Accepted to J. Chem. Phys. 10.1063/5.0132493

ACCEPTED MANUSCRIPT

The Journal of Chemical Physics

AIP Publishing

This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.
PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0132493

Effects of flexibility in coarse-grained models for bovine serum albumin and immunoglobulin G

Frank Hirschmann,[1, a)] Hender Lopez,[2, b)] Felix Roosen-Runge,[3] Tilo Seydel,[4] Frank Schreiber,[1] and Martin Oettel[1]

[1)] *Institute for Applied Physics, University of Tübingen, Auf der Morgenstelle 10, 72076 Tübingen, Germany*

[2)] *School of Physics, Clinical and Optometric Sciences, Technological University Dublin, Grangegorman, D07 ADY7, Ireland*

[3)] *Department of Biomedical Sciences and Biofilms-Research Center for Biointerfaces (BRCB), Malmö University, 20506 Malmö, Sweden*

[4)] *Institut Max von Laue – Paul Langevin, 71 avenue des Martyrs, 38042 Grenoble, France*

(Dated: 6 February 2023)

We construct a coarse-grained, structure-based, low-resolution, 6-bead flexible model of bovine serum albumin (BSA, PDB: 4F5S), which is a popular example of a globular protein in biophysical research. The model is obtained via direct Boltzmann inversion using all-atom simulations of a single molecule, and its particular form is selected from a large pool of 6-bead coarse-grained models using two suitable metrics which quantify the agreement in the distribution of collective coordinates between all-atom and coarse-grained Brownian dynamics simulations of solutions in the dilute limit. For immunoglobulin G (IgG), a similar structure-based 12-bead model has been introduced in the literature (J. Phys. Chem. B **116**, 8045 (2012)) and is employed here to compare findings for the compact BSA molecule and the more anisotropic IgG molecule. We define several modified coarse-grained models of BSA and IgG, which differ in their internal constraints and thus account for a variation of flexibility. Studying denser solutions of the coarse-grained models with purely repulsive molecules (achievable by suitable salt conditions), the effect of packing and flexibility on dynamic and static behavior is studied. Translational and rotational self-diffusivity is enhanced for more elastic models. Finally we discuss a number of effective sphere sizes for the BSA molecule which can be defined from its static and dynamic properties. Here it is found that the effective sphere diameters lie between 4.9 and 6.1 nm, corresponding to a relative spread of about $\pm 10\%$ around a mean of 5.5 nm.

I. INTRODUCTION

Bovine serum albumin (BSA) is a globular protein[1] which is widely used in biochemical[2–5] and biophysical[6–10] research. It serves notably as a model system for studying static structure, phase behavior and dynamic properties of biomacromolecules.[11] A rather simple colloid picture has proven remarkably successful in describing BSA solutions. Intensities from static scattering experiments can be described well by assuming ellipsoidal form factors and structure factors from hard-sphere Yukawa models for BSA in NaCl solutions.[12,13] Trivalent cations in BSA solutions lead to strong attractions and reentrant phase behavior, but also here an isotropic colloidal model (sticky hard spheres) seems to be sufficient to describe static scattering.[14–18]

Self-diffusion in crowded (globular) protein systems is also often discussed in reference to isotropic colloidal models. One needs to distinguish between short-time and long-time diffusion coefficients, where short-time diffusion is mainly governed by hydrodynamic interactions, and long-time diffusion in general contains effects of both hydrodynamics and intermolecular interactions. Long-time self-diffusion in BSA solutions was measured by Muramatsu and Minton already in 1988 and discussed in a pure hard sphere picture.[19] Theoretical work by Tokuyama *et al.* on the self-diffusion for biomolecules considers the hard sphere model as sufficient for short-time diffusion, but relates the long-time diffusion coefficient to a reference system of soft spheres.[20] This bears some connection to the paradigm of universality in simple liquids put forward by Dyre and coworkers, where the reference system is also a soft system and not hard spheres.[21] The hard-sphere picture for short-time diffusion in BSA solutions has been well corroborated in neutron scattering experiments.[22,23] For long-time diffusion in BSA, experiments and Brownian dynamics simulations reported in Ref. 24 indicate some deviations from the "soft-sphere universality". Here the simulations were performed using a rigid model but with atomistically resolved anisotropy, with interatomic repulsions and Debye-Hückel interactions between effective charges. The latter work inspires to look at anisotropy effects in BSA systems.

A popular strategy in soft matter physics to model polymeric systems is to represent the molecule by a reduced number of effective beads with effective intermolecular and intramolecular interactions and, thus, to establish a coarse-grained picture. Immunoglobulin (IgG) molecules, due to their relevance in immunology and pharmaceutics, as well as due to their remarkably anisotropic shape, constitute a particular example where a number of studies have used such an approach to describe numerous different aspects such as self-association, diffusion, clustering, scattering intensity profiles, network formation and viscosity behavior.[25–39]

[a)] Author to whom correspondence should be addressed: frank.hirschmann@uni-tuebingen.de

[b)] Author to whom correspondence should be addressed: hender.lopezsilva@tudublin.ie

Accepted to J. Chem. Phys. 10.1063/5.0132493

ACCEPTED MANUSCRIPT
The Journal of Chemical Physics
AIP Publishing
This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.
PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0132493

The aims of the paper are as follows. Using atomistic simulations of single BSA molecules, we construct a coarse-grained (CG) 6-bead model. This approach is, in spirit, very similar to the construction of coarse-grained bead models for IgG in Ref. 27. For both the newly constructed CG model for BSA and the 12-bead model for IgG from Ref. 27 pair correlations and translational/rotational diffusion are examined using overdamped Brownian dynamics (BD) simulations. Here, bead interactions in the CG models are taken to be repulsive. For BSA solutions at physiological conditions (around neutral pH and salt concentration ~ 0.3 M), the Debye length is approximately 0.6 nm, which results in a strong screening of electrostatic interactions. In similar conditions, van der Waals interactions are also rather small, as reflected by a positive second virial coefficient.[40,41] We thus focus on the excluded volume effects in dense solutions. We analyze pair correlations with effective, isotropic potentials, which turn out to be rather soft and less frequently used in colloidal modelling. Diffusion from BD simulations is certainly not realistic in the short-time regime, due to the neglect of hydrodynamic interactions (HI), but here we focus on the ratio of long–term to short–term diffusion coefficients (which should be mainly determined by interactions between the proteins and not HI) and examine the effects of varying the flexibility of the coarse-grained molecules. Here, it is interesting to contrast the more globular BSA molecule with the IgG molecule which possesses three globular regions connected through a flexible hinge region, thus forming a Y-shape. For both molecules, one globular and one more anisotropic, employing the experimentally determined HI, our simulations reproduce the experimentally observed long-time diffusion coefficients to a good extent.

II. COARSE-GRAINED (CG) MODELS FOR BSA AND IGG

A. BSA: CG beads from protein domains

The BSA protein consists of 583 amino acid residues which can be divided into six biological domains (Fig. 1(a)) whereby the specific loci of the domains within the sequence can be obtained from the protein structure classification database CATH.[1,42] In this work we propose a structure-based CG model for BSA, in which each of the protein biological domains is represented by a single bead and the connectivity of the different beads capture the internal dynamics of the protein. Note that other, more general-purpose CG models, e.g. the MARTINI[43,44] model, have shown to be useful in a variety of situations, but are not adequate for evaluating the internal dynamics of proteins. The location of the CG sites is based on the center-of-mass of these domains (Fig. 1(b)), whereby any residue not part of a biological domain has been included to its nearest domain. The X-ray crystal structure and residue sequence used here to associate residues to a CG bead were obtained from the protein database (PDB ID: 4F5S, reported in Ref. 1). Note that previous studies have used more sophisticated methods such as neural network topology building,[45] or elastic normal-mode analysis[27] to determine the positioning

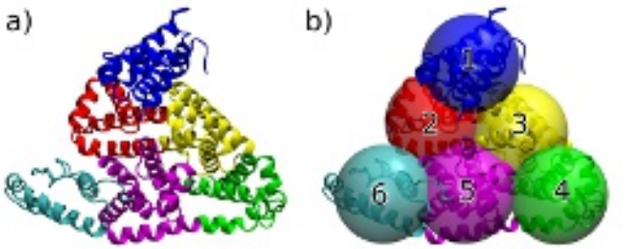


FIG. 1. (a) Tertiary structure[1] of BSA with its domains in different colors. (b) CG model of BSA showing the labelling of the different CG beads superimposed onto the 3D structure of BSA. Tab. I reports the composition of each of the CG beads. The beads are equally sized with $\sigma_{\text{bead}}^{\text{BSA}} = 29.52$ Å, see App. C for details.

of the CG sites, albeit, in the latter study the authors eventually also used the domain-based positioning of their beads. Fig. 1(b) shows the different CG beads and their labels superimposed onto the tertiary structure of BSA. The labels of the CG beads correspond to the sequence of the domains along the amino acid backbone of the protein. The exact mapping of the coarse-grained sites, their corresponding masses and cumulative charges are reported in Tab. I.

TABLE I. Composition, masses and charges of the six BSA CG sites. The charges are cumulative partial charges of the residues, whose protonation states have been determined by the program pdb2gmx[46] with default settings, which corresponds to the most common protonation state of amino acids at pH 7: Asp and Glu deprotonated, Arg and Lys protonated, His dependent on optimal hydrogen bond conformation.

Bead	Residues	Mass [kDa]	Charge [e]
1	1-107	12.2	-9
2	108-195	10.3	0
3	196-295	11.3	1
4	296-380	9.7	-6
5	381-493	12.8	2
6	494-583	10.0	-4

From the CG representation one sees that beads 2–6 all lie approximately in a plane while bead 1 sticks out of this plane, so overall the molecule appears pancake–like with a protruding end. The flat shape of the molecule has been corroborated by X-ray scattering, where a form factor for an oblate ellipsoid of half axes $1.7\times4.2\times4.2$ nm was fitted to scattering data.[12]

B. BSA: Parametrization of the intramolecular potentials for the CG model

There are various methodologies to derive intramolecular potentials between the CG beads, for reviews see Refs. 47–52. In bottom-up approaches, these potentials are obtained from all-atom (AA) simulations of the BSA molecule. A versatile approach, popular in the biophysical community, are elastic network models (ENM) in conjunction with normal mode

Accepted to J. Chem. Phys. 10.1063/5.0132493

ACCEPTED MANUSCRIPT
The Journal of Chemical Physics
AIP Publishing
This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.
PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0132493

analysis (NMA), which were first introduced by Tirion[53] and since then have become a popular and successful CG strategy which accurately describe domain motions and collective dynamics in proteins.[54] The original ENM by Tirion consists of nodes, which are connected by harmonic springs with a uniform spring constant, provided that they are within a certain cutoff distance. More advanced techniques have been developed by e.g. Hinsen[55] (here, spring constants decay with increasing distance of two nodes) or Lyman et al.[56] (spring constants are tuned for the CG molecule to yield correct mean-square distance fluctuations).

