Journal of Computer-Aided Molecular Design, 11 (1997) 3–8
ESCOM

J-CAMD 379

Prediction of the three-dimensional structure of the human Fas receptor by comparative molecular modeling

Jürgen Bajorath* and Alejandro Aruffo

Bristol-Myers Squibb Pharmaceutical Research Institute, 3005 First Avenue, Seattle, WA 98121, U.S.A., and Department of Biological Structure, University of Washington, Seattle, WA 98195, U.S.A.

Received 23 July 1996
Accepted 5 September 1996

Keywords: Protein structure prediction; Comparative modeling; TNFR superfamily; Cell surface receptor; Receptor–ligand interaction

Summary

The Fas antigen, a cell surface receptor belonging to the tumor necrosis factor receptor (TNFR) superfamily, triggers programmed cell death (apoptosis) in the immune system. The three-dimensional structure of Fas and molecular details of the interaction between Fas and its ligand are currently unknown. A three-dimensional model of the Fas extracellular region was generated by comparative modeling. Inverse folding analysis suggested good sequence–structure compatibility of the model and thus reasonable accuracy. The model was analyzed in the light of information provided by studies on TNFR and CD40, another member of the TNFR family, and the Fas ligand binding site was predicted.

Introduction

The Fas antigen is a type I transmembrane protein which is expressed on a variety of cell lines and highly expressed on activated lymphocytes [1]. In the immune system, Fas plays a major role in programmed cell death (apoptosis). The cytoplasmic region of Fas contains a death domain of ~80 residues which is capable of transmitting apoptotic signals following specific engagement of Fas on the cell surface [2]. The Fas ligand is predominantly expressed on activated T cells [3]. Cytotoxic T cells expressing the Fas ligand can specifically recognize cells expressing the Fas antigen. This interaction leads to signal transduction and targeted cell death in the immune system [2].

Fas and its ligand, like the B cell receptor CD40 and the CD40 ligand [4], are members of the tumor necrosis factor receptor (TNFR) and tumor necrosis factor (TNF) family, respectively [5,6]. The structural similarity of the CD40 ligand and TNF was recently confirmed by X-ray crystallography [7]. The prototypic TNFR fold consists of three to four homologous domains which are stabilized by canonical disulfide bonds in a ladder-like arrangement [8]. Other members of the TNFR superfamily, which typically share a sequence similarity of ~30% or less, are thought to display a similar protein architecture [5]. However, TNFR is currently the only member of this protein superfamily for which X-ray structures have been reported [8,9].

Mutagenesis studies on Fas have focused on its cytoplasmic region [2]. Naturally occurring mutations in Fas which impair apoptosis and are associated with lymphoproliferative syndrome in humans were found to map to the intracellular region of Fas or compromise the cell surface expression of the molecule [10,11]. Thus, the Fas ligand binding site has not yet been identified.

In this study, we have combined structure-oriented sequence comparison and comparative modeling techniques to generate a molecular model of the extracellular region of human Fas. Sequence–structure compatibility analysis of the model suggested that its quality is sound. The model was analyzed and the Fas ligand binding site was predicted, taking into account information provided by crystallographic and mutagenesis studies on TNFR and CD40. The structure prediction may aid in the design of experiments to better understand Fas and its interaction with ligand.

*To whom correspondence should be addressed.

Materials and Methods

An alignment of TNFR, CD40, and Fas sequences from different species was generated using the PileUp [12] and Pretty programs as implemented in the GCG program suite (Genetics Computer Group Inc., Madison, WI, U.S.A.) and modified manually to adjust the positions of cysteines and other structurally significant residues in TNFR [8]. Computer graphics and energy minimization calculations were carried out using the INSIGHTII (v. 95.0) and DISCOVER programs (MSI, San Diego, CA, U.S.A.) and the AMBER force field [13]. For model building of Fas, the structure of uncomplexed TNFR [9] was used as the template. Side-chain replacements were modeled using an automated low-energy rotamer search procedure [14]. In these calculations, a standard rotamer library was used and a cutoff distance of 8 Å was applied for the treatment of nonbonded interactions. The conformations of five regions in human Fas (see the Results section) which included insertions/deletions or were considered structurally variable were modeled by systematic conformational search using CONGEN [15]. The conformational space available to the selected residue segments was searched in main-chain torsion angle increments of 30° and by iterative side-chain conformational search. For each search, loop conformations with acceptable potential energy [15] were retained, and the conformation with the lowest solvent-accessible surface within 2 kcal/mol of the energy minimum was selected and included in the model. Two of the five CONGEN-modeled regions included cysteine residues involved in the formation of putative disulfide bonds (Cys^{43}–Cys^{57}, Cys^{119}–Cys^{124}) not present in TNFR. The conformations of these two regions were approximated by partial conformational search including the predicted disulfide bond constraints. The initially assembled Fas model was energy minimized until the root mean square (rms) derivative of the energy function was less than 1 kcal/(mol Å). The stereochemistry of the model was analyzed with PROCHECK [16] and its sequence–structure compatibility was assessed by energy profile analysis using PROSAII (v. 3.0) [17]. Pairwise residue interaction energies were calculated using β-carbon interactions, and a 50-residue window was used for energy averaging at each residue position. Color figures were generated with INSIGHTII.

