

The implementation of SOMO (SOLution MODeller) in the UltraScan analytical ultracentrifugation data analysis suite: enhanced capabilities allow the reliable hydrodynamic modeling of virtually any kind of biomacromolecule

Emre Brookes · Borries Demeler · Camillo Rosano ·
Mattia Rocco

Received: 14 November 2008 / Revised: 29 January 2009 / Accepted: 2 February 2009
© European Biophysical Societies' Association 2009

Abstract The interpretation of solution hydrodynamic data in terms of macromolecular structural parameters is not a straightforward task. Over the years, several approaches have been developed to cope with this problem, the most widely used being bead modeling in various flavors. We report here the implementation of the SOMO (SOLution MODeller; Rai et al. in *Structure* 13:723–734, 2005) bead modeling suite within one of the most widely used analytical ultracentrifugation data analysis software packages, UltraScan (Demeler in *Modern analytical ultracentrifugation: techniques and methods*, Royal Society of Chemistry, UK, 2005). The US-SOMO version is now under complete graphical interface control, and has been freed from several constraints present in the original implementation. In the direct beads-per-atoms method, virtually any kind of residue as defined in the Protein Data

Bank (e.g., proteins, nucleic acids, carbohydrates, prosthetic groups, detergents, etc.) can be now represented with beads whose number, size and position are all defined in user-editable tables. For large structures, a cubic grid method based on the original AtoB program (Byron in *Biophys J* 72:408–415, 1997) can be applied either directly on the atomic structure, or on a previously generated bead model. The hydrodynamic parameters are then computed in the rigid-body approximation. An extensive set of tests was conducted to further validate the method, and the results are presented here. Owing to its accuracy, speed, and versatility, US-SOMO should allow to fully take advantage of the potential of solution hydrodynamics as a complement to higher resolution techniques in biomacromolecular modeling.

Keywords Macromolecular hydrodynamics · Bead modeling · Analytical ultracentrifugation · Protein structure and dynamics · NMR spectroscopy · X-ray crystallography

AUC&HYDRO 2008—Contributions from 17th International Symposium on Analytical Ultracentrifugation and Hydrodynamics, Newcastle, UK, 11–12 September 2008.

Electronic supplementary material The online version of this article (doi:10.1007/s00249-009-0418-0) contains supplementary material, which is available to authorized users.

E. Brookes · B. Demeler
Department of Biochemistry, The University of Texas Health Science Center at San Antonio, San Antonio, TX, USA

C. Rosano
Nanobiotechnologie, Istituto Nazionale per la Ricerca sul Cancro (IST), Genoa, Italy

M. Rocco (✉)
Biopolimeri e Proteomica, Istituto Nazionale per la Ricerca sul Cancro (IST), IST c/o CBA,
Largo R. Benzi 10, 16132 Genoa, Italy
e-mail: mattia.rocco@istge.it

Introduction

Solution hydrodynamic parameters of macromolecules, such as the translational (D_t) and rotational (D_r) diffusion and sedimentation (s) coefficients, and the intrinsic viscosity ($[\eta]$), can be experimentally determined by well-established techniques. Since the size, detailed shape, and time-dependent conformation determine the macromolecules' frictional properties, the computation of these parameters from their structures has been a field of intense research. These calculations are, however, not straightforward. Well-defined geometrical objects, such as cylinders and ellipsoids, have been used initially to build very low

resolution models of proteins (Tanford 1961; Cantor and Schimmel 1980) and other biopolymers, and are still in use today in an enhanced version (Harding et al. 2004). A big step forward was the development of the theory for the computation of translational and rotational frictional coefficients and intrinsic viscosity of ensembles of non-overlapping spheres (beads) of differing radii (reviewed in García de la Torre and Bloomfield 1981; Spotorno et al. 1997; Carrasco and García de la Torre 1999). This procedure has been extended in a number of different ways to model proteins and other biomacromolecules of known 3D structure, ranging from shell modeling to grid-based methods (see Byron 2000). However, the calculation of the hydrodynamic parameters of ensemble of beads can be computationally demanding, requiring a compromise between the bead model resolution and the number and size of the beads employed. Furthermore, the effect of the so-called water of hydration (Halle and Davidovic 2003) should be correctly taken into account (see Rai et al. 2005).

Currently, three principal different bead modeling methods are available, implemented in public-domain computer programs. A “grid” method was implemented by O. Byron in the program AtoB (Byron 1997). Here, the protein is subdivided into equally sized cubes and each residue is assigned to a particular cube. Then, according to user choice, beads of either equal or differing radii are generated and placed in the center of gravity of each cube, the resolution of the final model depending on the spacing of the initial cubic grid. AtoB (Byron 1997) was tested against a large globular protein (aldolase) and a spherical hollow protein (apoferritin). Bead models were generated and the calculated $s_{(20,w)}^0$ and $[\eta]$ values agreed well with experimental values, provided that an appropriate grid spacing was used and after radial expansion of the beads to compensate for the water of hydration. Zipper and Durchschlag (1997, 1998) also followed a similar approach, and the usefulness of such methods appears to rest mainly in the modeling of very large structures.

At the other end of the spectrum lies the “shell modeling” approach implemented in the currently most widely used bead modeling program, HYDROPRO, developed by García de la Torre and collaborators (García de la Torre et al. 2000; García de la Torre 2001). In this approach, all atoms in a protein are first replaced by equally sized beads of a certain radius. Then, the surface of this “primary” model is covered with a “shell” of smaller beads, and the procedure is iterated, decreasing the shell beads’ radius, allowing extrapolation to zero bead size. This approach has undergone more extensive testing (García de la Torre et al. 2000; García de la Torre 2001), and the models can *on average* reasonably reproduce the hydrodynamic parameters determined experimentally (albeit without a critical evaluation of the literature data, see below). Furthermore,

to reach a consensus agreement across the test proteins, the primary beads’ radius was adjusted until a mean satisfactory value was found. In addition, to avoid excessive memory requirements and very long computing times, HYDROPRO currently has an upper limit of $\sim 3,000$ shell beads, whose radius is a function of the protein’s size, potentially limiting its precision when large structures are analyzed.

A third approach is to build a bead model with direct correspondence between the atoms in the macromolecules’ residues, such as amino acids in proteins, sugar units in carbohydrates and nucleosides in nucleic acids, and the beads that are used to represent them. This approach can overcome some of the limitations of the other methods, and was chosen for the development of SOMO (SOLUTION MOdeller; Rai et al. 2005), where, for instance, amino acid residues are represented each by two beads, one for the main-chain and another for the side-chain segments. The beads’ volumes are initially determined by summing the volumes of the atoms which they represent, and are then augmented by adding the volume of the water molecules which were experimentally found to be statistically bound to each residue (e.g., Kuntz and Kauzmann 1974). Beads are positioned according to the characteristics of the residues they represent, and the overlaps between them are then removed by proportional radial reduction, trying to preserve the original anhydrous surface envelope as much as possible. This is aided by an accessible surface area (ASA) computation initially performed on the atomic structure to separate exposed beads from buried beads. Buried beads can then be excluded from subsequent hydrodynamic parameters computations, which in the original SOMO implementation were carried out separately by the program SUPCW (Spotorno et al. 1997; Rai et al. 2005). SOMO was extensively tested against three small proteins, BPTI, RNase A and lysozyme, for which a very large body of hydrodynamic data exists (that were critically evaluated), plus two larger proteins, fibrinogen fragment D and citrate synthase dimer, with very good results (Rai et al. 2005). SOMO has already been instrumental in discriminating between alternative conformation of integrins in a recently published study (Rocco et al. 2008). However, the original SOMO implementation suffered from a number of drawbacks. SOMO consisted of a collection of separate, command-line driven executables running under the Linux operating system, with a rather rigid user interface, recognizing only residues hard-coded in the programs. To overcome these flaws, an entirely re-designed and enhanced version of the SOMO program was developed by the authors by integrating the basic functionality of SOMO under the open source software UltraScan (US). First, we added a graphical user interface (GUI) and replaced the hard coded residue representation by

