PROTEINS: Structure, Function, and Genetics 28:303–312 (1997)

A Multicopy Modeling of the Water Distribution in Macromolecular Crystals

Alberto D. Podjarny,[1]* Eduardo I. Howard,[2] Alexander Urzhumtsev,[1] and J. Raul Grigera[2]
[1]*IGBMC, UPR de Biologie Structurale, 67404 Illkirch, Cedex, France*
[2]*Instituto de Fisica de Liquidos y Sistemas Biologicos (IFLYSIB), Universidad de La Plata, 1900 La Plata, Argentina*

ABSTRACT **A multicopy protocol is proposed for modeling macromolecular hydration using diffraction experimental data (X-ray or neutron) to search for a better description of delocalized water sites than that given by point water models. The model consists of one macromolecule and several copies of each water molecule, refined simultaneously against diffraction data using molecular dynamics techniques. The protocol was applied to BPTI and an RNA tetradecamer. The sites defined by the different copies range from very ordered ones to continuous channels; they fit the density maps and agree with the diffraction amplitudes with an accuracy comparable with usual crystallographic methods. The delocalization of water in channels agrees with the high mobility observed in NMR experiments. Proteins 28:303–312, 1997** © 1997 Wiley-Liss, Inc.

Key words: water structure; macromolecular hydration; crystallographic refinement; multicopy model

INTRODUCTION

Native biomacromolecules are hydrated, and the presence of water in the medium of the "living" biomacromolecules is an essential condition of life. The main objectives of hydration studies concentrate on the properties of water molecules in the surroundings of the macromolecule. More precisely, the goal in the field is the description of the structural and dynamic properties of water around the biomacromolecule. Crystals are an appropriate case for these studies, because practically all water molecules are close to a biomacromolecule.

Crystallographic studies using X-ray and neutron diffraction have provided a large quantity of macromolecular models.[1] These models include well-defined water sites, reflected in the density maps as well-defined peaks, and in the final molecular model as atoms with low thermal Debye-Waller factor (B-factor). These highly localized sites are defined as "bound water" molecules. Because these molecules are in exchange with the bulk, the terms bound water or ordered water do not describe the real situation. Specific hydration site[2] or strongly preferred hydration site[3,4] seems to be a better name to describe the system.

Reliable assignment of all water molecules present in macromolecular crystals is not evident, because disordered solute atoms or noise may be easily taken as a water peak. Very high resolution studies, particularly those using neutron diffraction, are the most accurate ones. The water molecules are usually represented as "point models" of hydration. However, these very precise "frozen" sites may not be a good model for all water molecules, because many of them have some degree of mobility and therefore can give rise to continuous density. A recent report[5] using experimental phasing at high resolution gave an accurate description of hydration, showing, in particular, continuous density for water clusters. Such continuous density had already been observed in vitamin B12.[6]

The resulting problem of accurately determining the point model water positions has been treated extensively. Reports comparing either different determinations of the same molecule[3,7] or the application of different refinement programs to the same structure determination[8] showed the limitations of the point models for weak sites, because only a fraction of the sites agree. Furthermore, high-resolution studies on crambin at 130 K showed the existence of multiple networks, which were refined simultaneously. The alternative water positions, in different networks, were individually assigned.[9]

Nuclear magnetic resonance has been used also to obtain experimental information of water distribution around macromolecules. It has been applied to thin fibers,[10,11] crystalline powders,[12,13] single crystals,[14] and solutions.[15–17] The model emerging from these experiments accounts for only a few water molecules in specific hydration sites in exchange with other molecules "moving around." The exchange rate between the two fractions is often of the order of 10^{-6} s. The nonspecific hydration water has much larger mobility even in the crystals (10^{-10} s),

Permanent address for A. Urzhumtsev: IMBP RAN, Puschino, Moscow Region, 142297 Russia.
*Correspondence to: Alberto D. Podjarny, IGBMC, UPR de Biologie Structurale, BP 163, 1 rue Laurent Fries, 67404 Illkirch Cedex, France.
E-mail: podjarny@igbmc.u-strasbg.fr
Received 18 October 1996; Accepted 14 February 1997

© 1997 WILEY-LISS, INC.

and it approaches the rotation time of bulk water (10^{-11} s). This view is reinforced by recent nuclear magnetic relaxation dispersion (NMRD) studies,[18] which detect the existence of a small number of long-lived (10^{-8} to 10^{-5} s) and well-defined water molecules trapped in cavities or deep clefts. The nuclear magnetic resonance view of the system shows high residence times only for the most ordered water molecules detected by crystallography.[17] These specific hydration sites are usually buried or in deep clefts, whereas most surface waters are highly mobile.

Alternatively, the problem can be approached by using a purely modeling method, such as molecular dynamics simulation.[19] Within this scope, the description depends heavily on the reliability of the theoretical model, mainly the nature of the atom-atom interactions and the water model selected. The output of the method is not a point model but water trajectories, i.e., the position of the water molecules as a function of time. Analysis of the modeling results agrees with NMR because most surface waters are highly mobile with residence times lower than 200 ps.[20]

Most surface waters, which are difficult to assign crystallographically, have a high mobility according to NMR and modeling studies. Levitt and Park Britt[21] have pointed out that because the crystallographic signal comes from a very long time average, a water peak does not correspond to a single molecule but to a site that is occupied by different ones. This concept was applied to a specific case by Lounnas and Pettitt.[22] By comparing MD simulations and X-ray structures, they showed that the crystallographic "point model waters" correspond to "hydration sites" obtained by averaging the MD trajectories. This average produces the model water distribution as a function of position. The corresponding map resembles an electron density and, in fact, will be proportional to it, if the model describes the system properly. An attempt to merge modeling and X-ray data in a single simulation, to describe dynamically the water, was done by Grigera and coworkers.[23] In that work, calculated amplitudes were obtained from time averaged MD trajectories, and the agreement with observed amplitudes was used to introduce new forces. This modeling method is currently under study; in the present state, the crystallographic forces tend to freeze the position of water molecules, as happens during a crystallographic refinement. The method is currently time-consuming and in many cases cannot cover an appropriate fraction of the configurational space.

