Journal of Computer-Aided Molecular Design, 11 (1997) 547–556.
KLUWER/ESCOM

J-CAMD 416

Inclusion of conserved buried water molecules in the model structure of rat submaxillary kallikrein

Elsa F. Henriques[a], Maria J. Ramos[a,*] and Christopher A. Reynolds[b]

[a]*CEQUP/Departamento de Química, Faculdade de Ciências, Universidade do Porto, Rua do Campo Alegre 687, 4150 Porto, Portugal*
[b]*Department of Biological and Chemical Sciences, University of Essex, Colchester, Essex C04 3SQ, U.K.*

Received 27 April 1997
Accepted 4 July 1997

Keywords: Molecular modelling; Protein structure; Rat submandibular kallikrein; Molecular dynamics; Homology modelling

Summary

A new approach to the molecular modelling of homologous serine proteases is adopted, by including a set of 21 buried waters known to be preserved in enzymes sharing the primary specificity of trypsin, in the homology modelling of rat submaxillary gland kallikrein. Buried waters – water molecules sequestered from bulk solvent within a protein matrix – appear to be integral conserved components of all serine proteases of known structure and should be incorporated into serine protease models built on the basis of sequence/structural homology to this family. The absence of such waters might induce errors in a force field simulation, favouring the formation of nonexistent hydrogen bonds and locally inaccurate structure. The kallikrein model refinement has led to the conclusion that an additional buried water should be added to the original rigid matrix of 21 conserved water molecules. The structurally preserved protein cavities of such waters validate the modelled structure.

Introduction

Glandular (true tissue) kallikreins (E.C. 3.4.21.35) are a group of structurally and functionally related trypsin-like serine proteases that selectively cleave kininogen, releasing the vasoactive peptides bradykinin or lysylbradykinin [1]. These enzymes are acidic glycoproteins which comprise several forms with molecular weights varying from 25 000 to 40 000 Da.

Members of the extended subfamily of kallikrein-like serine proteases include tonin, which cleaves angiotensin II from angiotensinogen [2], the γ-subunit of nerve growth factor, which processes the carboxyl terminus of its precursor [3], and rat submaxillary gland kallikrein, which has the substrate cleavage specificity of a true kallikrein [4,5].

The limited substrate cleavage specificity, characteristic of glandular kallikreins, contrasts with that exhibited by other serine proteases such as chymotrypsin, trypsin and elastase, and their rather restricted inhibition pattern towards the most frequently studied protein-proteinase inhibitors seems to be quite unique as far as trypsin-like serine proteases are concerned [6]. Unlike trypsin, arginine residues are preferred to lysine residues at the S1 subsite of these enzymes, whereas at the S2 subsite they favour large hydrophobic residues such as phenylalanine or leucine [7]. Another characteristic feature, the so-called 'kallikrein loop' following residue 95 (α-chymotrypsin related numbering system), is the largest amino acid insertion in kallikreins when compared to trypsin.

Regardless of the widely divergent molecular weights of tissue kallikreins, their structures are presumably closely related to porcine kallikreins, porcine pancreatic kallikrein being the best studied one [8]. The structural features of this particular enzyme suggest that the generally restricted binding capacity might be explained by steric hindrance with some external bulky loops surrounding the kallikrein binding site [9].

Ashley and MacDonald [10] reported the complete amino acid sequence of rat submaxillary gland kallikrein, which proved entirely homologous to rat pancreatic kallikrein [11]. The amino acid sequence of the predicted active form of this enzyme (237 residues) holds high levels of sequence homology to those of other kallikrein-like serine proteases [10,11].

*To whom correspondence should be addressed.

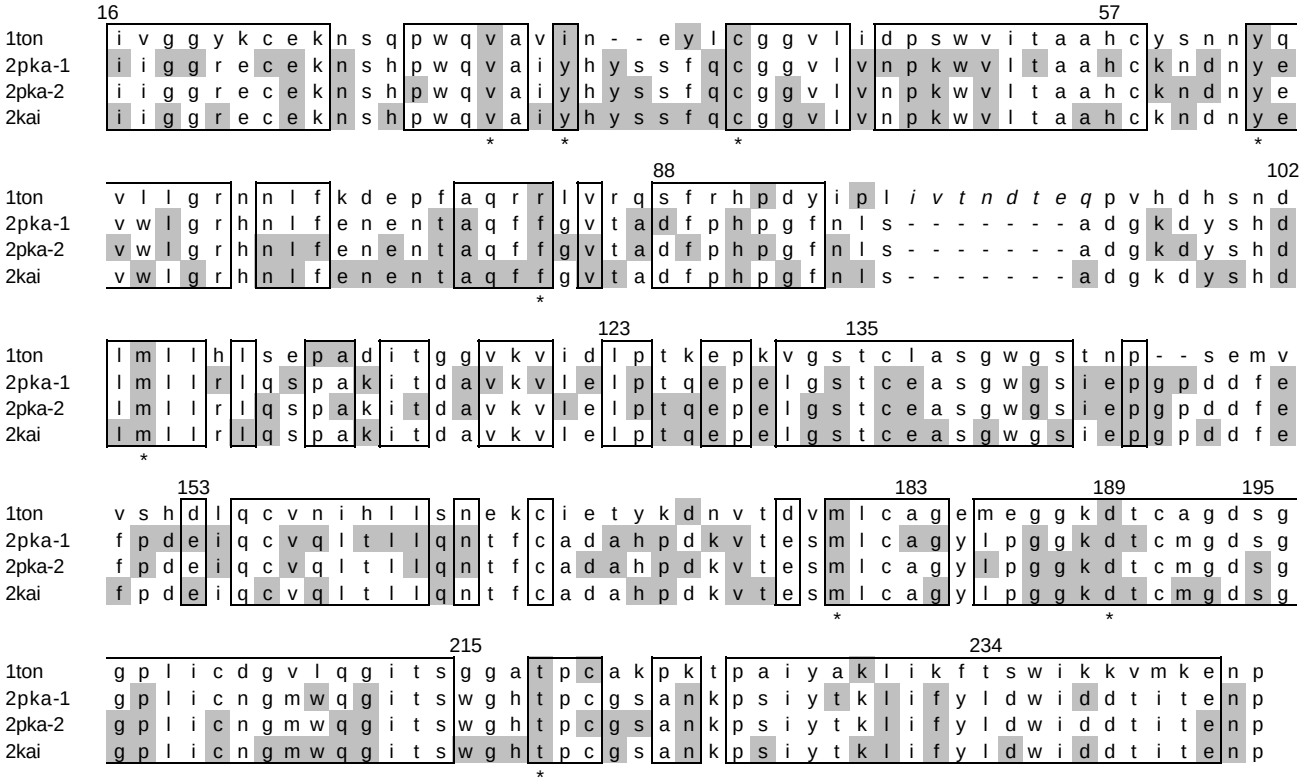


Fig. 1. Amino acid sequence alignment of rat tonin (1ton) and porcine pancreatic kallikrein (2pka – molecules 1 and 2 – and 2kai structures). Gaps (-) were introduced to optimize sequence alignments. The boxes delineate residue matches. Residues shadowed have nonplanar omega angles to adjacent residues. 'Conserved' nonplanar omega angles are marked with an asterisk.

