Modeling Complex Biological Macromolecules: Reduction of Multibead Models

Peter Zipper · Helmut Durchschlag

Received: 3 November 2007 / Accepted: 6 March 2008 /
Published online: 9 May 2008
© Springer Science + Business Media B.V. 2008

Abstract The shape of simple and complex biological macromolecules can be approximated by bead modeling procedures. Such approaches are required, for example, for the analysis of the scattering and hydrodynamic behavior of the models under analysis and the prediction of their molecular properties. Using the atomic coordinates of proteins for modeling inevitably leads to models composed of a multitude of beads. In particular, for hydrodynamic modeling, a drastic reduction of the bead number may become unavoidable to enable computation. A systematic investigation of different approaches and computation modes shows that the ‘running mean’, ‘cubic grid,’ and ‘hexagonal grid’ approaches are successful, provided that the extent of reduction does not exceed a factor of 100 and the grid approaches use beads of unequal size and the beads are located at the centers of gravity. Further precautions to be taken include usage of appropriate interaction tensors for overlapping beads of unequal size and appropriate volume corrections when calculating intrinsic viscosities. The applied procedures were tested with the small protein lysozyme in a case study and were then applied to the huge capsid of the phage ϕ and its trimeric building block. The appearance of the models and the agreement of molecular properties and distance distribution functions of unreduced and reduced models can be used as evaluation criteria.

Keywords Complex proteins · Multibead models · Bead reduction · Small-angle scattering · Hydrodynamics

1 Introduction

Over the years, the evaluation and interpretation of solution scattering and hydrodynamic data have been in permanent flux. While in the past, modeling was restricted to simple

P. Zipper
Physical Chemistry, Institute of Chemistry, University of Graz, Heinrichstrasse 28, 8010 Graz, Austria

H. Durchschlag (✉)
Institute of Biophysics and Physical Biochemistry, University of Regensburg, Universitätsstrasse 31,
93040 Regensburg, Germany
e-mail: helmut.durchschlag@biologie.uni-regensburg.de

whole-body approaches, later on, multibody approaches were in the spotlight. These procedures enabled the characterization of simple proteins or more complex proteins built up from several subunits, preferably in terms of several molecular parameters (e.g., [1–4]). More recently, advanced ab initio modeling approaches enabled the automated evaluation of data (e.g., [5, 6]). Nowadays, comprehensive high-resolution crystallographic or NMR data (atomic coordinates) are available for various kinds of biomolecules including thousands of proteins and protein complexes. Undoubtedly, this poses a challenge to all modelers because of several serious problems involved.

The scattering and hydrodynamic properties of rigid or flexible biological macromolecules of different nature, size, and shape can be calculated by a bead modeling methodology (e.g., by Debye modeling [1] or García de la Torre's approach [7, 8]). The particle under analysis has to be represented by an assembly of spherical elements yielding an overall size and shape resembling the target molecule as much as possible. This may be achieved either by a filling model strategy (using all beads, equal- or unequal-sized) or a shell modeling approach. The latter approach is hydrodynamically more rigorous and avoids problems like overlapping of beads. However, filling models are required for the concurrent prediction of scattering and hydrodynamic properties because the usage of all constitutive elements is a necessary prerequisite for handling scattering problems.

In this context, three major problems arise, in particular with hydrodynamic modeling: (1) the interaction of overlapping beads of unequal size, e.g., when bead models are exact copies of crystal structures, (2) the multitude of beads when applying huge numbers of coordinates or groups, and (3) the calculation of viscometric quantities. These problems require the application/development of rather sophisticated approaches: application of a special interaction tensor, usage of effective reduction steps, and consideration of corrections for the calculation of intrinsic viscosities. Far-reaching reduction procedures are essential, for example, for constructing bead models of huge proteins for hydrodynamic modeling from their atomic-level data.

The efficiency of various approaches addressed was tested with proteins of different size and symmetry, e.g., applying the small enzyme lysozyme as a test case and a bacteriophage capsid as a representative of a giant, highly complex molecule. These two types of proteins have also been chosen because, in both cases, comprehensive experimental data stemming from solution scattering (small-angle X-ray scattering [SAXS]), hydrodynamics (analytical ultracentrifugation, viscometry), and X-ray crystallography are available; several modeling attempts based on scattering or crystal data, including advanced ab initio modeling approaches, have also been undertaken. This allows predicted and observed scattering and hydrodynamic properties characterizing the size and shape of the molecules to be compared, in addition to a critical assessment of the diverse reduction steps to be applied.

While in the past, the comparison of calculated and experimental data of quite different proteins has established the universal validity of the approaches applied (e.g., [6, 9–18]), a systematic investigation of the reduction approaches including inherent limitations is missing to date. To test the accuracy of the reduction procedures applied, only the structures of models have to be considered to avoid being influenced by a serendipitous coincidence of predicted and observed quantities.

2 Methods

The atomic coordinates and the masses of lysozyme from hen egg white and the capsid of the *Escherichia coli* phage ϕ were obtained from the Protein Data Bank (PDB) [19]:

accession codes 6LYZ for lysozyme and 1FRS for the fr capsid; masses $M=14.313$ kg/mol for lysozyme (1 subunit consisting of 129 amino acid [AA] residues) and $M=2,472$ kg/mol for the fr capsid (180 subunits consisting of 129 AA residues each; subunits A, B, and C form a trimeric building block). Experimental SAXS and hydrodynamic parameters have been summarized recently ([16, 20] and references therein).

The programs PDB2AT and PDB2AM have been developed by us to transform the original crystal data to various kinds of bead models: unreduced ‘initial models’ based on atoms or atomic groups (IM-AT), ‘initial models’ based on AA residues (IM-AM; in these models, the number of beads is already reduced by about 85%), and reduced bead models at arbitrarily chosen reduction levels. Predecessors of the programs applied have been used and described previously [9, 10], particularly in the context of the first attempts to model multibead clusters (coarse graining of oligomeric protein models) for hydrodynamic purposes. Indeed, the development of reduction approaches is a necessary requirement for hydrodynamic modeling of structures on the atomic level; currently, calculations with standard personal computers are restricted to models of bead numbers up to about 10,000 [6]. In contrast to hydrodynamic modeling, usage of reduction processes is not a decisive factor for SAXS modeling (e.g., Debye modeling) because the limit of 10,000 beads does not apply to this case.