In a real crystal, a given hydration site can be occupied by different water molecules, whose position varies either in time or between different unit cells. Diffraction amplitudes from these sites can be simulated either by time or space averaging; the results should give equivalent information. Because averaging MD trajectories in time to generate calculated amplitudes posed the problems described above, space averaging by the multicopy method[24] was used.[25] In this method, a model is built, assuming that all unit cells are equal and including one solute molecule and several copies of each water molecule. All copies together describe the hydration site, and their dispersion gives an indication of the site disorder. The joining of copies of adjacent sites can be interpreted as an indication of a high degree of mobility between these sites. To adjust the model to the diffraction amplitudes, a slow cooling protocol is used, with the usual modeling forces and the addition of a crystallographic one, coming from the experimental diffraction data. The balance between model and diffraction forces, the time and the temperature of the simulation should be such that, in most of cases, one set of copies models one hydration site and corresponds to one feature of the difference map. Interatomic interactions are calculated only inside a given copy and between water molecules and the macromolecule. Because many variables are introduced, care has to be taken to avoid overinterpretation. This can be monitored by calculating the R-free factor, which should not increase.[26]

In this work the multicopy method has been applied to two different macromolecular crystals. The expectation is to obtain a model for hydration that is fully compatible with the experimental diffraction data and gives an accurate description of surface waters. The compatibility with amplitudes means that the outcoming model should have the R factor and the R-free factor[26] equal or less to that of the point model and that the model should fit density maps. Surface waters should be in extended sites, agreeing with the mobility observed in simulations and NMR studies.

METHODS

Multicopy Modeling Protocol

To apply the multicopy method for the analysis of hydration, it is necessary to have a set of water molecules within the system, the crystal in our case. In principle it is possible to use a previously determined set of hydration waters if available. If not, an automatic procedure has been used, which starts from the macromolecular model alone, determines the R and R-free factors, and calculates a starting water set.

The automatic procedure starts with several cycles of molecular dynamics simulated annealing refinement on the macromolecule alone, to obtain reliable R-free factors.[26] Water positions were found with a peak searching program[27] and a water assignment program (Arnez, personal communication). Peaks larger than a predefined threshold (typically $2\sigma(\rho)$, where $\sigma(\rho)$ is the standard deviation of the density) in the solvent region were selected from difference maps. These difference maps were calculated from

TABLE I. Agreement Factors for the Original Published Model, the Starting Model, the Output of the Single-Copy MD Run, and the Output of the Multicopy MD Run, Calculated Under the Same Conditions*

Solute	Data	R-value of original structure (%) (no. of original waters)	No. of waters	No. of copies	Starting model Start R(%)	Starting model R-free (%)	MD on single water copy R(%)	MD on single water copy R-free (%)	Multi water copy R(%)	Multi water copy R-free (%)	Multi water copy Weight water-solute
BPTI*	Neutron	22.2 (63)	63	9	22.2	—	21.7	27.0	20.2	26.9	1
BPTI	Neutron	22.2 (63)	63	9	22.2	—	21.7	27.0	21.2	26.5	1/9
BPTI	Neutron	22.2 (63)	63	5	22.2	—	21.7	27.0	19.8	27.1	1
BPTI	Neutron	22.2 (63)	63	5	22.2	—	21.7	27.0	20.8	26.3	1/5
BPTI*	X-ray	22.1 (63)	107	5	26.4	28.9	23.3	25.3	22.1	24.3	1
BPTI	X-ray	22.1 (63)	107	5	26.4	28.9	23.3	25.3	22.1	24.9	1/5
RNA*	X-ray	16.8 (91)	120	5	20.7	27.6	16.6	26.2	15.5	26.2	1/10
RNA	X-ray	16.8 (91)	120	5	20.7	27.6	16.6	26.2	15.0	27.0	1/5

*The models shown in the figures are indicated by an *. For BPTI neutron data, all 3950 available reflections available in the PDB between 10 and 1.8 Å are used. The starting model is the published one with 63 water molecules,[29] without any previous simulated annealing. No R-free was calculated originally for this structure. Subsequent R-free values correspond to molecular dynamics on the water alone. The number of parameters in the water multicopy model is 2268 (the protein is kept fixed during the dynamics refinement).

the coefficients (Fobs-Fcalc).exp(i.ϕ(calc)) where Fobs are the measured amplitudes and Fcalc.exp(iϕ(calc)) are the structure factors calculated from the macromolecular model. A minimum distance to the protein and between peaks of 2.4 Å was imposed. The water molecules that are too close to the macromolecular model or overlap with other ones are removed. Note that this step can remove close peaks that correspond to real alternative water sites; a better procedure is under development. The output gives the single copy model for which R factor and an R-free factor are calculated.

Multicopy Method and X-Ray Forces

The multicopy method including X-ray forces was implemented in XPLOR.[28] The steps are the following. First, single-copy refinement: A first refinement of the single copy water set was performed, including positional refinement, molecular dynamics, and temperature factor refinement. The molecular dynamics step is usually short (total time of 0.05 ps, cooling from 350 to 300 K, steps of 25 K, 50 cycles of 0.5 fs per step) to avoid a substantial drifting of the water molecules from their original sites. This drift would uncouple the water position from the difference map peak. Second, multicopy refinement: The resulting water set was used to generate several copies. Each water molecule interacts with the macromolecule and with the water molecules of the same copy, but not with the water molecules belonging to other copies.

A positional refinement is performed first, followed by a slow cooling procedure. An initial temperature of 350 K is applied to the system by assigning randomly different initial velocities to each water molecule. This makes the trajectories of the different copies not equal. Then the system is slowly cooled for a total time of 0.25 ps (steps of 5 K, 50 cycles of 0.5 fs per step) until it reaches the desired temperature (300 K). During the slow cooling, the solute is kept rigid. The structure factors are calculated as the sum of the contribution of all the copies, each one contributing with weight 1/N, with N equal to the number of copies. The total time and the temperature range are set to allow dispersion of the copies while keeping them in the neighborhood of the original site. Because the difference map is reproduced by the average of the copies, each of them can be farther from the original site than in the single-copy case. The water model used is TIP3, and periodic conditions are applied to the unit cell box. Crystallographic symmetries are strictly enforced. Temperature factors for individual waters are refined after the single-copy and multicopy steps, provided the R-free factor does not increase.

