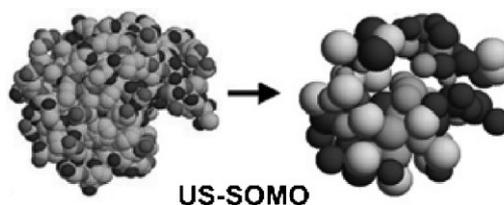


# Developments in the US-SOMO Bead Modeling Suite: New Features in the Direct Residue-to-Bead Method, Improved Grid Routines, and Influence of Accessible Surface Area Screening<sup>a</sup>

Emre Brookes, Borries Demeler, Mattia Rocco\*

The US-SOMO suite provides a flexible interface for accurately computing solution parameters from 3D structures of biomacromolecules through bead-modeling approaches. We present an extended analysis of the influence of accessible surface area screening, overlap reduction routines, and approximations for non-coded residues and missing atoms on the computed parameters for models built by the residue-to-bead direct correspondence and the cubic grid methods. Importantly, by taking the theoretical hydration into account at the atomic level, the performance of the grid-type models becomes comparable or exceeds that of the corresponding hydrated residue-to-bead models.



## Introduction

Bead modeling in various flavors is the most widely used methodology to compute, starting from the three-dimensional (3D) structure of macromolecules, their infinite-dilution hydrodynamic parameters, such as the translational diffusion coefficient  $D_t^0$ , the sedimentation coefficient  $s^0$ , the rotational relaxation time  $\tau^0$ , and the intrinsic viscosity  $[\eta]$ .<sup>[1]</sup> It is based on the ability to realistically calculate the frictional forces and torques experienced by a collection of rigidly connected, arbitrarily

placed spheres theoretically immersed in a solvent; importantly, the spheres can be of different radii provided that they do not overlap.<sup>[1–4]</sup> The methods employed to represent a macromolecule with beads range from shell-modeling (covering the surface with closely touching small beads, computing their properties, repeating the operation while decreasing the beads size at each iteration, and extrapolating to zero-bead size; see ref.<sup>[5]</sup>), to direct correspondence methods (representing selected parts of the macromolecule, such as whole residues or main-chain/side-chain segments, each with a bead of appropriate volume and position; see ref.<sup>[6]</sup>), and to grid modeling schemes (applying a cubic grid of defined size, and representing all atoms within each cubelet with a suitably sized bead; see refs.<sup>[7,8]</sup>). While the shell-modeling procedure has been available for some time as a self-standing, freely available program, HYDROPRO,<sup>[5]</sup> the direct correspondence and grid methods have been only recently implemented within one of the most widely-used, public domain analytical ultracentrifugation data analysis programs, UltraScan,<sup>[9]</sup> giving rise to the UltraScan-Solution MOdeler (US-SOMO) suite.<sup>[10]</sup> In the process, the direct correspondence method was further extensively tested with excel-

M. Rocco

Biopolimeri e Proteomica, Istituto Nazionale per la Ricerca sul Cancro (IST), IST c/o CBA, Largo R. Benzi 10, I-16132 Genova, Italy  
Fax: +39 0105737 325; E-mail: mattia.rocco@istge.it

E. Brookes, B. Demeler

Department of Biochemistry, The University of Texas Health Science Center at San Antonio, San Antonio, Texas, TX 78229, USA

<sup>a</sup> Supporting information for this article is available at the bottom of the article's abstract page, which can be accessed from the journal's homepage at <http://www.mbs-journal.de>, or from the author.

lent results,<sup>[10]</sup> but some issues, like the comparative performance of alternative bead overlap reduction methods, or the treatment of improperly recognized residues, were not fully addressed. Furthermore, the grid method was present in a basic form only, and performed poorly in comparison with the direct correspondence method. In this paper, we describe and analyze the most recent implementations in the US-SOMO suite. These include an extended analysis of the influence of accessible surface area (ASA) screening, overlap reduction routines, and approximations for non-coded residues and missing atoms within coded residues on the computed parameters for models built by the direct correspondence and cubic grid methods. In anticipation of the addition of a small-angle X-ray scattering (SAXS) simulation module (under development), we have introduced the assignment of the theoretically-bound water of hydration, a distinctive feature of our direct correspondence method,<sup>[6]</sup> at the atom level instead of at the residue level. This has allowed the additional possibility of directly hydrating the beads in the cubic grid approach, which was further developed with the inclusion of an ASA screening and outward translation (OT; see refs.<sup>[6,10]</sup>) of the exposed beads during overlap removal, resulting in a performance comparable to, and sometimes even better, than that of the direct correspondence method. A general restyling of the GUI, making the program even easier to use, and a number of other modifications and improvements, are also described.

## Experimental Part

### Technical Details

US-SOMO is written in C++ and linked against the UltraScan<sup>[9]</sup> 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 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, and were critically evaluated and reduced to standard conditions (water at 20 °C) previously.<sup>[10]</sup>

### Protein Structures

The high-resolution structures of the test proteins were taken from the Protein Data Bank (<http://www.rcsb.org/pdb/home/home.do>). Most were completed/adjusted previously,<sup>[10]</sup> with the exception

of human  $\alpha$ -lactalbumin, to which a biantennary Nag<sub>4</sub>Man<sub>3</sub>Gal<sub>2</sub> carbohydrate chain was added in this work, linked to Asn45<sup>[11]</sup> and positioned such that its frictional contribution is minimized.

