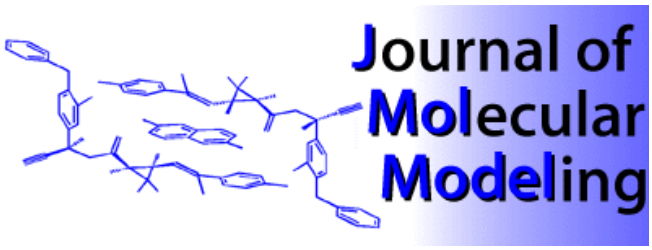
J. Mol. Model. **1997**, *3*, 287 – 293



Molecular Modeling of Immunoglobulin Superfamily Proteins: CTLA-4 (CD152) - Comparison of a Predicted and Experimentally Determined Three-Dimensional Structure

Jürgen Bajorath[1,3,]*, Peter S. Linsley[1], William J. Metzler[2]

[1]Bristol-Myers Squibb Pharmaceutical Research Institute, 3005 First Avenue, Seattle, Washington 98121, USA; Tel: (206) 727-3612; Fax: (206) 727-3602 (bajorath@protos.bms.com)

[2]Bristol-Myers Squibb Pharmaceutical Research Institute, PO Box 4000, Princeton, New Jersey, 08540-4000, USA,

[3]Department of Biological Structure, University of Washington, Seattle, Washington 98195, USA.

Received: 25 June 1997 / Accepted: 23 July 1997 / Published: 30 July 1997

Abstract

The CD28 and CTLA-4 (CD152) receptors on T cells recognize CD80 and CD86 ligands on antigen presenting cells. These interactions provide and control costimulatory signals required for effective T cell activation. CD28 and CTLA-4 belong to the immunoglobulin superfamily (IgSF) and contain a single extracellular ligand binding domain. The three-dimensional (3D) structure of the binding domain of CTLA-4 was modeled previously using a combination of structure-based sequence comparison, IgSF consensus residue analysis, conformational search, and inverse folding calculations. Recently, the 3D structure of CTLA-4 was determined by NMR. Comparison of the modeled and experimentally determined CTLA-4 structure has made it possible to assess the accuracy of our predictions. We found that the overall accuracy of the model was sound and sufficient for a meaningful application of the model in experimental studies. Major errors in the model are limited to the conformation and position of some loops. Our studies on CTLA-4 provide an example for the opportunities and limitations of comparative protein modeling in the presence of low sequence similarity.

Keywords: Immunoglobulin superfamily, comparative protein structure prediction, sequence-structure analysis, molecular model, solution structure, model-structure-comparison

Introduction

T cell activation is critically dependent on the presence of costimulatory signals which complement T cell receptor engagement by peptide-MHC complexes [1]. Costimulatory signals triggered by interactions between CTLA-4 (CD152) and CD28 on T cells and CD80 and CD86 on antigen presenting cells are critical for T cell activation [1]. CD28/CTLA-4 and CD80/CD86 are members of the IgSF. Both CTLA-4 and CD28 contain a single ligand binding domain which forms disulfide-linked homodimers on the cell surface [2]. Previ-