Another versatile approach, adopted here, uses three standard classes of effective interaction potentials between CG beads: bonds, angles and dihedrals. Bond potentials act between two CG beads, while angles and dihedrals apply forces on three and four CG beads, respectively. Within the class of bond potentials, one can define Urey-Bradley (UB) bonds, where the two beads in the bond correspond to the outer beads of an angle potential. All potentials are taken to be harmonic in their specific generalized coordinate (bead distance for bonds, the angle between three beads, or the dihedral angle in the four bead configuration) with specific spring constants. The use of these potentials is motivated by the basic motions seen in chainlike molecules, and for IgG the respective bottom-up parametrization has been performed in Ref. 27, which we will use and compare with our findings for BSA. The determination of the spring constants was realized by direct Boltzmann inversion which can be briefly summarized as follows: In all–atom simulations (details in App. A), trajectories of the center-of-mass coordinates of the six residue groups reported in Tab. I (corresponding to bead trajectories) are determined and histograms in the generalized coordinates for the interaction potentials are recorded. By assuming independent thermal sampling of the generalized coordinates, these histograms can be separately fitted to shifted Gaussians, with the spring constants being related to the width of the Gaussian. More details are found in App. B.

The derived force field is not only dependent on the selected CG beads, but is also a function of the connectivity between the beads in terms of bonds, angles and dihedrals and their generalized coordinates. For a chosen connectivity model (called topology in the following), the coordinates are not statistically independent, and employing too many effective potentials easily results in an overconstrained molecule, which is too rigid. Nevertheless, it is useful to consider such an overconstrained topology as a starting point, so we propose a (nearly) fully bond-connected topology, which we call FULL. Fig. 2(b) shows a visualization of the bond topology and Tab. II reports the connectivity definitions of the different bonds, angles and dihedrals. For the complete CG model with the FULL topology, the spring constants for bonds, angles and dihedrals are determined as described before (model FULLo, referring to an optimal choice of force constants in the FULL topology).

TABLE II. Connectivity of the FULL topology, which is realized in the models FULLo and FULLrigid. The numbering of the CG beads corresponds to the one shown in Fig. 2. The FULL topology is determined by the following rules: (i) each bead is bond-connected to its two nearest neighbors, (ii) each linearly bonded sequence (i,j,k) of beads gives rise to an angle potential for the angle (i,j,k) and beads i,k become connected by a UB bond (if not already bonded), (iii) Four subsequent beads will be connected by a dihedral potential (i.e. $(1,2,3,4)$, $(2,3,4,5)$, $(3,4,5,6)$). Three additional dihedrals are added in a cyclic manner (i.e. $(4,5,6,1)$, $(5,6,1,2)$, $(6,1,2,3)$) to stabilize the three–dimensional structure.

Bonds	UB-Bonds	Angles	Dihedrals
(1,2)	(1,(2),6)	(1,2,3)	(1,2,3,4)
(1,3)	(1,(3),4)	(1,2,6)	(2,3,4,5)
(2,3)	(2,(3),4)	(1,3,2)	(3,4,5,6)
(3,4)	(3,(2),6)	(1,3,4)	(4,5,6,1)
(4,5)	(3,(4),5)	(2,1,3)	(5,6,1,2)
(5,6)	(4,(5),6)	(2,3,4)	(6,1,2,3)
(6,2)	(5,(6),2)	(3,2,6)	
		(3,4,5)	
		(4,5,6)	
		(5,6,2)	

We define another CG model based on the FULL topology, which will be useful later on, namely the model FULLrigid, which has the same topology but for which all force constants are increased by a factor of 100, thus resulting in a molecule that can be considered quasi-rigid (see Tab. IV for an overview of the models discussed in this article.).

Deleting bonds and/or angles and/or dihedrals from the FULL topology results in a considerable number of less constrained topologies. As the parametrization of the interactions (bonds, angles and dihedrals) is independent of the topology, any new CG model for these new topologies have the same spring constants for their retained connections as in model FULLo. An educated guess for a reasonable connectivity comes from the domain sequence along the amino acid backbone (as reflected by the bead number). Direct bonds shall only exist along the backbone. In the following we call this domain structure–based topology STRUCT. See Fig. 2(a) for a visualization of the resulting bond topology and Tab. III for a listing of bonds, angles and dihedrals. The complete CG model with the STRUCT topology and the spring constants from FULLo is called STRUCTo.

TABLE III. Connectivity of the STRUCT topology, which is realized in the models STRUCTo, STRUCTflex, and STRUCTnd. The numbering of the CG beads corresponds to the one shown in Fig. 2. The topology is determined by the rules: (i) subsequent beads will be connected by a bond (e.g. $(1,2)$, $(2,3)$ etc.), (ii) three subsequent beads will be connected by an angle potential (e.g. $(1,2,3)$), (iii) dihedrals from FULL are retained (with the exception of STRUCTnd).

Bonds	UB-Bonds	Angles	Dihedrals
(1,2)		(1,2,3)	(1,2,3,4)
(2,3)		(2,3,4)	(2,3,4,5)
(3,4)		(3,4,5)	(3,4,5,6)
(4,5)		(4,5,6)	(4,5,6,1)
(5,6)			(5,6,1,2)
			(6,1,2,3)

Accepted to J. Chem. Phys. 10.1063/5.0132493

ACCEPTED MANUSCRIPT
The Journal of Chemical Physics
AIP Publishing
This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.
PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0132493

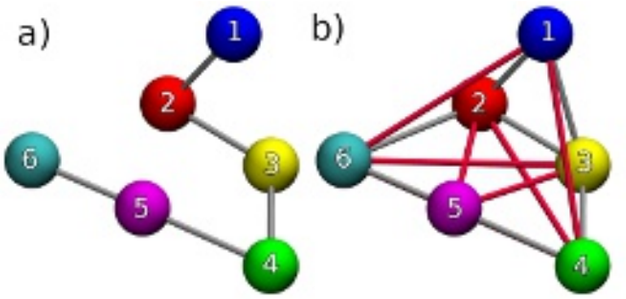


FIG. 2. Representation of the coarse-grained BSA molecule, showing the bonds (grey connectors) and UB-bonds (red connectors) for the two proposed topologies, (a) STRUCT and (b) FULL. Bead sizes have been scaled down for better visibility.

Based on the STRUCTo model, we define two further CG models by manipulating the spring constants. In the model STRUCTflex the spring constants are multiplied by a factor of 0.75 to represent a more flexible molecule. Secondly, in the model STRUCTnd, all dihedral potentials are set to zero which potentially destabilizes the overall three–dimensional structure of the molecule.

To summarize, all proposed models for BSA (and IgG), their underlying connectivities and features are reported in Tab. IV.

TABLE IV. Overview of the five proposed coarse-grained BSA models, as well as the two IgG models. The topologies FULL and STRUCT are defined in Tab. II and Tab. III.

Model	Topology	Spring constant multiplier	Note
FULLo	FULL	1	-
FULLrigid	FULL	100	-
STRUCTo	STRUCT	1	-
STRUCTnd	STRUCT	1	no dihedrals
STRUCTflex	STRUCT	0.75	-
IGGo	-	1	from Ref. 27
IGGrigid	-	100	from Ref. 27

As explained above, the STRUCT topology is solely based on the main backbone structure, but other connectivities based on the structure of the protein can be defined. To evaluate the adequacy of the STRUCT topology to accurately capture the CG dynamics of the all-atom model and compare it to other possible connectivities, we propose a systematic coarse-graining strategy in which a number of unique topologies are created by randomly deleting bonds, angles and dihedrals from the FULL topology. More specifically, the protocol is as follows: (i) all force field constituents (bonds, angles, dihedrals) are put in a single list. (ii) A random number r is determined and r constituents are deleted from this list. (iii) Uniqueness is enforced by skipping already generated topologies and a minimum number of 5 bonds is required. All the spring constants for these models are those from FULLo.

Subsequently, for all topologies with their associated spring constants, we run coarse-grained overdamped Brownian dynamics simulations of BSA molecules with only internal interactions (corresponds to volume packing fraction $\Phi = 0$) and topologies yielding unstable molecules are omitted. We define two metrics to evaluate the quality of a CG model: The first uses normalized histograms (probability distributions) of all generalized coordinates presented in the FULL topology. These are compared to the corresponding probability distributions from all-atom simulations by calculating their common area (overlap), see App. D. For this, a simple metric $\overline{\overline{S}} \in [0,1]$ is defined as the average of all calculated common areas. A value of 1 is optimal in the sense that the CG simulations reproduce the ensemble-averaged configurational states recorded in the all-atom simulations to at least the level of all observed coordinate distributions. The second metric is based on the commonly used root-mean-square deviation (RMSD) of CG bead and all-atom residue group positions $\vec{r}_i$,

$$\mathrm{RMSD}(t) = \left\langle \left(\frac{1}{N_{\mathrm{sites}}} \sum_{i=1}^{N_{\mathrm{site}}} |\vec{r}_i(0) - \vec{r}_i(t)|^2 \right)^{1/2} \right\rangle . \qquad (1)$$

By definition, Eq. (1) entails optimal translational and rotational superposition of the structure at time t to itself at $t = 0$. The second metric is a time-averaged difference in RMSDs from coarse-grained and all-atom simulations:

$$\Delta\mathrm{RMSD} := \left| \overline{\mathrm{RMSD}_{\mathrm{AA}}(t)} - \overline{\mathrm{RMSD}_{\mathrm{CG}}(t)} \right| , \qquad (2)$$

for more details on the averaging in Eqs. (1) and (2) we refer to App. D.

