Protein Engineering vol.10 no.2 pp.159–167, 1997

Protein engineering with monomeric triosephosphate isomerase (monoTIM): the modelling and structure verification of a seven-residue loop

N.Thanki, J.Ph.Zeelen, M.Mathieu, R.Jaenicke[1], R.A.Abagyan[2], R.K.Wierenga[3] and W.Schliebs[3]

EMBL, Postfach 102209, D69012 Heidelberg, [1]Institut für Biophysik und Physikalische Biochemie, Universität Regensburg, Dg3040 Regensburg, Germany and [2]The Skirball Institute for Biomolecular Medicine, New York University, New York, NY 10016, USA

[3]To whom correspondence should be addressed

Protein engineering experiments have been carried out with loop-1 of monomeric triosephosphate isomerase (monoTIM). Loop-1 of monoTIM is disordered in every crystal structure of liganded monoTIM, but in the wild-type TIM it is a very rigid dimer interface loop. This loop connects the first β-strand with the first α-helix of the TIM-barrel scaffold. The first residue of this loop, Lys13, is a conserved catalytic residue. The protein design studies with loop-1 were aimed at rigidifying this loop such that the Lys13 side chain points in the same direction as seen in wild type. The modelling suggested that the loop should be made one residue shorter. With the modelling package ICM the optimal sequence of a new seven-residue loop-1 was determined and its structure was predicted. The new variant could be expressed and purified and has been characterized. The catalytic activity and stability are very similar to those of monoTIM. The crystal structure (at 2.6 Å resolution) shows that the experimental loop-1 structure agrees well with the modelled loop-1 structure. The direct superposition of the seven loop residues of the modelled and experimental structures results in an r.m.s. difference of 0.5 Å for the 28 main chain atoms. The good agreement between the predicted structure and the crystal structure shows that the described modelling protocol can be used successfully for the reliable prediction of loop structures.

Keywords: loop-design/monoTIM/monomeric triosephosphate isomerase/protein design

Introduction

Protein loops play an important role in molecular recognition. For example, in proteins with the triosephosphate isomerase (TIM)-barrel framework, consisting of eight (βα)-units, eight loops determine the shape of the active site pocket. These active site loops follow immediately after the β-strands of the (βα)-units and are numbered as loop-1 to loop-8, in agreement with the corresponding (βα)-unit. TIM-barrel proteins are known to perform many different enzymatic functions (Reardon and Farber, 1995). Apparently this topology is a good framework for active sites catalysing very different reactions. Mutagenesis experiments have shown that large sequence changes, including insertions and deletions in the active site loops, are allowed without interfering with folding and stability (Urfer and Kirschner, 1992; Borchert *et al.*, 1994). These properties make TIM-barrel proteins ideally suitable for protein engineering experiments, in particular the design of new loops with altered properties.

We have initiated a protein engineering project which is aimed at redesigning the active site loops of TIM. TIM is a dimeric, glycolytic enzyme catalysing the interconversion of dihydroxyacetone phosphate and D-glyceraldehyde-3-phosphate (Knowles, 1991). The catalytic residues of TIM are Lys13 (loop-1), His95 (loop-4) and Glu167 (loop-6), when using the numbering scheme of trypanosomal TIM (Noble *et al.*, 1993). In addition, in wild-type TIM loops-1–4 are involved in tight interactions across the dimer interface; consequently, these loops are very rigid in the wild-type dimer. Our first design experiment was the monomerization of dimeric trypanosomal TIM. This was achieved by replacing the major interface loop (loop-3) with a shorter segment. The modelling was done with ICM (Molsoft LLC, New York; Abagyan *et al.*, 1994). The resulting variant, called monoTIM, is indeed a stable monomeric protein with residual but significant TIM activity. The k_{cat} is 1000-fold reduced compared with wild-type TIM and the K_m is ~10 times higher (Borchert *et al.*, 1994). The crystal structure of monoTIM showed that there is good agreement between the predicted structure of loop-3 and its experimental structure (Borchert *et al.*, 1993). Four different crystal structures of monoTIM have now been determined (Borchert *et al.*, 1995). The different crystal forms were obtained either because the crystallization was done in the presence of another active site ligand or because a surface residue of the original monoTIM was changed by a point mutation. The solution properties of these point mutation variants are the same as observed for the reference monoTIM (Schliebs *et al.*, 1996). An analysis of the four different monoTIM structures has shown that in particular loop-1, including the catalytic lysine, is disordered in monoTIM (Borchert *et al.*, 1995). Nevertheless, site-directed mutagenesis studies of monoTIM have shown that Lys13 is essential for the optimal catalysis by monoTIM (Schliebs *et al.*, 1996). In one of the monoTIM structures, with the substrate analogue 2-phosphoglycollate (2PG) bound in the active site, the Lys13 residue is well defined and adopts a conformation similar to the wild type (Figure 1), but the subsequent residues of loop-1 (residues 14–19) are disordered. The increased flexibility of this loop could be an explanation for the low activity of monoTIM compared with wild type. In order to test this hypothesis, we describe here our attempts to rigidify loop-1, such that the Lys13 side chain points in the same direction as in wild type.

Loop modelling cannot be done with fully automatic procedures (Fetrow and Bryant, 1993). There are two principal approaches, which rely either on extracting loop conformations from structure databases or on conformational search algorithms (Fidelis *et al.*, 1994). Currently, predictions of loop conformations are still far from being reliable (Cardozo *et al.*, 1995). Two major problems are (i) sufficient sampling of the conformational space which becomes problematic for locally

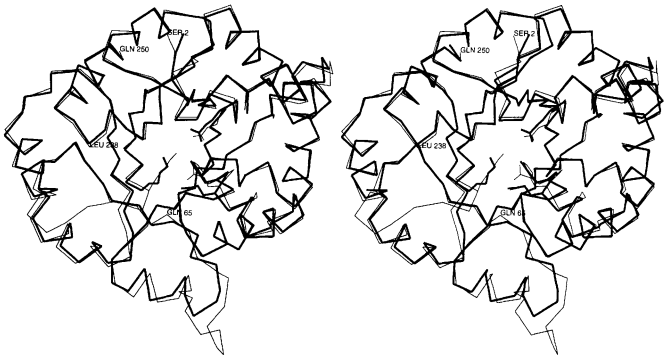

Fig. 1. Superposition of wild-type TIM (thin lines) and monoTIM-W (2PG) (thick lines). The N-terminus and C-terminus are labelled as Ser2 and Glu250, respectively. The side chains of the active site residues Lys13 (loop-1), His95 (loop-4), Glu97 (loop-4) and Glu167 (loop-6) are shown. Gln65 is at the beginning of loop-3 and Leu238 is in loop-8.