The relative weights of the water-solute interactions, water-water interactions, and the crystallographic terms were varied to search for the best R and R-free factors. For the water-macromolecule and the water-water interactions the possibilities of giving either full weight or dividing by the number of copies were considered.

RESULTS AND DISCUSSION

Hydration of BPTI: Neutron Data

One of the cases studied[29] was the bovine pancreatic trypsin inhibitor (BPTI). The multicopy water molecular dynamics method was first applied by using the published protein structure (with only one of the two alternative positions of Met 52), including the water set determined by conventional crystallographic methods.[30] The neutron diffraction data were used, with all the 3950 reflections from 10 to 1.8 Å.[30] A first run of single-copy water dynamics (heat-

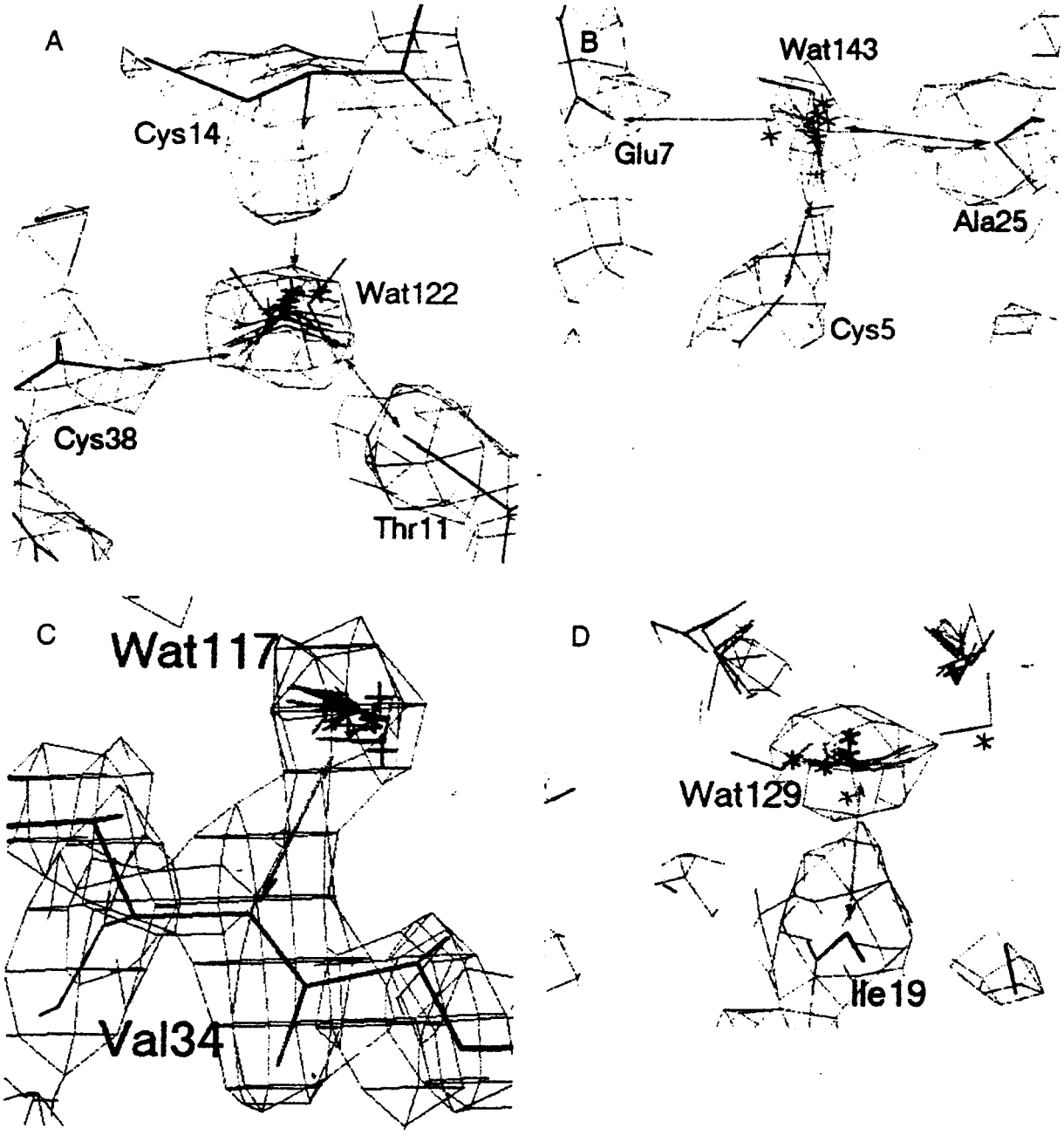

Fig. 1. Superposition of the BPTI water multicopy models and density maps. Waters obtained by using multicopy with neutron data are shown as angular brackets, and water obtained from multicopy with X-ray data are shown as crosses. Figures 1A–E show a 2Fobs-Fcalc map calculated at 1.8 Å (1σ contours) by using neutron data and the adjusted single-copy water model. Figure 1F shows a Fobs-Fcalc map calculated at 1.0 Å (1σ contours) by using X-ray data and the BPTI model alone after slow cooling. **A:** Water molecule W122, completely buried in the interior of the protein. **B:** Surface water molecule W143 in a partially buried restricted site. **C:** Surface water molecule W117 in a strongly restricted site. **D:** Surface water molecule W129 in a weakly restricted site. **E:** Surface waters W209, W127, W302, and W313 forming a disordered water channel near Lys 46, neutron map. **F:** Surface waters W209, W127, W302, and W313 forming a disordered water channel near Lys 46, X-ray map.

ing to 500 K and cooling to 300 K in steps of 25 K, 50 cycles of 0.5 fs per temperature step, total time 0.2 ps) was performed. Because in the neutron case the solvent contribution to the amplitudes is very significant, this step can be used to obtain an R-free factor. The strength of the solvent contribution to the diffraction signal also allowed a longer dynamics simulation without substantial drifting of the water molecules from their original site.

