Acta
Biochimica
Polonica

Vol. 44 No. 3/1997
549–556
QUARTERLY

Energy minimization of globular proteins with solvent effects included. Comparison of empirical solvation energy terms and explicit water treatment*

Andrzej Kierzek and Piotr Zielenkiewicz✉

Institute of Biochemistry and Biophysics, Polish Academy of Sciences,
A. Pawińskiego 5a, 02-106 Warsaw, Poland

Key words: accessible surface area, atomic solvation parameters, water shell

The effect of an empirical solvation energy term on energy minimization of ribonuclease T1 was established using different sets of Atomic Solvation Parameters. The results are compared to minimization *in vacuo* and in a 10 Å water shell. The best solvent model as judged from the comparison to the crystal structure was an empirical solvation potential derived from free energies of transfer of amino-acid side-chain analogues from vapour to water. The use of this model causes, however, energy and gradient oscillations, which make it inapplicable with standard protocols of molecular dynamics simulations. The empirical solvation model which was found by other authors (von Freyberg *et al.*, 1993, *J. Mol. Biol.* 233, 275–292) to give good results in the NMR structure refinement led to distortions of the ribonuclease native structure. The model based on statistical analysis of crystal structures did not perform better than minimization *in vacuo*.

Interaction of protein atoms with solvent is commonly belived to be a main force in protein folding as well as in stabilization of the native structure [1, 2] and that is why solvent effects should be included in any model which attempts to give a realistic picture of the protein molecule on the atomic level. In the case of molecular mechanics and dynamics simulations the most obvious way is to add water molecules explicitly to the system modelled. Many simulations of this kind have been reviewed by Daggett & Levitt [3] and, as it is described there, each of them meets two basic difficulties. First, the preparation of the system is extremely complex. There are several models of water and one should test which of them provides the best prediction of physico-chemical properties for a given system. This may require calculation of water diffusion coefficient and careful analysis of water structure after test experiments. Application of periodic boundary conditions requires the system to be neutral and therefore ion molecules must be usually added to screen charged groups of the protein. The second difficulty of explicit water treatment

---

*This work was in part supported by a grant 6 6380 9203 from the State Committee for Scientific Research.

✉Corresponding author: Piotr Zielenkiewicz, tel: (+22)+658-47-03, fax: (+39)+12-16-23, E-mail: piotr@ibbrain.ibb.waw.pl

Abbreviations: ASA, accessible surface area; ASP, atomic solvation parameter; CVFF, consistent valence force field; DK, Delarue & Koehl parameter set; MD, molecular dynamics; r.m.s., root mean square.

is of course the computational cost of force field calculations for a system containing a few thousand atoms. It is also clear that the so called hydrophobic interactions are due to their entropic nature, difficult to be modeled with the use of atom-atom potentials.

In order to deal with these problems an alternative approach is proposed. In agreement with the formalism which was first introduced by Eisenberg & McLachlan [4], it is assumed that the contribution of each protein atom to the solvation energy is proportional to its accessible area, and the total solvation energy is the sum of these contributions. This is expressed by the following equation:

$$E_{hyd} = \sum_{j} \mathrm{ASP}_j \mathrm{ASA}_j$$

where ASA_j is the accessible surface area of j-th atom and ASP_j denotes atomic solvation parameter specific for the atom type. It is possible to simulate the solvent effect by using E_{hyd} as an additional force field term. This leads to a decrease in the number of atoms in the system and provides a very simple model of solvation. Efforts were also made toward the development of programs dedicated to rapid estimation of gradients of accessible surface area which is the most time-consuming step if analytical methods [5] are in use. Le Grand & Merz [6] reported an algorithm which estimates ASA gradients in the time-scale of calculation of other force field terms. This algorithm applied to calculations of solvation term for molecular dynamics should increase the time-scale of simulations.

Several tests of the approach presented above have been made. Several solvation potentials were added to the ECEPP/2 program [7, 8] and tested in minimization of 39 near native structures of bovine trypsin inhibitor. The solvation potential defined by ASP values fitted to vapour to water transfer energies of small aliphatic and aromatic compounds [9] proved among the best in the light of r.m.s. deviations of minimized conformations from the crystal structure.