In the aforementioned programs, the crystal data are read from the PDB file line by line and are simultaneously interpreted and processed: the spatial coordinates of each atom or atomic group are stored; appropriate values of volume, mass, and weight are assigned to each atom or atomic group and these data are also stored. Running mean (RM) algorithms control the further processing, though with slight differences depending on the program used [9, 10]. In PDB2AT, atoms or atomic groups in sequential order are merged into a single bead until their total volume equals or just exceeds that of a bead with arbitrarily chosen bead diameter (BD). Calculation of volumes are based on Traube’s additivity principle, summing up volume increments for the basic atoms [15, 18, 21]. The resulting bead, having a volume, mass, and weight identical to the summed up contributions of the merged atoms or groups, is placed at the local center of gravity of the ensemble of atoms or groups it replaces; usually the radius r_b of this bead is slightly larger than $BD/2$. With the setting $BD=0$, no merging occurs and an unreduced initial model (IM-AT) is generated. In PDB2AM, at least all atoms or atomic groups belonging to the same AA residue are merged, leading to a reduced initial model (IM-AM) representing the spatial arrangement of the AA residues. This occurs if the compression index ICOM (the number of AA residues to be merged) equals 1, whereas with settings of $ICOM \geq 2$ exactly ICOM AA residues in sequential order are merged into a single bead. The weights mentioned above can be either mass, volume, or the number of electrons; these weights are taken into account accordingly when centers of gravity or radii of gyration are computed.

The coordinate files generated by PDB2AT or PDB2AM can be used directly for any calculations to be performed with the models. On the other hand, they may also be used as input for an alternative or complementary reduction program, MAP2GRID. This in-house program maps a given structure into a three-dimensional cubic or hexagonal grid (CG, HG) of arbitrarily chosen lattice parameter, the voxel size (VS) [9–12]. All structural elements falling into the same lattice cell are merged into a single bead which may be placed either at the local center of gravity or on an exact lattice point. While usually the beads obtained by merging are of unequal size, a special option forces the program to generate reduced models consisting of equal-sized beads on lattice points.

All models obtained were visualized by the program RASMOL [22]. For the detailed analyses and comparisons of the models, essentially the following information was of

interest: pair-distance distribution function, $p(r)$; molar mass, M ; volume, V ; maximum particle diameter, $d_{\max}$; radius of gyration, R_G ; translational diffusion coefficient, D ; sedimentation coefficient, s ; and intrinsic viscosity, $[\eta]$. In context with the present study (persistence of structural features at different reduction levels), calculation of scattering intensities, $I(h)$, the Fourier transforms of $p(r)$, is of subordinate importance. As functions in reciprocal space, intensities do not reflect any structural changes in such a direct way as the real-space functions $p(r)$.

Structural parameters ($d_{\max}$, R_G) and $p(r)$ profiles were calculated from the coordinates, radii, and if applicable, weights of the beads, essentially using algorithms described previously [1, 23]. Hydrodynamic parameters (s , D , $[\eta]$) were obtained by modified versions of García de la Torre's program HYDRO [7, 8], applying also the most recent version HYDRO++ [24]. Further details concerning the prediction of structural and hydrodynamic parameters can be found in previous papers [6, 17].

Hydrodynamic modeling requires several problems and pitfalls to be addressed explicitly. Recent versions of HYDRO and obviously also HYDRO++ make use of our ad hoc expression for the interaction tensor of overlapping nonequal spheres [9, 25, 26]; for numerical reasons, however, this expression is not applicable if third-order or higher-order hydrodynamic interactions are considered. In this study, s , D , and $[\eta]$ were calculated by means of a special, modified version of HYDRO that enabled us to process even huge bead models composed of up to about 10,000 beads [6] and to compute $[\eta]$ in different ways with respect to the so-called volume correction (cf. Refs. [24, 27]):

$$[\eta]_{\text{corr}} = [\eta]_{\text{uncorr}} + f_{\eta} \frac{2.5N_A V}{M}$$

where N_A is Avogadro's number, V is the total amount of bead volume, and M is the total amount of bead mass. In our version of HYDRO, the factor f_{η} successively adopts the values 0, 1, and $1/N_b^{1/3}$ where N_b represents the number of beads. The case of $f_{\eta} = 0$ corresponds to the absence of any volume correction. The uncorrected value, $[\eta]_{\text{uncorr}}$, usually underestimates the intrinsic viscosity of the model. The opposite is true for the value of $[\eta]_{\text{corr}}$ that is obtained with $f_{\eta} = 1$ by applying the 'full volume correction' (FVC) [27] based on the total volume V of all N_b beads. Following the idea that a more satisfying value of $[\eta]_{\text{corr}}$ might be obtained if the correction is related to the surface S of the bead model and considering that S is proportional to $N_b^{2/3}$, we chose $f_{\eta} = 1/N_b^{1/3}$ as basis for a 'reduced volume correction' (RVC). For this correction, only the effective volume $V/N_b^{1/3}$, corresponding to the volume of $N_b^{2/3}$ beads, is taken into account.