Then two sets of multicopy runs were done, one with 9 copies and the other with 5 copies of the original set of 63 water molecules, as described in

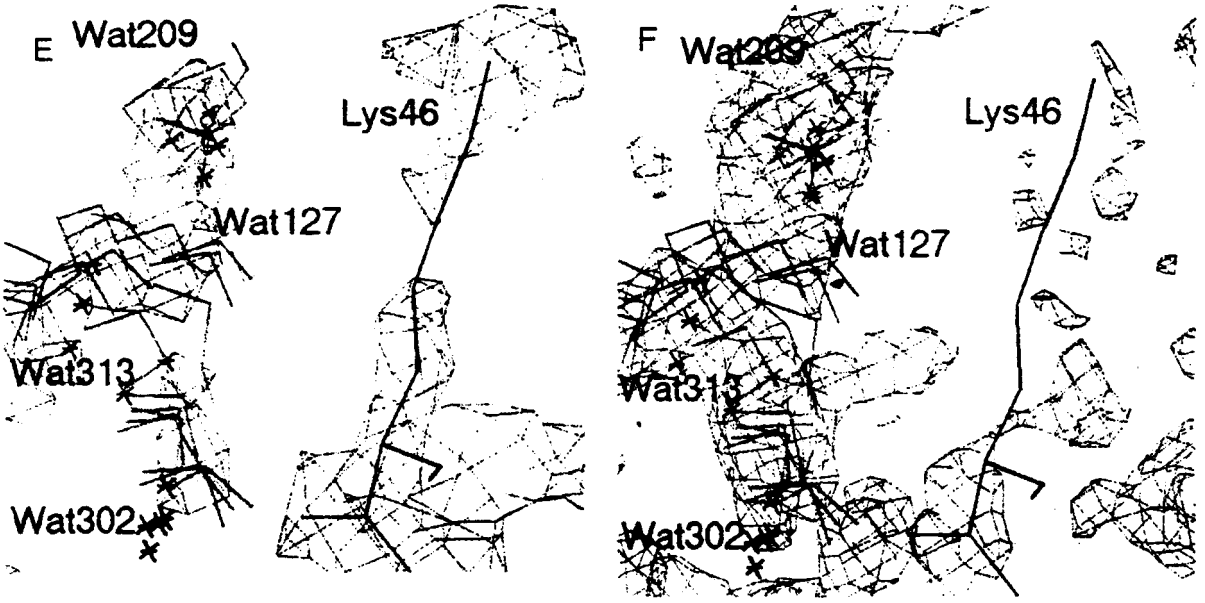

Figure 1. (Continued.)

multicopy refinement in Methods. A search for the optimal weights for the different interactions (water-water and water-protein) and for the X-ray energy term was conducted; most R and R-free values were lower than the single-copy ones. The results of the runs with different weights and different number of copies are similar; the cases with best R and R-free factors are shown in Table I. It can be seen that the multicopy method improves the R-factor, as expected from the increased number of parameters, but that the effect on the R-free is only marginal. Therefore, the main effect of the method is not to increase the agreement with the experimental data but to propose a model of equivalent accuracy that describes better the water mobility.

In what follows, results from the lowest R-factor 9 copies run are shown. Because molecules of different copies do not interact, they may occupy very close positions. For a highly specific ordered hydration site, both the model and the crystallographic added forces will keep all water copies close to the site. This is the case for buried water molecules, like W122. Figure 1A shows the multicopy model for W122 superposed with the map obtained by neutron diffraction at 1.8 Å using the refined single water copy model, calculated from the coefficients (2Fobs-Fcalc).exp(i.ϕ(calc)). This buried molecule is at a proper distance to make hydrogen bonds with residues Cys14, Cys38, and Thr11. These three hydrogen bonds not only restrict the overall position of the molecule but the orientation of most of the copies; seven of nine have similar orientations. The crosses represent the water molecules obtained by using the X-ray data and will be discussed later. This molecule has been localized by X-ray crystallography, in three different crystalline forms by neutron diffraction, and by low temperature X-ray diffraction.[29,31–33]

Waters that occupy clefts (accessible area <20 Å^2) tend also to be more ordered. Figure 1B shows the result of multicopy modeling of water W143 near residue Ala25. In the original model this water is partially buried (accessible area = 12 Å^2) and can make hydrogen bonds to Glu7, Cys5, Tyr23, and Ala 25. After multicopy modeling, the different copies are somewhat displaced in two groups, one with seven and one with two copies, optimizing the fit to the map and the hydrogen bonding distances.

The dispersion of surface waters covers a wide range. Some surface waters are quite ordered. An example is W117, hydrogen bonded to the NH of Val 34, for which the multicopy model shows a compact form (Fig. 1C). Other sites are more extended, like that of water molecule W129, hydrogen bonded to the NH of Ile19 (Fig. 1D). In this case, the different water copies are still in a small volume but dispersed in different positions, all making the same hydrogen bond. Most surface water have higher dispersions and are not identified in specific sites. Figure 1E shows such a case near Lys46, where the multicopy model forms a continuum following the 2Fobs-Fcal map. This disordered region is an example in which the anisotropy of the map makes it very difficult to fill the region with fixed point waters. The water molecules of different copies are distributed following this anisotropic volume, defining a region in which they can move.

These examples suggest that the multicopy method gives a better sampling of space than the single-copy one, indicating several possibilities compatible with the amplitude data. In some cases of surface waters,

TABLE II. Analysis of the Dispersion Value of the Water Molecules*

Water site	No. of waters	Average copy RMS (std) (all copies)	Average site RMS (std) (1 σ-radius)
1) Classification according to accessible area			
Buried sites	4	0.43 (0.21)Å	0.26 (0.08)Å
Cleft sites	9	0.66 (0.33)Å	0.38 (0.14)Å
Surface sites	50	0.80 (0.31)Å	0.49 (0.21)Å
2) Classification according to common sites			
Common sites	15	0.61 (0.36)Å	0.34 (0.16)Å
Other	48	0.81 (0.31)Å	0.51 (0.20)Å
3) Classification according to temperature factor			
B <20 Å²	9	0.58 (0.24)Å	0.38 (0.16)Å
B >20 Å²	54	0.78 (0.33)Å	0.48 (0.20)Å

*The average over the different water sites is given, with the corresponding standard deviation in parenthesis.
Three classifications are made: 1) Buried waters (accessible area = 0 Å²), cleft water (0 < accessible area < 20 Å²) and surface water (accessible area > 20 Å²). Accessible areas are calculated including symmetry related molecules. 2) Sites common to the 3 BPTI forms[32] and the rest. 3) Waters with original B <20 Å² and waters with original B >20 Å². The copy RMS is calculated over the distance of all copies to their center of mass, and the site RMS is calculated over the copies within a 1 σ(copy RMS) radius from the center of mass.

