ELSEVIER

Molecular modeling of CD28 and three-dimensional analysis of residue conservation in the CD28/CD152 family

Jürgen Bajorath,*† William J. Metzler,‡ and Peter S. Linsley*

**Bristol-Myers Squibb Pharmaceutical Research Institute, Seattle, Washington, †Department of Biological Structure, University of Washington, Seattle, Washington, ‡Bristol-Myers Squibb Pharmaceutical Research Institute, Princeton, New Jersey.*

CD28/CD152-CD80/CD86 receptor–ligand interactions result in costimulatory signals critical for optimal T cell activation. CD28/CD152 and CD80/CD86 are members of the immunoglobulin superfamily (IgSF). Despite common receptor–ligand interactions, both receptor and ligand pairs share only limited sequence identity. A detailed molecular model of the extracellular Ig-like domain of human CD28 was constructed using a combination of different modeling methods. The model was based on the solution structure of CD152 and sequence comparison of the CD28/CD152 family. Assessment of the model revealed good stereochemical quality and sequence–structure compatibility. The CD28 model was used to map surface residues, N-linked glycosylation sites, and to compare residue conservation in CD28 and CD152. The location of N-linked glycosylation sites in CD28/CD152 restricts the surface area available for binding. Rigorous sequence conservation in CD28 and CD152 is limited to core IgSF consensus positions and surface residues implicated in ligand binding. Other surface residues vary greatly in CD28/CD152. Residues critical for ligand binding are surrounded by surface patches conserved only in either CD28 or CD152. © *1997 by Elsevier Science Inc.*

Keywords: CD28/CD152 (CTLA-4), T cell costimulation, molecular modeling, model quality, computer graphics analysis, sequence conservation, surface residues, ligand binding

INTRODUCTION

The CD28/CD152 (CTLA-4) family of T cell surface proteins belongs to the immunoglobulin superfamily (IgSF)[1]. Effective T cell activation depends on costimulatory signals triggered by interaction between CD28 on T cells with CD80/CD86 ligands on antigen-presenting cells[2]. CD28-mediated T cell costimulation triggers T cell proliferation, while CD152 negatively regulates immune responses[3]. Both CD28 and CD152 contain single Ig V(ariable)-like ligand-binding domains, which are expressed on the cell surface as disulfide-linked homodimers[4].

The solution structure of monomeric CD152[5] has shown that the ligand-binding domain is most similar to the Ig V fold, despite some significant departures from standard Ig V geometry[6]. This similarity has been predicted on the basis of sequence analysis and model building[7,8]. The CD152 structure has provided a basis for the construction of a detailed molecular model of CD28. Stereochemical and inverse folding analysis suggested that the CD28 model was of good quality and sufficiently accurate to analyze some of its details. The model was used to map N- linked glycosylation sites and residues critical for ligand binding.

A noteworthy feature of the CD28/CD152 family is that, despite sharing common ligands, CD28 and CD152 display limited sequence identity. Only ~25% of the residues in the extracellular region are rigorously conserved in CD28/CD152 molecules from different species. Together with CD152, the CD28 model has been used to analyze surface residue patterns of the CD28/CD152 family in three dimensions.

Our analysis of residue conservation beyond IgSF characteristics showed that conservation of surface residues in CD28/CD152 is essentially limited to the CD80/CD86 binding site located on the A′–G–F–C–C′ β sheet surface. Other regions show great variability of surface residues. Potential N-linked glycosylation sites in CD28 and/or CD152 do not map to the conserved surface patch on the A′–G–F–C–C′ face, but limit the availability of other regions for protein–protein interactions, in particular, the A–B–E–D β sheet surface.

MATERIALS AND METHODS

Structural template and sequence comparison

The energy-minimized averaged nuclear magnetic resonance (NMR) coordinates of monomeric CD152[5] provided the tem-

Color Plates for this article are on pages 108–111.

Address reprint requests to: J. Bajorath, Bristol-Myers Squibb Pharmaceutical Research Institute, 3005 First Avenue, Seattle, Washington 98121.

Received 12 February 1997; accepted 11 April 1997.

655 Avenue of the Americas, New York, NY 10010

1093-3263/97/$17.00
PII S0263-7855(97)00020-9

plate for CD28 modeling. CD28/CD152 extracellular region sequences from different species were extracted from the SwissProt database and initially aligned using Pileup[9]. The β strands were assigned on the basis of the CD152 NMR model and incorporated into the sequence alignment. The alignment was manually modified to match appropriate residues at structurally important positions.

