ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/227806047

A proposed structural model of domain 1 of fasciclin III neural cell adhesion protein based on an inverse folding algorithm

ARTICLE *in* PROTEIN SCIENCE · MARCH 2008

Impact Factor: 2.85 · DOI: 10.1002/pro.5560040314 · Source: PubMed

CITATIONS
2

READS
14

4 AUTHORS, INCLUDING:

Jacquelyn Fetrow
University of Richmond
87 PUBLICATIONS **2,982** CITATIONS

SEE PROFILE

All in-text references underlined in blue are linked to publications on ResearchGate, letting you access and read them immediately.

Available from: Jacquelyn Fetrow
Retrieved on: 24 January 2016

Protein Science (1995), *4*:472-483. Cambridge University Press. Printed in the USA.
Copyright © 1995 The Protein Society

A proposed structural model of domain 1 of fasciclin III neural cell adhesion protein based on an inverse folding algorithm

LAURIE A. CASTONGUAY,[1,3] STEPHEN H. BRYANT,[2] PETER M. SNOW,[1] AND JACQUELYN S. FETROW[1]

[1] Department of Biological Sciences and Center for Molecular Genetics, University at Albany, State University of New York, 1400 Washington Avenue, Albany, New York 12222
[2] National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Maryland 20894

(RECEIVED October 26, 1994; ACCEPTED December 22, 1994)

Abstract

Fasciclin III is an integral membrane protein expressed on a subset of axons in the developing *Drosophila* nervous system. It consists of an intracellular domain, a transmembrane region, and an extracellular region composed of three domains, each predicted to form an immunoglobulin-like fold. The most N-terminal of these domains is expected to be important in mediating cell–cell recognition events during nervous system development. To learn more about the structure/function relationships in this cellular recognition molecule, a model structure of this domain was built. A sequence-to-structure alignment algorithm was used to align the protein sequence of the fasciclin III first domain to the immunoglobulin McPC603 structure. Based on this alignment, a model of the domain was built using standard homology modeling techniques. Side-chain conformations were automatically modeled using a rotamer search algorithm and the model was minimized to relax atomic overlaps. The resulting model is compact and has chemical characteristics consistent with related globular protein structures. This model is a de novo test of the sequence-to-structure alignment algorithm and is currently being used as the basis for mutagenesis experiments to discern the parts of the fasciclin III protein that are necessary for homophilic molecular recognition in the developing *Drosophila* nervous system.

Keywords: fasciclin III; homology modeling; immunoglobulin superfamily; inverse folding algorithm; neural adhesion protein; protein modeling; protein structure prediction; sequence-to-structure alignment algorithm; threading algorithm

Cell–cell interactions mediated by cell-surface molecules are thought to be central to the process of axonal guidance in the developing nervous system (reviewed by Goodman et al., 1984). Fasciclin III is one of several molecules identified in *Drosophila melanogaster* that possess characteristics consistent with a role in this process. This protein is expressed on a subset of neurons during *Drosophila* development and it may play a role in axonal guidance, mediating recognition events between neuronal cells (Patel et al., 1987). Fasciclin III has been found to mediate adhesion in a homophilic fashion when transfected into cells under the control of an inducible promoter (Snow et al., 1989). In addition, when fasciclin III-expressing cells are mixed with cells expressing fasciclin I (an unrelated homophilic adhesion molecule), the cells sort into fasciclin I-expressing and fasciclin III-expressing aggregates, with little mixing of the two cell types (Elkins et al., 1990). This demonstration of fasciclin-mediated cell sorting is consistent with recognition roles for both molecules in the developing nervous system.

Fasciclin III is an integral membrane protein, consisting of a 138-residue intracellular domain, a 24-residue transmembrane region, and a 326-residue extracellular domain (Snow et al., 1989). Homology studies suggest that the extracellular domain may be further subdivided into three distinct subdomains, each of which are predicted to fold into a structure characteristic of members of the immunoglobulin superfamily (Grenningloh et al., 1990). Sequentially, extracellular domain 3 is in closest proximity

Reprint requests to: Jacquelyn S. Fetrow, Department of Biological Sciences, University at Albany, SUNY, 1400 Washington Avenue, Albany, New York 12222; e-mail: jacque@isadora.albany.edu.
[3] Present address: Sterling Winthrop, Inc., 1250 S. Collegeville Road, PO Box 5000, Collegeville, Pennsylvania 19426-0900.

to the cell membrane and domain 1, the most N-terminal extracellular subdomain, is the most distant from the transmembrane region.

The immunoglobulin fold, originally defined as a structural feature of antibodies of the immune system, consists of a classic antiparallel β-sheet sandwich. Some of the loops between strands are highly variable and are called hypervariable regions, whereas other loops and turns are more conserved. The basic Ig motif can be further subdivided into constant (C) and variable (V) types, but all exhibit the same antiparallel β-sheet structure (Alzari et al., 1988). A number of cellular adhesion proteins have been included within the superfamily as they have been demonstrated to contain this structural motif (Williams & Barclay, 1988). For example, the crystal structures of CD2 and CD4, two T cell recognition molecules, have been solved and fold into an Ig-like structure (Ryu et al., 1990; Wang et al., 1990; Jones et al., 1992), as had been predicted based upon sequence homology (Killeen et al., 1988; Williams & Barclay, 1988).

We are interested in defining the regions of fasciclin III that are involved in mediating homophilic interactions. In several cases, adhesion molecules, such as CD2 and CD4 in the immune system, utilize the most membrane-distal domain to mediate adhesive interactions with their specific counter-receptors (Peterson & Seed, 1987; Clayton et al., 1989). Thus, we hypothesize that the analogous domain in fasciclin III might also be involved in intermolecular interactions. As a first step in testing this prediction, we have built a model of the first, most membrane-distal domain (domain 1) of fasciclin III (fasIIId1).

Based on a prediction that it folds into an Ig-like motif, a model of fasIIId1 was built using standard homology model building techniques (Greer, 1991); however, as the homology model was being built, techniques for aligning sequence and structure, the so-called "threading" or "inverse folding" algorithms, were introduced (for review, see Fetrow & Bryant, 1993). These methods employ linear profiles or three-dimensional contact potentials to correctly align an amino acid sequence with a given structural motif and have been successful in matching amino acid sequence to three-dimensional structure in cases where little sequence homology exists. These algorithms have been tested on proteins whose structures had previously been determined; few unknown structures have yet been predicted using these algorithms.

In this paper, we present a three-dimensional model of fasIIId1. This model was made by building the fasIIId1 sequence onto the backbone motif of immunoglobulin McPC603 (Satow et al., 1987) using the best alignment of sequence to structure determined by the threading algorithm of Bryant and Lawrence (1993). A side-chain packing algorithm (P. Shenkin, H. Farid, & J.S. Fetrow, manuscript in prep.) was used to determine optimal packing of side-chain atoms and energy minimization was used to eliminate remaining atomic overlaps. The model thus obtained is as compact as other protein molecules that fold into immunoglobulin motifs, such as the immunoglobulin MCP and the T-cell surface glycoprotein CD4. Other characteristics of the model, such as buried hydrophobic surface area and internal hydrogen bonding, are also similar to that of the native proteins. By comparison with CD2 and CD4, the model suggests regions important for homophilic recognition by fasciclin III and is being used as the basis for further mutagenesis studies to determine the function of this protein (P.M. Snow, unpubl. results).

