Journal of Computer-Aided Molecular Design, **14:** 719–730, 2000.
KLUWER/ESCOM

Construction of a full three-dimensional model of the transpeptidase domain of *Streptococcus pneumoniae* PBP2*x* starting from its C^{α}-atom coordinates

Paul A.M. van Hooft & Hans-Dieter Höltje*
Institut für Pharmazeutische Chemie, Heinrich-Heine Universität Düsseldorf, Universitätsstrasse 1, D-40225 Düsseldorf, Germany

Received 28 September 1999; Accepted 9 June 2000

Key words: homology modeling, molecular dynamics, PBP2*x*, penicillin-binding proteins

Summary

A new method is described for generating all-atom protein structures from C^{α}-atom information. The method, which combines both local structural trace alignments and comparative side chain modeling with *ab initio* side chain modeling, makes use of both the virtual-bond and the dipole-path methods. Provided that 3D structures of structurally and functionally related proteins exist, the method presented here is highly suitable for generating all-atom coordinates of partly solved, low-resolution crystal structures. Particularly the active site region can be modeled accurately with this procedure, which enables investigation of the binding modes of different classes of ligands with molecular dynamics simulations. The method is applied to the trace of *Streptococcus pneumoniae*, in order to construct an all-atom structure of the transpeptidase domain. Since after generation of full coordinates of the transpeptidase domain the structure had been solved to 2.4 Å resolution, new X-ray coordinates for the worst modeled loop (residues T370 to M386; 17 out of a total number of 351 residues constituting the transpeptidase domain) were incorporated, as kindly provided by Dr. Dideberg. The structure was relaxed with molecular dynamics simulations and simulated annealing methods. The RMS deviation between the 144 aligned C^{α}-atoms and the corresponding ones in the originally solved 3.5 Å resolution crystal structure was 0.98. The 351 C^{α}-atoms of the whole transpeptidase domain of the final model showed an RMS deviation of 1.58. The Ramachandran plot showed that 79.3% of the residues are in the most favored regions, with only 1.0% occurring in disallowed regions. The model presented here can be used to investigate the three-dimensional influences of mutations around the active site of PBP2*x*.

Introduction

The 'Holy Grail' in protein modeling has always been the construction of 3D models of protein structures from scratch [1]. Although there is considerable progress in the field [2], current methods are still not accurate enough. Fortunately, it is often possible to obtain C^{α}-atom information from poorly solved structures, especially where large proteins are concerned. Many methods have been developed for generating full protein structures from C^{α}-atom information [3–11]. The first step in these methods frequently is the reconstruction of the backbone geometry, since it is well-recognized that the backbone geometry is mainly governed by hydrogen-bonding interactions [12], while a crude protein structure might be determined by forces optimizing the hydrophobic packing interaction. Two groups of algorithms can be distinguished in these methods [13]: one based on the fitting of backbone fragments contained in a library [6–8] and one based on geometric considerations [9–11]. Some examples of the latter group are represented by the virtual-bond chain method [14] (which makes use of the fact that the distance between two constituent C^{α}-atoms is independent of ϕ and ψ) and the dipole-path

*To whom correspondence should be addressed.

method [13]. Whereas in the former method the conformational space of a polypeptide chain is described by the (torsional) angles between the C^{α}-atoms, the latter method aims at reconstructing a backbone from C^{α}-carbon coordinates based on peptide-group dipole alignment. In essence, though, these methods completely rely on the X-ray structure resolution, which makes application to partially solved low-resolution crystal structures not very useful. It is therefore more desirable to incorporate some degree of structure optimization during the process of generating full protein coordinates.

The procedure presented here starts with local trace alignments of secondary structure elements and some parts of loops. These structurally aligned segments are incorporated into an initial model of the protein (which can be generated with any protein modeling package) by folding the corresponding regions. Subsequently, all amino acids, except for the glycine and proline residues, are mutated into alanine residues, after which the structure is subjected to a simulated annealing molecular dynamics simulation. This procedure was chosen, since glycine shows considerable differences in geometrical backbone properties compared to other amino acids, and proline is not able to act as a hydrogen bond donator. This stage of the procedure may be seen as an extension of the method described by Claessens et al. [15], who built a polyalanine peptide by searching a crystallographic database for optimally matching segments, and used the resulting protein as a starting structure for MD simulations. Because the method starts with a local trace alignment, there are also similarities with the fragment-matching methods developed by Jones and Thirup [6], Reid and Thornton [7], and Holm and Sander [8]. Likewise, in order to match a fragment of varying length starting at a random C^{α}-position Levitt [16] searched a database of highly refined structures. By repeating this procedure several times he generated a set of different protein structures matching the trace, which were averaged afterwards. However, like in other methods [4a, 7–9, 17, 18] problems arise at junctions by overlapping one or two peptide groups. Such difficulties can be avoided, by simply folding the corresponding regions of a previously generated initial model according to the aligned regions.

Moreover, with most of these methods a starting structure is generated, which exactly matches the only experimentally known structural data, i.e. the structure of the trace atoms. If these data correspond to poorly resoluted structures, it is desirable to first allow the backbone to obtain a better conformation by performing molecular dynamics simulations, while *restraining* the C^{α}-atoms to their corresponding positions. In this way only minor deviations from the poorly solved crystal data are allowed. Since restraining force constants are also atom-specific, it is possible to allow the various parts of the protein to be relaxed to different degrees, while simultaneously allowing for enough conformational freedom for non-homology derived side chains. The method, combining both local trace alignments and comparative modeling with *ab initio* side chain modeling, therefore allows for thoroughly sticking to the X-ray coordinates, but at the same time resulting in more feasible structures.