implementing user-modifiable reference tables which code for the atomic groups and residues present in the Protein Data Bank (PDB; Berman et al. 2000) structures. A full range of options controlling details of the modeling process and of the hydrodynamic computations (performed with an integrated version of SUPCW; see Spotorno et al. 1997 and Rai et al. 2005) can be accessed through dedicated menus. In the process, we have also corrected some mistakes that went unnoticed in the original SOMO release, and added new features. Finally, a module for the creation of bead models based on the AtoB grid method (Byron 1997), further developed by M. Nöllman (Centre de Biochimie Structurale, CNRS-INSERM, Montpellier, FR) for the original SOMO program (Rai et al. 2005), has been coded for US-SOMO. This allows either a further reduction of resolution starting from a previously generated bead model, or the direct generation of grid-based models from PDB files or small-angle X-ray scattering (SAXS)-derived dummy atoms models. This first US-SOMO release was tested with an expanded number of X-ray crystallography and NMR spectroscopy structures, as presented by García de la Torre (2001), whose experimental hydrodynamic parameters were, however, critically re-evaluated. In the Electronic Supplementary Material (ESM) of this paper, a detailed description of the operation and main features of US-SOMO is presented. The very satisfactory results of US-SOMO in reproducing most experimental parameters of the test proteins are here reported and discussed, highlighting its potential as a powerful tool for many hydrodynamic modeling applications.

Methods

US-SOMO implementation: general layout, reference tables and options

In Fig. 1 we present the main GUI panel of the new SOMO implementation under UltraScan, which can be accessed from the US “Simulation” drop down menu. The program is divided into three sub-menus. “Modify Lookup Tables:” refers to the reference files needed to operate the program; settings of various modeling and computational options are listed under “Modify SOMO Options for:”; and “Run SOMO Program:” refers to the various runtime operations. The right-side window updates the user about the operation(s) in progress. The US-SOMO operations are described in detail in the ESM.

At the core of the program lies its capability to read and interpret PDB-formatted structural files. US-SOMO will upload a PDB file and recognize only the relevant records, discarding all others. Currently, these include the ATOM, HETATM, MODEL, ENDMODEL, TER, and END records. Within the

ATOM and HETATM records, US-SOMO extracts and loads the atom name, the residue name, the chain identifier, the residue sequence number, and the x , y and z coordinates into appropriate data structures. The atom and residue names are then compared with the records present in the *somo.residue* table, which can be edited by the user through a pop-up window accessed by pressing the “Add/Edit Residue” button in the main panel.

Each residue type present in the PDB file must be correctly described in the *somo.residue* table, and, in order to have maximum flexibility in coding for all possible residues, two other tables were defined. In the first one, *somo.hybrid*, the different atomic groups are listed, together with their fundamental properties, i.e., the mass and the atomic van der Waals (VdW) radius, according to the “hybridizations” described by Tsai et al. (1999). The current content of the *somo.hybrid* file is shown in Table S1 in the ESM, and users can edit the current definitions or add new atomic groups through the “Add/Edit Hybridization” menu (not shown). The atomic groups listed in *somo.hybrid* are then used to build the *somo.atom* table through the “Add/Edit Atom” menu (not shown). A brief excerpt of the current entries in this table is shown in Table S2 in the ESM, which shows that the PDB coding for atoms does not discriminate between different hybridization states. For instance, CB is bound to three hydrogen atoms in alanine (C4H3), to just one in leucine, isoleucine and threonine (C4H1), and to two in all other amino acids (C4H2), implying a mass difference as shown in Table S2. A more profound difference is found, as an example, between the CG in leucine, having four single bonds and one hydrogen atom bound (C4H1), and that in histidine, having two single and one double bonds, and no hydrogens bound (C3H0). In this case, not only is the molecular weight different, but also the atomic van der Waals (VdW) radius is different. Thus, the *somo.atom* file allows selection of the correct atom name/molecular weight/VdW radius combination for each of the atoms within a residue. A full description of the operations necessary to enter/edit a residue in the *somo.residue* table is presented in the ESM.

Technical details

US-SOMO is written in C++ and linked against the UltraScan (Demeler 2005) and Qt (TrollTech.com: Qt—a cross-platform application framework. <http://www.trolltech.com/>) libraries. The code is licensed under the GPL license (The GNU General Public License Version 3. <http://www.gnu.org/copyleft/gpl.html>) and can be downloaded from the UltraScan wiki (The UltraScan Trac Wiki. <http://wiki.bcf.uthscsa.edu/ultrascan/>). Binaries for all major

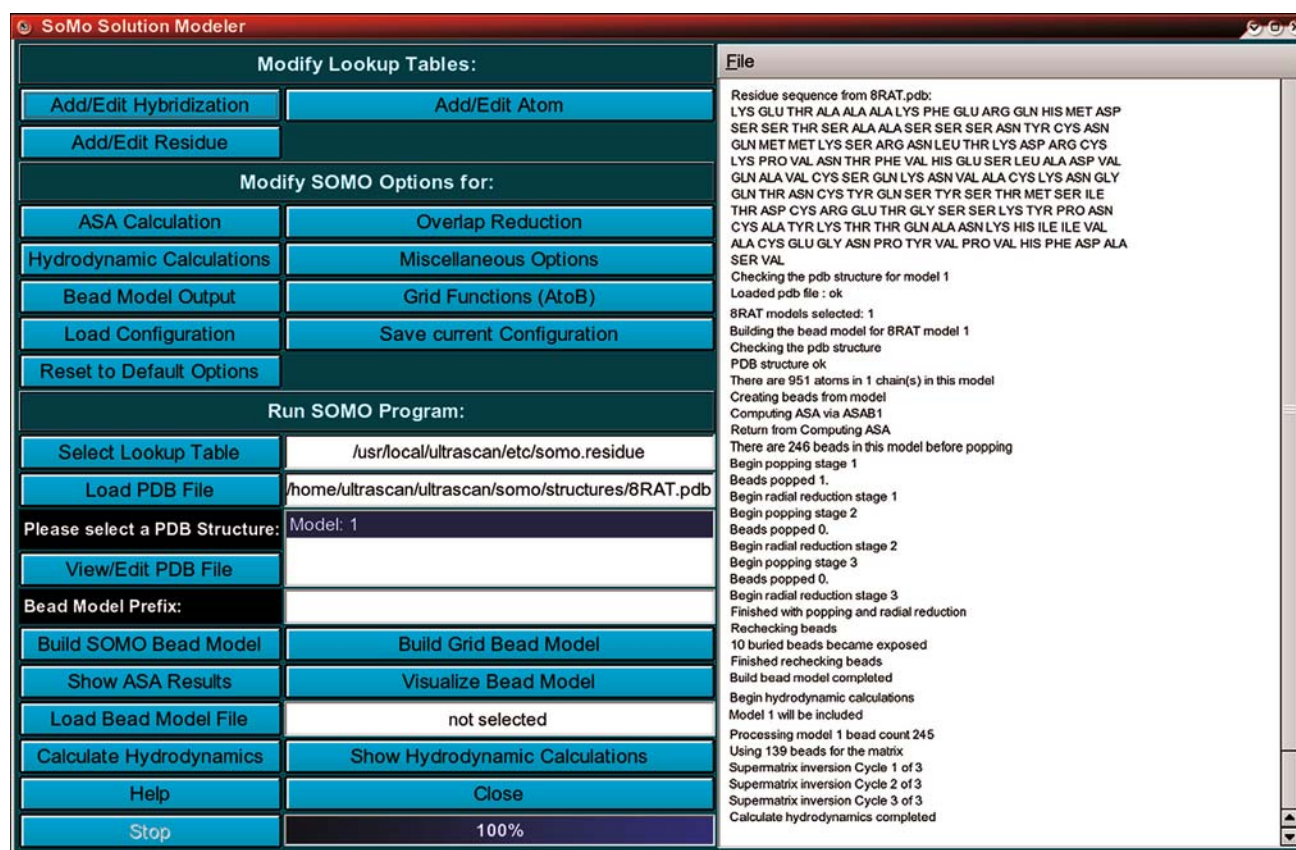


Fig. 1 Main panel of the US-SOMO program, shown after processing the 8RAT.pdb file. The font size in the right-side window has been artificially reduced to show the entire process

platforms (Linux/X11, Microsoft Windows, Macintosh OS-X) can be downloaded from the UltraScan website at <http://www.ultrascan.uthscsa.edu>.