	13							20
wild type loop-1 sequence	N	W	K	C	N	G	S	Q	Q	S	L	S	E	L
wild type secondary structure	β							α	α	α	α	α	α	α
disordered in monoTIM-W(2PG)			*	*	*	*	*	*						
ml1TIM loop-1 sequence	N	W	**K**	**-**	**S**	**G**	**S**	**P**	**D**	**S**	L	S	E	L
ml1TIM secondary structure	β							α	α	α	α	α	α	α

Fig. 2. The loop-1 sequence and secondary structures as calculated by DSSP (Kabsch and Sander, 1983), in wild-type TIM and ml1TIM. The disordered residues in the reference molecule [monoTIM-W (2PG)] are indicated by asterisks. The part of loop-1 in ml1TIM which was unfixed in the BPMC calculations is shown in bold (the first and last residues are Lys13 and Ser20, respectively).

deformed regions larger than five residues and (ii) a sufficiently accurate energy function. Our approach, as described here, is an iterative protocol, based on a conformational search algorithm, as implemented in ICM (Abagyan *et al.*, 1994), in combination with a careful analysis of the calculated low-energy conformations of the loop, followed by sequence modifications. The conformational search calculations by ICM consist of a Biased Probability Monte Carlo (BPMC) method (Abagyan and Totrov, 1994) using optimized random moves in combination with a loop closure procedure. This modelling method samples the torsion space of the loop residues and the surrounding side chains. This protocol is a further improvement of the previous loop prediction algorithm which was successfully used in the original design of the monomeric TIM (Borchert *et al.*, 1993). A detailed free energy function including the vacuum energy, electrostatic solvation and the side-chain entropic contribution is now considered (Abagyan and Totrov, 1994).

The end result of our cyclic design procedure is a new sequence with a predicted structure of loop-1. Subsequently this variant, ml1TIM, has been expressed in *Escherichia coli* and purified. Here we describe the design procedure and the characterization of the solution properties of this new variant and its crystal structure at 2.6 Å resolution. Comparison of the modelled and experimental structures of loop-1 show that there is good agreement between these structures.

Materials and methods

The loop design protocol with ICM

The iterative loop design procedure includes several sequence modifications and loop simulations. The following steps can be identified. (i) Assign a conformation and sequence to the loop. (ii) Run a BPMC-loop simulation. In a complete BPMC run ~10^6 conformations are sampled. Accepted conformations are saved in a Monte Carlo trajectory. Several low-energy conformations may be obtained, which are saved in increasing order of energy on a stack (Abagyan and Argos, 1992). These stack conformations are examined and compared later. (iii) Analyse the lowest energy conformation for energetic strain, including a cavity analysis. (iv) Assess the flexibility of the loop from the rearrangements possible near the lowest energy conformation by visual inspection of the structural changes which occur in the Monte Carlo trajectories. (v) Based on this analysis, suggest sequence changes and go back to the first step.

The starting model has been derived from monoTIM-W(2PG) (1TTI in the PDB). In this structure of monoTIM, complexed with 2PG (Borchert *et al.*, 1995), Lys13 is well defined but the subsequent loop-1 residues are disordered (Figures 1 and 2). The waters and the 2PG atoms were removed from this model and hydrogen atoms were added. The structure was regularized before starting the modelling calculations. The Monte Carlo simulations were performed at 1000 K for optimal sampling efficiency.

The loop-1 modelling was aimed at rigidifying loop-1 in such a way that the Lys13 side chain would point in the same direction as in the reference structure. As can be seen in Figure 1, loop-1 is in a rather extended conformation in wild type, leading into helix-1. The N-terminus of this helix (starting at residue 18, Figure 2) is rather solvent exposed in monoTIM; the first hydrophobic residue which anchors this helix into the rest of the protein is Leu21. A start conformation of loop-1 was calculated with ICM using the sequence Lys13–Cys14–Asn15–Gly16–Ser17–Pro18–Asp19–Ser20. This differs from wild type at positions 18–19 (Figure 2). In wild type the sequence is Gln18–Gln19; these residues are at the beginning of helix-1 (Figure 2). The N-capping of helix-1 in dimeric wild type is by the side chain of Asp85 of the other subunit. This interaction is missing in monoTIM. According to the definition of Richardson and Richardson (1988), Ser17 is at the N-cap position, being the first residue whose Cα atom is on the helical spiral. Therefore, position 18 is at the N-cap + 1 position. A proline at the N-cap + 1 position is known to favour helix initiation, as is an aspartate at the N-cap + 2 position (Richardson and Richardson, 1988). The sequence Ser17–Pro18–Asp19 is therefore in complete agreement with the residue preferences at the beginning of an α-helix.

In the Monte Carlo simulations, the torsion angles of residues Lys13 to Ser20 were completely free (eight residues), with the following exceptions. The phi(Lys13) dihedral angle was kept fixed at the value observed in the monoTIM-W(2PG) structure. Also, the NZ(Lys13) atom was restrained at the position observed in the monoTIM-W(2PG) structure. The main chain dihedral angles of residues 17–20 were restrained to be in a helical conformation. The torsion angles of the side chains of the residues within a 6 Å shell around the loop residues (13–20) were completely free. The side chain dihedral angles of Glu97 (loop-4) and Leu238 (loop-8) were explicitly unfixed in the calculations. The Glu97 side chain interacts in wild type with Lys13. Leu238 contributes to a hydrophobic cluster in which also two residues of the N-terminus of helix-1 (Leu21 and Leu24) as well as Trp12 of loop-1 participate. The rest of the molecule was kept fixed.

The actual loop-1 modelling protocol can be subdivided into several steps. First, the appropriate length of the loop was

considered. The loop has to cover a distance of 9.2 Å between Cα(Trp12) and Cα(Leu21). After the first BPMC run, the eight-residue loop, consisting mainly of polar residues, had multiple conformations with close energy values, in agreement with the high mobility observed in the crystal structures. Therefore, it was decided to test if the loop could be one residue shorter: Cys14 was deleted and Asn15 was replaced by a glycine. The new sequence of the loop was therefore Lys13–Gly15–Gly16–Ser17–Pro18–Asp19–Ser20. The BPMC run with this seven-residue loop resulted in a stable conformation with good packing, clearly indicating that the loop can be one residue shorter. The next step was to optimize the sequence further. First, we attempted to introduce a hydrophobic side chain (a leucine) at position 16, whereas at the same time Leu24 was changed into an alanine. It was hoped that a low-energy conformation could be found with Leu16 pointing inwards into a hydrophobic pocket created by the sequence change L24A. Other residues participating in this hydrophobic cluster are Trp12, Leu21 and Leu238. However, an analysis of several low-energy structures after a BPMC run with this sequence showed that a good packing could not be achieved. Although it was thought that a hydrophobic anchor would, theoretically, be the best way to stabilize the loop, the absence of a good low-energy model with a well packed hydrophobic anchor suggested that the particular environment of loop-1 was not suitable for this approach. Subsequently, several more BPMC runs were done with the sequence Lys13–(Gly or Ala)15–(Gly or Ala)16–Ser17–Pro18–Asp19–Ser20, in order to test if some of the glycines could be replaced by a residue with a side chain. Eventually, the lowest energy conformation of loop-1 had phi/psi values for Gly15 which were in the region of the Ramachandran plot allowed for non-glycine residues. In this conformation a side chain at this position would be pointing into the solvent. Therefore, Gly15 was replaced by a serine. The final sequence of loop-1 is therefore Lys13–Ser15–Gly16–Ser17–Pro18–Asp19–Ser20. The stability of this loop was subsequently tested by doing some further BPMC runs with the residues 11–22 completely free, except for restraining the main chain dihedrals of residues 18–22 to a helical conformation. For these calculations, as in all previous runs, the side chains (but not the main chains) within a 6 Å shell of the loop residues were also freely rotatable. The lowest energy conformations of loop-1 of this BPMC run were essentially the same as obtained previously. This result provided enough encouragement to make this monoTIM variant (referred to as ml1TIM) and determine its properties. The five lowest energy conformers of the stack were saved for further analysis. The lowest energy conformer, which has been deposited in the PDB (1MTM), is separated by 4 kcal from the next lowest energy conformer. This 1MTM structure is the reference structure for the comparisons with the experimental structure.