## US-SOMO Implementation – Improved GUI Layout and New Functionalities

US-SOMO is controlled by a GUI interface that has been redesigned to better accommodate the new capabilities and to improve its ease of use. In particular, important changes were made to the grid (AtoB) method. First, the theoretical hydration can now be assigned at the atomic level instead of at the residue level, by selecting one or more atoms within a residue using the residue editor (see the electronic Support Information for more details). Second, we have introduced an ASA screening of the initial beads placed in the cubelets, with the possibility of performing an outward translation when removing the overlaps between solvent-exposed beads. For a more detailed description of the new US-SOMO features, including snapshots of the new main panel, and of some of the other panels, see the electronic Supporting Information (Figure S1-S13).

## Results and Discussion

### Comparison Between Different Overlap Removal Procedures

When bead models are made with beads of different size, overlaps between them should not be present; otherwise, the hydrodynamic interaction tensor employed in the computation is no longer valid.<sup>[2,3]</sup> In the original SoMo method,<sup>[6]</sup> implemented in the first installation of US-SOMO,<sup>[10]</sup> only a hierarchical overlap removal procedure was fully tested. The hierarchical approach consists of calculating the amount of the overlap between bead couples, ranking them in decreasing overlap amount order, and sequentially removing the overlaps by proportional reduction of the beads' radii. The program also included a synchronous overlap removal option, but it was not fully tested. In that routine all radii of overlapping beads are simultaneously reduced by a fixed step (% of their radius), then the ensemble is re-screened for remaining overlaps, and the procedure is repeated until no overlaps remain. In both overlap removal methods, an optional OT procedure for the surface exposed beads can move the beads' centers outwardly, by an amount equal to the radial reduction, along a line connecting them with the center of gravity of the ensemble, effectively preserving the model's original surface envelope. We have now compared the performance of the hierarchical and synchronous overlap removal routines, computing the hydrodynamic parameters for a series of proteins ranging from  $\approx 13\,700$  to  $\approx 236\,000$  Da in molecular mass. The tests were all performed with the default ASA thresholds, 20 Å<sup>2</sup> for the residues' ASA screening and 50% for the beads' ASA re-check (A20 R50). As can be

**Table 1.** Comparison between experimental (Exp.) and computed (Comp.) values of  $D_{t(20,w)}^0$ ,  $\tau_{h(20,w)}^0$  and  $[\eta]$  for several test proteins using the two different overlap removal procedures (HI, hierarchical; SY, synchronous) within the direct correspondence SoMo bead modeling method (ASA cutoff  $20 \text{ \AA}^2$ ; peptide bond rule on; fusion threshold 70%; outward translation on; ASA re-check threshold 50%).

Protein	PDB	MW	$D_{t(20,w)}^0$ , (% diff. from exp.)			$\tau_{h(20,w)}^0$ , (% diff. from exp.)			$[\eta]$ , (% diff. from exp.)		
			Exp.	Comp., HI	Comp., SY	Exp.	Comp., HI	Comp., SY	Exp.	Comp., HI	Comp., SY
RNase A	8RAT	13 682	11.6 ± 0.3	11.8 (+1.7)	11.8 (+1.7)	8.05 ± 0.51	7.79 (−3.2)	7.74 (−3.9)	3.30 ± 0.04	3.24 (−1.82)	3.22 (−2.42)
α-Lactalbumin	1A4V <sup>a)</sup>	15 793	10.9	10.9 (0)	10.9 (0)	10.3 ± 2.7	10.25 (−0.5)	10.14 (−1.6)	na <sup>b)</sup>	3.63 (nd <sup>c)</sup> )	3.59 (nd)
Myoglobin (CO)	1DWR	17 521	10.7	10.9 (+1.9)	10.9 (+1.9)	10 ± 1	10.0 (0)	9.88 (−1.2)	na	3.24 (nd)	3.20 (nd)
Chymotrypsinogen A	2CGA	25 666	9.5	9.64 (+1.5)	9.64 (+1.5)	na	13.8 (nd)	13.7 (nd)	3.21	3.13 (−2.5)	3.10 (−3.4)
β-Lactoglobulin	1BEB	36 608	7.85 ± 0.08	7.92 (+0.9)	7.92 (+0.9)	23.2	25.1 (+8.2)	24.8 (+7.1)	na	3.92 (nd)	3.88 (nd)
Ovalbumin	1OVA	43 157	7.73 ± 0.04	7.78 (+0.7)	7.78 (+0.7)	20.9	26.2 (+25.4)	25.9 (+24.1)	4.0 ± 0.5	3.48 (−13.0)	3.44 (−14.0)
Hemoglobin CO	1HCO	64 557	6.9	6.98 (+1.2)	6.98 (+1.2)	na	35.6 (nd)	35.4 (nd)	na	3.28 (nd)	3.25 (nd)
Hemoglobin oxi	1GZX	64 573	7.21	6.99 (−3.1)	6.99 (−3.1)	35.4	34.9 (−1.1)	34.7 (−1.8)	3.16 ± 0.2	3.18 (+0.6)	3.16 (0.0)
Citrate synthase	1CTS	97 838	5.8	5.86 (+1.0)	5.86 (+1.0)	na	59.5 (nd)	59.1 (nd)	3.95	3.51 (−11.1)	3.49 (−11.7)
G3PD apo	2GD1	143 540	5.0	5.08 (+1.6)	5.08 (+1.6)	na	90.2 (nd)	89.6 (nd)	na	3.59 (nd)	3.56 (nd)
G3PD holo	1GD1	146 431	5.3	5.10 (−3.8)	5.10 (−3.8)	na	88.9 (nd)	88.6 (nd)	3.45	3.48 (+0.9)	3.46 (+0.3)
Lactate dehydrogenase	5LDH	148 636	5.06 ± 0.15	5.14 (+1.6)	5.14 (+1.6)	na	89.1 (nd)	88.3 (nd)	na	3.46 (nd)	3.43 (nd)
Aldolase	1ADO	157 122	na	4.71 (nd)	4.71 (nd)	na	115.8 (nd)	115.1 (nd)	3.63	4.07 (+12.1)	4.04 (+11.3)
Catalase	4BLC	235 762	4.1	4.37 (+6.6)	4.37 (+6.6)	na	140.7 (nd)	139.8 (nd)	3.9	3.40 (−12.8)	3.37 (−13.6)

