Modelling of protein hydration

2002 J. Phys.: Condens. Matter 14 2439

(http://iopscience.iop.org/0953-8984/14/9/331)

INSTITUTE OF PHYSICS PUBLISHING JOURNAL OF PHYSICS: CONDENSED MATTER

J. Phys.: Condens. Matter **14** (2002) 2439–2452 PII: S0953-8984(02)32441-X

Modelling of protein hydration

Helmut Durchschlag[1] **and Peter Zipper**[2]

[1] Institute of Biophysics and Physical Biochemistry, University of Regensburg, Universitätsstrasse 31, D-93040 Regensburg, Germany
[2] Physical Chemistry, Institute of Chemistry, University of Graz, Heinrichstrasse 28, A-8010 Graz, Austria

E-mail: helmut.durchschlag@biologie.uni-regensburg.de

Received 8 January 2002
Published 22 February 2002
Online at stacks.iop.org/JPhysCM/14/2439

Abstract

Starting from the atomic coordinates of proteins, by modern surface calculation programs the exact surface topography of proteins can be visualized in terms of dot surface points. In combination with tricky hydration algorithms, based on the hydration numbers of individual amino acid (AA) residues and appropriate selection criteria, this may be used for advanced modelling studies including a fine tuning of the input parameters for such computer simulations. 'Bead modelling' can be used for both the individual AA residues and individual water molecules placed at preferred positions on the protein envelope. Problems of special concern are connected with the properties and the localization of the water molecules bound to the protein surface. Qualified assumptions regarding number, density/volume, position and behaviour of the water molecules at the protein–water interface are required.

1. Introduction

Proteins are biopolymers of high molar mass, consisting predominantly of about 20 different amino acids (AAs) linked by peptide bonds. They are among the most important functional components of living cells; for example, enzymes catalyse many biochemical reactions of the metabolism. Their properties and behaviour are governed by their structures [1]. In solution, the bonds in the protein backbone allow considerable flexibility and motions of the chains. Among the various classes of proteins, the water-soluble globular proteins play a major role. The extent of solubility is determined by the AA composition, in particular by the distribution of polar (ionic or neutral and hydrophilic) and nonpolar (hydrophobic) AAs on the protein surface, the molecular shape and the environmental conditions (pH, temperature, presence of cosolvents, ionic strength etc). Hydrophilic proteins are surrounded by water molecules and are able to occlude hydrophobic compounds. The polar water molecules interact with proteins, to form a hydration layer. Water molecules are generally excluded from the protein interior.

Essentially all ionized groups (Asp, Glu, Lys, Arg) in water-soluble proteins are on the protein surface, i.e. exposed to the solvent, whereas nonpolar side chains (Ala, Leu, Val, Ile, Pro, Phe, Met, Trp, Gly) predominate in the protein interior. The water molecules on the molecule surface are attached by hydrogen-bonding to the polar groups on the protein surface. These attached water molecules can fix adjacent water molecules, but with a continuous decrease of order. This means that globular proteins in aqueous solution are surrounded by tightly bound water molecules, which have properties differing from those of bulk water (more ordered, less mobile, enhanced density [1–6]). In solution nearly all surface water molecules are in rapid exchange with bulk water, the average residence times of the ordered surface water ranging from nanoseconds to picoseconds. Water–water hydrogen bonds pull water molecules away from hydrophobic surface patches.

Hydration and the occurrence of manifold protein–solvent interactions are crucial for stability, structure and function of proteins in aqueous solution [1, 7–13]. Biological activity can only be generated if a minimum amount of water is provided. As follows from a comparison of the results obtained by various physicochemical and modelling techniques, the hydration layer of most simple proteins contains about 0.35 g of water/gram of protein ([6] and references therein), corresponding to approximately two water molecules per AA residue.

Precise modelling of protein structures including the contributions from hydration is vital for various reasons: comparison of the results obtained from quite different techniques and approaches, statements on any differences of the structure determined in the crystal from that in solution (i.e. under quasi-physiological conditions), understanding hydration and hydrodynamics of proteins, establishing the structural basis for drug-design projects by predicting putative protein conformations etc.

In this paper an attempt is made to reexamine the problem of hydration effects critically, using a well characterized enzyme—citrate synthase—in a case study. The general strategy is similar to that advocated by us previously [6, 14, 15], i.e. starting from the atomic (AT) or AA coordinates of the protein, we calculated the surface topography of the dry protein structure, assembled appropriate numbers of water molecules on the protein envelope, predicted various solution scattering and hydrodynamic parameters of the hydrated structures by advanced modelling techniques and, finally, checked the resultant molecular properties by comparison with experimental data. Embedded in this general framework, we now use a series of novel tuning parameters to adjust the procedures to be applied more appropriately. The approaches used in this study include variation of the hydration algorithms and hydration values for individual AA residues. By applying the mentioned tools, a reduction of a very complicated problem to smaller, more manageable problems can be achieved.

2. Development of hydrated protein models

The AT data, which represent the anhydrous protein model, were converted to the hydrated model step by step, as follows from the flowchart given in figure 1.

2.1. Sources of data

In the past, a wealth of physicochemical information has been obtained for the dimeric enzyme citrate synthase, covering both the high-resolution crystal structure [16, 17] as well as solution scattering [18, 19] and hydrodynamic behaviour [20, 21]. The crystal data of pig-heart citrate synthase can be taken from the protein data bank (PDB) [22], accession code 1CTS, and the exact value for its molar mass M of 97.838 kg mol^{-1} can be derived from the AA sequence stored in the SWISS-PROT data bank [23], accession number P00889.