* *correspondence to J. Bajorath, BMS-PRI*

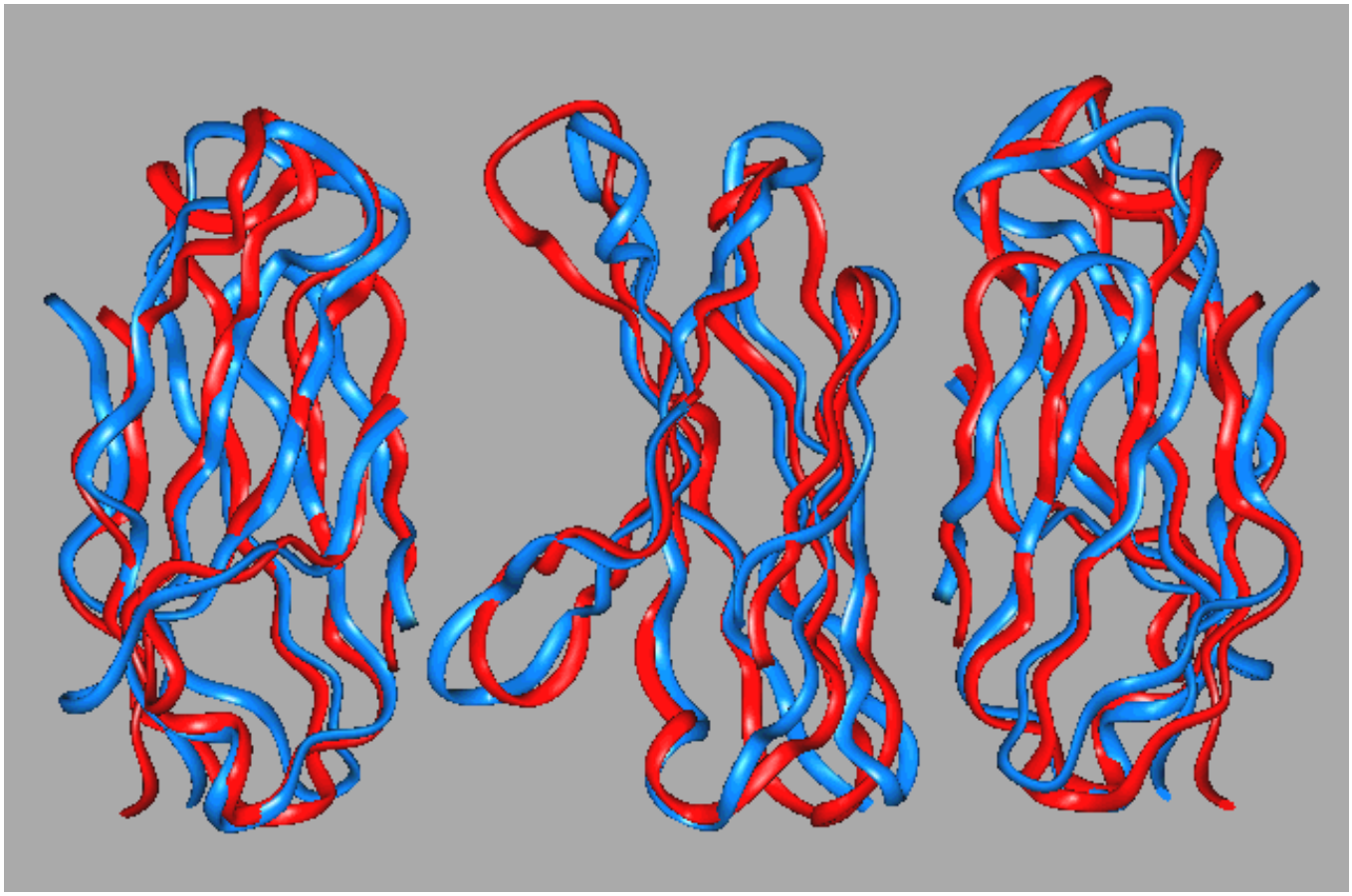


Figure 1. *Comparison of the CTLA-4 model (red) and NMR structure (blue). A backbone superposition is shown in three different orientations. The left view focuses on the A'GFCC' β-sheet surface and the right view on the opposite ABED face. The center image provides a side view.*

ously, we have built a detailed molecular model of the extracellular Ig domain of CTLA-4 by combining sequence analysis, structure comparison, and a variety of computer modeling methods [3]. This structure prediction was challenging since CTLA-4 displays only ~20% sequence identity to proteins with known 3D structure. At this low level of sequence identity, standard homology modeling methods may fail, since accurate alignments of the target sequence with potential template structures are difficult to generate.