different water copies merge in a continuous channel. Table II shows the difference between the RMS dispersion of buried sites (accessible area= 0 Å²), cleft sites (0 Å² < accessible area < 20 Å²), and surface sites (accessible area < 20 Å²). It is clear that buried sites are the most compact, followed by cleft sites, whereas surface sites are the least compact. For the surface waters with accessible area > 20 Å², there is no correlation between RMS multicopy dispersion and accessible area. Two RMS values of the dispersion are calculated in every case. The first one corresponds to the RMS distance of all copies to their center of mass (Copy RMS). The second one (Site RMS) corresponds to the RMS distance of the copies that are within 1σ (calculated from the copy RMS distribution) of the center of mass and is meant to give the dispersion near the center of the site. Table II shows also that sites common to all three BPTI forms are more compact than the others and that sites with original B-factors less than 20 Å² are more compact than the others. For temperature factors > 20 Å² the correlation between multicopy dispersion and B-factor is lost. This is probably because the isotropic B does not reflect the shape of the site in the same way as the multicopy model. The histogram showing the number of sites with a given RMS dispersion of the water copies around their center of mass is shown in Figure 2, curve A.

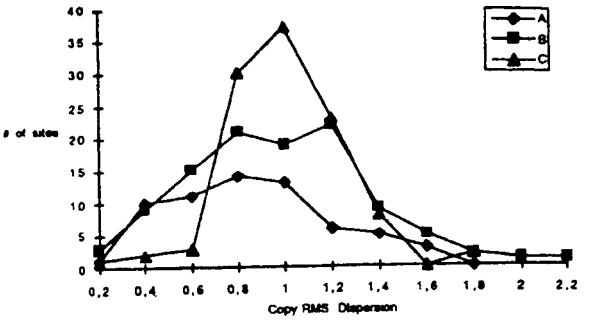

Fig. 2. Number of sites as a function of the RMS dispersion of the different copies for (**A**) the multicopy simulation against 1.8 Å BPTI neutron data described (total of 63 sites); (**B**) the model of the multicopy simulation against 1 Å BPTI X-ray data (total of 107 sites) ; and (**C**) the model of the multicopy simulation against 2.2 Å RNA X-ray data (total of 120 sites).

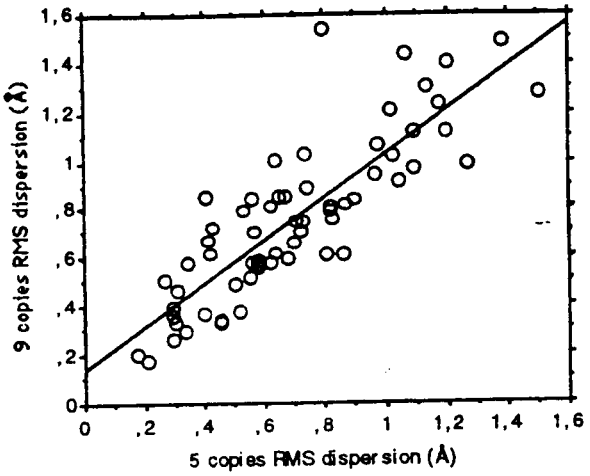

Fig. 3. Comparison of the RMS copy dispersions for the five copies and the nine copies runs of BPTI against neutron data. Each point represents a site, with coordinates (x = 5 copies RMS dispersion, y = 9 copies RMS dispersion). The line y = 0.89x + 0.141 represents the linear regression; the correlation value is 0.84.

Well-ordered sites (RMS < 0.6 Å) include most of the buried and cleft sites, and some ordered surface sites.

As noted above, the results of runs with different number of copies are similar. For example, the correlation between the RMS copy dispersion around the center of mass of the five copies run and the nine copies run with lowest R-factors is 84% (see Fig. 3). This figure gives also an idea of the reproducibility of the dispersion values, which is better for well-ordered sites. In all these runs, following the standard XPLOR protocol, the charges on the residues Asp, Glu, Arg, and Lys were turned off. When these charges were on, the results are qualitatively similar, except in the immediate neighborhood of charged residues.

These results can be compared with the water mobility, by using the residence times measured by

TABLE III. Comparison of the Residence Times From NMR and NMRD, the Multicopy Site RMS and the Multicopy Site Shape for the Water Molecules Shown in Figure 1*

Water	NMR and NMRD residence time	Copy RMS	Site RMS (1 σ-radius)	Site shape
W122 (buried)	10–200 μs[18,34]	0.29 Å	0.2 Å	Compact
W143 (cleft)	300 ps[20]	0.37 Å	0.23 Å	Compact
W117 (surface)	400 ps[20]	0.37 Å	0.25 Å	Compact
W129 (surface)	<50 ps[20]	0.96 Å	0.49 Å	Extended
W127, W209, W302, W313 (surface)	50 ps[20]	0.90 Å (mean)	0.59 Å (mean)	Channel

*For NMRD mobility, W122 is specifically identified by a mutation.[34] For NMR mobility, W143, W117, and W129 are identified by H-bonds to Ala 25, Ile 19, and Val 34, and W209 is near Cε of Lys 46. The mobility of water near these residues is given in Ref. 20.

NMR and NMRD. Table III shows the published water mobility values and the characteristics of the multicopy sites for the nine copies run. Multicopy delocalization and shape are useful to distinguish between ordered surface waters and channel surface waters. It gives an indication of mobility for the highly delocalized water in channels. Note that cleft waters tend to have a small multicopy site due to space restrictions and that surface waters near charged residues could be localized while interchanging rapidly with the bulk.

Note that low temperature diffraction experiments of BPTI show a much more detailed density map, a sign of higher order.[33] Therefore, the degree of delocalization of water molecules in channels is a function of their kinetic energy.