Experimental crystal structures (X-ray, neutron diffraction, NMR) are now available for some individuals of the large serine protease family. Their structural similarity encourages the building of computer three-dimensional models for other family members of unresolved structure, using the available ones as templates. It has been shown that where sequence identity between the target and the template structure is high (over 70%), comparative molecular modelling is highly successful, although some difficulties arise with insertions, deletions and local low similarity [12].

Buried waters – water molecules sequestered from bulk solvent within a protein matrix – have been found in many proteins, and it has been suggested that they play an important common role in both the structure and function of proteins [13,14], and therefore they have been regarded as intrinsic to their structures. In particular, they appear to be integral conserved components of all serine proteases of known structure [15] and should be incorporated into serine protease models built on the basis of sequence/structural homology to this family. The absence of such waters might induce errors in a force field simulation, favouring the formation of nonexistent hydrogen bonds and locally inaccurate structure.

Our goal to predict the three-dimensional structure of rat submaxillary kallikrein using a set of conserved buried waters to improve the accuracy of the model is, to the best of our knowledge, a new approach to molecular homology modelling techniques. The X-ray crystal structures of rat tonin [16] and porcine pancreatic kallikrein [8,9] were the chosen templates, together with 21 buried water molecules known to be preserved in enzymes sharing the primary specificity of trypsin [15], with positions predeterminated by structural homology. In order to obtain a more realistic structure, molecular dynamics (MD) was used to refine the model in the presence of bovine pancreatic trypsin inhibitor [9]. Simulations which also included a 'cap' of solvent water molecules proved to be crucial as the results led us to believe that an extra water should be added to the initial set of 21 buried ones, suggesting that trypsin-like serine proteases might, in fact, contain 22 preserved water molecules. The conserved protein environments of these waters in the refined structure strongly suggest the model has been correctly built.

Methods

Sequence and structure alignments

Based on the sequence alignments reported by Ashley and MacDonald [10] and Swift et al. [11] and on the available serine protease X-ray crystallographic PDB structures [17,18], we considered rat tonin (E.C. 3.4.21.35)

and porcine pancreatic kallikrein A (E.C. 3.4.21.35) as the best possible templates for modelling the rat submaxillary kallikrein. Rat tonin (Ton) holds the highest sequence identity (74%) to rat submaxillary kallikrein (RSK) but presents an altered substrate specificity; in addition, its only X-ray PDB structure (1ton) is not in the active conformation, but one that has been perturbed by the binding of Zn^{2+} in the active site [16]. Porcine pancreatic kallikrein (PPK), which is 59% homologous with RSK, is the single true kallikrein of known experimental structure; moreover, both X-ray PDB structures (2pka and 2kai) are in the active form [8,9].

Ton and PPK sequences were aligned manually and their structures superposed using combined criteria of sequence identity/similarity and structural homology. Graphical inspection of the structures and coordinate manipulation were carried out with the aid of QUANTA [19] running on a Silicon Graphics IRIS Indigo workstation. The resulting sequence alignment shown in Fig. 1 is in agreement with the one presented by Fujinaga and James [16], except for the region in between residues 146 and 148 (α-chymotrypsin related numbering system). Residue matches are outlined (151 satisfy the chosen criteria). As the peptide bond is normally expected to be planar, Fig. 1 also highlights in grey what is sometimes regarded as a 'bad' conformational feature, namely the nonplanarity of the omega angle $\Omega = \tau(C^{\alpha}\text{-C-N-}C^{\alpha})$; the acceptable range is considered to be $-174.2° < \Omega < 174.2°$. For this purpose, the PPK amino acid sequence is repeated 3 times in Fig. 1, one for each different available conformation of the enzyme (two in 2pka – molecules 1 and 2 – and one in 2kai). Investigation of the main chain angles ϕ (phi) and ψ (psi) revealed that the Ton structure, which has the best resolution (1.8 Å), exhibits the smallest number of deviated values from the allowed main chain conformations [20].

It became clear that the Ton structure was the best template for the bulk skeleton of the RSK model, for it has the highest sequence homology together with a minimum of bad conformational features. However, for the regions where significant structural differences between Ton and PPK arise, one of the PPK structures was used to model the equivalent ones in RSK. For this latter purpose, we considered that the 2kai structure was the most appropriate, because the enzyme binds to a *reversible substrate analogue* inhibitor, the bovine pancreatic trypsin inhibitor (PTI) which was originally detected by Kraut et al. [21] as a kallikrein inactivator; in 2pka the complex is formed with the small irreversible inhibitor benzamidine. It is precisely in the region around the RSK active site that the model must be built upon the PPK conformational pattern; consequently, it was decided to model the enzyme in the complex form [RSK:PTI] in order to allow a better understanding of its selective binding and catalytic specificity.

When the two structures 1ton and 2kai were superimposed, only 41 α-carbon atom pairs (out of 227) did not superimpose within 1.0 Å, and the corresponding residues are signalled with an asterisk in Fig. 2 (PPK sequence).