Recently, the solution structure of an extracellular fragment of CTLA-4 was determined [4]. Thus, we are now able to compare the CTLA-4 model with an experimentally determined structure. CTLA-4 modeling provides a blind test prediction. Such tests are essential for a realistic assessment of modeling accuracy and important for the improvement of modeling methods. In this communication, we report the comparison of the modeled and an experimentally determined 3D structure of CTLA-4. The CTLA-4 Ig domain type was correctly predicted and the quality of our model was sound. Structure-oriented sequence and consensus residue analysis proved to be powerful tools to accurately predict β-strands and their periodicity (exposed and buried residues). Major limitations of the model were inaccurately modeled loop conformations and relative segmental shifts of some structural elements. Despite its limitations, the CTLA-4 model was sufficiently accurate to identify the CD80/CD86 binding site.

Methods

Superpositions of structures were calculated using InsightII (MSI, San Diego, CA). For comparison, the CTLA-4 molecular model, as reported in [3], and the averaged energy minimized NMR coordinates [4] were used. Secondary structure and intermolecular contacts were analyzed with Procheck [5]. Computer graphics were generated with InsightII. Model coordinates have been deposited with the Journal and are available from the authors via electronic mail. CTLA-4 NMR coordinates have been submitted to the Brookhaven Protein Data Bank (pdb code "1AH1").

Results

Model building of CTLA-4 has been described in detail [3]. Briefly, CTLA-4 sequences from different species were

```
          1 * * *    *    15   * *   25         * *  *40
modCTLA-4:MhVAQ PAVVlas sRG IASFVCEyas PgkatEVR VTVLRQA dsQ-
nmrCTLA-4:xHVAQ PAVVLAS SHG IASFVCEYAS PGKATEVR VTVLRQA DSQ-

           A       A'          B                C
          ---    -----     ----------          -------



            **       55      *  64**    73    * * *   *   88
modCTLA-4:VTEVCA ATYMMG NELT FLDDS ICTGTSs gn QVNLTIQ GLRAMDT
nmrCTLA-4:VTEVCA xxxxxx xxxx xxxxx ICTGTSS GN QVNLTIQ GLRAMDT

           C'                        D         E
          ------                   -------    -------



          * * * *     101   106     * *118
modCTLA-4:GLYICKVE LMypppyy lGIGNgTQIYVI
nmrCTLA-4:GLYICKVE LMYPPPYY-LGIGNGTQIYVI

             F                 G
          -------          ------------
```

Figure 2. *Alignment of residues in the CTLA-4 model (modCTLA-4) and NMR structure (nmrCTLA-4) based on backbone superposition of the structures. Defined ß-strands are underlined and labeled. IgSF V-set consensus residues and additional largely conserved core residues, which were used to align the CTLA-4 sequence with template structures [3], are labeled with asterisks. Residues are shown in upper case if their alpha carbon positions are within 3.5 Å in the model and NMR structure. Residues omitted from the comparison are denoted x in nmrCTLA-4. Residue numbering is according to [12].*

analyzed in light of IgSF consensus residues [6,7] and with the aid of an alignment of representative IgSF structures. Although CTLA-4 displays only ~20% sequence identity to proteins with known 3D structure, these comparisons suggested that the structure of the extracellular domain of CTLA-4 was similar to Ig V(ariable)-domains. Due to the low sequence similarity between CTLA-4 and available V-type structures, matching of IgSF core residue [3] was used to align the CTLA-4 sequence with relevant template structures. In the absence of a closely related structural template, the antibody variable light (VL) chain of REI [8] was selected as a representative Ig V-fold for modeling of CTLA-4. Predicted core β-strands were copied to the model and alternative alignments of β-strands were assessed by energy profile analysis [9]. Loop and side chain conformations were modeled based on database conformations or specific conformational search protocols [3,10,11]. The model was refined by energy minimization and its sequence-structure compatibility [9] was assessed. This analysis suggested that significant errors in the core region of the model were absent [3]. The solution structure of an extracellular fragment of human CTLA-4 (residues 1-123) was determined by multidimensional heteronuclear NMR [4]. In the NMR structure, the conformation of residues at the N- and C- termini and residues encompassing the C" region, including adjacent loops, were experimentally not well defined, indicating high flexibility in solution. Thus, these regions were omitted from the comparison and only residues 2-50 and 67-118, which form β-sheets (ABED) and (A'GFCC'), were included.

