfci.harvard.edu
springer@sprsgi.med.harvard.edu

Key words: IgSF domain, integrin-binding sites, integrin ligand, MAdCAM-1, X-ray crystallography

Received: **20 March 1998**
Revisions requested: **8 April 1998**
Revisions received: **29 April 1998**
Accepted: **1 May 1998**

Structure 15 June 1998, **6**:793–801
http://biomednet.com/elecref/0969212600600793

Introduction

Cell adhesion molecules, including members of the immunoglobulin superfamily (IgSF), the integrin family, the selectin family and carbohydrate ligands for selectins, have important roles in the immune response and immune surveillance [1,2]. Ligands on the vascular endothelium bind to receptors on blood leukocytes and signal the trafficking of the immune system cells to sites of infection or injury. The adhesion of leukocytes to endothelial cells [3] and subsequent transendothelial migration [4] involves multiple 'receptor–ligand' pairs. Binding of selectins to carbohydrate ligands initiates the first step of adhesion activity by mediating tethering of the leukocyte and transient adhesion to the vessel wall. These interactions permit rolling of the cell along the vessel wall in response to hydrodynamic forces. The binding of the leukocyte β2 integrins αLβ2 (LFA-1) and αMβ2 (Mac-1) to their ligands, which include intercellular adhesion molecules (ICAM-1 and 2), supports firm adhesion [1,5] and migration across the endothelium into tissues. The leukocyte α4 integrins α4β1 (VLA-4) and α4β7 (LPAM-1) bind to their preferred ligands, vascular cell adhesion molecule 1 (VCAM-1) and

mucosal addressin cell adhesion molecule 1 (MAdCAM-1). These molecules function at an intermediate stage in leukocyte adhesion to the endothelium, namely they support both rolling interaction and firm adhesion [6–8].

MAdCAM-1 is selectively expressed on high-endothelial venules of Peyer's patches, other gut associated lymphoid tissues including appendix and mesenteric lymph nodes, and in venules of the lamina propria [9,10]. The molecule regulates traffic of a subset of lymphocytes into mucosal tissues [2]. Human MAdCAM-1 contains, beginning from its N terminus, two IgSF domains, a mucin-like region of 115 residues, a transmembrane domain, and a 43-residue cytoplasmic domain [11]. MAdCAM-1 has a unique dual function among cell adhesion molecules [8,12–15]: it binds the integrin α4β7 through its IgSF domains, and when appropriately *O*-glycosylated binds L-selectin through its mucin-like region. The interaction with L-selectin mediates a transient adhesion to the endothelium that allows the cell to roll in response to hydrodynamic forces exerted on it in the bloodstream. The interaction with α4β7 mediates slower rolling and, after activation, firm adhesion.

MAdCAM-1, VCAM-1, ICAM-1 and ICAM-2 represent a subclass of the IgSF that are specialized for binding to integrins. These proteins can be recognized by their greater amino acid sequence homology to one another than to other IgSF members. The members of this subclass differ in their specificities for integrins and the domains on integrins to which they bind. For example, the N-terminal IgSF domain of ICAM-1 and ICAM-2 binds to the I domain of the integrin αLβ2, whereas the third IgSF domain of ICAM-1 binds to the I domain of the integrin αMβ2. The I domain is a domain of about 200 amino acids that is inserted in the α subunit of some, but not all, integrins. In those integrins in which it is present, the I domain plays an important role in ligand binding. MAdCAM-1 and VCAM-1 bind to the α4 integrins α4β7 and α4β1, respectively. The α4 subunit lacks an I domain. Not only the N-terminal IgSF domain of MAdCAM-1 and VCAM-1, but also their second IgSF domain is involved in the integrin interaction [16–18]. The similarities and differences in these binding properties may reflect specializations in these IgSF members and their integrin receptors.

Recently, the structures of the first two of five IgSF domains of ICAM-1 and the entire extracellular portion of ICAM-2 containing two IgSF domains have been determined by X-ray crystallography [19–21]. Together with the structure of the first two domains of VCAM-1 [22,23], they have revealed remarkable structural features of these cell adhesion molecules, including the disposition of an acidic residue in domain 1 (D1) that is critical for integrin recognition. MAdCAM-1 is unique among the IgSF integrin ligands in its specificity for α4β7. Furthermore, among this IgSF subclass it alone is expressed on a long mucin-like stalk and can bind to L-selectin. The amino acid sequence of MAdCAM-1 has some unique features: a long acidic insertion in domain 2 (D2), compared to D2 of VCAM-1; an arginine residue predicted by homology to VCAM-1 to be buried in the hydrophobic core of D1; and an unusually high content of proline in D2 (12.3% compared to 5.1% in average proteins) [24].

Here we describe the crystal structure of the N-terminal fragment of human MAdCAM-1 containing the two IgSF domains to 2.2 Å resolution. The structure shows some remarkably distinctive aspects even compared to VCAM-1 [22,23], the most closely related IgSF family member with 23.5% amino acid sequence identity, and suggests novel integrin-binding features.

Results and discussion

Structure determination

A fragment containing the two IgSF domains of human MAdCAM-1 was expressed in lectin-resistant Chinese hamster ovary (CHO) Lec.3.2.8.1 cells. The N-terminal sequence Val-Lys-Pro-Leu revealed that the signal sequence is 22 rather than 18 residues as previously predicted [11]. The fragment crystallized in space group $C222_1$ with unit-cell dimensions a = 65.8 Å, b = 101.1 Å, c = 70.0 Å and a calculated solvent content of 50.1%. The structure was determined using conventional multiple isomorphous replacement methods and was refined with X-PLOR (Table 1; [25]). The final R free and R factor are 0.280 and 0.223, respectively. The model contains 209 amino acid residues, one *N*-acetyl glucosamine (NAG) residue and 77 water molecules. The only regions in the model that remain weak in the electron-density map are residues 151–157 that form a loop, and the five C-terminal residues 205–209 that form the beginning of the mucin-like region. Of all the residues, 92.3% are in the favored region of the Ramachandran plot; only 3.6% are in the disallowed region and occupy a special place in the structure, as discussed below.

Domain 1 and the key integrin-binding residue

The N-terminal IgSF domain (D1) of MAdCAM-1 (Figure 1) is small in size (90 residues). IgSF domains have been previously classified into V, C1, C2 and I sets according to edge strand content and key framework residues [26]. The V and C1 sets of IgSF domain are so named based on similarity to antibody V and C domains, respectively. The V set consists of ABED strands on one β sheet and A′GFCC′C″ strands on the other β sheet, whereas the C1 set contains an ABED sheet and a GFC sheet. The C2 set is a variation of the C1 set. It lacks a D strand, but contains a normal C′ strand. The I set is an intermediate between the V and C1 sets. It has an ABED sheet and an A′GFCC′ or A′GFC sheet. The I set has recently been subdivided into I1 and I2 sets, analogous to the C1 and C2 division in that the I1 set has a D strand whereas the I2 set does not [20]. D1 of MAdCAM-1 may be classified as an I1 set,

Table 1

X-ray data collection and phasing statistics for MAdCAM-1.