The just recently available program HYDRO++ was used to calculate two further alternative values for the intrinsic viscosity: Another value of $[\eta]_{\text{corr}}$ is obtained by applying the new approach of an 'adjusted volume correction' (AVC) [24]; in this approach, f_{η} is calculated by taking into account explicitly the degree of fragmentation and the anisometry of each model under analysis, provided that the number of beads does not exceed 2,000. The second alternative value obtained from HYDRO++, $[\eta]_{\text{CBS}}$, is computed by application of the 'cubic substitution' (CBS) approach [24, 28, 29]; in this approach, each bead of radius r_b in a given model is replaced with a cubic array of eight beads having a radius of $0.4668 r_b$; the intrinsic viscosity is calculated from the resulting CBS model without any volume correction. HYDRO++ restricts the application of the CBS approach to models with $N_b \leq 250$. In this study, however, the CBS was also performed by means of an in-house program and $[\eta]_{\text{CBS}}$ values were calculated for models with $N_b \leq 1,250$ by usage of our modified version of HYDRO; in this case, of course, without any volume correction.

3 Results and Discussion

In the case of the small enzyme lysozyme (a monomeric protein of simple shape), various reduction procedures were used to test the applicability of different approaches to reduce the number of beads. In the following, these findings were applied to the huge capsid of the phage ϕ and its trimeric building block to demonstrate the efficiency of the reduction processes.

3.1 Lysozyme as a Test Case

Figure 1 and Table 1 summarize in condensed form various reduction approaches (RM, CG, and HG models) and computation modes. The initial models IM-AT (#00) and IM-AM (#04), based on atoms/atomic groups or AA residues, were generated by means of the programs PDB2AT and PDB2AM, respectively.

Application of the RM approaches (#01–06) in an extent reducing the initial number of beads (IM-AT: $N_b=1,001$) to about 100 (corresponding to a reduction factor of 10) leads to realistic models and yields values for structural parameters ($d_{\max}$, R_G) and hydrodynamic quantities (D , s , $[\eta]$) which are in agreement with the values for the unreduced model. This holds both for varying the BD or the compression index ICOM.

Use of the program MAP2GRID (#07–18) allows the application of CG and HG approaches to achieve models of reduced bead number. Thereby models of similar quality as obtained by the RM approaches are afforded, provided that the beads of unequal (U) size have been located at centers of gravity (G). Locating the beads on lattice (L) points yields less satisfactory results, regardless of the use of beads of unequal or equal size.

In the case of lysozyme, the applied approaches and programs (PDB2AT, PDB2AM, and MAP2GRID) are similarly effective for constructing appropriate bead models. Because lysozyme is of simple shape, of course, no severe differences for applying the RM, CG, or

Fig. 1 Bead models of lysozyme generated by the application of various reduction approaches and computation modes to the crystal structure (PDB entry: 6LYZ). Selected RM models are shown in the *first row*, CG models in the *second row*, and HG models in the *third row*; all models based on atomic groups are given on the *left hand side*, and models based on AA residues are displayed on the *right hand side*. For further details see Table 1

Table 1 Bead models of lysozyme generated by various reduction approaches (RM, CG, and HG) and computation modes together with the calculated radii of gyration and hydrodynamic parameters

#	Method	N_b	r_b (nm)	$d_{\max}$ (nm)	$MSD \times 10^6$ ^a	R_G (nm)	$R_{G,HYDRO}$ ^b (nm)	$D \times 10^7$ ^b (cm ² /s)	$s \times 10^{13}$ ^b (s)	$[\eta]$ (cm ³ /g)
RM: BD (nm) ^c										
00 (IM-AT)	0	1,001	0.036–0.196	5.005		1.3860	1.3860	12.32	2.157	2.50 ^d , 2.86 ^e
01	0.1	1,000	0.084–0.196	5.005	0	1.3860	1.3860	12.32	2.157	2.50 ^d , 2.86 ^e , 2.47 ^f
02	0.5	223	0.250–0.283	5.007	6.0	1.3801	1.3801	12.57	2.202	2.46 ^d , 2.72 ^e , 2.40 ^g
03	1.0	32	0.464–0.508	4.863	83.8	1.3597	1.3597	13.13	2.299	2.45 ^d , 2.53 ^e , 2.30 ^g
RM: ICOM ^h										
04 (IM-AM)	1	129	0.242–0.378	5.105	4.4	1.3881	1.3881	12.45	2.180	2.59 ^d , 2.80 ^e , 2.53 ^g
05	3	43	0.397–0.532	4.946	47.6	1.3691	1.3691	12.97	2.271	2.47 ^d , 2.60 ^e , 2.35 ^g
06	5	26	0.492–0.590	4.885	199	1.3542	1.3543	13.16	2.305	2.47 ^d , 2.57 ^e , 2.30 ^g
CG: VS=0.5 nm ⁱ										
07	G_U	195	0.1220–0.3798	5.060	3.7	1.3838	1.3838	12.34	2.161	2.60 ^d , 2.85 ^e , 2.53 ^g
08	L_U	195	0.1220–0.3798	5.448	174	1.4228	1.4228	11.54	2.021	3.11 ^d , 3.34 ^e , 3.03 ^g
09	L_E	195	0.2748	5.575	157	1.4194	1.5382	11.14	1.951	3.43 ^d , 3.63 ^e , 3.41 ^g
HG: VS=0.5 nm ⁱ										
10	G_U	223	0.1217–0.3675	5.092	5.0	1.3829	1.3829	12.34	2.162	2.59 ^d , 2.85 ^e , 2.52 ^g
11	L_U	223	0.1217–0.3675	5.420	134	1.4211	1.4211	11.44	2.004	3.17 ^d , 3.42 ^e , 3.09 ^g
12	L_E	223	0.2628	5.600	113	1.4183	1.5358	11.05	1.936	3.48 ^d , 3.70 ^e , 3.48 ^g
CG: VS=0.5 nm ^j										
13	G_U	110	0.2425–0.4539	5.105	4.3	1.3864	1.3864	12.48	2.186	2.59 ^d , 2.79 ^e , 2.52 ^g
14	L_U	110	0.2425–0.4539	5.067	133	1.4147	1.4145	12.02	2.106	2.83 ^d , 3.01 ^e , 2.75 ^g
15	L_E	110	0.3326	5.248	117	1.4118	1.4283	11.96	2.094	2.87 ^d , 3.05 ^e , 2.80 ^g

	HG; VS=0.5 nm ^j									
	G,U		114		0.2425-0.4535		5.105		4.7	
	L,U		114		0.2425-0.4535		5.281		139	
	L,E		114		0.3287		5.453		125	

All models have the same total volume $V=16.953 \text{ nm}^3$, identical to the sum of the atomic volumes of lysozyme. In analogy to the usage in programs HYDRO and HYDRO++, the centers of gravity and radii of gyration were calculated by using the bead volumes as weighting factors. The cases BD=0 and ICOM=1 correspond to all-atom and all-AA representations, respectively.