Figure 1 shows the superposition of the model and NMR structure. As predicted, the CTLA-4 structure is most similar to the Ig V-type and monomeric. The overall backbone root mean square deviation (rmsd) for 100 residues, including all loops, is 2.90 Å. Exclusion of the B-C and F-G loops (see below) from the rms comparison reduces the overall backbone rmsd between model and NMR structure to 2.14 Å. The model includes the canonical IgSF disulfide bond (C21-C94) and correctly predicts an additional non-IgSF disulfide bond formed between C48 and C68. Multiple structure-oriented sequence analysis [3] with a focus on buried IgSF consensus residues [6,7] was used to align the CTLA-4 sequence relative to the selected template structure. Superposition of the model and NMR structure confirmed that the IgSF core residues were correctly predicted so that alignment errors in well defined β-strand regions were absent (Figure 2**)** After optimal superposition, alpha carbon atoms of 75% of all modeled residues were within 3.5 Å distance to the corre-

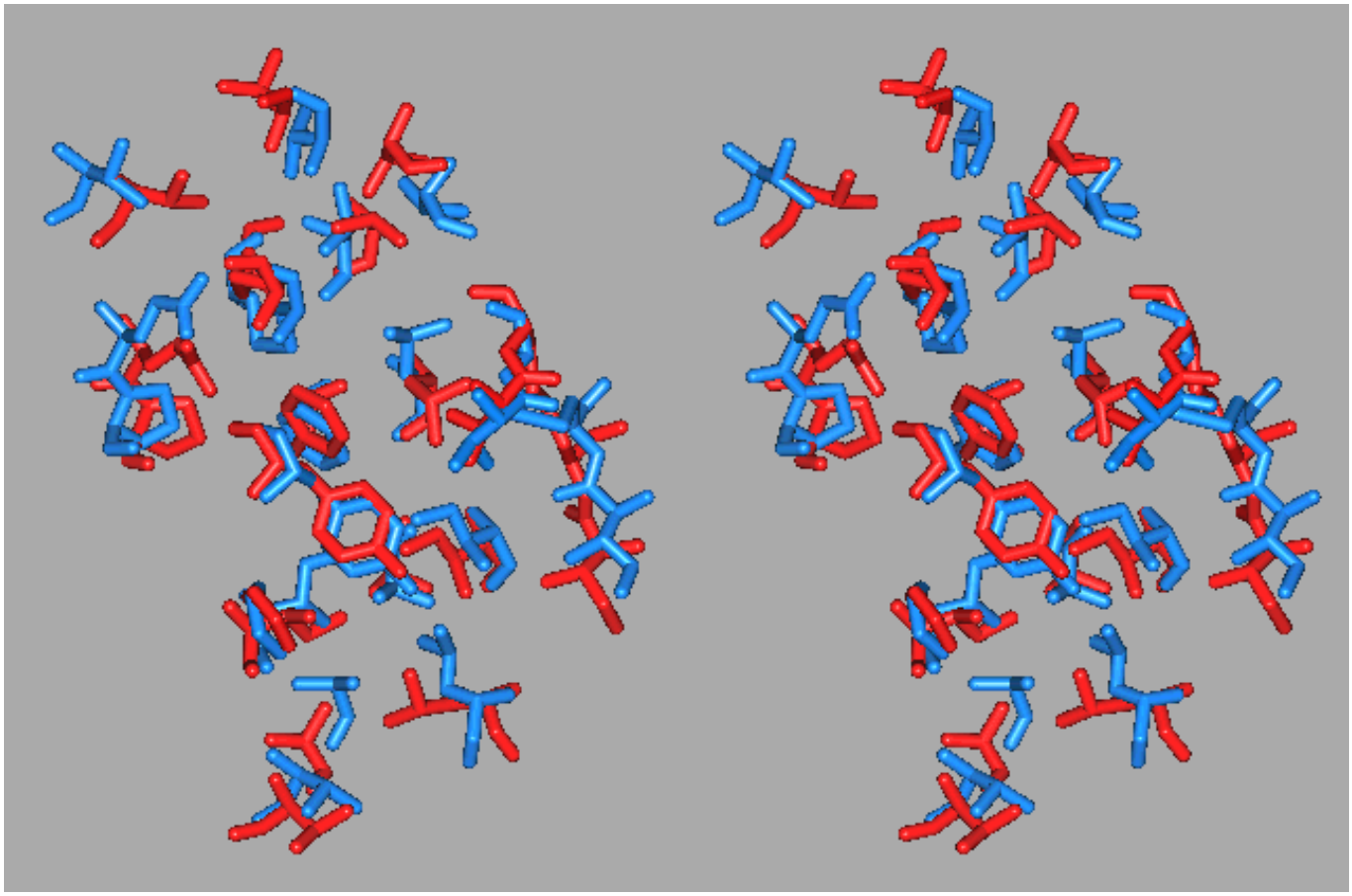


Figure 3. *Stereo superposition of predicted (red) and experimentally determined (blue) IgSF consensus/core residues in CTLA-4. The view focuses on the A'GFCC' face. The IgSF consensus positions in CTLA-4 were correctly identified.*

sponding experimentally observed positions (see Figure 2 for details). Only one of the 24 predicted IgSF consensus/core positions (A11), which were used to anchor the alignment, displayed a larger deviation (3.75 Å) due to a positional shift of the A'-strand (see below). Nine well defined β-strands (Figure 2) were correctly predicted. Differences between predicted and experimentally observed β-strands were, if present, limited to the conformation of one residue at the N- or C-terminus of a strand. Figure 3 shows the comparison of the predicted and experimentally determined IgSF consensus/core positions. Superposition of 24 IgSF consensus residues resulted in a backbone rmsd of 1.61 Å.