In 1992 Wesson & Eisenberg [10] performed a 110 ps MD simulation of mellitin peptides adding an empirical solvation energy term to the CHARMM program. They developed a new set of parameters based on vapour to water transfer energies of amino-acid side chain analogues. The authors reported better performance of the new approach as compared to *in vacuo* simulations, in prediction of behaviour of mellitin monomers and tetramers in solvent. The same solvation potential was also added to the AMBER force field and tested on alanyl dipeptide [11].

von Freyberg *et al.* [12] tested solvation potentials in the energy refinement of bovine trypsin inhibitor and tendamistat NMR structures. They have found that minimization of the ECEPP/2 energy plus accessibility of all atoms or apolar atoms are superior to minimization *in vacuo*. In contrast to the results of Vila *et al.* [7] solvation terms based on the free energies of transfer performed poorly.

Delarue & Koehl [13] proposed a parameter set based on statistical analysis of a representative set of protein structures. For each atom type and each accessibility range they have calculated frequency with which atoms of this type have the value of accessibility within the range x. Using the Boltzman-device [14] formalism they have converted frequencies into energies obtaining for each atom type an energy *vs* ASA plot. Atomic solvation parameters were fitted to reproduce this function.

Recently Fraternali & van Gunsteren [15] added a mean solvation model to the GROMOS force field. Instead of using empirical atomic solvation parameters they fitted parameters to reproduce potentials of interaction between protein and explicit water molecules as calculated by GROMOS force field. The model gave similar results as an explicit solvation model in molecular dynamics of several middle size proteins.

As one can see from the brief review presented above, only Fraternali & van Gunsteren [15] compared their mean solvation model with explicit water treatment. The aim of this work is to provide an analogous comparison for other sets of parameters. This addresses the question whether atomic solvation parameters developed according to empirical data, and shown to perform well in aggregation free energy calculations [16] and knowledge-based protein structure modelling [13], can also be useful in molecular dynamics simulations and protein structure refinement.

The aim of this work is to compare the performance of solvation models which use atomic solvation parameters derived from empirical data to explicit water models. As the test experiment, minimization of the ribonuclease T1 crystal structure (Protein Data Bank [17], entry 9RNT) was chosen. In contrast to the works mentioned above, this one tests energy minimization as a tool for obtaining equilibrium structure (also as a starting point for further dynamics simulations) rather than the method of structure refinement. It also shows how forces generated by different solvation models act on the structure of a medium size globular protein.

Results of the following calculations were compared:

- 1) Minimization *in vacuo* with respect to a Consistent Valence Force Field (CVFF) [18] potential.
- 2) Minimization of ribonuclease structure surrounded by 1437 water molecules forming a 10 Å shell, with respect to the CVFF potential.
- 3) Minimization with respect to the CVFF potential plus E_{hyd} as defined by Wesson & Eisenberg [10] ASP set (WWE set).
- 4) Minimization with respect to the CVFF potential plus solvation potential calculated using a Delarue & Koehl [13] parameter set (DK set).
- 5) Minimization with respect to the CVFF potential and total accessible surface area (TOTAL model [12]).

Since the aim of this study was energy minimization rather than molecular dynamics, a simple shell model of explicit solvent is sufficient, and more sophisticated models with periodic boundary conditions are not required.

For minimization *in vacuo* and with explicit water molecules a steepest descent minimizer followed by conjugate gradient minimization have been used. The systems, with empirical solvation potentials added, were minimized using the steepest descent method in order to decrease gradient oscillations which appeared during the calculations. We have also checked whether it is possible to start the molecular dynamics simulations using relaxed structures.

Three sets of ASP used in this work are shown in Table 1. They represent three different classes of parameters:

- i) parameters fitted to vapour-water transfer energies,
- ii) structure based parameters,

Table 1. Atomic solvation parameters and van der Waals radii used in this work

	WWE set[a]	DK set[b]	TOTAL set[c]	Radii[d]
C	0.012	0.00092	1.0	2.00
Carboxyl or Carbonyl C	0.012	0.00092	1.0	1.50
Aromatic C	0.012	0.00092	1.0	1.85
Side chain N	−0.116	−0.00125	1.0	1.50
Main chain NH	−0.116	0.00048	1.0	1.50
NH3	−0.186	−0.00125	1.0	1.50
Charged O	−0.175	−0.004	1.0	1.40
Main chain O	−0.116	0.00048	1.0	1.40
Hydroxyl O	−0.116	−0.00125	1.0	1.40
Sulphur	−0.018	0.00092	1.0	1.85

All atomic solvation parameters in Table are expressed in kcal/mol per Å^2. [a] Solvation parameters from Table 3 of Wesson & Eisenberg [10]. [b] Solvation parameters from Fig. 4 of Delarue & Koehl [13]. [c] Solvation parameters from Table 2 of von Freyberg *et al.* [12]. [d] The van der Waals radii of united atoms from Table 2 of Shrake & Rupley [21].