IM-AT: initial model, atomic groups, *IM-AM*: initial model, AA residues, *RM*: running mean, *BD*: bead diameter, *ICOM*: compression index, *CG*: cubic grid, *HG*: hexagonal grid, *VS*: voxel size, *G*: beads at centers of gravity, *L*: beads on lattice points, *U*: beads of unequal size, *E*: beads of equal size, N_b : number of beads, r_b : radius of beads, d_{max} : maximum particle diameter (vertex-to-vertex distance), *MSD*: mean square deviation of the $p(r)$ function, R_G : radius of gyration, R_G : radius of gyration by program HYDRO (a parameter calculated by our version of HYDRO or by HYDRO++), *D*: translational diffusion coefficient (a parameter calculated by our version of HYDRO or by HYDRO++), η : sedimentation coefficient (a parameter calculated by our version of HYDRO or by HYDRO++), f/η : intrinsic viscosity (a parameter calculated by our version of HYDRO or by HYDRO++)

^a This parameter was calculated by using the $p(r)$ function of the initial model IM-AT as reference.

^b For this parameter, our version of HYDRO and program HYDRO++ (run with the option ICASE=12) yielded identical results.

^c The program PDB2AT was applied to the PDB file 6lyz.pdb using increasing values of the BD.

^d Value obtained by means of our version of HYDRO using our patch for the RVC.

^e Value obtained by means of HYDRO++ using the AVC (option ICASE=12).

^f Value obtained by means of our version of HYDRO without volume correction after performing a CBS.

^g Value obtained by HYDRO++ using the CBS approach (option ICASE=21).

^h The program PDB2AM was applied to the PDB file 6lyz.pdb using increasing values of the compression index ICOM.

ⁱ The program MAP2GRID was applied to the initial model IM-AT using a constant value of the VS and placing the resulting beads at centers of gravity or lattice points.

^j The program MAP2GRID was applied to the initial model IM-AM using a constant value of the VS and placing the resulting beads at centers of gravity or lattice points.

HG approaches can be expected. On the other hand, this proves the applicability of all three approaches for constructing bead models of reduced bead number for hydrodynamic modeling purposes.

Apart from the visual inspection of the models obtained and the comparison of structural and hydrodynamic parameters, the evaluation of the distance distribution functions, $p(r)$, may be used as additional comparative indicator. As may be taken from Fig. 2, this indeed is the case. The values for the mean square deviation (MSD) of the $p(r)$ function from the reference $p(r)$ may be used as quantitative measure of the accuracy of the reduction procedures achieved (Table 1).

A critical comparison of the values obtained for the intrinsic viscosity $[\eta]$ reveals that the values obtained by HYDRO and our patch for the RVC are in good agreement with the values produced by HYDRO++ and using the AVC but in an even better agreement with the values obtained by application of the CBS approach. This follows convincingly from a comparison of the mean ratios Q_{AVC} and Q_{CBS} between the tabulated values of $[\eta]$ (see Table 3). In contrast to these values are the mean ratios Q_{uncorr} and Q_{FVC} , which were derived from the $[\eta]$ values (data not shown) calculated without any volume correction or with FVC. These ratios are considerably smaller or larger than 1.

3.2 The Capsid of the Bacteriophage fr

Reduction of highly complex and symmetric models such as those of the 180-meric capsid of the phage fr requires the application of highly sophisticated procedures to maintain the most pertinent structural peculiarities including symmetry characteristics during the reduction process. To handle this intricate problem, the reduction has to be performed in several steps. For a start, the first essential step in the reduction process could be the reduction of a building block, a subunit or multiples thereof (application of PDB2AT and

Fig. 2 Distance distribution functions $p(r)$ of selected reduced models of lysozyme obtained by RM or CG reduction approaches. The RM reductions are governed by the BD (program PDB2AT) or the compression index (ICOM; program PDB2AM); the CG reductions are governed by the VS (program MAP2GRID). All CG reductions were performed with VS=0.5 nm, starting from the crystal structure in atomic coordinates (AT; equivalent to RM with BD=0) or in AA coordinates (AM; obtained by RM with ICOM=1) and leading to models with beads at local centers of gravity (G) or on lattice points (L); the beads are either of unequal (U) or equal (E) size

Fig. 3 Reduced bead models for the capsid of bacteriophage fr and its subunits. **a** and **c** show the reduced models of the trimeric building block (subunits A, B, and C, displayed in red, blue, and green, respectively) of the capsid obtained by applying RM reductions to the crystal structure of each subunit (PDB entry: 1FRS) using different BD (**a**) or compression indices ICOM (**c**). Application of the appropriate symmetry relations to the subunit models provided the capsid models displayed in **b** and **d**, together with their central slabs. The models shown in the *leftmost columns* (BD=0 or ICOM=1) correspond to the original crystal structure (in atomic coordinates) (**a**, **b**) or to the AA representation (**c**, **d**) of the capsid and its building block

Table 2 Bead models of the capsid of phage ϕ and of its trimeric building block generated by various reduction approaches (RM, CG, and HG) and computation modes together with the calculated radii of gyration and hydrodynamic parameters