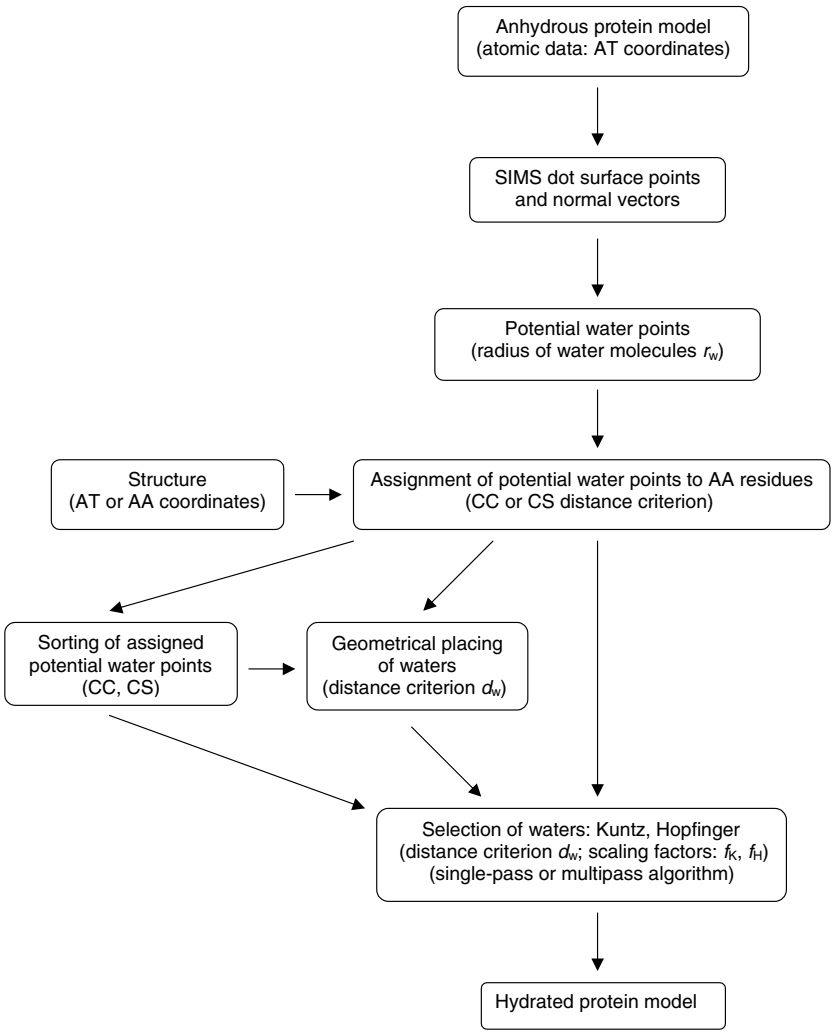

Figure 1. Flow chart illustrating the general strategy for the development of hydrated protein models, starting from the AT coordinates and utilizing the surface calculation program SIMS [31] and our hydration programs HYDCRYST and HYDMODEL [6, 14, 15] for creating space-filling models. Each box describes a stage in the procedures to be applied and shows how additional information is included. CC and CS are the centre–centre or centre–surface distances between water and the atoms or AA residues of the protein; d_w is the minimum distance between adjacent water molecules.

Hydration values for individual AA residues have been presented by many authors, among them the values by Kuntz [24] and Hopfinger [25] being the most prominent and comprehensive ones. While the values proposed by Kuntz are based on nuclear magnetic resonance studies of polypeptides (determination of the amount of unfrozen water at $-35\,^{\circ}$C) [24,26], the hydration numbers suggested by Hopfinger [25] originate from theoretical calculations carried out to describe the thermodynamic behaviour of water–polypeptide group interactions [27, 28]. The two sets of values are similar but not identical (e.g. showing major differences for the values of Pro and Phe at neutral pH) and predict the hydration of proteins (δ_1) in good agreement with experiments [7, 25].

2.2. Construction of dry models from crystal data

The AT or AA coordinates of citrate synthase were converted to dry (anhydrous) models as described previously [29, 30]. In short, all atoms of the crystal structure, or alternatively all AA residues, were represented by spheres ('beads') of appropriate volume. In the case of the AA approximation, the sphere volumes corresponded to the sum of the AT volumes of the corresponding residues and the spheres were placed at their mass centres of gravity.

2.3. Calculation of surface topographies

Among several surface calculation programs tested [6, 14, 15], the program SIMS [31] turned out to be particularly useful for creating a huge number of dot surface points (N_{dot}) and corresponding normal vectors. In the case of citrate synthase, approximately 25 000–170 000 dots were calculated from the original PDB file, the number of dots dependent on the nominal dot density (d_{dot} = 0.1–5.0 Å^{-2}) applied. For surface calculations, a probe radius of r_{probe} = 1.4 Å and a smoothing probe sphere of r_{sm} = 0.4 Å was used.