Data set	Native	Se-Met	K_2PtCl_4	UO_2Ac
Data collection				
Resolution (Å)	2.2	3.2*	2.8	4.0
Reflections				
total number	87,803	79,151	51,386	13,359
unique	10,955	3917	5949	2094
Completeness (%)	92.2	99.9	99.2	100.0
$I/\sigma(I)$	13.9	18.6	12.0	12.7
R_{merge} (%)†	7.4	5.7	5.4	7.5
MIR phasing				
Concentration (mM)			1	0.2
Soak time (days)			6	1.5
Number of sites		1	1	1
R_{iso} (%)‡		8.9	10.1	10.2
R_{cullis} (acentric)§		0.84	0.88	0.90
Phasing power#		1.16	1.25	0.80
Figure of merit¶	0.40			
Model refinement				
Resolution (Å)	15.0–2.2			
R_{work} (%)	22.3			
R_{free} (%)	28.0			
Rms deviation bonds (Å)	0.012			
Rms deviation angles (°)	2.087			

*Data up to 3.5 Å were used for phasing.
†$R_{merge} = 100\ \Sigma_h\Sigma_i|I_i(h) - <I(h)>| / \Sigma_h\Sigma I_i(h)$, where $I_i(h)$ and $<I(h)>$ are the ith and mean measurement of the intensity of reflection h.
‡$R_{iso} = \Sigma|F_{ph} - F_p|/\Sigma F_p$, where F_{ph} and F_p are the derivative and native structure-factor amplitudes, respectively.
§R_{cullis}(Cullis R factor) $= \Sigma||F_{ph} \pm F_p| - F_{h(calc)}|/\Sigma|F_{ph} \pm F_p|$, where $F_{h(calc)}$ is the calculated heavy-atom structure factor amplitude.
#Phasing power $= <F_h>/E$, where $<F_h>$ is the root mean square heavy-atom structure-factor amplitude and E is the residual lack of closure error. ¶Figure of merit is the mean probability of a phase being correct.

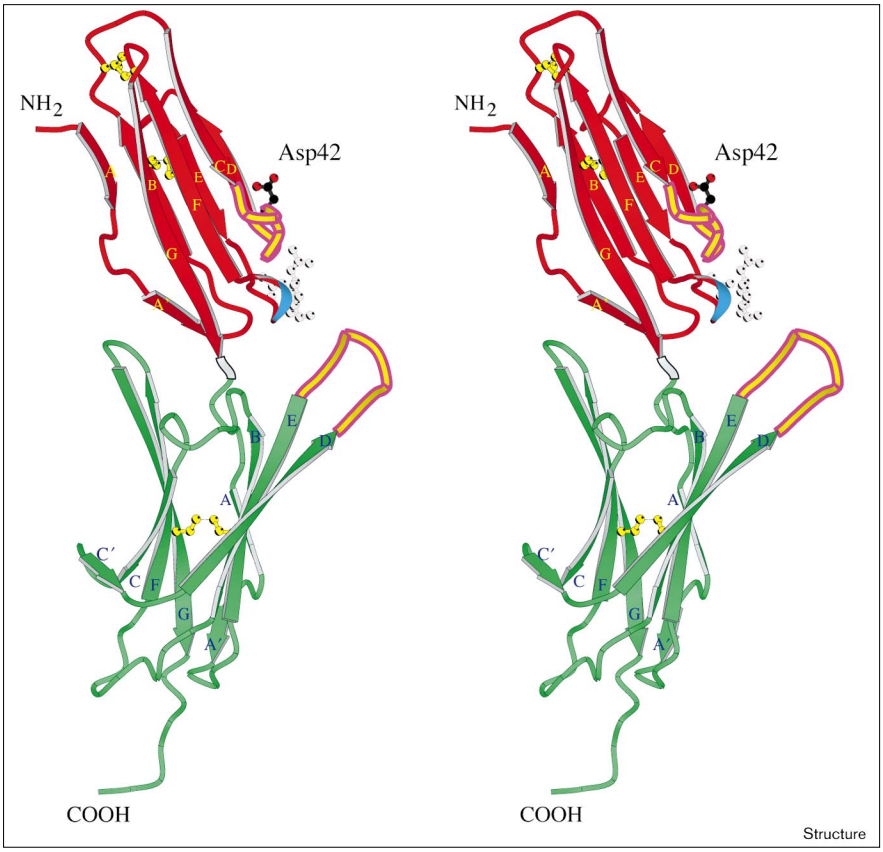
although it lacks a C′ strand. The A strand of D1 runs about half way down the edge of the ABE sheet, then crosses over the edge of the β sandwich and joins the A′GFC sheet at a kink which usually involves a *cis*-proline, as seen in the V set Ig framework. There is a conserved intersheet disulfide bond in the core of the domain. Furthermore, as in other IgSF integrin ligands [20,22,23,26], there is an 'extra' disulfide bond between Cys26 in the BC loop and Cys76 at the end of the F strand that consolidates the tip of the Ig domain. The residue Asp42, that is of key importance for integrin binding, is located in the CD loop [11,17,18]. The C and D strands form an 'edge' of the domain, as they are edge strands in the A′GFC and the ABED β sheets, respectively. The sidechain of Asp42 points up so as to be free for integrin binding.