Since a number of arbitrary decisions have to be made during the course of the method (as is the case with most automated protein prediction methods [2]), though, it is hardly possible to judge the method quantitatively. The quality of the final model, therefore, strongly depends on the intuition and skill of the modeler. The only way to judge the quality of such a method would be comparison of the final model with a better solved X-ray structure. Penicillin binding protein 2*x* (PBP2*x*) of *Streptococcus pneumonia* serves as an excellent example for this purpose. This protein was crystallized in 1993 [19] and its structure solved to 3.5 Å in 1996. Due to this low resolution, only the coordinates of the trace atoms have been deposited in the PDB [21]. After having generated full protein coordinates of the transpeptidase domain based on these data, however, the structure had been solved to 2.4 Å resolution [21], enabling more precise recognition of side chain orientations.

PBP2*x** is the only high molecular mass (hmm) PBP that has been solved so far. Like all PBPs (both hmm and low molecular mass, lmm) it takes part in the final step of bacterial cell wall synthesis, which involves cross-linking between the penultimate D-Ala residue of the D-Ala-D-Ala C-terminal ends of the newly synthesized peptidoglycan strands and the amino terminal of meso-diaminopimelate (transpeptidation) (for a review see [22]). Since only the hmm PBPs seem to be essential [23], though, it is highly desirable to understand the catalytic mechanism of these enzymes. Moreover, since *Streptococcus pneumonia* is able to achieve resistance to β-lactam containing antibiotics by modifying essential PBPs into low-affinity variants [24], it is important to study the consequences of these mutations on the affinity of this enzyme for these compounds three-dimensionally.

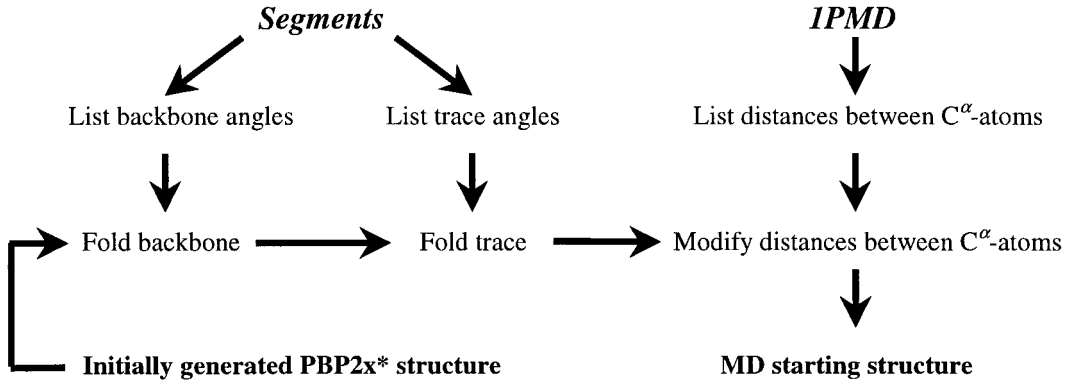


Figure 1. Flow diagram for the generation of the complete structure of the transpeptidase domain of PBP2x* from its C^{α}-atom coordinates (1PMD).

Methods

*Generation of a complete initial structure of the transpeptidase domain of PBP2x**

The program used for visualization was Insight II Versions 97.0 and 98.0 [25]. All calculations (force field: CVFF) were performed on a Silicon Graphics Origin2000 parallel computer. An initial all-atom structure of PBP2*x** was generated by dynamic programming upon simultaneously solving both the equations defining the bond angle constraint in terms of the peptide plane angles [26] and the virtual torsion constraint [14]. During the process of iteratively matching four consecutive C^{α}-atoms with segments of known protein structures contained in a built-in database, frequently occurring conformations were statistically favored. Peptide flips were introduced, because local secondary structure was not taken into account. Therefore, the trace atoms of the secondary structure elements of the transpeptidase domain of PBP2*x* (defined as ranging from S266 to D616 as postulated by Pares et al. [20]) were aligned locally structurally with the corresponding regions in five β-lactamases of known 3D structure (Figure 2). Similarly, some (parts of) loops could also be aligned. Using the Dayhoff PAM 250 substitution matrix [27] and the enhanced Needleman and Wunsch algorithm for pairwise alignment [28] as implemented in the Homology module of InsightII, the segments that were thus aligned are summarized in Table 1. A superposition of these parts with the crystal structure is shown in Figure 3.

The geometry of the thus created segments was incorporated into the initially generated structure by subsequently incorporating the angles and dihedrals between the main chain heavy atoms of the newly generated segments, refolding the trace according to the X-ray structure, and finally modifying the distance between two consecutive C^{α}-atoms of the modified regions. The last step of the procedure was performed by subsequently breaking the peptide bonds connecting two amino acids, adjusting the distance between the two C^{α}-atoms, and finally recovering the peptide bond. Afterwards, the structure was minimized. Following this route the trace geometry was completely recovered.

