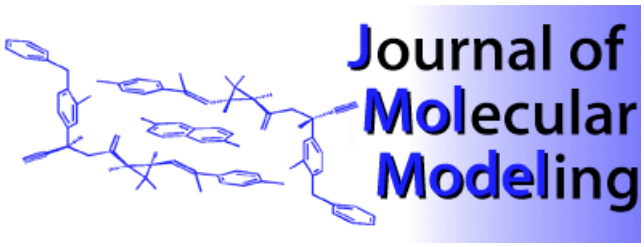
J. Mol. Model. **1998**, *4*, 1 – 11



REVIEW

Three-Dimensional Models of Immune Cell Surface Proteins and Identification of Binding Sites

Jürgen Bajorath[1,2]

[1]Bristol-Myers Squibb Pharmaceutical Research Institute, 3005 First Avenue, Seattle, WA 98121, USA

[2]Department of Biological Structure, University of Washington, Seattle, WA 98195, USA.
E-mail: jbajorath@panlabs.com

Received: 8 September 1997 / Accepted: 27 October 1997 / Published: 9 January 1998

Abstract The extracellular regions of many cell surface proteins of the immune system contain distinct domains that may be linked in many different ways and are often only loosely tethered to the transmembrane segment. In efforts to identify regions critical for binding, molecular models of these domains are used to select residues for mutagenesis and to map binding sites. Many immune cell surface proteins belong to protein superfamilies and display only limited sequence identity compared to proteins of known three-dimensional (3D) structure, often 30% or less. Therefore, detailed 3D structures are difficult to predict, and structure-based sequence analysis and model assessment are particularly important components of the model building process. In some cases, experimentally determined structures have made it possible to assess the accuracy of predictions, which illustrates the opportunities and shortcomings of the approach. Herein the model-based identification of binding sites in cell surface proteins is described and representative examples are discussed.

Keywords Binding sites, Cell surface receptors, Comparative molecular models, Protein superfamilies, Sequence similarity, Structural similarity, Structure-function analysis

Introduction

Cell surface receptor-ligand interactions are critical components of cellular communication and signal transduction [1]. In the immune system, events such as cell adhesion or signaling play, as described below, an important role in the initiation and/or regulation of immune responses. Significant effort has been spent to understand these interactions and the molecular level of detail [2]. We attempt to use 3D structures or models to guide the mutagenesis analysis of novel receptor-ligand interactions. Residues which determine binding specificities may also be identified by other means, for example, mutagenesis based on multiple sequence comparison or even random approaches. However, we aim to obtain a meaningful "picture" of a novel binding site which reveals some of its characteristics. Three-dimensional structures remain to be determined for many immunologically relevant proteins, although significant progress has been made in this area [2]. Therefore, molecular models, albeit less accurate, play an important role in our studies. This

Present address: MDS Panlabs, Computational Chemistry, 11804 North Creek Pkwy. S., Bothell, WA 98011-8805, USA; Tel. (425) 487-8297; Fax (425) 487-8262

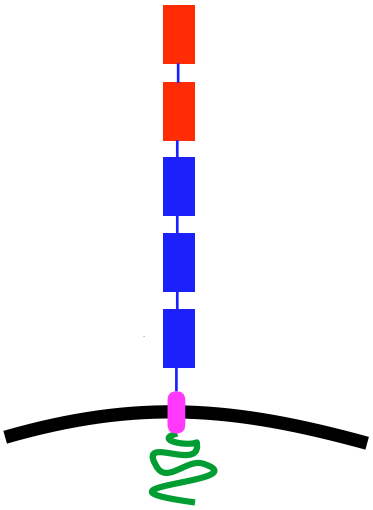


Figure 1 *Schematic representation of Activated Leukocyte Cell Adhesion Molecule (CD166) [52]. The extracelluar region of CD166 contains five Ig-like domains, two V(ariable)-type domains (red) and three C(onstant)-type domains (blue). These domains are followed by a transmembrane segment (magenta) and a cytoplasmic region (green). The modular organization of CD166 is characteristic for many immune cell surface proteins*

review focuses on comparative protein models which have been used to study immune cell surface receptor-ligand interactions and to identify and visualize binding sites. It should highlight, with the help of specific examples, some of the key problems and opportunities in this area of research. This contribution does not aim to provide a comprehensive account of comparative/homology modeling, which has been reviewed elsewhere [3-6].

Immune Cell Surface Proteins

A characteristic feature of immune cell surface proteins is the modular design of their extracellular regions [7]. Many of these proteins contain a number of structurally distinct extracellular domains with different functions (e.g., homotypic interactions, ligand binding, spacers), which are relatively independent of the membrane. Thus, extracellular domains involved in ligand binding can often be expressed in soluble recombinant form such as immunoglobulin (Ig) fusion proteins [8]. This is in contrast to other transmembrane proteins, for example G protein-coupled receptors [9], where binding sites are more intimately connected with the membrane environment and thus more difficult to study. **Figure 1** shows a schematic representation of an immune cell surface protein.

In order to identify binding sites, residues in soluble recombinant domains are targeted by site-specific mutagenesis using polymerase chain reaction (PCR) techniques [10]. Mutagenesis may have different consequences. Mutation of candidate residues may either not affect receptor-ligand interactions, affect binding directly (in the case of important contact residues), or sufficiently perturb the 3D structure to compromise binding indirectly. In the latter case, residues may be distant from the actual binding site, and their identification would thus lead to an incorrect prediction. Therefore, the gross structural integrity of mutant proteins must be tested, for example, using panels of conformationally sensitive monoclonal antibodies (mAbs) against the binding domain [10]. Only mutant proteins are considered structurally intact that consistently bind at wild type levels to mAbs recognizing different 3D epitopes [10, 11]. The binding activities of structurally sound mutant proteins can then be tested using a variety of assay systems [10, 11].

Sequences and Structures

Another characteristic feature is that many immune cell surface proteins belong to protein superfamilies such as the immunoglobulin superfamily (IgSF) [12], the tumor necrosis factor (TNF) and TNF receptor (TNFR) superfamilies [13], or calcium-dependent (C-type) lectins [14], for which structural prototypes have been determined. Thus, three-dimensional structures of many cell surface protein domains may be modeled based on available template(s). However, a major difficulty is the limited sequence identity shared by members of superfamilies, often ~30% or less [12-14]. This low level of sequence identity causes at least two problems. First, similar structural templates are difficult to identify. For example, several IgSF domain types show relatively subtle differences at the sequence level but significant structural variations [12, 15]. Second, topologically accurate alignments of target and template sequences, which are critical for comparative modeling, are difficult to generate [5, 16].