Fig. 3 shows the values of $\overline{\overline{S}}$ on the left y-axis for the 4730 topologies generated, as well as the five predefined ones, where each point corresponds to one particular topology and they are sorted in ascending order for $\overline{\overline{S}}$ from left to right on the x-axis. The points for the models FULLo, FULLrigid, STRUCTo, STRUCTflex and STRUCTnd are highlighted. In the same figure the right-hand y-axis shows the corresponding ΔRMSD for all topologies. Both similarity measures overall coincide in establishing, which topologies represent better the flexibility of the all-atom model. A perfect match corresponds to $\Delta\mathrm{RMSD} = 0$ and $\overline{\overline{S}} = 1$. $\overline{\overline{S}}$ clearly shows that FULLrigid does not well in capturing the flexibility of the molecule. The model FULLo performs better, but is still too rigid as compared to the all-atom model. STRUCTnd has also a comparably low value of $\overline{\overline{S}}$, because the neglect of dihedrals compromises the three-dimensional stability (this will be important below). The models STRUCTo and STRUCTflex are in the group of the best performers, i.e. $\overline{\overline{S}}_{\mathrm{STRUCTo}} = 0.9157$ and $\overline{\overline{S}}_{\mathrm{STRUCTflex}} = 0.8902$, with only 22 topologies being in the interval $\overline{\overline{S}} \in [0.89, 1]$. Furthermore, according to the measure $\overline{\overline{S}}$, STRUCTo is nearly optimal, with only one randomly generated topology prevailing with $\overline{\overline{S}} = 0.9161$.

Obviously, disabling interactions, starting from the FULL topology, can lead to better as well as to worse performance. Further analysis of all generated force field compositions shows that if the total number of force field constituents falls below 12 (down from 30, which is the sum of all elements of the FULL topology, Tab. II), the resulting topology

Accepted to J. Chem. Phys. 10.1063/5.0132493

ACCEPTED MANUSCRIPT
The Journal of Chemical Physics
AIP Publishing
This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.
PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0132493

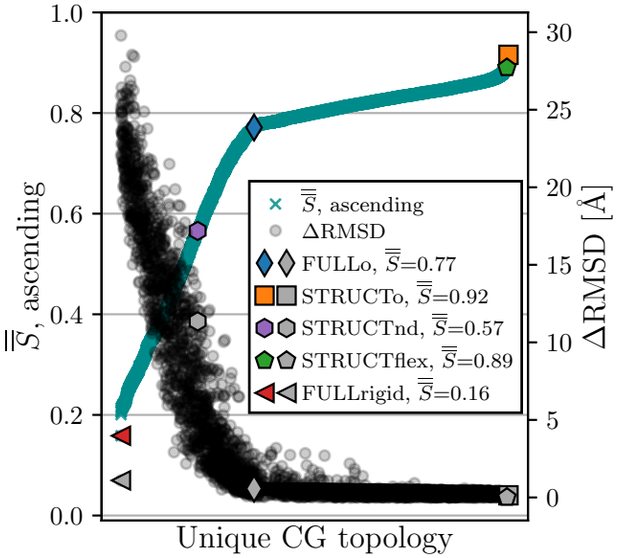


FIG. 3. The similarity measure $\overline{\overline{S}}$ (Eq. (D3)) and the ΔRMSD (Eq. (2)) measure for each of the 4735 structures studied. Each vertical pair of cross and circle corresponds to a particular model generated from the FULL topology. The results are shown in ascending order for $\overline{\overline{S}}$ from left to right. Filled symbols show the results for the five predefined models based on the FULL and STRUCT topology (see Tab. IV). Grey/black symbols represent ΔRMSD values, colored symbols $\overline{\overline{S}}$.

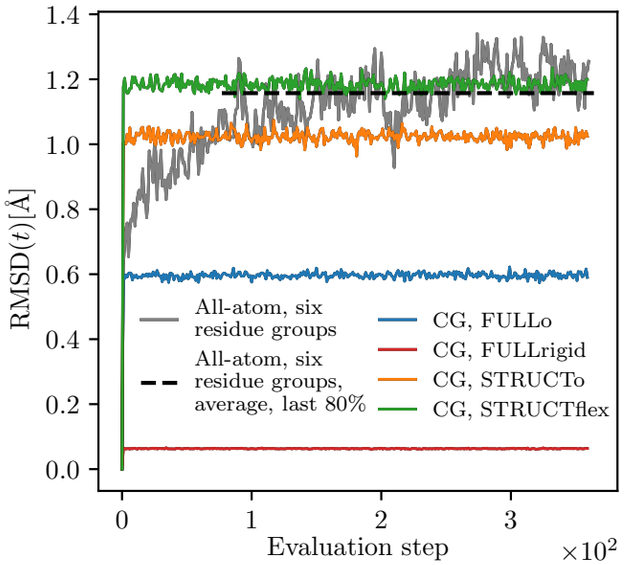


FIG. 4. $\mathrm{RMSD}_{\mathrm{AA}}(t)$ (Eq. (1)) of the COM of the six residue groups defined in Tab. I, averaged over 21 all-atom trajectories where the first 10 ns have been omitted, in total spanning 90 ns (grey), its time-average over the last 80% of time steps, $\overline{\mathrm{RMSD}_{\mathrm{AA}}(t)} = 1.16\,\text{Å}$ (dashed black), as well as $\mathrm{RMSD}_{\mathrm{CG}}(t)$ for four proposed CG models at $\Phi = 0$ (colored), spanning $10\,\tau_{\mathrm{BD}}$. The Brownian time τ_{B} is defined in Eq. (C3). Timescales between AA and CG simulations are not straightforwardly convertible, so we refer to "Evaluation steps" on the x-axis. Note that the result for STRUCTnd is not shown, since its RMSD(t) saturates at approximately 12 Å, much higher than the other models.

will certainly be in the lower third of poor performing topologies. On the other hand, if one compares topologies located the top interval $\overline{\overline{S}} \in [0.86, 1]$ (10% of generated topologies) to an adjacent interval below $\overline{\overline{S}} \in [0.71, 0.86]$ (62% of generated topologies) the former all have significantly less UB-bonds, angles and dihedrals on average (pointing to the mentioned over-parameterization of the FULL connectivity), but for the case of bonds it depends on which bonds are deleted. Here it is found, that those top 10%, on average, have less bonds not reflecting the amino acid sequence of BSA (i.e. (1,3) and (6,2)), but approximately equal numbers of bonds reflecting the sequence (i.e. (1,2), (2,3), etc.), compared to the previously mentioned adjacent, poorer performing set of topologies. This points to the effectiveness of the intuitive choice of selecting the CG bond connectivity according to the amino acid sequence. We refer to the supplementary material for the data supporting these findings.

For completeness, we show in Fig. 4 $\mathrm{RMSD}_{\mathrm{CG}}(t)$ of the models FULLo, FULLrigid, STRUCTo, STRUCTflex together with $\mathrm{RMSD}_{\mathrm{AA}}(t)$. As expected, a model with too many connections or an overestimation of the spring constant can lead to rigid-body-type structures that do not properly capture the elasticity of the original structure. On the other hand, an underestimation of the spring constants seems to better reproduce the flexibility of the all-atom BSA simulation.

As outlined above, the flexibility of CG molecules has been either manipulated by changing the combination of interaction potentials, or by scaling the spring constants by a constant value. A more systematic investigation of the latter approach, where the spring constants of FULLo and STRUCTo have been altered in small steps and $\overline{\overline{S}}$ and ΔRMSD has been calculated, can be found in the supplementary material. To summarize, the findings are as follows: First, the FULLo model can be improved by reducing all spring constants by a factor of 0.4 in order to yield an equally high $\overline{\overline{S}}$ value like STRUCTo, though, from a computational point of view, it is more efficient to utilize STRUCTo over FULLo. Second, reducing all spring constants of the model STRUCTo by a factor of 0.75 (which is per definition equivalent to the model STRUCTflex) minimizes the resulting ΔRMSD, whereas it is not needed to rescale the spring constants of STRUCTo in order to maximize the resulting $\overline{\overline{S}}$. This confirms the well-known observation that there is no unique similarity measurement and that one or several have to be chosen depending on the possible application of the CG model.[57]

We refer to the supplementary material for detailed data on the performance of the various models and topologies with respect to generalized coordinate overlap areas, which are broken up according to bonds, angles and dihedrals, on the number of total force field constituents compared to their performance, on the occurrence probabilities of specific constituents in specific performance intervals, as well as the aforementioned, more detailed investigation of spring constant manip-

Accepted to J. Chem. Phys. 10.1063/5.0132493

ACCEPTED MANUSCRIPT
The Journal of Chemical Physics
AIP Publishing
This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.
PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0132493

ulation.

Additionally, we conducted an Essential Dynamics (ED) analysis[58] on our all-atom and CG trajectories to evaluate how internal dynamics of BSA (with respect to coordinate-covariances) are represented by our CG models. For comparison, we also implemented a simple, fully-bonded ENM, whose spring constant has been chosen such that the coarse-grained mean-squared fluctuation of AA simulations are reproduced. To summarize our findings: The model STRUCTo (which yields the highest $\bar{\bar{S}}$ of all BSA models in Tab. IV) exhibits an amount of total positional fluctuation that lies within the error range of the coarse-grained all-atom fluctuations. STRUCTo also features the best agreement of dynamic cross-correlations compared to AA. The simple ENM gives a reasonable average overlap-parameter ($\bar{\bar{S}} \simeq 0.80$), but fails to reproduce the proper directions of motion (eigenvectors of the coarse-grained positional covariance matrix). Detailed data and description of the performed ED analysis are available in the supplementary material.