Table 1. Regions for which the C^{α}-atoms were aligned with the β-lactamases, as well as their homology (Dayhoff PAM 250 substitution matrix [27] and enhanced Needleman and Wunsch algorithm for pairwise alignment [28]) and RMSD values

Region	Length	β-Lactamase	Homology	RMS
S_{267}–K_{287}	21	2BLT	−4.29	0.96
Y_{288}–T_{305}	18	1BTL	4.44	0.77
Q_{330}–E_{334}	5	1CBL	26.00	0.49
P_{335}–N_{350}	16	4BLM	19.38	0.54
H_{394}–K_{406}	12	4BLM	8.46	0.43
A_{410}–K_{420}	11	1BTL	7.27	0.51
T_{459}–A_{469}	11	1BTL	4.55	0.55
A_{508}–L_{517}	10	1BLT	6.00	0.52
V_{544}–A_{551}	8	4BLM	20.00	0.83
I_{569}–S_{576}	8	4BLM	3.75	0.68
F_{583}–Q_{591}	9	4BLM	3.33	0.82
E_{602}–K_{615}	14	1BTL	−3.57	0.60

*Molecular dynamics simulations of the transpeptidase domain of PBP2x**

For several reasons, application of a standard tethering force was found to be unsuitable during the relax-

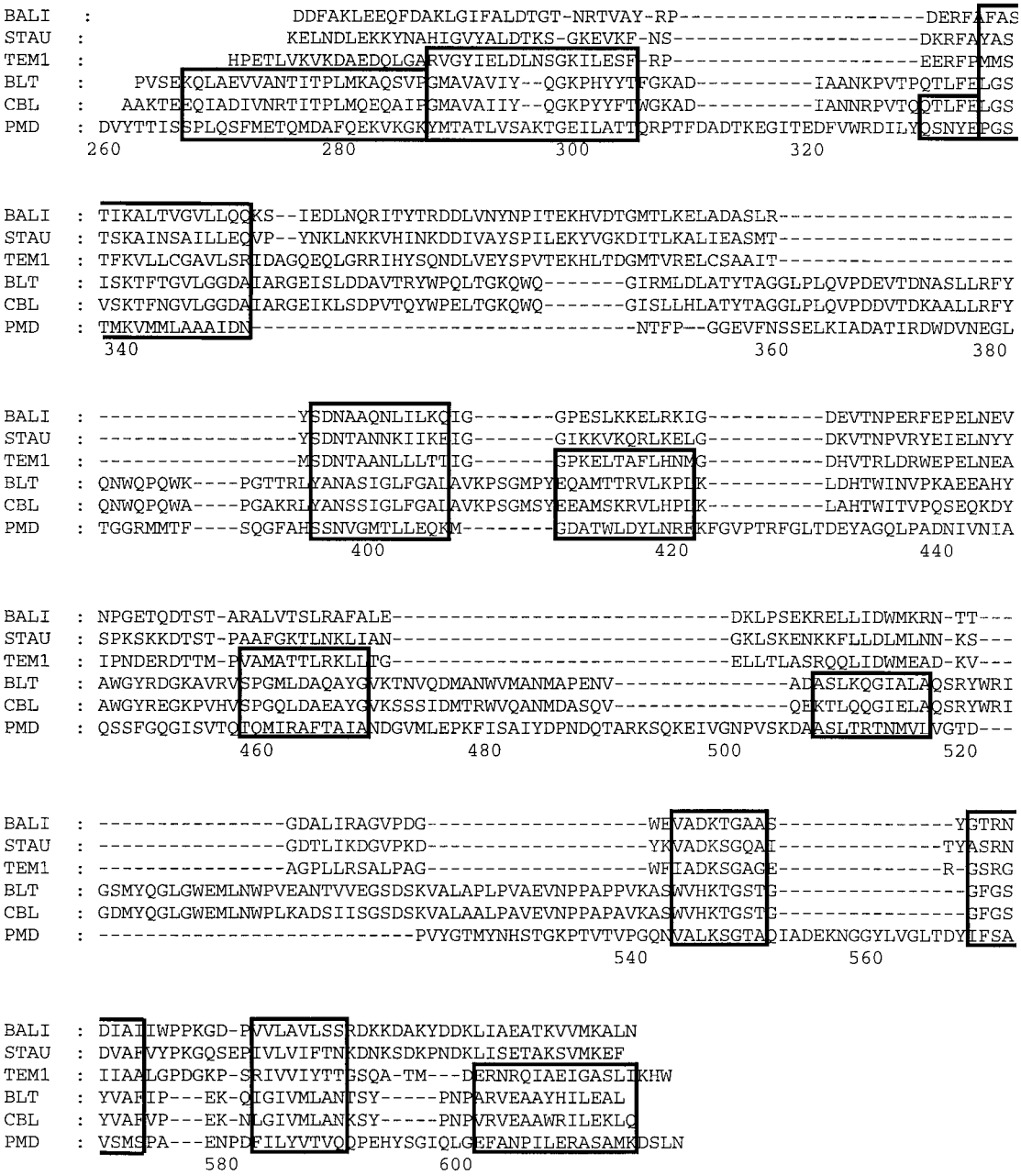


Figure 2. Alignment of the transpeptidase domain of PBP2*x** with the five β-lactamases of known 3D structure.

ation procedure. The main cause was that, in contrast to standard relaxation of homology modeled protein structures, in our case a problem had to be treated, where only *incomplete* information was available, i.e., the C^{α} coordinates. It was necessary to slowly relax the 3D homology derived regions in a positionally constrained framework, while simultaneously allowing for enough flexibility. Different force constants were, therefore, needed. These criteria were met upon application of a *restraining* force. Consequently, it was decided to run several MD simulations of the transpeptidase domain, during which the homology derived regions were relaxed slowly, while restraining the C^{α}-carbon atoms in the remaining regions to their respective X-ray structure positions, and keeping both the C-terminal and N-terminal domains fixed.