Experimental hydrodynamic data

All experimental hydrodynamic parameters of the proteins used to test US-SOMO were taken from the literature, but with a critical evaluation of the conditions used and of the correctness, whenever possible, of their reduction to standard conditions (water at 20°C). A full list is presented in the ESM, with the appropriate references.

Protein structures

The high-resolution structures of the test proteins were taken from the PDB (<http://www.rcsb.org/pdb/home/home.do>). Whenever possible, we sought structures deriving from the same species from which the solution data were available. This explains some differences between the structures we have employed and those previously used (García de la Torre et al. 2000; García de la Torre 2001). The completeness of each structure was checked and ensured at two levels: missing atoms within side chains

were automatically added by the WHATIF webserver (Vriend 1990; <http://swift.cmbi.ru.nl/servers/html/index.html>) while missing residues were mostly manually modeled using O (Jones et al. 1991). A relatively long C-terminal sequence in nitrogenase MoFe was generated by Robetta (Chivian et al. 2005; <http://robetta.bakerlab.org/>) using the ab initio protocol (Bonneau et al. 2002), and then pasted in the original structure using O.

Results and discussion

The US-SOMO implementation was firstly thoroughly tested against the original SOMO software (Rai et al. 2005). In the process, several minor bugs were fixed, the most significant involved an incorrect formulation of the outward translation when reducing exposed side chains beads (see Rai et al. 2005). Of the two ASA algorithms implemented, SurfRace (Tsodikov et al. 2002) was found to be very reliable for small, compact structures, but presented some problems with larger, multisubunit structures. Therefore, in all subsequent work we used the ASAB1 option based on the Lee and Richards (1971) rolling sphere method, which is also the only option implemented for

re-checking the beads' exposure after overlap reduction. Another change affects the threshold detection for activating bead fusion ("popping") which is now done by computing the intersection volume of pairs of beads. The pair of beads is fused when the volume of either bead multiplied by the user defined percentage overlap is greater than the volume of intersection. The volume of the fused bead is the total volume of the pair of beads.

The testing against protein structures was performed in three phases. In the first, multiple structures for the same protein, originating from both X-ray crystallography and NMR spectroscopy, were used. For the latter, averages of the hydrodynamic parameters computed for each of the multiple conformations present in the models were performed. The test proteins chosen, for which an extensive body of experimental hydrodynamic data exist, are the same utilized in Rai et al. (2005), bovine pancreatic trypsin inhibitor (BPTI), bovine pancreatic ribonuclease (RNase), and hen egg white lysozyme, to which myoglobin was added. A second set included the other proteins utilized by García de la Torre and collaborators (García de la Torre et al. 2000; García de la Torre 2001), excluding some less characterized proteins (trypsin, pepsin and subtilisin). Instead, we have examined in more detail hemoglobin, glyceraldehyde-3-phosphate dehydrogenase (G3PD) and lactate dehydrogenase (LDH), for which data and structures coming from different species exist. For these two sets, we computed and compared $D_{t(20,w)}^0$, $s_{(20,w)}^0$, $\tau_{c(20,w)}^h$, and $[\eta]$, whose experimental values were critically assessed as reported in Tables S3–S5 of the ESM. Finally, the $\tau_{c(20,w)}^h$ values only were computed for the full protein set presented in Table 2 of García de la Torre (2001), after recalculation of the reduction to standard conditions of the experimental values as presented in Table S6. In all our modeling, we kept some options fixed, including: (1) a popping threshold of 40% for exposed side chains and of 60% for exposed main chain beads (this differs from what was used by Rai et al. (2005) because of the new definition of overlap threshold implemented in US-SOMO, see above); (2) the hierarchical overlap removal procedure was used in all cases, with outward translation for the exposed side chain beads; (3) the computations of the hydrodynamic parameters were done with stick boundary conditions, referred to the diffusion center, and with exclusion of the buried beads from both the full computations and from the volume correction; (4) the molecular weights and partial specific volumes used for the computation of $s_{(20,w)}^0$ and $[\eta]$ were those computed by US-SOMO from the composition.

In Table 1, the comparisons between experimental and calculated $D_{t(20,w)}^0$ and $s_{(20,w)}^0$ for BPTI, RNase, lysozyme and myoglobin are presented. Taking full advantage of the ease by which some modeling options can be now set in

US-SOMO, we explored the influence of ASA thresholds on these parameters. Practically, increasing the residues' ASA threshold labels more beads as buried, and increasing the ASA re-check threshold also keeps more beads in the buried category. The effect on the two parameters examined in Table 1 is, therefore, entirely due to the number and position of the beads employed in the computations. However, as we will see later in Table 2 (and Tables 4, 5), it has an additional impact on the $\tau_{c(20,w)}^h$ and $[\eta]$ values because of the exclusion of the buried beads from the volume correction. The three conditions examined are residues' ASA thresholds of 10, 20 and 40 Å² (A10, A20, A40), coupled respectively with beads' ASA re-check thresholds of 30, 50 and 60% (R30, R50, R60). For comparison, the A10/R30 condition is equivalent to that employed by Rai et al. (2005) in their modeling study.

The first interesting result from Table 1 is that the $D_{t(20,w)}^0$ value of three out of the four test proteins examined here is reproduced extremely well, within the 3% error of experimental data. Moreover, it is independent of the structure used to generate the models, with no appreciable differences between X-ray and NMR models. The lone exception is lysozyme, for which the X-ray-derived structures perform slightly worse than in the other cases examined, while the NMR structure is in excellent agreement (<1%). This was already noticed by Rai et al. (2005), who suggested that this effect is mainly due to the high number of long, hydrophilic residues on lysozyme surface, not fully extended in crystal structures due to crystal packing. As for the $s_{(20,w)}^0$, the results are mixed, with an excellent agreement for RNase ($\leq 3\%$) and for the NMR-derived model of lysozyme ($\leq 2\%$), while the X-ray-derived models of the latter suffer from the same problem seen with $D_{t(20,w)}^0$. The poor agreement of the myosin CO $s_{(20,w)}^0$ data are instead likely due to a suspicious experimental value ("?" in Table 1), since a minor change is observed in the $D_{t(20,w)}^0$ data between the CO and apo forms. As it will be discussed in more detail below, the computed value of the partial specific volume \bar{v}_2 could also affect the reliability of these numbers.

The other important evidence derived from Table 1 is the very small effect of greatly reducing the number of the beads used in the computations by increasing the ASA thresholds. Practically, halving the number of beads decreases the accuracy by roughly 1%. This is a surprising result, and it will be further discussed below in conjunction with some graphical images of larger protein models. From the data presented in Table 1, it seems safe to use a residue ASA threshold of 20 Å² coupled with a bead ASA re-check threshold of 50%, effectively obtaining a factor of ~ 10 in the reduction of the number of frictional points with respect to the starting atomic structures.