C. Coarse-grained model for IgG

The family of Immunoglobulin G (IgG) are Y-shaped molecules about 2.2 times heavier than BSA molecules. The three arms of the "Y" are loosely connected in the hinge region, such that IgG is expected to be more anisotropic and flexible than BSA. The molecule consists of 2×6 domains, building the right and the left part of the "Y", seen from its symmetry axis (see Fig. 5(a)). In Chaudhri *et al.*, a 12-bead coarse-grained model has been defined using a similar methodology as we used here to construct the BSA models (see Fig. 5(b)).[27] Each domain is the center of a bead and intramolecular interactions are parametrized from distributions of the generalized coordinates from AA simulations (except dihedrals). The chosen topology is in between a fully connected and a structure-based topology. The three arms of the "Y" are fully connected within themselves, but the right and the left part of the "Y" are connected by fewer (and weaker) bonds. Dihedrals are introduced to stabilize the 3D structure of the molecule. There has been no investigation of different topologies and their associated metrics $\bar{\bar{S}}$ and ΔRMSD. We employ the model as described in the main text and in the supplementary material of Ref. 27. We denote this model IGGo. From IGGo we derive the model IGGrigid, where all spring constants are increased by a factor of 100 (see Tab. IV).

To quantify molecule compactness of the coarse-grained IgG models compared to the proposed BSA models, we compute the radius-of-gyration

$$R_g = \sqrt{\frac{1}{N_{\text{beads}}} \left\langle \sum_{i=1}^{N_{\text{beads}}} (\vec{r}_i - \vec{r}_{\text{COM}})^2 \right\rangle} \tag{3}$$

in Fig. 6 from CG simulations for different packing fractions (see App. C for simulation details). R_g for both IgG models is considerably higher than for all BSA models. Note that molecules from IGGo are less compact than from IGGrigid for all packing fractions. In the rigid model, the molecules are forced by their increased spring constants to be in an extended equilibrium structure, and this structure does not change under crowding. In contrast, for BSA, all proposed models show a similar compactness, except the one lacking the stabilizing dihedral potentials, STRUCTnd. The equilibrium conformations for this one, along with its flexible counterpart IGGo, are also influenced by packing effects, leading to more compact CG molecules at higher Φ.

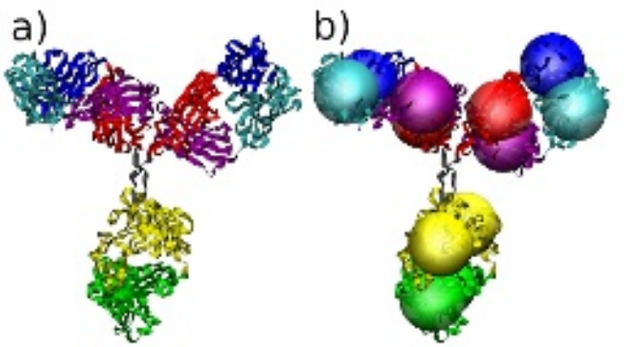


FIG. 5. (a) Tertiary structure[59] of IgG and (b) overlaid CG model representation. The different colors represent the domains of the molecule, VH (blue), CH1 (red), CH2 (yellow), CH3 (green), VL (cyan) and CL (purple). IgG consists of two identical heavy chains (VH-CH1-CH2-CH3) and two identical light chains (VL-CL). Domains are coarse-grained to beads which are equally sized with $\sigma_{\text{bead}}^{\text{IgG}} = 30.38\,\text{Å}$, refer to App. C for details on the coarse grained simulations.

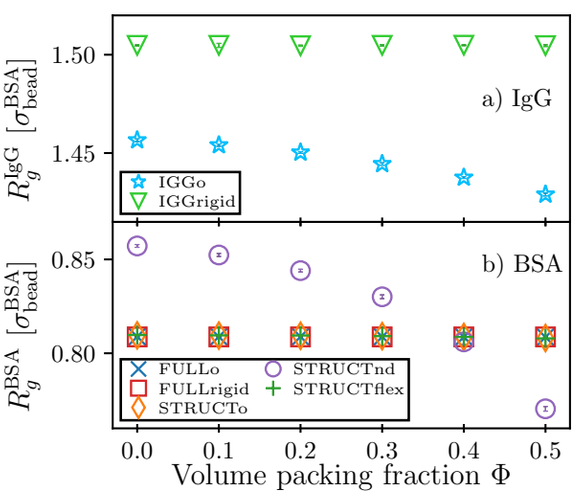


FIG. 6. Radius-of-gyration R_g in coarse-grained simulations for all molecule models for (a) IgG and (b) BSA, according to Eq. (3).

Accepted to J. Chem. Phys. 10.1063/5.0132493

ACCEPTED MANUSCRIPT
The Journal of Chemical Physics
AIP Publishing
This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.
PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0132493

III. PAIR CORRELATIONS IN EQUILIBRIUM

A. Effective isotropic model for BSA via the orientation averaged potential of mean force

Having parametrized several coarse–grained models of BSA, we can investigate properties of dilute and dense solutions using CG simulations. For these, the bead-bead interactions are taken to be purely repulsive in the form of the Weeks-Chandler-Anderson (WCA) potential, see App. C, Eq. (C2). Hence we focus on excluded volume effects which we expect to dominate in the high-salt regime. As described, the coarse-grained model is fairly anisotropic and it is not obvious that a simple isotropic model like (sticky) hard spheres should describe well the structure and dynamics of BSA in solution. With regard to structure, equilibrium pair correlations between centers-of-masses of molecules are a standard measure in the theory of liquids and a good isotropic BSA model should reproduce those of the anisotropic model. As a candidate pair potential for the isotropic model, we employ the free energy for two anisotropic (CG) molecules with their center-of-mass (COM) distance fixed at r, which is described by an effective potential (potential of mean force) $U_{\rm eff}(r)$:[60]

$$\begin{aligned}\beta U_{\rm eff}(r) &= -\ln\left(g_0(r)\right)\\ &= -\ln\left(\int_{\Omega_1}\int_{\Omega_2} e^{-\beta U(r,\Omega_1,\Omega_2)} d\Omega_2 d\Omega_1\right),\end{aligned} \tag{4}$$

where $\beta = 1/(k_{\rm B}T)$ and $g_0(r)$ is the COM-pair-correlation function of a system containing only two molecules (dilute limit), and $U(r,\Omega_1,\Omega_2)$ represents the full interaction energy of the two CG molecules where all the internal degrees of freedom of molecule i are denoted by Ω_i.

We have determined the effective potential for the special case of a rigid molecule (which approximately corresponds to the model FULLrigid). In this case, Ω_i corresponds to the orientation of molecule i (as expressed by three Euler angles for each molecule). The six–dimensional integral is calculated with 15 grid points per angle and Gauss–Legendre quadrature. The effective potential is a very soft, purely repulsive potential, see the inset in Fig. 7(a). It can be well fitted by a Mie-potential[61] which is cut-off at its minimum and shifted to 0. It represents a generalization of a WCA potential and it is given by:

$$U_{\rm Mie}(r) = \begin{cases} C\varepsilon\left[\left(\frac{\sigma_{\rm Mie}}{r}\right)^n - \left(\frac{\sigma_{\rm Mie}}{r}\right)^m\right] + \varepsilon & \text{if } r \leqslant r_{\rm cut} \\ 0 & \text{if } r > r_{\rm cut} \end{cases}, \tag{5}$$

with $C = \frac{n}{n-m}\left(\frac{n}{m}\right)^{\frac{m}{n-m}}$ and $r_{\rm cut} = \left(\frac{n}{m}\right)^{1/(n-m)}\sigma_{\rm Mie}$. We set $\varepsilon = 2k_BT = \varepsilon_{\rm WCA}$ and fit the effective potential to the Mie form between $\sigma_{\rm bead}^{\rm BSA}$ and $r_{\rm cut} = 3.0\,\sigma_{\rm bead}^{\rm BSA}$ (maximum interaction range of the two CG anisotropic molecules). Fitting the exponents gives rather small values, $n \approx 2.36$ and $m \approx 1.07$, and these determine the effective repulsive length scale $\sigma_{\rm Mie}^{\rm BSA} \approx 1.63\,\sigma_{\rm bead}^{\rm BSA}$.

From the effective potential, other effective diameters can be defined. A first, popular choice is the Barker-Henderson diameter:[62]

$$\sigma_{\rm BH}^{\rm BSA} = \int_0^\infty \left[1 - \exp(-\beta U_{\rm eff}(r))\right] {\rm d}r\,. \tag{6}$$

A second one may be termed "B_2–diameter" $\sigma_{\rm B_2}^{\rm BSA}$ where B_2 is the second virial coefficient and $\sigma_{\rm B_2}^{\rm BSA}$ is the diameter of a hard sphere which would have the same B_2 as BSA:

$$\begin{aligned}B_2 &= 2\pi\int_0^\infty \left[1-\exp(-\beta U_{\rm eff}(r))\right] r^2 {\rm d}r\\ &\stackrel{!}{=} \frac{2}{3}\pi\left(\sigma_{\rm B_2}^{\rm BSA}\right)^3\,.\end{aligned} \tag{7}$$

Both diameters are temperature-dependent. For the chosen temperature $\beta\varepsilon = 2.0$ we find $\sigma_{\rm BH}^{\rm BSA}/\sigma_{\rm bead}^{\rm BSA} \approx 2.02$ and $\sigma_{\rm B_2}^{\rm BSA}/\sigma_{\rm bead}^{\rm BSA} \approx 2.08$, both about 25% larger than $\sigma_{\rm Mie}^{\rm BSA}$.

Pair correlations are determined for the anisotropic CG models (see App. C for simulation details) as well as for the effective isotropic system (soft spheres), interacting pairwise via $U_{\rm eff}(r)$, using standard Brownian-dynamics simulations of $N = 2197$ particles. Results are compared for equal number densities ρ of CG molecules and soft spheres. Number densities are converted to BSA packing fractions via $\Phi = \pi(\sigma_{\rm bead}^{\rm BSA})^3\,\rho$.