However, due to the *local* alignment procedure of the secondary structure elements described above, the N-H. . . O=C hydrogen bonding network present in the five-stranded β-sheet was not arranged optimally. Before embarking on MD simulations of the transpeptidase domain as a whole, the N-H. . . O=C hydrogen bonding network was, therefore, optimized by subjecting the transpeptidase domain to an annealed MD simulation, in which all of the amino acids of PBP2*x** (except for the proline and glycine residues) had been transformed into alanine residues. Of the resulting poly-Ala-Gly-Pro structure, in which the hydrogen bonds were solely represented by the backbone N-H. . . O=C interactions, a simulated annealing MD run was performed at 1000 K, while *restraining* the C^{α}-atoms to their X-ray positions with a force of 10–15 kcal/mol/Å. In addition, each omega angle present in the non-proline residues was forced to adopt a *trans* configuration with 5 kcal/mol. After 20 ps initialization and 80 ps simulation at this temperature, the system was cooled down to 0 K. Subsequently, the system was repeatedly heated to 1000 K. It was frequently observed that the complete peptide group between two consecutive C^{α}-atoms was rotated over 180°, while the *trans* configuration was conserved. With the exception of P308, PROCHECK showed that none of the peptide moieties was changed into *cis* conformation. With this procedure, 10 structures of the poly-Ala-Gly-Pro protein were obtained, and finally minimized (first 500 cycles steepest descent, followed by conjugate gradients). A superposition of these structures showed nicely that some regions in the protein are more flexible than others (results not shown).

The lowest energy conformation was taken as starting conformation for the next simulated annealing calculation, in which the system was repeatedly cooled down from 600 K to 100 K. A constant restraining force constant was used, except for the force constant applied to conserve the *trans* conformation in the non-proline residues. The latter was reduced to 2 kcal/mol.

Afterwards, the native amino acid composition of the transpeptidase domain (including the side chain conformations of the newly modeled segments) was recovered. The natural amino acid composition was not recovered for the C-terminal and N-terminal domains, ensuring unrestricted movement of the side chains in the border regions of the transpeptidase domain during the MD simulations. After minimization of the side chains a protein structure was obtained with a more realistic backbone structure.

At this point the dihedral angles of the side chains in the methionine cluster (comprising M339 to A347, F392, M400, M407, W412, and F419) were adjusted manually according to those in the article by Pares et al. [20]. After minimization the resulting structure was taken as template for MD simulations of the transpeptidase domain.

First, a simulated annealing calculation was performed at 610 K, during which the atoms mentioned above were *restrained* to their corresponding positions in the template using 2–5 kcal/mol. After 80 ps (including 20 ps initialization) the system was cooled down to 310 K. After allowing the structure to relax during 15 ps, its conformation was saved. Subsequently, it was heated to 610 K again during another 15 ps. By repeating this procedure several times, 5 structures were obtained, which were minimized. With the aim of relaxing the transpeptidase domain as much as possible, the best conformation was taken as a starting conformation for a series of MD simulations at 310 K of 220 ps each, while different *restraining* force constants on different regions were applied. At this point a structural water molecule [29] was included in the MD simulations. Indeed, a water molecule could be easily placed in the hole between P335, T338, Y586 and S571. During the simulations that followed the water molecule was kept in place by restraining the intermolecular hydrogen bonds with these four amino acids with 1.0–2.0 kcal/mol/Å.

Furthermore, the transpeptidase domain was divided into two sets of regions: one in which the residues were restrained to their crystal structure positions 1.0–2.0 kcal/mol/Å in all simulations, and one in which the residues were to be relaxed completely. Thus, a series of six MD simulations was performed, in which the C^{α}-atoms of the residues present in the first set (colored green in Figure 4) were position-restrained with a force of 1.0 to 2.0 kcal/mol/Å, and the C^{α}-atoms of the residues present in the second set (represented in Figure 4 by both the orange and magenta colored regions) were relaxed slowly by decreasing the restraining force from 1.0–2.0 kcal/mole/Å to 0 kcal/mole/Å in steps of 0.2–0.4 kcal/mole/Å. As explained in Figure 4, some parts of the regions generated by BIOSYM (or any other program) were also allowed to move freely. By choosing the regions to be relaxed and restrained as shown in Figure 4, the residues of the secondary structure elements constituting the core of the protein were optimized in the framework of the loops, whereas the side chains contained in these loops were also free to

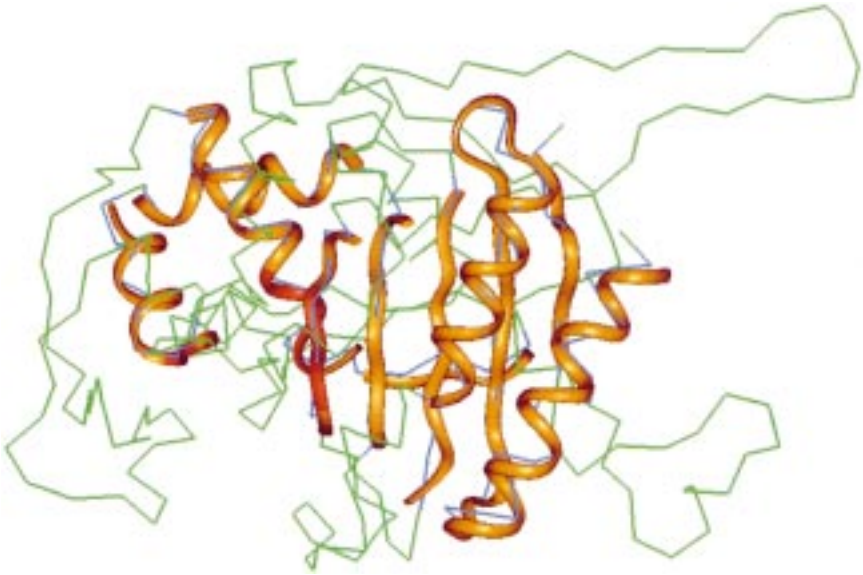


Figure 3. Superposition of 1PMD.pdb onto the total structure composed of fragments obtained by alignment of 1PMD.pdb with the five β-lactamases of known 3D structure (shown as ribbon). The corresponding regions in 1PMD.pdb are colored blue. The parts of the ribbon covering the active site residues are colored red.