In order to produce accurate alignments in the presence of low sequence identity, sequence comparison should be carried out in light of 3D structures and, if possible, structure comparison [15, 16]. Residues which contribute to the hydrophobic core, structurally constrained positions, and other signature motifs (e.g., disufide bonds, metal coordination spheres) are identified at the 3D level and special consideration is given to match those important positions with residues of similar character (e.g., large hydrophobic or polar) when aligning (multiple) template and target sequences [16]. **Figure 2** shows a representative structure-oriented sequence alignment for members of the TNFR superfamily and **Figure 3** shows shematic viewof the TNFR fold. The initial sequence comparison is a critical step, since inaccurate alignments cause

```
 Domain 1     DS1              DS1 DS2 DS3          DS2       DS3
               *     *          *   **  *            *     ** *
TNFR Human   SVCPQGK YIHPQNNSI  C   CTK CHKGTYLYND  CPGPGQDTDCR
TNFR Mouse   SLCPQGK YVHSKNNSI  C   CTK CHKGTYLVSD  CPSPGRDTVCR
TNFR   Rat   NLCPQGK YAHPKNNSI  C   CTK CHKGTYLVSD  CPSPGQETVCE

Fas  Human   TQNLEGL H...HDGQF  C   HKP CPPGERKARD  CTVNGDEPDCV
Fas  Mouse   KNCSEGL Y...QGGPF  C   CQP CQPGKKKVED  CKMNGGTPTCA

CD40 Human   TACREKQ YLI...NSQ  C   CSL CQPGQKLVSD  CTE.FTETECL
CD40 Mouse   VTCSDKQ YLH...DGQ  C   CDL CQPGSRLTSH  CTA.LEKTQCH


 Domain 2
               *     *          *   **  *            *     ** *
TNFR HUMAN   E CESG.SFTASENHLRH CLS CSK CRKEMGQVEISSCTVDR DTVCG
TNFR Mouse   E CEKG.TFTASQNYLRQ CLS CKT CRKEMSQVEISPCQADK DTVCG
TNFR   Rat   V CDKG.TFTASQNHVRQ CLS CKT CRKEMFQVEISPCKADM DTVCG

Fas  Human   P CQEGKEYTDKAHFSSK CRR CRL CDEGHGLEVEINCTRTQ NTKCR
Fas  Mouse   P CTEGKEYMDKNHYADK CRR CTL CDEEHGLEVETNCTLTQ NTKCK

CD40 Human   P CGES.EFLDTWNRETH CHQ HKY CDPNLGLRVQQKGTSET DTICT
CD40 Mouse   P CDSG.EFSAQWNREIR CHQ HRH CEPNQGLRVKKEGTAES DTVCT


 Domain 3
               *     *          *   **  *            *     ** *
TNFR Human     CRKNQ YRHYWSENLFQCFN CSL CLNG.TVHLS  CQEKQ NTVCT
TNFR Mouse     CKENQ FQRYLSETHFQCVD CSP CFNG.TVTIP  CKETQ NTVCN
TNFR   Rat     CKKNQ FQRYLSETHFQCVD CSP CFNG.TVTIP  CKEKQ NTVCN

Fas  Human     CKPNF FCNSTV..CEHCDP CTK CEHG..IIKE  CTLTS NTKCK
Fas  Mouse     CKPDF YCDSPG..CEHCVR CAS CEHG..TLEP  CTATS NTNCR

CD40 Human     CEEGW HCTSEA..CESCVL HRS CSPGFGVKQI  ATGVS DTICE
CD40 Mouse     CKEGQ HCTSKD..CEACAQ HTP CIPGFGVMEM  ATETT DTVCH
```

Figure 2 *Structure-oriented sequence alignment of TNFR extracellular domains with other family members (Fas, CD40). The sequences were aligned relative to residue positions which are important for the integrity of the TNFR fold (labeled with asterisks). Three TNFR domains (Domain 1, 2, 3) are shown. Gaps were introduced to ensure that corresponding structurally important residues in all three domains are aligned properly. Dots indicate the positions of deletions in the compared sequences. Cysteines involved in the formation of the canonical TNFR disulfide bonds are labeled DS1, DS2, and DS3, respectively*

topological errors in models which can not be corrected later, independent of how the model is constructed.

Model Building and Analysis

On the basis of the initially generated alignment, regions thought to be structurally conserved in template and target, often well-defined secondary structure elements, are selected to provide the core of the model. Side chain replacements in these regions are carried out by, for example, computer graphical replacements of residues in similar conformations (conservative replacements) or rotamer conformational searches

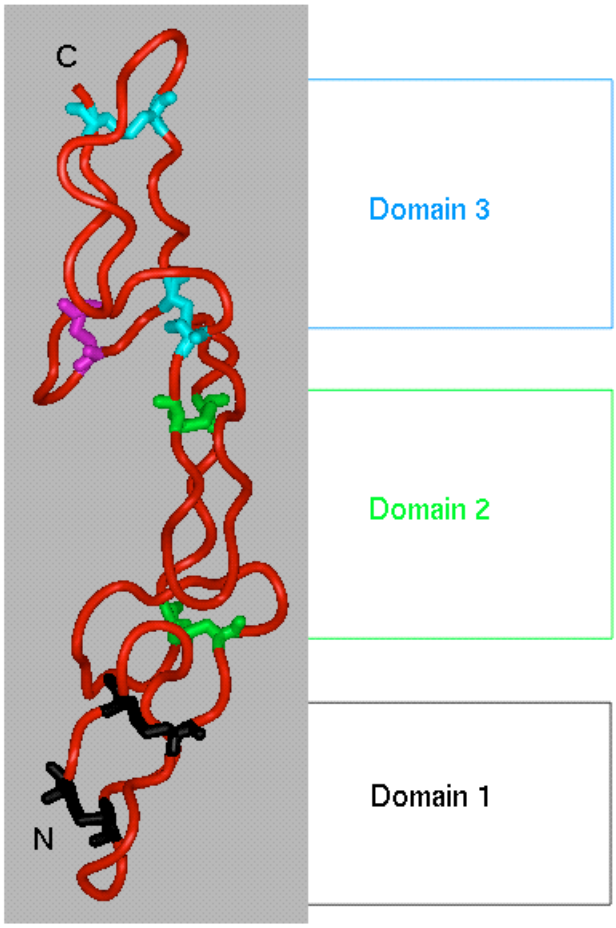


Figure 3 *Ribbon representation of the TNFR-like domains in the CD40 model, as discussed in the text. The representation highlights the stacked domain arrangement thought to be conserved in members of the TNFR superfamily (see Figure 2). Canonical TNFR disulfide bonds in each domain are color-coded. CD40 is predicted to contain an additional disulfide bond (magenta)*

(non-conservative) [17]. Conserved core regions of the model are then complemented with loop, insertions, or deletions, whose conformations are constructed either by extraction of suitable fragments from known 3D structures [18, 19] or conformational search calculations [20], including backbone and side chains. The stereochemistry and intramolecular contacts of the model are refined by energy minimization. As an approximate empirical value, if minimization is carried out until the root mean square (rms) derivative of the energy function is ~1 kcal/mol/Å, a backbone rms deviation (rmsd) of ~1 Å relative to the initially assembled model is frequently observed if a well-refined X-ray structure was available as template. Following energy refinement, the stereochemical quality of the model should be confirmed [21].

For protein models built in the presence of low sequence identity, the assessment of sequence-structure compatibility by inverse folding techniques [22] plays an important role. A variety of methods are available [22]. For example, the energy profile method [23] is able to identify local or global topological errors/misalignments at the level of 3D structures [23]. **Figure 4** shows energy profiles for different versions of a CD152 model (discussed below in more detail) with alternative local sequence alignments between target and template(s). Lower average energy values indicate better sequence-structure compatibility. Structures with overall negative energy profile are not expected to contain significant errors in core regions [23]. The comparison illustrates that energy profiles may also help during model building to discriminate between alternative alignments which are ambiguous at the sequence level.