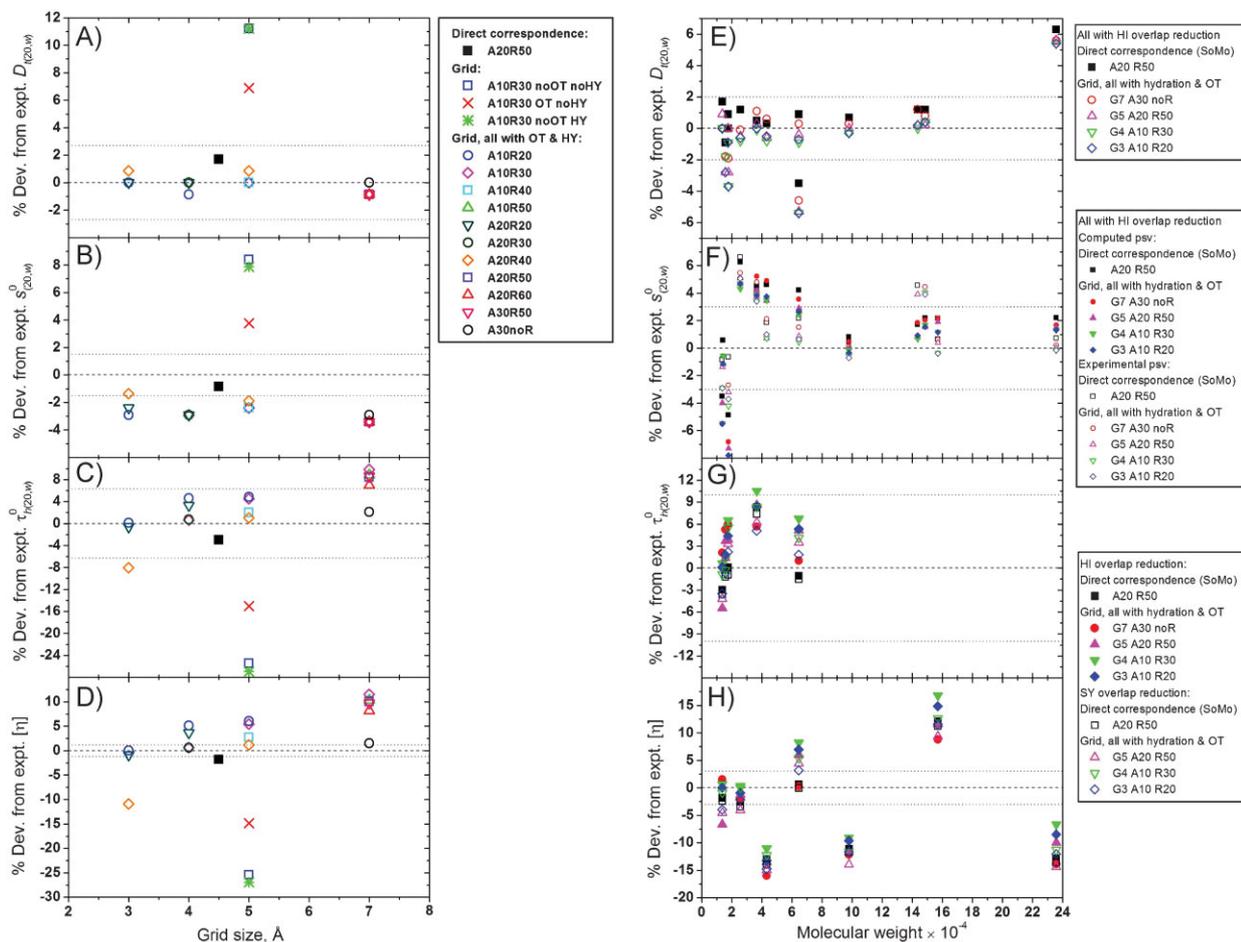
<sup>a)</sup>Modified with the addition of a Nag<sub>4</sub>Man<sub>3</sub>Gal<sub>2</sub> biantennary carbohydrate chain attached to Asn45.<sup>[11]</sup> <sup>b)</sup>na: not available; <sup>c)</sup>nd: not done.

seen in Table 1, the two methods perform identically when compared with experimental  $D_{t(20,w)}^0$  data, reproducing most values within 1–2%. Given this level of overall performance, it appears that flaws in the experimental data or their analysis are likely responsible for the few instances where predicted hydrodynamic parameters compared less favorably to experimentally measured values. As for  $\tau_{h(20,w)}^0$  and  $[\eta]$ , the synchronous method produces slightly higher and lower values, respectively, than the hierarchical procedure. With one exception, having a suspiciously low value, all available  $\tau_{h(20,w)}^0$  values are reproduced within 10%, and several within 3%, well within usual experimental error. For  $[\eta]$ , the matches are more erratic, again likely pointing out potential problems at the experimental data level, although a role for local flexibility or slightly different overall conformations between solution and crystals cannot be ruled out. On visual inspection, few differences are noticeable between the two methods, as evidenced by the two models derived from the RNase A 8RAT crystal structure presented in Figure S14 in the electronic Supporting Information (panel A, hierarchical procedure; panel B, synchronous procedure). A more careful examination reveals that the synchronous procedure produces more evenly-sized beads, especially in the buried category (orange color; these beads are not used in the hydrodynamic computations). In summary, the performance of synchro-

nous and hierarchical procedures appear to be equivalent, but since the latter is faster, it should be preferentially used when employing the direct correspondence method to generate bead models.