Side chain modeling

Model building, structural manipulations, and computer graphical analysis were performed using InsightII (95.0) (MSI, San Diego, CA). Side chains of residues conserved in human CD152 and CD28 were copied to the model. Conservative replacements of core residues were modeled in conformations similar to the original side chain. Nonconservative residue replacements were modeled using a low-energy rotamer search procedure[10].

Loop modeling

Loop conformations in CD28 were modeled in different ways. Loop backbone conformations predicted to be conserved in CD28 and CD152 were retained in the CD28 model. Two β turns were manually modeled and regularized. Other loop conformations were predicted using systematic conformational search with CONGEN[11]. In these calculations, possible loop conformations with acceptable potential energy were sampled and the conformation with lowest solvent-accessible surface within 3 kcal/mol of the energy minimum was selected.

Model refinement

The initial model was subjected to energy minimization with Discover (MSI) using AMBER[12] force field parameters, a distance-dependent dielectric constant (1r), and a 10-Å cutoff distance for nonbonded interactions. During these calculations, harmonic, all-atom constraints of initially 200 kcal/mol/Å^2 were gradually released and, finally, unconstrained minimization was continued until the maximum derivative of the energy function was ~20 kcal/mol/Å.

Model assessment

The stereochemical quality of the model was confirmed using PROCHECK[13]. Sequence–structure compatibility was assessed by PROSAII energy profile analysis,[14] using a 50-residue window for energy averaging at each position.

Structure comparison and representation

Structural models were superimposed and root-mean-square deviations (RMSDs) were calculated using ALIGN[15]. Color plates were produced with InsightII or RIBBONS[16] and processed as Silicon Graphics RGB images.

RESULTS AND DISCUSSION

CD152 structural template

Color Plate 1 shows a schematic representation of the solution structure of the monomeric extracellular Ig domain of CD152[5]. The CD152 structure most closely resembles an Ig V fold. A typical V fold consists of two tightly packed β sheets of four (A–B–E–D) and six (A′–G–F–C–C′–C″) β strands. CD152 shows some significant departures from standard Ig V geometry. A non-Ig disulfide bond (C48–C68) tethers the D strand to the C′ strand, reducing the interactions between D strand residues and the E–F loop. The E–F loop adopts a conformation different from that seen in many V-type domains. In addition, the region corresponding to the C′–C″ loop and C′ strand in Ig V domains (adjacent to the tethered C′ strand) is conformationally very flexible, and no C″ strand is formed.

Structure-oriented sequence comparison

Color Plate 2 shows an alignment of CD28/CD152 extracellular region sequences from different species with CD152 β strand assignments incorporated. Approximately 25% of the residues are rigorously conserved in CD28 and CD152. The conservation includes the non-Ig disulfide bond and the MYPPPY motif in the F–G region, which is critical for CD80/CD86 binding[17]. These residues are unique to the CD28/CD152 family. Sequence variation is greatest in the region between the C′ and D strands. With one exception (G strand, see below), all insertions and deletions in the CD28/CD152 sequence alignment map to loop regions in the CD152 structure.

Core regions of the CD28 model

Nine β strands of CD152 comprising the two β sheets of the Ig fold (here A–B–E–D and A′–G–F–C–C′) were used as core regions of the CD28 model. Buried residues in these regions are conserved or conservatively replaced throughout the CD28/CD152 family. Thus, the core structure should be similar in CD28/CD152 molecules, and structural manipulations in the core β strands were, with two exceptions, limited to side chain modeling as described in Materials and Methods. The exceptions occur in the A/A′ and G strands, where CD28 shows residue insertions compared to CD152.

In many Ig domains, including CD152, the A strand is split into the A and A′ strands and switches between the two β sheets. The A strand hydrogen bonds to the B strand and the A′ strand hydrogen bonds to the C-terminal section of the G strand[6]. The A/A′ strand transition is usually marked by a proline residue, which is conserved in CD28/CD152 (Color Plate 2). CD28 displays a single-residue insertion (S5) preceding the proline. Therefore, the A/A′ transition in CD28 was modeled involving two residues (S–P) as seen, for example, in some antibody variable light chains. To accommodate this transition, residue P6 in CD28 was modeled in cis conformation.

