Eur Biophys J
DOI 10.1007/s00249-015-1042-9

ORIGINAL PAPER

Computing translational diffusion and sedimentation coefficients: an evaluation of experimental data and programs

Mattia Rocco[1] · **Olwyn Byron**[2]

Received: 17 March 2015 / Revised: 4 May 2015 / Accepted: 13 May 2015

Abstract Hydrodynamic characterisation of (bio)macromolecules is a well-established field. Observables linked to translational friction, such as the translational diffusion ($D^0_{t\,(20,w)}$) and sedimentation ($s^0_{(20,w)}$) coefficients, are the most commonly used parameters. Both can be computed starting from high-resolution structures, with several methods available. We present here a comprehensive study of the performance of public-domain software, comparing the calculated $D^0_{t\,(20,w)}$ and $s^0_{(20,w)}$ for a set of high-resolution structures (ranging in mass from 12,358 to 465,557 Da) with their critically appraised literature experimental counterparts. The methods/programs examined are AtoB, SoMo, BEST, Zeno (all implemented within the US-SOMO software suite) and HYDROPRO. Clear trends emerge: while all programs can reproduce $D^0_{t\,(20,w)}$ on average to within ±5 % (range −8 to +7 %), SoMo and AtoB slightly overestimate it (average +2 and +1 %, range −2 to +7 and −4 to +5 %, respectively), and BEST and HYDROPRO underestimate it slightly more (average −3 and −4 %, range −7 to +2 and −8 to +2 %, respectively). Similar trends are observed with $s^0_{(20,w)}$, but the comparison is likely affected by the necessary inclusion of the partial specific volume in the computations. The somewhat less than ideal performances could result from the hydration treatment in BEST and HYDROPRO, and the bead overlap removal in SoMo and AtoB. Interestingly, a combination of SoMo overlapping bead models followed by Zeno computation produced better results, with a 0 % average error (range −4 to +4 %). Indeed, this might become the method of choice, once computational speed considerations now favouring the 5 Å-grid US-SOMO AtoB approach are overcome.

Keywords Hydrodynamics · Analytical ultracentrifugation · Multi-resolution modelling · Dynamic light scattering

Electronic supplementary material The online version of this article (doi:10.1007/s00249-015-1042-9) contains supplementary material, which is available to authorized users.

✉ Mattia Rocco
mattia.rocco@hsanmartino.it

[1] Biopolimeri e Proteomica, IRCCS AOU San Martino-IST, Istituto Nazionale per la Ricerca sul Cancro, Largo R. Benzi 10, 16132 Genova, Italy

[2] School of Life Sciences, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow G12 8QQ, UK

Introduction

The study of the hydrodynamic properties of (bio)macromolecules is a well-established field. Before the advent of the high-resolution structures era, it was mainly used to determine their overall size and shape in solution, by measuring observables such as the translational diffusion coefficient D_t, the sedimentation coefficients s, the rotational correlation time τ_c and the intrinsic viscosity $[\eta]$ (Scheraga and Mandelkern 1953). These properties, reduced to standard conditions (water at 20 °C) and preferably extrapolated to infinite dilution ($D^0_{t\,(20,w)}$, $s^0_{(20,w)}$, $\tau^0_{c(20,w)}$), could then be compared with those calculated for simple geometrical bodies (spheres, ellipsoids of revolution, cylinders), for which exact or very accurate expressions exist (Broersma 1960; Perrin 1936; Simha 1940), representing the structure under scrutiny. Soon it was realised that these simple geometrical objects could seldom reproduce well the shape of (bio)macromolecules, even at low resolution, and methods to calculate the hydrodynamic properties of ensembles of

Published online: 12 June 2015

geometrical objects started to be developed (Kirkwood 1949, 1954). Owing to its simplicity and versatility, bead modelling, i.e. the use of arrays of spheres to represent a three-dimensional object, eventually emerged as the major player in the field, thanks principally to the developmental efforts of V. Bloomfield, J. García de la Torre, R. F. Goldstein, S. C. Harvey and W. A. Wegener (e.g. Bloomfield et al. 1967a, b; García de la Torre and Bloomfield 1977a, b, c, 1978; Goldstein 1985; Harvey 1979; Harvey and Cheung 1980; Wegener 1982; Wegener et al. 1980). An important feature of all bead modelling procedures is that the effect of each bead on all the other beads must be taken into account by a "hydrodynamic interaction" tensor. Several versions of this tensor have been developed, the most important of which are the Rotne–Prager–Yamakawa tensor valid for overlapping beads of equal size (Rotne and Prager 1969; Yamakawa 1970) and its extension to non-overlapping beads of different sizes developed by García de la Torre and Bloomfield (1977a). However, no hydrodynamic interaction tensor has been formulated for overlapping beads of different sizes, thus overlap avoidance or removal is necessary in several bead modelling methods. The hydrodynamics are then computed by solving, usually by matrix inversion procedures, the resulting system of N equations (where N is the number of beads in the model) with $3N$ unknowns, meaning that computing times approximately grow as N^3 (García de la Torre and Bloomfield 1981).

An early key observation, which affects all hydrodynamic modelling procedures, was that it is not possible to correctly predict the hydrodynamics of biomacromolecules if only their anhydrous volume is considered [see, e.g. page 586 in Cantor and Schimmel (1980)]: a "hydration layer" of "tightly bound" waters (i.e. 0.3–0.4 g/g of protein) was found to be required to satisfactorily reproduce the measured parameters. This was seemingly confirmed in a seminal work by Kuntz and Kauzmann (1974), who measured by nuclear magnetic resonance (NMR) spectroscopy the amount of "un-freezable" water molecules bound to each amino acid, thus establishing a direct relation with the hydration layer required by hydrodynamics. However, this picture of a static water layer moving with a biomacromolecule is at odds with measured water residence times on its surface, orders of magnitude shorter even than tumbling times [ps versus ns (Denisov and Halle 1995)]. An elegant solution to the hydration layer conundrum was more recently put forward by Halle and Davidovic (2003), who postulated that the observed effect is a consequence of the increase in water molecule density, in comparison with the bulk value, as they approach a charged/polar surface, resulting in a local change of viscosity. It turns out that this local viscosity change, which would be very hard to properly take into account in computations, is quite well compensated for by the traditional static hydration layer representation.

A big step forward in hydrodynamic modelling was driven by the exponential growth of high-resolution biomolecular structures in the Protein Data Bank (PDB; http://www.rcsb.org) (Berman et al. 2000). It became important to be able to accurately calculate hydrodynamic properties starting from high-resolution structures, either to verify that they were compatible with solution data [e.g. integrins, see Rocco et al. (2008)] or to discriminate between models or larger complexes assembled from individual, atomic-resolution structures [e.g. Nöllmann et al. (2004)]. In this context, it is important to stress that, for a proper comparison, the structures on which the computations are performed must be complete; i.e., they should contain the very same residues and ligand/prosthetic groups, with all atoms in place, as for the samples on which the solution studies were made.

Building on early work (Teller et al. 1979; Venable and Pastor 1988), Byron (1997) developed AtoB, an automated method for construction of bead models from atomic-resolution coordinates wherein the biomacromolecule is subdivided into equally sized cubes and each residue is assigned to a particular cube. Thereafter, beads of either equal or differing radii are generated and placed at the centre of gravity of the residues assigned to each cube, their overlaps are removed by radial reduction, and the resolution of the final model depends on the spacing of the initial cubic grid.

This was followed in 2000 by what became the workhorse in the field: the shell-modelling approach of García de la Torre and co-workers [HYDROPRO (García de la Torre et al. 2000; Ortega et al. 2011)], utilising a two-step scheme in which the original atoms are first replaced by overlapping, "dummy" beads all of the same radius (ad hoc adjusted to account for the hydration) and the resulting model is then sequentially covered by layers (shells) of increasingly much smaller, touching beads. Hydrodynamic computations are then undertaken for the shell models, and the results are extrapolated to zero shell bead radius.

Recognising that both approaches had their drawbacks, our groups jointly developed the SoMo (Solution Modeler) method, with the aim of producing relatively coarse-grained models in which the correspondence between residues and beads could be preserved, allowing also a better treatment of the hydration issue (Rai et al. 2005). In the SoMo method, each residue in a protein is represented by two beads, one for the main- and one for side-chain parts, appropriately positioned (fewer or more beads can be used for e.g. carbohydrates, nucleic acids, lipids, ligand or prosthetic groups). The volume of each bead is calculated by summation of the volume of the constituent atoms plus, importantly, that of the theoretically "bound" waters of hydration. An accessible surface area (ASA) screen is

first performed on the original atomic-resolution structure, identifying exposed and buried side- and main-chain segments. The beads corresponding to the exposed side-chain segments are placed first, and their overlaps are removed, either in a hierarchical or a synchronous way, by proportionally reducing each bead radius but at the same time translating their centres outward by the same amount, in an attempt to preserve the original surface envelope. The exposed main-chain beads are then placed, and the overlaps between them (or with previously treated beads) are removed by reducing their radii only, without outward translation (OT). Finally, the beads corresponding to buried residues are placed and their overlaps with all other beads removed again by reduction of their radii only. An additional ASA screen is then performed on the bead model thus generated in order to check the final exposed/buried status of each bead, and the hydrodynamic computations are carried out only on the exposed set, thus greatly reducing the computational load.

Improved versions of both AtoB (including ASA screens, OT, and a direct hydration procedure) and SoMo were subsequently implemented in the public-domain UltraScan SOlution MOdeler suite (US-SOMO; http://somo.uthscsa.edu/), with an advanced graphical user interface (GUI) (Brookes et al. 2010b). A series of user-editable "master" tables facilitate the proper conversion of residues into beads, resulting however in a more demanding program requiring that each atom/residue in a structure be properly defined in the conversion tables. To partially obviate this coding necessity, approximations can be used in the bead generation step with minimal influence on the final hydrodynamic computations if the non-coded parts are limited to a small percentage of the total biomacromolecule (Brookes et al. 2010a). Over the years, US-SOMO (Brookes et al. 2010a, b) has grown to include a number of other tools, utilities and non-hydrodynamic data analysis and simulation methods, such as a discrete molecular dynamics [DMD (Ding and Dokholyan 2006; Dokholyan et al. 1998)], a general small-angle scattering (SAS) and a dedicated high-performance liquid chromatography/small-angle X-ray scattering (HPLC-SAXS) module (Brookes et al. 2013). A very detailed description of the hydrodynamic computation and analysis tools currently available in US-SOMO and of their operation can be found in Brookes and Rocco (2015).