### Improvements in the Grid (AtoB) Procedure

The first tests of the new features that we have introduced in the grid (AtoB) method were conducted using the 8RAT RNase A structure, for which the parameters computed from the direct correspondence (SoMo) bead models were all in excellent agreement with the experimental data (Table 1). In Figure 1, panels A–D, we have plotted the % deviation of the computed parameters for various grid settings from the experimental  $D_{t(20,w)}^0$  (panel A),  $s_{(20,w)}^0$  (panel B),  $\tau_{h(20,w)}^0$  (panel C), and  $[\eta]$  (panel D); the dash horizontal lines indicate perfect matching, while the two dot lines report the % standard deviation (SD) of the experimental data. Values computed for a SoMo model with A20 R50 are also reported for comparison (solid black squares). To improve the matching, an experimental partial specific volume  $\bar{v}$  ( $0.702 \text{ ml} \cdot \text{g}^{-1}$ ,<sup>[12]</sup> after correction<sup>[13]</sup> to 20 °C) was used in the computations of  $s_{(20,w)}^0$  (see below). Since the resolution of the SoMo models is between 4–5 Å, a grid size of 5 Å was initially chosen for the tests, with a 10 Å<sup>2</sup> residues' ASA threshold and a 30% beads' ASA re-check



**Figure 1. Panels A-D:** performance of the improved grid method with hierarchical overlap reduction, using the RNase A 8RAT X-ray structure, as a function of grid size for different ASA residue screening (A) and bead re-check (R) cutoffs, and, at 5 Å grid-size only, of including/excluding the outward translation (OT) and the atomic hydration (HY; see legend for symbols details). The data are reported as the % difference between the computed and experimental  $D_{t(20,w)}^0$  (panel A),  $s_{(20,w)}^0$  (panel B),  $\tau_{h(20,w)}^0$  (panel C), and  $[\eta]$  (panel D) values. The dash horizontal lines indicate perfect match with the experimental values; the dotted lines are the experimental % SD. The filled squares are the values computed for the SoMo bead models with A20 R50 (see side legends for symbols details). **Panels E-H:** performance of the improved grid method for a wide range of protein sizes (see Table 1), as a function of grid size (G) for different ASA residue screening (A) and bead re-check (R) cutoffs, all with atomic hydration and outward translation (OT). The data are reported as the % difference between the computed and experimental  $D_{t(20,w)}^0$  (panel E),  $s_{(20,w)}^0$  (panel F),  $\tau_{h(20,w)}^0$  (panel G), and  $[\eta]$  (panel H) values. The dash horizontal lines indicate perfect match with the experimental values; the dot lines are the expected experimental % SD. In panel E, the filled squares are the values computed for the SoMo bead models with A20 R50, and the open symbols are for grid models with hierarchical (HI) overlap reduction. In panel F, the filled symbols are for data produced using the computed  $\bar{v}$ , while the open symbols are for data produced with experimental  $\bar{v}$  values, all with hierarchical overlap removal only. In panels G and H, the filled and open symbols are for models with hierarchical (HI) and synchronous (SY) overlap reduction, respectively. See side legends for symbols details.

(A10 R30) cutoffs, and hierarchical overlap removal. Blue squares, red crosses, and green stars indicate models without both OT and atomic hydration (HY), with OT but not HY, and without OT but with HY, respectively. As can be seen in Figure 1, panels A-D, in the absence of both OT and HY, or with HY only, the models perform similarly and quite poorly, resulting in deviations of  $\approx +11\%$  for  $D_{t(20,w)}^0$ ,  $\approx +8\%$  for  $s_{(20,w)}^0$ ,  $\approx -25\%$  for  $\tau_{h(20,w)}^0$ , and  $\approx -27\%$  for  $[\eta]$ . Introducing the OT, without the HY, significantly improves the matching, nearly halving the % deviation in all cases.

However, a dramatic improvement is achieved when both the OT and HY are employed simultaneously (magenta diamonds), obtaining a perfect match for  $D_{t(20,w)}^0$ , and bringing the  $s_{(20,w)}^0$ ,  $\tau_{h(20,w)}^0$ , and  $[\eta]$  values within 3%, 4% and 5%, respectively (for  $s_{(20,w)}^0$ , we suggest that the  $\bar{v}$  uncertainty has the greatest effect on the deviation from the experimental value). Next, we investigated the effect of changing the screening and re-checking cutoffs, using A10 R20 (blue circles), A10 R40 (cyan squares), and A20 R40 (orange diamonds). Almost no influence was found for

$D_{t(20,w)}^0$  and  $s_{(20,w)}^0$ , while there was a moderate effect for the other two properties (Figure 1, panels A-D). Importantly, the excellent performance for  $D_{t(20,w)}^0$  and  $s_{(20,w)}^0$  was maintained across all the grid sizes investigated (3–7 Å), irrespective of the cutoffs employed (Figure 1, panels A and B; see side legend for the symbols' correspondence). Instead,  $\tau_{h(20,w)}^0$  and  $[\eta]$  (Figure 1, panels C and D) showed a noticeable effect, with some A and R combinations performing poorly at low grid values (A20 R40, orange diamonds), while others clearly improved when a higher grid value was chosen (A30 no R, black circles). We ascribe this subtle effect to the different volume corrections (see refs.<sup>[14,15]</sup>) employed, resulting from the ASA screening and re-check, rather than being linked to the total number of beads used in the hydrodynamic computations. Indeed, if this were the case, also the  $D_{t(20,w)}^0$  and  $s_{(20,w)}^0$  values should be affected, but we did not notice any significant changes. Further work is clearly needed to better define an effective volume correction to be used with this kind of intermediate-sized beads, different from either the very small beads used in shell-modeling approaches, and the low-resolution bulk models for which it was originally devised (see ref.<sup>[16]</sup> for a recent discussion on this subject).