Model	Method	N_b	r_b (nm)	$d_{\max}$ (nm)	$MSD \times 10^6$ ^a	R_G (nm)	$R_{G,HYDRO}$ ^c (nm)	$D \times 10^7$ ^d (cm ² /s)	$s \times 10^{13}$ ^d (s)	$[\eta]$ (cm ³ /g)
RM: BD (nm) ^e										
Trimer	0	2,898	0.084–0.196	9.249		2.603				
	0.6	405	0.194–0.325	9.306	1.5 ^f	2.597				
	1.2	57	0.308–0.605	9.004	15.8 ^f	2.571				
	1.8	18	0.682–0.902	8.161	1.010 ^f	2.348				
Capsid	2.4	9	0.799–1.201	8.037	1.570 ^f	2.279				
	0 ^g	173,880	0.084–0.196	29.043		12.181	12.177			
	0.6	24,300	0.194–0.325	28.931		12.180	12.177			
	1.2	3,420	0.308–0.605	28.365	0.31 ^h	12.175	12.180	1.656	44.55	5.92 ⁱ
	1.8	1,080	0.682–0.902	28.026	2.8 ^h	12.129	12.128	1.653 ^j	44.45 ^j	5.76 ⁱ , 5.89 ^k , 5.84 ^l
	2.4	540	0.799–1.201	27.097	7.5 ^h	12.116	12.115	1.709 ^j	45.97 ^j	5.26 ⁱ , 5.34 ^k , 5.44 ^l
RM: ICOM ^m										
Trimer	1	387	0.243–0.378	9.302		2.603				
	6	66	0.468–0.610	8.912	9.1	2.580				
	12	33	0.633–0.751	8.399	154	2.513				
	24	18	0.633–0.923	8.203	1,270	2.325				
Capsid	48	9	0.993–1.162	7.248	1,080	2.318				
	129	3	1.584	6.592	1,290	2.297				
	1 ⁿ	23,220	0.243–0.378	29.027		12.181	12.178			
	6	3,960	0.468–0.610	28.727	0.19	12.177	12.175	1.657	44.56	6.03 ⁱ
	12	1,980	0.633–0.751	27.819	0.36	12.162	12.162	1.660	44.65	5.74 ⁱ , 7.41 ^{k,o}
	24	1,080	0.633–0.923	28.104	3.9	12.125	12.125	1.651 ^j	44.40 ^j	5.77 ⁱ , 5.91 ^k , 5.86 ^l
HG:G,U: VS (nm) ^q										
Capsid	48	540	0.993–1.162	27.582	4.2	12.124	12.120 ^j	1.677 ^j	45.12 ^j	5.54 ⁱ , 5.64 ^k , 5.68 ^l
	129	180	1.584	27.688	3.2	12.120	12.120 ^j	1.699 ^j	45.71 ^j	5.48 ⁱ , 5.47 ^k , 5.78 ^o
	0.9	7,326	0.255–0.660	29.150	0.17	12.179	12.179	1.671	44.96	6.12 ⁱ
	1.5	2,337	0.255–0.923	29.150	0.22	12.175	12.175	1.621	43.61	6.15 ⁱ
	2.1	1,062	0.255–1.213	29.094	0.22	12.168	12.168 ^j	1.626 ^j	43.75 ^j	6.04 ⁱ , 6.50 ^k , 6.06 ^l
	2.7	655	0.255–1.472	29.150	0.30	12.160	12.160 ^j	1.628 ^j	43.79 ^j	6.03 ⁱ , 6.23 ^k , 6.11 ^l
	3.3	420	0.255–1.762	29.150	0.75	12.149	12.149 ^j	1.639 ^j	44.08 ^j	5.96 ⁱ , 6.06 ^k , 6.10 ^l
	3.9	269	0.255–2.039	28.748	1.7	12.134	12.134 ^j	1.652 ^j	44.43 ^j	5.87 ⁱ , 5.94 ^k , 6.09 ^l
HG:L,U: VS (nm) ^q										
Capsid	4.5	217	0.349–2.180	28.977	3.9	12.112	12.112 ^j	1.646 ^j	44.27 ^j	5.96 ⁱ , 6.01 ^k , 6.17 ^o
	0.9	7,326	0.255–0.660	30.127	0.27	12.193	12.193	1.637	44.04	6.50 ⁱ
	1.5	2,337	0.255–0.923	30.678	1.8	12.216	12.216	1.562	42.03	6.83 ⁱ
	2.1	1,062	0.255–1.213	31.242	5.4	12.245 ^j	12.245 ^j	1.543 ^j	41.52 ^j	7.02 ⁱ , 7.62 ^k , 6.99 ^l
	2.7	655	0.255–1.472	33.245	9.6	12.261	12.261 ^j	1.508 ^j	40.57 ^j	7.51 ⁱ , 7.72 ^k , 7.48 ^l
	3.3	420	0.255–1.762	32.591	17.7	12.288	12.288 ^j	1.492 ^j	40.13 ^j	7.77 ⁱ , 7.89 ^k , 7.77 ^l
	3.9	269	0.255–2.039	34.568	33.1	12.445	12.445 ^j	1.496 ^j	40.24 ^j	7.76 ⁱ , 7.84 ^k , 7.83 ^l
	4.5	217	0.349–2.180	34.501	41.8	12.465	12.465 ^j	1.455 ^j	39.15 ^j	8.38 ⁱ , 8.45 ^k , 8.40 ^o

Capsid	HG,L,E: VS (nm) ^g									
0.9	7.326	0.461	30.539	0.23	12.192	12.272	1.616	43.47	6.71 ⁱ	
1.5	2.337	0.674	31.310	1.5	12.213	12.361	1.531	41.19	7.26 ⁱ	
2.1	1.062	0.877	32.330	4.6	12.239	12.515 ^j	1.501 ^j	40.37 ^j	7.63 ^k , 8.33 ^k , 7.69 ^l	
2.7	655	1.030	34.795	8.3	12.250	12.794 ^k	1.441 ^j	38.77 ^j	8.58 ^k , 8.77 ^k , 8.69 ^l	
3.3	420	1.194	34.041	16.2	12.272	12.926 ^l	1.420 ^j	38.20 ^j	8.98 ^k , 9.07 ^k , 9.11 ⁱ	
3.9	269	1.385	36.546	32.8	12.430	12.894 ^j	1.428 ^j	38.42 ^j	8.86 ^k , 8.95 ^k , 9.00 ^l	
4.5	217	1.488	36.652	41.5	12.441	13.497 ^j	1.367 ^j	36.76 ^j	10.07 ⁱ , 10.10 ^k , 9.93 ^p	

Abbreviations are explained in Table 1. All models of the trimeric building block and of the capsid have the same total volume: $V=49,945 \text{ nm}^3$ (trimer) and $V=2,996.7 \text{ nm}^3$ (capsid). In generating the models in this case, the masses of the beads were used as weighting factors.