Construction of the mutant

Site-directed mutagenesis was performed by the polymerase chain reaction using the overlap-extension procedure (Higuchi, 1990). As a template the plasmid containing the monoTIM-W gene (Schliebs *et al.*, 1996) was used. The oligonucleotides used as internal mutagenic primers were L1-B (5′-**AGAATCCGGGGATCCGCTC**TTCCAGTTGGCTGCTGCG-3′) and L1-A (5′-**GAGCGGATCCCCGGATTCT**TTGTCGGAGCTTATTGAT-3′). The region of overlap containing the sequence for the new loop-1 is in bold. A new *Bam*HI site (underlined) was incorporated to facilitate the isolation of correct clones. As outer PCR primers have been used oligonucleotides corresponding to TIM-sequence (5′-CAAACCTCATTGACACATGAAG-3′) and plasmid pET3a-sequence (5′-CGATGCGTCCGGCGTAGAGGATC-3′), respectively. Amplified DNA fragments carrying the point mutations were digested with *Xba*I and *Kpn*I and subcloned into expression plasmid pTIM (Borchert *et al.*, 1994). The DNA sequence of the ml1TIM gene has been verified by double-strand sequencing (USB kit). The protein was expressed in *E.coli* strain BL21(DE3) as described previously (Schliebs *et al.*, 1996).

Purification and biochemical characterization

A 10 mg amount of pure protein per litre of culture could be obtained following the purification protocol described earlier for other monoTIM-point mutation variants (Schliebs *et al.*, 1996). The purity was assessed by SDS–PAGE. Protein concentrations were estimated with the Bradford reagent using bovine serum albumin as a standard. The assays for measuring TIM activity, steady-state kinetic analysis [using UltraFit (Biosoft, Cambridge, UK) and GraFit (Erithacus, Staines, UK)] and the estimation of thermal stability using CD spectroscopy were carried out as described previously (Schliebs *et al.*, 1996). The sedimentation analysis was performed in a Beckman Spinco Model E analytical ultracentrifuge, equipped with a UV scanning system. The high-speed sedimentation equilibrium experiments (at 16 000 and 24 000 r.p.m.) were done at room temperature and evaluated from ln c versus r^2 plots. Prior to the experiment the protein sample (0.6 mg/ml) was dialysed against a solution of 20 mM triethanolamine (TEA), pH 7.6, containing also 20 mM NaCl, 1 mM reduced dithiothreitol (DTT), 1 mM EDTA and 1 mM sodium azide.

Crystallization and structure determination

Suitable crystallization conditions were found with the hanging drop method after initially screening 48 different conditions (Zeelen *et al.*, 1994). Well diffracting crystals grow reproducibly after 1 week at room temperature (20°C) by mixing 2 µl of protein solution (5 mg/ml in 10 mM TEA–HCl-buffer, 25 mM NaCl, 1 mM DTT, 1 mM EDTA, 1 mM sodium azide, 10 mM 2PG, pH 7.5) with 2 µl of well solution (100 mM Tris–HCl, pH 8.5, 1.0 M Li_2SO_4, 0.7 M ammonium sulphate, 1 mM DTT, 1 mM EDTA, 1 mM sodium azide). The hanging drop is equilibriated with 1 ml of well solution. A dataset (the maximum resolution is 2.6 Å) was collected at station X11 (DESY synchrotron, Hamburg, Germany) on a MAR image plate and processed with DENZO (Gewirth *et al.*, 1995). The data collection statistics are shown in Table I. The space group is $P3$. The data could not be merged in space groups with higher symmetry. Cell dimensions are $a = b = 165.2$ Å, $c = 51.2$ Å. There are six molecules per asymmetric unit, resulting in a V_m of 2.7 Å^3/D. The molecular replacement calculations were done with AMORE (Navaza, 1994), using a monoTIM structure (1TTJ in the PDB) as a search model, after deleting the residues of loop-1 and loop-6. The rotation function shows two peaks clearly above the background. The translation function indicates three positions for each of the two rotation function peaks. The packing is such that the six molecules of the asymmetric unit are assembled into two trimers, with the two local threefold axes parallel to the crystallographic threefold axis (the existence of trimers also agrees with the packing in another (poorly diffracting) crystal form, in which the threefold axis of the same trimer coincides with a crystallo-

Table I. Crystallographic data

Space group	*P*3
Cell dimensions	165.2 Å 165.2 Å 51.2 Å 90.0° 90.0° 120.0°
No. of molecules per asymmetric unit	6
Data collection statistics:	
Observed reflections	124 266
Unique reflections	45 363
Overall *R*-merge[a]	7.2%
Overall completeness	94.6%
Last shell *R*-merge	29.7%
Last shell completeness 2.64–2.60 Å	96.2%
Refinement data statistics:	
No. of protein atoms[b]	1826 (×6)
No. of ligand atoms[b]	9 (×6)
No. of solvent atoms[b]	53
R-factor	23.1%
R-free[c]	24.7%
R.m.s. bond length deviations	0.01 Å
R.m.s. bond angle deviations	1.61°
Average *B*-factor all atoms	35.5 Å^2
Backbone atoms	32.5 Å^2
Side-chain atoms	37.7 Å^2
Ligand (2PG)	29.7 Å^2
Waters	32.8 Å^2

[a]*R*-merge = $(\Sigma_h\Sigma_i|I_{h,i} - <I>_h|/\Sigma_h\Sigma_i|<I>_h|)\times 100$.
[b]The protein and ligand (2PG) atoms were refined with strict n.c.s. The n.c.s. relationships were not imposed on the 53 water molecules.
[c]The *R*-free was calculated with a 5% subset of the data which was never included in any refinement calculations.