The grid procedure was then further tested using the set of proteins of Table 1. In Figure 1, panels E-H, the % deviation from the experimental  $D_{t(20,w)}^0$  (panel E),  $s_{(20,w)}^0$  (panel F),  $\tau_{h(20,w)}^0$  (panel G), and  $[\eta]$  (panel H) are reported; again, the dash horizontal lines indicate perfect matching, while the two dot lines report a reasonable % SD expected for each kind of experimental data. Only the best values for the residues' ASA screening and for the beads' ASA re-check cutoffs deduced from the RNase A tests (Figure 1, panels A-D) were employed. For the  $D_{t(20,w)}^0$  and  $s_{(20,w)}^0$  tests, only the hierarchical overlap removal procedure is reported, while for  $\tau_{h(20,w)}^0$  and  $[\eta]$  also the results of the synchronous option are shown. In addition, for  $s_{(20,w)}^0$  we have also compared the values obtained with the computed and with the experimental  $\bar{v}$  values, taken from literature<sup>[12,17–19]</sup> and corrected, when necessary, to 20 °C (8RAT, 0.702 ml · g<sup>-1</sup>; 1VXG, 0.743 ml · g<sup>-1</sup>; 2CGA, 0.731 ml · g<sup>-1</sup>; 1BEB, 0.749 ml · g<sup>-1</sup>; 1OVA, 0.744 ml · g<sup>-1</sup>; 1HCO, 0.752 ml · g<sup>-1</sup>; 1CTS, 0.740 ml · g<sup>-1</sup>; 5LDH, 0.740 ml · g<sup>-1</sup>; 1ADO, 0.740 ml · g<sup>-1</sup>; 4BLC, 0.730 ml · g<sup>-1</sup>). In all panels, values obtained by the direct correspondence SoMo method are also reported (filled and empty black squares). To begin with, the excellent agreement between experimental and computed  $D_{t(20,w)}^0$  values (within 2%) across the whole range investigated (Figure 1, panel E) is noteworthy, with two exceptions presumably resulting from questionable experimental values. Furthermore, the grid procedure performs as well or even better than the SoMo method for all grid sizes investigated. A similar situation is found for the  $s_{(20,w)}^0$  values, except they show a greater variation at low protein size, while better matching at larger sizes (Figure 1, panel F).

It is interesting to note that, in general, the agreement is better when experimental values for  $\bar{v}$  are used in place of computed values, with two evident exceptions, G3PD and lactate hydrogenase. Clearly, more work is needed toward a better computation of  $\bar{v}$ , possibly taking into account the actual experimental conditions. Taken together, the data in Figure 1, panels E-F, seem to suggest that  $D_{t(20,w)}^0$  should be preferentially employed rather than  $s_{(20,w)}^0$  when their values are used for macromolecular modeling applications.

For  $\tau_{h(20,w)}^0$ , fewer experimental values are available, none above 70 kDa molar mass. In this restricted range, the overall performance is reasonably good, with most computed values within 6% of the experimental data (Figure 1, panel G). In general, the grid procedure appears to work better with the synchronous overlap reduction (SY) option, and the dependence on the ASA cutoff values observed with RNase A (Figure 1, panel C) is confirmed. As for  $[\eta]$ , it is difficult to assess the reliability of the experimental data, which show a great variation (Figure 1, panel H). Given that, the only useful information that can be gathered is a confirmation of the dependence of the calculated values on the ASA cutoffs.

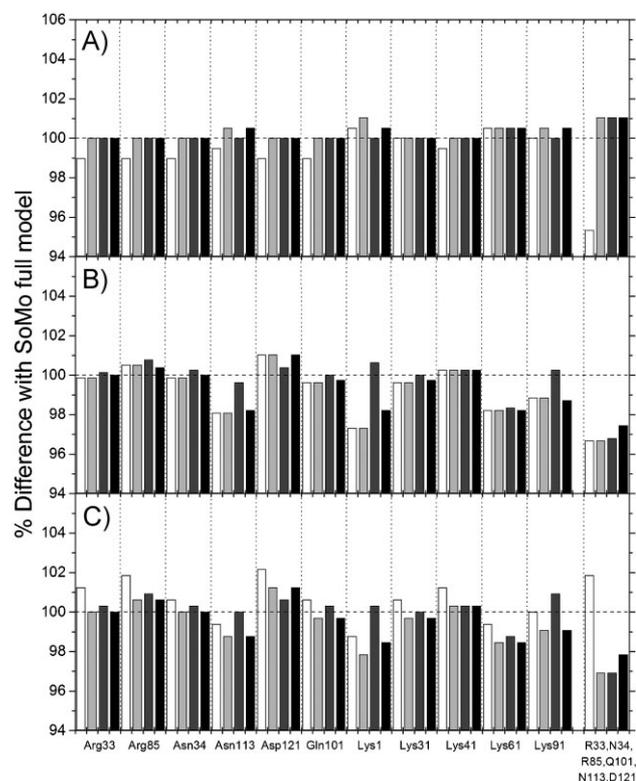
To conclude this section, the grid procedure with ASA screening, OT and direct hydration is now performing as well, sometimes even better, than the SoMo direct correspondence method. Based on the  $\tau_{h(20,w)}^0$  tests, the synchronous overlap removal procedure appears to produce somewhat better results. Indeed, the bigger influence of the overlap removal procedure on the overall quality of the grid models is evident in Figure S14, panels C and D, further suggesting that the synchronous option should be preferentially employed for these kinds of models, notwithstanding the longer computational times. Finally, it is presently difficult to indicate best values for the ASA screening and re-check cutoffs, as they also depend on the grid size employed. The values reported in Figure 1, panels E-H, were empirically found to work best for the corresponding grid sizes, and US-SOMO employs the A10/R30 combination with the 5 Å grid size as the default settings. Further work is in progress to better clarify this issue.