Limitations and Accuracy Requirements

The modeling procedure outlined above is designed to produce molecular models for a meaningful application in mutagenesis analysis. What are the principal limitations of the approach? Most importantly, the procedure extrapolates from known structural data and prediction of novel folds is not possible. In addition, the prediction of extracellular fragments including multiple domains is more problematic than single domain modeling, since relative domain orientations and associations may be flexible and are in general difficult to predict from sequence. The core regions of models generated as described above are usually template-biased, i.e. closer to the template than the actual structure [24]. While this bias limits overall modeling accuracy, examples discussed below show that it may not be critical for experimental application of the model, provided a closely related template could be identified. However, it is important to note that a number of modeling errors of relatively small magnitude (e.g., mismodeled loop and/or side chain conformations which alter the geometry of the protein surface) may be sufficient to make a meaningful analysis impossible. What are the major challenges? For a meaningful model-based selection of residues for mutagenesis, buried and exposed residues must be correctly predicted and their spatial arrangement (e.g, distances, interactions) well approximated. To identify and map binding sites, it is also required to predict the shape and chemical nature of a protein surface (e.g., flat regions, crevices, hydrophobic patches) with some certainty. Identification of such features critically influences the selection of residues based on computer graphical analysis of the model (e.g., which region is likely to bind a carbohydrate or, alternatively, recognize a protein surface?). In the following, representative examples are discussed.

Selectin Ligand Binding Domains

The selectins are a family of cell adhesion molecules which are involved in the initial attachment of leukocytes to vascular endothelium, an early event in the course of an inflamma-

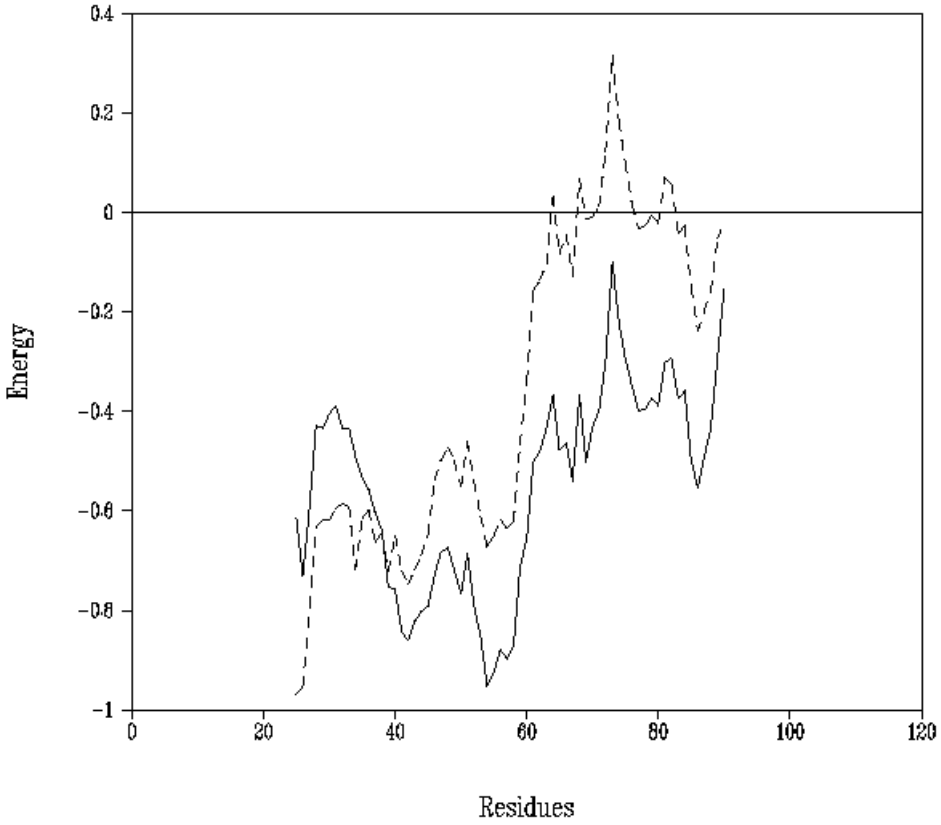


Figure 4 *Comparison of energy profiles for two versions of the CD152 model based on different local sequence alignments. The final CD40 model is discussed in the text. Energy is given as E/kT (E, pairwise residue interaction energy in kcal/mol (1 kcal = 4.18 kJ); k, Boltzmann constant; T, absolute temperature in K) and plotted against residue number. For energy averaging, a fifty residue window is used at each position. The solid line shows the profile of the final model, while the dashed line shows the profile of an intermediate model. This model contains an alignment error in a ß-strand which results in higher energy values (less favorable sequence-structure compatibility)*

tory reaction [25, 26]. The selectins are transmembrane proteins and contain an N-terminal extracellular C-type lectin domain which recognizes sialylated Lewis X (-like) oligosaccharide ligands [26]. Based on the X-ray structure of the mannose-binding protein (MBP) [27], the first structure of a C-type lectin, we and others have built molecular models of the E- and P-selectin ligand binding domains [28-31]. The sequence identity between the C-type lectin domains of the selectins and MBP is ~30% which is, as stated above, typical for many members of protein superfamilies. Model building of the selectins was complicated by the finding that MBP displayed a novel protein fold with an unusually high content of non-regular secondary structure [27]. The selectin models were used, in independent studies, to identify ligand binding sites by mutagenesis approaches [28-30]. These conclusions of these studies were confirmed by the later determined X-ray structure of E-selectin [32]. **Figure 5** shows a comparison of the E-selectin model and X-ray structure [31], and **Figure 6** focuses on the predicted and experimentally determined carbohydrate binding site, which was found to be conserved in E- and P-selectin [28-30]. The overall backbone rmsd for defined structural elements in X-ray structure and model was ~2 Å. **Figure 6** shows that the selectin ligand binding site was predicted with reasonable accuracy. Despite these findings, the structural models could not predict with certainty details of selectin-carbohydrate interactions, and in part different conclusions were reached [32-34]. This illustrates another limitation of the approach. Even if residues

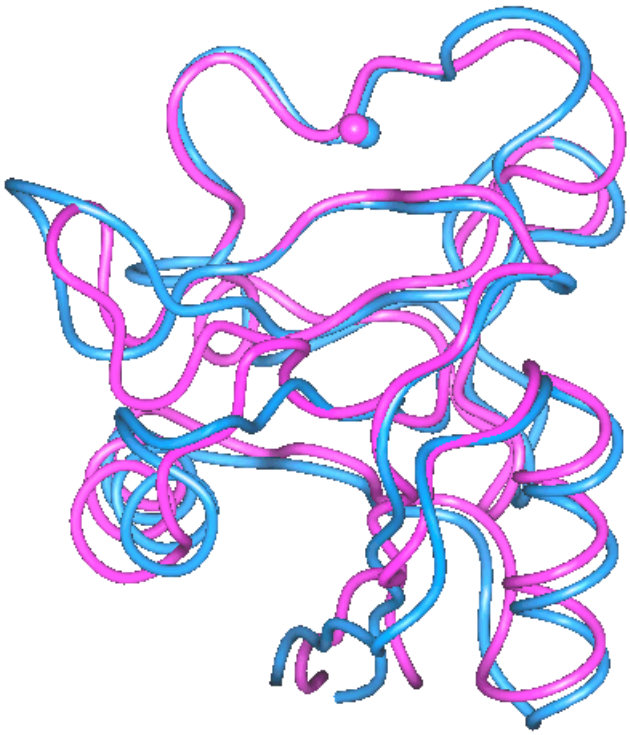


Figure 5 *Backbone superposition of the E-selectin C-type lectin domain X-ray structure (magenta) and the molecular model (blue). Calcium positions are shown as spheres*