Results

Comparison of alignments determined by homology and by the threading algorithm

The sequences of fasIIId1 and CD4 were aligned manually using the very limited sequence homology (Fig. 1). Alignment of fasIIId1 to the structure of MCP determined automatically by the threading algorithm (Bryant & Lawrence, 1993) is shown in Figure 2. The overall sum of the pairwise interaction energies calculated by the threading algorithm is comparable to that ob-

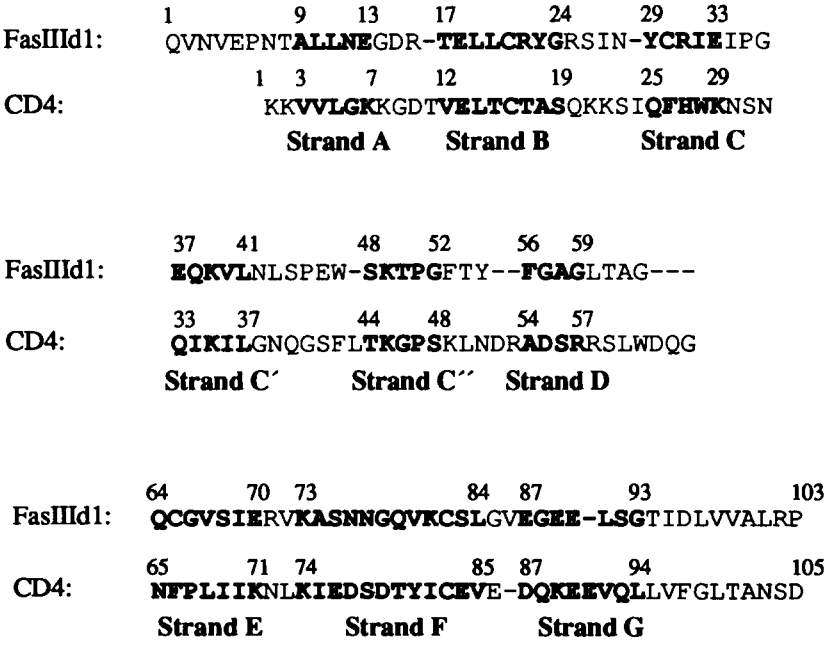

Fig. 1. Alignment of the amino acid sequence of the T cell surface glycoprotein, CD4 (Ryu et al., 1990; Wang et al., 1990), to the sequence of fasciclin III domain 1 (fasIIId1), as suggested by Grenningloh et al. (1990). In this alignment, SCRs, shown in bold face, were taken as the conserved β-strands in standard immunoglobulin structures and, by convention, are designated as strands A–G. First and last residue numbers of each SCR are indicated.

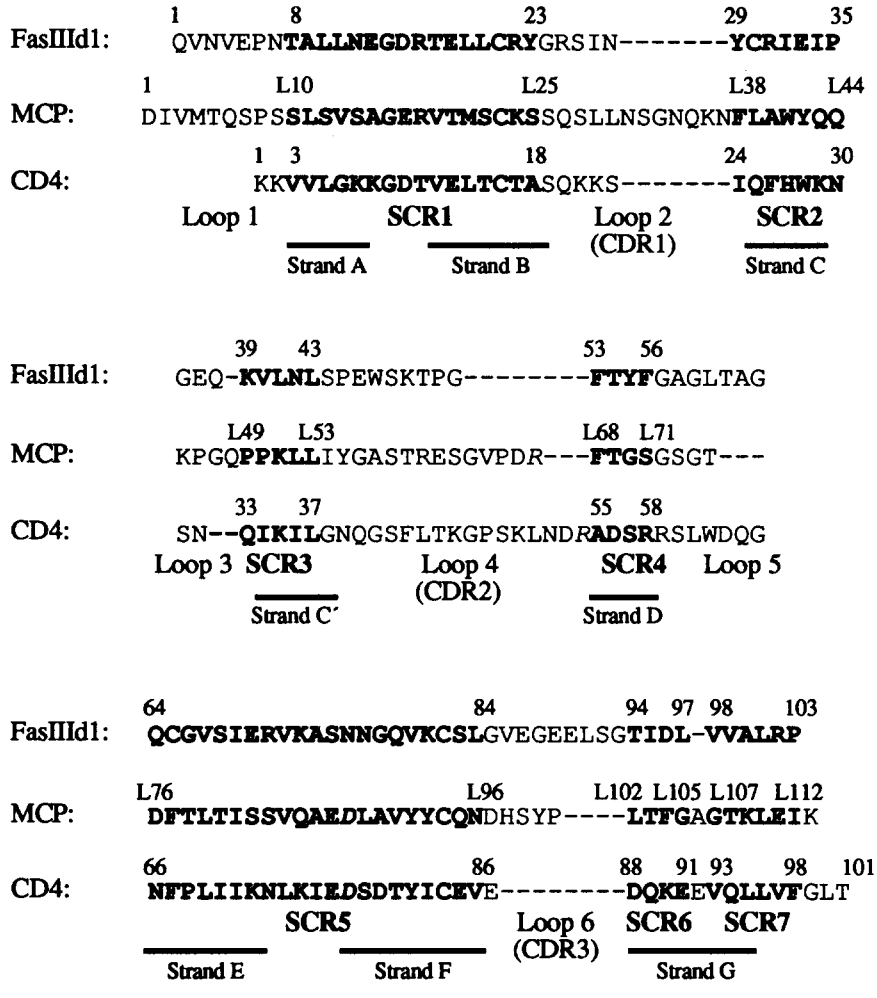

Fig. 2. Amino acid sequence alignment of fasciclin III domain 1 (fasIIId1), and the T cell surface glycoprotein, CD4 (Ryu et al., 1990; Wang et al., 1990), to immunoglobulin McPC603 light chain (MCP, Satow et al., 1987), as determined by the sequence-to-structure alignment ("threading") program of Bryant and Lawrence (1993). SCRs, shown in bold face, are slightly modified from the definition described in Bryant and Lawrence (1993), to account for the structural differences between standard immunoglobulins and the CD4 and CD2 structures. First and last residue numbers of each SCR are indicated. Conventional β-strands of the immunoglobulin superfamily (A–F) were previously determined for CD4 (Ryu et al., 1990; Wang et al., 1990) and MCP (Satow et al., 1987) and are underlined. Italicized residues in MCP and CD4 form the conserved salt bridges in these molecules.

tained with other Ig sequences (Bryant & Lawrence, 1993) and is quite unlikely for non-immunoglobulin sequences (Fetrow & Bryant, 1993).

The threading and sequence homology alignment methods yield similar alignments in some parts of the protein. SCR1 (structurally conserved region 1), as defined for input into the threading algorithm, encompasses both strands A and B and the intervening turn (Fig. 2). The sequence alignment puts a gap in the intervening turn, so that strand B is similarly aligned by both techniques, whereas strand A is shifted by one amino acid. Likewise, SCR5 (Fig. 2, threading alignment) encompasses strands E and F (Fig. 1, sequence alignment) and the intervening turn; the alignment determined by sequence homology and by the threading algorithm is the same in this region. Strand C corresponds quite closely with SCR2 and the alignment is the same in both cases.

The alignments differ slightly in two regions. Strand C′ (Fig. 1) is analogous to SCR3 (Fig. 2), but the sequence and threading alignments differ from each other by two residues. Thus, loops 3 and 4 in fasIIId1 are longer and shorter, respectively, than the analogous loops in CD4. Likewise, strand D (Fig. 1) is similar to SCR4 (Fig. 2), but the alignments determined by threading and homology differ by three residues in this region.

Two major differences exist between the alignments determined by sequence homology and by threading. Strand C″ (Fig. 1, sequence alignment) is not assigned as an SCR in the threading alignment and, thus, falls within loop 4 (Fig. 2, threading alignment). This β-strand is part of the immunoglobulin variable domain and is hydrogen bonded to the rest of the sheet in most, but not all, antibody structures; thus, it was not assigned as an SCR in the threading algorithm and was built as a loop, rather than as a β-strand. Strand G (Fig. 1) should correspond to SCRs 6 and 7 (Fig. 2), but the alignments for these two regions are quite different. The sequence defined as strand G by the homology alignment is defined as loop 6 by the threading algorithm and SCRs 6 and 7 (a β-strand with a bulge) fall later in the sequence.

FasIIId1 contains four cysteine residues that could form disulfide bonds in the tertiary structure of the protein. Both alignments by homology and by threading put these cysteines in similar positions, such that cysteines 21 and 30 can disulfide bond with cysteines 82 and 65, respectively. The canonical Ig variable domain structure contains one disulfide bond between strands B and F, which would correspond to a disulfide between residues 21 and 82 in fasIIId1 (Alzari et al., 1988). Thus, although the positions of the cysteines could not be used to discriminate between the sequence similarity and threading alignments, it does lend credence to their authenticity.

Models built from the two alignment procedures will yield similar predictions for fasIIId1 structure/function relationships

in many regions of the molecule, but different predictions in other regions, as described above. Because of the limited sequence homology between these two proteins (only 20% of all SCR residues are identical) and because the threading algorithm accounts for tertiary interactions, the alignment determined by the threading algorithm (Fig. 2) was chosen as the basis for the model structure.

Comparison of the CD4 threading alignment to CD4 crystal structure

As a control, the CD4 sequence was aligned to the MCP structure by the threading algorithm and this alignment (Fig. 2) was compared to the actual crystal structure of CD4 (Ryu et al., 1990; Wang et al., 1990). The alignment is perfect except in SCRs 2, 3, and 7, where the alignment is off in each case by one amino acid (Fig. 3). SCRs 2 and 3 could be misaligned because loop 3 in CD4 is a short β-turn. A comparison of the actual CD4 and MCP crystal structures shows the error. Because of the initial length assignment of the SCRs, the two central residues in the β-turn of CD4 crystal structure are also members of SCRs 2 and 3, respectively, causing the CD4 loop 3 length to be zero. Thus, the initial assignment of SCRs 2 and 3 in MCP are slightly longer than they should be for correct modeling of CD4. This result emphasizes the importance of proper SCR assignment.