Among the alternatives to bead modelling, two are of particular interest because their implementation is available in public-domain programs [see Rai et al. (2005) for a more detailed discussion]. The first involves the approximate analogy between certain hydrodynamic and electrostatic properties (Mansfield and Douglas 2008). Hydrodynamic and electrostatic properties are determined, respectively, by the Navier–Stokes and Laplacian equations. However, a specific orientational averaging of the Navier–Stokes equations brings them into the form of Laplace's equation. In particular, in this scheme, the hydrodynamic radius and the intrinsic viscosity of a macromolecule become proportional, respectively, to the capacitance and polarisability of a perfect conductor that has the same shape as the macromolecule, and can thus be computed with errors not exceeding 1 and 5 %, respectively. Moreover, the Laplacian operator can be related to random paths whose step size has a finite variance, and thus it can be formally expressed as an average over random walk trajectories. These concepts have been implemented in the software Zeno (http://www.stevens.edu/zeno/) (Kang et al. 2004), in which an arbitrary-shaped object (e.g. the biomacromolecule) is enclosed within a sphere and random walks are then launched from this sphere, hitting the object or returning to the launch surface. A Monte Carlo numerical path integration generates a large number of random walks, and since the Laplacian operator governs the statistics of these walks, sums taken over them yield the electrostatic capacitance and polarisability tensor of the object, from which $D^0_{t\,(20,w)}$ and $[\eta]$ are then calculated (Mansfield and Douglas 2008). While Zeno offers advantages for the computation of biomacromolecular hydrodynamics, it does absolutely require a properly hydrated model. Zeno was added to the US-SOMO suite to provide an alternative to standard matrix inversion procedures [e.g. García de la Torre et al. (1994)] for the computation of a limited range of hydrodynamic parameters for existing bead models.

The second alternative methodology that we discuss here is the boundary element procedure, as implemented in the software BEST (Aragon 2004, 2011) (http://esmeralda.sfsu.edu/), which is based on the direct evaluation of the stress forces acting on surface elements of an arbitrarily shaped body, considering the velocity field of the solvent flow as an integral over the particle surface. This is made possible by discretising the surface into very small elements, for which the Oseen–Burgers hydrodynamic interaction tensor (exact in the limit of an infinitesimal element) (Burgers 1938; Oseen 1927) can be applied. Because truly approaching this limit would require an enormous computing effort, the computations are performed on several discretised surfaces of decreasing size (the boundary elements), and as for shell-bead modelling (García de la Torre et al. 2000; Ortega et al. 2011), the results are extrapolated to zero element size. In practice, the surface of the original structure is first computed using MSROLL (Connolly 1993), which produces a very finely tessellated surface. In a second step, the number of triangular surface elements is reduced by coalescing them into larger patches, producing tessellated models at different resolution levels. The number and resolution of the models can be manually selected by the user (typical 4–6 models with 2000–6000 elements), but the latter can be also heuristically determined based on

the molecular weight of the input structure. The computations are then carried out for each model, and the results are extrapolated to zero element size. As with HYDRO-PRO (García de la Torre et al. 2000; Ortega et al. 2011), the hydration contribution needs to be taken into account. This takes place in the MSROLL step, with an increase in the radii of the atoms in the starting structure by an ad hoc amount determined by comparing experimental and computed parameters for a number of test structures (Aragon 2004). BEST was recently made available under the US-SOMO GUI with the calculations (which can be demanding) being farmed out to a remote supercompute cluster. The results returned for each model can then be plotted as a function of 1/(total element number) with a graphic utility in US-SOMO, allowing extrapolation to an infinite number of elements and then producing the final values of the hydrodynamic parameters.

It is also important to stress that all the hydrodynamic computation methods described so far operate in the so-called rigid-body approximation; that is, they assume that the structure/model has no moving regions. Flexibility/conformational variability issues require advanced treatments not dealt with in this work [for an introduction, see Rocco and Brookes (2014); Rocco and Byron (2015)].

In this paper, we present a detailed evaluation of the hydrodynamic modelling programs described above, all implemented under the US-SOMO GUI except HYDRO-PRO, for which the Windows version (http://leonardo.inf.um.es/macromol/programs/hydropro/hydropro.htm) WinHydropro (Ortega et al. 2011) was used. The comparison was limited to translational frictional properties, since measurements of $D^0_{\mathrm{t}\,(20,\mathrm{w})}$ and $s^0_{(20,\mathrm{w})}$, although less shape sensitive than $\tau^0_{\mathrm{c}(20,\mathrm{w})}$ and $[\eta]$, are by far the most commonly performed and thus utilised in modelling studies. The experimental data for a number of globular proteins (including monomeric as well as multimeric entities) spanning a molecular weight range of 12,358–465,557 Da, whose atomic-resolution structures are downloadable from the PDB, were taken from the literature and critically assessed. The evaluation reveals that all software examined can reproduce the translational friction hydrodynamics on average to within ~±5 % of the experimental value, but with significant differences between them. Guidelines for the choice and usage of the available programs are then offered.

Methods

Protein crystal structures, selected to be from the same species for which reliable solution data were available in the literature, were taken from the PDB [(Berman et al. 2000); http://www.rcsb.org/pdb/home/home.do]. Most of these structures had already been utilised in previous studies by our groups (Brookes et al. 2010a, b; Rai et al. 2005) and had had missing atoms and/or residues modelled using WHATIF [(Vriend 1990); http://swift.cmbi.ru.nl/servers/html/index.html] and O (Jones et al. 1991). The same procedures were applied, when necessary, to the additional proteins utilised in this study, all of which were taken from the list in Hahn and Aragon (2006). The presence of carbohydrate chains on the surface was limited to a single example (α-lactalbumin). As before (Brookes et al. 2010b; Rai et al. 2005), the literature data were carefully evaluated to verify that correction to standard conditions and extrapolation to infinite dilution were rigorously performed. This included digitising data contained in figures, using Paint Shop Pro 5.03 (JASC Software, now distributed by Corel Corp., Ottawa, ON, Canada) and then performing linear regressions using TableCurve2D 4 (SPSS, now distributed by Systat Software Inc., San Jose, CA, USA), allowing retrieval of the standard deviation (SD) associated with the extrapolated values, which was often missing in the primary source. An Excel spreadsheet containing all the literature data analyses is presented as Electronic Supplementary Material.

Hydrodynamic bead modelling computations provide the infinite-dilution translational frictional coefficient $f^0_{\mathrm{t}\,(20,\mathrm{w})}$, from which $D^0_{\mathrm{t}\,(20,\mathrm{w})}$ can be obtained as

$$D^0_{\mathrm{t}(20,\mathrm{w})} = \frac{k_\mathrm{b}T}{f^0_{\mathrm{t}(20,\mathrm{w})}}, \tag{1}$$

where k_b is Boltzmann's constant and T is the absolute temperature (293.15 K). BEST directly provides $D^0_{\mathrm{t}\,(20,\mathrm{w})}$, while Zeno computes the hydrodynamic radius $R_{\mathrm{h}(20,\mathrm{w})}$, related to $f^0_{\mathrm{t}\,(20,\mathrm{w})}$ by

$$R_{\mathrm{h}(20,\mathrm{w})} = \frac{f^0_{\mathrm{t}(20,\mathrm{w})}}{6\pi\,\eta_{(20,\mathrm{w})}}, \tag{2}$$

where $\eta_{(20,\mathrm{w})}$ is the viscosity of water at 20 °C. From $D^0_{\mathrm{t}\,(20,\mathrm{w})}$, $s^0_{(20,\mathrm{w})}$ can then be obtained as

$$s^0_{(20,\mathrm{w})} = \frac{D^0_{\mathrm{t}(20,\mathrm{w})}M(1-\bar{v}_{(20,\mathrm{w})}\rho_{(20,\mathrm{w})})}{RT}, \tag{3}$$

where $\rho_{(20,\mathrm{w})}$ is the density of water at 20 °C, R is the gas constant, and M and $\bar{v}_{(20,\mathrm{w})}$ are the molecular weight and partial specific volume (also under standard conditions), respectively, of the biomacromolecule. However, while M can be calculated quite accurately from the (e.g. amino acid) composition, doing so for $\bar{v}_{(20,\mathrm{w})}$ is notoriously less reliable [see Brookes et al. (2010b)]. Whenever reliable experimental $\bar{v}_{(20,\mathrm{w})}$ values were provided in the literature, they were employed in the computations in this evaluation. Otherwise, the values calculated by US-SOMO (Brookes

et al. 2010b) were used. Of particular importance, the literature $D^0_{t\,(20,w)}$ values considered were directly determined and not derived from $s^0_{(20,w)}$ values, as was done in a number of cases in the comparisons reported by Hahn and Aragon (2006).

US-SOMO was run on an AMD Athlon 64 X2 Dual Core processor PC with 8 GB RAM, operating under the Ubuntu 14.04 64-bit Linux OS. The synchronous overlap removal routines were used for both SoMo and AtoB bead model generation, with 70 % overlap threshold for bead fusion, and outward translation of the beads' centres when removing the overlaps between the exposed side-chain beads, and between the initially exposed beads, respectively. ASA screen cut-offs were the default settings for both methods (20 $Å^2$ for residues and 50 % of their total surface for beads in SoMo; 10 $Å^2$ for the initial beads and 50 % of their total surface for the final beads in AtoB; probe radius 1.4 Å in all cases). The supermatrix inversion method with computations relative to the diffusion centre, stick boundary conditions and exclusion of the buried beads was used for both SoMo and AtoB bead models. Zeno computations were performed with 10^6 steps.