move. In order to retain the accessibility of the active site, though, it was necessary to constrain most of the C^{α}-atoms of B3 and B4 of the five-stranded β-sheet to their X-ray coordinates as well. During these simulations, the heavy atoms of the side chains contained in the methionine cluster were also relaxed stepwisely. Each of these six 220 ps MD simulations was initialized for 20 ps. Subsequently, every 20 ps the conformation was saved, all of which were subjected to a minimization procedure afterwards (500 steps of steepest descent minimization, followed by conjugate gradients minimization until the energy was less than 1.0 kcal/mol). One of the ten resulting conformations was chosen as starting conformation for the next MD run (criteria used: low value of internal energy, low RMS deviation of the trace, high percentage of amino acids in most favored regions, sufficient length of simulation time). In the last MD simulation the orange colored regions were completely free to move within the framework of the large loops surrounding the transpeptidase domain. Since it was observed that the last helix of the transpeptidase domain (containing a proline residue) started to unfold when no restraining force was applied at all, in the last simulation the C^{α}-atoms of this helix were also restrained with a force constant of 0.2–0.4 kcal/mole/Å. A summary of the most important parameters obtained for the various MD simulations is given in Table 2.

At this point, the model was sent to the authors of the crystal structure for verification, since in the meantime the crystal structure had been solved to a resolution of 2.4 Å [21], allowing localization of the side chains. Almost all of the side chains appeared to be in the correct orientation [30], except for F359 and M386. Major differences appeared only in the loops. Fortunately, Dr O. Dideberg supplied the new X-ray coordinates of the loop spanning residues T370 to M386. So, finally, after substituting these new coordinates for the corresponding amino acids, and folding the rest of the structure according to 1PMD.pdb, the last six MD simulations were performed again. This time the heavy atoms of T370 to M386 were also restrained, in order to maintain their X-ray coordinates as much as possible. In the last simulation one of the hydrogens of the positively charged amine group of K547 and the carbonyl oxygen atom of S548 were constrained to each other using a force of 1–2 kcal/mole/$Å^2$. In order to preserve the hydrogen bond which the charged amine group of K547 forms with the backbone carbonyl oxygen atom of S548 during the last simulation, a distance constraint of 1.5–3 kcal/mole/$Å^2$ was added between one of its hydrogen atoms and the S548 carbonyl oxygen atom. Without this constraint the positively charged nitrogen of K547 was found to be attracted by the dipole of the opposite helix. However, since the crystal structure

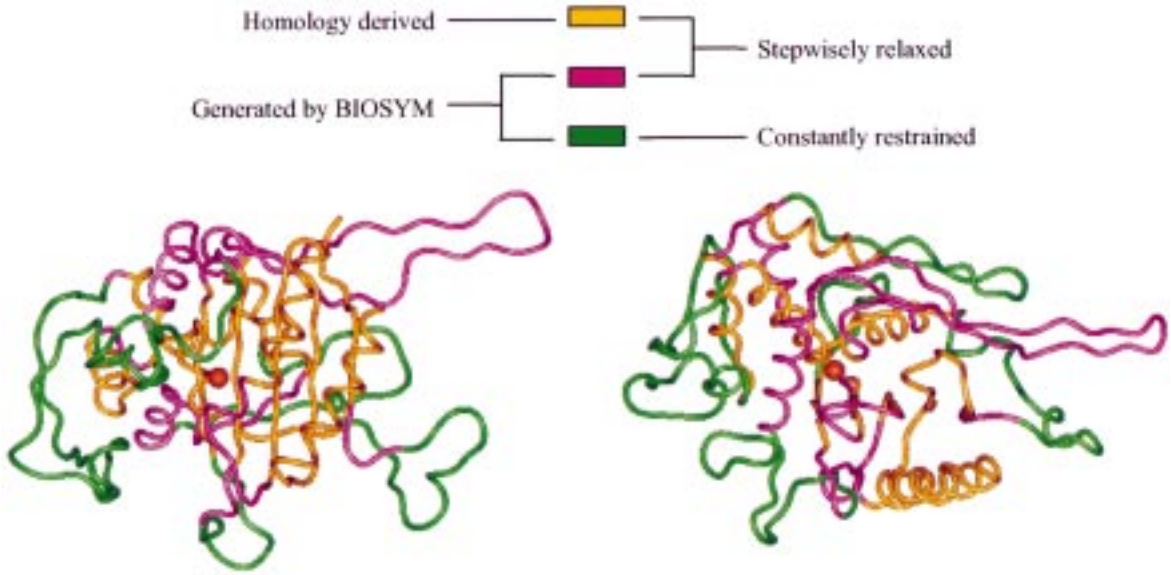
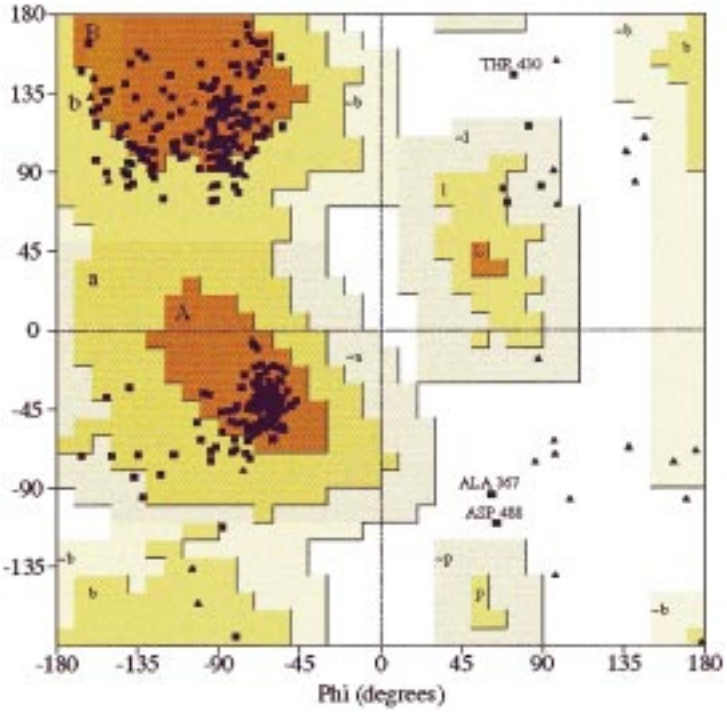


