Protein Science (1995), *4*:1644–1647. Cambridge University Press. Printed in the USA.
Copyright © 1995 The Protein Society

FOR THE RECORD

Molecular model of the N-terminal receptor-binding domain of the human CD6 ligand ALCAM

JÜRGEN BAJORATH, MICHAEL A. BOWEN, AND ALEJANDRO ARUFFO
Bristol-Myers Squibb Pharmaceutical Research Institute, Seattle, Washington 98121

(RECEIVED May 22, 1995; ACCEPTED June 14, 1995)

Abstract: CD6-ligand interactions have been implicated in the regulation of T-cell adhesion and activation. CD6 is a member of the scavenger receptor family, whereas its human ligand (ALCAM) belongs to the immunoglobulin superfamily. The extracellular region of ALCAM includes five immunoglobulin-like domains. As a fusion protein, the N-terminal extracellular domain of ALCAM (ALCAMD1) binds specifically to CD6. We report the construction, assessment, and analysis of a molecular model of ALCAMD1. The model defines the CDR-analogous loops, the location of *N*-linked glycosylation sites, and residues that form the β-sheet faces of the immunoglobulin-like domain. Predicted structural characteristics of the A′GFCC′C″ face of the model are consistent with the presence of monomeric and dimeric forms of ALCAMD1, which has implications for the receptor-ligand interactions.

Keywords: comparative modeling; CD6 ligand; immunoglobulin superfamily; T-cell functions

The CD6 receptor (Aruffo et al., 1991) is a member of the emerging scavenger receptor cysteine rich (SRCR) (type I membrane) protein superfamily (Aruffo et al., 1991; Resnick et al., 1994). CD6 is predominantly expressed on thymocytes, subpopulations of B-cells, and mature T-cells. Antibody crosslinking studies have implicated CD6 in co-regulating T-cell functions (Morimoto et al., 1988). A cDNA encoding a human cell surface protein that specifically binds to CD6 has recently been cloned (Bowen et al., 1995). This CD6 ligand, a type I membrane glycoprotein, is expressed by thymic epithelial cells (Wee et al., 1994; Patel et al., 1995), activated T-cells, B-cells, and monocytes (Bowen et al., 1995) and was named ALCAM (activated leukocyte cell adhesion molecule). ALCAM is a member of the immunoglobulin (Ig) superfamily (IgSF) and displays significant sequence homology to the extracellular domains of BEN (DM-GRASP), a chicken brain protein (Burns et al., 1991; Pourquie et al., 1992), and neurolin, its fish homologue (Lawssing et al., 1994). The extracellular region of these proteins includes five Ig-like (two V[ariable]-like, followed by three constant C[onstant]-like) domains. An Ig fusion protein including only the N-terminal ALCAM domain (ALCAMD1) is capable of binding to CD6. Thus, major structural determinants of ALCAM–CD6 interactions must be located within ALCAMD1 (although other domains may contribute to high-avidity CD6 interactions). As a first step to identify molecular regions in ALCAMD1 critical for the CD6-ALCAM interaction, we have attempted to construct a detailed three-dimensional model of ALCAMD1.

An initial sequence search in the Brookhaven Protein Data Bank revealed weak sequence identity of ALCAMD1 to immunoglobulin variable (light) chains. For random alignments of complete domains, the maximum sequence identity was ~15%. Among these sequences was the V-like domain of CD8 (Leahy et al., 1992), which displayed fewer insertions and deletions relative to ALCAMD1 than other Ig V-domains. A more structure-oriented alignment of ALCAMD1 and CD8 was generated manually to avoid gaps in β-strands. The majority of im-

```
                    * * * *         * *
CD8 :PLDRTWN LGE TVELKCQVL LSNPTS GCSWLFQP RGAAAS PTFLLYL SQNK
     (A')        (B)              (C)             (C')
     1         10          20          30          40
ALC :WYTVNSA YGD TIIIPCRLD vpqnlm FGKWKYEk pdgs-- PVFIAFr sstk
                          CDR1                          CDR2

                   **           * *       * * * *
CD8:  PKAAE GLDTQ RFSGKRL -GDT FVLTLS DFRRENE GYYFCSAL SNSI
      (C'')       (D)           (E)             (F)

      50          60          70          80          90
ALC:  KSVqv ddvp  EYKDRLn lsen YTLSIS NARisde kRFVCMLV tedn
                          CDR4                          CDR3

CD8:  MYFSHFVPVFLPA
        (G)
        100
ALC:  VFEAPTIVKVFKQ
```