Fig. 7(a) shows the center-of-mass pair-correlation function $g_{\rm COM}(r)$ for BSA (FULLrigid) as well as for the soft-sphere effective systems for BSA packing fractions between 0.1 and 0.4. The curves at the same Φ show good agreement. Note that the height of the first correlation peak is being exceeded only by $\sim 3\,\%$ by the spheres. This first peak also lies further outward for the anisotropic BSA systems, indicating that the obtained $U_{\rm eff}^{\rm BSA}$ slightly underestimates sterical incompatibilities. Not shown are results for $g_{\rm COM}(r)$ from the other coarse-grained models. However, FULLo, STRUCTo and STRUCTflex differ only marginally from FULLrigid. Only STRUCTnd shows a smaller first peak and consequently weaker oscillations; see the supplementary material. We conclude that the internal degrees of freedom of the CG molecules are not very important in terms of the pair correlations, as long as the overall 3D structure of the molecule is quite stable, and, secondly, the effect of anisotropy is very well captured simply in the softness of the isotropic, repulsive potential in the effective system.

B. Effective isotropic model for IgG

We have computed the effective isotropic potential also for the 12-bead (IGGrigid) model of IgG (with WCA repulsions between the beads as for BSA), the result is shown in the inset of Fig. 7(b). The effective potential $U_{\rm eff}^{\rm IgG}$ is much softer and extends to further distances than the one for BSA. For small COM-distances the potential eventually levels off, indicating the possibility of penetrating molecule configurations. The plateau for small r arises, since the COM is not located inside a bead.

COM pair correlations have been computed for the coarse grained IgG models as well as for the effective soft sphere

Accepted to J. Chem. Phys. 10.1063/5.0132493

ACCEPTED MANUSCRIPT
The Journal of Chemical Physics
AIP Publishing
This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.
PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0132493

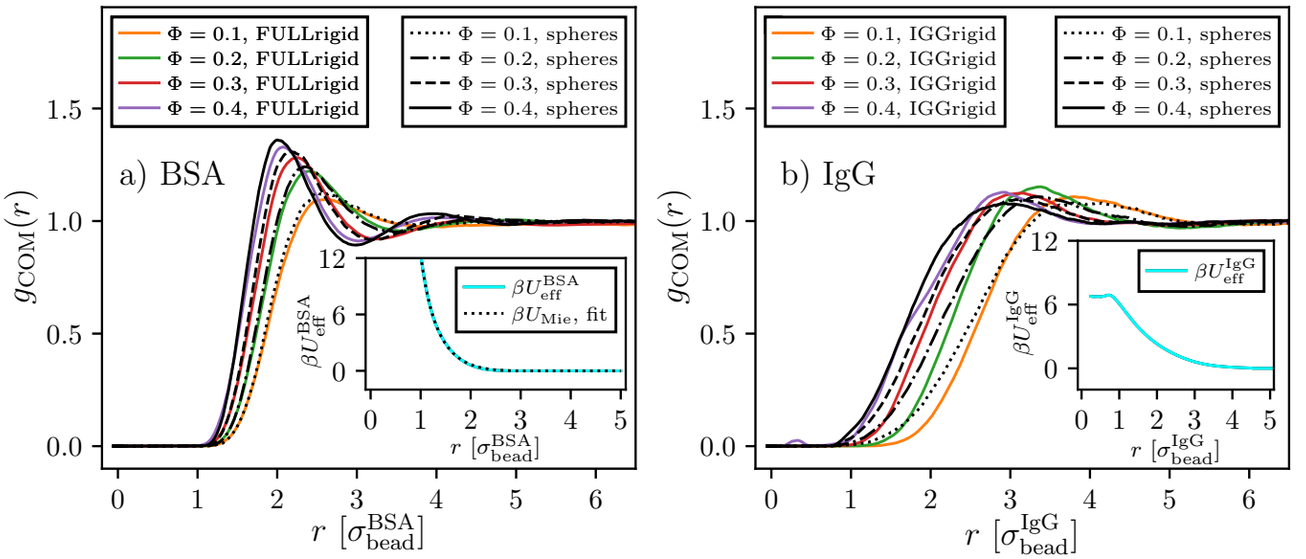


FIG. 7. Center-of-mass pair-correlation function $g_{\mathrm{COM}}(r)$ for models (a) FULLrigid and (b) IGGrigid (colored lines) and soft spheres (black lines). Data has been recorded with a bin width of $\Delta r = 0.01\sigma_{\mathrm{bead}}^{\mathrm{BSA}}$ ($0.01\sigma_{\mathrm{bead}}^{\mathrm{IgG}}$, respectively) and then averaged over ten bin widths. Insets: $\beta U_{\mathrm{eff}(r)}$ obtained by evaluation of Eq. (4) (cyan) and fitted modified Mie-potential Eq. (5) (black dashed line) with $\sigma_{\mathrm{Mie}}^{\mathrm{BSA}} = 1.63\,\sigma_{\mathrm{bead}}^{\mathrm{BSA}}$, $n = 2.36$, and $m = 1.07$. The effective potential for IgG could not be fitted with Eq. (5).

model and compared for equal number densities ρ of molecules. Number densities are converted to IgG packing fractions via $\Phi = 2\pi(\sigma_{\mathrm{bead}}^{\mathrm{IgG}})^3\rho$. The results for soft spheres and IGGrigid at packing fractions 0.1...0.4 are shown in Fig. 7(b). (As before for BSA, we do not observe a significant difference between the coarse-grained models, here IGGo and IGGrigid.) There is only qualitative agreement between the pair correlations of the anisotropic CG model and of the effective soft sphere model. A common feature is that, upon increasing the volume fraction, the first peak moves to smaller distances (expected for a repulsive system), however, this happens without an increase in the peak height (uncommon for a repulsive system and different from BSA). Noticeable disagreement occurs both for lower and higher packing fraction. At lower packing fraction the soft sphere potential is too soft in that $g_{\mathrm{COM}}^{\mathrm{IgG}}(r)$ is considerably overestimated for smaller r. At higher packing fraction $g_{\mathrm{COM}}^{\mathrm{IgG}}(r)$ of the anisotropic CG system begins to develop shoulders, indicative of a second repulsive length scale within the molecule. This is not seen in the results from the soft-sphere system.

For the highest packing fraction $\Phi = 0.4$, in the anisotropic CG model, $g_{\mathrm{COM}}^{\mathrm{IgG}}(r)$ develops a small bump at small r, arising from interpenetrated molecules. This is presumably an artefact, pointing to the shortcomings of the relatively basic, low-resolution, CG force field and specifically having only intermolecular WCA repulsions.

IV. DIFFUSION

Using the coarse-grained simulations described in App. C, we can address the intermediate and long-time diffusion in the Brownian limit. Note that in the Brownian Dynamics simulations each bead has a constant mobility Γ which defines the bead diffusion coefficient $D_{t,0}^{\mathrm{bead}} = k_{\mathrm{B}}T\Gamma$. We did not consider explicit hydrodynamic interactions (HI) in these calculations, but estimate their effect in the case of translational diffusion using a multiplicative ansatz using known results in the short-time limit.

A. Translational diffusion

Fig. 8 shows the translational center-of-mass mean square displacement (MSD) as a function of time (in logarithmic scale) for the STRUCTo and IGGo models at infinite dilution ($\phi = 0$) and at two different volume fractions ($\phi = 0.2$, and 0.4). As expected, for the infinite dilution case the MSD vs. time curve is linear at all times and shows a single slope for all time scales.

For the systems with $\phi > 0$, three regimes are observed: (i) at short times a linear dependence, (ii) at long times another linear dependence with a different slope, and (iii) a non-linear behavior between the short and long-time regimes. At short times, the proteins have not suffered collisions, so the diffusion coefficient (from the slope of the MSD vs. time curve) for any ϕ is the same as for the infinite dilute case (as expected if HI are absent). In contrast, if HI are included, the short-time diffusion coefficient will depend on ϕ as HI slow down

Accepted to J. Chem. Phys. 10.1063/5.0132493

ACCEPTED MANUSCRIPT
The Journal of Chemical Physics
AIP Publishing
This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.
PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0132493

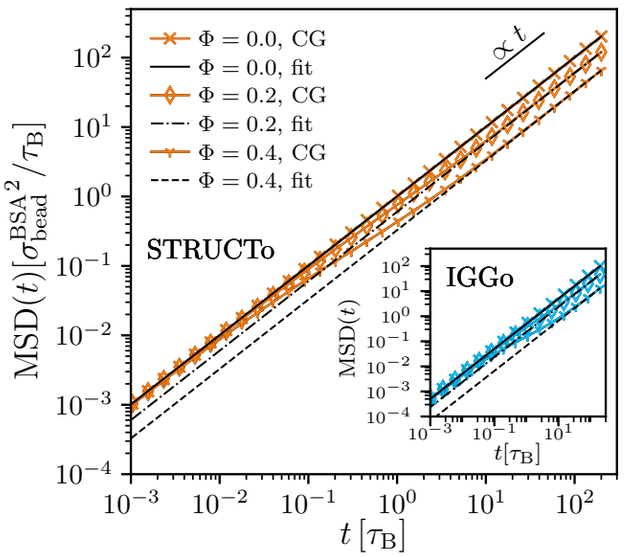


FIG. 8. Ensemble-averaged, center-of-mass MSD(t) of BSA (STRUCTo) for $\phi = 0$, 0.2 and 0.4. Inset: IgG (IGGo) for the same packing-fractions. Black solid/dashed lines correspond to fits to the long-time regime according to Eq. (8).

the motion of the proteins and this slow-down depends on the concentration of molecules.[63] At times between the short and long-time regime, the MSD becomes non-linear over time, which is typical for this transient cross-over regime. Finally, at long times, the diffusion process becomes linear again, but now the diffusion coefficient decreases as Φ increases. Compared to BSA, the slow-down of translational diffusion of IgG (inset of Fig. 8) is more pronounced at equivalent packing fractions due to its protruding, elongated shape. Similar behavior was observed for the other models studied (see supplementary material for all other MSDs).

Fig. 9(a) (BSA) and Fig. 9(b) (IgG) show the normalized, translational, center-of-mass, long-time diffusion coefficients as a function of the volume fraction for the different models studied, obtained via fitting the well known relation

$$\mathrm{MSD}(t) = 6D_t^l t \tag{8}$$

to the long-time regime (ii) of the measured MSD. The normalization is given by the diffusion coefficient in the dilute limit, $D_{t,0}$, which in turn is related to the bead diffusion coefficient by $D_{t,0} = D_{t,0}^{\mathrm{bead}}/N$, where N is the number of beads in the molecule (this is equivalent to the polymeric Rouse model[66]).