BEST calculations were carried out mainly on the Alamo supercompute cluster of the University of Texas at San Antonio, a hybrid composed of 16 nodes with dual AMD Opteron 2378 Quad-Core processor, 4 GB RAM/core, and of 20 nodes with one AMD Phenom 9750 Quad-Core processor, 2 GB RAM/core; jobs were processed on available nodes by the cluster's scheduler. More recently, they were run on the Stampede system at the Texas Advanced Computing Center (TACC) of the University of Texas at Austin, which is a 10-PFLOP Dell Linux cluster based on 6400+ Dell PowerEdge server nodes, each outfitted with 2 Intel Xeon E5 (Sandy Bridge) processors and an Intel Xeon Phi coprocessor (MIC Architecture). The majority of the 6400 nodes are configured with two Xeon E5-2680 processors and one Intel Xeon Phi SE10P coprocessor (on a PCIe card). These computer nodes are configured with 32 GB of "host" memory with an additional 8 GB of memory on the Xeon Phi coprocessor.

MSROLL (Connolly 1993) parameters were: probe radius 1.5 Å, starting finesse angle 0.6° and 60,000 maximum output triangles. Six tessellated models were generated for each structure examined, and two methods were employed to define the number of plates in each model: manual, where the upper and lower limits were always 6000 and 2000 plates, respectively, and automatic, using the heuristic approach implemented in BEST, $30 \times \sqrt{M}$ and $18 \times \sqrt{M}$, respectively. The results generated were examined in the US-SOMO BEST interface, applying the Q test criterion (Dean and Dixon 1951) to find potential outliers. In a few cases, and for high-M structures only, the data computed for the point with the lower number of plates were manually discarded [see Brookes and Rocco (2015) for a more detailed description of the BEST interface in US-SOMO]. WinHydropro was run on an Intel Core i5-3470 3.2 GHz PC with 6 GB RAM, operating under the Windows 7 Professional OS, utilising the default settings: shell models from atomic level, atomic element radius 2.84 Å, automatic radius of shell elements with 6 models generated (with a maximum number of 2000 shell beads) and automatic extrapolation of the parameters to zero shell bead size.

Results

The results of the comparison between experimental and calculated $D^0_{t\,(20,w)}$ and $s^0_{(20,w)}$ values utilising the various modelling approaches are reported in Tables 1 and 2, respectively. The test proteins are identified with a progressive number, their common name (sometimes abridged) and the name of the PDB file utilised for the hydrodynamic modelling. The molecular weight computed from the composition is also reported, followed by the $D^0_{t\,(20,w)}$ or $s^0_{(20,w)}$ values with their associated experimental precision (±SD), respectively, and the literature reference number(s) from which they were taken. In Table 2, an extra column reports the $\bar{v}_{(20,w)}$ values, either experimental or, in parentheses, computed from the composition by US-SOMO (Brookes et al. 2010a, b). The next eight columns report the percentage difference (Δ%) between the experimental values and those computed by the various methods. To allow for a comparison with previous results obtained by BEST (Hahn and Aragon 2006), we have made a distinction between monomeric and multimeric proteins, as indicated by the horizontal line dividing the tables into two sections. At the bottom of Tables 1 and 2, respectively, three and two different kinds of Δ% averages are reported. For Table 1, we report the mean Δ% values for monomeric proteins only and for all proteins considered, without discarding any single Δ% value; a third average was then made for the Δ% of all proteins except five whose Δ% values were clearly severely either over- or underestimated by most methods. In this case, in the absence of other evidence, we assumed that either the experimental value was unreliable, or that the solution structure of the test protein was different from the crystal. Since this comparison was aimed at finding the performance of each method in reproducing reliable experimental values starting from a reliable three-dimensional structure, we feel that simply discarding potential "global" outliers is justified. In any case, this only slightly affected the $D^0_{t\,(20,w)}$ Δ% all-proteins average. Since for the $s^0_{(20,w)}$ values the Δ% were consistently higher than for the $D^0_{t\,(20,w)}$, and just three clear big outliers were apparent, in Table 2 only two Δ% averages are reported, for the

Table 1 Proteins used for the comparison, with their PDB code, molecular weight (mol. wt.), experimental diffusion coefficient (±SD) ($D^0_{t(20,w)}$ expt.) taken from the literature (Ref.), and percent difference between the computed and experimental $D^0_{t(20,w)}$ values (Δ% $D^0_{t(20,w)}$ comp.) for each of the different methods used. The horizontal line between entries 13 and 14 separates monomeric from multimeric solution forms

#	Protein[a]	PDB	Mol. wt.	$D^0_{t(20,w)}$ expt., F[b]	Ref.[c]	Δ% $D^0_{t(20,w)}$ comp.							
						SoMo SMI	SoMo Zeno	SoMo ov Zeno	AtoB G5 SMI	AtoB G2 SMI	HP auto	BEST man	BEST heur
1	Cytochrome c	1HRC	12,357.5	12.10 ± 0.50	1–3	+1.9	+0.8	−1.7	−1.7	−1.1	−3.3	−3.7	−3.3
2	Ribonuclease A	8RAT	13,683.8	11.06 ± 0.31	4[d]	+7.1	+5.8	+3.1	+5.3	+5.9	−0.9	−1.3	−1.3
3	α-Lactalbumin	1A4V[e]	15,784.7	10.90 ± n.a.	5[f]	+0.2	−0.9	−3.7	−2.5	−2.5	−12.2	−7.0	−7.1
4	Lysozyme	1AKI	14,306.7	11.40 ± 0.42	4[f]	+4.4	+3.5	+0.9	+1.6	+2.8	−4.1	−3.2	−3.1
5	Myoglobin CO	1DWR	17,522.0	10.70 ± n.a.	5[f]	+1.5	+0.9	−1.9	−0.5	−0.7	−4.5	−4.3	−4.3
6	Soybean trypsin inh.	1AVU	19,962.8	9.47 ± 0.18*	2, 6	+7.4	+6.7	+4.5	+5.4	+5.9	+1.7	+1.3	+1.2
7	β-Trypsin	1TPO	23,335.9	9.40 ± 0.30*	7	+8.1	+7.4	+4.3	+5.3	+5.7	+0.5	+1.3	+1.3
8	Trypsinogen	1TGN	23,182.7	9.68 ± 0.23	8	+5.1	+4.3	+1.2	+2.4	+3.0	−2.6	−1.7	−1.7
9	α-Chymotrypsin (mon)	4CHA	25,236.5	10.20 ± 0.02*	9	−5.8	−5.9	−7.8	−7.4	−7.2	−11.9	−11.2	−11.1
10	Chymotrypsinogen A	2CGA	25,659.0	9.49 ± 0.02	5[f]	+1.6	+1.2	−0.9	−0.4	+0.4	−5.7	−4.4	−4.6
11	Carbonic anhydrase B	2CAB	28,820.5	8.89 ± 0.03	10	+4.5	+3.9	+1.5	+1.8	+2.6	−1.2	−0.8	−0.8
12	Pepsin	4PEP	34,588.6	8.71 ± 0.05	11, 12	−1.4	−2.2	−3.9	−3.3	−2.6	−7.7	−6.8	−6.9
13	H. serum albumin	1AO6	66,428.6	6.31 ± 0.09	13	−2.1	−1.7	−2.9	−2.1	−1.9	−6.0	−4.3	−4.6
14	Superoxide dismutase	2SOD	31,442.2	8.27 ± n.a.	14	+2.1	+1.8	−0.1	+1.5	+1.3	−2.5	−1.8	−1.7
15	β-Lactoglobulin	1BEB	35,224.7	7.85 ± 0.08	5[f]	+0.9	+0.5	−1.0	+0.4	+0.3	−3.6	−2.7	−2.7
16	α-Chymotrypsin (dim)	4CHA	50,473.5	7.20 ± 0.29	15, 16	+4.3	+3.9	+2.2	+3.1	+3.1	−1.8	−0.6	−0.8
17	Triosephosphate isom.	1YPI	53,331.4	7.21 ± 0.10	3	−1.4	−1.4	−3.3	−2.4	−1.9	−5.8	−4.2	−5.0
18	Hemoglobin CO	1HCO	64,559.7	6.89 ± 0.09	5[f]	+1.3	+0.9	−0.4	−0.1	+0.1	−2.9	−1.7	−2.3
19	Citrate synthase	1CTS	97,845.5	5.80 ± n.a.	4[f]	+1.0	+0.9	+1.2	0.0	+0.3	−4.0	−1.7	−2.2
20	Inorganic pyrophosph.	1FAJ	117,339.0	5.65 ± 0.12*	17	−4.8	−4.4	−5.5	−5.3	−5.3	−8.0	−6.7	−6.0
21	G3PD apo	2GD1	143,787.8	5.00 ± 0.20	5[f]	+1.6	+1.8	+0.4	+0.2	+0.6	−1.6	−0.6	−1.0
22	G3PD holo	1GD1	146,437.7	5.20 ± 0.20	5[f]	−1.9	−1.7	−3.1	−2.9	−2.7	−5.4	−3.8	−4.2
23	LDH pig H + NAD	5LDH	148,942.6	5.06 ± 0.15	5[f]	+1.6	+1.6	+0.6	+0.4	+0.6	−1.2	−1.4	−0.8
24	LDH pig M + NAD	9LDH	149,063.5	5.25 ± 0.16	5[f]	0.0	+0.2	−1.1	−1.5	−1.3	−4.4	−2.9	−3.0
25	Aldolase	1ADO	157,131.2	4.50 ± 0.26	18−21	+4.7	+5.6	+3.8	+4.0	+4.2	+1.8	+1.6	+1.8
26	Catalase	4BLC	235,775.1	4.10 ± n.a.*	5[f]	+6.6	+7.3	+6.1	+5.9	+6.1	+0.2	–	–
27	β-Galactosidase	1BGL	465,257.6	3.13 ± 0.04	22	+4.2	+4.8	+4.2	+3.5	+1.6	+1.0	–	–