^a If not stated otherwise, this parameter was calculated by using the $p(r)$ functions of the RM models corresponding to ICOM=1 as reference.

^b This parameter was calculated by using the correct weighting of the beads.

^c If not stated otherwise, this parameter was calculated by our version of HYDRO or, for models with $N_b > 10,000$, by an in-house program in a HYDRO-like manner (i.e., using the bead volumes as weighting factors).

^d If not stated otherwise, this parameter was calculated by our version of HYDRO.

^e The models of the trimeric building block (chains A, B, and C) were obtained by applying the program PDB2AT to the PDB file 1frs.pdb using increasing values of the BD; these models were then used to construct corresponding models of the capsid by employing the transformation matrices given in 1frs.pdb.

^f Value obtained by using the $p(r)$ function of the unreduced model (BD=0) as reference.

^g This model represents an initial model (IM-AT) of the capsid.

^h Value obtained by using the $p(r)$ function of the RM model corresponding to BD=0.6 nm as reference.

ⁱ Value obtained by means of our version of HYDRO using our patch for the RVC.

^j Value obtained by our version of HYDRO and by HYDRO++ (run with the option ICASE=12).

^k Value obtained by means of HYDRO++ using the AVC (option ICASE=12).

^l Value obtained by means of our version of HYDRO without volume correction after performing a CBS.

^m The models of the trimeric building block (chains A, B, and C) were obtained by applying the program PDB2AM to the PDB file 1frs.pdb using increasing values of the compression index ICOM; these models were then used to construct corresponding models of the capsid by employing the transformation matrices given in 1frs.pdb.

ⁿ This model represents an initial model (IM-AM) of the capsid.

^o This value is doubtful because it nearly equals the result obtained for FVC.

^p Value obtained by means of HYDRO++ using the CBS approach (option ICASE=21).

^q These reduced capsid models were obtained by applying the program MAP2GRID to the appropriately reoriented initial model (IM-AM) of the capsid using increasing values of the VS.

Table 3 Mean ratios of intrinsic viscosities $[\eta]$ obtained by various calculative approaches

Mean ratios of $[\eta]$ values	Lysozyme	Phage capsid
$Q_{AVC} = \langle [\eta]_{AVC} / [\eta]_{RVC} \rangle$	1.08 ± 0.03	1.04 ± 0.06
$Q_{CBS} = \langle [\eta]_{CBS} / [\eta]_{RVC} \rangle$	0.97 ± 0.02	1.02 ± 0.02
$Q_{uncorr} = \langle [\eta]_{uncorr} / [\eta]_{RVC} \rangle$	0.87 ± 0.05	0.97 ± 0.01
$Q_{FVC} = \langle [\eta]_{FVC} / [\eta]_{RVC} \rangle$	1.57 ± 0.23	1.24 ± 0.05

AVC: adjusted volume correction, *RVC*: reduced volume correction, *CBS*: cubic substitution, *FVC*: full volume correction

PDB2AM). Then, the reduced building blocks have to be assembled in such a way as to give the final complex (usage of appropriate transformation matrices). Finally, if required, the complex can be reduced further to yield a manageable number of beads (the initial model IM-AM, obtained by PDB2AM, is reoriented appropriately and then further reduced by MAP2GRID).

Fig. 4 Distance distribution functions $p(r)$ of models for the capsid of phage ϕ built from reduced models of the subunits. The crystal structure of each of the subunits A, B, and C (in atomic coordinates) was reduced separately by applying RM approaches yielding beads of nominal diameter BD (a) or representing ICOM AA residues (b). The displayed $p(r)$ functions correspond to models composed of 24,300, 3,420, 1,080, and 540 (a) or 23,220, 3,960, 1,980, 1,080, 540, and 180 beads (b), respectively

Reduced bead models of the trimeric building block (composed of subunits A, B, and C) and of the capsid of the phage fr are illustrated in Fig. 3. These models were obtained by RM approaches, varying either the BD (Fig. 3a) or the compression index ICOM (Fig. 3b). Both procedures prove the possibility to reduce the bead number drastically (by a factor of 322 compared to the original crystal structure), concomitantly retaining essential molecule features regarding size and symmetry, and irrespective of the usage of atomic coordinates or AA residues. The appearance of the models, however, changes considerably at the higher reduction levels.

Quantification of the RM results in terms of structural and hydrodynamic parameters and MSD values is shown in Table 2. Similar to the example lysozyme, too extensive reduction steps should be avoided if correct values for the molecular parameters are desired. In the case of the BD variation, reductions by a factor of 10 provide no problem, but factors >100 should be avoided. Owing to the reduced number of beads in the initial model (factor of about 10), the variation of ICOM should not exceed a factor of 10. Again, the mentioned calculation methods for $[\eta]$ are in accord. As follows from the mean ratios Q presented in Table 3, also with this protein, the results obtained by our RVC agree best with those computed by using the CBS approach.

An inspection of the $p(r)$ functions of the capsid (Fig. 4), built from reduced models of the subunits, reveals oscillations in the functions for the higher reduction steps, although the overall impression of all functions is satisfactory. This, again, is in agreement with the conclusions drawn on the basis of the parameter predictions.