For all BSA models, except STRUCTnd, the normalized diffusion coefficients are almost indistinguishable from each other. Also, the data clearly demonstrates the influence of molecular flexibility on translational self-diffusion. Both the two most elastic topologies STRUCTnd (BSA) and IGGo (IgG) diffuse (significantly) faster in crowded conditions than the other coarse-grained models for the respective molecule. This appears to be easily rationalized in view of the molecules' increased ability to deform and bend their structure in response to their immediate environment. Consequently, steric clashes between two molecules are less likely to diminish their translational motion. This flexibility with respect to deformations can be seen well in the reduction of R_g with increasing packing fraction, see Fig. 6. Also, the effect of the pronounced anisotropic distribution of exclusion volume in IgG compared to BSA molecules is captured: At highest packing $\Phi = 0.5$ translational motion in IgG mixtures is extremely slow, whereas for BSA significant diffusion can still be observed.

In order to estimate the effect of hydrodynamic interactions (which in general slow down diffusion), we employ a multiplicative ansatz for the diffusion coefficient, separating short-time and long-time contributions and resulting in a rescaling of our obtained coefficients. Such an ansatz has been shown to deliver accurate results for e.g. hard-sphere systems and also biomacromolecules.[67,68] In detail, let $D_{t,\mathrm{HI}}^l$ resp. $D_{t,\mathrm{HI}}^s$ be the long-time resp. short-time translational diffusion coefficients with full HI. Then the ratio between long-time and short-time coefficients is approximately the same in simulations with full HI and in simulations without HI, i.e. $D_{t,\mathrm{HI}}^l/D_{t,\mathrm{HI}}^s \approx D_t^l/D_{t,0}$.[69,70] Furthermore, the reduced diffusion coefficients $D_t^l/D_{t,0} = \hat{D}_t^{l,\mathrm{hs}}(\Phi_{\mathrm{hs,eff}}^l)$ and $D_{t,\mathrm{HI}}^s/D_{t,0} = \hat{D}_{t,\mathrm{HI}}^{s,\mathrm{hs}}(\Phi_{\mathrm{hs,eff}}^s)$ shall be described by the coefficients from a hard-sphere system with effective packing fractions $\Phi_{\mathrm{hs,eff}}^l = \alpha_i\Phi$ and $\Phi_{\mathrm{hs,eff}}^s = \beta_i\Phi$. Here, α_i and β_i are proportionality constants (not necessarily equal) relating the effective packing fractions to the packing fractions of the dry protein volume and i stands for BSA/IgG. Thus:

$$\hat{D}_{t,\mathrm{HI}}^l(\Phi) \approx \hat{D}_t^{l,\mathrm{hs}}(\alpha_i\Phi)\,\hat{D}_{t,\mathrm{HI}}^{s,\mathrm{hs}}(\beta_i\Phi)\,. \tag{9}$$

The dimensionless coefficients α_i can also be expressed by the ratio

$$\alpha_i = (\sigma_{\mathrm{hs,eff}}^{i,l}/\sigma_{\mathrm{s.v.}}^i)^3, \tag{10}$$

where $\sigma_{\mathrm{s.v.}}^i$ is the diameter of the effective sphere with the same volume as the dry molecule volume calculated from its specific volume, see App. C. Thus, this relation defines an effective hard-sphere diameter for long-time diffusion $\sigma_{\mathrm{hs,eff}}^{i,l}$. Likewise, an equivalent relation for β_i,

$$\beta_i = (\sigma_{\mathrm{hs,eff}}^{i,s,\mathrm{HI}}/\sigma_{\mathrm{s.v}}^i)^3\,, \tag{11}$$

defines an effective hard-sphere diameter for short-time diffusion.

For the translational diffusion in hard-sphere systems analytic solutions are available. First, the long-time diffusion of hard spheres without hydrodynamic interactions can be described via:[69,71]

$$\hat{D}_t^{l,\mathrm{hs}}(\Phi_{\mathrm{hs}}) = \frac{1}{1+2\Phi_{\mathrm{hs}}\chi}, \tag{12}$$

with χ being the hard-sphere pair-correlation contact value in the Percus-Yevick approximation:

$$\chi = \frac{1+(1/2)\Phi_{\mathrm{hs}}}{(1-\Phi_{\mathrm{hs}})^2}. \tag{13}$$

Accepted to J. Chem. Phys. 10.1063/5.0132493

ACCEPTED MANUSCRIPT
The Journal of Chemical Physics
AIP Publishing
This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.
PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0132493

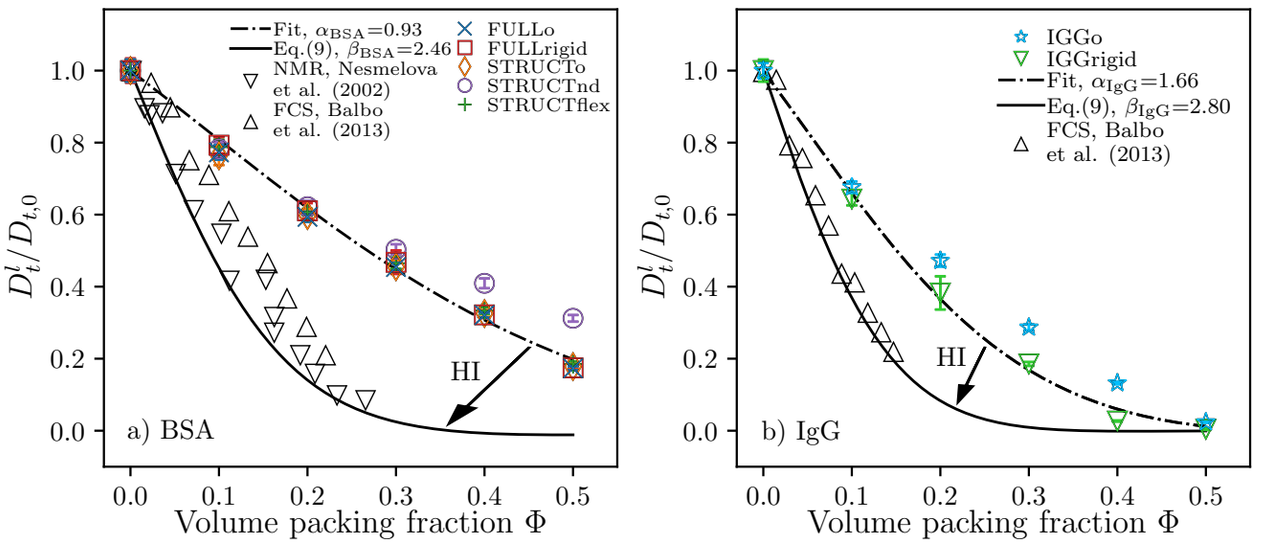


FIG. 9. Normalized, translational, center-of-mass, long-time diffusion coefficients for the different BSA (a) and IgG (b) models proposed. Colored symbols are unrescaled values from coarse-grained BD simulations. The black, dash-dotted lines represent fits of $\hat{D}_t^{l,\mathrm{hs}}(\Phi_{\mathrm{hs}} = \alpha_i\Phi)$ according to Eq. (12) to the FULLrigid and IGGrigid models, respectively. By rescaling these fits according to Eq. (9), one can heuristically include the otherwise neglected hydrodynamic slowing-down for a complete quantitative picture (black solid lines). For that, previously derived values have been used ($\beta_{\mathrm{BSA}} = 2.46$ from Ref. 23 and $\beta_{\mathrm{IgG}} = 2.80$ from Ref. 64). Also shown are experimental values from fluorescence correlation spectroscopy (FCS) measurements (black up-triangles) by Balbo *et al.* and (only for IgG) pulsed-gradient spin-echo NMR measurements (black down-triangles) by Nesmelova *et al.*.[24,65]

By fitting Eq. (12) to the measured diffusion constants for FULLrigid and IGGrigid (with $\Phi_{\mathrm{hs}} = \Phi^l_{\mathrm{hs,eff}}$), it is found that the long-time self-diffusion without HI of the coarse-grained protein models is well described by the effective sphere model (black, dash-dotted lines in Fig. 9(a) and 9(b)), yielding $\alpha_{\mathrm{BSA}} = 0.934$ and $\alpha_{\mathrm{IgG}} = 1.66$. As effective diameters we thus calculate $\sigma^{\mathrm{BSA,l}}_{\mathrm{hs,eff}} = 52.44\,\text{Å}$ and $\sigma^{\mathrm{IgG,l}}_{\mathrm{hs,eff}} = 82.36\,\text{Å}$. Note that we only fitted the rigid models for BSA and IgG, respectively. These effective "dynamic" diameters from our model complement the "static" ones determined before. For a full list see Tab. V and the discussion in Sec. V.

Second, short-time diffusion with hydrodynamic interactions in a hard-sphere system can be described with:[72]

$$\hat{D}^{s,\mathrm{hs}}_{t,\mathrm{HI}}(\Phi_{\mathrm{hs}}) = \frac{1}{1+H(\Phi_{\mathrm{hs}})}, \tag{14}$$

with

$$H(\Phi_{\mathrm{hs}}) = \frac{2b^2}{1-b} - \frac{c}{1+2c} - \frac{bc(2+c)}{(1+c)(1-b+c)}, \tag{15}$$

where $b = \sqrt{\frac{9}{8}\Phi_{\mathrm{hs}}}$ and $c = \frac{11}{16}\Phi_{\mathrm{hs}}$.

Since we do not have simulation results with hydrodynamic interactions, for the determination of β_i we use experimental results from Refs. 23 (BSA) and 64 (IgG). There, translational and rotational short-time diffusion coefficients have been extracted from neutron scattering experiments via an analytical separation method[23,73] and translational short-time diffusion coefficients are described well by Eq. (14) with $\Phi_{\mathrm{hs}} = \beta_i\Phi$ and $\beta_{\mathrm{BSA}} = 2.46$ resp. $\beta_{\mathrm{IgG}} = 2.80$.

We use these values together with Eq. (9) to rescale our diffusion coefficients in order to heuristically include the effect of hydrodynamic interactions, see the solid black line in Fig. 9(a) and Fig. 9(b). The resulting values are in reasonable agreement with experimental measurements.