Major deviations between the model and NMR structure occur in the conformation and position of loops. Table 1 shows the comparison of modeled and experimentally determined loop conformations. Three of 6 loop conformations were modeled with an accuracy of better than 1 Å rmsd. Incorrectly modeled was the conformation of the B-C loop, residues of which support CD80/CD86 binding [12]. The CTLA-4 signature sequence motif M99YPPPY104, which is critical for CD80/CD86 binding [12], was accurately mapped to the tip of the F-G loop. Figure 4 shows a comparison of the modeled and observed F-G loop. In the NMR experiment, the conformation of proline P103 was determined as cis, while the conformations of P101 and P102 could not be determined. In the NMR structure, P101-P103 are modeled as trans-trans-cis conformers. In the CTLA-4 model, cis-trans-cis conformations are predicted for P101-P103. Although the loop conformations are quite similar, with a backbone rmsd of 1.54 Å, the spatial positions of the F-G loops differ in the context of the 3D structures (Figure 4), due to a shift of a part of the G-strand. Other significant segmental shifts are observed in the A'-B and D-E regions. As discussed below, such positional deviations largely depend on the choice of the template structure(s) for modeling.

Figure 5 shows a side-by-side comparison of the modeled [3] and experimentally determined [4] CD80/CD86 binding site. CTLA-4 residues which, when mutated, affect the binding to CD80 and CD86 [4,12] were mapped on both the model and NMR structure. In the model, all residues critical for binding map to surface positions. Although deviations of side chains and the F-G loop conformation are observed, the CD80/CD86 binding site was correctly predicted. Thus, CTLA-4 modeling accuracy was sufficient to design and rationalize mutagenesis experiments.

Table 1. *Comparison of modeled and experimentally determined loop conformations.*

Loop	Residues	Method	b-rmsd(Å)	a-rmsd(Å)
A'-B	11-16	ST	1.46	2.22
B-C	24-33	CS	2.85	4.43
C-C'	40-43	CS	0.75	1.34
D-E	73-76	ST	0.59	1.40
E-F	83-90	ST	0.74	1.65
F-G	98-106	CS	1.54	3.28

Loop conformations were compared by direct superposition; b-rmsd gives the backbone and a-rmsd the all atom root mean square deviation. CS means that loops were modeled by conformational search [10] following a protocol described in [3]. ST means that loops were modeled based on the backbone conformation of corresponding loops in REI [8].