The threading algorithm suggests that CD4 contains a β-bulge in strand G, as is found in MCP (SCRs 6 and 7, Fig. 2), but none is actually found (Fig. 3). This result demonstrates a limitation of this implementation of threading algorithm.

Given the small errors in the alignment of CD4 sequence to MCP structure, it is necessary to consider the validity of the model of fasIIId1. Because the threading algorithm takes tertiary contacts into account and because the sequence similarity between fasIIId1 and MCP is so very limited, the alignment determined by the threading algorithm is almost certain to be more valid than a model built from sequence homology. In addition, the CD4 alignment errors are small and localized to two regions. The alignment of fasIIId1 to the MCP crystal structure as determined by the threading algorithm has a very high probability of being correct (Fetrow & Bryant, 1993), but it is less well determined in the C-terminal region; thus, we are least confident of the model structure in this region.

Backbone, loop, and side-chain modeling of fasIIId1

Using the Homology module of InsightII (Biosym Technologies, Inc., San Diego, California), the SCRs of the fasIIId1 sequence were built onto the backbone coordinates of the MCP light chain using the alignment shown in Figure 2. If homologous loops of the same length were found in immunoglobulins, the backbones of these loops were used in the model. For those loops without a similar homologous loop, possible loop structures from the Brookhaven Protein Data Bank (Abola et al., 1987) were found using the structural similarity method of Jones and Thirup (1986), a list of 10 best-fit loops was screened visually, and the resulting "best" loops are shown in Table 1. The conformation of two of the six loops, numbers 1 and 5, came directly from CD4 or MCP. The backbone conformation of loops 2 and 6 came from the homologous regions of other immunoglobulin chains, whereas the conformations of loops 3 and 4 came from heterologous proteins; therefore, we are least confident of the backbone conformations of loops 3 and 4. Loops 2, 4, and 6 correspond to the Ig hypervariable regions (CDR1, 2, and 3, Fig. 2) that are involved in antigen-antibody recognition in the Igs; thus, these loops could be involved in molecular recognition, if the fasIII self-recognition mechanism was similar to the antibody-antigen recognition mechanism.

To determine side-chain conformations, a simulated annealing search strategy (Metropolis et al., 1953) was used to search a rotamer library (Ponder & Richards, 1987) of side-chain conformations (P. Shenkin, H. Farid, & J.S. Fetrow, manuscript in prep.). To be sure that the conformational space was searched adequately, four independent conformational searches were run. Three of these four runs yielded identical side-chain rotamers in the protein core, though conformations of side chains at the surface differed among the four runs. The fourth conformational search yielded a different core packing for cysteines 21 and 30 and leucines 41, 60, and 84 in the hydrophobic core. Careful study demonstrated that this alternative packing put the cysteine side chains in relative conformations similar to that of ideal disulfide-bonded cysteines. The virtual dihedral angle formed by atoms β-Sγ-Sγ-Cβ in this alternative packing were $-75°$ and $-101°$ for cysteines 21–82 and cysteines 30–65, respectively, compared to 180° and 54° for the other calculated configurations. In disulfide bonds, the optimal dihedral angle is $+90°$ or $-90°$ and these angles are favored by as much as 10 kcal/mol (Creighton, 1993). Because all known Ig structures contain at least one disulfide in a position analogous to cysteines 21 and 82 (Alzari et al., 1988), we were fairly certain that at least this pair of cysteines should form a disulfide; therefore, the set of side-chain conformations that put the four cysteines in the more optimal disulfide bond orientation was chosen for further minimization and analysis.

	SCR2	SCR3	SCR7
MCP SCR:	^{L38}FLAWYQQ^{L44}	^{L49}PPKLL^{L53}	^{L107}GTKLEI^{L112}
CD4 (Thread):	^{24}IQFHWKN^{30}	^{33}QIKIL^{37}	^{93}VQLLVF^{98}
CD4 (Structure):	^{25}QFHWKNS^{31}	^{32}NQIKI^{36}	^{92}EVQLLV^{97}

Fig. 3. Comparison of CD4 alignment to the MCP immunoglobulin structure obtained by the threading algorithm to the actual crystal structure of CD4. Only the misaligned SCRs are shown; the alignment is correct in other regions. MCP is the amino acid sequence found in each SCR in the McPC603 crystal structure. Numbers are the first and last residue numbers of each SCR. CD4 (Thread) is the alignment determined by the threading algorithm. CD4 (Structure) is the structural alignment determined by visual observation of the crystal structures of MCP and CD4.

Table 1. *Sequences of the loop regions in fasIIId1 as determined by threading algorithm and the sequences of the loop structures on which they were modeled*

Loop[a]	FasIIId1 residues[b]	FasIIId1 sequence[c]	Size range[d]	Loop model residues[e]	Loop model sequence[f]
Loop 1	1-7	QVNVEPN	2-9	2mcp, L3-L9	VMTQSPS
Loop 2	24-28	GRSIN	5-12	2hfl, L26-L30	SSSVN
Loop 3	36-38	GEQ	0-6	3est, 37-39	SGS
Loop 4	44-52	SPEWSKTPG	7-21	2sns, 101-109	EALVRQGLA
Loop 5	57-63	GAGLTAG	4-7	2cd4, 59-65	RSLWDQG
Loop 6	85-93	GVEGEELSG	1-15	2mcp, H101-H109	NYYGSTWYF

[a] Arbitrary number assigned to each loop (see Fig. 2).
[b] Amino acid residues of each loop in fasciclin III domain 1 as determined by the threading program.
[c] Amino acid sequence of each loop in fasciclin III domain 1. Standard one-letter codes are used to represent the amino acids.
[d] Minimum and maximum loop sizes used as parameters in the threading program. Loop sizes were allowed to vary, as described by Bryant and Lawrence (1993).
[e] Protein and amino acid residues from which the backbone coordinates of each loop were assigned. Possible loop candidates were selected from the Brookhaven Protein Data Bank (Abola et al., 1987) using the method of Jones and Thirup (1986), as implemented in the Homology program. The Brookhaven code is used to describe each protein.
[f] Amino acid sequence of the loop from which coordinates were assigned. Standard one-letter codes are used for the amino acids.

Minimizations were then performed using the program Discover (Biosym Technologies, Inc.) and the CVFF force field (Hagler et al., 1979). Minimizations were not designed to find a global minimum, but merely to eliminate steric overlaps and relax any strain present in the model, so were done in vacuo, without charges, and were not taken to completion. A strategy of fixing the backbone and allowing the side chains to relax first, then allowing the backbone and side chains to minimize, then constraining the backbone and allowing side chains to relax further was developed and tested on the CD4 and MCP crystal structures, then applied to the model structure.

Using this strategy, four minimizations were performed on the fasIIId1 model, one with both disulfide bonds reduced, one with both disulfide bonds oxidized, and one each with one disulfide bond oxidized and one reduced. The final energy between the four models was quite similar, though the actual numbers are meaningless. No unfavorable interaction energies were found between residues in any of the models (data not shown). The fasIIId1 model presented here was minimized with both disulfide bonds oxidized (Fig. 4; Kinemages 1, 2).