Table 1 Comparison between experimental and calculated $D_{l(20,w)}$ and $s_{(20,w)}$ values for US-SOMO bead models derived from test proteins

Protein	PDB (method)	Beads	$D_{l(20,w)}^{exp}$, F (% SE)	$D_{l(20,w)}^{calc}$, F (% difference; used beads)			$s_{(20,w)}^{exp}$, S			$s_{(20,w)}^{calc}$, S (% difference)		
				A10 R30 ^a	A20 R50 ^a	A40 R 60 ^a	A10 R30 ^a	A20 R50 ^a	A40 R60 ^a	A10 R30 ^a	A20 R50 ^a	A40 R60 ^a
BPTI	1BPI (X)	109	14.9 ± 0.52 (±3.5%)	14.5 (-2.7; 75)	14.6 (-2.0; 54)	14.7 (-1.3; 33)	NA	nd	nd	nd	nd	
	4PTI (X)	109		14.6 (-2.0; 75)	14.6 (-2.0; 56)	14.7 (-1.3; 33)		nd	nd	nd	nd	
	1PIT (N20)	108 ^b		14.6 (-2.0; 74 ^b)	14.6 (-2.0; 55 ^b)	14.8 (-2.0; 33 ^b)		nd	nd	nd	nd	
RNase	1RBX (X)	246	11.6 ± 0.31 (±2.7%)	11.8 (+1.7; 136)	11.8 (+1.7; 99)	11.9 (+2.6; 64)	2.00 ± 0.03 (±1.5%)	1.94 (-3.0)	1.95 (-2.5)	1.96 (-2.0)		
	8RAT (X)	246		11.8 (+1.7; 135)	11.8 (+1.7; 97)	11.9 (+2.6; 62)		1.94 (-3.0)	1.95 (-2.5)	1.96 (-2.0)		
	2AAS (N32)	245 ^b		11.8 (+1.7; 131 ^b)	11.8 (+1.7; 97 ^b)	11.8 (+1.7; 64 ^b)		1.94 (-3.0)	1.94 (-3.0)	1.95 (-2.5)		
Lysozyme	1AKI (X)	247	11.4 ± 0.42 (±3.7%)	11.8 (+3.5; 136)	11.9 (+4.4; 97)	11.9 (+4.4; 59)	1.88 ± 0.02 (±1.1%)	2.00 (+6.4)	2.00 (+6.4)	2.02 (+7.4)		
	1HEL (X)	246		11.7 (+2.6; 128)	11.8 (+3.5; 92)	11.8 (+3.5; 56)		1.98 (+5.3)	1.99 (+5.9)	2.00 (+6.4)		
	6LYS (X)	246		11.8 (+3.5; 130)	11.8 (+3.5; 93)	11.8 (+3.5; 61)		1.99 (+5.9)	1.99 (+5.9)	2.00 (+6.4)		
Myoglobin (CO)	1E8L (N50)	246 ^b		11.3 (-0.9; 139 ^b)	11.3 (-0.9; 104 ^b)	11.3 (-0.9; 62 ^b)		1.90 (+1.1)	1.91 (+1.6)	1.91 (+1.6)		
	IDWR (X)	295	10.7	10.8 (+0.9; 151)	10.8 (+0.9; 115)	10.8 (+0.9; 74)	2.18 (?)	1.94 (-11.0)	1.94 (-11.0)	1.95 (-10.6)		
	1SPE (X)	300		10.8 (+0.9; 138)	10.8 (+0.9; 103)	10.9 (+1.9; 70)		1.97 (-9.6)	1.97 (-9.6)	1.98 (-9.2)		
Myoglobin (apo)	1MYF (N12)	304 ^b		10.8 (+0.9; 147 ^b)	10.8 (+0.9; 112 ^b)	10.8 (+0.9; 76 ^b)		1.98 (-9.2)	1.99 (-8.7)	1.99 (-8.7)		
	1VXG (X)	303	10.8	10.7 (-0.9; 150)	10.8 (0.0; 110)	10.8 (0.0; 77)	2.05	1.95 (-4.9)	1.96 (-4.4)	1.97 (-3.9)		

X X-ray crystallography, N NMR spectroscopy, with number of models, NA not available, *nd* not done^a SOMO ASA parameters^b Average values

Table 2 Comparison between experimental and calculated $\tau_{c(20,w)}$ and $[\eta]$ values for US-SOMO bead models derived from test proteins

Protein	PDB (method)	Beads	$\tau_{c(20,w)}^{exp}$, ns (% SE)	$\tau_{c(20,w)}^{calc}$, ns (% difference; used beads)				$[\eta]^{exp}$, cm ³ /g (% SE)	$[\eta]^{calc}$, cm ³ /g (% difference)			
				A10 R30 ^a	A20 R50 ^a	A40 R60 ^a	A40 R60 ^a		A10 R30 ^a	A20 R50 ^a	A40 R60 ^a	A40 R60 ^a
BPTI	1BPI (X)	109	3.92 ± 0.34 (±8.7%)	4.48 (+14.3; 75)	4.36 (+11.2; 54)	4.22 (+7.7; 33)	NA	nd	nd	nd	nd	
	4PTI (X)	109		4.42 (+13.0; 75)	4.32 (+10.2; 56)	4.17 (+6.4; 33)		nd	nd	nd	nd	
	1PIT (N20)	108 ^b		4.46 ± 0.17 (+13.8)	4.33 ± 0.17 (+10.5)	4.16 ± 0.17 (+6.1)		nd	nd	nd	nd	
RNase	1RBX (X)	246	8.05 ± 0.51 (±6.3%)	8.06 (+0.1; 136)	7.86 (-2.4; 99)	7.57 (-6.0; 64)	3.30 ± 0.04 (±1.2%)	3.34 (+1.2)	3.24 (-1.8)	3.08 (-6.7)		
	8RAT (X)	246		8.07 (+0.2; 135)	7.84 (-2.6; 97)	7.46 (-7.3; 62)		3.36 (+1.8)	3.25 (-1.5)	3.05 (-7.6)		
Lysozyme	2AAS (N32)	245 ^b		8.07 ± 0.23 (+0.2)	7.91 ± 0.24 (-1.7)	7.61 ± 0.25 (-5.5)		3.36 (+1.8)	3.27 (-0.9)	3.10 (-6.0)		
	1AKI (X)	247	7.76 ± 0.47 (±6.1%)	7.92 (+2.1; 136)	7.63 (-1.7; 97)	7.31 (-5.8; 59)	2.99 ± 0.01 (±0.3%)	3.18 (+6.4)	3.05 (+2.0)	2.88 (-3.7)		
	1HEL (X)	246		7.99 (+3.0; 128)	7.76 (0.0; 92)	7.47 (-3.7; 56)		3.19 (+6.7)	3.08 (+3.0)	2.91 (-2.7)		
	6LYS (X)	246		8.01 (+3.2; 130)	7.75 (-0.1; 93)	7.47 (-3.7; 61)		3.20 (+7.0)	3.09 (+3.3)	2.94 (-1.7)		
Myoglobin (CO)	1E8L (N50)	246 ^b		9.18 ± 0.27 (+18.3)	8.96 ± 0.28 (+15.5)	8.66 ± 0.27 (+11.6)		3.66 (+22)	3.56 (+19)	3.40 (+14)		
	1DWR (X)	295	10 ± 1 (±10%)	10.3 (+3.0; 151)	10.1 (+1.0; 115)	9.74 (-2.6; 74)	NA	nd	nd	nd		
	1SPE (X)	300		10.4 (+4.0; 138)	10.2 (+2.0; 103)	9.73 (-2.7; 70)		nd	nd	nd		
Myoglobin (apo)	1MYF (N12)	304 ^b		10.5 ± 0.3 (+5.0)	10.1 ± 0.5 (+1.0)	9.84 ± 0.50 (-1.6)		nd	nd	nd		
	1VXG (X)	303	NA	10.4 (+4.0; 150)	10.2 (+2.0; 110)	9.82 (-1.8; 77)	NA	nd	nd	nd		

X X-ray crystallography, N NMR spectroscopy, with number of models, NA not available, *nd* not done

^a SOMO ASA parameters

^b Average values

Table 3 Further comparison between experimental and calculated $D_{l(20,w)}$ and $s_{(20,w)}$ values for US-SOMO bead models derived from test proteins