Discussion

CTLA-4 provides an example for a problem faced in the modeling of many IgSF proteins: the target shows only low sequence identity to proteins with available 3D structure, here ~20%. In the absence of a more homologous structure, the antibody VL domain of REI was selected as template for model building. For at least two reasons, low sequence similarity generally limits the accuracy of comparative structure predictions. First, topologically correct alignments of sequences of target proteins to template structures are difficult to generate when sequence similarities are low. Second, the magnitude of structural deviations in core regions of related structures increases sharply with decreasing sequence similarity [13]. Rigid body shifts of corresponding secondary structure elements occur in structures of distantly related proteins, but also in some highly homologous structures [14]. In comparative model building, these effects, which are difficult, if not impossible to predict, bias a model towards the core region structure of a selected template [15,16] and limit modeling accuracy by changing the spatial relation of residues [16]. These limitations are, however, less severe than those resulting from topological misalignments which, in addition, lead to an incorrect modeling of core and surface residues. If no closely related template structure can be identified, averaging of several possible templates may reduce the magnitude of positional displacements but is likely to introduce other structural inconsistencies.

In the case of CTLA-4, IgSF consensus residue analysis made it possible to correctly align the CTLA-4 sequence to a prototypic IgSF V-type fold. The absence of substantial errors in the sequence-to-structure alignment resulted in the accurate prediction of exposed and buried residues in well-defined β-strand regions, a prerequisite for a successful application of the model in experimental studies. The initially

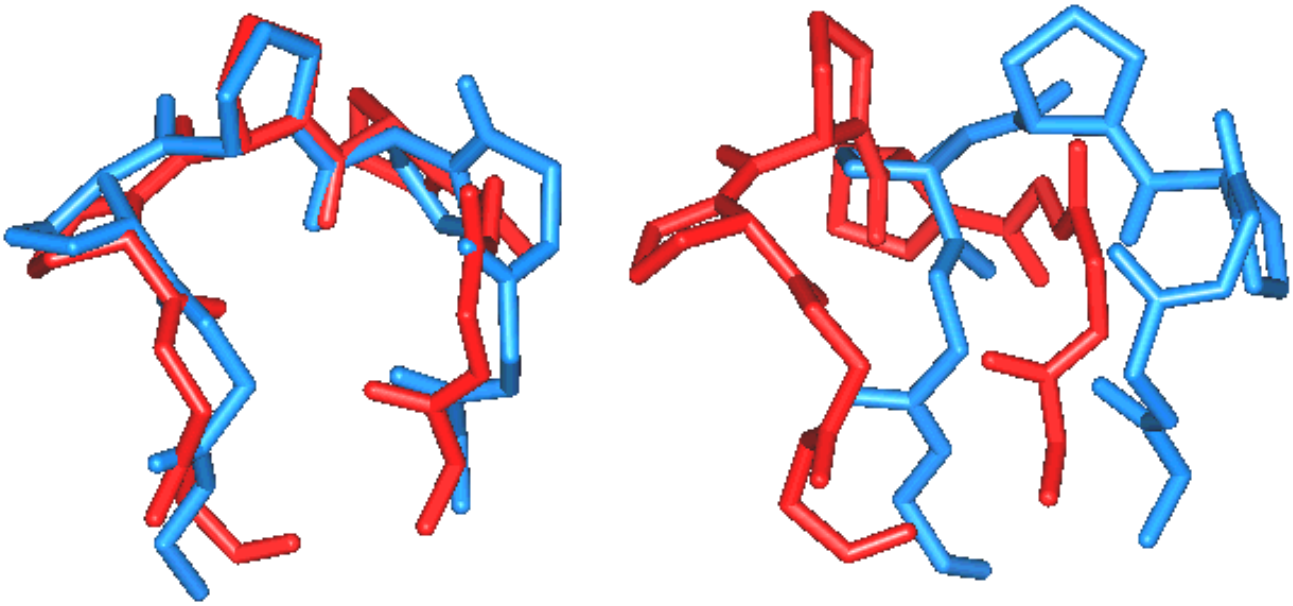


Figure 4. *Comparison of the modeled (red) and observed (blue) F-G loop conformations in CTLA-4. The loop contains the M99YPPPY104 motif which is critical for ligand binding. On the left, a direct superposition of the loop backbones is shown. For clarity, only proline side chains are displayed. On the right, the loops are shown after backbone superposition of the entire model and NMR structure.*