graphic threefold axis). Subsequently, refinement of the structure with X-PLOR (Brünger, 1992) was initiated. For all the refinement calculations a subset of 5% of the data was used for *R*-free calculations, to monitor the quality of the refinement protocol. Strict non-crystallographic symmetry (n.c.s.) was enforced during the refinement. The n.c.s. relationships of the six molecules were recalculated at several stages of the refinement by rigid body refinement calculations. The first step of the refinement was the rigid body refinement of the six molecules. Using data between 8 and 3.5 Å, the *R*-factor dropped from 47.9 to 34.0%. At this stage the maps, calculated with CCP4 programs (CCP4, 1979), clearly indicated how to build loop-6 and 2PG in each of the six molecules of the asymmetric unit. In some of the molecules there was also density for loop-1 residues. The residues of loop-6 and the 2PG molecule were included in the model and the refinement calculations (simulated annealing as well as Powell minimization), interleaved with model building sessions, using O (Jones *et al.*, 1991), proceeded to an *R*-factor of 23.7% (*R*-free is 26.1%), at a resolution of 2.9 Å. Subsequently the refinement was completed at 2.6 Å resolution. At this resolution the residues of loop-1 became clearly visible. Incorporating these residues into the model and further refinement calculations, including group B-factor refinement, resulted in the final model (still with strict n.c.s.) with an *R*-factor of 23.1% and *R*-free of 24.7% (Table I). The structure has good geometry (Table I) and the main chain dihedrals of all residues are in the allowed regions of the Ramachandran plot. This structure, referred to as the ml1TIM structure, has been analysed with WHAT IF (Vriend, 1990), O and ICM. The TIM-barrel framework β-strands and α-helices were used to calculate the superpositions shown in Figures 1, 7, 8, 9 and 10. Figures 6, 7 and 10 were obtained with ICM.

Results

Solution properties

The kinetic properties of ml1TIM are given in Table II. The k_{cat} and K_m values are essentially the same as for monoTIM. The competitive inhibition by two substrate analogues (phosphate and 2PG) and the transition state analogue phosphoglycolohydroxamate (PGH) have been measured. K_i (phosphate) and K_i(2PG) are the same as for monoTIM, but K_i(PGH) is lower. K_i(PGH) is 50 mM for monoTIM but 18 mM for ml1TIM. Therefore, ml1TIM has a higher affinity than monoTIM for PGH. Sedimentation equilibrium runs with a solution of ml1TIM showed that the protein is homogeneous and monomeric with a molecular mass of 23.0±0.9 kDa. There is no indication of dimer or trimer formation, even in the bottom of the cell, i.e. at concentrations up to 2.8 mg/ml. This holds despite the fact that it crystallizes as a trimer. Apparently, the trimer formation is induced by the crystallization conditions. The thermal stability of ml1TIM was measured in the absence and presence of 2PG by measuring the CD-signal at 222 nm as a function of temperature (Figure 3). The measured T_m values of ml1TIM are 49 and 56°C in the absence and presence of 1 mM 2PG, respectively. These values are essentially identical with the T_m values of monoTIM. Therefore, the new, shorter loop-1 has not changed the overall stability of monoTIM.

Structural properties

Ml1TIM crystallizes with two trimers per asymmetric unit. The molecules within each trimer are related by a local threefold axis. The crystals diffract to 2.6 Å resolution. The refinement of model was completed at this resolution, using strict n.c.s. There are no indications of structural differences between the six molecules of the asymmetric unit. The crystals were grown in the presence of 10 mM 2PG. The electron density maps clearly indicate the mode of binding of 2PG in the active site of each ml1TIM molecule in the asymmetric unit. In agreement with the presence of 2PG in the active site, it is observed that loops-5–7 of each ml1TIM molecule are in the closed conformation as observed in the other liganded monoTIM (Schliebs *et al.*, 1996) and wild-type TIM structures (Noble *et al.*, 1993).

As described in the Materials and methods section, the loop-1 residues were initially left out of the model. Only after refinement at 2.6 Å resolution of a model with a complete chain tracing (except for loop-1, but including 2PG) loop-1 residues were built into the corresponding electron density. In the final structure the loop-1 conformation is clearly defined, although the *B*-factors are relatively high (Figure 4). Figure 5 shows the superposition of the final structure of loop-1 and omit-density, calculated after a simulated annealing test X-PLOR refinement run with the final coordinate set but without the seven residues of loop-1 (*R*-factor = 23.3%; *R*-free = 25.6%). In the trimer the loop-1 faces the local threefold axis, however the loop-1 residues are not interacting with neighbouring molecules of the trimer: there are no atom–atom contacts of loop-1 atoms with atoms of any other neighbouring molecule within a cut-off distance of 4 Å. The arrangement of loop-1 within the context of the trimeric arrangement is shown in Figure 6. Loop-1 atoms also do not interact with molecules of any neighbouring trimer.

Table II. Kinetic parameters of ml1TIM

	k_{cat} (GAP) (min^{-1})	K_m (GAP) (mM)	K_i (phosphate) (mM)	K_i (PGH) (mM)	K_i (2PG) (mM)
Wild-typeTIM	3.7×10^5	0.25±0.05	6.9±1.1	0.008± 0.001	0.026±0.007
MonoTIM	3.1×10^2	4.1±0.6	14.2±2.1	0.50±0.07	0.052±0.007
ml1TIM	3.0×10^2	5.7±0.6	9.4±2.9	0.18±0.06	0.045±0.003

The measurements concern the conversion of D-glyceraldehyde-3-phosphate into dihydroxyacetone phosphate.

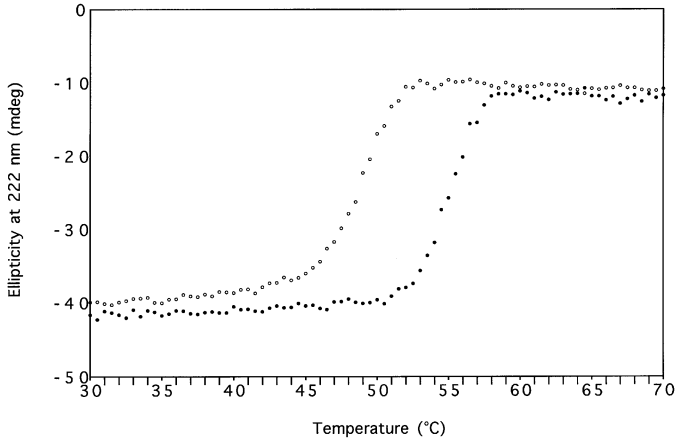

Fig. 3. The temperature dependent denaturation curves for ml1TIM in the absence (○, top row) and presence (●) of 1mM 2PG. The data were obtained by measuring the ellipticity at a wavelength of 222 nm at increasing temperatures with a scan rate of 20°C/h. The protein concentration in the 0.2 cm pathlength cuvette was 0.4 mg/ml in 20 mM 3-(*N*-morpholino)-propanesulphonic acid (MOPS) buffer, pH 7.0, containing also 1 mM DTT, 1 mM EDTA and 1 mM azide.

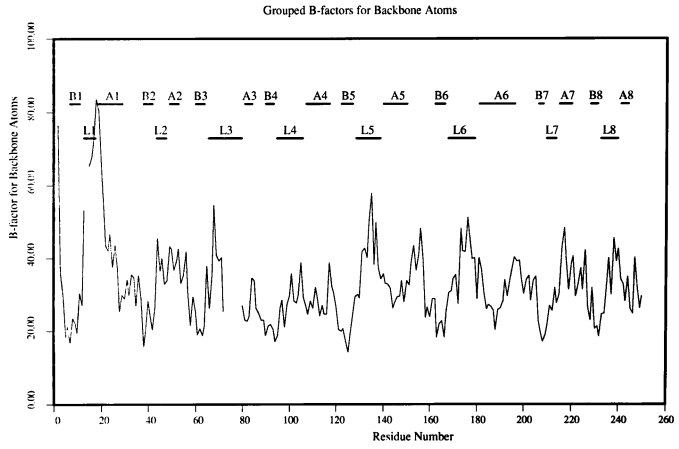

Fig. 4. *B*-factor plot of the main chain *B*-factors (in $Å^2$) of the ml1TIM structure. The discontinuities in the plotted line are due to the discontinuous numbering scheme near loop-1 and loop-3.