### Testing the Approximate Methods for Missing Atoms within Coded Residues or for Non-Coded Residues

Protein structures present in the databanks are sometimes incomplete, often because local flexibility prevented the crystallographers from correctly placing side chains or whole residues. In other occasions, residues were removed or mutated in proteins engineered for crystallography, while the native species was used for gathering solution data. Clearly, all residues should be present in their complete form in an atomic structure if its solution properties are to be correctly computed (apart from flexibility effects not dealt with in this study). While

missing side-chains can be manually built or automatically added by dedicated programs (e.g. WHATIF<sup>[20]</sup>, <http://swift.cmbi.ru.nl/servers/html/index.html>), a rapid way of computing the hydrodynamics with comparable accuracy could still be useful. Since US-SOMO will not allow processing of incomplete residues in normal operational mode, the simplest solution is to altogether skip the residue with missing atoms, with the possibility of entering the correct global molecular weight and  $\bar{v}$  values affecting the sedimentation coefficient and intrinsic viscosity computations. A more refined procedure involves pretending that the residue is complete, as coded in the US-SOMO 'somo.residue' table, and approximately placing the beads. Obviously, if the atoms used in defining the beads' positions are present, there is no effect of the missing portions. If the position-defining atoms are instead missing, the program can use the remaining atoms to place the bead. For amino acids, where usually two beads are used (one for the main-chain and one for the side-chain regions), a single atom in each region is then sufficient to place a bead. If one of the two regions is missing, a single bead is then used, placed at the center of mass of all remaining atoms. Both options (skip or approximate the coded residue with missing atoms) are now fully implemented in US-SOMO, and are selectable from the PDB parsing options menu (see the electronic Supporting Information).

To test the two alternative options, we have manually removed portions of several surface-exposed side-chains from the RNase A 8RAT PDB file (indicated in Figure S14, panel E). The results are reported in Figure 2, where the differences with the values for  $s_{(20,w)}^0$  (panel A),  $\tau_{h(20,w)}^0$  (panel B), and  $[\eta]$  (panel C) computed for the intact 8RAT structure are presented as white bars (skip the residue without entering the correct global molecular weight and  $\bar{v}$ ), light gray bars (skip the residue and enter the correct global molecular weight and  $\bar{v}$ ), and dark gray and black bars (use the approximate method). The side-chains of the selected residues were removed completely (black bars) or leaving the CB atom (white, light gray and dark gray bars). Model generation was done with the SoMo method using the default parameters (A20 R50, OT, hierarchical overlap removal). Besides single-residue tests, we have also performed the computations for a model where the missing atoms affected 6 residues at the same time (last group). As can be seen in Figure 2, simply skipping the truncated residues, without correcting for molecular weight and  $\bar{v}$ , is reasonably safe for  $s_{(20,w)}^0$  (panel A, white bars) and if just a single residue side-chain has missing atoms. In fact, while most of the single mutations have computed  $s_{(20,w)}^0$  within 1% of the complete model, the 6-residue group showed a  $\approx -5\%$  change. Interestingly, skipping the residues and entering the correct global molecular weight and  $\bar{v}$  produces  $s_{(20,w)}^0$  values within 1% in all cases (light gray bars), a performance nearly identical with that of the more



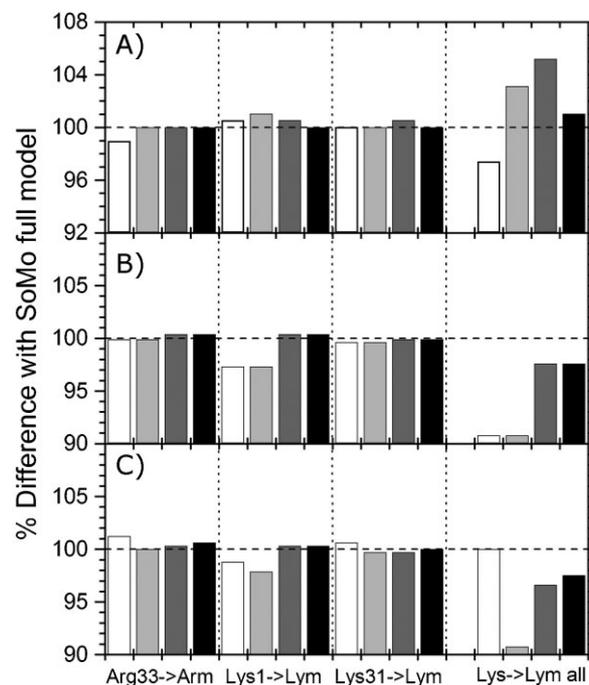
**Figure 2.** Test of the approximate methods for dealing with missing atoms within US-SOMO-coded residues, listed as the % difference compared to the parameters computed from the SoMo model of the intact protein (horizontal dash lines). The residues in the RNase 8RAT crystal structure whose side chains were either truncated at the CB level (white, light grey and dark grey bars) or completely removed (black bars) are indicated at the bottom (the vertical dotted lines help in separating the groups). In the last group, several residues (single letter code) were affected at the same time in the model. White bars, simple incomplete residue skipping, without any correction to the global molecular weight and  $\bar{v}$  values; light grey bars, residue skipping with corrected global molecular weight and  $\bar{v}$  values; dark grey and black bars, SoMo approximate method. Panel A,  $s_{(20,w)}^0$ ; panel B,  $\tau_{h(20,w)}^0$ ; panel C,  $[\eta]$ .