Results

Figure 1A shows the multiple structure-oriented sequence alignment of the TNFR, CD40, and Fas extracellular regions. These sequences share a limited similarity of ~30%. The majority of extracellular region residues in Fas is part of the TNFR-homologous region, and three TNFR-homologous repeat domains (D1–D3) can clearly be assigned. Conserved cysteine positions and other structurally significant or constrained positions of the TNFR fold, as revealed by X-ray crystallography [8], were used to align the sequences and to determine the positions of insertions and deletions. All the insertions and deletions map to loop regions in the TNFR structure. Figure 1B shows the structural outline of a TNFR repeat domain and the positions of the disulfide bonds and other structurally important residues used in the sequence alignment.

Based on the alignment, the three-dimensional model was constructed using the X-ray structure of uncomplexed TNFR as the template. The backbone conformations of five segments (37–44, 70–74, 92–96, 119–126, 134–139) were approximated by conformational search as described in the Materials and Methods section and included in the model. Intramolecular contacts and stereochemistry of the initially assembled model, including Fas residues 30–150, were refined by energy minimization. The refined model was subjected to energy profile analysis to assess its sequence–structure compatibility. Figure 2 shows the energy profile of the model compared to the energy profile of the TNFR X-ray structure. The calculated average pairwise residue interaction energy values are negative at each residue position in the Fas model and the TNFR structure. Under the averaging conditions used in these calculations (see the Materials and Methods section), this energy profile trace is characteristic of a correctly folded structure [17] with no substantial topological errors.

Figure 3 shows the Fas model and illustrates its domain structure. In D2 and D3, all TNFR cysteines are conserved in Fas. Like CD40, Fas includes an additional pair of cysteines in a D3 loop region which are predicted to form a disulfide bond. In D1, two cysteines involved in the formation of different disulfide bonds are not conserved in human Fas, but are present in murine Fas. The spatial proximity of the affected disulfide bonds in the TNFR X-ray structure suggests that the formation of a disulfide bond between unpaired cysteines 37 and 44 is possible in human Fas. The conformation of this region including the disulfide bond was approximated by conformational search calculations. In human Fas, D2 and D3 each include a potential N-linked glycosylation site. The D2 glycosylation site is conserved in murine Fas.

Residues which substantially contribute to the TNFR–TNF [8] and the CD40–ligand interactions [18] were mapped on corresponding positions in the Fas model. These residues cluster on the surface of D2 and on the upper part of D3 and form a coherent region. This region of the model includes eight residues (Lys^{78}, Lys^{84}, Arg^{86}, Arg^{87}, Leu^{90}, Glu^{93}, His^{95}, His^{126}) which are conserved in murine and human Fas, but not in TNFR and CD40 (Fig. 4). Seven of these residues are located in D2 and only one (His^{126}) is in D3. Six residues (except for Arg^{86}

and His[95]) map to positions which correspond to residues implicated in the TNFR– and/or CD40–ligand interactions. These findings are significant since residues conserved in Fas, but not in TNFR and CD40, are likely to determine the Fas specificity.