2.4. Hydration of dry models

The dot surface points and normal vectors of the resultant surface topography served as a pool of starting points for potential positions of water molecules on the protein envelope. Special hydration algorithms (program HYDCRYST for AT coordinates and program HYDMODEL for AA coordinates) were applied to select preferential positions for bound water. Selected bound water molecules were represented by spheres of average volumes (V_w = 24.5 Å^3) different from those of bulk water (V_w = 29.9 Å^3). The number of water molecules (N_w) assigned to each accessible AA residue was based on the values given by Kuntz [24] and Hopfinger [25] for neutral pH. In previous calculations [6, 14, 15] a radius of water molecules of r_w = 1.4 Å and a minimum distance between adjacent water molecules of d_w = 2.8 Å turned out to be useful. Selected water molecules were then added to the dry models. For details see [6, 14, 15].

2.5. Fine tuning of protein hydration

Different degrees of hydration were achieved by introducing the scaling factors f_K and f_H, acting directly on the hydration values given by Kuntz [24] and Hopfinger [25], respectively, for different AA residues. Factors of unity signify that each accessible AA residue on the protein surface attains the original hydration value given by these authors, or for steric constraints a smaller value. The water molecules obtained by employing factors of unity ascertain a minimum hydration ($N_{w,1}$); factors of two mean that given AA residues may obtain twice as many water molecules as obtained by the original Kuntz or Hopfinger values, unless interdicted by steric reasons; usage of factors of about ten leads to a maximum hydration ($N_{w,max}$), yielding a surface more or less covered by a monolayer of water.

The sequence and the details of the steps to be performed for the application of hydration algorithms turned out to be similarly important tools for the fine tuning of hydration (cf figure 1). The process of the selection of water molecules, the placing of water at definite positions on the protein surface, and some kind of sorting mechanism are adjustable parameters which have to be tested critically.

2.6. Calculation of parameters of hydrated models

Based on the AT or AA coordinates and the coordinates of the selected water molecules, parameters such as radius of gyration, R_G, and hydrated volume, V, can be calculated. The use of the hydration numbers given by Kuntz [24] or Hopfinger [25] allows the calculation of a theoretical hydration of accessible AA residues and the prediction of a maximum hydration in the case where the total number of AA residues would be hydrated.

2.7. Hydrodynamic modelling

For hydrodynamic modelling, the resulting hydrated models had to be reduced to a practicable number of spheres (N_{beads}) because the initial number of beads was too large (>1000). This was accomplished by a cubic grid approach [29]. Sedimentation and diffusion coefficients, s and D, and intrinsic viscosity $[\eta]$ were calculated by means of the program HYDRO [32, 33]; the delicate problem of overlapping non-equal spheres was handled as described in [29, 34].

2.8. Modelling of scattering functions

For the calculation of scattering profiles and SAXS parameters, the process of model reduction is not needed. SAXS curves, $I(h)$, were calculated by Debye's formula and were converted to distance distribution functions, $p(r)$, by Fourier transformation [35]. In contrast to hydrodynamic calculations, for the prediction of the SAXS behaviour, the spheres were weighted according to the calculated number of excess electrons (difference between the number of electrons of atoms or AA residues or bound water molecules and the number of electrons in the same volume of bulk water) [6, 14, 15, 36]. As usual, from SAXS functions also a variety of structural parameters such as radius of gyration, $R_{G,SAXS}$, and hydrated volume, V, can be derived if required.

3. Results and discussion

The PDB file of the crystal data of citrate synthase can be exploited to construct space-filling models of the anhydrous protein, and, by use of the surface calculation program SIMS, a legion of surface points may be visualized on the protein envelope (figures 2(*a*), (*c*)). Exploiting the normal vectors to these points and applying the programs HYDCRYST or HYDMODEL, a plethora of potential HOH points may be generated. The models presented in figures 2(*b*), (*d*)–(*f*) for different dot densities (0.1–3.0 Å^{-2}) clearly show variable numbers of possible water points at a definite distance from the protein surface.

By application of the hydration numbers for AA residues according to Kuntz [24] or Hopfinger [25] and several qualified selection criteria for the preferentially bound water molecules, the number of water molecules fixed to the protein surface can be reduced to reasonable numbers (about 2000 in the case of citrate synthase), compatible with experimental estimates for hydration of about 0.35 g of water/gram of protein [6]. As shown previously [14, 15], this may also be checked by predictions for solution scattering (R_G, V) and hydrodynamic parameters (s, D, $[\eta]$). However, for a fine tuning of hydrated protein models and resultant molecular parameters, the applied selection rules and hydration numbers must be scrutinized more critically. This also involves the possibility of triggering these input requirements somehow (cf figure 1).

The assignment of potential water points to individual AA residues is a geometric one and uses the structure of the protein in AT or AA coordinates and a distance criterion based on the centre–centre (CC) or centre–surface (CS) distance between water and the atoms or AA