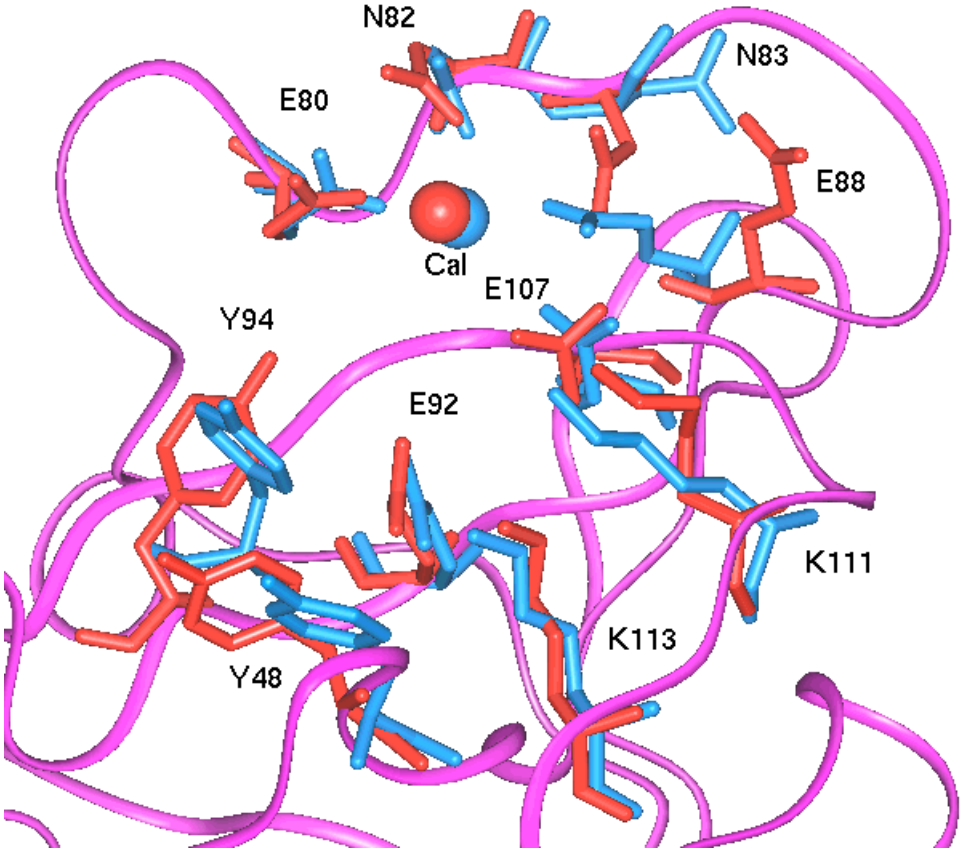


Figure 6 *Close-up view of the E-selectin ligand binding site in the vicinity of a conserved calcium. Residues which are, on the basis of mutagenesis, critical for carbohydrate binding are shown in crystallographically determined (red) and predicted (blue) position and conformation, superimposed on a solid ribbon representation of the X-ray structure (magenta)*

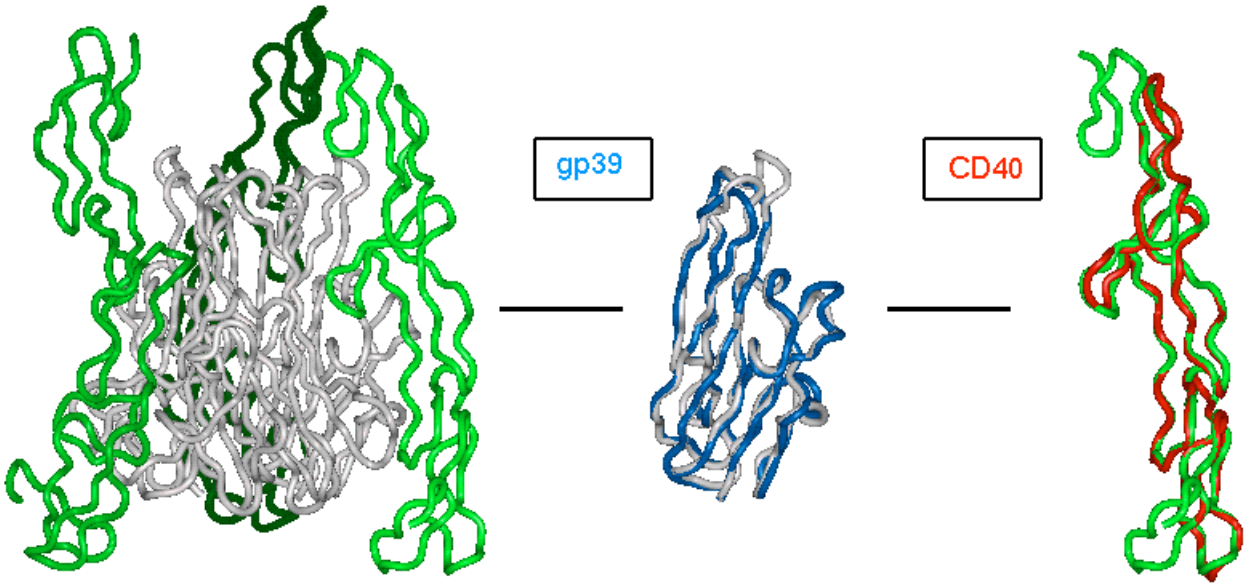


Figure 7 *The TNF-TNFR X-ray structure and molecular models of gp39 and CD40. The TNF homotrimer is shown in gray and the three receptors are colored green. The receptors bind to symmetry-related binding sites at the interfaces between adjacent monomers. CD40 (red) and gp39 (blue) were modeled based on TNFR and TNF, respectively, and are shown superimposed on their respective templates. For clarity, only a gp39 monomer is shown*

critical for ligand binding can be identified and the binding site accurately predicted, precise docking of ligands remains a difficult problem. The detailed comparison of the MBP and E-selectin X-ray structures provided an improved basis for model building of other C-type lectins [35].