As a control, CD4 and MCP crystal structures were subjected to the same minimization strategy. In addition, the CD4 and MCP side chains were built onto the crystal structure backbone using the same algorithm applied to the fasIIId1 model and these "side-chain modeled" molecules were also minimized. For both CD4 and MCP, the relative energies between the minimized crystal structure and the minimized side-chain modeled structure were similar (data not shown) and the RMS differences to the

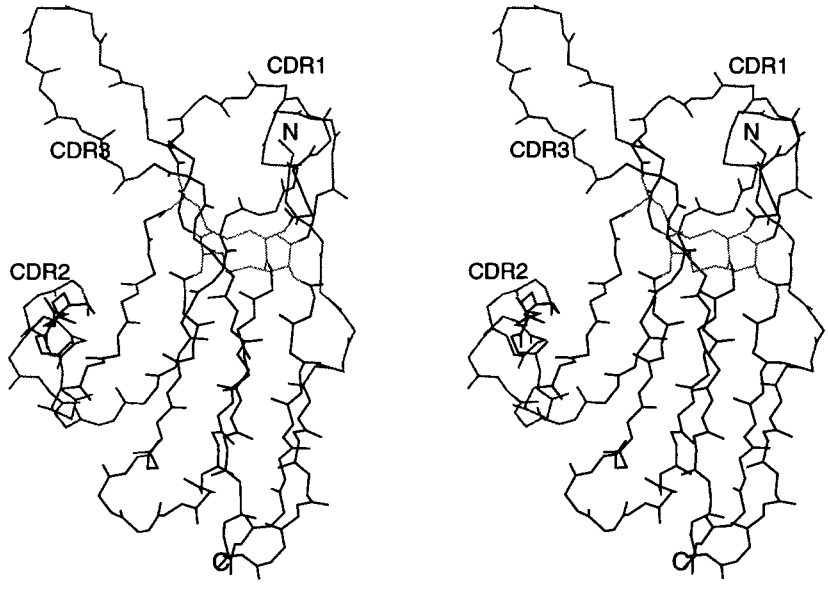

Fig. 4. Backbone of the fasIIId1 model presented in stereo. SCRs are shown in dark lines, loops are indicated by lighter lines. N- and C-termini of the molecule are indicated by N and C, respectively; loops corresponding to the CDRs (CDR1, loop 2; CDR2, loop 4; CDR3, loop 6) are marked (see Kinemage 1). Two disulfide bonds that provide support for the authenticity of the model are shown in gray in this figure and in yellow in Kinemage 2.

crystal structure are similar (Table 2). In addition, unfavorable residue–residue interaction energies were not found, similar to the results obtained for the fasIIId1 model.

Characterization of the fasIIId1 model

In addition to the atomic interaction energies discussed above, the surface area, compactness, and hydrogen bonding patterns were calculated for the fasIIId1 model structure, the native MCP and CD4 crystal structures, and the "energy minimized" and "side-chain modeled" MCP and CD4 structures. A comparison of the side-chain modeled and the energy minimized immunoglobulins shows that neither of these techniques greatly perturbs the structures of these molecules (Table 2). Careful visual examination of these structures confirms this quantitative result, showing that the side-chain modeling and energy minimization strategies used here are reasonable for these molecules.

Comparison of the fasIIId1 model and the crystal structures shows that the model looks like a "real" globular protein; 51% of all exposed surface area in the fasIIId1 model structure is hydrophobic (Table 2). These values compare favorably to the MCP and CD4 crystal structures, where 55 and 50%, respectively, of all exposed surface area is hydrophobic (Table 2); furthermore, these values compare favorably to protein structures overall, where approximately 45–60% of all exposed surface area is generally hydrophobic (Richards, 1977). The fasIIId1 model buries about 63% of its available surface area on folding, comparable to other proteins (Table 2). In addition, it was determined that no buried, charged side chains were found in the model (data not shown).

The compactness of the model structure is slightly higher than that of the MCP and CD4 crystal and model structures (Table 2); however, the value is well within the range computed for a large database of native proteins (Zehfus & Rose, 1986). Taken together, these results suggest that the fasIIId1 model shown in Figure 3 is a reasonable structure typical of a single domain of a globular protein.

Discussion

Fasciclin III was originally designated as a member of the immunoglobulin superfamily through analysis using the ALIGN program of Dayhoff and colleagues (1983), as described by Williams and Barclay (1988) (Grenningloh et al., 1990). This analysis depends on limited sequence similarity between fasciclin III and other members of the superfamily, and has indicated that this protein is likely composed of three immunoglobulin folds in the extracellular domain. The most amino-terminal of these folds (domain 1) is of the V type, whereas the two most carboxy-terminal folds (domains 2 and 3) were predicted to belong to the C2 category.

Although such an analysis is suggestive of inclusion within the superfamily, the limited sequence similarity between fasciclin III and other members of the family makes this designation tenuous. To substantiate this assignment further, as well as to provide a rational basis for experiments designed to probe the

Table 2. *Hydrophobic and hydrophilic surface areas and compactness of the fasciclin III domain 1 model (fasIIId1), the native T cell receptor protein (CD4), and the native immunoglobulin McPC603 (MCP)*

Protein[a]	Nonpolar SA exposed[b] (%)	Total SA buried[c] (%)	Nonpolar SA buried[d] (%)	Z[e]	RMS (Å)[f]
FasIIId1 (reduced)	51.7	63.1	68.5	1.72	
FasIIId1 (oxidized)	51.2	63.0	68.7	1.72	0.75
CD4 (native)	50.4	67.5	72.3	1.58	
CD4 (minimized)	54.2	67.8	70.6	1.55	1.08
CD4 (SC modeled)	51.3	66.5	71.1	1.58	1.62
MCP (native)	54.8	68.7	71.8	1.59	
MCP (minimized)	58.4	68.8	70.0	1.57	0.97
MCP (SC modeled)	56.9	68.0	70.2	1.59	1.27

[a] FasIIId1 is the model structure of fasciclin III domain 1 built using the threading algorithm (Bryant & Lawrence, 1993), with loops and side chains modeled, and energy minimized as described in the Methods. CD4 is the crystal structure of the T cell surface receptor, 2CD4 (Wang et al., 1990), residues 1–98. MCP is the crystal structure of immunoglobulin McPC603 (Satow et al., 1987), residues L1–L113. As a control, values were calculated on CD4 and MCP structures that were modified in three ways: "native" is the native crystal structure of the protein, "minimized" is the crystal structure subjected to the same minimization protocol as the fasIIId1 model, and "SC modeled" is that model in which the side chains of each crystal structure were subjected to the same side-chain modeling algorithm and minimization protocols as fasIIId1.

[b] "Nonpolar SA exposed" is the percentage of the total solvent-accessible surface area that is nonpolar (carbon or sulfur atoms only). The solvent-accessible surface area was calculated by the method of Lee and Richards (1971).

[c] "Total SA buried" is calculated as $100 - (SA_{folded}/SA_{standard} * 100)$, where SA_{folded} is the solvent-accessible surface area in the folded or modeled protein and $SA_{standard}$ is the solvent-accessible area in the standard state. Standard-state accessible surface areas are the stochastic standard states described by Rose et al. (1985).

[d] "Nonpolar SA buried" is calculated in the same manner as "Total SA buried" except the surface area of only the hydrophobic atoms, carbon, and sulfur were included in the calculation.

[e] The compactness of the protein, Z, calculated by the method of Zehfus and Rose (1986) as described in the Methods section.

[f] The calculated RMS difference for N, CA, C, O, and CB atoms between two structures. FasIIId1 (reduced) is the RMS difference between model structures containing two disulfides and zero disulfides. CD4 structures are compared to the native CD4 crystal structure, and MCP structures are compared to the native MCP crystal structure.

regions of the molecule that are important in mediating intermolecular interactions, we have used a sequence-to-structure alignment (inverse folding) algorithm (Bryant & Lawrence, 1993) to formulate a model of the most amino-terminal domain of fasciclin III (fasIIId1).

The model based on the threading algorithm supports the hypothesis that fasIIId1 folds into an immunoglobulin V-type structure

Because this algorithm uses tertiary contacts to align a sequence to a proposed structure and does not require sequence similarity between the unknown protein and the proposed structure, it provides an independent means for testing the prediction that fasIIId1 belongs to the Ig superfamily of protein domains. The threading-based and homology-based models are similar in many regions (Figs. 1, 2), providing additional evidence that the fasIIId1 sequence may form an immunoglobulin fold.

Many Ig superfamily domains are characterized by the presence of a conserved disulfide bond at a characteristic position within the fold, between cysteine residues in strands B and F (Alzari et al., 1988). The model predicts that a disulfide bond will be formed between residues 21 and 82 in fasIIId1 (Fig. 4; Kinemage 2). In addition, a second disulfide bond is predicted to form between cysteine residues 30 and 65 in strands C and E (Fig. 4; Kinemage 2). These two predictions may be tested by comparison of the Raman spectra or CD spectra of purified fasciclin III under oxidizing and reducing conditions.

Immunoglobulins and CD4 contain a conserved salt bridge that is not found in the fasIIId1 model. In MCP, Arg L67 forms an electrostatic interaction with Asp L88; in CD4, Arg 54 interacts with Asp 78 (italicized residues in Fig. 2). In fasIIId1, Gly 52 and Asn 76 are structurally homologous to these residues (Fig. 5). Gly 52 has no side chain. Asn 76 is thus exposed to, and can hydrogen bond to, the solvent.