De Novo Modeling of Hydration of BPTI: X-Ray Data

The application of the multicopy method to modeling the hydration of BPTI described in the previous section used the information of high resolution diffraction experiments, including the point hydration model.[29] A completely different run was done to check it, with the following differences: 1) the available X-ray diffraction data (17,614 reflections[30]) were used from 8 to 1.0 Å resolution; 2) the protein model was obtained from the original BPTI model after several cycles of heating (3000 K) and slow cooling. This enabled the calculation of an R-free factor; and 3) the water model was obtained automatically as described in Methods, selecting peaks larger than $2\sigma(\rho)$. The procedure gave a total number of 107 water molecules, 44 more than the original model.

The multicopy simulation described in Methods was then applied by using five copies of these water molecules. The results of the runs with different weights and different number of copies are similar. Most R and R-free values were lower than the single-copy ones; the cases with best R and R-free factors are shown in Table I. The model with the lowest R and R-free values was chosen. The root mean square (RMS) dispersion of the multiple copies around the site center of mass is shown in Figure 2, curve B. It shows a similar number of ordered sites with low RMS dispersion (<0.6 Å) of the different copies as the one obtained by using neutron data (curve A), whereas there are more sites with higher dispersion, reflecting the additional 44 water molecules. The results of this simulation is shown in Figure 1 as crosses. Note the agreement between X-ray (crosses) and neutron results (angular brackets), which is good for the sites shown in Figure 1A–D and reasonable for the channel region shown in Figure 1E.

The agreement of the neutron (angular brackets) and X-ray multicopy models (crosses) with the X-ray difference map, calculated from the protein alone, is shown in Figure 1F. The region displayed is the same as the one in Figure 1E, near Lys 46. Note that the neutron model follows closely the X-ray map. This represents a cross-validation of the multicopy model and also implies that the results do not depend on small variations of the protein conformation.

The overall picture emerging from these results shows that most of the water molecules present in the system are distributed in some regions, likely moving along them, and a few very precisely located in preferential hydration sites. A general picture of surface BPTI hydration is given in Figure 4, which shows a view of the BPTI molecule with the water molecules according to the results of the first simulation using neutron data (right), compared with the original water distribution (left). For most surface waters, the multicopy run expands individual water molecules into continuous distributions that tend to follow the protein surface, like the channel near Lys 46. On the other hand, the copies of ordered waters in clefts, like W143, stay close to the single sites.

Application to RNA Fragment

The multicopy modeling method was applied to the RNA fragment $U(UA)_6A$[35] to describe its hydration. Because the crystals diffract only to 2.2 Å resolution, this case is much less favorable than the BPTI one from the point of view of the signal strength from the water molecules . The data between 8 and 2.2 Å resolution (3714 reflections) were used for the simulations.

The automatic water determination protocol described in Methods was applied. First, a simulated annealing process was run for the RNA alone (with-

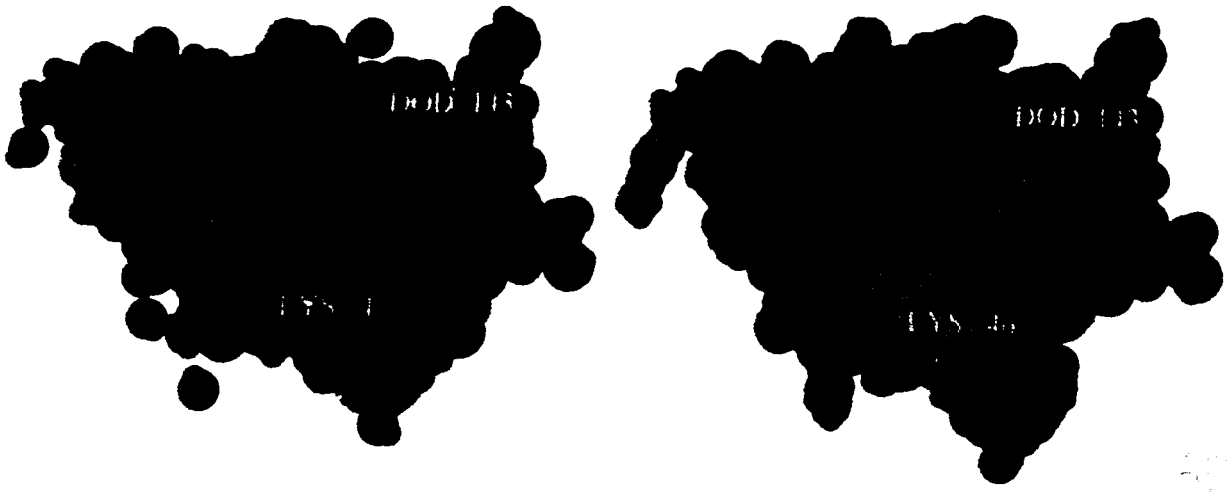

Fig. 4. Comparison of the original water distribution for BPTI (**left**) and the result of the nine copies run against neutron data (**right**). Lys 46 and W143 are marked.