Application of the HG approach and varying computation modes also produces realistic models for the capsid (Fig. 5). Table 2 presents evidence that the GU mode (beads of unequal size at centers of gravity) and reductions not exceeding a factor of 10 with regard to the

Fig. 5 Reduced bead models for the capsid of bacteriophage fr obtained by mapping the AA representation of the capsid (initial model [IM-AM] with $N_b=23,220$) into HG of different VS and placing the resulting unequal beads at local centers of gravity (G) or on exact lattice points (L). The central slabs displayed in the second row visualize the inner surface of the hollow sphere-like particles

starting model (IM-AM) or about 100 in comparison to the original crystal structure yield reliable values regarding the prediction of structural and hydrodynamic parameters. And the same is true for the distance distribution functions $p(r)$ that agree to a great extent in the case of the GU mode, but not when applying the LU approach (beads of unequal size on lattice points) (Fig. 6). In the latter case, serious discrepancies in both the form and position of the maximum of the $p(r)$ function and the resulting particle diameter, $d_{\max}$, come to the fore. Similar pronounced discrepancies occur upon applying the LE approach (beads of equal size on lattice points, profiles not shown). Again, quantifying the discrepancies of the $p(r)$ functions in terms of the MSD values proves the fact that the GU mode should be preferred.

In principle, the same information obtained from $p(r)$ can be obtained by inspecting the scattering intensities $I(h)$ (profiles not shown). The real-space information, however, is more intuitive, straightforward, and instructive than the reciprocal-space profiles [1, 5]. In this context, it should also be mentioned that the $p(r)$ function of a hollow sphere represents an extreme case because $p(r)$ resembles a triangle, whereas the $p(r)$ of a solid sphere or other globular particles is bell-shaped. By contrast, the corresponding $I(h)$ profiles of hollow and solid spheres differ mainly in the height of their (error-prone) submaxima. In

Fig. 6 Distance distribution functions $p(r)$ of reduced models for the capsid of phage ϕ obtained by mapping the initial model of the capsid (IM-AM) into HG of different VS. The resulting beads of unequal size are placed at the local centers of gravity of the lattice cells (a) or on exact lattice points (b). In both cases, the displayed $p(r)$ functions correspond to models composed of 23,220, 7,326, 2,337, 1,062, 655, 420, 269, and 217 beads, respectively

the case of the capsid under consideration, the intensities obtained for coarse-grained models tend to deviate at $h > 1.5 \text{ nm}^{-1}$ [16].

4 Conclusions

In the past, we have modeled several proteins of different size and symmetry, tentatively applying different approaches for the required reduction steps [9, 10]. Among the proteins tested, the application of the RM and/or CG approaches to monomeric proteins (malate synthase, cellobiose dehydrogenase), dimeric proteins (citrate synthase, phosphorylase), tetrameric proteins (glyceraldehydes-3-phosphate dehydrogenase, lactate dehydrogenase, catalase), and the dodecameric subunit of earthworm hemoglobin [9–11, 13, 15, 17, 18] can be mentioned; HG approaches have been applied advantageously in the case of proteins of hexagonal symmetry such as the giant, 180-meric annelid hemoglobins [11–14, 17].

In this study, we have applied the RM, CG, and HG approaches for hydrodynamic modeling the small test protein, lysozyme, and a huge protein, the capsid of the phage ϕ , and its trimeric building block. This was performed by analyzing different calculation modes and input variables in a systematic manner. The choice between RM, CG, and HG approaches depends on the symmetry properties of the protein to be analyzed. When using CG or HG procedures, beads of unequal (U) size that are located at centers of gravity (G) are the optimum modes to be used. As may be taken from a comparison of the MSD values, in any case, one should abstain from using lattice points. While the placement of equal-sized beads (LE) on lattice points yields slightly lower MSD values than the use of unequal-sized beads (LU), the LE approach nevertheless turns out to be the worst choice for hydrodynamic calculations because the appropriate weights of the equal-sized beads cannot be utilized by HYDRO or HYDRO++.

Several bead reduction steps revealed that a total reduction up to a factor of 100 (starting from the initial atomic coordinates) is possible without changing the structural features and the predicted molecular properties. This aspect is of utmost importance if hydrodynamic properties of multibead models are to be analyzed. In this case, it is imperative to reduce the size and complexity of huge bead assemblages to overcome the limitations imposed by numerics and computer software and hardware. Together with the application of appropriate interaction tensors and volume corrections, this *modus operandi* provides the scientific basis for the investigation of complex proteins and the hydrodynamic elucidation of their molecular properties. Of course, scattering modeling approaches may also benefit from reduction procedures, e.g., for accelerating scattering calculations of multibead models based on the Debye formula. Application of such coarse-grained models may be advantageous for examining a wide range of questions in structural biology and SAXS-based refinements as well.

The intrinsic viscosities $[\eta]$ computed by our approach of RVC have turned out to be about equivalent to the results obtained from the CBS approach. This is another aspect of utmost importance because the latter approach was claimed to yield the most reliable results for filling models [24] but is applicable only to models of up to about 1,250 beads (when using HYDRO++ this limit even reduces to 250 beads) and requires much more computing time (by a factor of about 500) than our approach. The results obtained by the approach of AVC [24], on the other hand, were found to deviate on the whole a little bit more from ours, and in some exceptional cases, they turned out to reach or even exceed the values calculated by using the FVC. At this stage, our strategy concerning RVC has been tested in detail only for the globular particles analyzed. Pilot tests on a few prolate and oblate particles (e.g.,

cellobiose dehydrogenase, trimeric malate synthase), however, revealed that our approach might also be applicable to these particle geometries.

Acknowledgements The authors are much obliged to J. García de la Torre for the use of the computer programs HYDRO and HYDRO++, and to R.A Sayle for RASMOL.