RSK sequence alignment to the two template enzymes was based on sequence identity and functional homology, and residue matches are shown outlined in Fig. 2. The

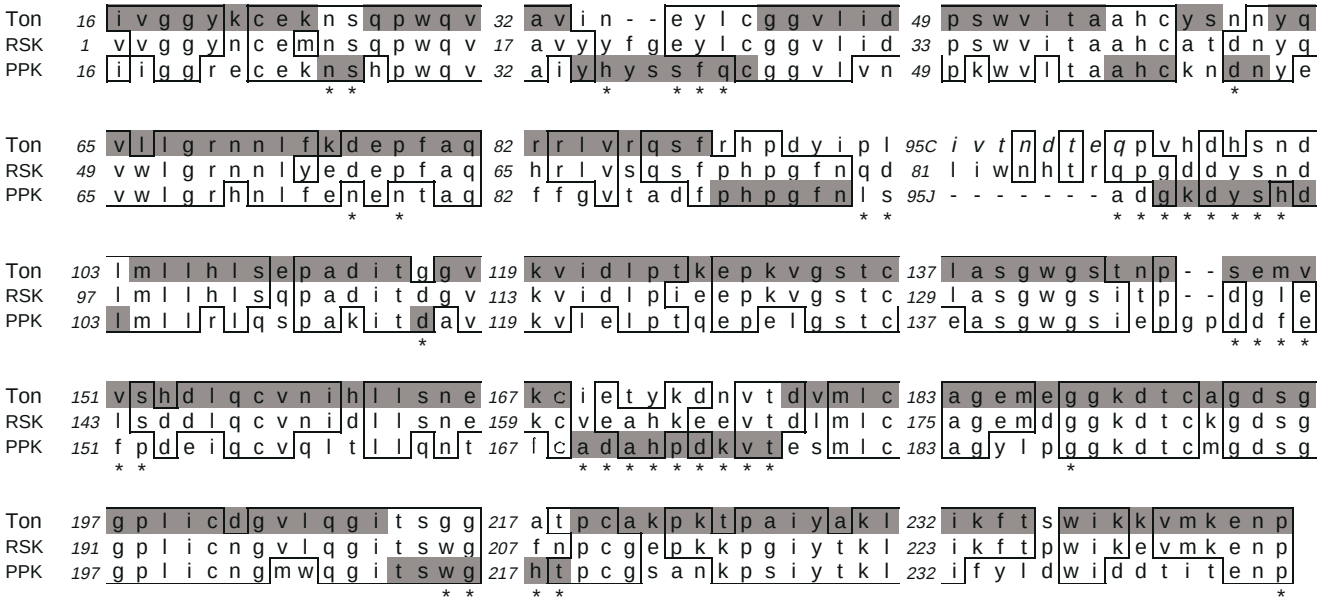


Fig. 2. RSK sequence alignment to the two template enzymes Ton and PPK. Gaps (-) were introduced to optimize sequence alignments. The boxes delineate residue matches. Residues in italic do not have crystallographic defined coordinates. Regions in the template structures that do not superimpose within 1 Å (C^{α} distance) are marked with an asterisk (PPK sequence). Fragments copied from the template structures to the RSK model are highlighted in grey.

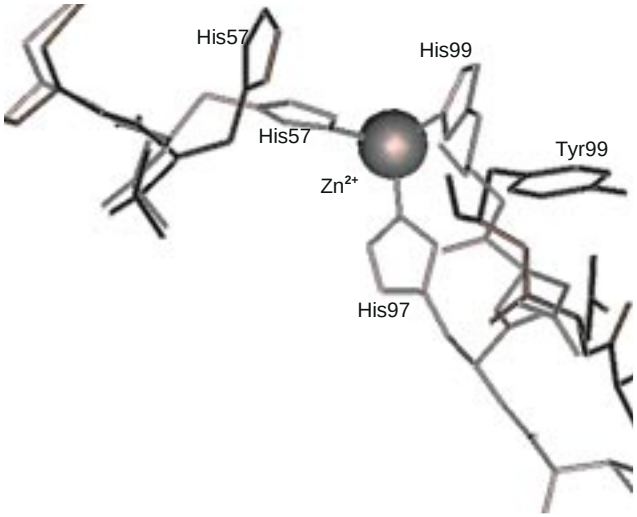


Fig. 3. Superposition of the neighbouring regions of His^{57} in the Ton (grey) and PPK (black) template structures. The van der Waals sphere represents the Zn^{2+} cation (in tonin).

alignment is in agreement with those presented by Swift et al. [11] and Ashley and MacDonald [10], although nothing had been concluded by these authors for the alignment in the region 146–148.

Modelling scheme

Most of the secondary structure elements of the RSK model were homology copied from the two selected structures, 1ton and 2kai. The main chain atom coordinates were copied directly from the known structure to the corresponding region in the model. Side chain atom coordinates were copied as far as they made sense with respect to the RSK residue type. Any unresolved side chain atom coordinates were built by CHARMm [22] during a regularization process.

The fragments copied from the template structures to the RSK model are displayed in grey in Fig. 2. Some aspects of the copying scheme need to be discussed in more detail (RSK sequential residue numbers are shown in parentheses):

(1) *Region Ala:56(40) to Cys:58(42):* In the Ton structure, the Zn^{2+} binds to residue His^{57} – a key residue of the catalytic triad of all known serine proteases – and consequently this residue exhibits a perturbed side chain conformation, as shown in Fig. 3. Therefore, the corresponding RSK region had to be modelled from the PPK structure.

(2) *Region 96(90) to 103(97):* The Ton structure has residues 97 and 99 bound to the Zn^{2+} ion (see Fig. 3) and the fragment presents a very different conformation from the equivalent one in the PPK structure. In addition, residue 99(93) – a histidine in Ton but a tyrosine in PPK and RSK) – is considered a key residue in the kallikrein S2 subsite [9], thus justifying the choice of the PPK fragment as the template.

(3) *Region 213(203) to 218(208):* If this region was copied from the Ton structure, residue 215(205), another key residue in the kallikrein S2 subsite, would completely block the access of PTI to the active centre.

(4) *Residues N:25(10) and S:26(11):* These residues are in the neighbourhood of residue 116(110) which is a glycine in Ton but an aspartic acid in both PPK and RSK; since these residues have very different stereo properties, the RSK residues were copied from PPK.

(5) *Residue 61(45):* In the PPK structure this residue is in the neighbourhood of residues 35(21) and 36(22), which have equivalent positions in RSK but do not even exist in Ton. The stereo requirements are necessarily different and, consequently, both the RSK match residue and its neighbourhood were copied from PPK.

Position 236(227) is a proline in RSK, an amino acid with very specific main chain structural characteristics which cannot adequately be modelled from a different residue type and, consequently, its coordinates were generated by CHARMm.