Figure 7 shows the superposition of the modelled loop–1 and the experimental loop–1. For the framework superposition the r.m.s. difference for the 28 main chain atoms (including the carbonyl oxygen atoms) is 0.9 Å. This r.m.s. value drops to 0.5 Å when these 28 main chain atoms are directly superimposed. There are no peptide flips between the modelled structure and the crystal structure. As can be seen in Figure 1, loop-8 and loop-2 are spatially close to loop-1. A comparison of the Cα-traces of 1MTM (the reference structure) and ml1TIM (the experimental structure), as depicted in Figures 8 and 9, shows no conformational differences for loop-8. However, loop-2, but also loop-3 and loop-4, adopt somewhat different conformations. Loop-1 is not hydrogen bonded to loop-8, but in both 1MTM and ml1TIM, loop-1 is hydrogen bonded to loop-2 and loop-3. Specifically, O(Trp12) interacts with NE2(Gln65) of loop-3 and O(Gly16) is hydrogen bonded to NE2(His47) of loop-2. In the crystal structure, Gln65 and His47 have moved towards loop-1 and, as shown in Figure 10, the loop-1 residues Lys13–Ser15–Gly16 of loop-1 have shifted in a concerted fashion with Gln65 and His47. This movement of loop-1 towards Leu238 (~1 Å near Ser15) is correlated with differences in main chain dihedrals (Table III), which cause the O(Ser15) to point inwards in ml1TIM. In the model, O(Ser15) points along the surface and it contacts Leu238 (loop-8) at van der Waals distance. Leu238 is in the same position in 1MTM as in ml1TIM. In the crystal structure, O(Ser15) is rotated inwards to avoid clashes with the Leu238 side chain. A simulated annealing refinement test run (by X-PLOR) was done to confirm this interpretation of the map. For this purpose the peptide plane of Ser15 was flipped and subsequently the fragment Lys13–Ser15–Gly16 was optimally fitted into the electron density map. The simulated annealing test run with this model resulted in a structure almost identical with the refined model, confirming the position and orientation of the Ser15 atoms.

Discussion

Structure of the new loop-1

The BPMC procedure, as implemented in ICM, was used to model a new loop-1 in monoTIM. Eventually, the design resulted in a new loop-1 sequence which is one residue shorter than the wild-type sequence. The suggested sequence consists of polar residues, without inward pointing hydrophobic residues which could have anchored the loop to the core of the protein. In the modelling calculations the structure of this seven-residue loop-1 (Lys13–Ser15–Gly16–Ser17–Pro18–Asp19–Ser20) was optimized. Subsequently, the new loop-1 sequence was introduced in monoTIM-W using site-directed mutagenesis. The new variant (ml1TIM) was purified and could be crystallized in the presence of 2PG. The modelled loop-1 has been derived from a monoTIM crystal structure [monoTIM-W(2PG)], which was also crystallized in the presence of 2PG. In both of these experimental structures the 2PG is bound in the active site and the Lys13 side chains are well defined and interact with the carboxyl moiety of 2PG. In the monoTIM-W(2PG) structure the residues after Lys13 are disordered but in the ml1TIM structure they have adopted a defined conformation, as aimed for in the modelling protocol. In ml1TIM loop-1 is not involved in crystal contacts, therefore the observed structure of loop-1 is only determined by interactions with neighbouring residues of the same molecule. The agreement between the predicted structure and the modelled structure of loop-1 is very good. The *B*-factors of loop-1 are relatively high (Figure 4). The highest loop-1 *B*-factors are for the residues forming the N-terminus of helix-1 (Table III), despite the fact that the optimal

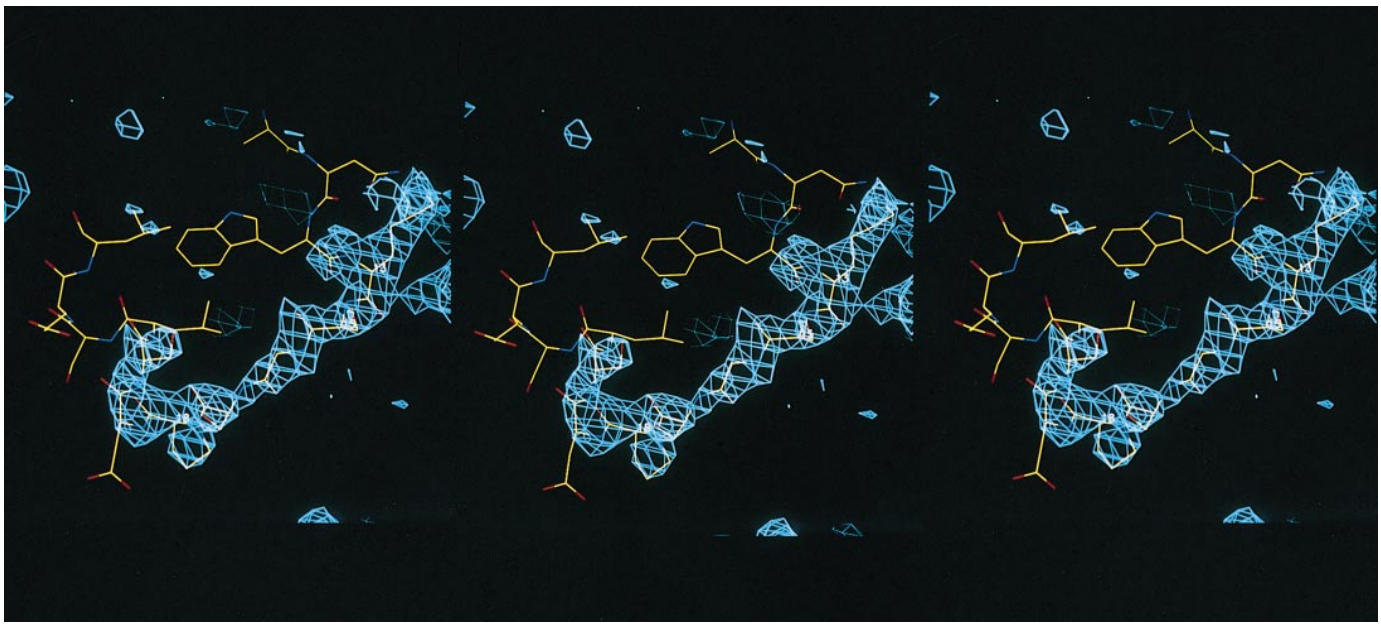

Fig. 5. Loop-1 superimposed on an $(F_o - F_c)\alpha_c$ difference map. The map was calculated from a model obtained after a simulated annealing refinement X-PLOR run in which the seven loop-1 residues had been deleted from the model. The map, calculated with all data between 30 and 2.6 Å, is contoured at 2.2 sigma.