efficient approximate method (dark gray bars), which worked well also when entire side-chains were missing (black bars). However, the performance of the skip methods is not so effective when computing  $\tau_{h(20,w)}^0$  and  $[\eta]$  (Figure 2, panels B and C, respectively; white and light gray bars). In these cases, the approximate method instead works quite well for single-residue mutations and when the side chains are truncated at the CB atom level (dark gray bars), but when the entire side-chain is removed, the results are less satisfactory (black bars). The simultaneous absence of several side-chains (last group in Figure 2, panels B and C) noticeably degrades the performances of all methods for the  $\tau_{h(20,w)}^0$  and  $[\eta]$  computations. To summarize this section, while the computation of translational frictional properties

( $D_{t(20,w)}^0$  and  $s_{(20,w)}^0$ ) can be performed with reasonable precision when the approximate method is used for coded residues with missing atoms, this does not hold quite as well for  $\tau_{h(20,w)}^0$  and  $[\eta]$ , especially if several incomplete residues are present. In any case, structures should be completed whenever possible for best hydrodynamic parameters computations.

Another complication can arise when residues present in the structure have not yet been included in the “somo.residue” file (called “non-coded” residues in US-SOMO). While the residue editor included in US-SOMO allows for coding of any arbitrary residue, this operation is not always straightforward, requiring a detailed knowledge of several physico-chemical and structural parameters of the residue in question. Therefore, we have also explored the possibility of approximate non-coded residues using average parameters computed as mean values among the residues already included in the “somo.residue” table. These parameters can be edited and fine-tuned to the kind of non-coded residue present (e.g., amino acid, nucleotide, carbohydrate, etc.) from a dedicated menu (see the electronic Supporting Information). A single bead is used in the approximate method, and the bead positioning issue is taken care of by placing the bead in the center of mass of all atoms present. The peptide-bond rule (see electronic Supporting Information) is also disallowed in this case. This option, and the ability to instead skip non-coded residues, are now fully integrated in the current US-SOMO release, and are accessible from the PDB parsing options menu (see the electronic Supporting Information).

To test these methods, we have again manually edited the RNase A 8RAT PDB file changing the names of some exposed residues (Arg and Lys) to non-coded residues (“Arm” and “Lym”). As we have done for the missing atoms within coded residues case, the simplest way to deal with non-coded residues is to skip them, with or without entering global molecular weight and  $\bar{v}$  values. The alternative is the approximate method using average parameters described above. Models were then generated using either option, and their hydrodynamic parameters were compared with those of the original model. The results are shown in Figure 3, where the % difference with the original 8RAT structure is reported for  $s_{(20,w)}^0$  (panel A),  $\tau_{h(20,w)}^0$  (panel B), and  $[\eta]$  (panel C). The results from calculations where non-coded residues were skipped, with and without molecular weight and  $\bar{v}$  correction, are represented by the light gray and white bars, respectively. Results from the approximate method, with and without molecular weight and  $\bar{v}$  correction, are represented by the black and dark gray bars, respectively. Only the results of three single “mutations” (Arg33 → Arm33, Lys1 → Lym1, Lys33 → Lym33) are reported, plus the mutation of all the Lys → Lym residues in the same model (last group). As can be seen in Figure 3, for single mutations the approximate



**Figure 3.** Comparison between the skip and approximate methods for dealing with non-coded residues within US-SOMO and the parameters computed from the original protein SoMo model (horizontal dash lines), plotted as the % difference. The residues in the RNase 8RAT crystal structure that were renamed with a non-coded name are indicated at the bottom (the vertical dotted lines help in separating the groups). In the last group, all lysines were renamed at the same time in the model. White bars, simple non-coded residue skipping, without any correction of the global molecular weight and  $\bar{v}$  values; light grey bars, non-coded residue skipping with corrected global molecular weight and  $\bar{v}$  values; dark grey bars, SoMo approximate method for non-coded residues without any correction of the global molecular weight and  $\bar{v}$  values; black bars, SoMo approximate method for non-coded residues with corrected global molecular weight and  $\bar{v}$  values. Panel A,  $s_{(20,w)}^0$ ; panel B,  $\tau_{h(20,w)}^0$ ; panel C,  $[\eta]$ .

method worked very well in all cases, with the skip methods performing slightly worse. However, when all Lys were changed to “Lym” residues, only the approximate method with the global molecular weight and  $\bar{v}$  correction worked reasonably well (the perfect match of the  $[\eta]$  value for the skip method without the global molecular weight and  $\bar{v}$  correction is most likely just a coincidence due to internal compensations). Again, while the approximate method can be used when a quick estimate of the hydrodynamic parameters is needed, best results can be obtained only with proper coding of all residues.