In the case of VCAM-1, the critical integrin-binding residue Asp40 is also located in the CD loop [22,23]. In contrast, in the structures of ICAM-1 [20,21] and ICAM-2 [19] the key integrin binding glutamate residue is located at the end of the C strand on a rather flat surface. The aspartate residues in the CD loops of VCAM-1 and MAdCAM-1 are located in a kink that may be considered a vestige of the C′ strand found in some of the I1 set domains. There is no such vestige of C′ strand in ICAM-1 or ICAM-2. Thus, the structural environments around the acidic residue in VCAM-1 and MAdCAM-1 are distinct from those in ICAM-1 and ICAM-2. These differences correlate with binding to α4 integrins, that lack I domains, and to the integrin LFA-1 that does contain an I domain. There is, however, a key distinction even between the integrin recognition motifs in MAdCAM-1 and VCAM-1. In the overall superposition of D1 of MAdCAM-1 and VCAM-1, Asp42 in MAdCAM-1 and Asp40 in VCAM-1 lie 8 Å apart (Figures 2 and 3a; 53 of 90 residues gave a root mean square deviation [rmsd] = 1.35 Å). The aspartate residue is also less protruded in MAdCAM-1. These differences appear to be a consequence of the two-residue deletion in the CD loop of MAdCAM-1 compared to VCAM-1, and mainchain hydrogen bonds to Arg70 discussed below. If five-residue fragments centered at the critical aspartic acid residue are used for the superposition, however, the local conformation of this 'W-shaped' motif matches well (inset, Figure 3a). This W-shaped motif ensures that the acidic sidechain of the aspartate residue is markedly exposed for integrin binding, and may be a key structural determinant for α4 integrin binding. Furthermore, if the aspartate residues in the W-shaped motifs of MAdCAM-1 and VCAM-1 occupy a similar position in the integrin-binding site, then the remainder of D1 would be oriented quite differently. This may partly account for their preference for binding to α4β7 and α4β1 integrins, respectively.

A buried arginine residue

Arg70 plays a critical role in maintaining the conformation of the CD loop of D1. This arginine is fully buried in the domain and is surrounded by seven hydrophobic residues: Leu23, Trp38, Leu57, Val59, Ala66, Val85 and Leu87 (Figure 3b). Arg70 is located two residues before the conserved Cys68 on the F strand that forms the intersheet disulfide bond (Figure 1). The density for the sidechain of Arg70 was unambiguous even in the earliest electron-density maps. There are no negatively charged groups in the vicinity of Arg70 to neutralize its strong positive charge. Instead, the guanido group is placed in an optimal position to donate three hydrogen bonds to mainchain carbonyl oxygens: one to Ala66 (2.7 Å) in the beginning of the F strand, and the other two to Asp42 (2.5 Å) and Thr43 (2.6 Å) in the CD loop (Figure 3b). Buried, unneutralized arginines are rare, but have previously been found to have a structural role by forming multiple hydrogen bonds to backbone carbonyl oxygens [27]. In the case of MAdCAM-1, the hydrogen bonds to the two successive carbonyl groups of Asp42 and Thr43 appear to constrain the backbone to a γ-turn-like conformation around Asp42 (Figure 3b). The turn is characterized by a hydrogen bond between the CO group of Leu41 and the NH group of Thr43 forming a seven-membered ring. The hydrogen

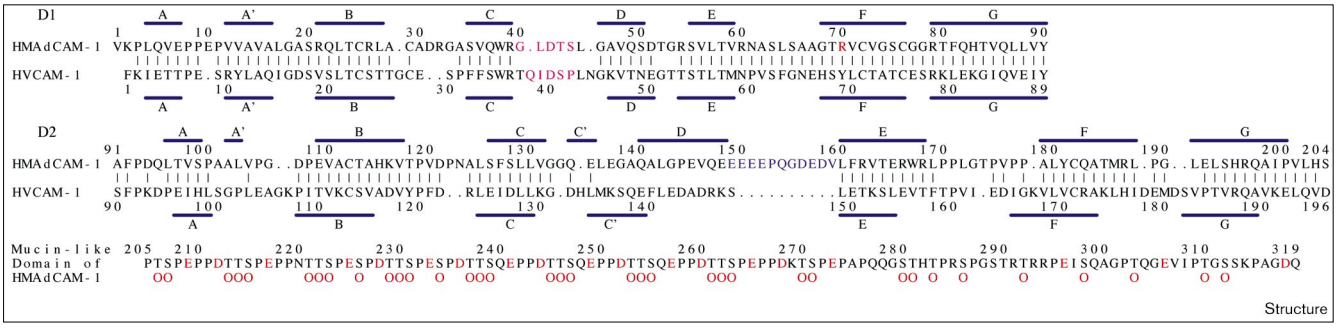
Figure 1



Stereo view ribbon drawing of the N-terminal two-domain fragment of human MAdCAM-1. The key Asp42 sidechain in domain 1 (D1) is shown in ball-and-stick representation with oxygens in red. Two loops involved in integrin binding, the CD loop on D1 and the DE β ribbon on domain 2 (D2) are highlighted in yellow delineated with red. The disulfide bonds and the *N*-acetylglucosamine (NAG) sugar group are also drawn in ball-and-stick form, in yellow and gray, respectively. Only the first NAG unit has been identified (with high B factors) whereas the others appear to be disordered. Site-directed mutagenesis to remove the glycosylation site found no significant effect on integrin-binding affinity (J Rosebrook, N Green, N Cochran, KT, J-hW, TAS and MJB, unpublished data). Residue Tyr90, the 'pivot' between D1 and D2 is shown in gray; a short piece of helix in D1 is shown in blue. (The figure was prepared using the program MOLSCRIPT [45].)