B. Rotational diffusion

In analyzing rotational diffusion for flexible bodies, one usually has to resort to concepts developed for rotational diffusion of rigid bodies. For these, a principal coordinate system can be defined in which the rotational diffusion tensor of one molecule is diagonal.[74] In practice, the principal coordinate system can be found by analyzing the time-dependent covariance matrix of quaternion coefficients, specifying the orientational state of a rigid molecule.[75] A somewhat simpler estimate of a rotational diffusion constant D_r often employed in the literature (e.g. Refs. 24, 76, and 77) can be obtained by measuring the auto-correlation functions of several unit vectors $\hat{e}_i$, attached to the center-of-mass of each molecule, pointing to each of the beads. D_r can then be obtained via:[78]

$$\langle P_l[\hat{e}_i(t)\cdot\hat{e}_i(t+\tau)]\rangle = \exp\left[-l(l+1)D_r(\Phi)t\right]. \tag{16}$$

Here, P_l is the Legendre polynomial of degree l, where in experiments different techniques probe different relaxation

Accepted to J. Chem. Phys. 10.1063/5.0132493

ACCEPTED MANUSCRIPT
The Journal of Chemical Physics
AIP Publishing
This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.
PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0132493

modes $l = 1, 2$. An average rotational diffusion constant $\overline{D_r}$ is determined by an average over all molecular axes: $\overline{D_r} = (1/N_{\text{axes}}) \sum_{\text{ax}} D_r^{\text{ax}}$.

Note that a more common choice for the unit vectors $\hat{e}_i$ are the protein's principal axes of inertia, but those do not necessarily coincide with the principal axes of rotational diffusion, making this choice somewhat arbitrary.[76] The authors of the latter study also investigated the anisotropic rotational diffusion tensor of a globular protein (ubiquitin) and found that the average over its eigenvalues coincided with the simple, isotropic diffusion coefficient. Also note that Eq. (16) has been originally derived for completely rigid, Brownian rotors. The issues arising from this fact are discussed below.

Fig. 10(a) shows the decay of $\langle P_1[\hat{e}_1(t) \cdot \hat{e}_1(t+\tau)] \rangle$ of axis 1 for FULLrigid (BSA), which can be described well by an exponential fit. De-correlation of the molecules' orientations is fully observed even for the highest packing fractions—very much in contrast to IGGrigid in Fig. 10(b). Here, for $\Phi = 0.5$, the molecules are in an arrested state, not showing any significant rotational movement even for the longest observed times.

Obviously, not all molecular axes are equivalent to each other. For the two rigid models, the correlator for all axes for the specific packing fraction $\Phi = 0.2$ is depicted in the insets of Fig. 10. Here, the deviations between the axes are moderate. To quantify the spread of rotational diffusion between axes, we calculate a relative standard-deviation of diffusion coefficients, averaged over all packing fractions:

$$\langle \sigma_{\text{model}}^{D_r,\text{rel}} \rangle_\Phi = \left\langle \left(\frac{1}{N_{\text{ax}}} \sum_{\text{ax}} \left(D_r^{\text{ax}} - \overline{D_r} \right)^2 \right)^{1/2} \Big/ \overline{D_r} \right\rangle_\Phi \qquad (17)$$

TABLE V. Relative standard-deviation of rotational diffusion coefficients for BSA and IgG, Eq. (17).

CG model	$\langle \sigma_{\text{model}}^{D_r,\text{rel}} \rangle_\Phi$
FULLo	7.80 %
FULLrigid	7.83 %
STRUCTo	8.01 %
STRUCTnd	45.40 %
STRUCTflex	7.94 %
IGGo	13.80 %
IGGrigid	4.27 %

From the results in Tab. V one sees that STRUCTnd for BSA exhibits a rather large variance of the rotational diffusivity about different axes. Further analysis shows that this is mainly because of a single axis, directed to bead 1, whose auto-correlation function is only poorly captured by the exponential fit and whose D_r actually grows with increasing packing fraction (refer to the supplementary material for fits and detailed D_r for all axes). This particular axis connects the COM to bead 1, which sticks out of the molecular plane (compare to Fig. 1(b)). In CG simulations, the equilibrium distance between the molecules' COM and bead 1 is the shortest among all distances of the COM to the beads and it also varies the most, pointing to a rather floppy mode. Since Eq. (16) is technically only valid for rigid objects, we omit this particular axis from the calculation of $\overline{D_r}$ for all BSA models.

Fig. 11 shows the packing fraction dependence of all obtained axis-averaged rotational diffusion constants. The qualitative picture is very similar to the discussed translational diffusion: Flexible topologies rotate faster than rigid ones and the elongated IgG molecules are locked into position earlier than the more globular BSA molecules. Rotational diffusion constants for the $l = 2$ mode are slightly shifted to higher values for all models, but qualitatively show the same picture (see supplementary material).

There is a qualitative difference between rotational diffusion on the one hand and translational diffusion and equilibrium pair correlations on the other hand. For the latter, fits and interpretation in terms of an isotropic colloidal model can be formulated, resulting in an effective diameter as a single parameter (as shown, the procedure worked reasonably well for BSA and somewhat worse for IgG). The packing fraction dependence of the rotational diffusion constant, however, is directly linked to the anisotropy of the molecule since there would be no such dependence for an isotropic model in Brownian Dynamics (if HI are taken into account, such a dependence is present, of course).

To compare translational to rotational slowing-down of diffusion, we present in Fig. 12 the ratio of normalized diffusion constants, $\left(D_t^l / D_{t,0}\right) / \left(\overline{D_r} / \overline{D_{r,0}}\right)$. The data suggests that for both molecules, translational motion for low to intermediate packing fractions is more likely to be hindered by crowding than rotational motion, whereas for the highest packing fractions those differences level out again. This diffusional imbalance is more pronounced for BSA than for IgG. Recent neutron scattering experiments on bovine γ-globulin (antibodies which are similar to IgG) revealed that short-time diffusion of the lobes of the antibodies is less affected by crowding than the global diffusion of the molecules.[79] Although this is not directly comparable to our long-time diffusion constants, lobe diffusion is also quantified in our simulations by $\overline{D_r}$ and global diffusion by D_t^l. Hence, experiments on short-time and our simulations on long-time diffusion come to a similar, qualitative picture.

V. SUMMARY AND OUTLOOK

In this study a low-resolution, 6-bead, domain-based, flexible, coarse-grained model for BSA has been formulated. The corresponding CG force field parameters have been determined via direct Boltzmann inversion using configurations from fully atomistic MD simulations. Subsequently, the "fitness" of many different CG models (combinations of force field components, or topologies) has been quantified via two simple metrics, comparing equilibrium structure and dynamics of coarse-grained, overdamped Brownian dynamics and all-atom simulations. As a result, the coarse-grained motion from atomistic simulations proved to be represented to a very good extent by an intuitive CG model, where beads are connected according to the amino acid sequence of BSA.

Using Brownian dynamics simulations of the CG model,

Accepted to J. Chem. Phys. 10.1063/5.0132493

ACCEPTED MANUSCRIPT
The Journal of Chemical Physics
AIP Publishing
This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.
PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0132493

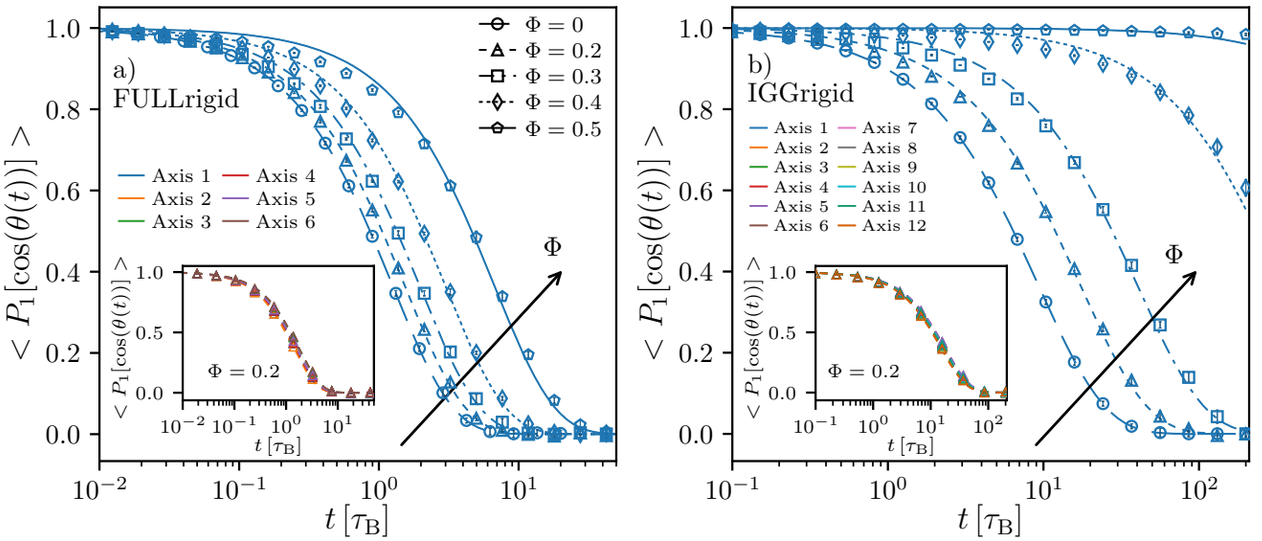


FIG. 10. Ensemble-averaged, rotational auto-correlation function of (a) FULLrigid and (b) IGGrigid in coarse-grained simulations (left-hand side of Eq. (16)) for $l = 1$ vs. time together with fits (solid and dashed lines, right-hand side of Eq. (16)). Main plot: Data of axis 1 (center-of-mass to bead no. 1) for various packing fractions. Insets: Data of all axes for specific packing fraction $\Phi = 0.2$. Data points have been thinned by a factor of four for better visibility.