Discussion

The presence of TNFR signature residues in Fas and CD40 supports the view that the TNFR fold represents a prototype for the three-dimensional structures of other

A

```
Fas  Human      RLSSKSVNAQVTDINSKGLELRKTVTTVE
Fas  Mouse        SQLRVHTQGTNSISESLKLRRRVHETD
                 1         10         20

 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
             30            40              50            60
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
              70        80          90         100          110
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
                       120          130          140          150
CD40 Human     CEEGW HCTSEA..CESCVL HRS CSPGFGVKQI  ATGVS DTICE
CD40 Mouse     CKEGQ HCTSKD..CEACAQ HTP CIPGFGVMEM  ATETT DTVCH
                            #


Fas  Human   EEGSRSN
Fas  Mouse   KQSPRNR
                157
```

Fig. 1. Structure-oriented sequence comparison of Fas with TNFR and CD40 proteins from different species. (A) The alignment of the sequences is shown. The extracellular region sequence of Fas includes three TNFR repeat domains (domains 1, 2, and 3). Sequence numbers are given for mature human Fas. The sequences were aligned relative to residue positions which are important for the integrity of the TNFR fold (labeled with asterisks). Gaps were introduced to ensure that corresponding structurally important residues in all three domains are in the same columns. Dots indicate the positions of deletions in the compared sequences. Cysteines involved in the formation of the canonical TNFR disulfide bonds 1, 2, and 3 are labeled with DS1, DS2, and DS3, respectively. TNFR residues which bury 40 Å^2 or more solvent-accessible surface upon formation of the crystallographic TNFR–TNF complex are labeled with '&'. Residues in human CD40 which are, on the basis of mutagenesis studies, important for the interaction with the CD40 ligand are labeled with '#'. Possible N-linked glycosylation sites in human Fas have an exclamation mark. (B) A Richardson-type [21] stereoview of TNFR domain 3 (PDB entry '1ncf'). Secondary structure elements are defined according to Kabsch and Sander [22]. Turns and extended loop regions are shown as tubes and β-strands as green arrows. The N- and C-termini of the domain are labeled. Structurally important residues according to A are shown. Disulfide bonds 1, 2, and 3 are colored blue, pink, and magenta, respectively. The core tyrosine (corresponding to residue Phe^{118} in human Fas) is shown in white, and the serine (corresponding to Thr^{131} in human Fas), which forms a conserved hydrogen bond, is shown in gold. The conserved asparagine and threonine (Asn^{146} and Thr^{147} in Fas) residues are colored yellow and red, respectively.

B

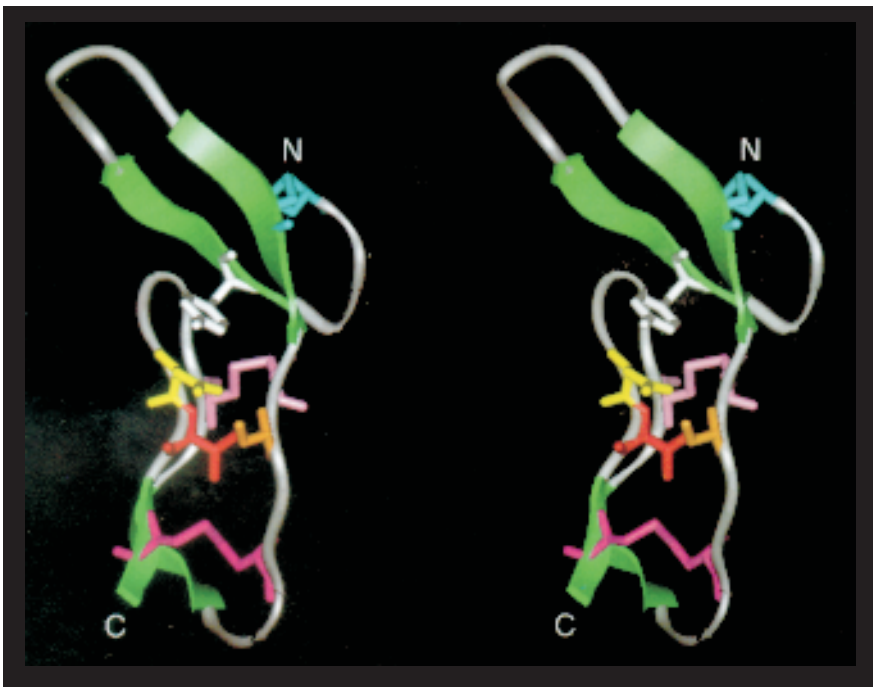

Fig. 1. (continued).

members of this protein superfamily. The alignment of the sequences with emphasis on these positions reflects topological correspondence and provided a meaningful basis for the comparative modeling [19] of human Fas. Sequence–structure compatibility analysis is not expected to detect some local errors, for example, in the modeling of loop or side-chain conformations. However, good sequence–structure compatibility, as obtained for the Fas model, ensures that the model is overall correctly folded and sufficiently accurate to analyze some details. Such