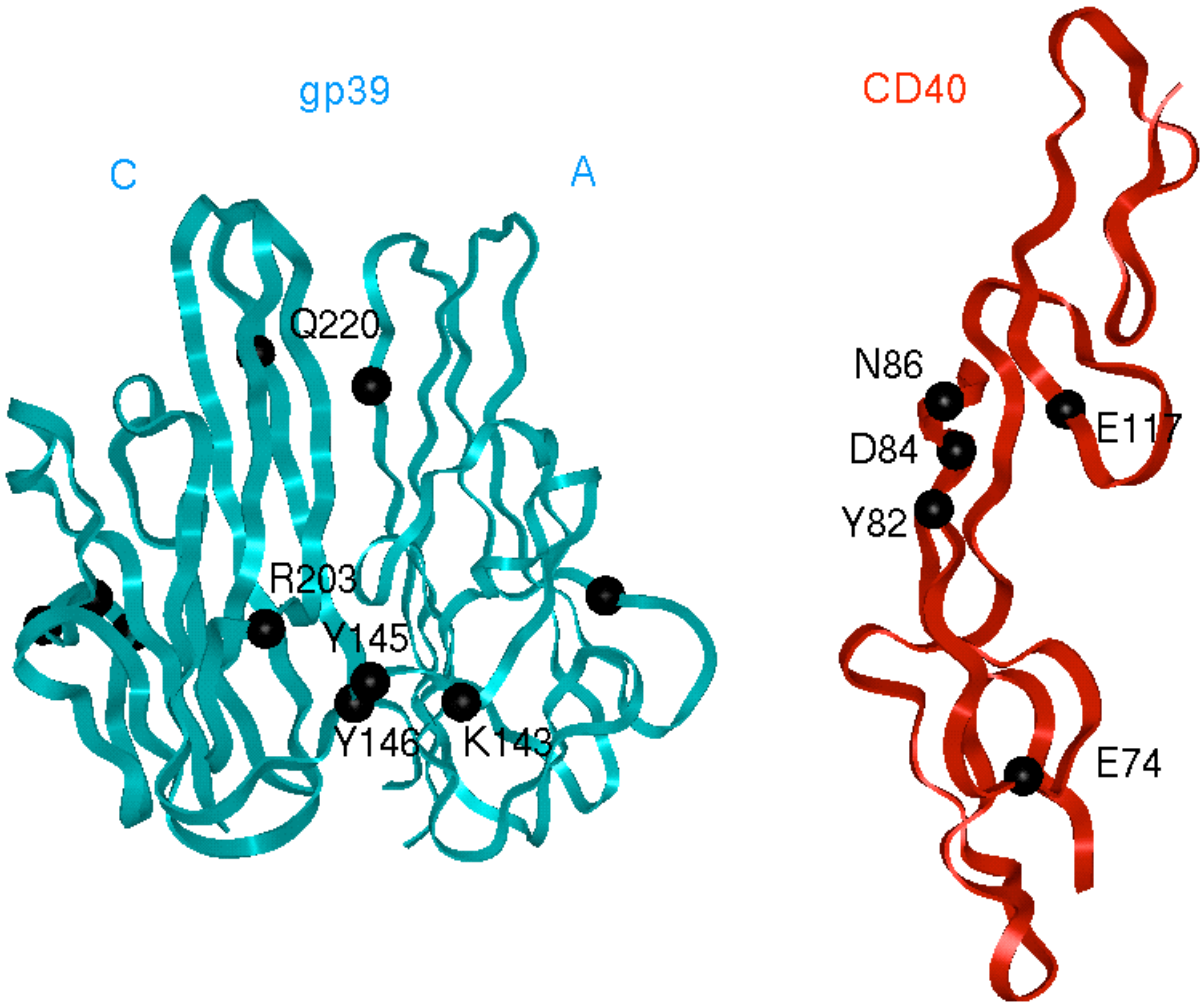


Figure 8 *Residues in gp39 (blue) and CD40 (red) which, when mutated, affect the gp39-CD40 interaction. The alpha carbon positions of these residues are shown as black spheres. For gp39, monomers A and C (small ribbon) are shown (B is omitted for clarity). The view is facing the predicted contact areas*

TNF and TNFR Superfamily Members

The B cell receptor CD40 and its ligand (gp39, CD40L) on T cells are members of the TNFR and TNF superfamilies [13], respectively. Their interaction plays a critical role in B cell activation and the regulation of humoral immune responses [36]. Initial models of the human [10] and mouse [37] CD40 ligand were built based on TNF crystal structures [38, 39]. Like TNF, the CD40 ligand was predicted to be a homotrimer [10, 37] and the CD40 binding site was mapped by mutagenesis to the interface between adjacent monomers [10, 40]. The later determined crystal structure of the human TNF-TNFR complex [41] made it possible to refine the predictions and extend the analysis to CD40 using initially a schematic [40] and subsequently a detailed molecular model [42]. **Figure 7** shows the TNF-TNFR complex [41] and the gp39 and CD40 models. In a subsequent analysis, the gp39 and CD40 molecular models were superimposed on their respective crystallographic templates in the TNF-TNFR complex and residues involved in the putative interface were identified. The regions including these residues were systematically remodeled by conformational search [20, 40] and residues which were consistently involved in gp39-CD40 contacts, independent of how the details of the interface were modeled, were selected for mutagenesis. This protocol proved to be very helpful for identifying residues important for binding [40, 42]. **Figure 8** shows the predicted binding sites in CD40 and gp39, obtained by mapping of residues important for the interaction. The X-ray structure of gp39 has confirmed the location of the binding site and the role of critical gp39 residues [43]. A more detailed comparison of the X-ray and model structuresis yet to be carried out. The gp39 model has also made it possible to classify many gp39 mutations, which naturally occur in immune compromised patients with defective CD40-ligand interactions [36], according to their predicted effects on structure or binding [44].

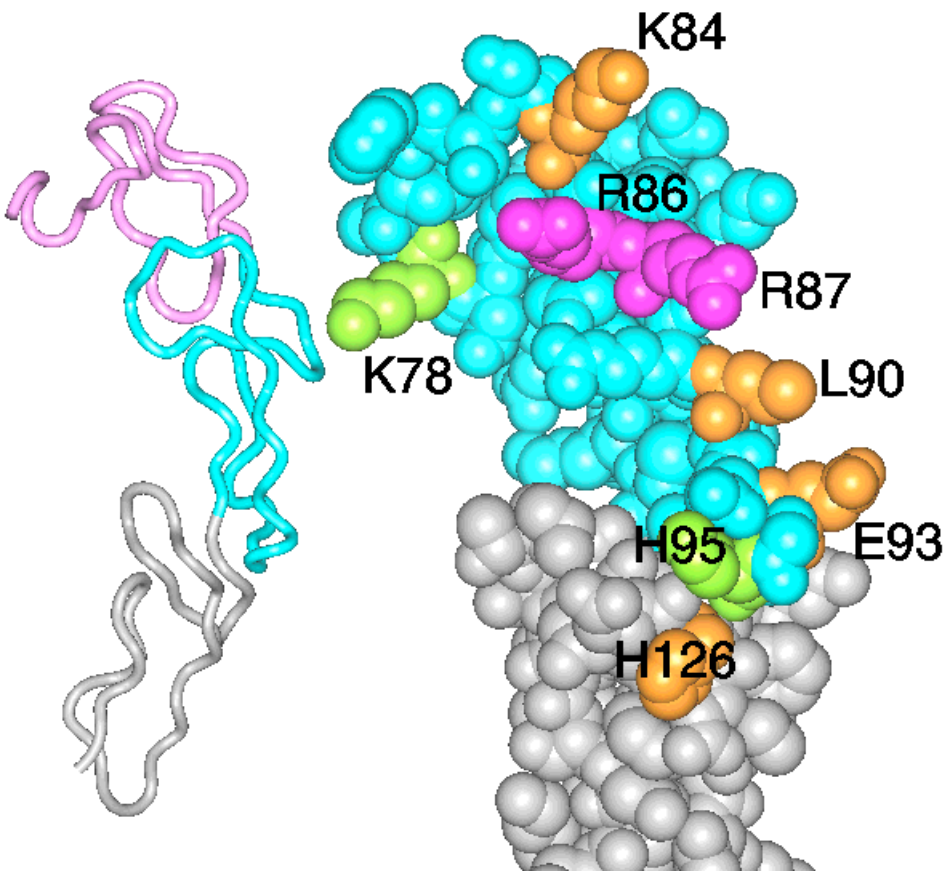


Figure 9 *Fas molecular model and mapping of residues important for ligand binding. On the left, a ribbon outline of the Fas model is shown. The TNFR-like domains are colored pink, blue, and gray, respectively. On the right, a close-up view of domains two (blue) and three (gray) is shown in similar orientation. Tested residues are mapped and color-coded according to their importance for ligand binding (magenta, critical for binding; gold, support binding; green, not important for binding)*