## Conclusion

The development and comprehensive testing of US-SOMO as reported in this and our previous work<sup>[10]</sup> has now

resulted in a reliable and flexible program that can predict with high confidence the rigid-body hydrodynamic parameters of biomacromolecules, starting from their high-resolution structures. US-SOMO now offers the full choice between two alternative methods of building a bead model, the direct correspondence, residue-to-bead (SoMo), and the cubic grid approach (AtoB), with a similar level of precision. Most needs in rigid-body hydrodynamics can thus be covered by US-SOMO, from small structures to large complexes, with a tunable accuracy (grid size, number of beads per residue) that can be tailored to the computer power available and the size of the macromolecule. Furthermore, incomplete or non-coded residues can now be tolerated to a certain degree, providing users not able or willing to deal with the additional complications posed by such deficiencies with a rapid way, although potentially less accurate, to estimate the hydrodynamic parameters. The new SAXS simulator module, which is in an advanced phase of development, will also permit the evaluation of the properties of proposed solution structures against SAXS data, using the same kind of bead models. Batch mode operations, which can process sequential models in a NMR-type file, or independent model files, already permit the study of multiple conformations provided either from experimental data (e.g., NMR) or from artificially generated data (e.g. Monte Carlo, discrete molecular dynamics), offering a first approach to the study of truly flexible or partially disordered structures. While this can provide reasonable estimates of conformational parameters (i.e. the radius of gyration) or of translational frictional properties, a more accurate description of the full hydrodynamics of locally- or segmentally-flexible biomacromolecules (e.g. side-chain dynamics, multi-domain proteins, long DNA stretches, polysaccharides) calls for different approaches, such as Brownian dynamics (BD) simulations. In this respect, the direct correspondence models produced by US-SOMO are naturally suited to be used in current BD schemes, such as the publicly available SIMUFLEX suite.<sup>[21]</sup> While we are actively exploring the direct implementation of our own BD algorithm within US-SOMO, an interface linking our models to the SIMUFLEX suite will be provided in a future release, thus virtually covering all major aspects of biomacromolecular hydrodynamics.

**Acknowledgements:** The development of the UltraScan software is supported by the *National Institutes of Health* through grant RR022200 (to BD). We thank *Jeremy Mann* for technical support.

Received: December 24, 2009; Revised: February 18, 2010;  
Published online: May 17, 2010; DOI: 10.1002/mabi.200900474

**Keywords:** biopolymers; computer modeling; molecular dynamics; structure–property relationships; ultracentrifugation

- [1] O. Byron, *Methods Enzymol.* **2000**, *321*, 278.
- [2] J. García de la Torre, V. A. Bloomfield, *Q. Rev. Biophys.* **1981**, *14*, 81.
- [3] B. Spotorno, L. Piccinini, G. Tassara, C. Ruggiero, M. Nardini, F. Molina, M. Rocco, *Eur. Biophys. J.* **1997**, *25*, 373 (Erratum *Eur. Biophys. J.* **1998**, *26*, 417).
- [4] B. Carrasco, J. García de la Torre, *Biophys. J.* **1999**, *76*, 3044.
- [5] J. García de la Torre, M. L. Huertas, B. Carrasco, *Biophys. J.* **2000**, *78*, 719.
- [6] N. Rai, M. Nöllmann, B. Spotorno, G. Tassara, O. Byron, M. Rocco, *Structure* **2005**, *13*, 723.
- [7] O. Byron, *Biophys. J.* **1997**, *72*, 408.
- [8] P. Zipper, H. Durchschlag, *Progr. Colloid Polym. Sci.* **1997**, *107*, 58.
- [9] B. Demeler, "UltraScan. A Comprehensive Data Analysis Software Package for Analytical Ultracentrifugation Experiments." in: *Modern Analytical Ultracentrifugation: Techniques and Methods*, D. J. Scott, S. E. Harding, A. J. Rowe Eds., Royal Society of Chemistry, Cambridge, UK 2005, p. 210 ff.
- [10] E. Brookes, B. Demeler, C. Rosano, M. Rocco, *Eur. Biophys. J. Biophys. Lett.*, **2010**, *39*, 423.
- [11] G. Picariello, P. Ferranti, G. Mamone, P. Roepstorff, F. Addeo, *Proteomics* **2008**, *8*, 3833.
- [12] K. Gekko, H. Noguchi, *J. Phys. Chem.* **1979**, *83*, 2706.
- [13] D. K. McRorie, P. J. Voelker, "Self-Associating Systems in the Analytical Ultracentrifuge", Beckman Instruments Inc., Fullerton, CA 1993, p. 41.
- [14] J. García de la Torre, V. Rodes, *J. Chem. Phys.* **1983**, *79*, 2454.
- [15] J. García de la Torre, B. Carrasco, *Eur. Biophys. J.* **1998**, *27*, 549.
- [16] J. García de la Torre, D. Amorós, A. Ortega, *Eur. Biophys. J. Biophys. Lett.* **2010**, *39*, 381.
- [17] J.-Y. Wu, J. T. Yang, *J. Biol. Chem.* **1970**, *245*, 212.
- [18] F. J. Castellino, R. Barker, *Biochemistry (USA)* **1968**, *7*, 2207.
- [19] J. B. Sumner, N. Gralén, *J. Biol. Chem.* **1938**, *125*, 33.
- [20] G. Vriend, *J. Mol. Graph.* **1990**, *8*, 52.
- [21] J. García de la Torre, J. G. Hernández Cifre, A. Ortega, R. Rodríguez Schmidt, M. X. Fernandes, H. E. Pérez Sánchez, R. Pamies, *J. Chem. Theory Comput.* **2009**, *5*, 2606.