Fig. 1. Structure-oriented sequence alignment of ALCAMD1 (ALC) and CD8. The β-strands of CD8 and the corresponding (predicted) ALCAMD1 β-strands are underlined (the A-strand in CD8, predicted to be absent in ALCAMD1, is not shown). Residue numbers are given for ALCAMD1. The predicted CDR-analogous regions in ALCAMD1 are annotated, and IgSF V-set consensus residues are labeled (*). ALCAMD1 loop regions modeled by conformational search are given in lower case. Residues identical in CD8 and ALCAMD1 and the most conservative replacements are shown in bold.

Reprint requests to: Jürgen Bajorath, Bristol-Myers Squibb Pharmaceutical Research Institute, 3005 First Avenue, Seattle, Washington 98121; e-mail: bajorath@protos.bms.com.

munoglobulin superfamily V-set consensus residues (Williams & Barclay, 1988) are conserved in ALCAMD1. The positions of these residues and the pattern of buried (core) and exposed residues in the CD8 structure (defining the register of β-strands) aided in the generation of the alignment. ALCAMD1 is predicted to lack the A-strand, but to include the A′-strand. Taking the most conservative mutations into account, the sequence similarity between CD8 and ALCAMD1 in the structure-oriented alignment (Fig. 1) is ~35%.

Direct assignment of ALCAMD1 residues to framework β-strands was possible for six of nine strands based on conservation of IgSF consensus residues and other sequence similarities (Fig. 1). For β-strands C″, D, and G, exact assignments were difficult on the basis of sequence comparison alone. Alternative local alignments were possible, effectively changing the register of the strands and the size of one or two adjacent loops. Therefore, several models with alternative alignments in these regions were built and assessed by position-dependent energy profile analysis (see below). The energetically most favorable alignments were identified (Fig. 1) and used to build the final ALCAMD1 model (including residues 1–106). Five of the eight loop regions connecting the framework β-strands were predicted to have the same length in CD8 and ALCAMD1, but only one of these eight loops, the short A′-B loop, was similar in sequence. Therefore, the conformations of seven loop regions (Fig. 1) were remodeled by systematic conformational search (Bruccoleri & Novotny, 1992).

Figure 2 (top) shows the ALCAMD1 model (see Kinemage 1). The A′GFCC′C″ and BED β-sheet faces and the complementarity determining region (CDR)-analogous loops of ALCAMD1