Protein	PDB (all X-ray)	Beads	$D_{l(20,w)}^{exp}$, F	$D_{l(20,w)}^{calc}$, F (% difference; used beads)		$s_{(20,w)}^{exp}$, S		$s_{(20,w)}^{calc}$, S (% difference)		
				A10 R30 ^a	A20 R50 ^a	A40 R60 ^a	A10 R30 ^a	A20 R50 ^a	A40 R60 ^a	
α -Lactalbumin hu	1A4 V	240	10.9	11.3 (+3.7; 128)	11.4 (+4.6; 96)	11.4 (+4.6; 61)	1.76	1.74 (-1.1)	1.74 (-1.1)	1.75 (-0.6)
Chymotrypsinogen	2CGA	466	9.5	9.59 (+0.9; 239)	9.61 (+1.2; 171)	9.68 (+1.9; 88)	2.56	2.73 (+6.6)	2.73 (+6.6)	2.75 (+7.4)
β -Lactoglobulin	1BEB	642	7.85 \pm 0.08	7.88 (+0.4; 325)	7.89 (+0.5; 219)	7.92 (+0.9; 139)	2.87 \pm 0.06	3.02 (+5.2)	3.03 (+5.6)	3.04 (+5.9)
Ovalbumin	1OVA	748	7.73 \pm 0.04	7.77 (+0.5; 343)	7.78 (+0.6; 235)	7.81 (+1.0; 155)	3.47 \pm 0.12	3.65 (+5.2)	3.65 (+5.2)	3.67 (+5.8)
Hemoglobin CO	1HCO	1142	6.9	6.95 (+0.7; 521)	6.98 (+1.2; 346)	7.01 (+1.6; 206)	4.5	4.72 (+4.9)	4.74 (+5.3)	4.76 (+5.8)
Hemoglobin oxo	1GZX	1142	7.21	6.95 (-3.6; 488)	6.96 (-3.5; 338)	6.99 (-3.1; 207)	NA	4.67 (nd)	4.67 (nd)	4.69 (nd)
Hemoglobin deoxi	1HGA	1138	NA	7.00 (nd; 482)	7.01 (nd; 330)	7.04 (nd; 205)	4.34	4.76 (+9.7)	4.77 (+9.9)	4.79 (+10)
Citrate synthase	1CTS	1678	5.8	5.83 (+0.5; 670)	5.84 (+0.7; 458)	5.86 (+1.0; 286)	6.1 \pm 0.1	6.15 (+0.8)	6.16 (+1.0)	6.18 (+1.3)
G3PD apo	2GDI	2636	5.0	5.05 (+1.0; 1062)	5.06 (+1.2; 668)	5.09 (+1.8; 363)	7.6	7.95 (+4.6)	7.97(+4.9)	8.01 (+5.4)
G3PD holo	1GDI	2656	5.3	5.08 (-4.2; 1044)	5.08 (-4.2; 631)	5.11 (-3.6; 351)	8.1	8.14 (+0.5)	8.15 (+0.6)	8.18 (+1.0)
LDH pig H NAD	5LDH	2597	5.06 \pm 0.15	5.11 (+1.0; 1249)	5.12 (+1.2; 785)	5.13 (+1.4; 457)	7.80 \pm 0.08	8.05 (+3.2)	8.06 (+3.3)	8.09 (+3.7)
LDH pig M NAD	9LDT	2688	5.25 \pm 0.16	5.22 (-0.6; 916)	5.23 (-0.4; 594)	5.25 (0.0; 352)	7.79 \pm 0.08	8.46 (+8.6)	8.47 (+8.7)	8.51 (+9.2)
LDH dogfish M	4GPD	2528	NA	5.00 (nd; 1148)	5.00 (nd; 700)	5.02 (nd; 395)	7.54	7.69 (+2.0)	7.69 (+2.0)	7.71 (+2.3)
Aldolase rabbit	1ADO	2759	NA	4.68 (nd; 1192)	4.69 (nd; 800)	4.70 (nd; 433)	7.85 \pm 0.05	8.02 (+2.2)	8.03 (+2.3)	8.06 (+2.7)
	1ZAH	2749		4.77 (nd; 1145)	4.77 (nd; 787)	4.79 (nd; 443)		8.13 (+3.6)	8.13 (+3.6)	8.16 (+3.9)
Nitrogenase MoFe	1MIN	3849	4.6 \pm 0.2	4.35 (-5.4; 1399)	4.35 (-5.4; 883)	4.36 (-5.2; 513)	10.3 \pm 0.1	11.4 (+11)	11.4 (+11)	11.5 (+12)
Catalase NADP	4BLC	3909	4.1	4.36 (+6.3; 1573)	4.36 (+6.3; 960)	4.37 (+6.6; 512)	11.35 \pm 0.1	11.6 (+2.2)	11.7 (+3.1)	11.7 (+3.1)

hu human, NA not available, nd not done

^a SOMO ASA parameters

Table 4 Further comparison between experimental and calculated $\tau_{c(20,w)}$ and $[\eta]$ values for US-SOMO bead models derived from test proteins

Protein	PDB (X-ray)	Beads	$\tau_{c(20,w)}^{exp}$, ns (% SE)	$\tau_{c(20,w)}^{calc}$, ns (% difference; used beads)			$[\eta]^{exp}$, cm ³ /g	$[\eta]^{calc}$, cm ³ /g (% difference)		
				A10 R30 ^a	A20 R50 ^a	A40 R60 ^a		A10 R30 ^a	A20 R50 ^a	A40 R60 ^a
α -Lactalbumin hu	1A4 V	240	10.3 ± 2.7 (26%)	9.09 (-11; 128)	8.92 (-13; 96)	8.60 (-17; 61)	NA	nd	nd	nd
Chymotrypsinogen	2CGA	466	NA	nd (nd; 239)	nd (nd; 171)	nd (nd; 88)	3.21	3.25 (+1.2)	3.15 (-1.9)	2.94 (-8.4)
β -Lactoglobulin	IBEB	642	23.2	25.9 (+12; 325)	25.1 (+8.2; 219)	24.5 (+5.6; 139)	NA	nd	nd	nd
Ovalbumin	IOVA	748	20.9	26.5 (+27; 343)	26.0 (+24; 235)	25.2 (+21; 155)	4.0 ± 0.5	3.55 (-11)	3.45 (-14)	3.33 (-17)
Hemoglobin hu oxi	IGZX	1142	35.4	36 (+1.7; 488)	35.1(-0.8; 338)	33.9 (-4.2; 207)	3.16 ± 0.2	3.28 (+3.8)	3.18 (+0.6)	3.05 (-3.5)
Hemoglobin ho oxi	IG0B	1132	40.5 ± 0.8 (2%)	35.4 (-13; 461)	34.7 (-14; 348)	33.4 (-18, 208)	NA	nd	nd	nd
Citrate synthase	ICTS	1678	NA	nd (nd; 670)	nd (nd; 458)	nd (nd; 286)	3.95	3.65 (-7.6)	3.54 (-10)	3.42 (-13)
G3PD	IGDI	2656	NA	nd (nd; 1044)	nd (nd; 631)	nd (nd; 351)	3.45	3.50 (+1.4)	3.40 (-1.4)	3.26 (-5.5)
LDH bovine H	4GPD	2528	NA	nd (nd; 1148)	nd (nd; 700)	nd (nd; 395)	3.8 ± 0.04	3.87 (+1.8)	3.75 (-1.3)	3.63 (-4.5)
Aldolase rabbit	IADO	2776	NA	nd (nd; 1192)	nd (nd; 800)	nd (nd; 433)	3.63 ± 0.12	4.20 (+16)	4.10 (+13)	3.94 (+8.5)
Catalase	IZAH	2749	NA	nd (nd; 1145)	nd (nd; 787)	nd (nd; 443)	3.9	4.02 (+11)	3.93 (+8.3)	3.78 (+4.1)
	4BLC	3909	NA	nd (nd; 1573)	nd (nd; 960)	nd (nd; 512)		3.51 (-10)	3.41 (-13)	3.26 (-16)

hu human, ho horse, NA not available, nd not done

^a SOMO ASA parameters

In Table 2, the results of the comparisons between experimental and computed $\tau_{c(20,w)}^h$ and $[\eta]$ values are presented for the same proteins of Table 1. The first thing to notice is the larger error present in the experimental $\tau_{c(20,w)}^h$ values, between 6 and 10%, with respect to the $D_{t(20,w)}^0$, $s_{(20,w)}^0$ and $[\eta]$ values. The second is that the ASA threshold values have a relevant effect on the calculated parameters: increasing the ASA threshold decreases the computed $\tau_{c(20,w)}^h$ and $[\eta]$ values, because fewer beads are included in the volume correction. Examining in detail the $\tau_{c(20,w)}^h$ values, it seems that the A20/R50 values produce the best match between experimental and computed data, well below the experimental errors. The exceptions are BPTI, for which it seems that the experimental data might underestimate the rotational tumbling (supported by the lack of differences between X-ray and NMR structures), and the NMR-derived lysozyme model(s). For the latter, this effect was again noticed and tentatively explained by Rai et al. (2005) as deriving from the opposite effect of the long, hydrophilic and flexible surface side-chains on translational and rotational diffusion. As for $[\eta]$, the two available datasets confirm that choosing a 20 Å² residue ASA threshold coupled with a 50% beads ASA re-check threshold produces an excellent match between experimental and computed data. Again, the NMR-derived model(s) of lysozyme are in poor agreement for the reasons given above.