Fasciclin III possesses four potential N-glycosylation sites (Snow et al., 1989), each of which has been shown to be utilized when the molecule is expressed in a soluble form in the baculovirus system (P. Snow, unpubl. obs.). One of these asparagine residues, Asn 42, is located within the amino-terminal domain. In the model, the side chain of this residue is on the exterior of the protein, with its side chain extending into the solvent (Fig. 6; Kinemage 2), consistent with the observation that this residue can serve as a substrate for glycosylation. Asn 42 is found in strand C′ (SCR3), which is part of the CC′FG face of the molecule (Fig. 6; Kinemage 2). This observation suggests that, if Asn 42 is also utilized as a glycosylation substrate during embryonic development, the CC′FG face of fasIIId1 would be sterically hindered, precluding involvement in homophilic interactions. Alternatively, the carbohydrate moiety may project from the protein in such a manner that it does not interfere with such interactions (see below).

Structure and function correlations in fasIIId1

Comparison to immunoglobulins

Various regions within the immunoglobulin superfamily fold have been shown to be important in mediating interactions between molecules. Extrapolation of these observations to the fasIIId1 model may allow predictions to be made concerning regions within the fold that interact to mediate homophilic binding in the developing *Drosophila* nervous system.

Several different surfaces of antibody molecules are involved in subunit interactions within the antibody as well as between the antibody and its cognate antigen. Interactions between constant domains are mediated by β-strands ABED of these domains, whereas analogous interactions between variable domains rely upon interactions between the CC′FG face (Edmundson et al., 1975; Amzel & Poljak, 1979). The location of a poten-

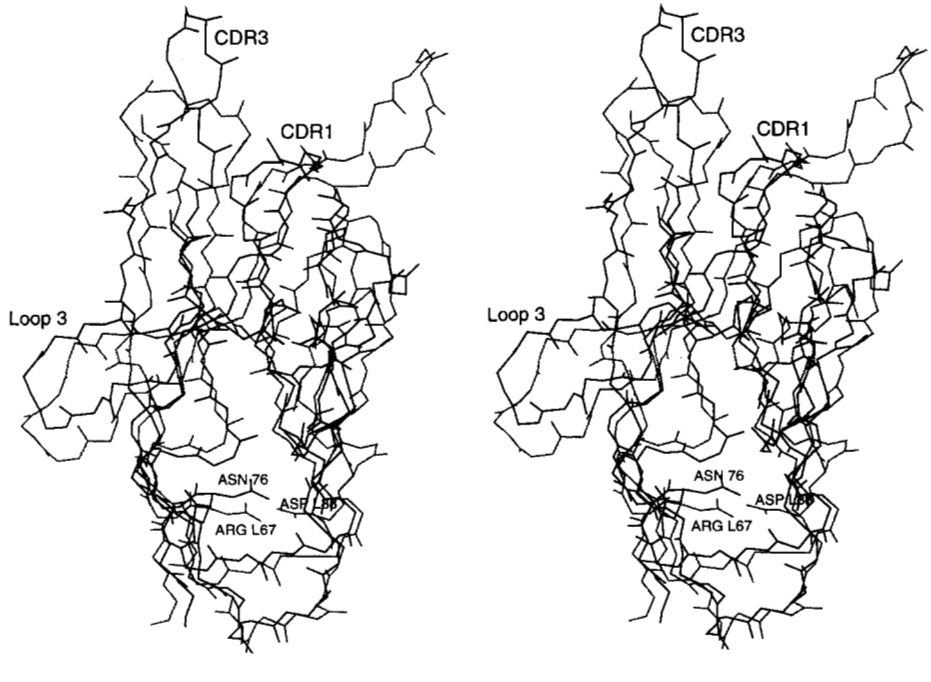

Fig. 5. FasIIId1 model and the crystal structure of residues L1–L112 of immunoglobulin McPC603 are superimposed and shown in stereo. FasIIId1 is shown by dark lines; MCP is shown by lighter lines. Conserved salt bridge in MCP and structurally homologous residues in fasIIId1 are shown. CDRs are marked for orientation.

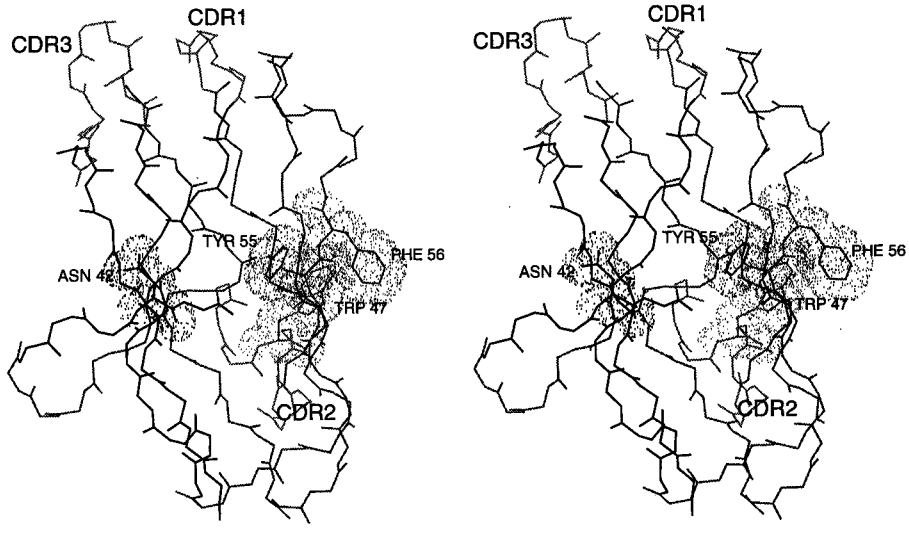

Fig. 6. Location of asparagine 42, the residue in domain 1 that is N-glycosylated, and the cluster of surface aromatic residues (Trp 47, Tyr 55, and Phe 56) in the model are indicated by van der Waals surfaces. β-Strands C, C′, F, and G are shown in dark lines; CDRs are shown in light gray lines; remainder of the molecule is shown in medium gray. (See Kinemage 2, view 2.)

tial glycosylation site on the CC′FG face of the fasIIId1 model suggests that this face may not be involved in dimerization. The ABED face, however, contains a cluster of aromatic residues, Tyr 55, Phe 56, and Trp 47, at the protein surface (Fig. 6; Kinemage 2). One might speculate that hydrophobic residues on the surface of the protein may be important for dimerization, analogous to interactions observed between constant domains of immunoglobulins. This hypothesis is easily tested because Trp 47 is solvent-exposed in the monomer and would become buried if this was a dimerization or recognition face. Tryptophan fluorescence of the molecule should change upon intra- or intermolecular recognition.

The specific antigen recognition sites in immunoglobulins are the complementarity determining regions (CDRs), which are labeled as loop 2 (CDR1), loop 4 (CDR2), and loop 6 (CDR3) in Figures 1 and 2. The CDR1 and CDR2 loops in fasIII are small compared to the normal immunoglobulins. In addition, two of these regions are contained within another distinct hydrophobic pocket on the surface of the model. This pocket consists of residues Ile 27, Leu 84, Val 86, and Ile 95, and these residues are found in loop 2 (CDR1), SCR5 (strand F), loop 6 (CDR3), and SCR6 (strand G), respectively (Fig. 7; Kinemage 2). Because this cluster of hydrophobic residues is found on the part of the molecule that is homologous to the antigen-binding region of immunoglobulins, it suggests that, if fasIIId1 homophilic interactions are similar to antibody-antigen interactions, hydrophobicity might be a key determinant of this interaction.

Comparison to CD4

Our understanding of the molecular basis of interactions mediated by cell surface adhesion proteins belonging to the Ig superfamily, such as fasIII, is limited. In addition to immunoglobulins, two cell surface molecules expressed in the vertebrate immune system, CD2 and CD4, have served as the paradigms for the nature of such interactions.

CD4 interacts in a heterophilic manner with nonpolymorphic determinants of class II major histocompatibility complex (MHC) molecules expressed on the surface of antigen-presenting cells (APC). In addition, human immunodeficiency virus (HIV) envelope protein gp120 has been shown to bind to the CD4 molecule, and thus acts as a receptor for this virus (Dagleish et al., 1984; Klatzmann et al., 1984).