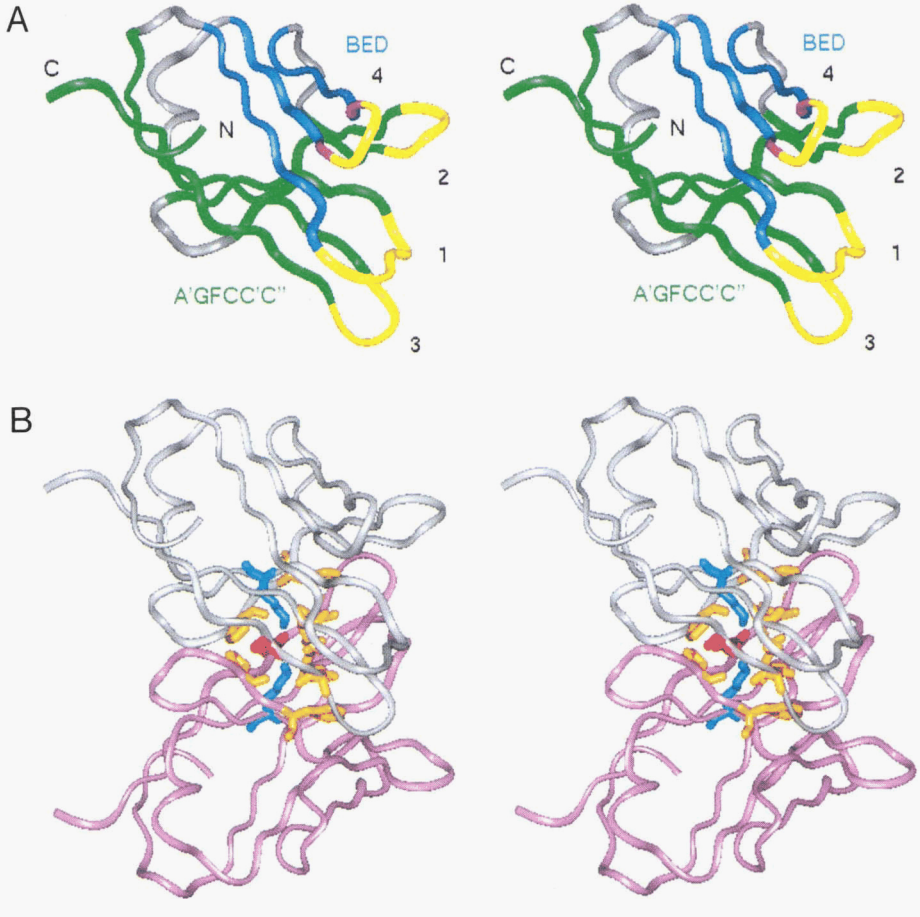

Fig. 2. Stereo representation of the ALCAMD1 model. **A:** ALCAMD1 monomer, backbone shown as a solid ribbon. The termini (N, C), the β-sheet faces (A′GFCC′C″, green; BED, blue), and the CDR-analogous loops (1–4, yellow) are labeled. Positions of potential N-linked glycosylation sites are colored in magenta. **B:** Possible Fv-like homodimer with monomers colored in gray and pink, respectively. ALCAMD1 at the top and the gray monomer at the bottom have the same orientation. The spatial arrangement of complementary negatively charged (red), positively charged (blue), and hydrophobic (gold) residues thought to be consistent with Fv-like dimerization (as discussed in the text) can be seen. Modeling methods: For computer graphics, model building, and energy minimizations, the InsightII/Discover package (version 2.3.0.; Biosym Technologies, San Diego) was used. Loop conformations were modeled by conformational search using CONGEN (version 2) (Bruccoleri & Novotny, 1992). For each loop, the conformation with minimum solvent-accessible surface within 3 kcal/mol of the energy minimum conformation was selected. Side-chain conformations of surface residues were modeled using low-energy side-chain rotamers (Ponder & Richards, 1987). Conservative side-chain replacements in core regions were modeled in conformations as similar as possible to the original residue. The initially assembled model was subjected to some minor energy minimization (until the RMS derivative of the energy function was ~1.5 kcal/mol/Å) to relieve stereochemical constraints. This resulted in a cumulative protein backbone RMS deviation (of the β-strands relative to the structural template) of less than 0.5 Å.

are highlighted. Details of the modeling procedure are given in the legend of Figure 2. When assessed with PROCHECK (Laskowski et al., 1993), the model displayed no unfavorable intramolecular contacts and good stereochemistry. The sequence-structure compatibility of the model was assessed by position-dependent energy profile analysis using PROSAII (version 2.0) (Sippl, 1993). The energy profile is shown in Figure 3. The negative average energy of the ALCAMD1 model and the trace of the profile indicate an overall correctly folded structure (Sippl, 1993). Several charged residues are exposed on both β-sheet faces and the CDR2- and CDR3-analogous regions. In the model, alternative ionic interactions are possible, dependent on selected side-chain rotamers, between residues R77-D80 and E58-R77, respectively. The latter interaction spatially corresponds to an intersheet salt bridge contact often seen in Ig-V domains (Williams & Barclay, 1988). ALCAMD1 contains two possible N-linked glycosylation sites that are located in the CDR4 region (i.e., the (D-E) loop, which is spatially adjacent to the CDR-analogous loops). Therefore, the N-linked glycosylation sites in ALCAMD1 map to a solvent-exposed region of the molecule and are proximal to the BED face of ALCAMD1, but opposite to the A'GFCC'C" face.