Table 1 continued

#	Protein[a]	PDB	Mol. wt.	$D^0_{t(20,w)}$ expt., F[b]	Ref.[c]	$\Delta\%\ D^0_{t(20,w)}$ comp.							
						SoMo SMI	SoMo Zeno	SoMo ov Zeno	AtoB G5 SMI	AtoB G2 SMI	HP auto	BEST man	BEST heur
	Mean Δ% monomeric					+2.5 ± 4.1	+1.8 ± 3.9	−0.6 ± 3.6	+0.3 ± 3.8	+0.8 ± 4.0	−4.5 ± 4.3	−3.5 ± 3.5	−3.6 ± 3.5
	Mean Δ% all					+1.9 ± 3.5	+1.7 ± 3.4	−0.1 ± 3.3	+0.4 ± 3.3	+0.6 ± 3.4	−3.6 ± 3.6	−2.9 ± 2.9	−3.0 ± 2.9
	Mean Δ% all without "*" values					+1.9 ± 2.5	+1.6 ± 2.4	−0.2 ± 2.4	+0.3 ± 2.4	+0.6 ± 2.3	−3.6 ± 3.0	−2.7 ± 2.1	−2.8 ± 2.2

SoMo SoMo models without overlaps, *SoMo ov* SoMo models with overlaps, *AtoB G5* AtoB models generated with a 5 Å grid, *AtoB G2* AtoB models generated with a 2 Å grid, *SMI* hydrodynamic computations with the supermatrix inversion procedure, *Zeno* hydrodynamic computations using the Zeno method, *HP auto* HYDROPRO shell models generation and hydrodynamic computations via the SMI method with the automatic determination of the number of beads in each shell model with a maximum of 2000 beads, *BEST man* BEST models and hydrodynamic computations with the manual setting for the determination of minimum and maximum number of plates, *BEST heur* BEST models and hydrodynamic computations with the heuristic approach for the determination of minimum and maximum number of plates

[a] For the species of origin of the protein considered, see the PDB headers and the literature cited (Ref.). *Mon* monomer, *dim* dimer

[b] Values marked with "*" were not considered when taking the mean values reported in the last row (mean Δ% all without "*" values)

[c] See the correspondence between the numbers listed and the references at the end of these table footnotes

[d] See references cited within this paper where the mean value was however erroneously stated as 11.6 F instead of 11.06 F

[e] Carbohydrate not present in the PDB structure, manually modelled

[f] See references cited within this paper

[1] Clark et al. (2002); [2]Walters et al. (1984); [3]Wilkins et al. (1999); [4]Rai et al. (2005); [5]Brookes et al. (2010b); [6]Hanlon et al. (2010); [7]Cunningham (1954); [8]Tietze (1953); [9]Schwert and Kaufman (1951); [10]Armstrong et al. (1966); [11]Edelhoch (1957); [12]Neurath et al. (1941); [13]Charlwood (1952); [14]Wood et al. (1971); [15]Kunitz and Northrop (1935); [16]Schwert (1949); [17]Wong et al. (1970); [18]Christen et al. (1965); [19]Glikina and Finogenov (1950); [20]Kawahara (1969); [21]Taylor et al. (1948); [22]Sund and Weber (1963)

monomeric and all-proteins sets, respectively, each without taking into account those outliers. Although potentially interesting, we have made no attempt to investigate the real reasons for the large discrepancies found in either $D^0_{t\,(20,w)}$ or $s^0_{(20,w)}$ for the few identified "outliers", since that would be beyond the scope of this work. Noteworthily, for all the outliers (except inorganic pyrophosphatase) large discrepancies (6–18 %) between the molecular weights calculated with the Svedberg equation from the experimental $D^0_{t\,(20,w)}$ and $s^0_{(20,w)}$ values and those computed from the composition reported in Tables 1 and 2 were found as well (data not shown), reinforcing the likelihood of issues with the experimental values themselves.

A graphical representation of the final average values presented in Tables 1 and 2 is reported in Fig. 1, panels a and b, respectively. There, the proteins are plotted by their progressive number listed in the tables, and the empty symbols connected with straight lines correspond to the Δ% values calculated by each method and included in the final mean in Table 1 and in all means in Table 2, while the solid symbols are the outliers described above. To aid in appreciating the relative magnitude of the Δ% between experimental and computed values and the precision of the experimental values, the average Δ% SD of the latter are represented as horizontal dotted lines in each panel. The vertical dashed lines indicate the boundary between the monomeric and multimeric proteins. Thus, the molecular weight of the proteins increases from left to right from ~12,400 to ~66,400 for protein number 1–13, and restarts from ~31,400 to ~465,600 for protein number 14–27.

We can now proceed with assessing the results of our modelling/computations, considering first the $D^0_{t\,(20,w)}$ results. We start from the SoMo bead modelling method with overlap reduction coupled with the standard supermatrix inversion procedure to compute the hydrodynamics (column 7 in Table 1, blue diamonds in Fig. 1, panel a). Outliers included, the average Δ% is on the positive side but still within the experimental average Δ% SD for the monomeric proteins, and is even lower than that obtained when all proteins are considered. Exclusion of the outliers does not affect this result, probably because of compensation between the values above and below the mean. By also looking at Fig. 1, panel a, it is apparent that the standard SoMo method tends to slightly overestimate (average +2.5 to +1.9 %) the $D^0_{t\,(20,w)}$ of the test proteins. The upper limit is around +6 %, but it should be borne in mind that any individual experimental value/protein structure among those utilised could be potentially "wrong", and therefore the average Δ% values are more informative of how well a method can reproduce the experimental parameters. If we utilise the same SoMo models without overlaps but resort instead to the Zeno computational method (Table 1,

Table 2 Proteins used for the comparison, with their PDB code, molecular weight (mol. wt.), partial specific volume ($\bar{v}_{(20,w)}$), experimental sedimentation coefficient (±SD) ($s^0_{(20,w)}$ expt.) taken from the literature (Ref.), and percent difference between the computed and experimental $s^0_{(20,w)}$ values ($\Delta\%\ s^0_{(20,w)}$ comp.) for each of the different methods used. The horizontal line between entries 13 and 14 separates monomeric from multimeric solution forms

#	Protein[a]	PDB	Mol. wt.	$\bar{v}_{(20,w)}$[b] cm^3/g	$s^0_{(20,w)}$ expt.[c] S	Ref.[d]	$\Delta\%\ s^0_{(20,w)}$ comp.							
							SoMo SMI	SoMo Zeno	SoMo ov Zeno	AtoB G5 SMI	AtoB G2 SMI	HP auto	BEST man	BEST heur
1	Cytochrome c	1HRC	12,357.5	0.724	2.00 ± 0.04*	2,3	−21.0	−14.2	−16.3	−23.5	−23.0	−17.5	−18.1	−17.8
2	Ribonuclease A	8RAT	13,683.8	0.709	2.00 ± 0.03	2[e]	−3.0	−4.0	−6.5	−5.0	−4.0	−10.0	−10.4	−10.4
3	α-Lactalbumin	1A4V[f]	15,784.7	(0.718)	1.76 ± n.a.*	3[e]	+13.6	+12.6	+9.4	+10.8	+10.8	0.0	+5.7	+5.6
4	Lysozyme	1AKI	14,306.7	(0.716)	1.88 ± 0.02	2[e]	+5.9	+5.1	+2.4	+3.2	+4.3	−2.7	−1.8	−1.6
5	Myoglobin apo	2V1 K	17,568.3	0.743	2.05 ± n.a.	3[e]	−1.0	−1.9	−4.6	−3.9	−3.4	−6.3	−6.7	−6.8
6	Soybean trypsin inh.	1AVU	19,962.8	(0.735)	2.29 ± n.a.	4	−2.6	−3.8	−5.7	−4.4	−3.9	−4.4	−8.7	−8.8
7	β-Trypsin	1TPO	23,335.9	(0.724)	2.54 ± 0.02	5	+6.3	+5.6	+2.4	+3.5	+3.9	−1.2	−0.5	−0.5
8	Trypsinogen	1TGN	23,182.7	0.73	2.48 ± 0.01	5	+7.7	+5.1	+2.0	+4.8	+5.6	−1.6	−0.9	−0.9
9	α-Chymotrypsin (mon)	4CHA	25,236.5	(0.733)	2.59 ± 0.02	6	+3.1	+3.0	+0.8	+1.5	+1.5	−3.5	−2.8	−2.7
10	Chymotrypsinogen A	2CGA	25,659.0	(0.732)	2.56 ± 0.03	3[e]	+6.6	+6.3	+4.1	+4.7	+5.5	−0.8	+0.4	+0.2
11	Carbonic anhydrase B	2CAB	28,820.5	0.73	3.01 ± 0.19	7−10	+1.0	−1.5	−3.9	−1.7	−1.0	−6.3	−6.0	−6.0
12	Pepsin	4PEP	34,588.6	(0.723)	3.19 ± 0.08	11	+6.3	+5.5	+3.6	+4.4	+5.0	−0.3	+0.5	+0.4
13	H. serum albumin	1AO6	66,428.6	(0.734)	4.28 ± 0.04	12	+5.1	+5.5	+4.3	+5.1	+5.4	+0.9	+2.8	+2.5
14	Superoxide dismutase	2SOD	31,442.2	(0.718)	3.03 ± 0.05	13	+1.7	+1.6	−0.4	+1.0	+1.0	−3.0	−2.1	−1.9
15	β-Lactoglobulin	1BEB	35,224.7	(0.745)	2.87 ± 0.06	3[e]	+4.9	+1.8	+0.3	+4.2	+4.2	−2.1	−1.4	−1.4
16	α-Chymotrypsin (dim)	4CHA	50,473.5	(0.733)	3.5 ± 0.02*	14	+18.0	+18.7	+16.8	+16.9	+16.9	+12.3	+13.7	+13.3
17	Triosephosphate isom.	8TIM	52,971.4	(0.742)	3.75 ± 0.05	15	+7.5	+6.7	+5.4	+6.1	+6.4	+2.9	+3.7	+3.4
18	Hemoglobin CO	1HCO	64,559.7	0.749	4.5 ± n.a.	3[e]	+4.4	+3.2	+1.9	+3.1	+3.3	−0.7	+2.5	+2.2
19	Citrate synthase	1CTS	97,845.5	0.733	6.1 ± 0.1	2[e]	+1.1	+3.3	+3.7	+0.2	+0.5	−1.6	+0.6	+0.1
20	Inorganic pyrophosph.	1FAJ	117,339.0	(0.743)	7.01 ± 0.04	16	−4.6	−4.2	−5.3	−5.1	−5.1	−8.0	−6.5	−5.8
21	G3PD apo	2GD1	143,787.8	(0.742)	7.6 ± 0.15	3[e]	+2.2	+2.5	+1.1	+0.9	+1.2	−0.8	0.0	−0.4
22	G3PD holo	1GD1	146,437.7	(0.741)	8.1 ± 0.16	3[e]	−1.5	−1.3	−2.7	−2.6	−2.3	−5.2	−3.5	−3.8
23	LDH pig H + NAD	5LDH	148,942.6	(0.746)	7.80 ± 0.08	3[e]	+2.8	+2.8	+1.8	+1.7	+1.9	0.0	−0.2	+0.4
24	LDH pig M + NAD	9LDH	149,063.5	(0.745)	7.79 ± 0.08	3[e]	+5.6	+5.9	+4.4	+4.1	+4.2	+1.2	+2.6	+2.4
25	Aldolase	1ADO	157,131.2	0.742	7.85 ± 0.05	3[e]	+2.5	+1.2	−0.5	+1.9	+2.0	−2.5	−2.7	−2.5
26	Catalase	4BLC	235,775.1	0.73	11.4 ± 0.15	3[e]	+2.1	+1.3	+0.2	+1.2	+1.5	−5.3	–	–
27	β-Galactosidase	1BGL	465,257.6	(0.725)	16.23 ± 0.13	17	+6.0	+6.6	+5.9	+5.4	+3.3	+2.8	–	–