The characteristic kallikrein loop – residues 95*A*(79) to 95*K*(89) – could not be modelled by homology because PPK exhibits a much shorter loop and the corresponding coordinates in the Ton structure are not defined (see Fig. 2). An attempt was made to model this region using a fragment search (looking for possible template peptides in a database), but the resulting loops presented several structural problems when attempts were made to dock the inhibitor PTI in the modelled enzyme. In view of this fact, we generated the coordinates for the missing fragment using CHARMm and anchored it to the ends manually. We forced the anchorage residues 79, 80 and 81 (RSK numbers) to adopt similar conformations to those of residues 95*A*, 95*B* and 95*K* in PPK (the corresponding residues from Ton would have orientated the loop in an unfavourable conformation for docking PTI).

The same five disulphide bridges known to be present in PPK and Ton were defined in RSK, for this enzyme exhibits the same 10 cysteine residues in equivalent posi-

TABLE 1
IDENTIFICATION OF THE DISULPHIDE BRIDGES ON THE [RSK:PTI] MODEL AND ON THE [PPK:PTI] CRYSTAL STRUCTURE

[RSK:PTI]	[PPK:PTI]
RSK	**PPK**
Cys:7–Cys:149	Cys:22–Cys:157
Cys:26–Cys:42	Cys:42–Cys:58
Cys:128–Cys:195	Cys:136–Cys:201
Cys:160–Cys:174	Cys:168–Cys:182
Cys:185–Cys:210	Cys:191–Cys:220
PTI	**PTI**
Cys:242–Cys:292	Cys:5–Cys:55
Cys:251–Cys:275	Cys:14–Cys:38
Cys:267–Cys:288	Cys:30–Cys:51

tions to the template structures. We then docked PTI by copying its crystallographic coordinates from the 2kai complex. Table 1 identifies all the disulphide bridges for the newly modelled [RSK:PTI] complex structure and the equivalent ones in the 2kai complex.

Regularization, a modelling technique that builds and/or relaxes structures in a specified region, was performed for each RSK residue copied from nonidentical ones, for the adjoining residues of fragments copied from the two templates, for the entire kallikrein loop and for the terminal residues of both RSK and PTI. As regularization is an energy minimization procedure that only takes into account the local geometry, bad interatomic contacts were not removed.

All hydrogens were CHARMm generated. Close contacts (1.5 Å bump cutoff distance) arose for residues F:207, F:225, R:215 and L:171 (RSK sequential numbering will be used from now on), and were promptly removed by reassigning their side chain conformations.

The regularization process imposed some small Ω angle deviations from planarity, other than those 'inherited' from the template structures, an undesirable minor side effect we found impossible to avoid. Nevertheless, the overall kallikrein conformation and structural functional characteristics were solidly preserved.

Figure 4 shows the C^{α} trace of our modelled [RSK:PTI] structure and its superposition to the template structures of Ton and PPK. Notice that the two superpositions, 4b and 4c, clearly reveal the effectiveness of the modelling scheme: the loops surrounding the RSK active site reproduce the conformational pattern of a true kallikrein, with marked differences from tonin. The kallikrein loop extends outwards, away from the docking region, and does not impede the inhibitor's approach.

Conserved buried waters

Sreenivasan and Axelsen [15] reported a study on 35 crystallographic structures of 30 different serine proteases,

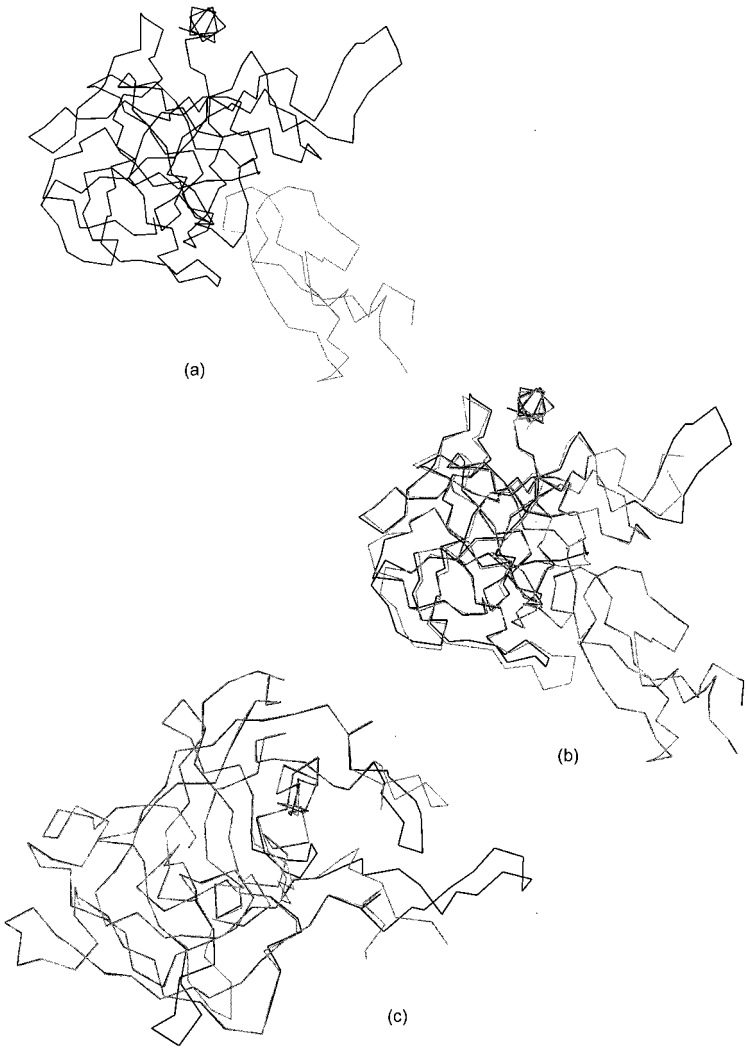


Fig. 4. C^{α} trace representation for (a) the [RSK:PTI] regularized structure; the enzyme is depicted in black and the inhibitor in grey; (b) [RSK:PTI] (black) superimposed with 2kai (grey); and (c) [RSK:PTI] (black, without the inhibitor) superimposed with 1ton (grey).

and they concluded the existence of 16 buried water molecules with conserved positions in all of the structures, plus five additional water sites associated with the S1 substrate binding region (Asp^{189}) of the trypsin-like members of that family. Such waters may play an important role in the process of folding, probably reducing the energetic cost of burying peptide groups with no hydrogen bond partners in the protein core. It was also observed that buried water protein environments (the set of non-hydrogen protein atoms within 5 Å of the water oxygen) exhibit a preponderance of main chain atoms and seem to be preserved beyond sequential homology. Therefore, our RSK model would be incomplete without the adequate set of conserved buried waters. Furthermore, the inclusion of a matrix of fixed positions (the buried water coordinates) into the protein model will help to judge its accuracy according to how well they fit in, and will impose a new type of structural constraint that otherwise would be impossible to define in MD.