The CD8 X-ray structure consists of homodimers (Leahy et al., 1992) and dimerization is, as in antibody variable fragments (Fv), mediated by conserved β-bulges in the C'- and G-strands (Chothia et al., 1985). ALCAMD1 displays a conserved C'-strand bulge sequence motif (P38-V39-F40-I41-A42). The putative ALCAMD1 G-strand bulge region sequence (E96-A97-P98-T99) departs from the consensus motif (F-G-X-G). This is also seen in CD8, which, nevertheless, shows a G-strand β-bulge that supports dimerization (Leahy et al., 1992). In addition to the C'-strand bulge residues in ALCAMD1, hydrophobic residues (F43, V85, V94) are found at positions that are buried at V-domain interfaces (Novotny & Haber, 1985). Residues K30 and E96 in ALCAMD1 are notable exceptions to the conservation of hydrophobic V-domain interface residues. Using the most conserved Ig-V framework segments (Novotny & Sharp, 1992), we have superposed the ALCAMD1 model on the X-ray structure of the REI VL-VL dimer (Epp et al., 1975) to study the predicted A'GFCC'C" face in more detail (see Kinemage 2). Intermolecular contacts in the initially obtained dimer were improved by some minor energy refinement. In this hypothetical Fv-like dimer, the two charged residues K30 and E96 formed complementary charge–charge interactions between monomers. Figure 2 (bottom) shows that these complementary ionic interactions are surrounded by hydrophobic residues interacting across the interface (F40-M87-V94 and P38, respectively). No burial of uncompensated charges, prohibitive for the formation of hydrophobic interfaces, was found. In accord with these observations, energy profile analysis suggested sequence-structure compatibility of the dimer. The bulge region sequences, hydrophobic residues, and charged residues implicated in dimer interactions are conserved in ALCAM's chicken homologue (Burns et al., 1991; Pourquie et al., 1992), suggesting the possibility of structural and/or functional relevance.

The analysis of the model suggests that Fv-analogous dimerization of ALCAMD1 may be possible, although structural characteristics of the A'GFCC'C" β-sheet face depart from classical Ig V-domains. However, the presence of charged residues on the A'GFCC'C" face suggests that ALCAMD1, in contrast to many antibody V-domains, should be stable as a monomer. Thus, Fv-like dimerization may, for example, be caused by receptor interactions or, alternatively, be driven by dimerization or oligomerization of ALCAM's other domains. However, if homodimerization plays a role, the A'GFCC'C" face of ALCAMD1, which represents the adhesion site in (IgSF members) CD2 (Arulanandam et al., 1993) and VCAM-1 (Jones et al., 1995), would not be available for specific interactions with CD6. Coordinates of the ALCAMD1 model are included as supplementary material in the Electronic Appendix.

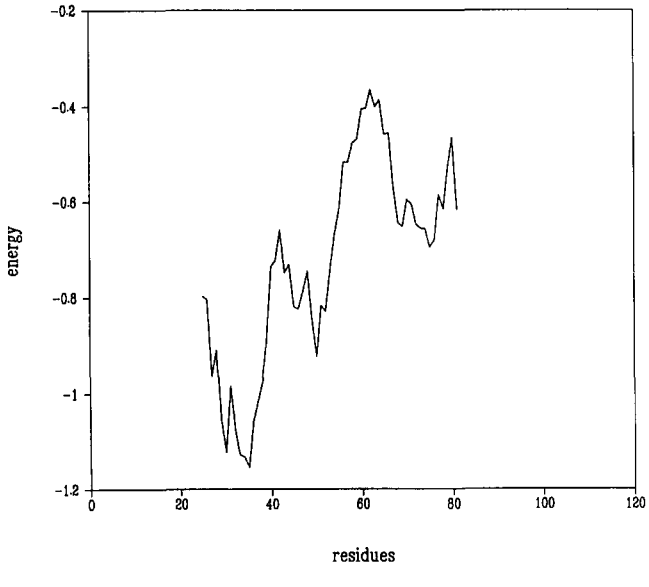

Fig. 3. Position-dependent energy profile of the ALCAMD1 model. Pairwise residue interaction energy is given as E/kT (E, interaction energy in kcal/mol; k, Boltzmann constant; T, absolute temperature in K). The β-carbon interactions were used to calculate the residue interactions energies. At each position, energy values were averaged using a 50-residue window (Sippl, 1993).

References

Aruffo A, Melnick MB, Linsley PS, Seed B. 1991. The lymphocyte glycoprotein CD6 contains a repeated domain structure characteristic of a new family of cell surface and secreted proteins. *J Exp Med 174*:949–952.

Arulanandam ARN, Withka JM, Wyss DF, Wagner G, Kister A, Pallai P, Recny MA, Reinherz EL. 1993. The CD58 (LFA3) binding site is a localized and highly charged surface area on the AGFCC'C" face of the human CD2 adhesion domain. *Proc Natl Acad Sci USA 90*:11613–11617.