- iii) parameters arbitrarily assigned to improve the energy refinement of NMR structures.

METHODS

Gradients of accessible surface area were calculated using the ANAREA program [5]. In the case of WWE, DK and TOTAL models a procedure which multiplies gradients by atomic solvation parameters from a given set was added by the authors. Consistent Valence Force Field terms have been calculated by the DISCOVER program (ver. 94.0 Biosym). At each step of the calculation the ANAREA program was run by the DISCOVER program as the external procedure and gradients of ASA multiplied by ASP values were added to the force field gradients by the DISCOVER program.

In all calculations with water explicitly treated, the dielectric constant was set at 1. In other calculations it was set at 80. The cutoff for the non-bonded interactions was 9.5 Å. The switching function with a spline width 0.5 Å and a buffer width 1.0 Å was used to turn off interactions smoothly. In order not to divide groups, one atom from each group was chosen as a representative one for the cutoff calculations. The list of the interacting atoms was updated whenever one of these atoms moved more than half of the cutoff distance.

Systems were prepared and results analyzed with InsightII molecular graphics program (Biosym).

RESULTS AND DISCUSSION

Convergence of minimization

Minimization *in vacuo* reached a gradient lower than 0.001 kcal/mol after 8883 steps. In the case of minimization with explicit water molecules it took 28803 iterations due to the large number of atoms in the system.

Experiments with WWE and DK solvation terms were stopped after about 20000 iterations because no further decrease of the energy was observed. The final structures have gradient values of 3.027 kcal/mol for the WWE model and 0.2 kcal/mol for the DK model.

After the first 200 steps of minimization with the TOTAL potential included, it was clear that the crystal structure of ribonuclease is going to be distorted. It will be shown below that the structure collected after 275 steps (gradient value 0.5 kcal/mol) significantly and unrealistically differs from the native one.

For each model 20 ps molecular dynamics was run in order to check whether the system is stable enough to perform it. In the case of the WWE model the molecule unfolded during the first steps of equilibration. For other models the calculations have been completed without distortion of the native structure.

In order to study reasons for poor performance of the WWE model, additional 1000 steps of minimization were performed. During this experiment conformations were collected after each 100 steps. Figure 1 shows an energy plot *vs* iteration number. The energy increased from 709 kcal/mol to 772 kcal/mol and then decreased to 739 kcal/mol. This shows that strong energy oscillations appeared during the minimization. Due to these oscillations it was not possible to perform MD simulation. If one takes into account that minimization with respect to the CVFF potential alone reached convergence after 8883 steps, it is clear that the standard minimization method which was used is not sufficient for minimization of the system with respect to the CVFF potential plus the solvation energy term modeled by the WWE parameter set.

Minimization experiments were stopped before they reached convergence also by von Fryberg *et al.* [12]. They set the iteration limit arbitrarily to 400 iterations. Wesson & Eisenberg [10] performed molecular dynamics simulations starting from the structures optimized without the solvation term. We have also additionally tried to start the simulation from the structure optimized with respect to CVFF potential alone, with the WWE solvation term added at the beginning of the MD simulation. This resulted in distortion of the structure after the first steps of equilibration. Thus MD simulation of tetramer of pep-