The other single-residue insertion in the core region of CD28 relative to CD152 (residue D103) occurs after the highly conserved MYPPPY motif in the F–G loop (Color Plate 2), i.e., at the beginning of the G strand. The CD28 sequence corresponding to the N-terminal part of the G strand is also much more hydrophilic than in CD152. Homolog scanning mutagenesis has suggested that the G strand is important for the regulation of CD28/CD152 binding avidity[17].

To retain the conformation of the conserved MYPPPY motif, the insertion in CD28 was modeled as a β bulge (D103–N105). Both CD28 and CD152 contain a highly conserved N-linked glycosylation site in the G strand (N108), which is

accommodated by a β bulge in the CD152 structure[5]. An additional β bulge in CD28 would further weaken G strand hydrogen bonding. Therefore, the prediction implies that this strand at the edge of the β sheet may be more flexible in CD28 than in CD152. However, it is possible that this region adopts a conformation other than the one predicted.

Regions with tentatively modeled structure

In addition to the N-terminal part of the G strand, the CD28 model contains segments for which no detailed predictions could be attempted. The segment connecting the CD28 Ig domain to the membrane-proximal cysteine (C120), which forms the disulfide bond responsible for covalent homodimerization of CD28 and CD152,[4] is unstructured in CD152 and was included in the model in extended conformation (V114–C120). The region connecting the C′ and D strands in CD152, which forms the C′–C″ loop and the C″ strand in other Ig V-type domains, is most variable in sequence in the CD28/CD152 family (Color Plate 2) and conformationally flexible in the CD152 structure (see above). Therefore, the average NMR backbone conformation of this region in CD152 was included in the CD28 model as an approximation.

Loop modeling

Two loop conformations in CD28 were modeled on the basis of the backbone conformations of corresponding loops in CD152, the F–G loop containing the MYPPPY motif and the E–F loop. Modeling of the E–F loop considered the presence of several key residues. In both CD28 and CD152, T86 is conserved. In CD152, this residue forms a bifurcated intraloop hydrogen bond to carbonyl oxygen atoms. In addition, both Q85/D in CD28/CD152 are capable of forming hydrogen bonding interactions with the conserved (and buried) Y89 and/or the positively charged residue (R or K) at position 36. Neither CD28 nor CD152 contains residues that form a salt bridge between the D strand and the E–F loop found in other V domains. The single-residue deletion in the C′–C′ loop of CD28 was modeled and regularized manually. The D–E loop was modeled as a type I and not type II′ β turn (as in CD152) because the turn in CD28 does not include a glycine.

Possible conformations of three other loops not conserved in CD28/CD152 were generated by conformational search. The A′–B and C′–D loops show a two-residue deletion and insertion, respectively, relative to CD152. The seven-residue B–C loop in CD28 shares the same length with CD152 but no sequence similarity. The B–C loop contributes to CD28/CD152 ligand binding and is spatially adjacent to the F–G loop, which is critical for binding[17]. Thirty possible B–C loop conformations were generated with potential energy ranging from −79.5 to 3.8 kcal/mol, and the conformation with lowest solvent-accessible surface within 3 kcal of the energy minimum was selected (−79.2 kcal/mol). Compared to the B–C loop in CD152, this conformation displayed a backbone RMSD of 2.1 Å. Color Plate 3 shows a comparison of the predicted and experimentally determined B–C and F–G loop conformations in CD28 and CD152. The B–C and F–G loops correspond to the complementarity determining region (CDR) loops 1 and 3 in antibodies, which are major determinants of antigen specificity[1].

Model refinement and assessment

After energy minimization the backbone RMSD of the refined from the initially assembled model was ~1 Å. When compared to CD152, 105 of 120 α carbon atoms superimpose with an RMSD of ~0.7 Å. No unfavorable intramolecular contacts were detected in the CD28 model, and no nonglycine residues were found in disallowed torsional space. Figure 1 shows an energy profile of the model which monitors average pairwise residue interaction energies[14]. The overall negative profile trace of the model confirmed its sequence–structure compatibility[14] and indicated that significant errors in the core regions of the model were unlikely.