Next, we examine a series of structures with increasing size. In Table 3, the $D_{t(20,w)}^0$ and $s_{(20,w)}^0$ values are reported, and it can be seen that most $D_{t(20,w)}^0$ values computed with A10/R30 are within 1% of the experimental values. The exceptions are catalase (+6.3%), α -lactalbumin (+3.7%), the oxi form of hemoglobin (-3.6%), the holo form of G3PD (-4.2%), and nitrogenase MoFe (-5.4%). Given the uncertainty associated with the experimental data, some of which are more than 60 years old (see ESM Table S3), we can consider this an excellent result. As for the computed $s_{(20,w)}^0$ values, most of them are within 5% of the experimental data. The deoxi form of hemoglobin and again nitrogenase MoFe are here the worst performers (+9.7 and +11%, respectively), while in the case of NAD-bound pig muscle lactate dehydrogenase (+8.6) the experimental value is suspiciously equal to that of the pig heart form. This is in contrast to the corresponding $D_{t(20,w)}^0$ values, where there is a net difference extremely well matched by the relative models. As done for Table 1, we have not investigated the effect of using experimental \bar{v}_2 values in the computations instead of calculated values, and similar effects could have affected the conversion to standard conditions of the original data. In this light, the agreement of the computed and experimental $s_{(20,w)}^0$ data in Table 3 can be considered satisfactory. Moreover, it can be seen that the A20/R50 combination works here as well as the original A10/R30.

$\tau_{c(20,w)}^h$ and $[\eta]$ data are also available for a restricted set of the same proteins, presented in Table 4. Using the A20/R50 combination, the $\tau_{c(20,w)}^h$ of all proteins, except ovalbumin, are within 15% of the experimental values, which is good considering the errors associated with the experimental data. More data are available for $[\eta]$, and here the results are mixed, with four protein models having computed values within 2% of the experimental data, and another four laying between 10 and 15% (still considering the A20/R50 framework). Again, some experimental data are suspicious, like the 4 cm³/g value for ovalbumin, but a full examination of these issues is beyond the scope of this paper.

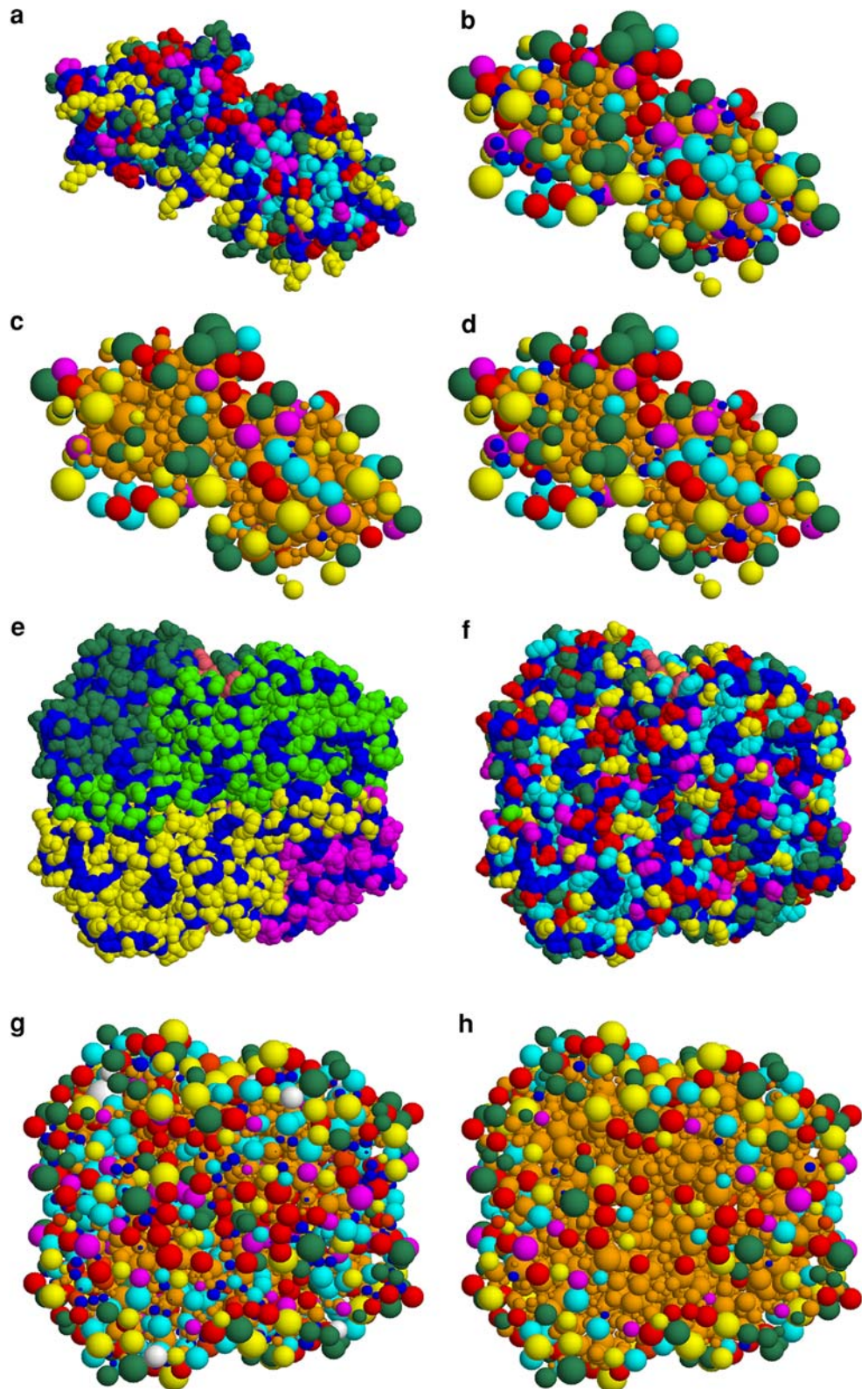
The results described in the previous section can be better interpreted by comparing the original atomic structures and the US-SOMO-generated models. In Fig. 2, panels a–d, the original β -lactoglobulin (1BEB.pdb) structure is shown (panel a) together with the three bead models generated by US-SOMO (panels b–d) whose parameters are reported in Tables 3 and 4. The color-coding, fully described in the Fig. 2 legend, refers to the characteristics of the residues' side chains and distinguishes also the buried beads (orange) from all the other beads. Note the increasing proportion of the buried beads in going from panel b (model generated with A10/R30) to panel d (A20/R50), to panel c (A40/R60). From the data in Table 3, the loss of prediction accuracy for $D_{t(20,w)}^0$ is only about 0.5% in going from model b (A10/R30) to model c (A40/R60), while the number of beads used to calculate the parameters drops from 325 to 139. Evidently, the translational motion of the protein is dominated by a restricted number of highly exposed frictional centers, clearly seen in Fig. 2 by comparing panels b and c. A similar situation is found with a larger protein made of four subunits, pig heart lactate dehydrogenase whose atomic structure (5LDH.pdb) is shown in panels e and f of Fig. 2. The two different representations were made to show both the subunits composition of LDH (panel e) and the US-SOMO residue coding (panel f). The two US-SOMO bead models in panels g and h were generated with A10/R30 and A40/R60, respectively. Again, the huge increase of “buried” beads between the two models corresponds only to a modest loss of accuracy of about 0.4% in $D_{t(20,w)}^0$ (Table 3). Overall, these data cast doubt on the necessity of an accurate modeling of protein surfaces for translational friction, which appears to be dominated by a subset of frictional centers. As for the rotational dynamics and intrinsic viscosity, the interpretation is complicated by the role of the excluded beads in the volume correction, and a more in depth analysis should await further studies.