Table 2 continued

#	Protein[a]	PDB	Mol. wt.	$\bar{v}_{(20,w)}$[b] cm^3/g	$s^0_{(20,w)}$ expt.[c]	S	Ref.[d]	$\Delta\%\ s^0_{(20,w)}$ comp.						
								SoMo SMI	SoMo Zeno	SoMo ov Zeno	AtoB G5 SMI	AtoB G2 SMI	HP auto	BEST man	BEST heur
	Mean Δ% monomeric (without * values)							+3.2 ± 4.0	+2.3 ± 4.2	−0.1 ± 4.2	+1.1 ± 4.1	+1.7 ± 4.0	−3.3 ± 3.2	−3.1 ± 4.2	−3.1 ± 4.2
	Mean Δ% all (without * values)							+2.9 ± 3.5	+2.3 ± 3.5	+0.6 ± 3.7	+1.4 ± 3.5	+1.7 ± 3.4	−2.4 ± 3.2	−1.9 ± 3.8	−1.9 ± 3.7

SoMo SoMo models without overlaps, *SoMo ov* SoMo models with overlaps, *AtoB G5* AtoB models generated with a 5 Å grid, *AtoB G2* AtoB models generated with a 2 Å grid, *SMI* hydrodynamic computations with the supermatrix inversion procedure, *Zeno* hydrodynamic computations using the Zeno method, *HP auto* HYDROPRO shell models generation and hydrodynamic computations via the SMI method with the automatic determination of the number of beads in each shell model with a maximum of 2000 beads, *BEST man* BEST models and hydrodynamic computations with the manual setting for the determination of minimum and maximum number of plates, *BEST heur* BEST models and hydrodynamic computations with the heuristic approach for the determination of minimum and maximum number of plates

[a] For the species of origin of the protein considered, see the PDB headers and the literature cited (Ref.). *Mon* monomer, *dim* dimer

[b] The values in parentheses were calculated by US-SOMO; the others are experimental, taken from the literature

[c] The values marked with "*" were not considered when taking the reported mean values

[d] See the correspondence between the numbers listed and the references at the end of these table footnotes

[e] See references cited within this paper

[f] Carbohydrate not present in the PDB structure, manually modelled

[1] Stellwagen (1968); [2]Rai et al. (2005); [3]Brookes et al. (2010b); [4]Rackis et al. (1962); [5]Cunningham (1954); [6]Ghirlando (2011); [7]Armstrong et al. (1966); [8]Coleman (1965); [9]Nyman (1961); [10]Rickli et al. (1964); [11]Edelhoch (1957); [12]Charlwood (1952; [13]Wood et al. (1971); [14]Schwert (1949); [15]McVittie et al. (1977); [16]Wong et al. (1970); [17]Sund and Weber (1963)

column 8; cyan diamonds in Fig. 1, panel a), the results are still slightly overestimated but somewhat improved (averages +1.8 % for the monomeric set, and +1.6 % for the complete set, outliers excluded). The upper limit is also somewhat decreased, to ~+5 %.

A strikingly relatively larger improvement is instead obtained with a new approach that we have taken in this work: the SoMo models are generated without overlap reduction, and the computations are carried out using Zeno, since the supermatrix inversion procedure cannot be used in this case. As can be seen (Table 1, column 9; green stars/crosses in Fig. 1, panel a), excellent results are obtained with this "hybrid" method: a very minor average underestimation (−0.6 %) of the monomeric protein set, and an almost exact average evaluation of the full set, with or without the outliers (−0.2 % for the latter). Even more striking, the individual values are almost all within the experimental average Δ% SD, evenly distributed, and none above ±4 %, outliers excluded.

A similar situation is encountered with the US-SOMO grid method AtoB, with a 5 Å cube size (Table 1, column 10; magenta up-triangles in Fig. 1, panel a). Practically, no differences are found in the various average values, which are around an excellent +0.3 %, but with a range of ~−3 to +5 %, outliers excluded, indicating a small tendency to overestimation. Reducing the AtoB cube size to 2 Å (Table 1, column 11; orange down-triangles in Fig. 1, panel a) had the somewhat unexpected result of producing slightly worse average Δ% values (+0.8 % for the monomeric set, +0.6 % for the entire set, outliers excluded; range ~−3 to +6 %).

We now consider the last of the bead modelling methods examined, the shell-modelling approach of HYDROPRO (Table 1, column 12; red circles in Fig. 1, panel a). The situation appears to be reversed with respect to the SoMo and AtoB methods, with a constant and appreciable underestimation especially for the monomeric proteins set (−4.5 % average Δ%). The situation somewhat improves when the full set is considered, with an average Δ% of −3.6 % (outliers excluded). However, there is a slight trend toward less negative Δ% values as the proteins get bigger, with a near-perfect calculation for catalase, a clear outlier for all other methods. This could partially result from the limited number of shell beads available in the default settings of HYDROPRO (≤2000; García de la Torre et al. 2000), producing rougher models as the structure size increases, perhaps somewhat offsetting the basic observed underestimation. Shape-dependent factors affecting the shell covering could also play a role, as likely happened for catalase (and for α-lactalbumin, whose results are also very different from those obtained with the other methods).

Finally, we examine the boundary element method BEST, using both the manual and heuristic approaches

to determine the upper and lower plate numbers and thus the intermediate values as well, for the six BEST models used in each extrapolation. A first observation is that the heuristic approach needs much more computing power as the molecular size increases. In fact, under our operating conditions, we could not compute the hydrodynamics of the last two proteins in our set using either the heuristic approach or the manual setting with a minimum of 2000 and a maximum of 6000 plates. Moreover, computing times are very dependent on the available computing architecture and queuing privileges (for the larger structures they were on the order of up to 3 weeks on the Alamo cluster, but on the order of 1–2 days on the Stampede cluster, which, however, has a 48 h time limit for processing and therefore some runs could not be completed). While we could have computed the values for the larger proteins by manually reducing the limits, we elected not to do so to avoid introducing other variables, and we preferred to have the best possible models to fully assess the performance of BEST even with a somewhat more limited set of proteins. In the end, this amounted to the omission of just one protein (β-galactosidase), since catalase was found to be an outlier by all other methods (except HYDROPRO, see comments above). Comparing first head-to-head the manual and heuristic BEST results (Table 1, columns 13 and 14, respectively), one can notice that for the monomeric set the differences are mostly irrelevant. Slightly more variability is found as the molecular size increases, but in the end the differences between the various average values were again irrelevant. In any case, as for HYDROPRO, there is a clear trend in BEST of underestimating the $D^0_{t\,(20,w)}$ values, with an average Δ% of −3.5 % for the monomeric protein data set, somewhat improving to −2.7 % for the whole set without outliers in the manual mode (see also Fig. 1, black squares; due to the closeness of most heuristic and manual values, only the latter are plotted). The range (~−7 to +2 %) is also evidence of the tendency of BEST to underestimate $D^0_{t\,(20,w)}$.

We can now compare the values of $s^0_{(20,w)}$ computed by the different methods, bearing in mind that these derive from the same $f^0_{t\,(20,w)}$ values from which $D^0_{t\,(20,w)}$ is extracted. Therefore, variations in performance for a given protein will depend mostly on either the value of $\bar{v}_{(20,w)}$ utilised (see Eq. 3) or on the experimental data themselves. If we start by looking at the data in Fig. 1, panel b, it is immediately evident that the computed $s^0_{(20,w)}$ Δ% values have a larger spread in comparison with their $D^0_{t\,(20,w)}$ counterparts, and for all modelling methods employed. Three large outliers are also striking. Overall, the trends do reflect those observed with $D^0_{t\,(20,w)}$ Δ% values. The SoMo method with overlap removal slightly overestimates the average $s^0_{(20,w)}$ Δ% values with either the supermatrix inversion computational method or the Zeno method, the latter performing