CD28 molecular model

Color Plate 4 shows the final CD28 model. On the basis of our model building analysis, regions of high prediction confidence include the β strands A′, B, C, C′, D, and E; the C-terminal half of G; and the D–E, E–F, and F–G loops. Lower confidence regions include two edge strands (A and, in part, G), the tentatively modeled regions (which are probably flexible), and *de novo*-modeled loop conformations, for example, the B–C loop. High-confidence regions represent the majority of the domain (Color Plate 4). Thus, despite its limitations, the model should be sufficiently accurate to analyze some details beyond the backbone level.

Mapping of surface residues

Residues conserved in CD152 and CD28 were mapped on the CD152 structure and the CD28 model, respectively (Color Plate 5). Rigorous conservation of surface residues in CD28/CD152 is limited to the F–G loop and a number of residues that are spatially adjacent on the A′–G–F–C–C′ face. A few residues conserved only in either CD152 or CD28 extend this region toward the C terminus of the domain. In contrast, the opposite (A–B–E–D) β sheet surface displays many nonconserved residues and only a small conserved surface patch at the A/A′ strand transition.

In the CD28 model, all five potential N-linked glycosylation sites map to exposed positions. N108 in the G strand is the only N-linked glycosylation site which is conserved in CD152. Several glycosylation sites in CD152 or CD28 map to the A–B–E–D face of the domain, but none are found in the conserved surface patch on the A′–G–F–C–C′ face. When mutated, several of these conserved residues, e.g., K95, E97, E33 in CD152[5,18] and P100, Y101 in CD28,[17] reduce or abolish CD80/CD86 binding. These data and the observed residue conservation suggest that the surface patch extending the MYPPPY region is involved in CD80/CD86 binding to CD28/CD152,[5] which is consistent with the predicted loca-tion of surface residues and N-linked glycosylation sites. Other residues adjacent to this site are conserved only in CD28 or CD152 (e.g., F90 in CD28 or Q43 in CD152). These residues may modulate ligand binding.

On the basis of our analysis, conservation of surface residues in CD28/CD152 is essentially limited to the region implicated in ligand binding. The great variability of surface residues in regions not implicated in ligand binding suggests that the CD28/CD152 family is ancient and that CD28 and CD152 have diverged early from a common ancestor.

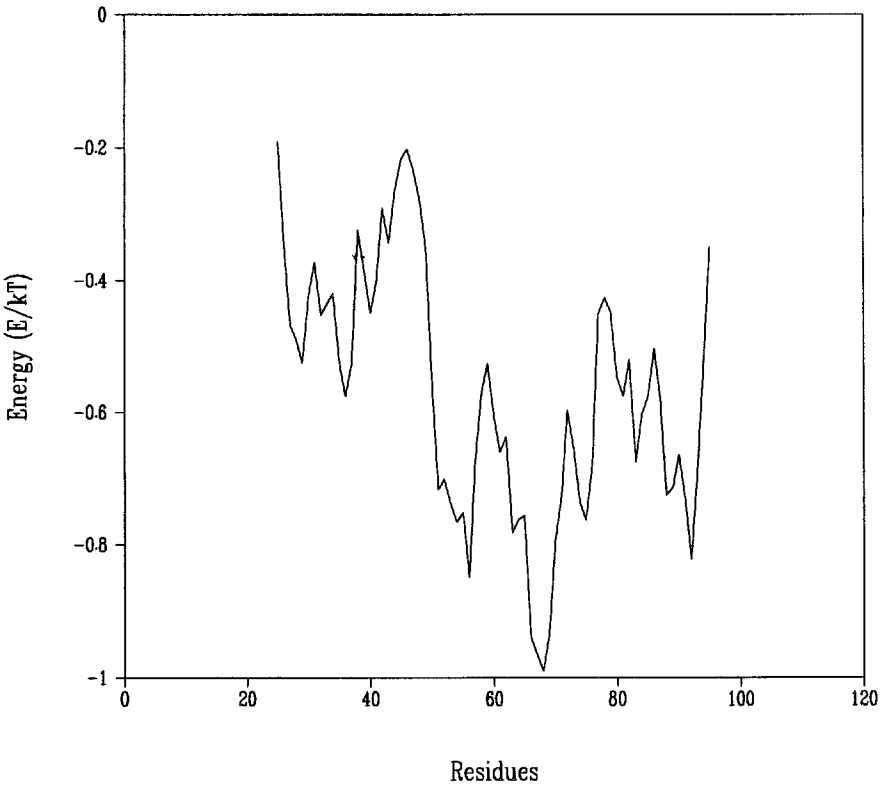

Figure 1. Energy profile of the CD28 model. Pairwise residue interaction energies were calculated using PROSAII, and average energies were plotted for each residue. Energy is given as E/kT *(*E*, residue interaction energy in kcal/mol;* k*, Boltzmann constant;* T*, temperature in K). A 50-residue window was used for energy averaging at each position.*