Bernstein FC, Koetzle TF, Williams GJB, Meyer EF Jr, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M. 1977. The Protein Data Bank: A computer-based archival file for macromolecular structures. *J Mol Biol 112*:535–542.

Bowen MA, Patel DD, Li X, Modrell B, Malacko AR, Wang WC, Marquardt H, Neubauer M, Pesando J, Francke U, Haynes BF, Aruffo A. 1995. Cloning, mapping, and characterization of activated leukocyte cell adhesion molecule (ALCAM), a CD6 ligand. *J Exp Med 181*:2213–2220.

Bruccoleri RE, Novotny J. 1992. Antibody modeling using the conformational search program CONGEN. *ImmunoMethods 1*:96–106.

Burns FR, von Kannen S, Guy L, Raper A, Kamholz J, Chang S. 1991. DM-GRASP, a novel immunoglobulin superfamily axonal surface protein that supports neurite extension. *Neuron 7*:209–220.

Chothia C, Novotny J, Bruccoleri RE, Karplus M. 1985. Domain association in immunoglobulin molecules: The packing of variable domains. *J Mol Biol 186*:651–663.

Epp O, Lattman EE, Schiffer M, Huber R, Palm W. 1975. The molecular

structure of a dimer composed of the variable portions of the Bence-Jones protein REI refined at 2.0 Å resolution. *Biochemistry 14*:4943–4952.

Jones EY, Harlos K, Bottomley MJ, Robinson RC, Driscoll PC, Edwards RM, Clements JM, Dudgeon TJ, Stuart DI. 1995. Crystal structure of an integrin-binding fragment of vascular cell adhesion molecule-1 at 1.8 Å resolution. *Nature 373*:539–544.

Laskowski RA, MacArthur MW, Moss DS, Thornton JM. 1993. PROCHECK: A program to check the stereochemical quality of protein structures. *J Appl Crystallogr 26*:283–291.

Lawssing U, Giordano S, Stecher B, Lottspeich F, Stuermer CAO. 1994. Molecular characterization of fish neurolin: A growth-associated cell surface protein and member of the immunoglobulin superfamily in the fish retinotectal system with similarities to chicken DM-GRASP/SC-1/BEN. *Differentiation 56*:21–29.

Leahy DJ, Axel R, Hendrickson WA. 1992. Crystal structure of a soluble form of the human T cell coreceptor CD8 at 2.6 Å resolution. *Cell 68*: 1145–1162.

Morimoto C, Rudd CE, Letvin NL, Hagan M, Schlossman SF. 1988. 2H1—A novel antigen involved in T lymphocyte triggering. *J Immunol 140*: 2165–2170.

Novotny J, Haber E. 1985. Structural invariants of antigen binding: Comparison of VL-VH and VL-VL dimers. *Proc Natl Acad Sci USA 82*: 4592–4596.

Novotny J, Sharp KA. 1992. Electrostatic fields in antibodies and antibody/antigen complexes. *Prog Biophys Mol Biol 58*:203–224.

Patel DD, Wee SF, Whichard LP, Bowen MA, Pesando JM, Aruffo A, Haynes BF. 1995. Identification and characterization of an 100 kDa ligand for CD6 on human thymic epithelial cells. *J Exp Med 181*:1563–1568.

Ponder JW, Richards FM. 1987. Tertiary templates for proteins. Use of packing criteria in the enumeration of allowed sequences for different structural classes. *J Mol Biol 193*:775–791.

Pourquie O, Corbel C, Le Caer JP, Rossier F, Le Douarin NM. 1992. BEN, a surface glycoprotein of the immunoglobulin superfamily, is expressed in a variety of developing systems. *Proc Natl Acad Sci USA 89*:5261–5265.

Resnick D, Pearson A, Krieger M. 1994. The SRCR superfamily: A family reminiscent of the Ig superfamily. *Trends Biochem Sci 19*:5–8.

Sippl MJ. 1993. Recognition of errors in three-dimensional structures of proteins. *Proteins Struct Funct Genet 17*:355–362.

Wee SF, Wang WC, Farr AG, Nelson AJ, Patel DD, Haynes BF, Linsley PS, Aruffo A. 1994. Characterization of a CD6 ligand(s) expressed on human- and murine-derived cell lines and murine lymphoid tissues. *Cell Immunol 158*:335–364.

Williams AF, Barclay AN. 1988. The immunoglobulin superfamily—Domains for cell surface recognition. *Annu Rev Immunol 6*:381–406.