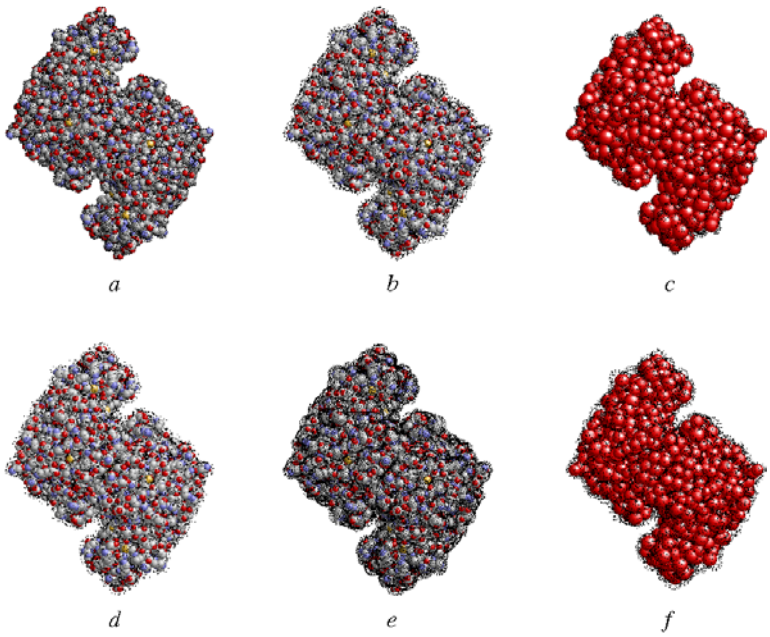

Figure 2. Selected space-filling models for anhydrous citrate synthase and molecular dot surface points created by the program SIMS for the anhydrous contour (*a*), (*c*) or, in combination with the programs HYDCRYST or HYDMODEL, for the contour of potential HOH points (*b*), (*d*)–(*f*). Models (*a*), (*b*), (*d*) and (*e*) are based on AT coordinates, and models *c* and *f* refer to AA coordinates. All dot surface points are calculated from the AT coordinates of citrate synthase; in model *c* only the dot surface points outside the spheres representing the AA residues are visible. For comparison of the obtained surface topographies, calculations with different dot densities were performed: $d_{dot} = 0.1$ (*d*), 1.0 (*a*)–(c), (*f*) or 3.0 (*e*) Å^{-2}. Graphics were produced with the program RasMol [37]. The basic atoms derived from the PDB file are shown in space-filling format and in the usual CPK colours (C in light grey, O in red, N in light blue, S in yellow etc); AA residues are displayed in red; dot surface points and potential HOH points are coloured black.

residues of the protein. The choice of different distance criteria, however, turned out to be of subordinate importance. Sorting (So) of assigned potential water points may be accomplished according to increasing CC or CS distances. A pure geometrical placing (Pl) of water on the surface may be achieved by use of d_w as a distance criterion (minimum distance between water molecules). The placing algorithm reduces N_w when using low scaling factors, whereas N_w is increased upon usage of high f values.

The water should be selected first for AAs of high preference to water. Otherwise steric constraints could give rise to an unrealistic preferred binding of water to AAs of low-binding behaviour. Therefore, multipass algorithms (selection step by step in the order of decreasing Kuntz or Hopfinger hydration numbers, or, alternatively, increase of hydration by increasing the scaling factors step by step) should be the method of choice as compared with single-pass procedures.

An efficient selection (Se) of water may be achieved on the basis of Kuntz or Hopfinger hydration numbers, using the scaling factors, f_K and f_H, and d_w as a distance criterion. Applying different values for f_K or f_H results in a quite different coverage of the protein envelope by surface water, corresponding to low ($f = 1$), intermediate ($f = 2$) or high (maximum) ($f = 10$) protein hydration (figure 3). The differences generated by usage of the hydration numbers by Kuntz or Hopfinger are minimal, their significance still decreasing with

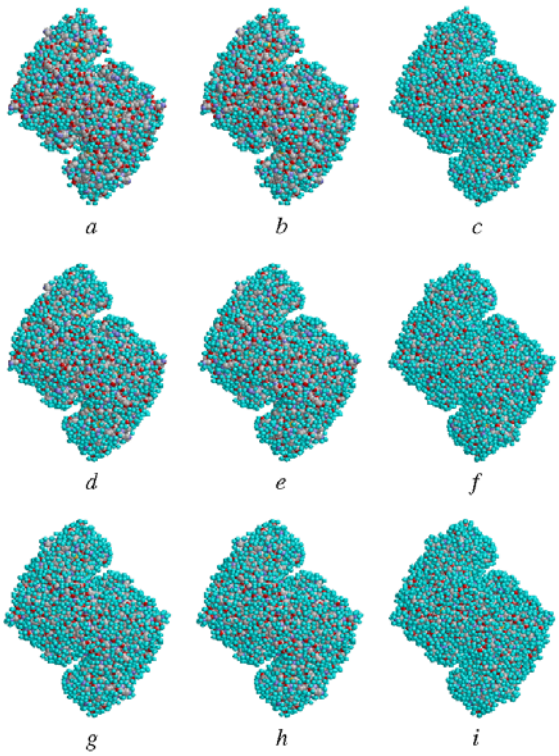

Figure 3. Selected space-filling AT models for hydrated citrate synthase, obtained after application of the hydration numbers given by Kuntz (*b*), (*c*), (*e*), (*f*), (*h*), (*i*) or Hopfinger (*a*), (*d*), (*g*), together with scaling factors f_K and f_H of 1 (*a*)–(*c*), 2 (*d*)–(*f*) or 10 (*g*)–(i). The surface calculations by SIMS are based on dot densities of $d_{dot} = 0.1$ (*a*), (*b*), (*d*), (*e*), (*g*), (*h*), 1.0 (*c*), 3.0 (*f*) or 5.0 (*i*) Å^{-2}. Hydration by the HYDCRYST algorithm used sorting and selection (So + Se) steps. Atoms are displayed in colours as in figure 2 and O atoms of bound water molecules are given in cyan.

increasing scaling factor. Of course, application of different dot densities, d_{dot}, influences the results additionally, showing more water molecules on the protein surface after application of higher dot densities.