Figure 4. Ribbon representation of the transpeptidase domain of PBP2*x**. The structural water molecule is shown as a red dot. The regions in which the C^{α}-atoms have been restrained constantly to their corresponding positions in the crystal structure are shown in green. Both the regions shown in orange and magenta have been relaxed slowly by decreasing the restraining force constant to zero gradually. For further details see text.

Figure 5. Ribbon representation of the superposition of the C^{α}-atoms of the transpeptidase domain (Ser266 to Asp616) of the crystal structure (1PMD.pdb, shown in blue) and the optimized structure of the model (shown in red). The C^{α}-atoms of the N-terminal and C-terminal domains have been superimposed in order to show the relative displacement of the transpeptidase domain.



Plot statistics

Residues in most favoured regions [A, B, L]	241	79.3%
Residues in additional allowed regions [a,b,1,p]	54	17.8%
Residues in generously allowed regions [~a, ~b, ~1, ~p]	6	2.0%
Residues in disallowed regions	3	1.0%
Number of non-glycine and non-proline residues (excl. Gly and Pro)	304	100%
Number of end-residues (excl. Gly and Pro)	2	
Number of glycine residues (shown as triangles)	29	
Number of proline residues	16	
Total number of residues	351	

Figure 6. Ramachandran plot of the optimized structure of the transpeptidase domain of PBP2*x**.

shown in reference [20] clearly shows the presence of a hydrogen bond with the backbone carbonyl group, it may be speculated that movement of this charged amine group towards the opposite helix is hindered by another structural water molecule.

A summary of the most important parameters of the second round of MD calculations is presented in Table 3. Figure 5 shows the relative displacement of the ribbon with respect to the crystal structure by superimposing both the N- and C-terminal domains. In Figure 6 a Ramachandran plot is shown of the final model. Finally, in Figure 7 a view of the active site is presented.

Discussion

The work presented here describes how several well-established protein modeling techniques can be combined for the generation of a full protein structure from the 3D coordinates of C^{α}-atoms by incorporating 3D coordinates of highly homologous proteins. By applying the method to the generation of an accurate full 3D structure of the transpeptidase domain of *S. pneumoniae* PBP2*x**, it was shown that particularly the active site region can be modeled accurately.

In the first stage of the procedure, it has been shown that the geometry of an optimally aligned seg-

Table 2. Analysis of the various MD simulations of the penicillin-binding domain of PBP2*x**

	Temperature (K)	Restrain/ tether	Force	Simulation time (ps)	Internal energy (kcal)	RMS trace (full transpeptidase domain)	RMS trace (aligned regions)	% Most favoured regions	% Dis-allowed regions	Distorted		
										Main chain bonds	Main chain angles	Planar groups
Poly-Ala-Gly-Pro	1000	restrain	10–15	390	2030.5	0.61	0.53	78.6	1.0	0	0	0
	600	restrain	10–15	220	2018.1	0.66	0.55	78.0	1.0	0	0	0
Full penicillin-binding domain	610	restrain	2–5	220	2425.5	0.85	0.57	80.6	1.0	10	0	3
	310	restrain	1.0–2.0	220	2395.0	0.93	0.74	79.6	1.0	6	0	3
	310	restrain	0.8–1.6	220	2300.6	1.23	0.96	80.6	0.7	9	0	2
	310	restrain	0.6–1.2	220	2292.6	1.32	1.02	82.2	0.7	8	0	2
	310	restrain	0.4–0.8	220	2282.5	1.45	1.05	82.2	0.7	8	0	0
	310	restrain	0.2–0.4	220	2288.0	1.52	1.00	82.2	0.7	7	1	2
	310	restrain	0	220	2260.1	1.64	1.29	79.3	1.0	8	2	4

Table 3. Analysis of the various MD simulations of the penicillin-binding domain of PBP2*x** after incorporation of the X-ray coordinates of T370 to M386

	Temperature (K)	Restrain/ tether	Force	Simulation time (ps)	Internal energy (kcal)	RMS trace (full transpeptidase domain)	RMS trace (aligned regions)	% Most favoured regions	% Dis-allowed regions	Distorted		
										Main chain bonds	Main chain angles	Planar groups
Full penicillin-binding domain	310	restrain	1.0–2.0	220	2372.9	1.04	0.83	77.0	0.7	5	0	4
	310	restrain	0.8–1.6	220	2354.7	1.17	0.88	80.9	0.7	6	0	3
	310	restrain	0.6–1.2	220	2331.4	1.26	0.89	82.6	0.7	6	0	2
	310	restrain	0.4–0.8	220	2282.5	1.39	0.90	80.3	0.7	6	0	3
	310	restrain	0.2–0.4	220	2290.0	1.50	0.92	79.3	1.0	8	0	0
	310	restrain	0	220	2282.8	1.56	0.98	79.3	1.0	7	0	1

ment of a highly homologous protein can serve to supply an incompletely solved low-resolution structure with missing coordinates, simply by adjusting some dihedral angles. Fortunately, it appeared that upon modification of the distance between two atoms along their bond vector, the rest of the protein was moved as well. This aspect of the procedure could be of general value in the homology modeling procedure, since insertion of a small loop between two secondary structure elements is often prevented by geometric factors.