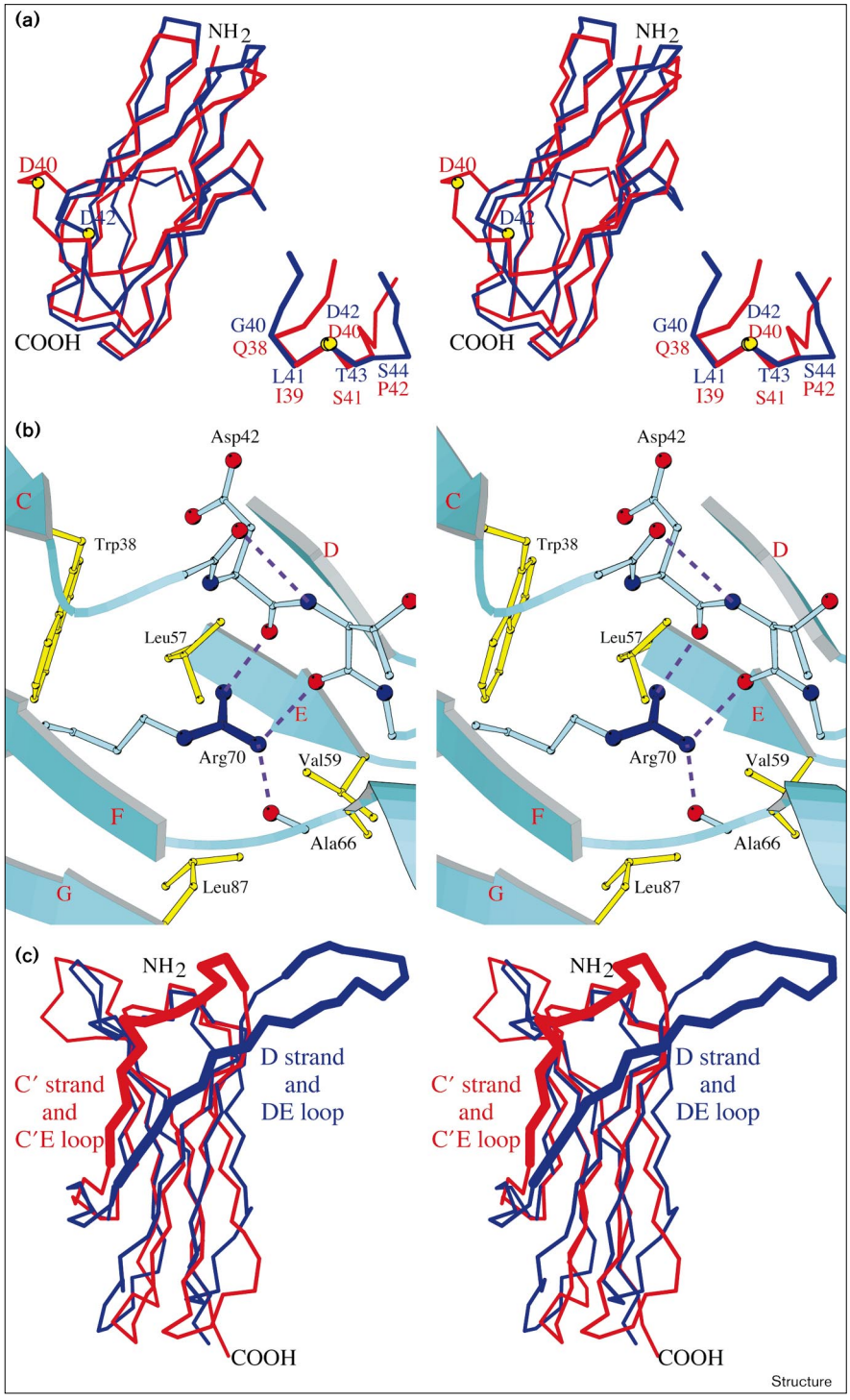
bond is weak, and barely within the allowed range, but this is generally true for this kind of seven-membered ring, an energetically less favorable conformation than the ten-membered ring found in a β turn [28]. The dihedral angles for the peptide segment of residues 42–44 are in the disallowed region on the Ramachandran plot. In particular, Asp42 has unusual mainchain dihedral angles ($\phi = 102°$ and $\psi = -69°$). The positive value of ϕ reflects steric strain and is often functionally relevant [28]. Arg70 not only helps to form the γ-turn-like structure, but also holds the CD loop and the EF loop of D1 together. These two structural elements define the integrin-binding region in D1, and thus

Figure 2



Structural alignment of the MAdCAM-1 and VCAM-1 amino acid sequences. Structures were aligned with defaults of 3DMALIGN in MODELLER [46]. The β strands defined by the program DSSP [47] are shown with blue lines. The W-shaped binding motifs in both MAdCAM-1 and VCAM-1 are highlighted in magenta. Residue Arg70 is shown in red and the DE β ribbon of D2 of MAdCAM-1 is highlighted in purple. The acidic residues in the mucin-like domain of MAdCAM-1 are shown in red; one 'O' is drawn in red beneath each predicted *O*-linked site [48]. (The figure was prepared using the program MOLSCRIPT [45].)

Figure 3



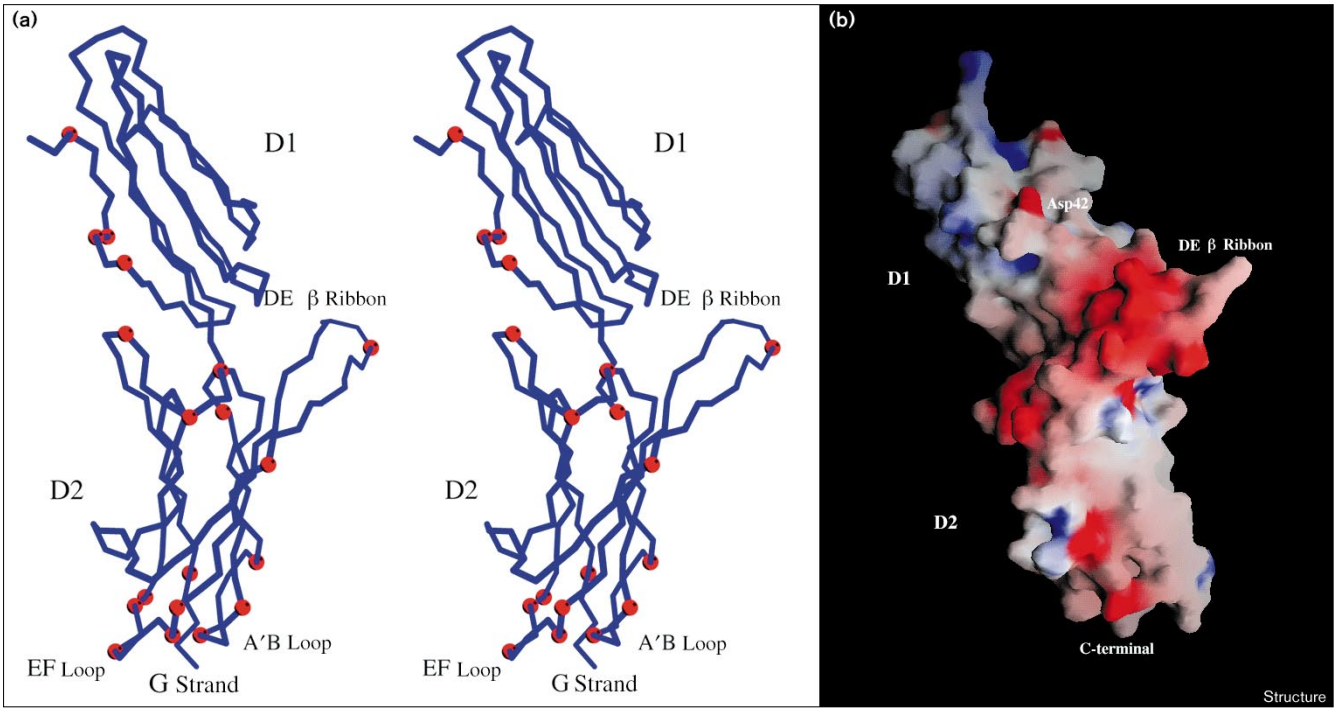
Superposition of domains 1 and 2 of MAdCAM-1 and VCAM-1 and the local environment of residue Arg70 in MAdCAM-1. **(a)** Stereo view superposition of D1 of MAdCAM-1 (blue) and VCAM-1 (red). The key integrin-binding residues, Asp42 in MAdCAM-1 and Asp40 in VCAM-1, are labeled. The structures were aligned as described in the text. Inset: a local superposition centered at the key aspartate residues from the two molecules, showing their common W-shaped binding structural motifs. **(b)** Stereo view of the environment of residue Arg70 of MAdCAM-1 (dark blue) and its hydrogen bonds (purple dashed lines) to three mainchain carbonyl oxygens. For clarity, only five out of the seven hydrophobic residues surrounding Arg70 are shown here (in yellow), and strands D and E are truncated. The γ-turn-like element centered at Asp42 is illustrated. **(c)** Stereo view superposition of D2 of MAdCAM-1 (blue) and VCAM-1 (red). The D strand and DE loop in MAdCAM-1, and part of the C′ strand and C′E loop in VCAM-1 are shown in thicker lines for comparison. (The figures were prepared using the program MOLSCRIPT [45]).