CD4 is composed of four Ig domains, two of the V type, one of the C2 type, and one that has been shown to be similar to C

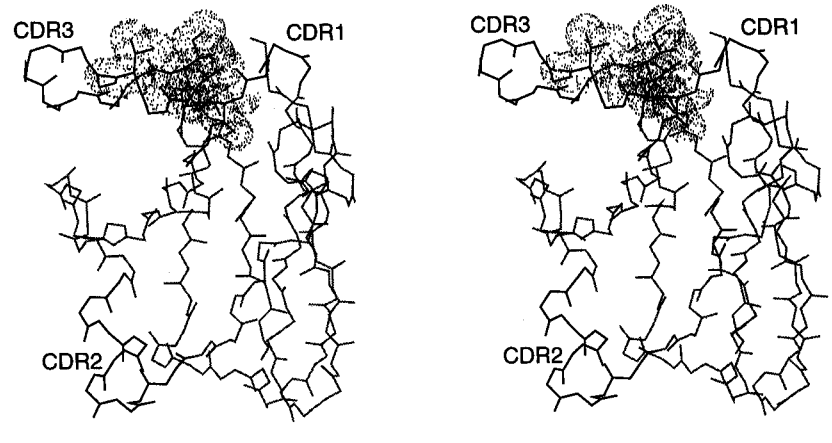

Fig. 7. Hydrophobic pocket found on the fasIIId1 surface consists of Ile 27, Leu 84, Val 86, and Ile 95. Van der Waals surfaces of the four hydrophobic residues are shown by the dotted surfaces; CDRs are indicated by darker lines. (See Kinemage 2, view 3.)

types, but with several unusual features (pseudo C) (Williams et al., 1989; Ryu et al., 1990; Wang et al., 1990). The crystal structure of domains 1 (V) and 2 (pseudo C) has been solved, and the regions that are important in mediating interactions with both MHC class II molecules and the HIV virus have been defined through mutagenesis and antibody inhibition studies.

MHC class II interactions involve a rather extended region of CD4, encompassing portions of both domains 1 and 2. Mutations in the CC′ (loop 3) and C″D (CDR 2) corners in domain 1 and in the A strand of domain 2 show the strongest effect (Clayton et al., 1989); furthermore, antibodies that bind to the BED face of domain 1 interfere with MHC class II interactions (Clayton et al., 1989; Lamarre et al., 1989). This interaction has been proposed to be mediated by a concavity formed between the two CD4 domains that might provide a "notch" for interlocking with a complementary surface on the MHC class II molecule projecting from the apposing cell surface (Wang et al., 1990). Such an arrangement may be sterically necessitated by the involvement of the T cell receptor, which concomitantly interacts with more distal, polymorphic MHC sequences.

There are notable similarities between fasciclin III and CD4, particularly in the relationship between their most amino-terminal two domains. In CD4, strand A of domain 2 is a direct continuation of strand G from domain 1, with no elbow or hinge, generating a stiff, apparently rod-like structure. Limited proteolysis studies of the extracellular domain of fasciclin III suggest that, similar to CD4, the two most amino-terminal domains are very tightly juxtaposed, as judged by their resistance to proteolytic degradation (P. Snow, manuscript in prep.). Thus, it is possible that homophilic interactions between fasIII moieties may also require such an extended interaction between both domains I and II.

Studies directed at defining regions of CD4 required for HIV binding suggest that residues 41–59, in a region analogous to CDR2, serve as the major determinants for this interaction (for a summary of data and references, see Ryu et al., 1990). This region is longer in CD4 than in immunoglobulin molecules and is highly solvent-accessible in the CD4 crystal structure (Ryu et al., 1990; Wang et al., 1990). A possible, additional route of T cell infection occurs when CD4 on uninfected cells interacts with the envelope protein on infected cells to mediate the formation of syncytia. A separate region of CD4 appears to be important in mediating this interaction, including the CDR3-like loop, which is part of a highly negatively charged region of CD4 (Camerini & Seed, 1990).

It is interesting in this regard that fasciclin III also appears to interact with a viral protein to initiate syncytia formation. When infected with a recombinant baculovirus encoding fasciclin III, the host cells undergo a marked transformation resulting in the formation of an extensive syncytium, which is not observed in cells infected by wild-type or other recombinant baculoviruses, even those encoding other insect cell adhesion molecules (P. Snow, unpubl. obs.). If this interaction is similar to that mediated by CD4, the FG loop in fasciclin III, which is also highly negatively charged, with several additional negative charges nearby, may be involved (Fig. 8; Kinemage 2). These negative charges may play a role in mediating interaction with an endogenous baculovirus fusion protein, inducing membrane fusion and syncytium formation. This hypothesis may be tested experimentally by altering the loop by site-directed mutagenesis.

Comparison to CD2

CD2 is a second cell-surface adhesion molecule expressed on the surface of T cells. This cell-surface glycoprotein mediates interactions between T cells and accessory cells by binding in a heterophilic manner to LFA-3 (CD58) and CD59 cell surface proteins (Hahn et al., 1992). The structure of CD2 has been solved and it is composed of an N-terminal V domain, which is separated by a flexible linker from a C-terminal C-type domain that is similar to domain 2 of CD4 (Jones et al., 1992).

The molecular basis of the interaction of CD2 with its two counter-receptors, LFA-3 and CD59, has been explored in detail. LFA-3 is also a member of the Ig superfamily and consists of two C2-type domains (Williams & Barclay, 1988). Mutational analysis has indicated that interactions between CD2 and LFA-3 are mediated by residues located on the CC′ loop, C′ strand,

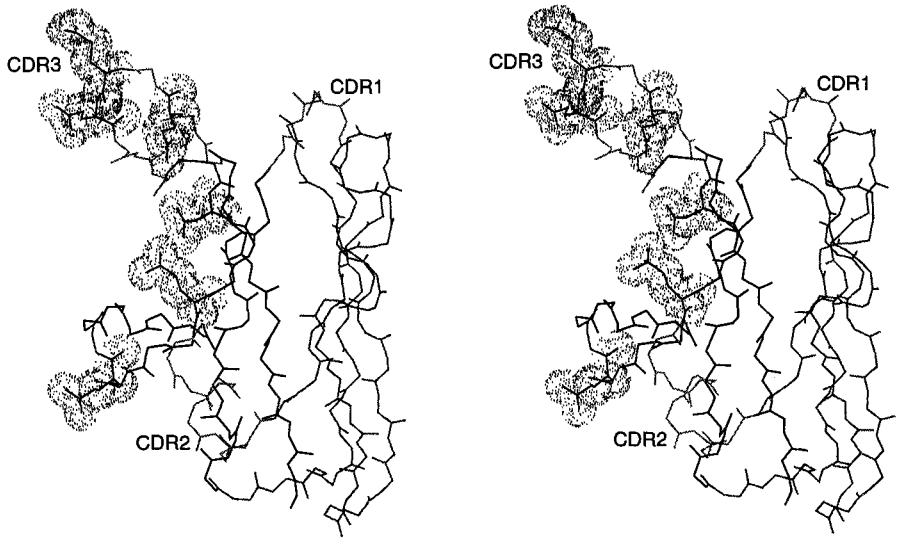

Fig. 8. One face of fasIIId1 contains a large number of negatively charged residues, including Glu 33, Glu 37, Glu 87, Glu 89, Glu 90, and Asp 96. Van der Waals surfaces of the six charged residues are shown by the dotted surfaces; strands CC′FG are indicated by darker lines; CDRs are indicated by very light gray lines. (See Kinemage 2, view 4.)

and FG loop in domain 1 of CD2, which lie diagonally opposed on the GFCC′C″ face (Peterson & Seed, 1987). Binding of CD59 requires only residues located in the FG loop (Hahn et al., 1992). As noted by Jones et al. (1992), the GFCC′C″ face is well suited to mediating binding because it is situated at the top of the molecule and is surrounded by the CC′ (loop 3) and FG (loop 6, CDR3) loops. In addition, this surface also contains a number of charged residues that help orient the two molecules so that their respective polypeptide chains can continue into the membranes of opposing cells.

Homophilic molecular interactions between CD2 molecules also are evident in the crystal between residues believed to be involved in counter-receptor binding, that is, on the GFCC′C″ face. These interactions are similar to those observed in interactions between variable domains in immunoglobulins as predicted (Springer, 1991), with several differences that result in a marked variation in the relationship between β-sheets from that observed in immunoglobulins (Jones et al., 1992).

CD2 differs from both CD4 and Ig variable domains in the regions of the CC′, C′C″ (CDR2), and FG (loop 6, CDR3) loops. These loops bound the functionally important GFCC′C″ sheet and have been shown to be involved in mediating dimerization between variable domains, as well as in CD2 homophilic dimerization as mentioned previously. The CC′ loop of CD2 is shorter than that found in variable domains, and equivalent to that found in CD4. As noted above, this loop has been found to be important in CD4 interactions with class II MHC molecules. The C′C″ loop is lengthened in CD2 and CD4 compared with variable domains, whereas the FG loop of CD2 is markedly larger than that observed in either CD4 or Ig variable domains. Thus, in CD4, which does not dimerize, both the CC′ and FG loops are shortened, and are involved in mediating interactions, whereas in variable domains and CD2 at least one of these loops is lengthened.