To be able to compare the number and positions of bound water molecules, the water created by different hydration steps (So, Pl, Se) of the hydration algorithms was superimposed on the space-filling models. Figure 4 reveals that, indeed, application of all steps and their sequence is of relevance. Optimum results may be obtained for the sequence So + Pl + Se.

The critical comparison of the results obtained for hydrated citrate synthase by different calculation steps and modelling approaches (tables 1 and 2) in terms of number of bound water molecules, N_w, and further calculated parameters such as δ_1 or R_G supports the conclusions drawn already on inspection of the images: differences obtained by applying Kuntz or Hopfinger hydration numbers are marginal, N_w increases with enhanced d_{dot} (particularly with enhanced f), choice of the scaling factors, f_K or f_H, is very important (values of about two are realistic), the type of hydration scheme used is of relevance, application of So + Pl + Se steps is recommended and usage of AT coordinates should be preferred to AA coordinates.

Differences in the arrangement of water molecules bound preferentially to the protein surface are reflected by obvious differences in R_G. This is shown, for example, for application of the hydration numbers by Kuntz or Hopfinger, respectively (figure 5). Of course, R_G

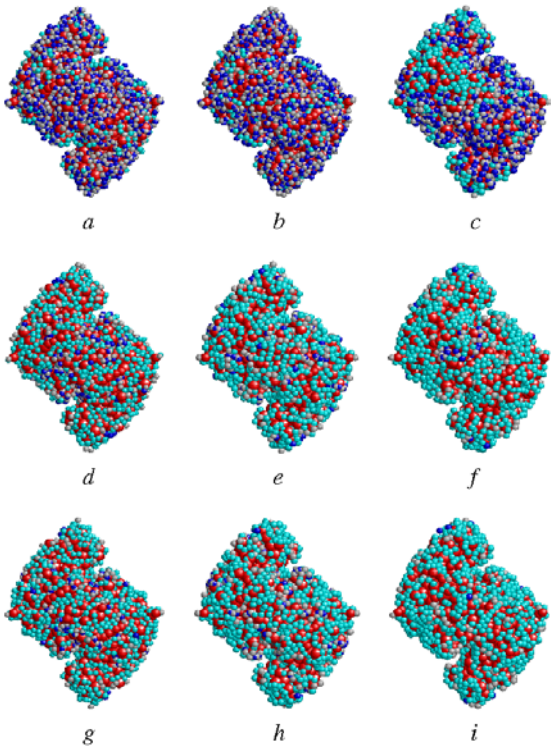

Figure 4. Pair-wise comparison of space-filling AA models for hydrated citrate synthase, obtained by use of the Kuntz and Hopfinger values for individual AA residues and different sorting (So), placing (Pl) and selection (Se) steps for the hydration process. Each image represents the superposition of two models. Bound water molecules common to both models are highlighted in cyan (position tolerance <0.1 Å), whereas water present in only one model is given in grey (model 1) or blue (model 2). AA residues are displayed in red. Images (*a*)–(*c*) refer to Kuntz values and Se versus So + Se, Se versus So + Pl + Se and So + Se versus So + Pl + Se steps, respectively. Images (*d*)–(*i*) use Kuntz or Hopfinger hydration numbers and Se (*d*), (*g*), So + Se (*e*), (*h*), and So + Pl + Se (*f*), (*i*) steps, respectively. The models are based on dot densities $d_{\mathrm{dot}} = 0.1$ (*a*)–(*f*) or 3.0 (*g*)–(*i*) Å^{-2} and values of unity for f_{K} and f_{H}. Hydration was performed by means of HYDMODEL.

increases with enhanced number of water molecules on the protein surface; however, it is only slightly different for the two hydration types (Kuntz and Hopfinger) and AT and AA approaches, and nearly identical for CC and CS distance criteria. Similarly, the predictions for the hydrodynamic parameters, s and D, differ with increasing numbers of N_{w}. This is shown in figure 6 for the diffusion coefficient D, the parameter initially obtained by means of HYDRO. The values obtained differ slightly for different voxel sizes required for creating efficient data reduction.

Finally, comparing the SAXS functions, $I(h)$ and $p(r)$, may be used as additional powerful tool for testing the validity of different calculation procedures. Both functions reveal marked differences for the chosen approaches, as may be taken, for example, from the normalized scattering curves, $I(h)$ versus h, in figure 7. $I(h)$ of the anhydrous protein models and of the models hydrated according to Kuntz or Hopfinger and different scaling factors and dot densities are significantly different. This obviously reflects quite different SAXS functions for low, intermediate and maximum hydration of proteins. Presumably as a consequence