Arg70 has a key role in their conformation. Indeed, site-directed mutagenesis has recently shown that substitution of Arg70 with alanine abolishes integrin binding (J Rosebrook, N Green, N Cochran, KT, J-hW, TAS and MJB, unpublished data). In VCAM-1, a different type of elaborate hydrogen-bond architecture, that includes a β-turn hydrogen bond between Thr37 and Asp40, defines the local conformation of the integrin-binding loop [23].

Domain 2 and the hydrophobic interface between domains

D2 of MAdCAM-1 is unique compared to D2 of other known IgSF integrin ligands in that it contains a D strand

Figure 4



The distribution of proline residues in MAdCAM-1 and a predicted orientation of the molecule. **(a)** Stereo view of the distribution of proline residues in MAdCAM-1. The Cα atom of each proline is highlighted in red. The figure was prepared using the program MOLSCRIPT [45]. **(b)** Electrostatic potential surface representation of the two N-terminal domains of MAdCAM-1 (blue, positive; red, negative). The figure was prepared using the program GRASP [49].

and belongs to the I1 set (Figures 1, 2 and 3c). Domains that lack a D strand were previously classified as belonging to the C2 set [29]. C1 and C2 set domains lack an A′ strand, whereas the D2s of ICAM-1, ICAM-2 and VCAM-1 contain an A′ strand and a β-sheet framework that more closely resembles the I set [26]; therefore, they have been classified as the defining members of the I2 set of Ig domains [20]. With 114 residues, D2 of MAdCAM-1 is unusually large for an I set domain (Figures 1 and 2).

The boundary between D1 and D2 occurs between Tyr90 and Ala91 in MAdCAM-1. Tyr90 is part of the G strand in D1. Ala91 forms two mainchain hydrogen bonds to Thr118 in the BC loop in D2 (2.7 Å and 3.2 Å with proper geometry). After Thr118 there is a *cis* proline residue, Pro119. The *cis* conformation of the proline favors the hydrophobic interaction between its pyrrolidine ring and a group of hydrophobic residues from both D1 and D2, including Ala14, Met186 and Leu188. The homologous residues at the beginning of D2 in VCAM-1, ICAM-1 and ICAM-2 also form strong mainchain hydrogen bonds to the BC loop. The structure of the BC loop, and the *cis*-proline that follows the hydrogen-bonded residue in the BC loop, are conserved in all IgSF integrin ligands, including four examples of the VCAM-1 molecule and three examples of ICAM-1 that differ in the angle between D1 and D2 [20–23]. There is only one MAdCAM-1 molecule in the asymmetric unit of the crystal studied here, and therefore no evidence for inter-domain movement. Nevertheless, because the domain interface is predominantly hydrophobic in nature, we predict that domain–domain movement should be possible. Furthermore, because of the conservation of the hydrogen bond of Ala91 to the BC loop, we predict that if movement does occur Tyr90 in MAdCAM-1 would act as pivot similar to Tyr89 in VCAM-1 [30].

Distribution of proline residues and resistance to force at the D2 junction with the mucin-like region

During leukocyte–endothelial adhesion in the vasculature, tensile force, which can denature IgSF domains [31–33], will be exerted on MAdCAM-1. Force can be distributed over multiple β strands within a domain, and over multiple contacts between IgSF domains. The boundary of D2 with the mucin-like domain in MAdCAM-1 is likely to be the most susceptible region to force, as the entire force would be exerted on the end of strand G at the bottom of D2. D2 of MAdCAM-1 has several specialized features that appear to allow it to resist this force, however. As D2 belongs to the I set, β strand G is not an edge strand, but has backbone

hydrogen bonds on one side to strand F and on the other side to strand A′. Notably, the I set IgSF domains of the muscle proteins twitchin and titin, which function to resist tensile stress, have the same characteristics [34,35]. Multiple IgSF domains in the nervous system adhesion molecules NCAM and L1 are also known, or predicted, to belong to the I set and have the A–A′ kink [30]. D2 of MAdCAM-1 has an unusually high proline content (12.3% compared to 5.1% for average proteins) [24]; these prolines are concentrated on the bottom of D2 (Figure 3a). The proline residues are predicted to add rigidity, and buttress the bottom of D2 against the force applied to the end of the G strand during adhesion to integrins.

The unusual negatively charged DE β ribbon of D2

The most striking structural feature of MAdCAM-1 is a negatively charged DE β ribbon that extends from the D and E strands of D2 (Figures 1 and 4). Seven of the 11 residues of this extension, from position 149–159, are glutamate or aspartate residues. Mutation of residues Glu148, Glu151, Glu152, or Glu157 to alanine diminishes binding of α4β7 (J Rosebrook, N Green, N Cochran, KT, J-hW, TAS and MJB, unpublished data). Deletion of residues 143–150 also abolishes binding to α4β7 [17], however, this deletion almost certainly disrupts D2 as it removes four residues of strand D. Residues 149–159 extend 20 Å from the main body of D2, and have a conformation and twist like a slightly 'unzipped' β ribbon. The B factors of these residues are high, indicating that this extension is mobile. The DE extension may function to electrostatically orient the MAdCAM-1 integrin-binding face. ICAM-2 has been proposed to be oriented by tripod-like carbohydrates that are N-linked near the base of D2 [19]. In contrast, MAdCAM-1 extends far above the membrane on a mucin-like stalk that contains residues 205–319 immediately following D2. Electron microscopic studies of proteins containing similar mucin-like domains have shown that the mucin-like region has an extended, rod-like shape with an extension on average of 1.8–2.0 Å per residue [36]. The mucin-like domain of human MAdCAM-1 is thus estimated to extend approximately 22 nM and present the integrin binding IgSF domains well above the cell surface. The first segment of the mucin-like domain, 65 residues immediately following the second IgSF domain, contains 15 acidic residues and no basic residues. As cell surfaces have a net negative charge, it is likely that repulsive electrostatic forces make the mucin-like fragment point away from, and roughly perpendicular to, the membrane. We propose that the DE ribbon functions as a negatively charged 'antenna' to orient the IgSF domains. The Coulombic field established by the cell surface and mucin-like region will repel the antenna on D2, and orient it facing away from the cell surface. This action is predicted to optimally orient D1 and D2 such that the key integrin-binding residue (Asp42 in the CD loop) and the surrounding region of D1 and the DE β ribbon in D2 will both be well positioned for interaction with integrins on other cells.