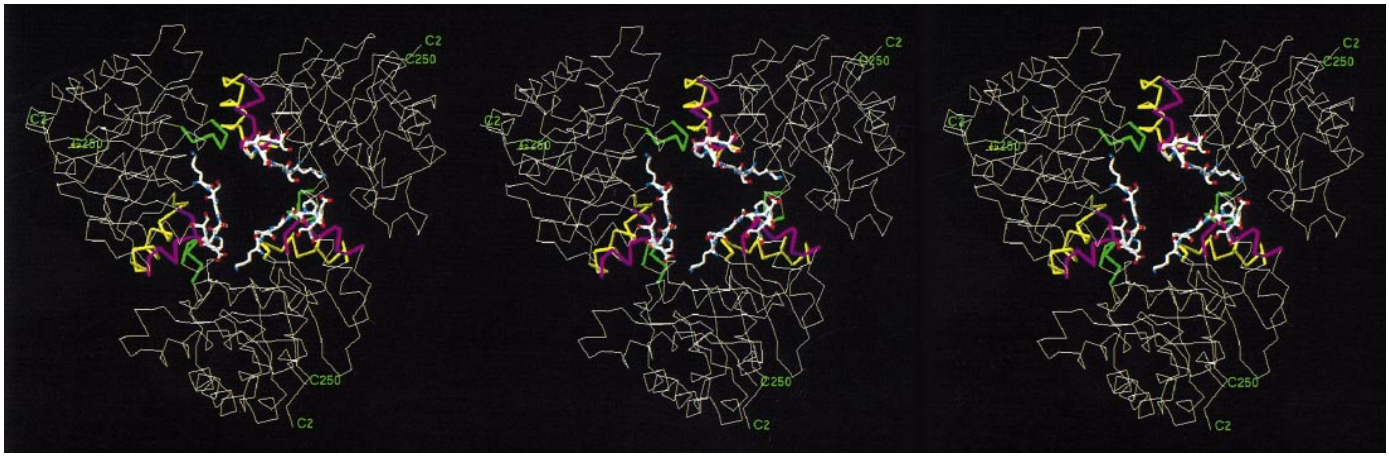

Fig. 6. Stereopicture of the ml1TIM trimer. Loop-1 (in white, red and blue) faces the intra trimer space; loop-2 (and helix-2) is in purple; loop-3 (and helix-3) is in yellow; loop-4 is in green. The N-terminus and C-terminus are labelled C2 and C250, respectively.

sequence (Pro18–Asp19–Ser20) was chosen for these helix-initiating residues. The loop-1 residues with high *B*-factors are from Lys13 to Leu21. This stretch of residues seems to be anchored to the rest of the protein by the hydrophobic side chains of Trp12 and Leu21, as suggested by the observation that the side chain *B*-factors of Trp12 and Leu21 are lower than the *B*-factors of the loop-1 main chain residues (Table III).

Environment of the ml1TIM loop-1

For the modelling of loop-1 it has been assumed that there are no conformational differences in the main chains of neighbouring loops because simultaneous large-scale sampling of several loops still presents a computational challenge. However, as shown in Figures 8 and 9, there are structural differences in loops-2–4 when comparing the crystal structure and the reference structure. An analysis of these differences shows that the loop-2 and loop-3 movements, in particular His47 (loop-2) and Gln65 (loop-3), correlate very well with the 1 Å shift of the Lys13–Ser15–Gly16 fragment (Figure 10). Indeed, a BPMC-run of loop-1 in the context of the ml1TIM structure produces a low-energy structure with the same main chain trace as seen in the ml1TIM structure. The structural differences for loops-2–4 are probably due to crystal contacts. Loop-4 is intimately involved in crystal contacts, due to strong interactions with loop-2 and loop-3 of another molecule of the same trimer (Figure 6). For example, the aromatic rings of Trp100 (loop-4) and Phe86 (loop-3) of two contacting molecules are stacked and the Tyr101 (loop-4) side chain binds in a pocket between loop-2 and loop-3 of the adjacent molecule. As can be seen in Figure 9, the largest movements in loop-4 are at residues Trp100 and Tyr101.

The ml1TIM active site

Previous monoTIM studies have shown that Lys13 and His95 are essential catalytic residues (Schliebs *et al.*, 1996), despite the observation that in some monoTIM structures these residues

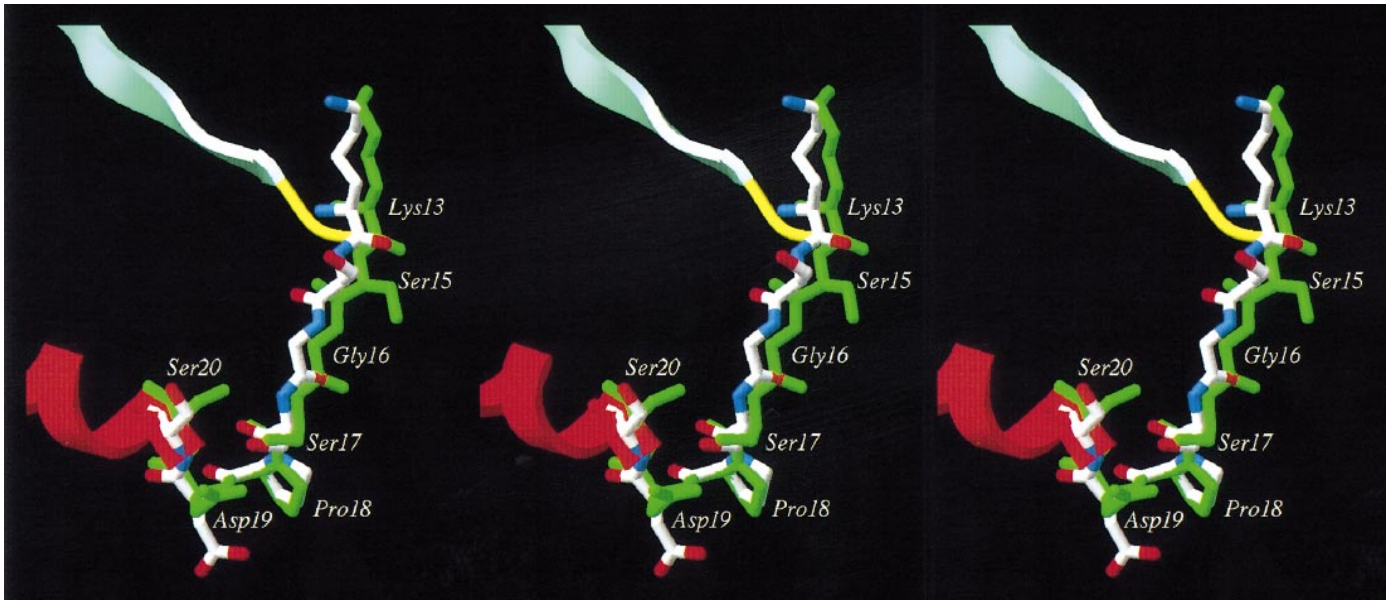

Fig. 7 Comparison of the loop-1 structure of ml1TIM (with white, blue and red) and 1MTM (green) (similar view as in Figure 1).