Predictably, RSK must comprise 21 buried water sites. None of the two template structures has the complete set of such cavities fully occupied. Ton has 18 water molecules identified as buried, although seven of them do not completely satisfy the 'buried' criteria proposed by Sreenivasan and Axelsen [15]. The PPK structure has only three buried waters crystallographically defined (probably due to its relatively poor 2.5 Å resolution), none of which corresponds to the three missing waters in Ton. Superposition of the two template structures onto the 2ptc [23] trypsin PDB structure, in which the only missing buried water is equivalent to an existing one in Ton, allowed us to visualize the complete set of equivalent sites. In agreement with Sreenivasan and Axelsen [15], numbers 1–21 were used to identify these sites. Graphical inspection of the water cavities in the superposed structures revealed that Ton buried water assigned to position (site) 20 by the authors [15] is best assigned to position 21.

TABLE 2
RESIDUE IDENTIFIERS FOR THE TEMPLATE SET OF 21 BURIED WATERS SELECTED FROM THE 1ton AND 2ptc STRUCTURES

Water site	PDB symbol[a]	Water site	PDB symbol[a]
1	1ton:14	12	1ton:15
2	1ton:10	13	2ptc:604
3	2ptc:406	14	1ton:39
4	1ton:22	15	1ton:1
5	1ton:7	16	1ton:4
6	1ton:36	17	1ton:40
7	1ton:12	18	1ton:75
8	1ton:6	19	1ton:32
9	1ton:37	20	2ptc:541
10	1ton:17	21	1ton:66
11	1ton:51		

[a] Waters are identified by the four-character PDB symbol for the protein in which they occur and by the residue number.

Identification of the complete set of water sites in the RSK model was finally done, together with a comparative study of their protein environments to those in the template and trypsin structures, in order to choose the best coordinate matrix of buried waters. The resulting matrix comprised the 18 water molecules existing in Ton plus three selected from trypsin. Table 2 identifies them by using the PDB symbol for the system where they originally occur and the corresponding water residue number.

The matrix was added to the [RSK:PTI] modelled structure and the water hydrogen positions were generated using CHARMm. Close interatomic contacts arose for water molecules in positions 13 and 20 (both copied from the trypsin structure). As can be seen in Fig. 5, there are a few significant differences for the protein environment of the waters in trypsin compared to both RSK and Ton (the template for those regions), which explains their rather imperfect fit in the model.

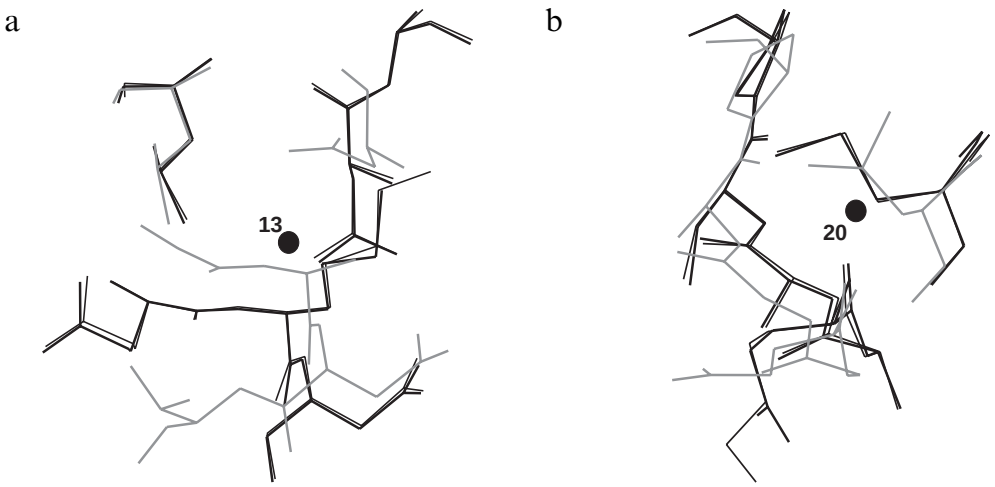


Fig. 5. Superposition of the protein environments of buried waters in sites 13 (a) and 20 (b) in structures [RSK:PTI] regularized (black thick line), 1ton (black thin line) and 2ptc (grey line). Both waters are depicted as filled circles. Hydrogens were omitted for visual simplification.

A minor translation of the two buried waters inside their RSK sites would be sufficient to solve the problem, but no attempt was made to do so manually because:

(1) Position 20 is one in the group of five associated with the S1 binding pocket (the other four are positions 17, 18, 19 and 21), and this pocket is more roomy in kallikreins than it is in trypsin – with probable effects on the size of the water cavity and relative position – thus making this particular buried water, along with the remaining four, a 'delicate' case.

(2) Position 13 is in a rather large cavity near the enzyme solvent-accessible surface, and Wade et al. [24] proved that it is possible for a buried water to be absent from a sufficiently large protein cavity, depending on the relative free energy of solvation of water in bulk solution or in the protein environment. The possible substitution of this water residue by a histidine side chain [15] does not seem to apply in this particular case, for no such substitute was found in the RSK protein environment of position 13 (nor in Ton).

Clearly, these aspects of the water structure are better handled within an MD simulation; thus, our present $[RSK:PTI]_{BW}$ (BW stands for buried waters) structural model suffered no further local rearrangements. Figure 6 shows the set of 21 conserved buried waters in the RSK model, numbered with their position identifiers, together with a Connolly molecular surface representation [25,26] of the correspondent protein cavities.

Calculation of the solvent-accessible surface area (1.4 Å probe radius, 0.05 proportionality constant) assigned all but water residue 15 as clearly buried. Position 15 was copied from the Ton matrix where it had been assigned as 'present but not buried' [15].

Structure refinement

All the refinement simulations were performed with the AMBER 4.0 [27] all-atom type force field [28], which has the explicit inclusion of H-bonding hydrogens.