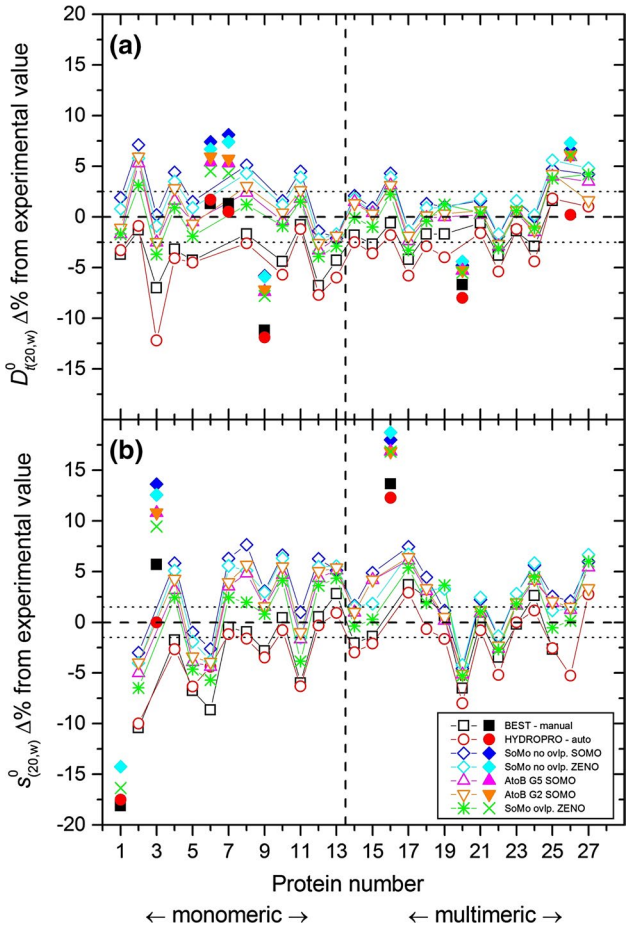

Fig. 1 Plots of the percent difference (Δ%) between calculated and experimental values of $D^0_{t\,(20,w)}$ (*panel* **a**) and $s^0_{(20,w)}$ (*panel* **b**) for the 27 proteins listed in Tables 1 and 2. The *open symbols* connected with *straight lines* are the points utilised to calculate the mean Δ% values reported at the end of Tables 1 and 2, while the *solid symbols* are the points excluded from the final mean in Table 1 and from all means in Table 2. The correspondence between symbols and modelling/computational methods utilised is shown in the *panel* **b** *inset*. The *dotted lines above* and *below* the *horizontal* Δ% = *0 dashed line* represent the average (SD) deviation of the experimental values. The *vertical dashed line* separates monomeric from multimeric proteins

somewhat better (~+3 versus ~+2 %, respectively). No differences are found between the monomeric and full protein data sets, but the overall ranges are quite large (~−5 to +8 and ~−4 to +7 %, respectively). The SoMo method without overlap removal coupled with the Zeno computations is again the best performer, with average Δ% values of −0.1 and +0.6 % for the monomeric and full protein data sets, respectively, and a range of ~−5 to +6 %. The AtoB grid method with either a 5 or 2 Å cube size is the second best performer, with similar average Δ% values of around ~+1 to +2 % for the monomeric and the full data sets. A ~−5 to +5 % Δ% range is observed in both cases. In the end, HYDROPRO and BEST perform

similarly, slightly underestimating the average Δ% values by ~−3 and ~−2 % for the monomeric and the full data sets, respectively (ranges ~−8 to +3 % for HYDROPRO and ~−9 to +6 % for BEST).

Finally, we would like to briefly report on the use of NMR structures instead of X-ray structures when performing hydrodynamic computations. We have previously performed such tests in our SOMO and US-SOMO articles (Brookes et al. 2010b; Rai et al. 2005). The US-SOMO implementation renders possible the automatic processing of all models included in an NMR structure, and the averaging of the computed hydrodynamic parameters. This capability is absent from both HYDROPRO and BEST, and therefore we have elected not to conduct the same extensive comparison on the restricted set of NMR structures available for the proteins we have examined, namely ribonuclease A (2AAS), lysozyme (1E8L) and myoglobin (1MYF). Nevertheless, we have performed the comparisons for the SoMo and AtoB methods, confirming what was already evident in our previous studies: there is practically no difference in the $D^0_{t\,(20,w)}$ and $s^0_{(20,w)}$ values computed for the X-ray and NMR structures of ribonuclease A, a very small difference for myoglobin, but a relatively large difference for lysozyme, with the NMR structure-derived values significantly closer to the experimental values than their X-ray counterparts (data not shown). These findings confirm what we have already reported and interpreted as an effect of the many extended chains on the surface of lysozyme compared with the other proteins (Brookes et al. 2010b; Rai et al. 2005).

Discussion

The results of our comparative analysis can now be discussed. An overestimation of the $D^0_{t\,(20,w)}$ values with respect to experimental data indicates that the model employed is somewhat smaller than it should be. Therefore, we interpret the performance of the SoMo models with the overlap removal as due to a reduction of the surface of the "hydrated" protein compared with the one that would be obtained from the original crystal structure, notwithstanding the outward translation procedure that was devised to avoid this problem (Rai et al. 2005). In fact, when we employ the SoMo models without the overlap reduction, coupled with the Zeno computational method, the best performance of all the methods examined is obtained. Importantly, since the volume of the beads in the SoMo models is derived from that of the constituent residues plus that of the theoretically "bound" water of hydration (Brookes et al. 2010b; Kuntz and Kauzmann 1974; Rai et al. 2005), this result suggests that our local hydration scheme is very effective. This is also confirmed by the excellent performance of the AtoB models, where the same hydration scheme is applied. Evidently, in the AtoB models the larger number of beads utilised allows a better preservation of the overall "hydrated" protein surface when the overlaps are removed, which is still performed with the outward translation for the exposed beads.

The results generated by HYDROPRO and BEST, which underestimate the $D^0_{t\,(20,w)}$ values slightly more than the SoMo models with overlap removal overestimate them, indicate that both the shell-bead and the triangulated models are larger than they should be to correctly represent the "hydrated" protein surface. Evidently, the atomic element radius utilised by HYDROPRO and the atom radii utilised by BEST have both been overestimated. We recall that these values were not derived from a detailed analysis of the average hydrated surface of proteins, but adjusted so as to match, on average, the hydrodynamics of a selected group of proteins, moreover without the careful evaluation of the available experimental data that we have instead conducted. Furthermore, the approach used in both HYDROPRO and BEST treats the hydration as a uniform layer, while the results of the SoMo models with overlaps discussed above seem to confirm that preferentially hydrating the polar/charged residues produces better results. It is possible that reduction of the atomic element radius in HYDROPRO and the atom radii in BEST, and perhaps accounting for preferential hydration, could bring the performance of these two methods on a par with that of the SoMo with overlaps/Zeno combination or the AtoB methods.

The other important observation that stems from our extensive comparisons is the relative inadequacy of the computed $s^0_{(20,w)}$ in reproducing the experimental data. This is particularly disappointing, given that $s^0_{(20,w)}$ is more easily determined (and usually with better statistical error, see Table 2; Fig. 1, panel b) than the corresponding $D^0_{t\,(20,w)}$ value. Apart from trivial experimental errors such as the "time bug" in AUC data analysis reported not long ago (Zhao et al. 2013), a likely source of the discrepancy is the $\bar{v}_{(20,w)}$ value utilised in converting $D^0_{t\,(20,w)}$ to $s^0_{(20,w)}$ (see Eq. 3). Although a great deal of effort has been invested in finding an accurate way of computing $\bar{v}_{(20,w)}$ from (bio) macromolecular composition [e.g. Durchschlag and Zipper 2005, 2008)], the results are still not always reliable. In particular, the contribution of the solvent composition and pH is sometimes not properly taken into account, nor is the interplay between the various types of residues. Experimental values, which in principle should always be sought and employed, are not easily determined either, since they require milligram quantities of material when using a densimetric approach, or the availability of water isotopes and careful experimental design if the differential sedimentation approach is chosen. In any case, an analysis of the data reported in Table 2 for the best performing method (SoMo

models without overlap removal coupled with the Zeno calculations, outliers excluded) revealed only a modest difference between the average $\Delta\%$ values for the $s^0_{(20,w)}$ data calculated when experimental $\bar{v}_{(20,w)}$ values were available (-1.0 ± 3.6; $n = 8$) versus computed $\bar{v}_{(20,w)}$ values ($+2.1 \pm 5.2$; $n = 17$), leaving this issue still open for further investigation.

Regarding the difference between NMR and X-ray crystallography derived structures, we confirm what we have previously reported (Brookes et al. 2010b; Rai et al. 2005). NMR-derived structures appear to be better suited to model translational friction, especially when the macromolecule is small and has many long, potentially flexible surface side-chains. However, this effect should vanish as the protein size increases.

Conclusions

Some general conclusions, and some guidelines for the use of the various hydrodynamic modelling approaches examined, can now be offered. First of all, it seems evident that, whenever possible, the experimental determination of $D^0_{t(20,w)}$ should be sought in preference to (or in conjunction with) $s^0_{(20,w)}$. Dynamic light scattering (DLS) is now a relatively affordable technique for the measurement of $D^0_{t(20,w)}$, provided that a size-exclusion chromatography step is performed to obtain a monodispersed solution of the (bio)macromolecule of interest, and that a concentration series is also performed. A potential source of error can be encountered when dealing with elongated, large particles, in which case determining the angular dependence of the DLS signals becomes mandatory. But this requires either much more sophisticated, costly equipment, or modification of simpler, multi-angle instruments relying on a single-angle fibre optic to collect the DLS signal. On the other hand, analytical ultracentrifugation (AUC) can directly provide $D^0_{t(20,w)}$ from boundary spreading analysis in a standard sedimentation velocity experiment. But the data processing required for this appears to be more error prone than when dealing with the simpler determination of $s^0_{(20,w)}$. Synthetic boundary experiments, seldom performed, could be a viable alternative, but, again, analysis of the data is less straightforward.

Regarding the hydrodynamic modelling methods, whose main characteristics and basic performances are outlined in Table 3, we would like to offer the following guidelines: The least demanding, most rapid method for calculating $D^0_{t(20,w)}$ and $s^0_{(20,w)}$ is HYDROPRO, which however suffers from two problems: using the current atomic element radius, it is the least accurate, and its reliability declines as the size of the (bio)macromolecule increases. Furthermore, any problems with the extrapolation to zero shell bead size are not immediately obvious to the user, since a graphical representation of the extrapolation is not provided. Although users could manually plot such graphs from the results generated by HYDROPRO, this will offset one of its advantages, namely the speed with which it will produce the final extrapolated results, which is practically independent of protein size in the default mode (see Table 3). Furthermore, HYDROPRO does not perform a check of the structure submitted, and while this avoids some of the drawbacks of the US-SOMO methods (see below), it could hide potential problems such as missing atoms/residues leading to avoidable errors in the hydrodynamic computation. Therefore, HYDROPRO in its current implementation is to be recommended only for a quick initial evaluation of the translational frictional properties of a (bio) macromolecule.