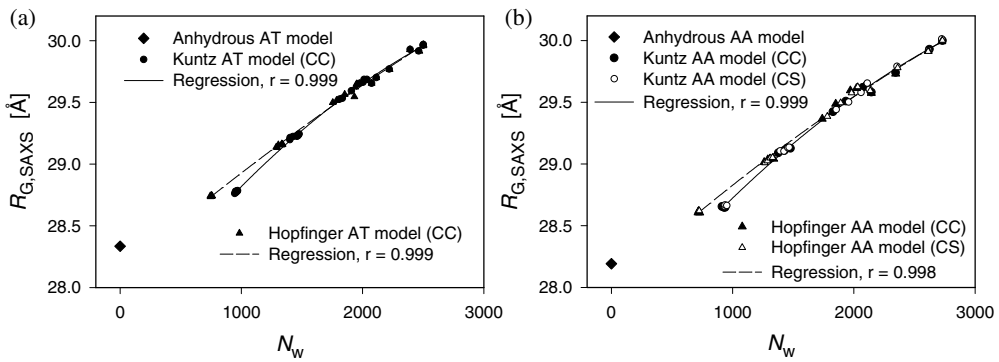

Figure 5. Radii of gyration, $R_{G,SAXS}$, from AT or AA coordinates and excess electrons versus number of bound water molecules, N_w, of models for hydrated citrate synthase obtained by use of Kuntz or Hopfinger hydration numbers. Calculations of water molecules are based on SIMS points and AT (*a*) or AA (*b*) coordinates, sorting and selection (So + Se) hydration steps, using CC or CS distance criteria. For the calculation of $R_{G,SAXS}$, the volume of a bound water molecule was assumed to be 24.5 Å^3. Second-order regression curves are added. For comparison $R_{G,SAXS}$ of the anhydrous model is also given.

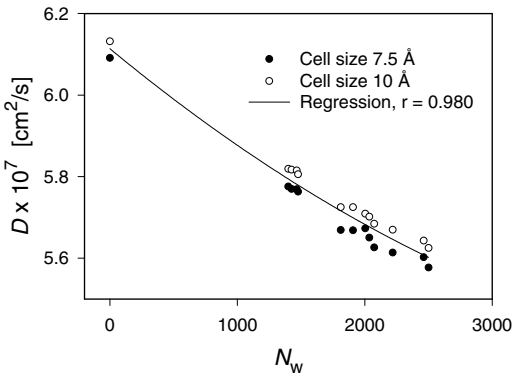

Figure 6. Diffusion coefficients, D, from hydrodynamic modelling utilizing the program HYDRO versus number of bound water molecules, N_w, of models for hydrated citrate synthase obtained by use of Kuntz hydration numbers. Calculations of water molecules are based on SIMS points and AT coordinates, sorting and selection (So + Se) hydration steps, using CC distance criteria. For the calculations of D, the volume of a bound water molecule was assumed to be 24.5 Å^3, and voxel sizes of 7.5 or 10 Å were applied for establishing reduced models. Second-order regression curves are added. For comparison D of the anhydrous model is also given.

of insufficiencies in the experimental SAXS curve, the experimental conditions applied and permanent fluctuations of water density on the protein surface, at present no clear-cut decision in favour of definite amounts of bound water can be made. However, the experimental results rather point to an intermediate binding of water to proteins than to very low or very high contributions of preferentially bound water. This finding is also in agreement with the results from other physicochemical techniques [6].

Table 1. Comparison of calculated parameters of hydrated citrate synthase as obtained by different modelling approaches.

Input [a]			Output					
			AT coordinates			AA coordinates		
SIMS d_{dot}^{b} ($Å^{-2}$)	Hydration type [c]	Scaling factor (f_K or f_H)	N_w	δ_1 (g g^{-1})	R_G (Å)	N_w^d	δ_1 (g g^{-1})	R_G^d (Å)
0.1	K	1	1399	0.258	29.20	1377 ± 17	0.253	29.08 ± 0.01
0.1	H	1	1296	0.239	29.13	1254 ± 14	0.231	29.00 ± 0.01
3.0	K	1	1465	0.270	29.23	1409 ± 43	0.259	29.13 ± 0.03
3.0	H	1	1335	0.246	29.16	1271 ± 38	0.234	29.05 ± 0.04
5.0	K	1	1477	0.272	29.24	1430 ± 50	0.263	29.14 ± 0.01
5.0	H	1	1334	0.246	29.16	1285 ± 49	0.237	29.06 ± 0.04
0.1	K	2	1811	0.333	29.53	1834 ± 12	0.338	29.43 ± 0.01
0.1	H	2	1754	0.323	29.50	1752 ± 23	0.323	29.38 ± 0.01
3.0	K	2	2003	0.369	29.66	2031 ± 26	0.374	29.62 ± 0.05
3.0	H	2	1939	0.357	29.62	1930 ± 31	0.355	29.58 ± 0.04
5.0	K	2	2035	0.375	29.69	2084 ± 28	0.384	29.66 ± 0.02
5.0	H	2	1974	0.363	29.65	1979 ± 33	0.364	29.62 ± 0.05
0.1	K	10	2074	0.382	29.66	2123 ± 59	0.391	29.58 ± 0.02
0.1	H	10	2074	0.382	29.66	2140 ± 40	0.394	29.57 ± 0.01
3.0	K	10	2463	0.454	29.92	2679 ± 148	0.493	29.97 ± 0.09
3.0	H	10	2468	0.454	29.91	2682 ± 147	0.494	29.96 ± 0.09
5.0	K	10	2500	0.460	29.97	2822 ± 109	0.520	30.06 ± 0.07
5.0	H	10	2502	0.461	29.97	2794 ± 144	0.514	30.04 ± 0.09

[a] Further input data: coordinates of 6888 atoms or 874 AAs.
[b] The dot densities $d_{dot} = 0.1$, 3.0 and 5.0 $Å^{-2}$ correspond to 24 524, 109 836 and 171 393 surface dots and normal vectors, respectively.
[c] K, Kuntz; H, Hopfinger.
[d] Parameters are averages of values calculated from different variants of the hydration algorithms (sorting, placing, selection, CC, CS).