Asp42 in the CD loop on D1 and the DE loop in D2 are positioned 20–30 Å away from one another on the same face of MAdCAM-1 (Figure 1). The DE loop might increase the on rate for binding to α4β7 by orienting MAdCAM-1 as described above, and also by guiding docking of a positively charged integrin domain onto MAdCAM-1. The DE loop might also directly bind α4β7. In murine VCAM-1, residues on the same face of D2 together with residues around the CD loop in D1 contribute to binding to both α4β1 and α4β7 [17]. Integrin binding to D1 and D2 in MAdCAM-1 and VCAM-1 could be analogous to binding of the α5β1 integrin to the Arg-Gly-Asp (RGD) motif in fibronectin type III domain 10 of fibronectin and to the 'synergy region' in the neighboring domain 9 [37].

The preference of α4β1 for VCAM-1 and α4β7 for MAdCAM-1 is mainly determined by the integrin β subunit. We have identified two important differences between VCAM-1 and MAdCAM-1: the orientation of the critical aspartate residue in D1 with respect to the rest of D1; and the presence of the DE antenna in D2 of MAdCAM-1. Both regions bear residues that are important for recognition by the α4β7 integrin. Further work will be required to determine the relative importance of these regions in integrin specificity, and to establish which sites the integrin α and β subunits bind.

Biological implications

Mucosal addressin cell adhesion molecule 1 (MAdCAM-1) is a multifunctional adhesion molecule expressed on the endothelium in mucosa. The molecule binds to receptors on leukocytes in the bloodstream, enabling trafficking of these immune cells to specific tissues. MAdCAM-1 is capable of initiating primary lymphocyte contact at two levels: to L-selectin, through its mucin-like region, as well as to the integrin α4β7, through its immunoglobulin-like (IgSF) domains. Furthermore, MAdCAM-1 binding to integrin α4β7 can also mediate transient rolling adhesion. MAdCAM-1 is the primary counter-receptor of integrin α4β7, and regulates lymphocyte trafficking to both normal and inflamed mucosal tissues. *In vitro*, pro-inflammatory cytokines have been shown to induce the expression of MAdCAM-1 on endothelium [38]. In addition, monoclonal antibodies to integrin α4β7 can block inflammatory bowel disease, suggesting that the α4β7–MAdCAM-1 interaction has an important role in the inflammatory process. Inhibition of MAdCAM-1 binding to integrin α4β7 can control the inflammatory process and is therefore an attractive target for drug design [39,40].

The crystal structure of the two N-terminal IgSF domains of MAdCAM-1 reveals many novel features important for interaction with its integrin receptor.

Comparison of MAdCAM-1 to other immunoglobulin superfamily integrin ligands shows important similarities and differences. These ligands share a similar overall immunoglobulin-like topology and all have their critical integrin-binding residues positioned at the edge of the first IgSF domain (domain 1). Intercellular adhesion molecule 1 (ICAM-1) and ICAM-2 bind to I-domain-containing integrins and have their key integrin-binding residues located on a flat surface. In contrast, MAdCAM-1 and vascular cell adhesion molecule 1 (VCAM-1) that bind to non-I-domain-containing integrins have their key integrin-binding residues located on protruding loops. MAdCAM-1 and VCAM-1 both contain an aspartate residue in domain 1 that is critical for integrin binding. An overall structure superposition of these domains, however, shows that this residue is shifted by 8 Å in the two structures. This, together with a slight difference in local conformation, may partly determine the specificity of their integrin-binding abilities. Furthermore, dramatic differences in structure between the domains 2 of MAdCAM-1 and VCAM-1, including the presence of an extended negatively charged β ribbon in MAdCAM-1, provide an additional mechanism for determining integrin specificity. The unique negatively charged β ribbon stretches out far from the body of domain 2. The ribbon may directly participate in integrin binding and help to optimally orient domains 1 and 2 for recognition by integrin during adhesion.

Materials and methods

Protein expression and purification

A recombinant MAdCAM-1 cDNA with a translation stop codon in position P232 of the precursor protein [11] was generated by the polymerase chain reaction (PCR) using *Pfu* DNA polymerase (Stratagene). The recombinant fragment was subcloned into the unique *Xho*I and *Not*I sites of the expression vector pBJ5/GS and used for transfection of the lectin resistant CHO Lec.3.2.8.1 cell line [19]. Clones secreting soluble recombinant MAdCAM-1 and growing in the presence of 100 μM L-methionine-sulfoximine were selected as described [19]. The supernatant of selected cell clones collected in the presence of 2 mM butyric acid was loaded on an anti-MAdCAM-1 antibody LS46 sepharose column. The column was extensively washed with 10 mM Tris-saline buffer (pH 8.0) and the protein eluted with five column volumes of 50 mM glycine-saline buffer (pH 3.2). Neutralized fractions containing the recombinant MAdCAM-1 were pooled and run through a Superdex 75 column (Pharmacia) with 10 mM Hepes buffer (pH 7.4). The purified protein was analyzed by SDS–PAGE and concentrated to about 17 mg/ml for crystallization. For preparation of the selenomethionine MAdCAM-1 derivative, CHO cell transfectants were grown in a Nunc Cell Factory device. Confluent cells were cultured for 10 h with Dulbecco's modified Engle's medium (DMEM) lacking glutamine and methionine, supplemented with 10% fetal calf serum (FCS) and containing 50 μg/ml of seleno-L-methionine (Sigma) for depletion of intracellular methionine. Culture medium was then replaced with the same medium and the supernatant was collected three days later for purification. The MAdCAM-1 protein concentration in the cell supernatant was about 1 mg/L and selenomethionine incorporation was around 90%, as revealed by amino acid analysis.