Nevertheless, we can now examine in detail the performance of the US-SOMO-generated models in matching the NMR-derived $\tau_{c(20,w)}^h$ values for a set of relatively small

proteins, as presented by García de la Torre (2001). To ensure a proper comparison, the reduction to standard conditions of the experimental data was again critically assessed, as reported in Table S6. It was found that most likely the effect of D₂O on the solution viscosity was not accounted for, leading to values different from those reported in Table 2 of García de la Torre (2001). In Table 5, the corrected $\tau_{c(20,w)}^{\text{exp}}$ are thus reported, and compared with those computed by US-SOMO ($\tau_{c(20,w)}^{\text{SOMO}}$) using the A20/R50 combination that performed better in the previously described testing phase. In addition, we have also reported in Table 5 the HYDROPRO-generated values, $\tau_{c(20,w)}^{\text{HP}}$, presented in Table 2 of García de la Torre (2001), with their % differences from the new recalculated experimental values. Furthermore, we have expanded the set of structures to include additional NMR-derived structures, for which the $\tau_{c(20,w)}^{\text{SOMO}}$ averages were computed. This was facilitated by the fully automated processing implemented in US-SOMO, allowing the generation of most of the dataset presented in Table 5 in a mere 5 h of work, including the retrieval of the structures from the PDB and the coding of new residues (ligands and co-factors) not originally present in the *somo.residue* file. A few structures needed more work because they were incomplete, requiring additional operations. In particular, in the 1LKI leukemia inhibitor factor X-ray structure the first eight N-terminal residues were missing, which were taken from the 1A7M NMR structure (three different conformations were selected, and averages computed). Likewise, in the 1STN staphylococcal nuclease SN X-ray structure, the first five N-terminal and the last eight C-terminal residues were missing, and were taken from the 1JOR NMR structures, again generating three different models whose parameters were then averaged. For comparisons, the original incomplete structures were also processed and their computed values are presented in Table 5. We must also underscore that the data reported in Table 2 of García de la Torre (2001) were computed on a single structure for each protein examined, while the full datasets could be processed in our study.

Examining in detail the data presented in Table 5, we notice that our models in general match better the experimental data than those produced by HYDROPRO, the apparent exceptions being interleukin-1 β , lysozyme and eglin-c. However, these data must be also interpreted in the light of the evidence, documented in Table 5, that the average data from the NMR-derived structures in many cases perform worse than the corresponding X-ray structures, when both can be compared. An examination of the structures (not shown) reveals that in these cases the residues at the N- and C-terminal ends are very disordered, giving rise to quite different conformations. It is thus likely that this disorder reflects true conformational flexibility, which cannot be

Fig. 2 Atomic structures, shown in space filling mode, of β -lactoglobulin (1BEB.pdb, panel **a**) and pig heart lactate dehydrogenase with NAD bound (5LDH.pdb, panels **e** and **f**) with their corresponding US-SOMO-generated bead models (β -lactoglobulin, panels **b–d**; LDH, panels **g** and **h**). The models in panels **b** and **g** were generated with A10/R30, that in panel **d** with A20/R50, and those in panels **c** and **h** with A40/R60 (see text for details). The color-coding in panel **e** is blue for the main chain atoms, and green, greenblue, magenta and yellow for the side chains atoms of the four LDH subunits; the NAD moieties are pink. In all other panels, the color coding is: blue, main-chain; cyan, hydrophobic; magenta, non-polar; yellow, basic; green, acid; white, fused beads; orange, buried beads



properly modeled in the rigid-body approximation used by the current implementation of US-SOMO and by HYDROPRO. Obviously, it would be possible to choose a single conformation matching the experimental parameters, but

this would be clearly incorrect. In any case, overall the data presented in Table 5 confirm the reliability of the US-SOMO hydrodynamic modeling, while suggesting that other approaches, like Brownian dynamics (Ermak and

Table 5 Comparison between NMR-derived correlation times, $\tau_{c(20,w)}^{\text{exp}}$, with those computed by US-SOMO with A20/R50, $\tau_{c(20,w)}^{\text{SOMO}}$, and by HYDROPRO, $\tau_{c(20,w)}^{\text{HP}}$ (García de la Torre, 2001), for a set of test proteins

Protein	PDB (method)	No. of structures (features)	Beads total/used	$\tau_{c(20,w)}^{\text{exp}}$, ns	$\tau_{c(20,w)}^{\text{SOMO}}$, ns (% difference)	$\tau_{c(20,w)}^{\text{HP}}$, ns (% difference)
Savinase	1SVN (X)	1	503/156	12.19 ± 0.03	11.35 (−6.9)	13.67 (+12.1)
HIV-1 protease + DMP323	1BVG (N)	1 (average)	371/133	12.77 ± 0.08	12.90 (+1.0)	13.73 (+7.5)
	1BVE (N)	28	370/131		12.78 ± 0.27 (+0.1)	13.73 (+6.5)
Leukemia inhibition factor (LIF)	1LKI (X)	1 (res. missing)	335/134	14.45 ± 0.12	(11.88)	12.73 (−11.9)
	1LKI + 1A7 M	1 (completed)	352/151		13.80 (−4.5)	
	1A7 M (N)	20	350/134		13.20 ± 0.59 (−8.7)	
Interleukin-1 β	6IIB (N)	1 (average)	297/111	11.82 ± 0.07	11.52 (−2.5)	11.88 (+0.5)
	7IIB (N)	32	295/117		11.27 ± 0.27 (−4.7)	
Staphylococcal nuclease SN-OB	2SOB (N)	10 (fragment)	200/98	10.59 ± 0.04	9.04 ± 0.35 (−14.6)	
Staphylococcal nuclease SN	1STN (X)	1 (res. missing)	262/108	12.75 ± 0.04	(9.36)	10.30 (−19.2)
	1STN + 1JOR	1 (completed)	287/125		13.2 (+3.5)	
	1JOR (N)	30 (mutant)	289/130		14.21 ± 0.81 (+11.5)	
Lysozyme	1E8L (N)	50	247/104	8.26 ± 0.29	8.95 ± 0.33 (+8.4)	8.6 (+4.1)
Trp repressor	1WRT (N)	15	407/218	22.42 ± 0.47	23.95 ± 2.28 (+6.8)	26.29 (+17.3)
Barstar C40/82A	1BTA (N)	1 (average)	173/71	7.23	6.35 (−12.2)	6.2 (−14.2)
	1BTB (N)	30	174/70		6.03 ± 0.20 (−16.6)	
Cytochrome <i>b5</i>	1NX7 (N)	30 (res. 3–83)	157/71	5.64	5.84 ± 0.2 (+3.5)	
	1HKO (N)	42 (res. 1–104)	207/103		(11.6 ± 0.54)	6.6 (+17.0)
Ubiquitin	1UBQ (X)	1	146/68	5.14 ± 0.13	5.43 (+5.6)	5.87 (+ 14.2)
	1D3Z (N)	10	146/66		5.46 ± 0.20 (+6.2)	
Calbindin-D9k apo	1CLB (N)	33	143/69	5.06 ± 0.05	5.16 ± 0.25 (+2.0)	5.44 (+7.5)
Calbindin-D9k holo	2BCB (N)	32	143/66		4.87 ± 0.19 (−3.8)	5.16 (+2.0)
Eglin <i>c</i>	1EGL (N)	25	135/72	6.15 ± 0.07	5.52 ± 0.24 (−10.2)	6.28 (+2.1)
Xfin-zinc finger DBD	1ZNF (N)	37	52/29	1.99 ± 0.02	2.01 ± 0.22 (+1.0)	2.15 (+8.0)

McCammon 1978) or discrete molecular dynamics (Dokholyan et al. 1998) simulations, should be used to properly account for local flexibility effects.