The other extrapolation-based method tested, BEST, appears to generate values of $D^0_{t(20,w)}$ and $s^0_{(20,w)}$ that suffer from an underestimation problem of the same magnitude as HYDROPRO. While attempting to adjust the radii values to deliver a better match with the experimental $D^0_{t(20,w)}$ and $s^0_{(20,w)}$ values was outside the scope of this work, it is completely conceivable that it could be achieved in this way. However, we feel that the computational accuracy at the core of the BEST calculations is not really required when dealing with translational frictional properties, and comes at the cost of a large increase in computing time, very much dependent on the computing architecture available, on protein size, and on the choice of the minimum and maximum number of plates used to produce the triangulated models (see Table 3; the apparently odd time decrease with increasing size with fixed min–max plate numbers is due to repeated failures in the MSROLL-coalescing steps when generating very finely triangulated surfaces). Moreover, the extrapolations should always be checked, a task which is however facilitated by the BEST implementation under US-SOMO. In addition, like HYDROPRO, BEST does not check the input structure for missing atoms etc., which while being less demanding on the operator, requires that such checks be conducted prior to submission to ensure that meaningful results are obtained. Therefore, provided that atom radii that correct the observed underestimation of translational frictional properties can be determined, BEST seems better suited for very specialised applications that really require its level of internal accuracy, such as when the computation of $\tau^0_{c(20,w)}$ and $[\eta]$ are sought.

The SoMo and AtoB methods offered inside the US-SOMO suite are more demanding, requiring the coding of atoms/residues in a series of lookup tables for best performance (Brookes et al. 2010b). However, these lookup tables already encode most of the atoms and residues commonly found in structures downloaded from the PDB [see e.g. supplementary material in (Brookes et al. 2010b)],

Table 3 Characteristics and performances of the main hydrodynamic modelling/computational methods discussed in this work

Program (conditions)	Structure check?	NMR?[a]	Modelling method	Computational method	$D^0_{t(20,w)}$ average Δ%[b]	$s^0_{(20,w)}$ average Δ%[c]	Computing time (minutes) for selected structures[d]		
							1AKI (14 kDa)	1AO6 (66 kDa)	1ADO (160 kDa)
SoMo (with overlaps) in US-SOMO	Yes[e]	Yes	BM, residue to bead	Zeno	−0.2 ± 2.4	−0.6 ± 3.7	0.5	7.8	20.3
AtoB (5 Å grid) in US-SOMO	Yes[e]	Yes	BM, grid	SMI	+0.3 ± 2.4	+1.4 ± 3.5	0.03	1	9
SoMo (no overlaps) in US-SOMO	Yes[e]	Yes	BM, residue to bead	SMI	+1.9 ± 2.5	+2.9 ± 3.5	0.02	0.2	0.5
BEST (manual) in US-SOMO	Yes[f]	No	BE (2000–6000 plates)	SI	−2.7 ± 2.1	−1.9 ± 3.8	356	100	93
BEST (heuristic) in US-SOMO	Yes[f]	No	BE (variable # of plates)	SI	−2.8 ± 2.2	−1.9 ± 3.7	170	261	1072
HYDROPRO (WinHydropro)	No	No	Shell BM (≤2000 beads)	SMI	−3.6 ± 3.0	−2.4 ± 3.2	0.3	0.3	0.3

BM bead modelling, *BE* boundary elements, *SMI* supermatrix inversion, *SI* surface integrals

[a] Automatic computation and averaging of hydrodynamic parameters possible for multiple structures in NMR-type files

[b] For all test proteins listed in Table 1, outliers excluded, (±SD)

[c] For all test proteins listed in Table 2, outliers excluded, (±SD)

[d] For the Zeno and SMI methods within US-SOMO (Windows version) and HYDROPRO (WinHydropro), computations were run on an Intel Core i5-3470 3.2 GHz PC with 6 GB RAM, operating under the Windows 7 Professional OS; for BEST within US-SOMO, they were run on the TACC Stampede cluster, and do not include waiting times in the queue (see "Materials and methods")

[e] Approximate methods available for non-coded or incomplete residues

[f] Checks performed but no influence on program execution

and approximate methods are offered to avoid this coding step when the non-coded parts are a small portion of the structure (Brookes et al. 2010a). This usually affects more the $s^0_{(20,w)}$ values, because of the $\bar{v}_{(20,w)}$ contribution, than $D^0_{t(20,w)}$. Nevertheless, the SoMo and AtoB methods offer several advantages. Although the SoMo method with overlap removal slightly overestimates $D^0_{t(20,w)}$ and $s^0_{(20,w)}$, the direct correspondence between residues and beads makes it suitable for studies addressing, for instance, complex formation or flexibility issues. The new combination of SoMo models without overlap removal and Zeno hydrodynamic computations proposed here [directly available with the next US-SOMO release, see Brookes and Rocco (2015)] appears to be able to provide the most accurate computation of $D^0_{t(20,w)}$ and $s^0_{(20,w)}$, but the calculations with the current implementation of Zeno are still significantly slower than with the supermatrix inversion procedure (see Table 3). This should change in the near future, with the arrival of a new, faster Zeno code (J. Douglas, NIST, personal communication), and could make this combination the method of choice for standard hydrodynamic computations of translational friction properties. In the meantime, the AtoB method with a 5 Å grid size appears to combine very good accuracy with reasonable computing times, even for the largest structures examined, making it the current method of choice. Moreover, the possibility of changing the grid size in AtoB facilitates the study of very large structures that would pose computational problems for all other methods examined.

In conclusion, owing to speed, ease of operation and relative accuracy, the determination of the translational frictional properties of (bio)macromolecules followed by hydrodynamic modelling has become a complementary, standard tool in multi-resolution modelling. With our detailed examination of the methods currently available, we hope that we have provided a clear view of their advantages and limitations. While further improvements in the software are to be expected, what is sorely missing, given the vintage of most literature data (dating back to 1927!), is a new, robust set of experimental data for a series of test proteins to be used as "standard candles" in the refinement of hydrodynamic modelling methods.

Acknowledgments M.R. designed and performed research, and wrote the paper. O.B. contributed analytical tools and wrote the paper. We are indebted to Emre Brookes (University of Texas at San Antonio, TX, USA) for his continuous and timely efforts in improving the US-SOMO software. We thank Camillo Rosano (IRCCS AOU San Martino-IST, Genova, Italy) for completing several protein structures.

M.R. was partially supported by the Italian Ministry of Health, "5 per mille 2011" funds. The authors declare no conflict of interest.

References

Aragon S (2004) A precise boundary element method for macromolecular transport properties. J Comp Chem 25:1191–1205

Aragon SR (2011) Recent advances in macromolecular hydrodynamic modeling. Methods 54:101–114

Armstrong JM, Myers DV, Verpoorte JA, Edsall JT (1966) Purification and properties of human erythrocyte carbonic anhydrases. J Biol Chem 241:5137–5149

Berman HM et al (2000) The protein data bank. Nucleic Acids Res 28:235–242

Bloomfield V, Dalton WO, van Holde KE (1967a) Frictional coefficients of multi-subunit structures. I. Theory. Biopolymers 5:135–148

Bloomfield V, van Holde KE, Dalton WO (1967b) Frictional coefficients of multi-subunit structures. II. Application to proteins and viruses. Biopolymers 5:149–159

Broersma S (1960) Rotational diffusion coefficient of a cylindrical particle. J Chem Phys 32:1626–1631

Brookes E, Rocco M (2015) Calculation of hydrodynamic parameters—US-SOMO. In: Uchiyama S, Arisaka F, Stafford WF III, Laue TM (eds) Analytical ultracentrifugation: instrumentation, software and application, Chapter 10. Springer, New York **In press**

Brookes E, Demeler B, Rocco M (2010a) 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. Macromol Biosci 10:746–753

Brookes E, Demeler B, Rosano C, Rocco M (2010b) 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. Eur Biophys J 39:423–435

Brookes E, Perez J, Cardinali B, Profumo A, Vachette P, Rocco M (2013) Fibrinogen species as resolved by HPLC-SAXS data processing within the UltraScan Solution Modeler (US-SOMO) enhanced SAS module. J Appl Crystallogr 46:1823–1833

Burgers JM (1938) On the motion of small particles of elongated form, suspended in a viscous liquid. In: Burgers JM, Jaeger FM, Houwink R, Van Nieuwenberg CJ, Saal RNJ (eds) Second report on viscosity and plasticity. Nordemann, Amsterdam, p 209

Byron O (1997) Construction of hydrodynamic bead models from high-resolution X-ray crystallographic or nuclear magnetic resonance data. Biophys J 72:408–415

Cantor CR, Schimmel PR (1980) Biophysical Chemistry, vol 2. W.H. Freeman, San Francisco

Charlwood PA (1952) Sedimentation and diffusion of human albumins. 1. Normal human albumins at a low concentration. Biochem J 51:113–118

Christen P, Göschke H, Leuthard F, Schmid A (1965) Über die Aldolase der Kaninchenleber: Molekulargewicht, Dissoziation in Untereinheiten. Über Aldolasen, 5. Mitteilung. Helv Chim Acta 48:1050–1056

Clark SM, Leaist DG, Konermann L (2002) Taylor dispersion monitored by electrospray mass spectrometry: a novel approach for studying diffusion in solution. Rapid Commun Mass Spec 16:1454–1462

Coleman JE (1965) Human carbonic anhydrase. Protein conformation and metal ion binding. Biochemistry 4:2644–2655

Connolly ML (1993) The molecular surface package. J Mol Graphics 11:139–141

Cunningham LW (1954) Molecular-kinetic properties of crystalline diisopropyl phosphoryl trypsin. J Biol Chem 211:13–19

Dean RB, Dixon WJ (1951) Simplified statistics for small numbers of observations. Anal Chem 23:636–638

Denisov VP, Halle B (1995) Protein hydration dynamics in aqueous-solution—a comparison of bovine pancreatic trypsin-inhibitor and ubiquitin by O-17 spin relaxation dispersion. J Mol Biol 245:682–697