An initial energy minimization of the $[RSK:PTI]_{BW}$ system was carried out, with harmonic constraints on the protein backbone (in order to preserve secondary structure), on the active-site Ser, His and Asp residues, and on the BW oxygens. The remaining close contacts were promptly removed and no undesirable structural deviations from the unrefined structure were observed. Buried waters 13 and 20 (see the previous section) were easily accommodated inside the protein without loss of consistency of the modelling procedure. The minimized energy of the model was negative and of similar magnitude to that calculated for the crystal structure of the [PPK:PTI] complex which has a similar number of residues. A control minimization was performed on the [RSK:PTI] model without the buried waters.

Whether buried waters preserved their positions under physiological conditions was to be tested with MD. Macromolecules in physiological systems are solvated. To simulate correctly the $[RSK:PTI]_{BW}$ structure in a full solvent representation with no protein intermolecular interactions, a box of water molecules greater than the sum of the complex major diameter (70 Å) and the nonbonded cutoff distance (12 Å) was needed, along with the approximate periodic boundary conditions. The computational burden of such a large system was technically forbidden. A feasible solution was to place a shell of water molecules around the complex, with no boundary conditions, so that solvent effects such as the flexibility increase of solvent-accessible residues could be accounted for [29,30]. An asymmetric cap of TIP3P 'Monte Carlo water' [27] was generated around the energy-minimized structure. A total of 430 solvent water molecules were included. Constant temperature dynamics [31] was subsequently performed for 20 ps at 500 K and then at 300 K during 100 ps, with constant dielectric function and SHAKE on bonds involving hydrogen. The average structure was calculated.

Results and Discussion

The modelled structure of rat submaxillary kallikrein bound to bovine pancreatic trypsin inhibitor consists of 295 amino acid residues (237 for the active enzyme and 58 for the inhibitor) and was homology built using rat tonin and porcine pancreatic kallikrein as template structures; the coordinates are available from the authors on request. It has one cis peptide bond at P:209, a particular feature found in the equivalent proline residue (P:219) of both template molecules [8,16], leading to the idea that this might be a unique kallikrein-like characteristic. A few other interesting general structural aspects were brought to our attention:

(1) Some nonplanar Ω angles appear in all of the four structures (see Fig. 1), the centre residue of the region where they occur being marked with an asterisk in Fig. 1; a similar study carried out on two X-ray trypsin PDB structures – 2ptc [23] and 1ppe [32] – revealed the same unexpected pattern on equivalent regions (data not shown), leading to the idea that trypsin-like serine proteases exhibit a set of conserved 'wrong' Ω values.

(2) The main structural differences between Ton and PPK structures lie on a few external nonhomologous loops surrounding the active site, namely the regions defined by residues 35–41, 95–101, 144–152, 169–177 and 215–218, which are probably involved in substrate specificity; similar results were reported by Fujinaga and James [16]. Our RSK model has been built accordingly.

A set of 21 buried water molecules was initially incorporated into the model. During MD most of the buried waters moved less than 1 Å within their respective cavities, an expected rearrangement considering the observed differences (~1–2 Å) between equivalent buried water positions among different serine proteases of known struc-

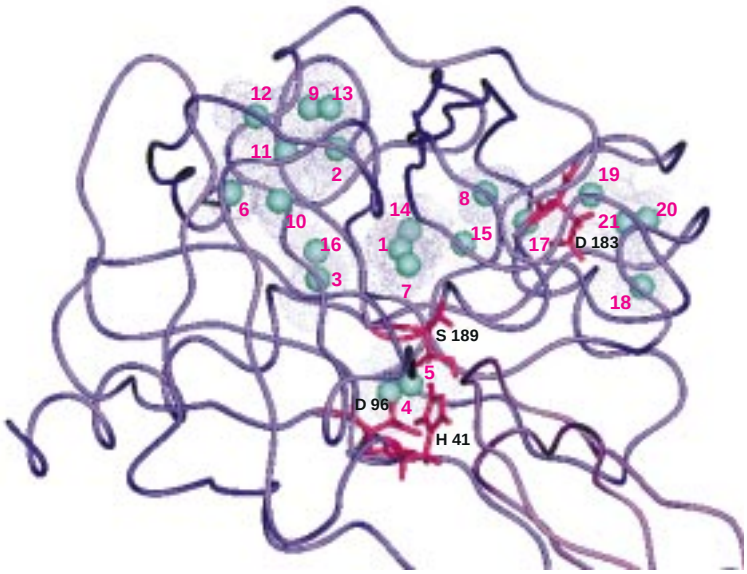
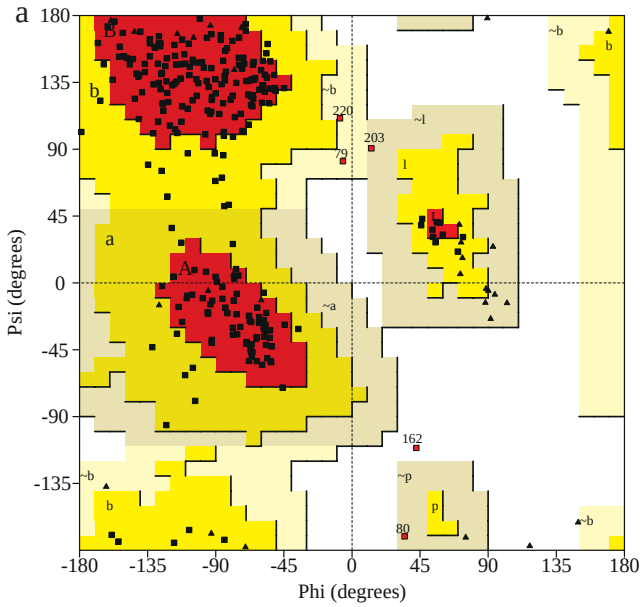


Fig. 6. Smooth C^{α} trace schematic representation of the $[RSK:PTI]_{BW}$ unrefined structure, with the set of 21 buried waters (oxygens only) in van der Waals spheres – numbered with the position identifiers – and a Connolly molecular surface of the correspondent protein cavities. The catalytic triad and the S1 key residue are fully represented (stick drawing) in red.

ture (e.g. the ones referred in the text). The exception occurred for the water molecule in position 18 which showed a comparatively large deviation (3.9 Å) from its original location. Surprisingly, the two positions occupied by this particular water (initial and after MD) are both simultaneously occupied by water molecules in the 2ptc