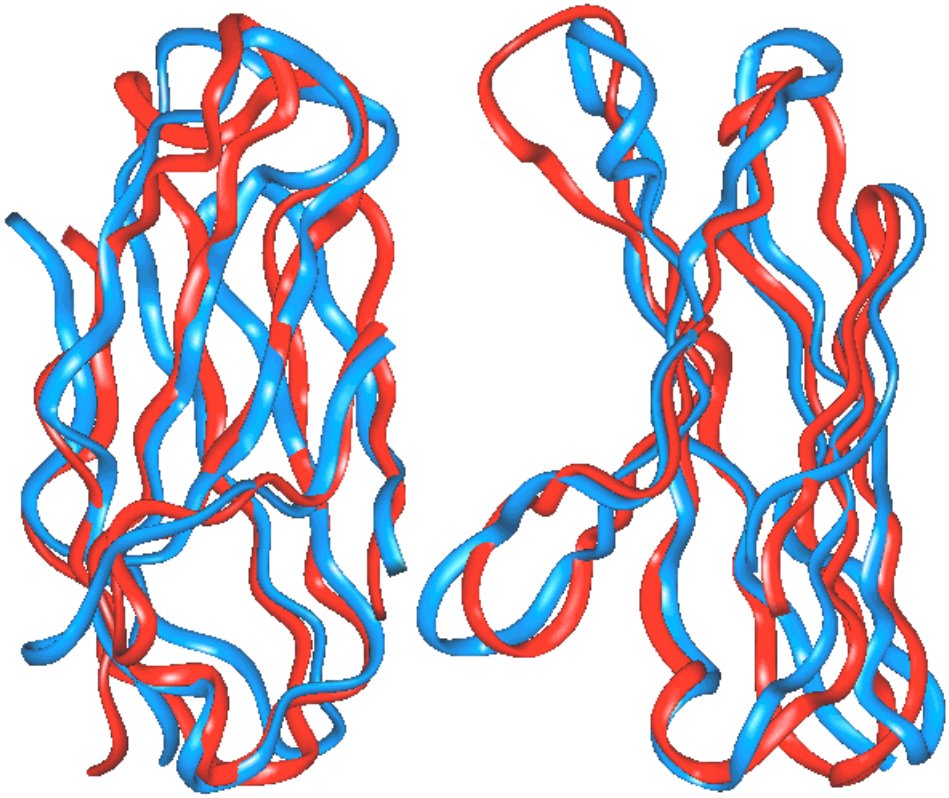


Figure 10 *Backbone superposition of the CD152 model (red) and NMR structure (blue). The left view focuses on the A'GFCC' ß-sheet surface and the right image provides a side view*

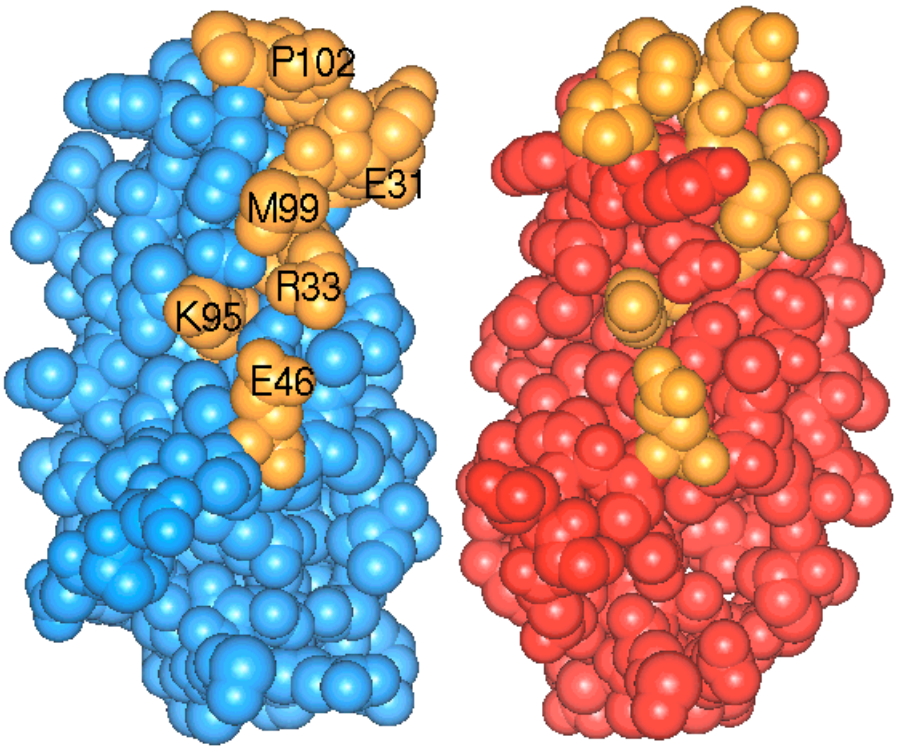


Figure 11 *Side-by-side comparison of the ligand binding site in the CD152 NMR (blue) and model (red) structure. The view focuses on the A'GFCC'ß-sheet surface and residues important for ligand binding are colored gold*

More recently, the structure-function analysis of TNFR proteins was extended to Fas, a receptor that triggers programmed cell death (apoptosis) in the immune system [45]. Following the protocol established for CD40 [42], a molecular model of the Fas extracellular region was generated [46]. Considering the presence of cross-species Fas-ligand interactions [47], residues on the surface of the Fas model conserved in mouse and human Fas (but not in CD40 or TNFR) were selected and mutated to serine. All mutant proteins bound to a panel of conformationally sensitive mAbs and six of eight mutants displayed reduced or abolished ligand binding [47]. **Figure 9** shows the Fas model and residues important for binding. Others extended the analysis to the Fas ligand and reached equivalent conclusions regarding the receptor-ligand interactions [48]. On the basis of these studies, residues at spatially corresponding positions determine the TNFR-TNF [41], CD40-gp39 [40], and Fas-ligand [47, 48] interactions. However, these residues are not conserved. This provides an explanation for the specificity of the respective receptor-ligand interactions [47].

IgSF Proteins

The IgSF [12, 15] is the most populated immune cell surface protein superfamily. Nearly half of the currently known immune cell surface proteins contain Ig or Ig-like domains [7]. Two IgSF proteins with different functions are discussed here. Interactions between CD28/CD152 on T cells and their ligands CD80/CD86 on antigen presenting cells, all of which are IgSF members, provide costimulatory signals critical for effective T cell activation [49]. Both CD28 and CD152 have a single extracellular Ig-domain [49]. Structure-based sequence analysis suggested that these domains were more similar to Ig-V(ariable) domains than other Ig-domain types [16]. Based on this idea, a detailed molecular model of the CD152 ligand binding domain was constructed and CD152 mutants were mapped with the aid of this model [16]. This modeling exercise was challenging, since CD152 displayed less than 20% sequence identity with proteins of known 3D structure. The subsequently determined NMR structure of CD152 [50] allowed a detailed comparison of model and experiment [51]. As predicted, CD152 is most similar to Ig-V domains. An encouraging results of the structure prediction is that no topological errors were observed (i.e., the periodicity of the ß-strands and IgSF core/signature residues were correctly identified) [51]. **Figure 10** shows a backbone superposition of the model and NMR structure and **Figure 11** the modeled and experimentally determined ligand binding site. Major errors were limited to the conformation and spatial position of several loops. Positional deviations are due to rigid body shifts of framework ß-strands and are part of the framework bias discussed above. An overall backbone rmsd of ~2 Å was observed with two mismodeled loops excluded from the comparison [51]. Despite its limitations, the CD152 model provided a reasonable approximation of the ligand binding site (**Figure 11)**. The largest deviation in the binding site region was a relative positional displacement of the loop connecting ß-strands F and G, which contains a sequence motif critical for ligand binding [50, 51].