Ding F, Dokholyan NV (2006) Emergence of protein fold families through rational design. PLoS Comput Biol 2:e85

Dokholyan NV, Buldyrev SV, Stanley HE, Shakhnovich EI (1998) Discrete molecular dynamics studies of the folding of a protein-like model. Fold Des 3:577–587

Durchschlag H, Zipper P (2005) Calculation of volume, surface, and hydration properties of biopolymers. In: Scott DJ, Harding SE, Rowe AJ (eds) Analytical Ultracentrifugation: Techniques and Methods. Royal Society of Chemistry, Cambridge, pp 389–431

Durchschlag H, Zipper P (2008) Volume, surface and hydration properties of proteins. Prog Colloid Polym Sci 134:19–29

Edelhoch H (1957) The denaturation of pepsin. I. Macromolecular changes. J Am Chem Soc 79:6100–6109

García de la Torre J, Bloomfield VA (1977a) Hydrodynamic properties of macromolecular complexes. I. Translation. Biopolymers 16:1747–1763

García de la Torre J, Bloomfield VA (1977b) Hydrodynamics of macromolecular complexes. II. Rotation. Biopolymers 16:1765–1778

García de la Torre J, Bloomfield VA (1977c) Hydrodynamics of macromolecular complexes. III. Bacterial viruses. Biopolymers 16:1779–1793

García de la Torre J, Bloomfield VA (1978) Hydrodynamic properties of macromolecular complexes. IV. Intrinsic viscosity theory with applications to once-broken rods and multisubunit proteins. Biopolymers 17:1605–1627

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, Navarro S, Lopez Martinez MC, Díaz FG, Lopez Cascales JJ (1994) HYDRO: a computer program for the prediction of hydrodynamic properties of macromolecules. Biophys J 67:530–531

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

Ghirlando R (2011) The analysis of macromolecular interactions by sedimentation equilibrium. Methods 54:145–156

Glikina MV, Finogenov PA (1950) Issledovanie myshechnoi aldolazy na razlichnykh stadiyakh izolirovaniya (Investigation of muscular aldolase in various stages of isolation). Biokhimiya (Moscow) 15:457–464

Goldstein RF (1985) Macromolecular diffusion constants: a calculational strategy. J Chem Phys 83:2390–2397

Hahn DK, Aragon SR (2006) Intrinsic viscosity of proteins and platonic solids by boundary element methods. J Chem Theor Comp 2:1416–1428

Halle B, Davidovic M (2003) Biomolecular hydration: from water dynamics to hydrodynamics. Proc Natl Acad Sci USA 100:12135–12140

Hanlon AD, Larkin MI, Reddick RM (2010) Free-solution, label-free protein-protein interactions characterized by dynamic light scattering. Biophys J 98:297–304

Harvey SC (1979) Transport properties of particles with segmental flexibility. I. Hydrodynamic resistance and diffusion coefficients of a freely hinged particle. Biopolymers 18:1081–1104

Harvey SC, Cheung HC (1980) Transport properties of particles with segmental flexibility. II. Decay of fluorescence polarization anisotropy from hinged macromolecules. Biopolymers 19:913–930

Jones TA, Zou J-Y, Cowan SW, Kjeldgaard M (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. Acta Cryst A 47:110–119

Kang EH, Mansfield ML, Douglas JF (2004) Numerical path integration technique for the calculation of transport properties of proteins. Phys Rev E Stat Nonlinear Soft Matter Phys 69:031918

Kawahara K (1969) Evaluation of diffusion coefficients of proteins from sedimentation boundary curves. Biochemistry 8:2551–2775

Kirkwood JG (1949) The statistical mechanical theory of irreversible processes in solutions of macromolecules (visco-elastic behaviour). Recl Trav Chim Pays Bas 68:649–660

Kirkwood JG (1954) The general theory of irreversible processes in solutions of macromolecules. J Polym Sci 12:1–14

Kunitz M, Northrop JH (1935) Crystalline chymo-trypsin and chymotrypsinogen I. Isolation, crystallization, and general properties of a new proteolytic enzyme and its precursor. J Gen Physiol 18:433–458

Kuntz ID, Kauzmann W (1974) Hydration of proteins and polypeptides. In: Anfinsen CB, Edsall JT, Richards FM (eds) Advances in Protein Chemistry, vol 28. Academic, Waltham, pp 239–345

Mansfield ML, Douglas JF (2008) Improved path integration method for estimating the intrinsic viscosity of arbitrarily shaped particles. Phys Rev E Stat Nonlinear Soft Matter Phys 78:046712

McVittie JD, Esnouf MP, Peacocke AR (1977) Denaturation-renaturation of chicken-muscle triosephosphate isomerase in guanidinium chloride. Eur J Biochem 81:307–315

Neurath H, Cooper GR, Erickson JO (1941) The shape of protein molecules II. Viscosity and diffusion studies of native proteins. J Biol Chem 138:411–436

Nöllmann M, He J, Byron O, Stark W (2004) Solution structure of the Tn3 resolvase-crossover site synaptic complex. Mol Cell 16:127–137

Nyman O (1961) Purification and properties of carbonic anhydrase from human erythrocytes. Biochim Biophys Acta 52:1–12

Ortega A, Amorós D, García de la Torre J (2011) Prediction of hydrodynamic and other solution properties of rigid proteins from atomic- and residue-level models. Biophys J 101:892–898

Oseen CW (1927) Neure Methoden und Ergebnisse in der Hydrodynamik. In: Hilb E (ed) Mathematik und ihre Anwendungen in Monographien und Lehrbüchern. Academisches Verlagsgellschaft, Leipzig

Perrin J (1936) Mouvement Brownien d'un ellipsoide. II. Rotation libre et depolarisation des fluorescences. Translation et diffusion de molecules ellipsoidales. J Phys Radium 7:1–11

Rackis JJ, Anderson RL, Mann RK, Sasame HA, Smith AK (1962) Soybean trypsin inhibitors: isolation, purification and physical properties. Arch Biochem Biophys 98:471–478

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

Rickli EE, Edsall JT, Ghazanfar ASS, Gibbons BH (1964) Carbonic anhydrases from human erythrocytes: preparation and properties of two enzymes. J Biol Chem 239:1065–1078

Rocco M, Brookes E (2014) Dynamical aspects of biomacromolecular multi-resolution modelling using the ultrascan solution modeler (US-SOMO) suite. In: Howard JAK, Sparkes HA, Raithby PR, Churakov AV (eds) The Future of Dynamical Structural Science, NATO Science for Peace and Security series A: Chemistry and Biology, Chapter 13. Springer, Berlin, pp 189–199. doi:10.1007/978-94-017-8550-1_13

Rocco M, Byron O (2015) Hydrodynamic modeling and its application in AUC. In: Cole J (ed) Analytical ultracentrifugation. Methods in Enzymology, Chapter 8, vol 562. Elsevier, Amsterdam **in press**

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 multiresolution approach. Structure 16:954–964

Rotne J, Prager S (1969) Variational treatment of hydrodynamic interaction in polymers. J Chem Phys 50:4831–4837

Scheraga HA, Mandelkern L (1953) Consideration of the hydrodynamic properties of proteins. J Am Chem Soc 75:179–184

Schwert GW (1949) The molecular size and shape of the pancreatic proteases. 1. Sedimentation studies on chymotrypsinogen and on α- and γ-chymotrypsin. J Biol Chem 179:655–664

Schwert GW, Kaufman S (1951) The molecular size and shape of the pancreatic proteases. III. α-chymotrypsin. J Biol Chem 190:807–816

Simha R (1940) The influence of Brownian movement on the viscosity of solutions. J Phys Chem 44:25–34

Stellwagen E (1968) Reversible unfolding of horse heart ferricytochrome c. Biochemistry 7:2893–2898

Sund H, Weber K (1963) Untersuchungen über Milchzuckerspaltende Enzyme. XIII. Grosse und gestalt der Beta-Galaktosidase aus *E. coli* (Studies on the lactose-splitting enzyme. XIII. Quantity and configuration of beta-galactosidase from *E. coli*). Biochemische Zeitschrift 337:24–34

Taylor JF, Green AA, Cori GT (1948) Crystalline aldolase. J Biol Chem 173:591–604

Teller DC, Swanson E, de Haën C (1979) The translational frictional coefficients of proteins. Methods Enzymol 61:103–124

Tietze F (1953) Molecular-kinetic properties of crystalline trypsinogen. J Biol Chem 204:1–11

Venable RM, Pastor RW (1988) Frictional models for stochastic simulations of proteins. Biopolymers 27:1001–1014

Vriend G (1990) WHAT IF: a molecular modeling and drug design program. J Mol Graphics 8:52–56

Walters RR, Graham JF, Moore RM, Anderson DJ (1984) Protein diffusion coefficient measurements by laminar-flow analysis: method and applications. Anal Biochem 140:190–195

Wegener WA (1982) Bead models of segmentally flexible macromolecules. J Chem Phys 76:6425–6430

Wegener WA, Dowben RM, Koester VJ (1980) Diffusion coefficients for segmentally flexible macromolecules: general formalism and application to rotational behaviour of a body with two segments. J Chem Phys 73:4086–4097

Wilkins DK, Grimshaw SB, Receveur V, Dobson CM, Jones JA, Smith LJ (1999) Hydrodynamic radii of native and denatured proteins measured by pulse field gradient NMR techniques. Biochemistry 38:16424–16431

Wong SCK, Hall DC, Josse J (1970) Constitutive inorganic pyrophosphatase of *Escherichia coli*. III. Molecular weight and physical properties of enzyme and its subunits. J Biol Chem 245:4335–4345

Wood E, Dalgleish D, Bannister W (1971) Bovine erythrocyte cuprozinc protein. 2. Physicochemical properties and circular dichroism. Eur J Biochem 18:187–193

Yamakawa H (1970) Transport properties of polymer chains in dilute solution: hydrodynamic interaction. J Chem Phys 53:435–443

Zhao H, Ghirlando R, Piszczek G, Curth U, Brautigam CA, Schuck P (2013) Recorded scan times can limit the accuracy of sedimentation coefficients in analytical ultracentrifugation. Anal Biochem 437:104–108