In conclusion, we have shown that the bead-modeling scheme implemented in US-SOMO could be a very valuable tool in biomacromolecular hydrodynamic studies. The limitations present in the original SOMO (Rai et al. 2005) have been removed, and the program is now controlled from a GUI. When using pre-set default parameters, and non-standard residues or ligands are absent from a structure, the computations of all the hydrodynamic parameters are fast, reliable and very accurate for at least the translational diffusion parameters. When using the proper ASA/ASA re-check combination, the computational time required is minimal on standard personal computers, ranging from seconds for structures in the range 5–50 kDa to a few minutes for structures up to 250 kDa like catalase. While defining new residues still requires a detailed knowledge of their physical-chemical characteristics, these operations are also greatly aided by the GUI interface, effectively allowing the modeling of any kind of biomacromolecule from proteins to nucleic acids, carbohydrates, lipids and their

complexes. The *somo.atom* and *somo.residue* files already contain, respectively, 300 and 64 entries covering amino acids, nucleotides, sugars, co-factors like heme and NAD/NADPH, prosthetic groups like *N*-acetyl and phosphate. These files will be constantly updated, we hope also through the help of the analytical ultracentrifugation and other hydrodynamic techniques communities, for which this enhanced and powerful tool was mainly developed.

Acknowledgments We thank M. Nöllmann for providing his newAtoB code, and O. Byron for suggestions. The development of the UltraScan and US-SOMO is supported by the National Institute of Health Grant # RR022200 (to B. D.). M. R. gratefully acknowledges support from the Istituto Superiore della Sanità, program Italia-USA.

References

- Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The Protein Data Bank. *Nucleic Acids Res* 28:235–242. doi:10.1093/nar/28.1.235
- Bonneau R, Strauss CE, Rohl CA, Chivian D, Bradley P, Malmstrom L, Robertson T, Baker D (2002) De novo prediction of

- three-dimensional structures for major protein families. *J Mol Biol* 322:65–78. doi:[10.1016/S0022-2836\(02\)00698-8](https://doi.org/10.1016/S0022-2836(02)00698-8)
- Byron O (1997) Construction of hydrodynamic bead models from high-resolution X-ray crystallographic or nuclear magnetic resonance data. *Biophys J* 72:408–415. doi:[10.1016/S0006-3495\(97\)78681-8](https://doi.org/10.1016/S0006-3495(97)78681-8)
- Byron O (2000) Hydrodynamic bead modeling of biological macromolecules. *Methods Enzymol* 321:278–304. doi:[10.1016/S0076-6879\(00\)21199-3](https://doi.org/10.1016/S0076-6879(00)21199-3)
- Cantor CR, Schimmel PR (1980) *Biophysical chemistry. Part II: techniques for the study of biological structure and function*. W.H. Freeman, San Francisco
- Carrasco B, García de la Torre J (1999) Hydrodynamic properties of rigid particles: comparison of different modeling and computational procedures. *Biophys J* 76:3044–3057. doi:[10.1016/S0006-3495\(99\)77457-6](https://doi.org/10.1016/S0006-3495(99)77457-6)
- Chivian D, Kim DE, Malmstrom L, Schonbrun J, Rohl CA, Baker D (2005) Prediction of CASP6 structures using automated Robetta protocols. *Proteins* 61(Suppl 7):157–166. doi:[10.1002/prot.20733](https://doi.org/10.1002/prot.20733)
- Demeler B (2005) UltraScan. A comprehensive data analysis software package for analytical ultracentrifugation experiments. In: Scott DJ, Harding SE, Rowe AJ (eds) *Modern analytical ultracentrifugation: techniques and methods*. Royal Society of Chemistry, UK, pp 210–229
- Dokholyan NV, Buldyrev SV, Stanley HE, Shaknovich EI (1998) Discrete molecular dynamics studies of the folding of a protein-like model. *Fold Des* 3:577–587. doi:[10.1016/S1359-0278\(98\)00072-8](https://doi.org/10.1016/S1359-0278(98)00072-8)
- Ermak DL, McCammon JA (1978) Brownian dynamics with hydrodynamic interactions. *J Chem Phys* 69:1352–1360. doi:[10.1063/1.436761](https://doi.org/10.1063/1.436761)
- García de la Torre J (2001) Hydration from hydrodynamics. General considerations and applications of bead modelling to globular proteins. *Biophys Chem* 93:159–170. doi:[10.1016/S0301-4622\(01\)00218-6](https://doi.org/10.1016/S0301-4622(01)00218-6)
- García de la Torre J, Bloomfield VA (1981) Hydrodynamic properties of complex, rigid, biological macromolecules: theory and applications. *Q Rev Biophys* 14:81–139
- García de la Torre J, Huertas ML, Carrasco B (2000) Calculation of hydrodynamic properties of globular proteins from their atomic-level structure. *Biophys J* 78:719–730. doi:[10.1016/S0006-3495\(00\)76630-6](https://doi.org/10.1016/S0006-3495(00)76630-6)
- Halle B, Davidovic M (2003) Biomolecular hydration: from water dynamics to hydrodynamics. *Proc Natl Acad Sci USA* 100:12135–12140. doi:[10.1073/pnas.2033320100](https://doi.org/10.1073/pnas.2033320100)
- Harding SE, Longman E, Carrasco B, Ortega A, García de la Torre J (2004) Studying antibody conformations by ultracentrifugation and hydrodynamic modeling. *Methods Mol Biol* 248:93–113
- Jones TA, Zou JY, Cowan SW, Kjeldgaard M (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr A* 47:110–119. doi:[10.1107/S0108767390010224](https://doi.org/10.1107/S0108767390010224)
- Kuntz ID, Kauzmann W (1974) Hydration of proteins and polypeptides. *Adv Protein Chem* 28:239–345. doi:[10.1016/S0065-3233\(08\)60232-6](https://doi.org/10.1016/S0065-3233(08)60232-6)
- Lee B, Richards FM (1971) The interpretation of protein structures: estimation of static accessibility. *J Mol Biol* 55:379–400. doi:[10.1016/0022-2836\(71\)90324-X](https://doi.org/10.1016/0022-2836(71)90324-X)
- Rai N, Nöllmann M, Spotorno B, Tassara G, Byron O, Rocco M (2005) SOMO (SOLUTION MOdeler): differences between X-ray- and NMR-derived bead models suggest a role for side chain flexibility in protein hydrodynamics. *Structure* 13:723–734. doi:[10.1016/j.str.2005.02.012](https://doi.org/10.1016/j.str.2005.02.012)
- Rocco M, Rosano C, Weisel JW, Horita DA, Hantgan RR (2008) Integrin conformational regulation: uncoupling extension/tail separation from changes in the head region by a multi-resolution approach. *Structure* 16:954–964. doi:[10.1016/j.str.2008.02.019](https://doi.org/10.1016/j.str.2008.02.019)
- Spotorno B, Piccinini L, Tassara G, Ruggiero C, Nardini M, Molina F, Rocco M (1997) BEAMS (BEAdS Modelling System): a set of computer programs for the generation, the visualization and the computation of the hydrodynamic and conformational properties of bead models of proteins. *Eur Biophys J* 25:373–384 Erratum 26:417
- Tanford C (1961) *Physical chemistry of macromolecules*. Wiley, New York
- Tsai J, Taylor R, Chothia C, Gerstein M (1999) The packing density in proteins: standard radii and volumes. *J Mol Biol* 290:253–266. doi:[10.1006/jmbi.1999.2829](https://doi.org/10.1006/jmbi.1999.2829)
- Tsodikov OV, Record MT Jr, Sergeev YV (2002) Novel computer program for fast exact calculation of accessible and molecular surface areas and average surface curvature. *J Comput Chem* 23:600–609. doi:[10.1002/jcc.10061](https://doi.org/10.1002/jcc.10061)
- Vriend G (1990) WHAT IF: a molecular modeling and drug design program. *J Mol Graph* 8:52–56. doi:[10.1016/0263-7855\(90\)80070-V](https://doi.org/10.1016/0263-7855(90)80070-V)
- Zipper P, Durchschlag H (1997) Calculation of hydrodynamic parameters of proteins from crystallographic data using multi-body approaches. *Prog Colloid Polym Sci* 107:58–71. doi:[10.1007/BFb0118015](https://doi.org/10.1007/BFb0118015)
- Zipper P, Durchschlag H (1998) Recent advances in the calculation of hydrodynamic parameters from crystallographic data by multi-body approaches. *Biochem Soc Trans* 26:726–731