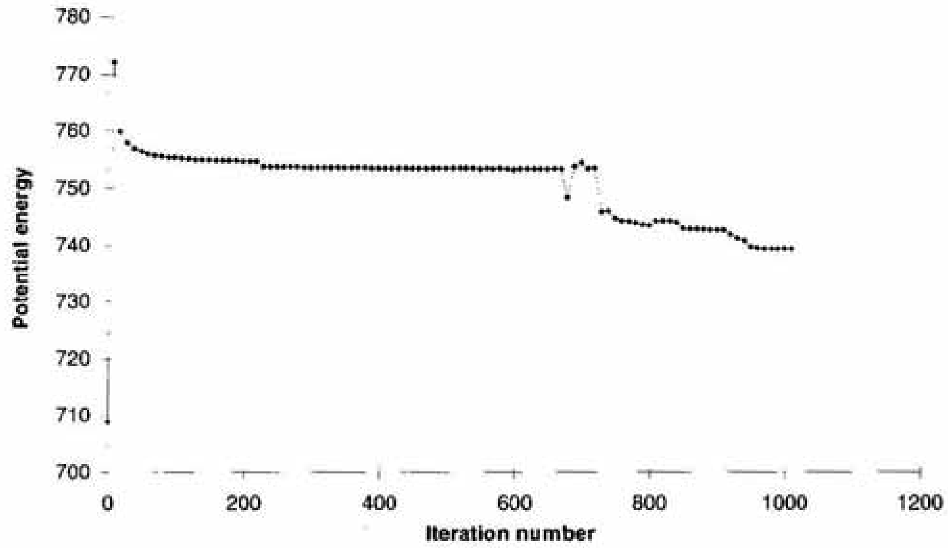

Figure 1. Oscillations of the CVFF + + WWE potential energy during the last 1000 steps of minimization.

The picture shows additional 1000 iterations of minimization of ribonuclease T1 performed in order to check reasons of the poor minimization performance of the CVFF + WWE potential energy function. One can clearly see strong oscillations of the energy value.

tides (the system used by Wesson & Eisenberg [10]) is a task quite different from simulation of dynamics of true globular protein.

The results presented above imply that application of the empirical solvation energy term modeled by the WWE parameter set, to MD simulations of globular proteins requires development of a special optimization strategy.

r.m.s. comparison for whole structures

Table 2 shows r.m.s. deviations of heavy atoms between each pair of structures. A structure closest to the crystal one was obtained by minimization of the CVFF+WWE potential. Explicit water treatment also provides a good approximation to the native structure. These two structures are similar, and both are significantly more similar to the crystal one than the conformation obtained by minimization *in vacuo*. In contrast, application of the DK potential did not improve *in vacuo*. The reason seems to be obvious — DK parameters are lower by three orders of magnitude than WWE ones, and forces provided by the DK model are too weak to change the structure significantly.

r.m.s. deviation greater than 2 Å between the crystal structure and the conformation obtained by minimization of the CVFF + + TOTAL potential indicates distortion of the native structure caused by this model.

Conservation of the beta-sheet structures

Ribonuclease T1 has 7 extended chain fragments, stretching between residues 4–6, 9–12, 39–42, 55–62, 76–81, 85–91, 100–103 [19, 20]. The first two fragments form a short two stranded antiparallel β-sheet, and a second antiparallel sheet is composed of the five remaining ones. As it is clear by visual inspection with molecular graphics programs, most of these strands are buried inside the ribonuclease globule.

The lowest r.m.s. deviation of the heavy atom positions belonging to β-sheets as compared with the crystal structure, was 0.46 Å and it was obtained by minimization with respect to the CVFF+WWE potential. The water shell model also conserved β-sheets better than the CVFF model alone. r.m.s. deviations were 0.56 and 0.67, respectively. Addition of the DK potential did not improve minimization significantly. The r.m.s. value for this experiment was 0.65. Minimization of the CVFF+TOTAL potential resulted in the r.m.s. deviation being 1.61, which was the highest change of the β-sheet conformation.

Conservation of the helix structure

There is one alpha helix structure in ribonuclease T1 located between residues 13–29. The structure of the helix is best conserved by the water shell model with the r.m.s. deviation value of 0.37. Addition of the WWE

Table 2. r.m.s. deviations (in Å) of heavy atoms for all pairs of structures

	Crystal structure	CVFF	Water shell	TOTAL	DK
CVFF	1.14				
Water shell	0.81	1.09			
TOTAL	2.56	2.54	2.68		
DK	1.01	0.99	0.99	2.59	
WWE	0.72	1.05	0.83	2.92	0.90

solvation model resulted in r.m.s. deviation equal to 0.43, close to that of minimization *in vacuo*, which ended with r.m.s. deviation of 0.48. Addition of the DK solvation potential resulted in r.m.s. deviation value of 0.52 and thus provided a worse approximation to the crystal structure than minimization of CVFF potential alone. The CVFF+TOTAL potential provided a significantly worse model than any of the remaining ones, producing the r.m.s. deviation of 1.40.