Fig. 8. Superposition of the Cα-traces of ml1TIM (thick lines) and 1MTM (thin lines). The N-terminus and C-terminus are labelled Ser2 and Glu250, respectively. The side chains of ml1TIM of the active site residues Lys13 (loop-1), His95 (loop-4), Glu97 (loop-4) and Glu167 (loop-6) are shown. Pro18 is at the N-cap + 1 position of helix-1, Gln65 is at the beginning of loop-3 and Leu238 is in loop-8.

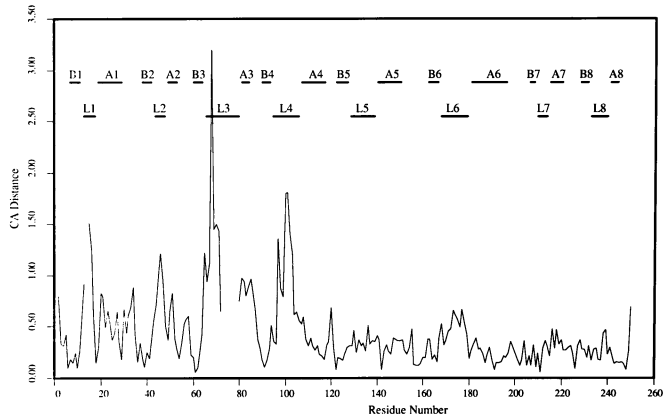

Fig. 9. Cα-distance plot for the comparison of the observed ml1TIM structure and the reference 1MTM structure. The distances are in Å. The discontinuities in the plotted line are due to the discontinuous numbering scheme near loop-1 and loop-3.

adopt conformations which are not compatible for catalysis. Therefore, it has been concluded that in solution, in the presence of substrate an active site geometry is induced which is compatible with catalysis (Schliebs *et al.*, 1996). The catalytic properties of ml1TIM and monoTIM are remarkably similar (Table II), despite the large sequence changes in loop-1. It should be noted that ml1TIM is not more active than monoTIM; apparently the rigidification of loop-1 in ml1TIM is not sufficient to restore wild-type catalytic activity. In ml1TIM the Lys13 side chain is pointing into the active site, in a similar position as in wild type. This is achieved without any strain in the loop-1 main chain conformation, as aimed for in the modelling exercise. Nevertheless, an important interaction of Lys13 is missing in ml1TIM: in wild-type dimeric TIM Lys13 interacts with loop-4 via a conserved salt bridge with Glu97 (loop-4). This salt bridge is not observed in ml1TIM and monoTIM-W(2PG); in fact, in none of the monoTIM structures is this salt bridge observed. Instead, in two monoTIM structures, including ml1TIM, Glu97 is salt bridged to His95 (Figure 8). Such a salt bridge interaction is not compatible with catalysis, because it has been shown that for catalysis a neutral histidine is required (Lodi and Knowles, 1991). Apparently, the side chains of Lys13 and Glu97 are not fixed (in solution) by the rigidification of loop-1 and therefore the catalytic activity of ml1TIM seems also to be due to an induced fit mechanism. Indeed, the modelling was aimed at a rigidification of the main chain of loop-1 and not of its side chains. In wild-type TIM the main chain and side chain conformations of Lys13 and Glu97 are stabilized at the dimer interface. Further protein engineering experiments, aimed at making monoTIM more active, should include fixing not only the main chain of loop-1 but also the side chains of Lys13 and Glu97.

Conclusion

In this protein engineering project we have predicted and verified the structure of a seven-residue loop. There is good agreement between the modelled and crystal structures, as the r.m.s. positional difference for the superposed 28 main chain

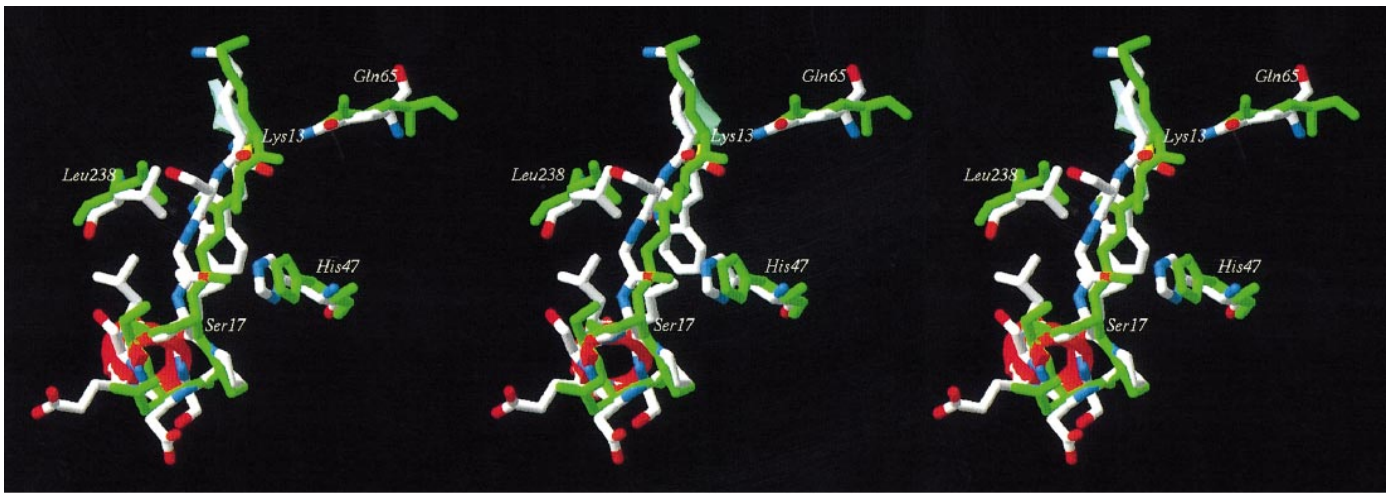

Fig. 10. The concerted shift of the Lys13–Ser15–Gly16 fragment and His47 (loop-2) and Gln65 (loop-3) when comparing ml1TIM and 1MTM (green). The side chains of Trp12 and Leu21 are part of a hydrophobic cluster, together with Leu24 and Leu238. Ser17 is at the N-cap position of helix-1.