Residues in most favoured regions [A,B,L]	200	82.6%
Residues in additional allowed regions [a,b,l,p]	37	15.3%
Residues in generously allowed regions [~a,~b,~l,~p]	4	1.7%
Residues in disallowed regions	1	0.4%
Number of non-glycine and non-proline residues	242	100.0%
Number of end-residues (excl. Gly and Pro)	3	
Number of glycine residues (shown as triangles)	30	
Number of proline residues	20	
Total number of residues	295	

Residues in most favoured regions [A,B,L]	207	85.5%
Residues in additional allowed regions [a,b,l,p]	35	14.5%
Residues in generously allowed regions [~a,~b,~l,~p]	0	0.0%
Residues in disallowed regions	0	0.0%
Number of non-glycine and non-proline residues	242	100.0%
Number of end-residues (excl. Gly and Pro)	3	
Number of glycine residues (shown as triangles)	30	
Number of proline residues	20	
Total number of residues	295	

Fig. 7. Ramachandran maps and corresponding plot statistics for the energy-minimized structures (a) [RSK:PTI], without buried waters, and (b) $[RSK:PTI]_{22BW}$, with buried waters. Major 'unacceptable' phi/psi values are signalled with the corresponding residue numbers.

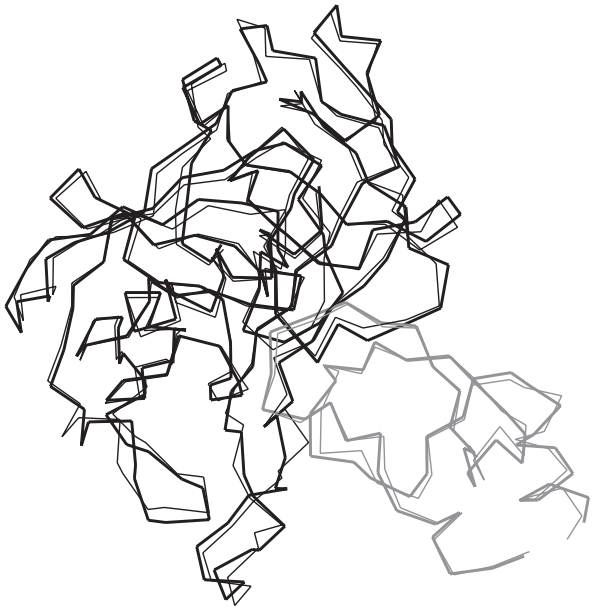


Fig. 8. C^{α} trace superposition of the $[RSK:PTI]_{22BW}$ energy-minimized structure (enzyme: black thin line; inhibitor: grey thin line) with the final MD-refined structure (enzyme: black thick line; inhibitor: grey thick line).

trypsin structure and were assigned by Sreenivasan and Axelsen [15] with identifiers 18 and 23, the latter being considered a characteristic of the trypsin only. Moreover, the PPK template structure has a water molecule (2kai: 104) that could easily be assigned to the new position. Upon further inspection, it was found that for some other well-characterized serine protease structures such as human thrombin [33], human elastase [34], bovine chymotrypsin [35] and rat mast cell protease [36], there was always a water molecule buried near the bottom edge of the S1 binding pocket (as is buried water 23 in trypsin). Consequently, it looks as if there should be such a buried water associated with the S1 cavity in serine proteases, although its exact position depends on the structural features of the particular cavity. As for the trypsin-like serine proteases, one should consider the possibility of extending the original set of 21 preserved buried waters by an extra water molecule, the new position being equivalent to Sreenivasan and Axelsen [15] identifier 23. Therefore, residue water 2ptc:414 (corresponding to position 23) should be included in Table 2 for the sake of completeness. The $[RSK:PTI]_{BW}$ model was corrected accordingly, by incorporating the new water residue, and the refinement process repeated.

After the initial energy minimization, the quality of the resulting structures, with and without the 22 buried waters, was assessed with PROCHECK [37]. The main Ramachandran maps [38] and the corresponding plot statistics are presented in Fig. 7. As can be seen, the inclusion of buried waters did in fact result in a significant structure improvement. This was particularly relevant for residues 203 and 220 (both Thr) which had their initial bad conformations corrected so that their side chains adopted more favourable

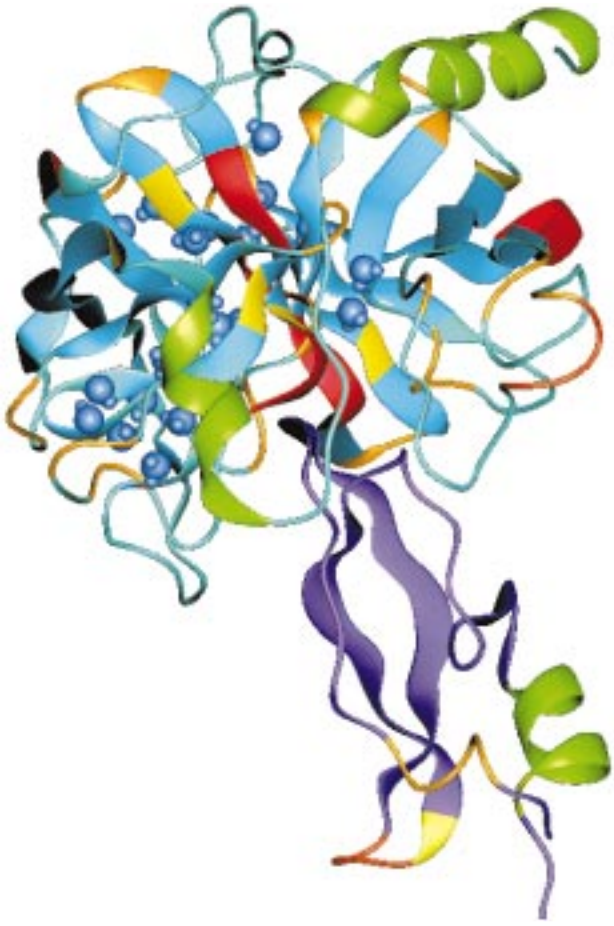


Fig. 9. Final MD-refined [RSK:PTI] complex model with the 22 conserved buried waters. Space-filled schematic representation of its secondary structure: thick helical ribbons for α-helices, thick extended ribbons for β-strands and thin ribbons for the remaining conformational features. Buried waters are represented by van der Waals spheres.

positions towards buried waters 14 and 4, respectively; moreover, the absence of buried water 4 resulted in the occupation of its cavity by the T:220 side chain.