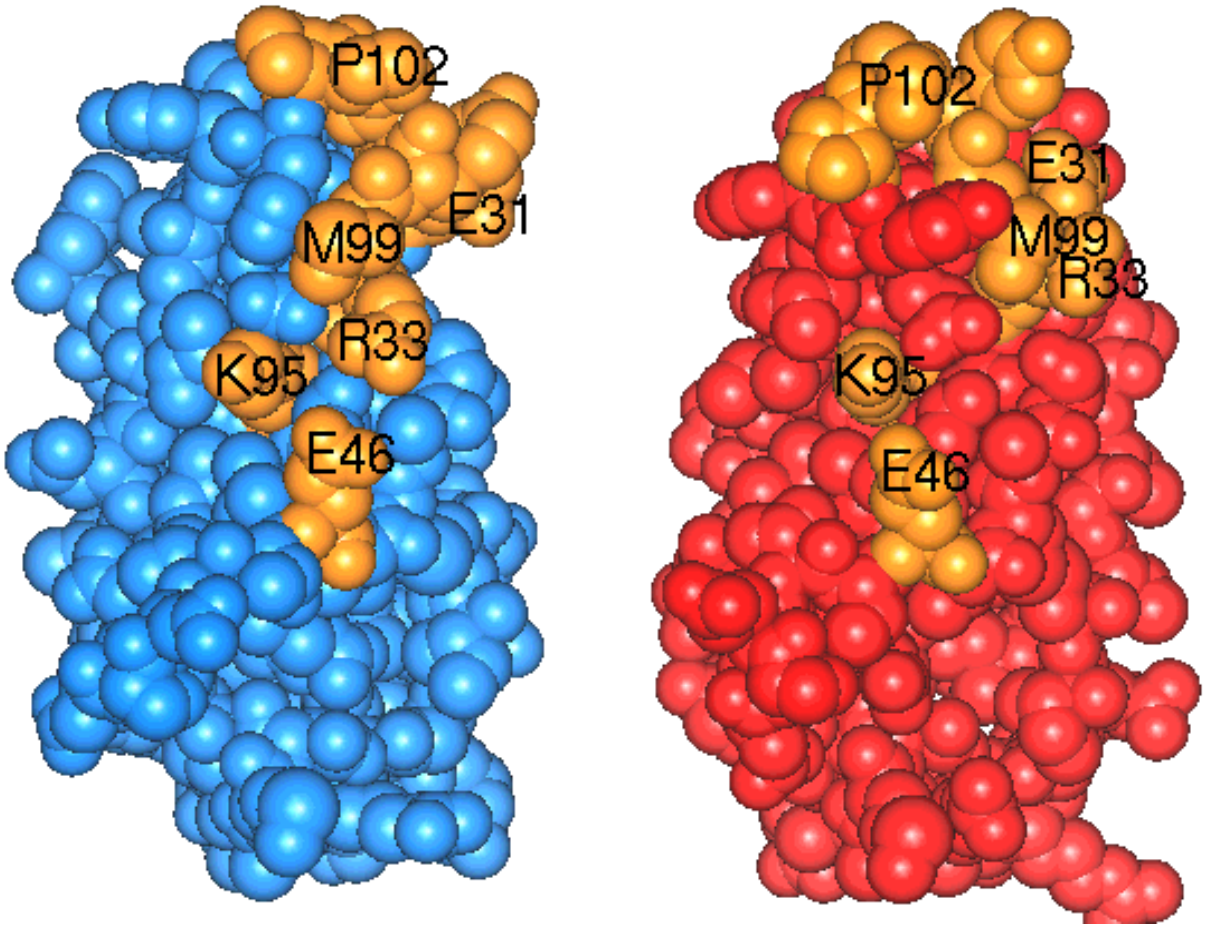


Figure 5. *The predicted (red) and experimentally determined (blue) CD80/CD86 binding sites in CTLA-4 are shown in space-filling representation. The view focuses on the A'GFCC' β-sheet surface. Residues important for the binding of both CD80 and CD86 (including the 99MYPPPY104 motif) are shown in gold.*

obtained alignment is perhaps the most critical determinant of the overall quality of a model, since alignment errors can not be compensated for later on in the modeling process. However, the comparison also emphasizes that modeling of (longer) loops remains a major problem in our IgSF structure predictions. In addition, rigid body shifts of secondary structure elements, as discussed above, limit the accuracy of the model. Detailed structure predictions in the presence of 20% or less sequence identity are an exception in the comparative modeling arena [15,17]. Therefore, to put the overall accuracy of our prediction in context, more frequent structure predictions in the range of 30%-40% sequence identity may be considered. In these cases, backbone rmsd errors of ~1.5Å are typically observed, provided siginificant alignment errors are absent, when ~80% of the residues are included in the comparison [17]. Our CTLA-4 model was sufficiently accurate to outline the CD80/CD86 binding site before an experimentally determined structure was available. Thus, despite its limitations, CTLA-4 modeling made it possible to recognize some important features of the 3D structure, although the sequence identity was lower than usually considered suitable for reliable model building. The present study supports our view that the model building approach originally described for CTLA-4 [3] and subsequently applied to CD86 [18] should make it possible to generate reasonable 3D models for other IgSF proteins.