Table III. Geometric data of the ml1TIM structure

Residue	Crystal structure (ml1TIM)		Predicted structure 1MTM		Comparison of ml1TIM and 1MTM $\Delta\phi/\Delta\psi$ (°)	Crystal structure (ml1TIM) Group *B*-factor (Å^2)	
	ϕ (°)	ψ (°)	ϕ (°)	ψ (°)		Main chain	Side chain
Asn11	−118.8	87.3	−104.5	104.9	14.3/17.6	30	29
Trp12	−73.4	92.3	93.1	117.1	19.7/24.8	27	24
Lys13	−82.7	−20.0	−82.0	−40.0	0.7/20.0	53	43
Ser15	−156.8	−172.6	−166.5	160.4	9.7/27.0	65	80
Gly16	120.1	162.5	−178.0	−171.1	61.9/26.4	68	68
Ser17	−103.9	147.8	−122.4	149.7	18.5/1.9	72	85
Pro18	−49.0	−41.0	−68.7	−28.7	19.7/12.3	83	88
Asp19	−70.6	−49.8	−69.2	−47.5	1.4/2.3	80	134
Ser20	−59.0	−41.8	−65.6	−36.8	6.6/5.0	66	85
Leu21	−64.6	−38.9	−74.8	−31.5	10.2/7.4	57	34
Ser22	−68.6	−23.8	−75.2	−30.9	6.6/7.1	44	64

atoms is as small as 0.5 Å. There are also several examples of modelling of CDR loops of the antigen-binding site of antibodies, where the predicted loop conformations have been compared with X-ray structures (Bassolino-Klimas *et al.*, 1992; Eigenbrot *et al.*, 1993; Bajorath and Sheriff, 1996). For these modelling studies, conformational search and homology modelling techniques were used. There are important differences between the monoTIM loop-1 modelling and the modelling of the CDR loops. The existence of canonical structures for five of the six CDR-loops should make a successful prediction easier. However, the interplay between the six different loops makes it more complicated to predict the structural details of the antigen binding site architecture correctly. The success of the prediction varies considerably for the different loops. For loops with six or seven residues, the main chain atoms were reported to agree within 0.4 Å (after direct superposition) and 1.1 Å after framework superposition in a recent study (Bajorath and Sheriff, 1996). However, in this case the good agreement for these loops is also due to the existence of canonical structures, because the predicted structures of these loops were directly transferred from the reference structure (Bajorath and Sheriff, 1996). Higher r.m.s. values were reported for a modelling study in which the CDR loop structures were predicted via a conformational sampling algorithm (Bassolino-Klimas *et al.*, 1992).

Another important but even more challenging problem is the correct prediction of loop structures in 'modelling by homology' structure predictions (Aehle *et al.*, 1995). The meeting on the Critical Assessment of Techniques for Protein Structure Prediction held in December 1994 (Mosimann *et al.*, 1995) revealed a gloomy picture: none of the loops in the homology modelling targets were predicted correctly. One of the serious problems with loop prediction in homology modelling, in addition to the sampling and energy accuracy, is backbone deformation of the loop ends and unpredicted structural differences in the loop environment (Cardozo *et al.*, 1995).

In the case of protein engineering experiments, the environment is conserved much better than in a typical case of modelling by homology, since the rest of the protein has exactly the same sequence and is more structurally conserved. Our loop-1 modelling studies show that, in this context, the conformational search techniques as implemented in ICM can predict the loop structure to a high level of accuracy.

Acknowledgements

We thank the staff of the EMBL outstation in Hamburg for help with data collection at the DESY synchrotron. The coordinates and structure factors of

ml1TIM have been deposited at the PDB (1ML1). This work was supported by EC grants BIOT-CT90-0182, CHRX-CT93-0173 and ERBCHGE-CT94-0062.

References

Abagyan, R.A. and Argos,P. (1992) *J. Mol. Biol.*, **225**, 519–532.
Abagyan,R.A. and Totrov,M.M. (1994) *J. Mol. Biol.*, **235**, 983–1002.
Abagyan,R.A., Totrov,M.M. and Kuznetsov,D.N. (1994) *J. Comput. Chem.*, **15**, 488–506.
Aehle,W., Sobek,H. and Schomburg,D. (1995) *J. Biotechnol.*, **41**, 211–219.
Bajorath,J. and Sheriff,S. (1996) *Proteins*, **24**, 152–157.
Bassolino-Klimas,D., Bruccoleri,R.E. and Subramaniam,S. (1992) *Protein Sci.*, **1**, 1465–1476.
Borchert,T.V., Abagyan,R., Radha Kishan,K.V., Zeelen,J.P. and Wierenga,R.K. (1993) *Structure*, **1**, 205–213.
Borchert,T.V., Abagyan,R., Jaenicke,R. and Wierenga,R.K. (1994) *Proc. Natl Acad. Sci. USA*, **91**, 1515–1518.
Borchert,T.V., Kishan,K.R., Zeelen,J.P., Schliebs,W., Thanki,N., Abagyan,R., Jaenicke,R. and Wierenga,R.K. (1995) *Structure*, **3**, 669–679.
Brünger,A. (1992) *X-PLOR Manual*. Yale University Press, New Haven, CT.
Cardozo,T., Totrov,M. and Abagyan,R. (1995) *Proteins*, **23**, 403–414.
CCP4 (1994) *Acta Crystallogr.*, **D50**, 760–763.
Eigenbrot,C., Randal,M., Presta,L., Carter,P. and Kossiakoff,A.A. (1993) *J. Mol. Biol.*, **229**, 969–995.
Fetrow,J.S. and Bryant,S.H. (1993) *Bio/technology*, **11**, 479.
Fidelis,K., Stern,P.S., Bacon,D. and Moult,J. (1994) *Protein Engng*, **7**, 953–960.
Gewirth,D., Otwinowski,Z. and Minor,W. (1995) *The HKL Manual*. Yale University Press, New Haven, CT.
Higuchi,R. (1990) In Innis,M.A., Geldford,D.H., Sninsky,J.J., White,T.J. (eds), *PCR Protocols: A Guide to Methods and Applications*. Academic Press, San Diego, pp.177–183.
Jones,T.A., Zou,J.-Y., Cowan,S.W. and Kjeldgaard,M. (1991) *Acta Crystallogr.*, **A47**, 110–119.
Kabsch,W. and Sander,C. (1983) *Biopolymers*, **22**, 2577–2637.
Knowles,J.R. (1991) *Nature*, **350**, 121–124.
Lodi,P.J. and Knowles,J.R. (1991) *Biochemistry*, **30**, 6948–6956.
Mosimann,S., Meleshko,R. and James,M.N.G. (1995) *Proteins*, **23**, 301–317.
Navaza,J. (1994) *Acta Crystallogr.*, **A50**, 157–163.
Noble,M.E.M., Zeelen,J.P. and Wierenga,R.K. (1993) *Proteins: Struct. Funct. Genet.*, **16**, 311–326.
Reardon,D. and Farber,G.K. (1995) *FASEB J.*, **9**, 497–503.
Richardson,J.S. and Richardson,D.C. (1988) *Science*, **240**, 1648–1652.
Schliebs,W., Thanki,N., Eritja,R. and Wierenga,R.K. (1996) *Protein Sci.*, **5**, 229–239.
Urfer,R. and Kirschner,K. (1992) *Protein Sci.*, **1**, 31–45.
Vriend,G. (1990) *J. Mol. Graphics*, **8**, 52–56.
Zeelen,J.P., Hiltunen,J.K., Ceska,T.A. and Wierenga,R.K. (1994) *Acta Crystallogr.*, **D50**, 443–447.

Received August 6, 1996; revised October 16, 1996; accepted October 24, 1996