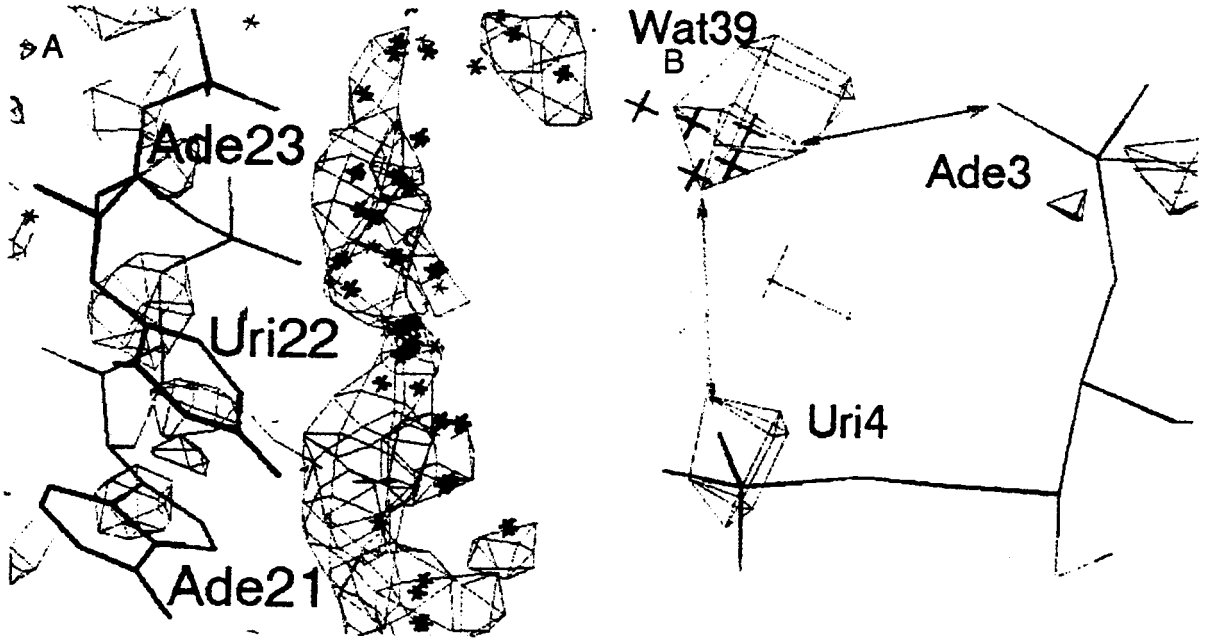

Fig. 5. Superposition of different RNA water multicopy model with a Fobs-Fcalc map calculated at 2.2 Å at 1.5 sigma level with the RNA model. The water molecules from the multicopy model are shown as crosses. **A:** Deep groove region showing a water channel in coincidence with the difference electron density map. **B:** An ordered water molecule hydrogen bonded to the oxygens of two phosphate groups.

out solvent) starting at 6000 K and followed by a slow cooling. Then a difference density map was calculated at 2.2 Å resolution. An analysis of this map detected 120 independent peaks larger than $2\sigma(\rho)$, which were assigned as waters.

Five copies of the water set were generated, and the multicopy molecular dynamics modeling was run. Note that in this case the charges on the phosphate backbone are on, but the water-protein interactions are divided by a factor of 10, which diminishes the electrostatic effect. During the search for the best weights, most R-free values were higher than the single-copy one, showing the difficulty of this case. The best R-factor and R-free values are shown in Table I. The model chosen had a lower R-factor and equal R-free factor compared with the single-copy values. The histogram of the number of sites with a given copy dispersion is given in Figure 3, curve C. The emerging picture is that of a very mobile hydration layer, with some ordered sites (RMS < 0.6 Å). As an example of the mobile hydration layer, Figure 5A shows a region of the deep groove with the difference map and the water molecules originated by the multicopy procedure. A clear

water channel is seen; such a channel had already been noted in the original structure determination,[36] whereas mobile water bridges have been observed for other nucleic acids (see Westhof[37] and references therein). Another example, showing a clearly ordered site, is given in Figure 5B. It shows a water molecule clearly bound to the backbone of the RNA bridging two anionic phosphate oxygen atoms of successive phosphate groups. Such bridges are frequent in RNA (see Westhof[37] and references therein).

Note that, in the described applications, the multicopy model follows closely the single-copy one. The possibility of extending the multicopy procedure to a set of waters that fills the unit cell is currently under study. To begin, we have run single-copy molecular dynamics simulations with complete waters sets, which gave an adequate starting point. The work is being pursued, and the details will be published elsewhere.

CONCLUSIONS

The modeling method described here allows to build up a multicopy hydration model fully compatible with the experimental diffraction data. In this model, there is only a small fraction of water molecules belonging to very specific hydration sites, whereas most of them are distributed in larger regions, sometimes forming continuous channels. Therefore, the water structure shown in X-ray and neutron diffraction studies seems to be a result of the model used, rather than of the experimental data. The single-copy model implies the selection of possible water molecules due to distance criteria and the modeling of the accepted ones as Gaussian spheres or ellipsoids, whereas the present work models much more closely the electron density maps with partially occupied sites. It stills starts from a single-copy model; work in progress will enable to start from alternative networks, thus recovering all information available in the electron density map.

The R-free measures show that the result of this modeling procedure is at least as good as the point water model, considering its agreement with diffraction amplitudes, and sometimes marginally better. It adds a larger sampling of space, specially visible for the surface water molecules that are expanded and merge into channels. This view agrees with the NMR and modeling data, which suggest a high mobility for these water molecules and, therefore, gives a description of protein hydration that can integrate the results of different techniques.

ACKNOWLEDGMENTS

This work was supported by the EU through the collaborative project CII-CT 93 0014, DG12 HSMU, the MENESRIP (France) through a travel grant to J.R.G., the CONICET through the IFLYSIB, the CNRS through the UPR 9004, the Institut National de la Santé et de la Recherche Médicale, and the Centre Hospitalier Universitaire Régional. J.R.G. is a member of the Carrera del Investigador of CONICET. We thank D. Moras for his continuous support and B. Rees, T. Simonson, and E. Westhof for useful discussions.