Crystallization and data collection

The crystals were grown using the conventional hanging-drop vapor diffusion method, with a crystallization solution containing 10% PEG 4K, 0.5 M Li_2SO_4 (pH 7.5–8.0). Each native and derivative X-ray diffraction data set was collected from a single crystal at room temperature on a 18 cm MAR Research area detector equipped with Rigaku RU200 rotating-anode X-ray generator. The crystal used to collect native data was about $0.3 \times 0.4 \times 0.6$ mm in dimensions and the crystal-to-detector distance was 85 mm. A total of 112° of data were collected with oscillation angle 1° and 10 min exposure per frame. The experimental setups for derivatives were varied depending on how far they diffracted. Reflections of each data set were integrated and scaled using the programs DENZO and SCALEPACK [41]. Subsequent analyses were performed using the CCP4 suite [42]. A summary of the data collection and processing statistics is presented in Table 1.

Structure determination and refinement

Initial phasing was achieved with selenomethionine-labeled MAdCAM-1. Although there is only one Se site in the 23 kDa molecular mass and diffraction data were collected at room temperature using an in-house X-ray source, the isomorphous difference Patterson map calculated at 3.5 Å resolution revealed a well resolved Se peak. The initial phases, though not good enough for calculating an interpretable map, helped to solve Pt and U derivatives using the difference Fourier technique. The heavy-atom parameters of these three derivatives were refined together and the multiple isomorphous replacement (MIR) phases were calculated to 2.8 Å resolution with the program MLPHARE [42] within the CCP4 suite. The electron-density maps were improved with procedures of density modification and phase extension to 2.2 Å using the program DM [43] in the CCP4 suite. Only solvent flattening and histogram matching were applied. The subsequently calculated electron-density map clearly showed the two domains of the molecule and most of their β strands. Some bulky sidechains which form the cores of the Ig domains were also recognized. Examination of electron-density map and model building were performed using the program O [44].

The initial model was built into a 2.5 Å map. Successive cycles of phase combination, difference Fourier map calculation and model rebuilding, allowed placement of about 70% of the polypeptide chain. Positional refinement, individual B-factor refinement and simulated-annealing refinement were carried out with X-PLOR using data to 2.2 Å [25], improving the structure progressively. Alternate model building and refinement eventually led to the completion of the structure. Only peaks above 3σ in the $F_o–F_c$ difference map and consistent with the $2F_o–F_c$ map were accepted as water molecules if their geometry met the requirement of forming hydrogen bonds to protein atoms or previously placed water molecules.

Accession numbers

The atomic coordinates of the human MAdCAM-1 crystal structure have been deposited with the Brookhaven Protein Data Bank with accession number 1mad.

Acknowledgments

The authors thank Jay R Luly, Julia Pinto, Brian Matthews and Walter Newman for helpful discussion; Mai-kun Teng for help in computation; Linda Chee for assistance with protein purification; Luisa M Stamm for assistance in some heavy-atom screening experiments; and Ellis L Reinherz for reading of the manuscript. This work was supported by NIH grants to JHW and TAS, and a grant from LeukoSite to JHW.

References

1. Springer, T.A. (1994). Traffic signals for lymphocyte recirculation and leukocyte emigration. *Cell* **76**, 301-314.
2. Picker, L.J. & Butcher, E.C. (1992). Physiological and molecular mechanisms of lymphocyte homing. *Annu. Rev. Immunol.* **10**, 561-591.
3. Butcher, E.C. & Picker, L.J. (1996). Lymphocyte homing and homeostasis. *Science* **272**, 60-66.
4. Lauffenburger, D.A. & Horwitz, A.F. (1996). Cell Migration: a physically integrated molecular process. *Cell* **84**, 359-369.
5. Springer, T.A. (1990). Adhesion receptors of the immune system. *Nature* **346**, 425-434.