Hydrogen bond patterns

In the crystal structure of ribonuclease T1 there are 81 formal hydrogen bonds for which the donor-acceptor distance is less than 2.5 Å and the acceptor-hydrogen-donor angle is greater than 120°. In each of the structures resulting from minimization, formal hydrogen bonds meeting these requirements have been found. To express statistically conservation of hydrogen bonding patterns two criteria were used:

- ✦ 1) The percentage of hydrogen bonds from the crystal structure which are present in the simulated structure.
- ✦ 2) The percentage of hydrogen bonds in the simulated structure which are also present in the crystal one.

In Table 3 the application of both criteria to the whole structures and main chain atoms only are presented.

Minimization with water explicitly treated provided a structure with a similar number of hydrogen bonds as in the crystal one, and the highest percentage of the H-bonds present in the native structure. The CVFF+ + WWE model yielded a structure with the lowest number of hydrogen bonds and low percentage of the H-bonds present in a crystal structure. This concerns the whole structure as well as main chain atoms, although the pattern of the H-bonds in the main chain is better conserved in all structures. Criterion 2 judges how many H-bonds in the given structure are native-like, providing information how often a given model fails in prediction of native H-bonds. By means of this criterion the CVFF+WWE model gave certainly the best results. Visual inspection of the structures indicates that hydrogen bonds formed mainly by the side chains on protein surface are affected by application of the WWE potential.

The native hydrogen bonding pattern is clearly strongly distorted when the CVFF + + TOTAL forces are applied.

How do the forces act?

The results presented above show which forces acting on protein atoms are generated by the models, and how they influence the structure of ribonuclease T1. It is clear that the TOTAL potential providing strong packing forces, acting equally on all atoms, simply "compresses" the native structure. Such forces are also present in minimization *in vacuo* although at a much smaller magnitude. They move the side chains protruding into the solvent towards the core of the protein. The DK potential provides the forces which distinguish atoms which are frequently present on the surface of known protein structures from those which are usually buried, but these forces are too weak to equilibrate the packing forces resulting from the vacuum environment of the protein. Both WWE and explicit solvent models keep side

Table 3. Hydrogen bonding patterns

	Crystal structure	CVFF + +DK	CVFF	Water shell	CVFF + +WWE	CVFF + +TOTAL
Number of H-bonds	81	53	56	82	46	65
Number of H-bonds in main chain	51	35	34	53	34	33
Criterion 1[a] for all atoms	100%	51%	54%	72%	48%	35%
Criterion 1 for main chain only	100%	62%	68%	84%	64%	56%
Criterion 2[b] for all atoms	00%	79%	78%	71%	86%	48%
Criterion 2 for main chain only	100%	91%	97%	94%	100%	33%

[a] The percentage of hydrogen bonds from crystal structure which are also present in structure analyzed. [b] The percentage of hydrogen bonds in structure analyzed which are present in the crystal one.

chains protruding into the solvent. They act differently on hydrophobic and hydrophilic atoms with forces strong enough to influence minimization. The WWE solvation model provides definitely the best approximation to the crystal structure, which means that forces acting on the structure are almost equivalent. The model exhibits a slight trend to unpack the structure, which is expressed by lower conservation of the surface-exposed helix than in the case of the explicit solvent model.

von Freyberg *et al.* [12] have reported that the WWE model fails in refinement of partially unfolded structures since it does not provide sufficiently strong packing forces. In turn, this work shows that the TOTAL model, which performed well in structure refinement of partially unfolded structures, distorts the native one. The technical implication is that the structure refinement and minimization as the first stage of MD simulations require a different approach when the effect of environment was to be included. From the theoretical point of view one should notice that forces generated by the WWE model are based on experimental measurements and that they conserve the native structure, which suggests that they provide a good approximation for interaction with water. This probably means that packing forces provided by the solvent are so weak that they may influence the structure only over large time scales, such as the time scale of protein folding.

CONCLUSIONS

✦i) Among the models investigated the empirical solvation model with the WWE parameter set provides the best conservation of the crystal structures. Thus its use in molecular dynamics seems very promising, but due to instabilities caused by this model specially designed protocols should be developed.

✦ii) The TOTAL parameters set, giving the best results in NMR structure refinement, gives poor results in energy minimization.

✦iii) Energy contributions provided by the DK solvation model are too low in comparison with other force field terms, to influence minimization significantly.

All calculations were performed on SGI Challenge L computer (R4400, 150 Mhz) at Institute of Biochemistry and Biophysics.