References

1. Linsley, P. S.; Ledbetter, J. A. *Ann. Rev. Immunol*. **1993**, *11*, 191.
2. Linsley, P. S.; Ledbetter, J.; Peach, R.; Bajorath, J. *Res. Immunol.* **1995**, *146*, 130.
3. Bajorath, J.; Linsley, P. S. *J. Mol. Model* **1997**, *3*, 117.
4. 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.
5. Laskowski, R. A.; MacArthur, M. W.; Moss, D. S.; Thornton, J. M. *J. Appl. Cryst.* **1993**, *26*, 283.
6. Williams, A. F.; Barclay, A. N. *Ann. Rev. Immunol*. **1988**, *6*, 381.

7. Bork, P.; Holm, L.; Sander, C. *J. Mol. Biol.* **1994**, *242*, 309.
8. Epp, O.; Lattman, E. E.; Schiffer, M.; Huber, R.; Palm, W. *Biochemistry* **1975**, *14*, 4943.
9. Sippl, M. J. *Proteins: Structure, Function, and Genetics* **1993**, *17*, 355.
10. Bruccoleri, R. E.; Novotny, J. *Immunomethods* **1992**, *1*, 96.
11. Bajorath, J.; Fine, R. M. *Immunomethods* **1992**, *1*, 137.
12. Peach, R. J.; Bajorath, J.; Brady, W.; Leytze, G.; Greene, J.; Naemura, J.; Linsley, P. S. *J. Exp. Med.* **1994**, *180*, 2049.
13. Chothia, C; Lesk, A. M. *EMBO J.* **1986**, *5*, 823.
14. Bajorath, J.; Harris, L. J.; Novotny, J. *J. Biol. Chem.* **1995**, *270*, 22081.
15. Bajorath, J.; Stenkamp, R.; Aruffo, A. *Protein Science* **1993**, *2*, 1798.
16. Mosimann, S.; Meleshko, R.; James, M. N. G. *Proteins: Structure, Function, and Genetics* **1995**, *23*, 301.
17. Sanchez, R.; Sali, A. *Curr. Opin. Struct. Biol.* **1997**, *7*, 206.
18. Bajorath, J. *J. Mol. Model.* **1997**, *3*, 216.