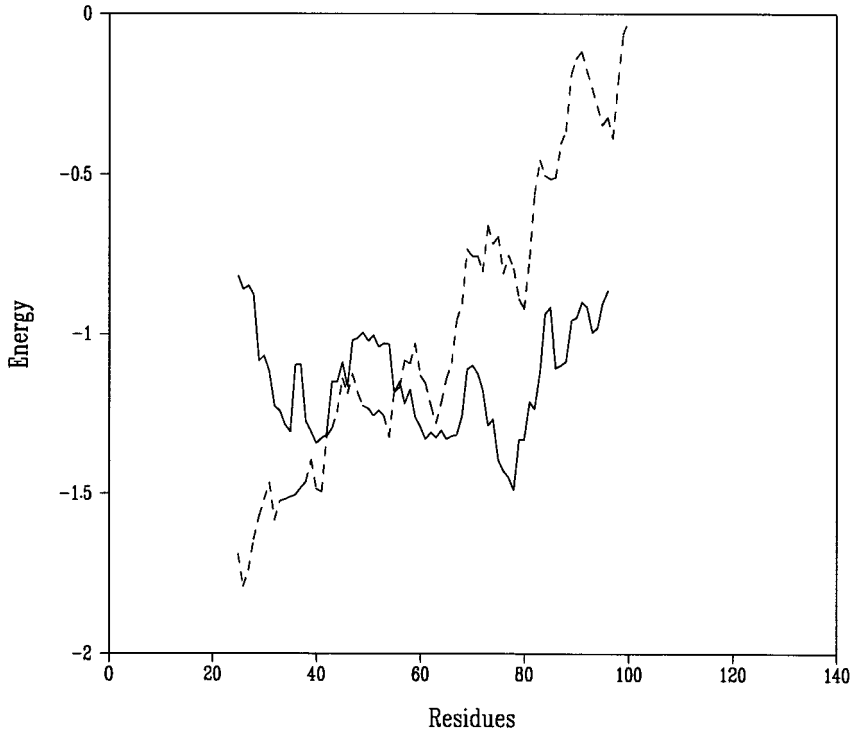

Fig. 2. Energy profile of the Fas model. The pairwise residue interaction energy is given as E/kT (E: interaction energy in kcal/mol; k: Boltzmann constant; T: absolute temperature in K). A 50-residue window was used for energy averaging at each residue position. The energy profile of the Fas model (solid line) is compared with the corresponding profile of domains 1–3 of TNFR (dashed line).