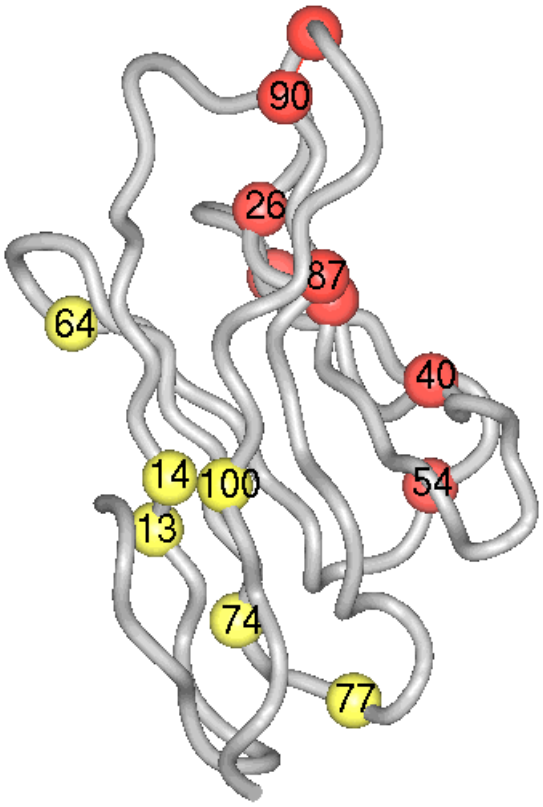


Figure 12 *Molecular model of the N-terminal Ig domain of human CD166 (see also Figure 1). The alpha carbon positions of residues which are important for CD6 binding are shown as red spheres. These residues strictly map to the A'GFCC'C" ß-sheet surface. Residues that differ in human and mouse CD166 are colored yellow and map to the opposite ß-sheet. Thus, the CD6 binding site is completely conserved in these homologues*

In contrast to CD28/CD152, the extracellular region of CD166 consists of five Ig-like domains [52] (**Figure 1**). CD6-CD166 interactions have been implicated in the adhesion of thymocytes to thymic epithelial cells [52] and the regulation of T cell functions [53]. Domain deletion analysis has demonstrated that the N-terminal Ig-V domain of CD166 is sufficient for CD6 binding [54]. A molecular model of the binding domain was built based on CD8 [55] as structural template [56]. Based on this model, residues on the exposed ß-sheet surfaces of the domain and in loops were systematically tested by alanine scanning mutagenesis [57]. As shown in **Figure 12**, residues critical for binding exclusively mapped to the ß-sheet surface formed by strands A', G, F, C, C, C" (A'GFCC'C" face) [57, 58]. Corresponding regions in other IgSF cell surface proteins also contain ligand binding sites [50, 59, 60]. The CD6 binding site was found to be rigorously conserved in mouse and human CD166 [61]. Residue differences in these homologues map to the ß-sheet opposite the binding site (**Figure 12**), which explains the finding that human CD166 binds both human and mouse CD6 [62].

Conclusion

What can be concluded from these studies? The anticipated accuracy of comparative protein modeling increases with the sequence similarity shared by protein with known and unknown structures [6, 63]. In this regard, immune cell surface proteins are challenging targets, and the results of model-based analyses are, as comparisons with experiment illustrate, approximate in nature. However, as demonstrated for the selectins, gp39, and CD152, molecular models have made it possible to identify and visualize binding sites in immune cell surface proteins and to reveal some important features of these receptors/ligands. It is important to note that studies combining protein modeling and experimental analysis as described here critically depend on a number of factors. For example, sufficiently accurate models can only be generated if sequence and sequence-structure comparisons provide meaningful alignments and if closely related template structures can be identified. Furthermore, binding site analysis critically depends on the ability to establish reliable experimental protocols and to obtain reagents. For example, the generation of specific mAbs may be required, which is time consuming and often difficult. Thus, combined modeling and experimental analysis is far from being a routine approach. However, in a number of cases, these studies have become, despite their inherent limitations, an important component in the analysis of novel receptor-ligand interactions.

References

1. Heldin, C.-H. *Cell* **1995**, *80*, 213.
2. Stuart, D. I.; Jones, E. Y. *Curr. Opin. Struct. Biol.* **1995**, *5*, 735.
3. Greer, J. *Meth. Enzymol.* **1991**, *202*, 239.
4. Blundell, T. L.; Sibanda, B. L.; Sternberg, M. J. E.; Thornton, J. M. *Nature* **1987**, *326*, 347.
5. Bajorath, J.; Stenkamp, R.; Aruffo, A. *Protein Science* **1993**, *2*, 1798.
6. Sanchez, R.; Sali, A. *Curr. Opin. Struct. Biol.* **1997**, *7*, 206.
7. Barclay, A. N.; Birkeland, M. L.; Brown, M. H.; Beyers, A. D.; Davis, S. J.; Somoza, C.; Williams, A. F. *The Leukocyte Antigen Facts Book*, **1993**, Academic Prss, London.
8. Hollenbaugh, D.; Doutwright, J.; McDonald, V.; Aruffo, A. *J. Immunol. Meth* **1995**, *188*, 1.
9. Bourne, H. R.; Sanders, D. A.; McCormick, F. *Nature* **1991**, 348, *125*.
10. Bajorath, J.; Chalupny, N. J.; Marken, J. S.; Siadak, A. W.; Skonier, J.; Gordon, M.; Hollenbaugh, D.; Noelle, R. J.; Ochs, H. D.; Aruffo, A. *Biochemistry* **1995**, *34*, 1834.