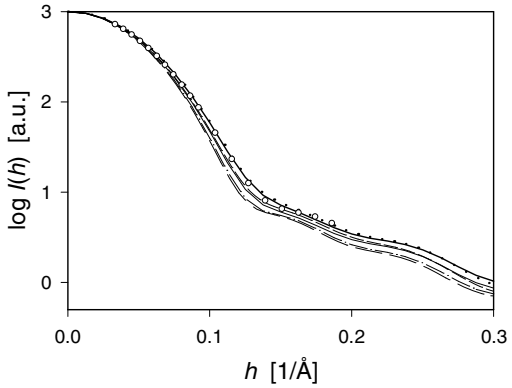

Figure 7. Normalized scattering curves, $I(h)$, of selected models for anhydrous and hydrated citrate synthase, where $h = 4\pi \sin\theta/\lambda$ (2θ = scattering angle, λ = wavelength). The selected models comprise only a few representative examples: anhydrous AT model (——); anhydrous AA model (· · · · · ·); Kuntz AT model, $d_{dot} = 0.1$ $Å^{-2}$, $f_K = 1$, CC (——); Kuntz AA model, $d_{dot} = 0.1$ $Å^{-2}$, $f_K = 1$, CC (- - - -); Kuntz AT model, $d_{dot} = 0.1$ $Å^{-2}$, $f_K = 10$, CC (— · —); Kuntz AT model, $d_{dot} = 5.0$ $Å^{-2}$, $f_K = 10$, CC (— — —); experimental SAXS function (○).

Table 2. Comparison of calculated bound water molecules of hydrated citrate synthase as obtained by different modelling approaches.

Input					N_w		
Type of coordinate	SIMS d_{dot}^{a} (Å $^{-2}$)	Hydration type[b]	Scaling factor (f_K or f_H)	Hydration schemes[c]	Present in both models[d]	Present only in model 1	Present only in model 2
			Comparison of Kuntz and Hopfinger values:				
AT	0.1	K versus H	1		1212	187	84
AT	1.0	K versus H	1		1197	228	108
AT	3.0	K versus H	1		1226	239	109
AT	5.0	K versus H	1		1221	256	113
AT	0.1	K versus H	10		2065	9	9
AT	1.0	K versus H	10		2207	12	16
AT	3.0	K versus H	10		2453	10	15
AT	5.0	K versus H	10		2474	26	28
			Comparison of different SIMS dot densities:				
AT	0.1 versus 1.0	K	1		642	757	783
AT	0.1 versus 5.0	K	1		387	1012	1090
AT	1.0 versus 5.0	K	1		369	1056	1108
AT	0.1 versus 1.0	K	10		758	1316	1461
AT	0.1 versus 5.0	K	10		439	1635	2061
AT	1.0 versus 5.0	K	10		423	1796	2077
AT	0.1 versus 1.0	K	10		1272[e]	802	947
AT	0.1 versus 5.0	K	10		1210[e]	864	1290
			Comparison of Kuntz and Hopfinger values:				
AA	0.1	K versus H	1	Se	1105	275	154
AA	0.1	K versus H	1	So + Se	1160	217	100
AA	0.1	K versus H	1	So + Pl + Se	1147	217	88
AA	3.0	K versus H	1	Se	1137	301	168
AA	3.0	K versus H	1	So + Se	1185	267	122
AA	3.0	K versus H	1	So + Pl + Se	1213	171	37

Table 2. (Continued.)

Input					N_w		
Type of coordinate	SIMS d_{dot}[a] (Å^{-2})	Hydration type[b]	Scaling factor (f_K or f_H)	Hydration schemes[c]	Present in both models[d]	Present only in model 1	Present only in model 2
			Comparison of different hydration schemes:				
AA	0.1	K	1	Se versus So + Se	287	1093	1090
AA	0.1	K	1	Se versus So + Pl + Se	218	1162	1146
AA	0.1	K	1	So + Se versus So + Pl + Se	584	793	780
			Comparison of different SIMS dot densities:				
AA	0.1 versus 5.0	K	1	So + Se	776	601	702
AA	0.1 versus 5.0	K	1	So + Pl + Se	102	1262	1292
AA	0.1 versus 5.0	K	1	So + Se	1103[e]	274	375
AA	0.1 versus 5.0	K	1	So + Pl + Se	494[e]	870	900
			Comparison of AT and AA coordinates:				
AT versus AA	0.1	K	1	So + Se	220	1179	1157
AT versus AA	5.0	K	1	So + Se	120	1357	1358
AT versus AA	0.1	K	10	So + Se	558	1516	1586
AT versus AA	5.0	K	10	So + Se	209	2291	2522
AT versus AA	0.1	K	10	So + Se	1005[e]	1069	1139
AT versus AA	5.0	K	10	So + Se	1215[e]	1285	1516

[a] The dot densities $d_{dot} = 0.1$, 1.0, 3.0 and 5.0 Å^{-2} correspond to 24 524, 45 129, 109 836 and 171 393 surface dots and normal vectors, respectively.
[b] K, Kuntz; H, Hopfinger.
[c] Se, selection; Pl, placing; So, sorting.
[d] Unless otherwise stated, all water molecules differing in their positions by less than 0.1 Å were considered as being common to both models.
[e] Based on an overlap of more than 50% of the volume of water molecules.