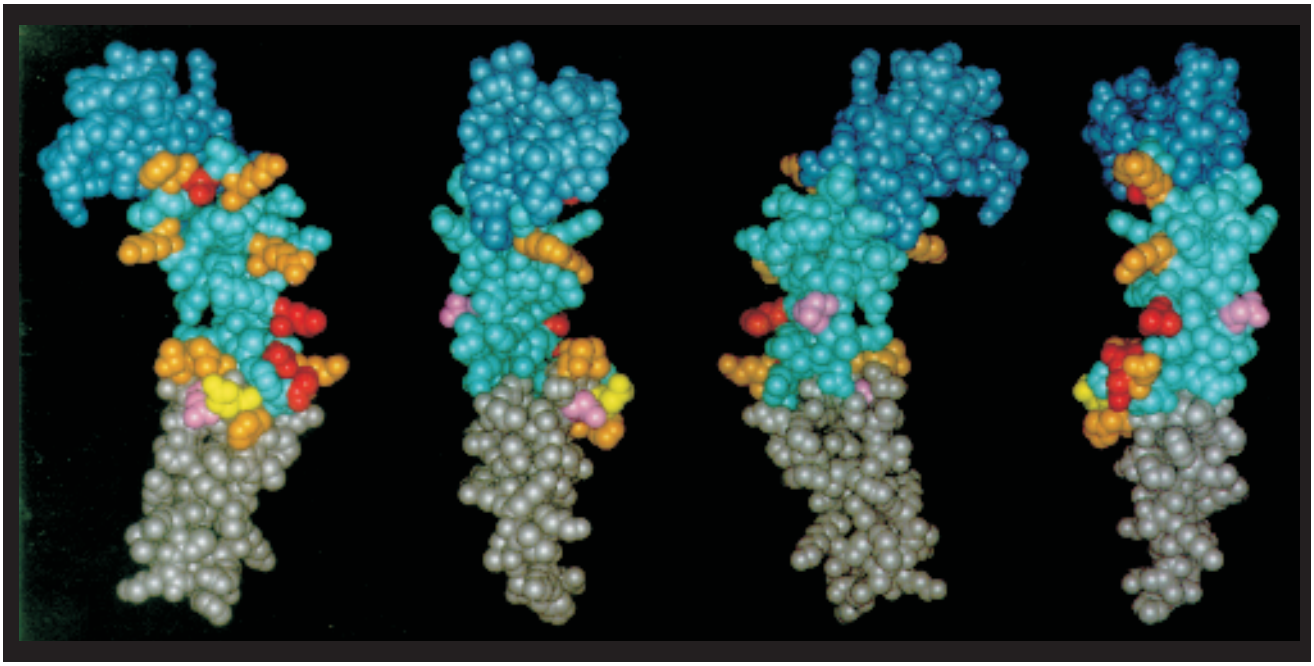

Fig. 3. Model of the extracellular region in human Fas. A space-filling representation is shown with domains D1, D2, and D3 colored in blue, light blue, and gray, respectively. From the left to the right, images were obtained by a subsequent rotation of 90° around the vertical axis to provide a complete view of the model. N-linked glycosylation sites are shown in pink. The positions of residues which are important for the TNFR– and CD40–ligand interactions were mapped on the corresponding positions of the model and colored in gold and yellow, respectively. Red indicates that the residues map to the same position in Fas.

model assessment is particularly important for modeling in the presence of limited sequence similarity [19].

TNFR–TNF and CD40–ligand interactions have been analyzed in some detail by mutagenesis of TNF [20], X-ray analysis of a TNFR–TNF complex [8], and mutational analysis of CD40–ligand interactions [18]. Residues

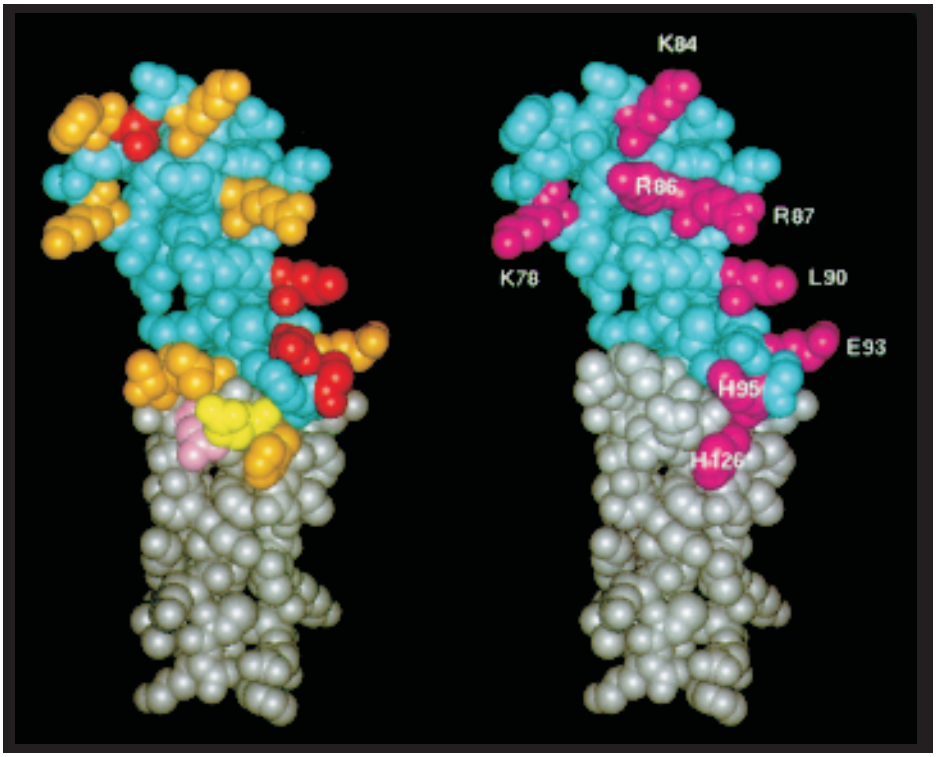

Fig. 4. A close-up view of D2 and D3. The left image is color-coded as in Fig. 3. On the right, residues in this region of the model which are conserved in human and murine Fas, but are different in TNFR and CD40 (see text), are colored magenta and labeled.

important for these interactions map to regions corresponding to D2 and D3 in TNFR and CD40. Therefore, since TNFR, CD40, and Fas belong to the same receptor family, these proteins can be expected to utilize corresponding regions for ligand binding.