In comparison, fasIIId1 resembles CD2 in several features. The CC′ loop is shortened compared with variable domains, and the FG loop is greatly lengthened, as well as being highly negatively charged. In addition, the GFCC′C″ sheet is strongly negatively charged, including residues Glu 87, Glu 88, and Glu 90 of loop 6 (CDR3), Glu 37 of loop 3 (C-C′ loop), Glu 33 of SCR2 (strand C), and Asp 96 of SCR6 (strand G) (Fig. 8; Kinemage 2). Such a highly charged surface could provide an ideal surface for mediating protein–protein interactions through electrostatic forces, as is found in CD2. For example, residues at positions 31, 33, and 37 (Arg, Glu, and Glu) occupy equivalent positions to residues in CD2 that have been found to mediate homophilic interactions in the crystal. It should be noted that the glycosylation site present in fasIIId1 is found on the C′ strand, but appears to be oriented away from the GFCC′C″ face, in a manner that would not interfere with intermolecular forces. It will be important to determine whether the orientation of domain 1 of fasciclin III is in an equivalent position with respect to the cell surface as is found in CD2. Such an orientation would favor interactions between fasciclin III molecules on apposed cell surfaces.

Fasciclin III has been crystallized (Strong et al., 1994), and the determination of its structure, coupled with the results of mutagenesis experiments suggested here, will allow us to directly examine this question. Such information will also allow us to test the accuracy of the sequence-to-structure alignment method for protein structure prediction on an unknown test case.

Methods

Standard sequence alignment

Based on sequence homology, it has been proposed that the structure of fasIIId1 is similar to the structure of an immunoglobulin variable domain (Grenningloh et al., 1990); thus standard sequence alignment and homology modeling techniques (Greer, 1991) were initially used to build the backbone structure of fasIIId1. Immunoglobulin McPC604 (Satow et al., 1987) and CD4, a T cell surface glycoprotein whose crystal structure is similar to an Ig variable domain (Ryu et al., 1990; Wang et al., 1990), were retrieved from the Brookhaven Protein Data Bank (Abola et al., 1987) and the conserved β-strands (strands A–G) of the Igs and CD4 were identified and defined as SCRs. The amino acid sequences of these proteins were aligned based on this structural superposition. Sequence patterns in fasIIId1 that were homologous to regions of the Igs or CD4 were identified and manually aligned, although the alignment was exceptionally difficult because of the overall low sequence similarity between fasIIId1, the Igs, and CD4. (Of the 59 residues in the β-strands of CD4, only 12 [20%] are identical to residues in fasIIId1.) This sequence alignment was used to build the fasIIId1 sequence onto the backbone structure of the N-terminus of 2CD4 (Wang et al., 1990), using the Homology modeling program (version 2.1, Biosym Technologies, Inc.). This model was not pursued further, but the alignment is shown here for purposes of comparison to the alignment obtained using the threading model.

Sequence alignment using a threading algorithm

Using the known Ig structures available in the Brookhaven Protein Data Bank (Abola et al., 1987), 1F19, 1FDL, 1MCP, 1MCW, 1REI, 2FB4, 2FBJ, 2HFL, 2RHE, 2FAB, 3HFM, 3MCG, and 4FAB and the T-cell surface receptor, CD4 (Wang et al., 1990), SCRs 1–7 were assigned to the sequence of immunoglobulin McPC603 (Satow et al., 1987). These SCRs are not just the β-strand regular secondary structures, but are regions of structural similarity among the Igs and CD4. The SCR assignments were slightly modified from those described by Bryant and Lawrence (1993) and take into account the differences between typical immunoglobulin domains and the CD4 structures. The loops (loops 1–6) are the structurally nonconserved regions between SCRs. Loop lengths were allowed to vary to allow for possible insertions and deletions in the fasciclin sequence. The amino acid sequence of fasIIId1, residues 1–103, was aligned to the backbone structure of residues L1–L113 of immunoglobulin McPC603 variable domain (Satow et al., 1987) using the "threading" algorithm described by Bryant and Lawrence (1993). For each possible alignment of fasIIId1 sequence to MCP structure, an "energy" was computed by summing all contact potentials between side chains, between side chains and peptide groups, and between side chains and solvent found in that particular model structure. Individual contact potentials were taken from a table of three-dimensional contact potentials found in a database of 147 globular proteins whose structures have been determined to atomic resolution (Bryant & Lawrence, 1993). To ensure that the contact potentials were not biased toward immunoglobulin structures, none of the 147 proteins were immunoglobulins. The best alignment or "thread" of the fasIIId1

amino acid sequence to the MCP immunoglobulin structure was the one that gave the lowest overall contact energy.

Model building of SCRs and loops

Using the sequence-to-structure alignment determined by the threading algorithm, the backbone of the fasIIId1 SCRs was built manually on the structure of MCP using the program Homology (version 2.1). The backbone conformation of each loop region was obtained in one of two ways. First, if a homologous loop from CD4, MCP, or any other immunoglobulin in the structure database was the same size as the loop in fasIIId1, then the backbone coordinates for the fasIIId1 loop were taken from that Ig loop. For two fasIIId1 loops (loops 3 and 4), no homologous immunoglobulin or CD4 loop was found. In these cases, a subset of the Brookhaven Protein Data Bank (Abola et al., 1987) was searched for loops of the same size as the fasIIId1 loop in which the SCRs on either side of the fasIIId1 loop would superimpose onto the adjacent regions of the test loop with a low RMS deviation (Jones & Thirup, 1986). The 10 best candidate loops thus found were built into the fasciclin backbone model and visualized for atomic overlaps between loop and SCR atoms. One "best" loop that had the smallest number of overlaps with the modeled fasIIId1 SCRs was chosen and the backbone coordinates for the fasIIId1 loop were taken from this best loop.

To relax the loop backbone structures and eliminate atomic overlaps between the loop and SCRs, the loops were subjected to energy minimization using Discover (version 2.9, Biosym Technologies, Inc.) with the CVFF forcefield (Hagler et al., 1979). First, the SCRs were held fixed and the loop backbone was constrained, and the loop side chains were subjected to 100 iterations of steepest descents minimization. Then, the protein SCRs were held fixed and the loop backbone and side chains were subjected to 10,000 steps of minimization using a conjugate gradients algorithm. These minimizations were done in vacuo without charges.

Side-chain packing

Side chains were automatically built onto the backbone using a simulated annealing search strategy (Metropolis et al., 1953) to search possible side-chain conformations. The resulting candidate configurations were ranked by the simple criteria of atomic overlap and combined rotamer probability (P. Shenkin, H. Farid, & J.S. Fetrow, manuscript in prep.). Allowed side-chain conformations were taken from the Ponder and Richards rotamer library (Ponder & Richards, 1987). Four annealing runs were completed and the four resulting side-chain configurations were compared. Although the four configurations were similar, one was chosen based on disulfide bond considerations (see Results). As a control, this algorithm was also run on the CD4 and MCP crystal structures.

Model optimization

The complete model, including the all-backbone and side-chain atoms, was then energy minimized using the the CVFF forcefield (Hagler et al., 1979) as implemented in the program Discover (version 2.9). The purpose of this minimization was not to achieve a global energy minimum for the structure, but simply to relax any major atomic overlaps or strain that was present. All minimizations were done in vacuo with no charges. The following strategy was used: first, the backbone of the model was tethered with a force constant of 100 kcal/Å^2 and the model was subjected to 50 iterations of minimization using a steepest descents algorithm; second, all atoms were allowed to move and the model was subjected to 100 iterations using a conjugate gradient algorithm; third, the backbone was again tethered with a force constant of 60 kcal/Å^2 and 5,000 iterations of conjugate gradient minimization were performed. Energies, hydrogen bonds, and major atomic overlaps were recorded every 100 iterations. This type of alternating strategy has been used to relax the backbone and side chains of other model structures without causing distortion between loop and SCR regions (Hsieh et al., 1994). As a control, the same minimization routine was applied to the CD4 and MCP crystal structures and the CD4 and MCP structures whose side chains were built onto the backbone crystal structure using the side chain-building algorithm described above.