11. Bodian, D. L.; Skonier, J. E.; Bowen, M. A.; Neubauer, M.; Siadak, A. W.; Aruffo, A.; Bajorath, J. *Biochemisty* **1997**, *36*, 2637.
12. Williams, A. F.; Barclay, A. N. *Ann. Rev. Immunol.* **1988**, *6*, 381.
13. Beutler, B.; van Huffel, C. *Science* **1994**, *264*, 667.
14. Drickamer, K. *J. Biol. Chem.* **1988**, *263*, 9557.
15. Bork, P.; Holm, L.; Sander, C. *J. Mol. Biol.* **1994**, *242*, 309.
16. Bajorath, J.; Linsley, P. S. *J. Mol. Model.* **1997**, *3*, 117.
17. Bajorath, J.; Fine, R. M. *Immunomethods* **1992**, *1*, 137.
18. Jones, T. A.; Thirup, S. *EMBO J.* **1986**, *5*, 819.
19. Bernstein, F. C.; Koetzle, T. F.; Williams, G. J. B.; Meyer, E. F., Jr.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. *J. Mol. Biol.* **1977**, *112*, 535.
20. Bruccoleri, R. E.; Novotny, J. *Immunomethods* **1992**, *1*, 96.
21. Laskowski, R. A.; MacArthur, M. W.; Moss, D. S.; Thornton, J. M. *J. Appl. Cryst.* **1993**, *26*, 283.
22. Wodak, S. J.; Rooman, M. J. *Curr. Opin. Struct. Biol.* **1993**, *3*, 247.
23. Sippl, M. J. *Proteins: Structure, Function, and Genetics* **1993**, *17*, 355.
24. Mosimann, S.; Meleshko, R.; James, M. N. G. *Proteins: Structure, Function, and Genetics* **1995**, *23*, 301.
25. Springer, T. A. *Nature* **1990**, *346*, 425.
26. Lasky, L. A. *Science* **1992**, *258*, 964.
27. Weis, W. I.; Kahn, R.; Fourme, R.; Drickamer, K.; Hendrickson, W. A. *Science* **1991**, *254*, 1608.
28. Erbe, D. V.; Wolitzky, B. A.; Presta, L. G.; Norton, C. R.; Ramos, R. J.; Burns, D. K.; Rumberger, J. M.; Narasinga Rao, B. N.; Foxall, C.; Brandley, B. K.; Lasky, L. A. *J. Cell. Biol.* **1992**, *119*, 215.
29. Erbe, D. V.; Watson, S. R.; Presta, L. G.; Wolitzky, B. A.; Foxall, C.; Brandley, B. K.; Lasky, L. A. *J. Cell. Biol.* **1993**, *120*, 1227.
30. Hollenbaugh, D. H.; Bajorath, J.; Stenkamp, R.; Aruffo, A. *Biochemistry* **1993**, *32*, 2960.
31. Bajorath, J.; Stenkamp, R. E.; Aruffo, A. *Bioconjug. Chem.* **1995**, *6*, 3.
32. Graves, B. J.; Crowther, R. L.; Chandran, C.; Rumberger, J. M.; Li, S.; Huang, K.-S.; Presky, D. H.; Familetti, P. C.; Wolitzky, B. A.; Burns, D. K. *Nature* **1994**, *367*, 532.
33. Bajorath, J.; Hollenbaugh, D.; King, G.; Harte, W., Jr.; Eustice, D. C.; Darveau, R. P.; Aruffo, A. *Biochemistry* **1994**, *33*, 1332.
34. Kogan, T. P.; Revelle, B. M.; Tapp, S.; Scott, D.; Beck, P. J. *J. Biol. Chem.* **1995**, 270, 14047.
35. Bajorath J.; Aruffo, A. *Protein Science* **1996**, *5*, 240.
36. Foy, T. M.; Aruffo, A.; Bajorath, J.; Buhlmann, J. E.; Noelle, R. J. *Ann. Rev. Immunol.* **1996**, *14*, 591.
37. Peitsch, M. C.; Jongeneel, C. V. *Int. Immunol.* **1993**, *5*, 233.
38. Eck, M. J.; Sprang, S. R. *J. Biol. Chem.* **1989**, *264*, 17595.
39. Eck, M. J.; Ultsch, M.; Rinderknecht, E.; de Vos, A. M.; Sprang, S. R. *J. Biol. Chem.* **267**, *1992*, 2119.
40. Bajorath, J.; Marken, J. S.; Chalupny, N. J.; Spoon, T. L.; Siadak, A. W.; Gordon, M.; Noelle, R. J.; Hollenbaugh, D.; Aruffo, A. *Biochemistry* **1995**, *34*, 9884.
41. Banner, D. W.; D'Arcy, A.; Janes, W.; Gentz, R.; Schoenfeld, H.-J.; Broger, C.; Loetscher, H.; Lesslauer, W. *Cell* **1993**, *73*, 431.
42. Bajorath, J.; Aruffo, A. *Proteins: Structure, Function, and Genetics* **1997,** *27*, 59.
43. Karpusas, M.; Hsu, Y.-M.; Wang, J.-H.; Thompson, J.; Lederman, S.; Chess, L.; Thomas, D. *Structure* **1995**, *3*, 1031.
44. Bajorath, J.; Seyama, K.; Nonoyama, S.; Ochs, H. D.; Aruffo, A. *Protein Science* **1996**, *5*, 531.
45. Nagata, S.; Golstein, P. *Science* **1995**, *267*, 1449.
46. Bajorath, J.; Aruffo, A. *J. Comp.-Aided Mol. Design* **1997**, *11*, 3.
47. Starling, G. C.; Bajorath, J.; Emswiler, J.; Ledbetter, J. A.; Aruffo, A.; Kiener, P. A. *J. Exp. Med.* **1997**, *185*, 1487.
48. Schneider, P.; Bodmer, J.-L.; Holler, N.; Mattmann, C.; Scuderi, P.; Terskikh, A.; Peitsch, M. C.; Tschopp, J. *J. Biol. Chem.* **1997**, *272*, 18827.
49. Linsley, P. S.; Ledbetter, J. A. *Ann. Rev. Immunol.* **1993**, *11*, 191.
50. 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.; Linsley, P. S. Nature Struct. Biol., **1997**, *4*, 527.
51. Bajorath, J.; Linsley, P. S.; Metzler, W. J. *J. Mol. Model.* **1997**, *3*, 287.
52. Bowen, M. A.; Patel, D. D.; Li, X.; Modrell, B.; Malacko, A. R.; Wang, W.-C.; Marquardt, H.; Neubauer, M.; Pesando, J. M.; Francke, U.; Haynes, B. F.; Aruffo, A. *J. Exp. Med.* **1995**, *181*, 2213.
53. Gangemi, R. M.; Swack, J. A.; Gaviria, D. M.; Romain, P. L. *J. Immunol.* **1989**, *143*, 2439.
54. Bowen, M. A.; Bajorath, J.; Siadak, A. W.; Modrell, B.; Malacko, A. R.; Marquardt, H.; Nadler, S. G.; Aruffo, A. *J. Biol. Chem*. **1996**, *271*, 17390.
55. Leahy, D. J.; Axel, R.; Hendrickson, W. A. *Cell* **1992**, *68*, 1145.
56. Bajorath, J.; Bowen, M. A.; Aruffo, A. *Protein Science* **1995**, *4*, 1644.
57. Skonier, J. E.; Bowen, M. A.; Emswiler, J.; Aruffo, A.; Bajorath, J. *Biochemistry* **1996**, *35*, 12287.
58. Skonier, J. E.; Bowen, M. A.; Emswiler, J.; Aruffo, A.; Bajorath, J. *Biochemistry* **1996**, *35*, 14743.
59. van der Merwe, P. A.; McNamee, P. N.; Davies, E. A.; Barclay, A. N.; Davies, S. J. *Curr. Biol.* **1995,** *5*, 74.
60. van der Merwe, P. A.; Crocker, P. R.; Vinson, M.; Barclay, A. N.; Schauer, R.; Kelm, S. *J. Biol. Chem.* **1996**, *271*, 9273.
61. Skonier, J. E.; Bowen, M. A.; Aruffo, A.; Bajorath, J. *Protein Science* **1997**, *6*, 1768.
62. Bowen, M. A.; Bajorath, J.; D'Egidio, M.; Whitney, G. S.; Palmer, D.; Kobarg, J.; Starling, G. C.; Siadak, A. W.; Aruffo, A. *Eur. J. Immunol.* **1997**; *27*, 1469.
63. Chothia, C.; Lesk, A. M. *EMBO J.* **1986**, *5*, 823.