However, the different receptor–ligand specificities must result from contributions of residues which are not conserved across the TNFR family. Therefore, D2 and D3 of the Fas model were screened for residues which are conserved in Fas, but different in TNFR and CD40. Eight such residues were identified. Seven of these residues map to D2, and six correspond to positions implicated in TNFR– and/or CD40–ligand interactions. All eight residues map to the same site of the model and are spatially adjacent. Thus, their predicted spatial arrangement provides an outline of a suitable Fas ligand binding site.

Conclusions

Topological sequence comparison and comparative modeling techniques were used to generate a molecular model of the extracellular TNFR-homologous region of Fas, a human immune cell surface receptor. The model provides a first approximate view of the Fas structure. Analysis of the model has allowed the prediction of the Fas ligand binding site.

References

1 Itoh, N., Yonehara, S., Ishii, A., Yonehara, M., Mizushima, S.-I., Sameshima, M., Hase, A., Seto, Y. and Nagata, S., Cell, 66 (1991) 233.

2 Nagata, S. and Golstein, P., Science, 267 (1995) 1449.

3 Lynch, D.H., Watson, M.L., Alderson, M.R., Baum, P.R., Miller, R.E., Tough, T., Gibson, M., Davis-Smith, T., Smith, C.A., Hunter, K., Bhat, D., Din, W., Goodwin, R.G. and Seldin, M.F., Immunity, 1 (1994) 131.

4 Foy, T.M., Aruffo, A., Bajorath, J., Buhlmann, J.A. and Noelle, R.J., Annu. Rev. Immunol., 14 (1996) 591.

5 Beutler, B. and Van Huffel, C., Science, 264 (1994) 667.

6 Smith, C.A., Farrah, T. and Goodwin, R.G., Cell, 76 (1994) 959.

7 Karpusas, M., Hsu, Y.-M., Wang, J.-H., Thompson, J., Lederman, S., Chess, L. and Thomas, D., Structure, 3 (1995) 1031.

8 Banner, D.W., D'Arcy, A., Janes, W., Gentz, R., Schoenfeld, H.-J., Broger, C., Loetscher, H. and Lesslauer, W., Cell, 73 (1993) 431.

9 Naismith, J.H., Devine, T.Q., Brandhuber, B.J. and Sprang, S.R., J. Biol. Chem., 270 (1995) 13303.

10 Fisher, G.H., Rosenberg, F.J., Straus, S.E., Dale, J.K., Middelton, L.A., Lin, A.Y., Strober, W., Lenardo, M.J. and Puck, J.M., Cell, 81 (1995) 935.

11 Rieux-Laucat, F., Le Deist, F., Hivroz, C., Roberts, I.A.G., Debatin, K.M., Fischer, A. and De Villartay, J.P., Science, 268 (1995) 1347.

12 Feng, D.-F. and Doolittle, R.F., J. Mol. Evol., 25 (1987) 351.

13 Seibel, G., Singh, U.C., Weiner, P.K., Caldwell, J. and Kollman, P., AMBER 3.0 revision A, University of California at San Francisco, CA, U.S.A., 1990.

14 Bajorath, J. and Fine, R.M., Immunomethods, 1 (1992) 137.

15 Bruccoleri, R.E. and Novotny, J., Immunomethods, 1 (1992) 96.

16 Laskowski, R.A., MacArthur, M.W., Moss, D.S. and Thornton, J.M., J. Appl. Crystallogr., 26 (1993) 283.

17 Sippl, M.J., Proteins, 17 (1993) 355.

18 Bajorath, J., Marken, J.S., Chalupny, N.J., Spoon, T.L., Siadak, A.W., Gordon, M., Noelle, R.J., Hollenbaugh, D. and Aruffo, A., Biochemistry, 34 (1995) 9884.

19 Bajorath, J., Stenkamp, R. and Aruffo, A., Protein Sci., 2 (1993) 1798.

20 Van Ostade, X., Tavernier, J. and Fiers, W., Protein Eng., 7 (1994) 5.

21 Richardson, J.S., Methods Enzymol., 115 (1985) 359.

22 Kabsch, W. and Sander, C., Biopolymers, 22 (1983) 2577.