References

1. Glatter, O., Kratky, O. (eds.): Small Angle X-ray Scattering. Academic, London (1982)
2. Durchschlag, H.: Small-angle X-ray scattering of proteins in relation to food systems, with special emphasis on enzymes and storage proteins. In: Baianu, I.C., Pessen, H., Kumosinski, T.F. (eds.) Physical Chemistry of Food Processes, Vol. 2 Advanced Techniques, Structures, and Applications, pp. 18-117. Van Nostrand Reinhold, New York (1993)
3. Harding, S.E., Rowe, A.J., Horton, J.C. (eds.): Analytical Ultracentrifugation in Biochemistry and Polymer Science. Royal Society of Chemistry, Cambridge (1992)
4. Scott, D.J., Harding, S.E., Rowe, A.J. (eds.): Analytical Ultracentrifugation: Techniques and Methods. Royal Society of Chemistry, Cambridge (2005)
5. Svergun, D.I., Koch, M.H.J.: Small-angle scattering studies of biological macromolecules in solution. Rep. Prog. Phys. **66**, 1735-1782 (2003)
6. Zipper, P., Durchschlag, H., Krebs, A.: Modelling of biopolymers. In: Scott, D.J., Harding, S.E., Rowe, A.J. (eds.) Analytical Ultracentrifugation: Techniques and Methods, pp. 320-371. Royal Society of Chemistry, Cambridge (2005)
7. García de la Torre, J., Navarro, S., López Martínez, M.C., Díaz, F.G., López Cascales, J.J.: HYDRO: a computer program for the prediction of hydrodynamic properties of macromolecules. Biophys. J. **67**, 530-531 (1994)
8. García de la Torre, J., Huertas, M.L., Carrasco, B.: Calculation of hydrodynamic properties of globular proteins from their atomic-level structure. Biophys. J. **78**, 719-730 (2000)
9. Zipper, P., Durchschlag, H.: Calculation of hydrodynamic parameters of proteins from crystallographic data using multibody approaches. Prog. Colloid & Polym. Sci. **107**, 58-71 (1997)
10. Zipper, P., Durchschlag, H.: Recent advances in the calculation of hydrodynamic parameters from crystallographic data by multibody approaches. Biochem. Soc. Trans. **26**, 726-731 (1998)
11. Zipper, P., Durchschlag, H.: Prediction of hydrodynamic and small-angle scattering parameters from crystal and electron microscopic structures. J. Appl. Crystallogr. **33**, 788-792 (2000)
12. Zipper, P., Krebs, A., Durchschlag, H.: Prediction of hydrodynamic parameters of *Lumbricus terrestris* hemoglobin from small-angle X-ray and electron microscopic structures. Prog. Colloid & Polym. Sci. **119**, 141-148 (2002)
13. Zipper, P., Durchschlag, H.: Modeling of protein solution structures. J. Appl. Crystallogr. **36**, 509-514 (2003)
14. Krebs, A., Durchschlag, H., Zipper, P.: Small angle X-ray scattering studies and modeling of *Eudistylia vancooverii* chlorocruorin and *Macrobdella decora* hemoglobin. Biophys. J. **87**, 1173-1185 (2004)
15. Durchschlag, H., Zipper, P.: Calculation of volume, surface, and hydration properties of biopolymers. In: Scott, D.J., Harding, S.E., Rowe, A.J. (eds.) Analytical Ultracentrifugation: Techniques and Methods, pp. 389-431. Royal Society of Chemistry, Cambridge (2005)
16. Zipper, P., Durchschlag, H.: Modelling of bacteriophage capsids and free nucleic acids. J. Appl. Crystallogr. **40**, s153-s158 (2007)
17. Durchschlag, H., Zipper, P., Krebs, A.: A comparison of protein models obtained by small-angle X-ray scattering and crystallography. J. Appl. Crystallogr. **40**, 1123-1134 (2007)
18. Durchschlag, H., Zipper, P.: Volume, surface and hydration properties of proteins. Prog. Colloid & Polym. Sci. in press (2008)
19. Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., Bourne, P.E.: The protein data bank. Nucleic Acids Res. **28**, 235-242 (2000)
20. Durchschlag, H., Zipper, P.: Modeling the hydration of proteins at different pH values. Prog. Colloid & Polym. Sci. **127**, 98-112 (2004)
21. Traube, J.: Ueber den Raum der Atome. Samml. chem. chem.-tech. Vortr. **4**, 255-332 (1899)
22. Sayle, R.A., Milner-White, E.J.: RASMOL: biomolecular graphics for all. Trends Biochem. Sci. **20**, 374-376 (1995)

23. Glatter, O.: Computation of distance distribution functions and scattering functions of models for small angle scattering experiments. *Acta Phys. Austriaca* **52**, 243-256 (1980)
24. García de la Torre, J., del Río Echenique, G., Ortega, A.: Improved calculation of rotational diffusion and intrinsic viscosity of bead models for macromolecules and nanoparticles. *J. Phys. Chem. B* **111**, 955-961 (2007)
25. Zipper, P., Durchschlag, H.: Prediction of hydrodynamic parameters from 3D structures. *Prog. Colloid & Polym. Sci.* **113**, 106-113 (1999)
26. Carrasco, B., García de la Torre, J., Zipper, P.: Calculation of hydrodynamic properties of macromolecular bead models with overlapping spheres. *Eur. Biophys. J.* **28**, 510-515 (1999)
27. García de la Torre, J., Carrasco, B.: Intrinsic viscosity and rotational diffusion of bead models for rigid macromolecules and bioparticles. *Eur. Biophys. J.* **27**, 549-557 (1998)
28. García de la Torre, J., Bloomfield, V.A.: Hydrodynamic properties of complex, rigid, biological macromolecules: theory and applications. *Q. Rev. Biophys.* **14**, 81-139 (1981)
29. García Bernal, J.M., García de la Torre, J.: Transport properties of oligomeric subunit structures. *Biopolymers* **20**, 129-139 (1981)