Two notable exceptions of surface residue variability outside the extended MYPPPY region were observed. These are the region connecting the extracellular domain to the membrane-proximal disulfide bond, which is conserved only in CD152 but not in CD28, and a region proximal to the N terminus in CD28, which is not conserved in CD152 (Color Plate 5). It is possible that these regions are important for the function of CD28/CD152. For example, the large surface patch on the A–B–E–D face, which is conserved in CD152, opposite the CD80/CD86 binding site, and which does not include glycosylation sites, may mediate interactions between disulfide-linked monomeric subunits. Such interactions may contribute to the regulation of binding avidity.

CONCLUSIONS

A comparative model of the extracellular domain of CD28 was generated on the basis of the solution structure of CD152, using a combination of different modeling methods. Assessment of the CD28 model suggested that its accuracy was sufficient to analyze some details. We have mapped surface residues in CD28 and compared (non)conserved surface patches in CD28 and CD152. The distribution of N-linked glycosylation sites restricts the surface area available for protein–protein interactions. Rigorous conservation of surface residues in this family is essentially limited to the region implicated in ligand recognition and suggests that the binding site extends to the A′–G–F–C–C′ face. Surface residues in other regions vary significantly, but some residue conservation unique to CD28 or CD152 is seen distant from the ligand binding site. These previously unidentified regions may be important for other molecular interactions.

REFERENCES

1 Williams, A.F. and Barclay, A.N. *Annu. Rev. Immunol.* 1988, **6,** 381
2 Linsley, P.S. and Ledbetter, J.A. *Annu. Rev. Immunol.* **1993, 11,** 191
3 Krummel, M.F. and Allison, J.P. *J. Exp. Med.* 1995, **182,** 459
4 Linsley, P.S., Ledbetter, J., Peach, R., and Bajorath, J. *Res. Immunol.* 1995, **146,** 130
5 Metzler, W.J., Bajorath, J., Fenderson, W., Shaw, S.-Y., Constantine, K.L., Naemura, J., Leytze, G., Peach, R.J., Lavoie, T.B., Mueller, L., and Linsley, P.S. Nature Structural Biology 1997, in press
6 Bork, P., Holm, L., and Sander, C. *J. Mol. Biol.* 1994, **242,** 309
7 Brunet, J.F., Denizot, F., Luciani, M.F., Roux-Dosseto, M., Suzan, M., Mattei, M.G., and Goldstein, P. *Nature (London)* 1987, **328,** 267
8 Bajorath, J. and Linsley, P.S. *J. Mol. Modeling* 1997, 3, 117
9 Feng, D.-F. and Doolittle, R.F. *J. Mol. Evol.* 1987, **25,** 351
10 Bajorath, J. and Fine, R.M. *Immunomethods* 1992, **1,** 137

11 Bruccoleri, R.E., Haber, E., and Novotny, J. *Nature (London)* 1989, **355,** 483
12 Seibel, G., Singh, U.C., Weiner, P.K., Caldwell, J., and Kollman, P. *AMBER 3.0,* revision A. University of California at San Francisco, San Francisco, 1990
13 Laskowski, R.A., MacArthur, M.W., Moss, D.S., and Thornton J.M. *J. Appl. Crystallogr.* 1993, **26,** 283
14 Sippl, M.J. *Proteins Struct. Funct. Genet.* 1993, **17,** 35
15 Satow, Y., Cohen, G.H., Padlan, E.A., and Davies, D.R. *J. Mol. Biol.* 1986, **190,** 593
16 Carson, M. *J. Appl. Crystallogr.* 1991, **24,** 958
17 Peach, R.J., Bajorath, J., Brady, W., Leytze, G., Greene, J., Naemura, J., and Linsley, P.S. *J. Exp. Med.* 1994, **180,** 2049
18 Morton, P.A., Fu, X.-T., Stewart, J.A, Giacoletto, K.S., White, S.L., Leysath, C.E., Evans, R.J., Shieh, J.-J., and Karr, R.W. *J. Immunol.* 1996, **156,** 1047