REFERENCES

1. Frey, M. Water structure of crystallized proteins: High-resolution studies. In: "Water and Biological Macromolecules" and references therein. Westhof, E. (ed.). London: Macmillan, 1993:98–147.
2. Berendsen, H.J.C. Specific interactions of water with biopolymers. In: "Water. A Comprehensive Treatise" Vol. 5. Franks, F. (ed.). New York: Plenum Press, 1975:293–330.
3. Karplus, P.A., Faerman, C. Ordered water in macromolecular structure. Curr. Opin. Struct. Biol. 4:770–776, 1994.
4. Westhof, E. Water: An integral part of nucleic acid structure. Annu. Rev. Biophys. Biophys. Chem. 17:125–144, 1988.
5. Burling, F.T., Weis, W.I., Flaherty, K.M., Brünger, A.T. Direct observation of protein solvation and discrete disorder with experimental crystallographic phases. Science 271:72–77, 1996.
6. Saenger, W. Structure and dynamics of water surrounding biomacromolecules. Annu. Rev. Biophys. Biophys. Chem. 16:93–114, 1987.
7. Malin, R., Zielenkiewicz, P., Sanger, W. Structurally conserved water molecules in ribonuclease T1. J. Biol. Chem. 266:4848–4852, 1991.
8. Geertrui, S.S., Smith, C.K., Turkenbourg, J.P., Dettmar, A.N., Van Meervelt, L., Moore, M.H. DNA-drug refinement: A comparison of the programs NUCLSQ, PROLSQ, SHELXL93 and X-PLOR, using the low temperature dTGATCA-Nogalamycin structure. Acta Crystallogr. D 52: 299–314. 1996.
9. Teeter, M.M, Roe, S.M, Heo, N.H. Atomic resolution 0.83 Å crystal structure of the hydrophobic protein crambin at 130°K. J. Mol. Biol. 230:292–311, 1993.
10. Migchelsen, C., Berendsen, H.J.C. Proton exchange and molecular orientation of water in hydrated collagen fibers. An NMR study of H_2O and D_2O. J. Chem. Phys. 59:296–305, 1962.
11. Grigera, J.R., Berendsen, H.J.C. The molecular details of collagen hydration. Biopolymers 18:47–57, 1979.
12. Hilton, B.D., Hsi, E., Bryant, R.G. 1H Nuclear magnetic resonance relaxation of water in lysozyme powders. J. Am. Chem. Soc. 99:8487–8490, 1977.
13. Grigera, J.R. Nuclear magnetic resonance spin-spin relaxation time in hydrated protein powders. A two sites dynamic exchange model. J. Chem. Phys. 83:2145–2147, 1979.
14. Grigera, J.R., Mogilner, I.G. Water bridges in myoglobin. In: "Biophysics of Water." Franks, F., Mathias, S. (eds.). Chichester: J. Wiley, 1982:30–41.
15. Kibinec, M.G., Wemmer, D.E. Solution NMR of water on protein surfaces. Curr. Opin. Struct. Biol. 2:828–831, 1992.
16. Belton, P.S. NMR studies of protein hydration. Prog. Biophys. Mol. Biol. 61:61–79, 1994.
17. Otting, G., Liepinsh, E., Wüthrich, K. Protein hydration in aqueous solution. Science 254:974–980, 1991.
18. Denisov, V.P., Halle, B. Protein hydration dynamics in aqueous solution. Faraday Discuss. 103:227–244, 1996.
19. Levitt, M., Sharon, R. Accurate simulation of protein dynamics in solution. Proc. Natl. Acad. Sci. USA 85:7557–7561, 1988.
20. Brunne, R.M., Liepinsh, E., Otting, G., Wüthrich, K., Van Gunsteren, W.F. Hydration of proteins: A comparison of experimental residence times of water molecules solvating the bovine pancreatic trypsin inhibitor with theoretical model calculations. J. Mol. Biol. 231:1040–1048, 1993.
21. Levitt, M., Park Britt, H. Water: Now you see it, now you don't. Structure 1:223–226, 1993.
22. Lounnas, V., Pettitt, B.M. A connected-cluster of hydration

around myoglobin: Correlation between molecular dynamics simulation and experiment. Proteins 18:133–147, 1994.
23. Grigera, J.R., Grigera, T.S., Howard, E.I., Podjarny, A.D. Molecular dynamics simulation of crystal water with X-ray constraints. Int. J. Quantum Chem. Quantum Biol. 21:109–116, 1994.
24. Burling, F.T., Brünger, A.T. Thermal motion and conformational disorder in protein crystal structures: Comparison of multi-conformer and time-averaging models. Isr. J. Chem. 34:165–175, 1994.
25. Howard, E.I., Grigera, J.R., Grigera, T.S., Podjarny, A.D. On the problem of modeling in macromolecular crystals using diffraction data. 2. Using molecular dynamics in the middle resolution range. Join CCP4 and ESF-EACBM Newsletter in Protein Crystallography 31:17–19, 1995.
26. Brünger, A.T. Free R value: A novel statistical quantity for assessing the accuracy of crystal structures. Nature 355: 472–475, 1992.
27. Kleywegt, G.J., Jones., T.A. xdlMAPMAN and xdlDATAMAN. Programs for reformatting, analysis and manipulation of biomacromolecular electron density maps and reflection data sets. Acta Crystallogr. D 52:826–828, 1996.
28. Brünger A.T. "X-PLOR. Version 3.1, A system for X-ray Crystallography and NMR." The Howard Hughes Medical Institute and Department of Molecular Biophysics and Biochemistry, Yale University. New Haven, CT: Yale University Press, 1992.
29. Wlodawer, A., Walter, S., Huber, R., Sjölin, L. Structure of bovine pancreatic trypsin inhibitor. Results of join neutron and X-ray refinement of crystal form II. J. Mol. Biol. 180:301–329, 1984.
30. Wlodawer, A., Walter, S., Huber, R., Sjölin, L. Protein Data Bank entry PDB5PTI, structure factor entries R5PTISFX and R5PTISFN, Upton, NY: Brookhaven National Laboratory, 1984.
31. Deisenhofer, J., Steigemann, W. Crystallographic refinement of the structure of bovine pancreatic inhibitor at 1.5 Å resolution. Acta Crystallogr. B 31:238–250, 1975.
32. Wlodawer, A., Nachman, J., Guilliland, G.L., Gallagher, W., Woodward, C.J. Structure of form III crystals of bovine pancreatic trypsin inhibitor. J. Mol. Biol. 198:469–480, 1987.
33. Parkin, S., Rupp, B., Hoe, H. Structure of bovine pancreatic trypsin inhibitor at 125K: Definition for carboxyl-terminal residues Gly57 and Ala59. Acta Crystallogr. D 52:18–29, 1996.
34. Denisov, U.P., Halle, B., Peters, J., Horlein, H.D. Residence times of the buried water molecules in bovine pancreatic trypsin inhibitor and its G36S mutant. Biochemistry 34: 9046–9051, 1995.
35. Dock-Bregeon, A.C., Chevrier, B., Podjarny, A.D., Moras, D., de Bear, J.S., Gough G.R., Ghiham, P.T., Johnson, J.E. High resolution structure of the RNA duplex [UUA6A]2. Nature 335:375–378, 1988.
36. Dock-Bregeon, A.C., Chevrier, B., Podjarny, A.D., Johnson, J.E., de Bear, J.S., Gough, G.R., Ghiham, P.T., Johnson, J.E., Moras, D. Crystallographic structure on an RNA helix [UUA6A]2. J. Mol. Biol. 209:459–474, 1989.
37. Westhof, E. Structural water bridges in nucleic acids. In: "Water and Biological Macromolecules" and references therein. Westhof, E. (ed.). London: MacMillan, 1993: 226–243.