The data presented in Table 2 show that the relaxation protocol used here served very well. The high percentage of amino acids in the most favored regions of 79–82% is maintained throughout all restrained dynamics, whereas the internal energy decreases steadily. Moreover, taking into account that the RMS deviation of the C^{α}-atoms in the transpeptidase domain of the *optimized* structures reflects the error of the 351 C^{α}-atoms of the transpeptidase domain of the model with respect to those of a crystal structure solved at 3.5 Å resolution, it may be concluded that the coordinates of the trace atoms are nicely maintained.

Nevertheless, the RMS deviation of the corresponding 351 C^{α}-atoms from those contained in the 3.5 Å resolution crystal structure is relatively high (1.56 of the final model). As follows from Tables 2 and 3, this is mainly due to the high flexibility of the large loop regions. Indeed, the RMS deviation of the originally aligned regions from the corresponding C^{α}-atoms in the crystal structure remains below 1.0. In spite of restraining the C^{α}-atoms in these loop regions, the resulting movement is likely to have been caused by the absence of solvent. This especially accounts for the long loop connecting B3 and B4. This loop would have completely moved toward the bulk of the protein, if no restraints had been applied. As a consequence, B3 and B4 would have been distorted as well, which in turn would have affected the position of the structural water molecule. Since it was not found to be very useful to restrain this loop only, additional restraints were placed on the C^{α}-carbon atoms of both β-sheets. Similarly, it was necessary to restrain the C^{α} atoms in the large loop spanning residues T352 to F392, because otherwise movement of this region towards the bulk of the protein resulting from van der Waals forces would have lowered the accessibility of the active site considerably. It has recently been confirmed that the side chains in this region are highly flexible and fold differently depending on the ligand bound [30].

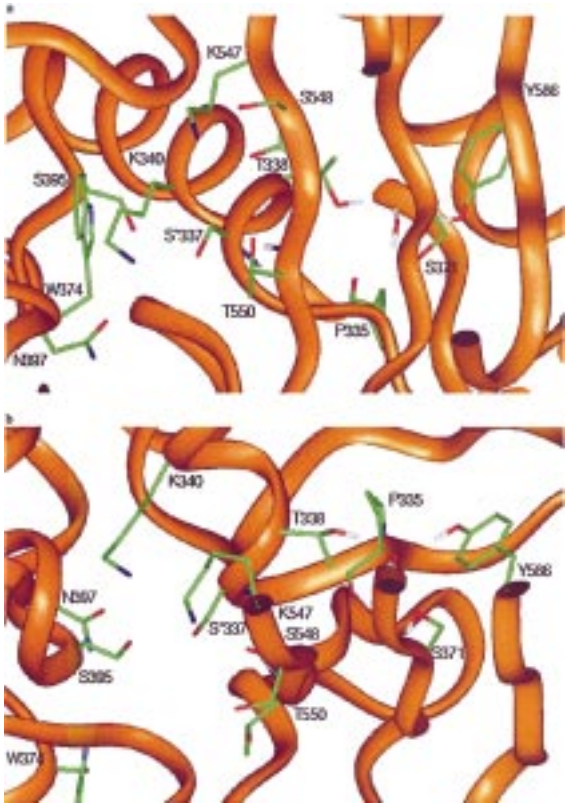


Figure 7. The active site of PBP2x*. (a) Front view. (b) Top view.

The Ramachandran plot presented in Figure 6 shows that 79.3% of the 304 non-glycine and non-proline residues contained in the transpeptidase domain have most favored geometry. Another 17.8% occur in the additional allowed regions. Moreover, since only 3 of these residues – all located at the ends of long loops – have bad geometry, it can be concluded that the quality of the model is good.

The good quality of the model was confirmed by Dr. Dideberg after comparison with the newly solved 2.4 Å resolution crystal structure [21]. Except for F359 and M386 all side chains are in the correct orientation. As expected, major differences occur in loop regions. The largest differences are represented by $C^{\alpha}{}_{480}$ (3.4 Å) and $C^{\alpha}{}_{495}$ (4.7 Å). The active site appears to be good (except for the left side, but this is the region for which new coordinates were supplied). The largest difference in this region occurs for $C^{\alpha}{}_{550}$ (RMSD = 1.8 Å), but this is exactly where the loop connecting B3 and B4 starts. Although the authors would very much like to be more specific here, the authors strongly depend on Dr. Dideberg's comments.

Finally, Figure 7 shows the active site of the final model, including the structural water molecule forming hydrogen bonds with P335, T338, S371 and Y586. Even without knowing the location of additional structural water molecules in the active site, it can be concluded that the dense hydrogen bonding network present in the Class A β-lactamases is also present in PBP2*x*. The active site is nicely flanked by the indole ring system of W374, which does not seem to be stabilized by any hydrogen bond in particular, though. As confirmed by experimental data [30], a weak hydrogen bond is present between the side chains of H394 and D373, whereas R384 does not interact with any particular amino acid. Summarizing, it can be concluded that the procedure developed here performs very well for the construction of the active site region of a protein that has been solved to low resolution, and of which the C^{α}-atom coordinates represent the only available structural data. Thanks to the application of restraining forces on some of the C^{α}-carbon atoms and the heavy atoms of the amino acids for which new coordinates were supplied, the active site is still accessible. Major differences occur in the loops, but these could have been lowered if a higher restraining force had been applied. However, this remains a matter of compromise, since this would also have affected the flexibility of these regions, which in turn would have decreased the conformational freedom of the aligned segments. Moreover, since it is well known that the quality of a model strongly depends on the quality of the alignment (which is also subject to the user's intuition [31]), it may be that a model can be improved by laboriously testing various alignments. Both this aspect and finding out which parts should be restrained represent the most laborious stages in the whole procedure.