6. Alon, R., Kassner, P.D., Carr, M.W., Finger, E.B., Hemler, M.E. & Springer, T.A. (1995). The integrin VLA-4 supports tethering and rolling in flow on VCAM-1. *J. Cell Biol.* **128**, 1243-1253.
7. Berlin, C., *et al.*, & Butcher, E.C. (1995). α4 Integrins mediate lymphocyte attachment and rolling under physiologic flow. *Cell* **80**, 413-422.
8. Bargatze, R.F., Jutila, M.A. & Butcher, E.C. (1995). Distinct roles of L-selectin and integrins alpha 4 beta 7 and LFA-1 in lymphocyte homing to Peyer's patch-HEV in situ: the multistep model confirmed and refined. *Immunity* **3**, 99-108.
9. Streeter, P.R., Lakey-Berg, E., Rouse, B.T.N., Bargatze, R.F. & Butcher, E.C. (1988). A tissue-specific endothelial cell molecule involved in lymphocyte homing. *Nature* **331**, 41-46.
10. Briskin, M.J., *et al.*, & Ringler, D.J. (1997). Human mucosal addressin cell adhesion molecule-1 is preferentially expressed in intestinal tract and associated lymphoid tissue. *Am. J. Pathol.* **151**, 97-110.
11. Shyjan, A.M., Bertagnolli, M., Kenney, C.J. & Briskin, M.J. (1996). Human mucosal addressin cell adhesion molecule-1(MAdCAM-1) demonstrates structural and functional similarities to the α4β7-integrin binding domains of murine MAdCAM-1, but extreme divergence of mucin-like sequences. *J. Immunol.* **156**, 2851-2857.
12. Briskin, M.J., McEvoy, L.M. & Butcher, E.C. (1993). MAdCAM-1 has homology to immunoglobulin and mucin-like adhesion receptors and to IgA1. *Nature* **461**, 461-464.
13. Berlin, C., *et al.*, & Butcher, E.C. (1993). α4β7 Integrin mediates lymphocyte binding to the Mucosal vascular addressin MAdCAM-1. *Cell* **74**, 185-195.
14. Hu, M.C.-T., Crowe, D.T., Weissmann, I.L. & Holzmann, B. (1992). Cloning and expression of mouse integrin βp(β7): a functional role in Peyer's patch-specific lymphocyte homing. *Proc. Natl Acad. Sci. USA* **89**, 8254-8258.
15. Berg, E.L., McEvoy, L.M., Berlin, C., Bargatze, R.F. & Butcher, E.C. (1993). L-selectin-mediated lymphocyte rolling on MAdCAM-1. *Nature* **366**, 695-698.
16. Briskin, M.J., Rott, L.S. & Butcher, E.C. (1996). Structural requirements for mucosal vascular addressin binding to its lymphocyte receptor α4β7. *J. Immunol.* **156**, 719-726.
17. Newham, P., *et al.*, & Humphries, M.J. (1997). α4 Integrin binding interfaces on VCAM-1 and MAdCAM-1. *J. Biol. Chem.* **272**, 19429-19440.
18. Viney, J.L., *et al.*, & Fong, S. (1997). Mucosal addressin cell adhesion molecule-1, a structural and functional analysis demarcates the integrin binding motif. *J. Immunol.* **157**, 2488-2497.
19. Casasnovas, J.M., Springer, T.A., Liu, J.-h., Harrison, S.C. & Wang, J.-h. (1997). The crystal structure of ICAM-2 reveals a distinctive integrin recognition surface. *Nature* **387**, 312-315.
20. Casasnovas, J.M., Stehle, T., Liu, J.-h., Wang, J.-h. & Springer, T.A. (1998). A dimeric crystal structure for the N-terminal two domains of ICAM-1. *Proc. Natl Acad. Sci. USA* **95**, 4134-4139.
21. Bella, J., Kolatkar, P.R., Marlor, C.W., Greve, J.M. & Rossmann, M.G. (1998). The structure of the two amino-terminal domains of human ICAM-1 suggests how it functions as a rhinovirus receptor and as an LFA-1 integrin ligand. *Proc. Natl Acad. Sci. USA* **95**, 4140-4145.
22. Jones, E.Y., *et al.*, & Stuart, D.I. (1995). Crystal structure of an integrin-binding fragment of vascular cell adhesion molecule-1 at 1.8 Å resolution. *Nature* **373**, 539-544.
23. Wang, J.-h., *et al.*, & Osborn, L. (1995). The crystal structure of an N-terminal two domain fragment of VCAM-1: a cyclic peptide based on the domain 1 C-D loop can inhibit VCAM-1/α4 integrin. *Proc. Natl Acad. Sci. USA* **92**, 5714-5718.
24. McCalden, P. & Argos, P. (1988). Oligopeptide biases in protein sequences and their use in predicting protein coding regions in nucleotide sequences. *Proteins* **4**, 99-122.
25. Brünger, A.T. (1992). *X-PLOR. Version 3.1: a System for Crystallography and NMR*. Yale University Press, New Haven, CT.
26. Harpaz, Y. & Chothia, C. (1994). Many of the immunoglobulin superfamily domains in cell adhesion molecules and surface receptors belong to a new structural set which is close to that containing variable domains. *J. Mol. Biol.* **238**, 528-539.
27. Borders Jr., C.L., *et al.*, & Pett, V.B. (1994). A structural role for arginine in proteins: multiple hydrogen bonds to backbone carbonyl oxygens. *Protein Sci.* **3**, 541-548.
28. Herzberg, O. & Moult, J. (1991). Analysis of the steric strain in the polypeptide backbone of protein molecules. *Proteins* **11**, 223-229.
29. Williams, A.F. & Barclay, A.N. (1988). The immunoglobulin superfamily: domains for cell surface recognition. *Annu. Rev. Immunol.* **6**, 381-405.
30. Wang, J.-h., Stehle, T., Pepinsky, B., Liu, J.-h., Karpusas, M. & Osborn, L. (1996). Structure of a functional fragment of VCAM-1 refined at 1.9 Å resolution. *Acta Cryst. D* **52**, 369-379.
31. Tskhovrebova, L., Trinick, J., Sleep, J.A. & Simmons, R.M. (1997). Elasticity and unfolding of single molecules of the giant muscle protein titin. *Nature* **387**, 308-312.
32. Rief, M., Gautel, M., Oesterhelt, F., Fernandez, J.M. & Gaub, H.E. (1997). Reversible unfolding of individual titin immunoglobulin domains by AFM. *Science* **276**, 1109-1112.
33. Kellermayer, M.S., Smith, S.B., Granzier, H.L. & Bustamante, C. (1997). Folding-unfolding transitions in single titin molecules characterized with laser tweezers. *Science* **276**, 1112-1116.
34. Pfuhl, M. & Pastore, A. (1995). Tertiary structure of an immunoglobulin-like domain from the giant muscle protein titin: a new member of the I set. *Structure* **3**, 391-401.
35. Fong, S., *et al.*, & Clarke, J. (1996). Structure and stability of an immunoglobulin superfamily domain from twichin, a muscle protein of the nematode C. elegans. *J. Mol. Biol.* **264**, 624-639.
36. Li, F., Erickson, H.P., Janes, J.A., Moore, K.L., Cummings, R.D. & McEver, R.P. (1996). Visualization of P-selectin glycoprotein ligand-1 as a highly extended molecule and mapping of protein epitopes for monoclonal antibodies. *J. Biol. Chem.* **271**, 6342-6348.
37. Leahy, D.J., Aukhil, I. & Erickson, H.P. (1996). 2.0 Å crystal structure of a four-domain segment of human fibronectin encompassing the RGD loop and synergy region. *Cell* **84**, 155-164.
38. Sikorski, E.E., Hallmann, R., Berg, E.L. & Butcher, E.C. (1993). The Peyer's patch high endothelial receptor for lymphocytes, the mucosal addressin, is induced on a murine endothelial cell line by tumor necrosis factor-α and IL-1. *J. Immunol.* **151**, 5239-5250.
39. Hesterberg, P.E., *et al.*, & Ringler, D.J. (1996). Rapid resolution of chronic colitis in the cotton-top tamarin with an antibody to a gut-homing integrin alpha 4 beta 7. *Gastroenterology* **111**, 1373-1380.
40. Picarella, D., Hurlbut, P., Rottman, J.R., Shi, X., Butcher, E. & Ringlor, D.J. (1997). Monoclonal antibodies specific for beta 7 integrin and mucosal addressin cell adhesion molecule-1 (MAdCAM-1) reduce inflammation in the colon of scid mice reconstituted with CD45RBhigh CD4+ T cells. *J. Immunol.* **158**, 2099-2106.
41. Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307-326.
42. *Collaborative Computational Project Number 4*. (1994). The CCP4 suit: program for protein crystallography. *Acta Cryst. D* **50**, 760-763.
43. Cowtan, K. (1994). DM: an automated procedure for phase improvement by density modification. *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography* **31**, 34-38.
44. Jones, T.A. & Kjeldgaard, M. (1997). *O– The Manual Version 6.10, Uppsala University*. Uppsala University, Uppsala, Sweden.
45. Kraulis, P. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Cryst.* **24**, 924-950.
46. Sali, A. & Blundell, T.L. (1993). Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **234**, 779-815.
47. Kabsch, W. & Sander, C. (1983). Dictionary of protein structure: pattern hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577-2637.
48. Hansen, J.E., *et al.*, & Brunak, S. (1995). Prediction of *O*-glycosylation of mammalian proteins: specificity patterns of UDP-GalNAc:polypeptide *N*-acetygalactosaminyltransferase. *Biochem. J.* **308**, 801-813.
49. Nicholls, A.J., Sharp, K.A. & Honig, B. (1991). Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* **11**, 281-296.