Figure 8 shows the C^{α} trace superposition of the model structure $[RSK:PTI]_{22BW}$ before and after MD refinement. For the C^{α} atomic positions the root mean square deviation is 0.85 Å, the major differences being on the predictably flexible kallikrein loop and on a nearby solvent-accessible PTI segment. The docking residues P4 to P4' and the neighbouring regions of the enzyme hardly suffered any change. No water molecule from the bulk solvent became buried nor did it replace any of the initially buried ones. The 22 buried waters ensured a structurally preserved protein environment, thus strongly suggesting that the model is most probably a good representative structure of the enzyme. As can be seen in Fig. 9, the RSK model exhibits the secondary structure features common to the serine protease family.

Conclusions

In conclusion, one should consider a set of 22 buried waters, as opposed to the original set of 21, to be incorporated into serine protease models built on the basis of sequence/structural homology: 16 buried water molecules with conserved positions in all homologous serine proteases, plus six additional water sites for the trypsin-like members of that family.

Acknowledgements

We would like to thank JNICT (PRAXIS XXI) for a research grant to E.F.H.

References

1 Murray, S.R., Chao, J., Lin, F.K. and Chao, L., J. Cardiovasc. Pharmacol., 15 (Suppl. 6) (1990) 7.
2 Moreau, T., Brillard-Bourdet, M., Bouhnik, J. and Gauthier, F., J. Biol. Chem., 267 (1992) 10045.
3 Thomas, K.A., Baglan, N.C. and Bradshaw, R.A., J. Biol. Chem., 256 (1981) 9156.
4 Lazure, C., Seidah, N.G., Thibault, G., Genest, J. and Chretien, M., In Proceedings of the VIIth American Peptide Symposium, 1981, pp. 517–519.
5 Gutkowska, J., Thibault, G., Cantin, M., Garcia, R. and Genest, J., Can. J. Physiol. Pharmacol., 61 (1983) 449.
6 Serveau, C., Moreau, T., Zhou, G.X., ElMoujahed, A., Chao, J. and Gauthier, F., FEBS Lett., 309 (1992) 405.
7 Kettner, C., Mirabelli, C., Pierce, J.B. and Shaw, E., Arch. Biochem. Biophys., 202 (1980) 420.
8 Bode, W., Chen, Z. and Bartels, K., J. Mol. Biol., 164 (1983) 237.
9 Chen, Z. and Bode, W., J. Mol. Biol., 164 (1983) 283.
10 Ashley, P.L. and MacDonald, R.J., Biochemistry, 24 (1985) 4512.
11 Swift, G.H., Dagorn, J., Ashley, P.L., Cummings, S.W. and MacDonald, R.J., Proc. Natl. Acad. Sci. USA, 79 (1982) 7263.
12 Mosimann, S., Meleshko, R. and James, M.N.G., Proteins Struct. Funct. Genet., 23 (1995) 301.
13 Edsall, J.T. and McKenzie, H.A., Adv. Biophys., 16 (1983) 53.
14 Hubbard, S.J., Gross, K.H. and Argos, P., Protein Eng., 7 (1994) 613.
15 Sreenivasan, U. and Axelsen, P.H., Biochemistry, 31 (1992) 12785.
16 Fujinaga, M. and James, M.N.G., J. Mol. Biol., 195 (1987) 373.
17 Abola, E.E., Bernstein, F.C., Bryant, S.H., Koetzle, T.F. and Weng, J., In Allen, F.H., Bergerhoff, G. and Sievess, R. (Eds.) Protein Data Bank in Crystallographic Databases – Information Content, Software Systems, Scientific Applications, Data Commission of the International Union of Crystallography, Bonn/Cambridge/Chester, 1987, pp. 107–132.
18 Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F., Brice Jr., M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T. and Tasumi, M., J. Mol. Biol., 112 (1977) 535.
19 QUANTA (v. 4.0), Molecular Simulations Inc., Burlington, MA, U.S.A., 1994.
20 Wilmott, C.M. and Thornton, J.M., Protein Eng., 3 (1990) 479.
21 Kraut, H., Frey, E.K. and Werle, E., Hoppe-Seyler's Z. Physiol. Chem., 189 (1930) 97.
22 Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S. and Karplus, M., J. Comput. Chem., 4 (1983) 187.
23 Marquart, M., Walter, J., Deisenhofer, J., Bode, W. and Haber, R., Acta Crystallogr., B39 (1983) 480.
24 Wade, R.C., Mazor, M.H., McCammon, A. and Quiocho, F.A., Biopolymers, 31 (1991) 919.
25 Connolly, M.L., Science, 221 (1983) 709.
26 Connolly, M.L., J. Appl. Crystallogr., 16 (1983) 548.
27 Pearlman, D.A., Case, D.A., Caldwell, J.C., Seibel, G.L., Singh, U.C., Weiner, P. and Kollman. P.A., AMBER 4.0, University of California, San Francisco, CA, U.S.A., 1991.
28 Weiner, S.J., Kollman, P.A., Case, D.A., Singh, U.C., Ghio, C., Alagona, G., Profeta Jr., S. and Weiner, P., J. Am. Chem. Soc., 106 (1984) 765.
29 Van Gunsteren, W.F. and Karplus, M., Biochemistry, 21 (1982) 2259.
30 Brooks III, C.L. and Karplus, M., J. Mol. Biol., 208 (1989) 159.
31 Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.C., DiNola, A. and Haak, R., J. Chem. Phys., 81 (1984) 3684.
32 Bode, W., Greyling, H.J., Huber, R., Otlewski, J. and Wilusz, T., FEBS Lett., 242 (1989) 285.
33 Rehse, P.H., Steinmetzer, T., Li, Y., Konnishi, Y. and Cygler, M., to be published.
34 Bode, W., Wei, A.Z., Huber, R., Meyer, E., Travis, J. and Neumann, S., EMBO J., 5 (1986) 2453.
35 Wang, D., Bode, W. and Huber, R., J. Mol. Biol., 185 (1985) 595.
36 Remington, S.J., Woodbury, R.G., Reynolds, R.A., Matthews, B.W. and Neurath, H., Biochemistry, 27 (1988) 8097.
37 Laskowski, R.A., McArthur, M.W., Moss, D.S. and Thornton, J.M., J. Appl. Crystallogr., 26 (1993) 283.
38 Ramakrishnan, C. and Ramachandran, G.N., Biophys. J., 5 (1965) 909.