We are grateful to Dr. D. Płochocka for critical reading of the manuscript.

REFERENCES

1. Kauzmann, W. (1959) Some factors in the interpretation of protein denaturation. *Adv. Prot. Chem.* **14**, 1–63.
2. Dill, K.A. (1990) Dominant forces in protein folding. *Biochemistry* **29**, 7133–7155.
3. Daggett, V. & Levitt, M. (1993) Realistic simulations of native-protein dynamics in so-

lution and beyond. *Annu. Rev. Biophys. Biomol. Struct.* **22**, 353–380.

4. Eisenberg, D. & McLachlan, A.D. (1986) Solvation energy in protein folding and binding. *Nature* **319**, 199–203.
5. Richmond, T.J. (1984) Solvent accessible surface area and excluded volume in proteins. Analytical equations for overlapping spheres and implications for the hydrophobic effect. *J. Mol. Biol.* **178**, 63–89.
6. LeGrand, S.M. & Merz, K.M. (1993) Rapid boolean approximation to molecular surface areas *via* the use of boolean logic and look-up tables. *J. Comp. Chem.* **14**, 349–352.
7. Vila, J., Williams, R.L., Vasquez, M. & Scheraga, H.A. (1991) Empirical solvation models can be used to differentiate native from near-native conformations of bovine pancreatic trypsin inhibitor. *Proteins Struct. Funct. Genet.* **10**, 199–218.
8. Williams, R.L., Vila, J., Perrot, G. & Scheraga, H.A. (1992) Empirical solvation models in the context of conformational energy searches: Application to bovine pancreatic trypsin inhibitor. *Proteins Struct. Funct. Genet.* **14**, 110–119.
9. Ooi, T., Ootobake, M., Nemethy, G. & Scheraga, H.A. (1987) Accessible surface areas as a measure of the thermodynamic parameters of hydratation of peptides. *Proc. Natl. Acad. Sci. U.S.A.* **84**, 3086–3090.
10. Wesson, L. & Eisenberg, D. (1992) Atomic solvation parameters applied to molecular dynamics of proteins in solution. *Protein Sci.* **1**, 227–235.
11. Schiffer, C., Caldwell, J.W., Stroud, R.M. & Kollmann, P.A. (1992) Inclusion of solvation free energy term with molecular mechanics energy: Alanyl dipeptide as a test case. *Protein Sci.* **1**, 396–400.
12. von Freyberg, B., Richmond, T.J. & Braun, W. (1993) Surface area included in energy refinement of proteins. A comparative study of atomic solvation parameters. *J. Mol. Biol.* **233**, 275–292.
13. Delarue, M. & Koehl, P. (1995) Atomic environment energies in proteins defined from statistics of accessible and contact surface areas. *J. Mol. Biol.* **249**, 675–690.
14. Sippl, M.J. (1990) Calculation of conformational ensembles from potential of mean force. An approach to the knowledge-based prediction of local structures in globular proteins. *J. Mol. Biol.* **213**, 859–883.
15. Fraternali, F. & van Gunsteren, W.F. (1996) An efficient mean solvation force model for use in molecular dynamics simulations of proteins in aqueous solution. *J. Mol. Biol.* **256**, 939–948.
16. Horton, N. & Lewis, M. (1992) Calculation of the free energy of association for protein complexes. *Protein Sci.* **1**, 169–181.
17. Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977) The protein data bank: A computer-based archival file for macromolecular structures. *J. Mol. Biol.* **112**, 535–542.
18. Maple, J.R., Dinur, U. & Hagler, A.T. (1988) Derivation of force fields for molecular mechanics and dynamics from *ab initio* energy surfaces. *Proc. Natl. Acad. Sci. U.S.A.* **85**, 5350–5354.
19. Martinez-Oyanedel, J., Choe, H.-W., Heinemann, U. & Saenger, W. (1991) Ribonuclease T_1 with free recognition and catalytic site: Crystal structure analysis at 1.5 Å resolution. *J. Mol. Biol.* **222**, 335–352.
20. Pace, C.N., Heinemann, U., Hahn, U. & Saenger, W. (1991) Ribonuclease T1: Structure, function and stability. *Angew. Chem. Int. Edn. Engl.* **30**, 343–360.
21. Shrake, A. & Rupley, J.A., (1973) Environment and exposure to solvent of protein atoms. Lysozyme and insulin. *J. Mol. Biol.* **79**, 351–371.