4. Conclusions

The behaviour of proteins in aqueous solution is influenced by the molecular characteristics of the biopolymer and the interactions with the solvent water. Though many attempts have been made in solution biophysics to elucidate the role of water, the solvation/hydration problem is still the most important obstacle for understanding the solution behaviour of biopolymers in detail, including the precise interpretation of results stemming from scattering and hydrodynamic techniques (see [6] and references therein).

The development of hydration strategies is based on two essential prerequisites: knowledge of the precise surface topography and of qualified assumptions regarding properties and localization of ordered water on the protein surface. The surface calculation program SIMS [31] allows the precise visualization of the surface topography of proteins. Based on experimental or theoretical findings for peptides, hydration contributions of individual AA residues have been derived by Kuntz [24], Hopfinger [25] and others. These hydration numbers may be exploited advantageously for the development of hydration algorithms, to account realistically for appropriate hydration contributions of ordered water on the protein surface. Though the residence times of ordered water are very short, the fact that both the constituents of simple proteins and the ordered water molecules exhibit average properties results in realistic modelling procedures. Several input parameters used for fine tuning of the results have now been tested critically, among them the scaling of the original hydration numbers and the sequence to be performed for placing water on the protein surface being the most important ones.

Acknowledgments

The authors are much obliged to Y N Vorobjev for making available the program SIMS and to J García de la Torre for the program HYDRO. Computer graphics were produced with the program RasMol by R A Sayle.

References

[1] Creighton T E 1993 *Proteins: Structures and Molecular Properties* 2nd edn (New York: Freeman)
[2] Perkins S J 1986 *Eur. J. Biochem.* **157** 169
[3] Gerstein M and Chothia C 1996 *Proc. Natl Acad. Sci. USA* **93** 10 167
[4] Svergun D I, Richard S, Koch M H J, Sayers Z, Kuprin S and Zaccai G 1998 *Proc. Natl Acad. Sci. USA* **95** 2267
[5] Ebel C, Eisenberg H and Ghirlando R 2000 *Biophys. J.* **78** 385
[6] Durchschlag H and Zipper P 2001 *Biophys. Chem.* **93** 141
[7] Kuntz I D Jr and Kauzmann W 1974 *Adv. Protein Chem.* **28** 239
[8] Thanki N, Thornton J M and Goodfellow J M 1988 *J. Mol. Biol.* **202** 637
[9] Otting G, Liepinsh E and Wüthrich K 1991 *Science* **254** 974
[10] Rupley J A and Careri G 1991 *Adv. Protein Chem.* **41** 37
[11] Oobatake M and Ooi T 1993 *Prog. Biophys. Mol. Biol.* **59** 237
[12] Levitt M and Park B H 1993 *Structure* **1** 223
[13] Murphy L R, Matubayasi N, Payne V A and Levy R M 1998 *Folding Design* **3** 105
[14] Zipper P and Durchschlag H 2002 *Physica* A **304** 283
[15] Durchschlag H and Zipper P 2002 *Prog. Colloid Polym. Sci.* **119** 131
[16] Remington S, Wiegand G and Huber R 1982 *J. Mol. Biol.* **158** 111
[17] Wiegand G, Remington S, Deisenhofer J and Huber R 1984 *J. Mol. Biol.* **174** 205
[18] Durchschlag H, Zipper P, Wilfing R and Purr G 1991 *J. Appl. Crystallogr.* **24** 822
[19] Durchschlag H, Zipper P, Purr G and Jaenicke R 1996 *Colloid Polym. Sci.* **274** 117
[20] Wu J-Y and Yang J T 1970 *J. Biol. Chem.* **245** 212
[21] Singh M, Brooks G C and Srere P A 1970 *J. Biol. Chem.* **245** 4636

[22] Sussman J L, Lin D, Jiang J, Manning N O, Prilusky J, Ritter O and Abola E E 1998 *Acta Crystallogr.* D **54** 1078
[23] Bairoch A and Apweiler R 2000 *Nucleic Acids Res.* **28** 45
[24] Kuntz I D 1971 *J. Am. Chem. Soc.* **93** 514
[25] Hopfinger A J 1977 *Intermolecular Interactions and Biomolecular Organization* (New York: Wiley)
[26] Kuntz I D 1971 *J. Am. Chem. Soc.* **93** 516
[27] Hopfinger A J 1971 *Macromolecules* **4** 731
[28] Hopfinger A J and Battershell R D 1976 *J. Med. Chem.* **19** 569
[29] Zipper P and Durchschlag H 1997 *Prog. Colloid Polym. Sci.* **107** 58
[30] Zipper P and Durchschlag H 1998 *Biochem. Soc. Trans.* **26** 726
[31] Vorobjev Y N and Hermans J 1997 *Biophys. J.* **73** 722
[32] García de la Torre J, Navarro S, López Martínez M C, Díaz F G and López Cascales J J 1994 *Biophys. J.* **67** 530
[33] García de la Torre J, Huertas M L and Carrasco B 2000 *Biophys. J.* **78** 719
[34] Carrasco B, García de la Torre J and Zipper P 1999 *Eur. Biophys. J.* **28** 510
[35] Glatter O and Kratky O (eds) 1982 *Small Angle X-Ray Scattering* (London: Academic)
[36] Durchschlag H and Zipper P 2002 *Prog. Colloid Polym. Sci.* **119** 121
[37] Sayle R A and Milner-White E J 1995 *Trends Biochem. Sci.* **20** 374