Characterization of the model

Various measures were used to determine the authenticity of the modeled fasIIId1 structure. First, the CD4 sequence was aligned to the MCP structure using the same threading algorithm (Bryant & Lawrence, 1993). Second, to compare the amount of accessible and buried hydrophobic surface area, the solvent-accessible surface area was calculated using the Access program of Lee and Richards (1971). The atom-by-atom output of the Access program was tabulated as hydrophobic (all carbon and sulfur atoms) and hydrophilic (all oxygen and nitrogen atoms) surface area. Third, to determine whether the protein was as well-packed as native proteins, the compactness was calculated using the algorithm of Zehfus and Rose (1986), which computes the ratio of the protein solvent-accessible surface area to the solvent-accessible surface area of a sphere that has a volume equivalent to that of the protein. Because a sphere is, by definition, maximally compact, this ratio is a measure of how much less compact the protein is than if it was maximally compact. Finally, the hydrogen bonds found within the model protein and the control crystal structures MCP and CD4 were tabulated from the output of the Discover program. For the purposes of comparison, each of these measures was computed for the modeled fasIIId1 structure, the CD4 and MCP crystal structures, and the CD4 and MCP structures whose side chains were built onto the known backbone structure using the algorithm of P. Shenkin, H. Farid, and J.S. Fetrow (manuscript in prep.).

Acknowledgments

This work was supported by a postdoctoral fellowship from the University at Albany Center for Molecular Genetics to L.A.C.

References

Abola EE, Bernstein FC, Bryant SH, Koetzle TF, Weng JC. 1987. Protein Data Bank. In: Allen FH, Bergerhoff G, Sievers R, eds. *Crystallographic databases – Information content, software systems, scientific applications*. Bonn/Chester/Cambridge: Data Commission of the International Union of Crystallography. pp 107–132.

Alzari PN, Lascombe MB, Poljak RJ. 1988. Three-dimensional structure of antibodies. *Annu Rev Immunol* 6:555–580.

Amzel LM, Poljak RJ. 1979. Three-dimensional structure of Igs. *Annu Rev Biochem* 48:961–997.

Bryant SH, Lawrence CE. 1993. An empirical energy function for threading protein sequence through the folding motif. *Proteins Struct Funct Genet 16*:92–112.

Camerini D, Seed B. 1990. A CD4 domain important for HIV-mediating syncytium formation lies outside the virus binding site. *Cell 60*:747–754.

Clayton LK, Sieh M, Pious DA, Reinherz EL. 1989. Identification of human CD4 residues affecting class II MHC versus HIV-1 gp120 binding. *Nature 339*:548–551.

Creighton TE. 1993. *Proteins: Structure and molecular properties, 2nd ed.* New York: W.H. Freeman. pp 18–19.

Dagleish AG, Beverley PCL, Clapham PR, Crawford DH, Greaves MF, Weiss RA. 1984. The CD4 (T4) antigen is an essential component of the receptor for the AIDS retrovirus. *Nature 312*:763–766.

Dayhoff MO, Barker WC, Hunt IT. 1983. Establishing homologies in protein sequences. *Methods Enzymol 91*:524–545.

Edmundson AB, Ely KR, Abola EE, Schiffer M, Panagiotopoulos N. 1975. Rotational allomerism and divergent evolution of domains in Ig light chains. *Biochemistry 14*:3953–3961.

Elkins T, Hortsch M, Bieber AJ, Snow PM, Goodman CS. 1990. *Drosophila* fasciclin I is a novel homophilic adhesion molecule that along with fasciclin III can mediate cell sorting. *J Cell Biol 110*:1825–1832.

Fetrow JS, Bryant SH. 1993. New programs for protein tertiary structure prediction. *Bio/Technology 11*:479–484.

Goodman CS, Bastiani MJ, Doe CQ, du Lac S, Helfand SL, Kuwada JY, Thomas JB. 1984. Cell recognition during neuronal development. *Science 225*:1271–1279.

Greer J. 1991. Comparative modeling of homologous proteins. *Methods Enzymol 202A*:239–252.

Grenningloh G, Bieber AJ, Rehm EJ, Snow PM, Traquina ZR, Hortsch M, Patel NH, Goodman CS. 1990. Molecular genetics of neuronal recognition in *Drosophila*: Evolution and function of immunoglobulin superfamily cell adhesion molecules. *Cold Spring Harbor Symp Quant Biol LV*:327–340.

Hagler AT, Lifson S, Dauber P. 1979. Consistent force field studies of intermolecular forces in hydrogen bonded crystals. II. A benchmark for the objective comparison of alternative force fields. *J Am Chem Soc 101*:5122–5130.

Hahn WC, Menu E, Bothwell ALM, Sims PJ, Bierer BE. 1992. Overlapping but nonidentical binding sites on CD2 for CD58 and a second ligand CD59. *Science 256*:1805–1807.

Hsieh M, Hensley P, Brenowitz M, Fetrow JS. 1994. A molecular model of the inducer-binding domain of the galactose repressor of *Escherichia coli*: Implications for the allosteric regulation of DNA binding. *J Biol Chem 269*:13825–13835.

Jones EY, Davis SJ, Williams AF, Harlos K, Stuart DI. 1992. Crystal structure at 2.8 Å resolution of a soluble form of the cell adhesion molecule CD2. *Nature 360*:232–239.

Jones TA, Thirup S. 1986. Using known substructures in protein model building and crystallography. *EMBO J 5*:819–822.

Killeen N, Moessner R, Arviewx J, Willis A, Williams AF. 1988. The MRC OX-45 antigen of rat leukocytes and endothelium is in a subset of the immunoglobulin superfamily with CD2, LFA-3 and carcinoembryonic antigens. *EMBO J 7*:3087–3091.

Klatzmann D, Barre-Sinoussi F, Nugeyre MT, Dauget C, Vilmer E, Griscelli C, Brun-Vezinet F, Rouzioux C, Gluckman JC, Chermann JC, Montagnier L. 1984. Selective tropism of lymphadenopathy associated virus LAV for helper-inducer T lymphocytes. *Science 225*:59–63.

Lamarre D, Ashkenazi A, Fleury S, Smith DH, Sekaly RP, Capon DJ. 1989. The MHC-binding and gp120-binding functions of CD4 are separable. *Science 245*:743–746.

Lee BK, Richards FM. 1971. The interpretation of protein structures: Estimation of static accessibility. *J Mol Biol 55*:379–400.

Metropolis N, Rosenbluth A, Rosenbluth M, Teller A, Teller E. 1953. Equation of state calculations by fast computing machines. *J Chem Phys 21*:1087–1089.

Patel NH, Snow PM, Goodman CS. 1987. Characterization and cloning of fasciclin III: A glycoprotein expressed on a subset of neurons and axon pathways in *Drosophila*. *Cell 48*:975–988.

Peterson A, Seed B 1987. Monoclonal antibody and ligand binding sites of the T cell erythrocyte receptor (CD2). *Nature 329*:842–846.

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.

Richards FM. 1977. Areas, volumes, packing and protein structure. *Annu Rev Biophys Bioeng 6*:151–176.

Rose GD, Geselowitz AR, Lesser GJ, Lee RH, Zehfus MH. 1985. Hydrophobicity of amino acid residues in globular proteins. *Science 229*:834–838.

Ryu SE, Kwong PD, Truneh A, Porter TG, Arthos J, Rosenberg M, Dai X, Xuong NH, Axel R, Sweet RW, Hendrickson WA. 1990. Crystal structure of an HIV-binding recombinant fragment of human CD4. *Nature 348*:419–426.

Satow W, Cohen GH, Padlan EA, Davies DR. 1987. Phosphocholine binding immunoglobulin FAB McPC603. An X-ray diffraction study at 2.7 angstroms. *J Mol Biol 190*:593–604.

Snow PM, Bieber AJ, Goodman CS. 1989. Fasciclin III: A novel homophilic adhesion molecule in *Drosophila*. *Cell 59*:313–323.

Springer TA. 1991. A birth certificate for CD2. *Nature 353*:704–705.

Strong R, Vaughn D, Bjorkman PJ, Snow PM. 1994. Expression and crystallization of a secreted form of fasciclin III, a homophilic cell adhesion molecule. *J Mol Biol 241*:483–487.

Wang J, Yan Y, Garrett TPJ, Liu J, Rodgers DW, Garlick RL, Tarr GE, Husain Y, Reinherz EL, Harrison SC. 1990. Atomic structure of a fragment of human CD4 containing two immunoglobulin-like domains. *Nature 348*:411–418.

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

Williams AF, Davis SJ, He Q, Barclay AN. 1989. Structural diversity in domains of the immunoglobulin superfamily. *Cold Spring Harbor Symp Quant Biol LIV*:637–647.

Zehfus MH, Rose GD. 1986. Compact units in proteins. *Biochemistry 25*: 5759–5765.