Now that a reliable model of the active site region of PBP2*x* has been generated, molecular dynamics studies can be performed in order to study the effect of mutations. At present, docking studies are performed with some penicillins and cephalosporins. The results obtained with these docking studies will be published elsewhere. In the meantime, the coordinates of the full transpeptidase domain will be deposited in the Protein Data Bank.

Acknowledgements

The authors would like to thank Dr. O. Dideberg for his cooperation in this modeling project. Without his supplying new coordinates for the loop region spanning amino acids T370 to M386, publication of the model would not have been possible. The authors would also like to thank Dr. M. Adikwu for checking the English.

References

1. Moult, J., Curr. Opin. Biotechnol., 10 (1999) 583.
2. Moult, J., Hubbard, T., Fidelis, K. and Pedersen, J.T., Proteins Struct. Funct. Genet., Suppl. 3 (1999) 2.
3. a. Schäfer, L., Cao, M. and Meadows, M.J., Biopolymers, 35 (1995) 603.
   b. Schäfer, L. and Cao, M., J. Mol. Struct., 33 (1995) 201.
   c. Jiang, X., Cao, M., Teppen, B., Newton, S.Q. and Schäfer, L., J. Phys. Chem., 99 (1995) 10251.
   d. Jian, G.X., Cao, M., Newton. S.Q. and Schäfer, L., Electronic J. Theor. Chem., 1 (1995) 11.
4. a. Gan, K., Alexander, P., Coxon, J.M., McKinnon, A.J. and Worth, G.H., Biopolymers, 41 (1997) 381.
   b. Gan, K., Alexander, P., Coxon, J.M., McKinnon, A.J. and Worth, G.H., Biopolymers, 41 (1997) 367.
5. Milik, M., Kolinski, A. and Skolnick, J., J. Comput. Chem., 18 (1997) 80.
6. Jones, T.A. and Thirup, S., EMBO J., 5 (1986) 819.
7. Reid, L.S. and Thornton, J.M., Proteins Struct. Funct. Genet., 5 (1989) 170.
8. Holm, L. and Sander, C., J. Mol. Biol., 218 (1991) 183.
9. Purissima, E.O. and Scheraga, H.A., Biopolymers, 23 (1984) 1207.
10. a. Rackovsky, S. and Scheraga, H.A., Macromolecules, 11 (1978) 1168.
    b. Rackovsky, S. and Scheraga, H.A., Macromolecules, 13 (1980) 1440.
11. a. Wako, H. and Scheraga, H.A., J. Protein Chem., 1 (1982) 5.
    b. Wako, H. and Scheraga, J.A., J. Protein Chem., 1 (1982) 85.
12. Dill, K.A., Biochemistry, 29 (1990) 7133.
13. Liwo, A., Pincus, M.R., Wawak, R.J., Rackovsky, S. and Scheraga, H.A., Protein Sci., 2 (1993) 1679.
14. Nishikawa, K., Momany, F.A. and Scheraga, H.A., Macromolecules, 7 (1974) 797.
15. Claessens, M., Van Cutsem, E., Lasters, I. and Wodak, S., Protein Eng., 2 (1989) 335.
16. Levitt, M., J. Mol. Biol., 226 (1992) 507.
17. Correa, P.E., Proteins, 7 (1990) 366.
18. Rey, A. and Skolnick, J., J. Comput. Chem., 13 (1992) 443.
19. Charlier, P., Buisson, G., Dideberg, O., Wierenga, J., Keck, W., Laible, G. and Hakenbeck, R., J. Mol. Biol., 232 (1993) 1007.
20. Pares, S., Mouz, N., Pétillot, Y., Hakenbeck, R. and Dideberg, O., Nat. Struct. Biol., 3 (1996) 284.
21. Gordon, E., Mouz, N., Di Guilmi, A.M., Martin, L., Duee, E., Vernet, T. and Dideberg, O., Sectoral Meeting: 'Controlling the Proliferation of the Microbial Cell Factory', Verona, Italy, April 19–21, 1999.
22. Ghuysen, J.-M., Annu. Rev. Microbiol., 45 (1991) 37.
23. Ghuysen, J.-M. and Dive, G., In R. Hakenbeck and J.-M. Ghuysen (Eds.) Bacterial Cell Wall, New Comprehensive Biochemistry, Vol. 27, Elsevier, Amsterdam, 1994, pp. 103–129.
24. Laible, G. and Hakenbeck, R., J. Bacteriol., 173 (1991) 6986.

25. Insight II, Molecular Simulations Incorporated, Version 97.0, Biosym/MSI, San Diego, CA, USA.
26. Payne, P.W., Protein Sci., 2 (1993) 315.
27. Dayhoff, M.O., Barker, W.C. and Hunt, L.T., Methods Enzymol., 91 (1983) 524.
28. Needleman, S.B. and Wunsch, C.D., J. Mol. Biol., 48 (1970) 443.
29. Mouz, N., Gordon, E., DiGuilmi, A.-M., Petit, I., Petillot, Y., Dupont, Y., Hakenbeck, R., Vernet, T. and Dideberg, O., Proc. Natl. Acad. Sci. USA, 95 (1998) 13403.
30. Dideberg, O., personal communication.
31. Bates, P.A. and Sternberg, M.J.E., Proteins Struct. Funct. Genet., Suppl. 3 (1999) 47.