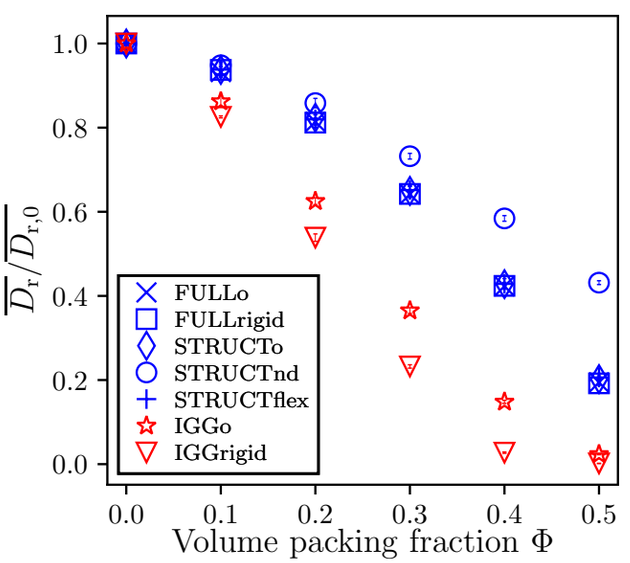


FIG. 11. Normalized, rotational diffusion coefficients obtained via fits of Eq. (16) to the coarse-grained simulations of all BSA/IgG models for the $l = 1$ mode, averaged over all defined molecular axes, except axis 1 for BSA.

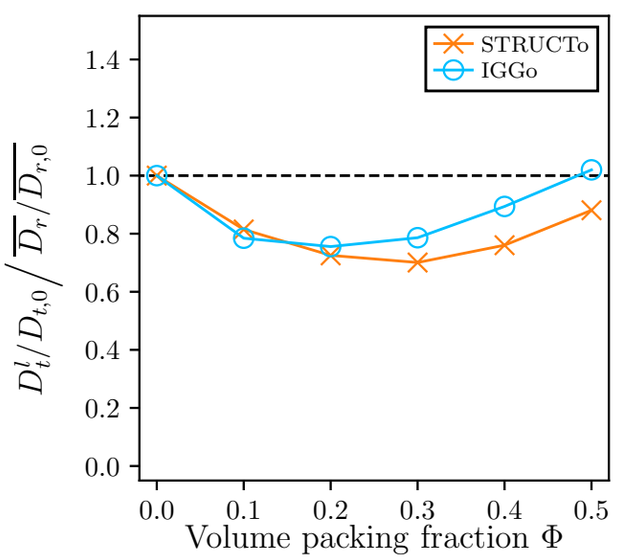


FIG. 12. Ratio of normalized translational and rotational diffusion coefficients $\left(D_t^l/D_{t,0}\right) / \left(\overline{D_r}/\overline{D_{r,0}}\right)$ for BSA (STRUCTo) in orange and IgG (IGGo) in cyan.

a variety of solution properties have been calculated and their dependence on molecular flexibility examined. For that purpose, modified BSA topologies were introduced which have been artificially made more elastic/rigid compared to the optimal structure. As a comparison to a more elongated CG molecule, we implemented the 12-bead IgG model of Chaudhri *et al.*.[27]

By calculating an orientation-averaged, effective potential U_{eff} for the CG models we showed that the equilibrium structure of coarse-grained BSA solutions could be well represented by effective isotropic spheres, up to relatively high den-

Accepted to J. Chem. Phys. 10.1063/5.0132493

ACCEPTED MANUSCRIPT
The Journal of Chemical Physics
AIP Publishing
This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.
PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0132493

sities. This procedure was less successful for IgG, which exhibits a too high degree of anisotropy to be captured by its effective potential. Measurements of radii-of-gyration revealed, that the most flexible molecule models become more compact with increasing packing fraction.

We also investigated the translational and rotational diffusivity of our CG models. Here, extracted normalized diffusion constants of BSA were always larger than those of IgG, however, molecular flexibility and connectivity also affected the mobility of the proteins, where molecules from floppy models diffused faster than molecules from rigid models.

Various effective diameters could be determined from the definition of the CG model itself, the equilibrium angular-averaged, effective potential and the translational diffusion coefficient. From the model itself one may define (i) $\sigma^{\mathrm{BSA}}_{\mathrm{s.v.}}$, which represents the volume of the equivalent sphere that includes the six bead volumes, determined by the specific volume of BSA (Eq. (C1)). From the equilibrium angular-averaged, effective potential one may define (ii) the "Mie diameter" $\sigma^{\mathrm{BSA}}_{\mathrm{Mie}}$ (see Eq. (5)), (iii) the Barker-Henderson diameter $\sigma^{\mathrm{BSA}}_{\mathrm{BH}}$ (see Eq. (6)) and (iv) the "B_2-diameter" $\sigma^{\mathrm{BSA}}_{\mathrm{B_2}}$ (see Eq. (7)). The long-time translational diffusion constant from Brownian dynamics can be interpreted (v) with an effective hard sphere diameter $\sigma^{\mathrm{BSA,l}}_{\mathrm{hs,eff}}$ (see Eq. (10)). These values can be compared with other values from the literature. The experimentally extracted short-time diffusion coefficient can be fitted to the one of a hard sphere system, resulting in (vi) the diameter $\sigma^{\mathrm{BSA},s,\mathrm{HI}}_{\mathrm{hs,eff}}$ (see Eq. (11)).[23] Ellipsoidal form factor fits in scattering experiments result in a effective diameter (vii) from small-angle X-ray scattering and (viii) from small-angle neutron scattering.[12,80] Finally, (ix), a "B_2-diameter" $\sigma^{\mathrm{BSA}}_{\mathrm{B_2,AA}}$ was defined in Ref. 81 from an all-atom representation of BSA where atoms interact purely hard. The values of these 9 diameters are collected in Tab. V.

TABLE VI. BSA molecule diameters σ^{BSA} obtained trough various methods described in the text.

BSA molecule diameter		Result	Numeric value [Å]
(i)	$\sigma^{\mathrm{BSA}}_{\mathrm{s.v.}}$	$1.82\,\sigma^{\mathrm{BSA}}_{\mathrm{bead}}$	53.65
(ii)	$\sigma^{\mathrm{BSA}}_{\mathrm{Mie}}$	$1.63\,\sigma^{\mathrm{BSA}}_{\mathrm{bead}}$	48.12
(iii)	$\sigma^{\mathrm{BSA}}_{\mathrm{BH}}$	$2.02\,\sigma^{\mathrm{BSA}}_{\mathrm{bead}}$	59.63
(iv)	$\sigma^{\mathrm{BSA}}_{\mathrm{B_2}}$	$2.08\,\sigma^{\mathrm{BSA}}_{\mathrm{bead}}$	61.40
(v)	$\sigma^{\mathrm{BSA,l}}_{\mathrm{hs,eff}}$	$1.78\,\sigma^{\mathrm{BSA}}_{\mathrm{bead}}$	52.44
(vi)	$\sigma^{\mathrm{BSA},s,\mathrm{HI}}_{\mathrm{hs,eff}}$	-	$(72.4)^{23}$
(vii)/(viii)	$\sigma^{\mathrm{BSA}}_{\mathrm{SAXS}}$ / $\sigma^{\mathrm{BSA}}_{\mathrm{SANS}}$	-	$(66.8\,/\,63.6)^{12,80}$
(ix)	$\sigma^{\mathrm{BSA}}_{\mathrm{B_2,AA}}$	-	$(69.4)^{81}$

The effective molecule diameters (i)-(v) from the model are smaller than the literature values (vi)-(ix). In fact, the determination of the bead-size $\sigma^{\mathrm{BSA}}_{\mathrm{bead}}$ of our model is based on the partial specific volume fraction v_2^0 of the pure BSA phase, given in the literature (see App. C), which neglects hydration effects stemming from solute-solvent interactions, and therefore should be seen as a lower boundary for estimating the exclusion volume. The bead size directly sets the range of the repulsive WCA interactions between the beads, and thus all other static and dynamic diameters can be expected to be somewhat too small if the bead size is too small. Therefore, in drawing conclusions regarding a systematic error in defining an effective diameter of BSA, one should consider the spread of just the model values derived in this study (approximately $\pm 10\%$ around a mean of about 5.5 nm). This spread reflects the partial success in colloidal modelling with isotropic systems. On the other hand, a systematic uncertainty of about 10% is still considerable in some contexts.

As an outlook for further research, several options appear to be interesting. First, investigations of BSA solutions with trivalent cations have revealed effective, patchy-like attractions, resulting in liquid-liquid phase separation and interesting diffusive dynamics.[15,18,82–84] Previous theoretical work employed Wertheim theory for patchy spheres.[85] In the coarse-graining approach, suitable charge and generally more realistic bead-bead interactions need to be formulated, and it would be interesting to investigate the interplay of patchiness and anisotropy, which is known to influence diffusion of proteins to a great extent.[86] Secondly, BSA can be used as a crowder or tracer in multicomponent systems. Recent work has investigated diffusion in a binary mixture of BSA and IgG, with an interpretation in terms of a binary hard sphere system being quite sensible.[87] An explicit investigation using the coarse-grained models could provide us with more insight on these findings. In addition, our CG model for BSA together with the established CG model for IGG opens opportunities for more realistic simulations of dynamical processes in solutions modelling blood, as BSA can be used as substitute for human serum albumin (HSA), which is the most abundant protein in blood. Finally, the possibility to "calibrate" the short-time diffusion by experimental hydrodynamic interactions and subsequently obtain experimentally confirmed long-time diffusion coefficients from the simulations points towards more accurate predictions of long-time diffusive dynamics in protein solutions in the future.

SUPPLEMENTARY MATERIAL

Generalized coordinate histograms of all-atom simulations; force field constants; $\overline{S}_{\mathrm{type}}$ diagrams; force field component statistics and occurrence probabilities; $\overline{\overline{S}}$ and ΔRMSD behavior upon spring constant rescaling; displacement covariance matrices and Essential Dynamics (ED) analysis; $g_{\mathrm{COM}}(r)$ and $\mathrm{MSD}(t)$ for all models; rotational correlators for STRUCTnd; D_r for all molecular axes, $\overline{D_r}$ for $l = 1, 2$.

ACKNOWLEDGEMENTS

The authors thank DFG and ANR for funding through the joint DFG-ANR project "ImmunoglobulinCrowding" (DFG SCHR 700/28-1, ANR-16-CE92-0009). Moreover, continued support by the DFG and ANR (ANR-21-CE06-0047), as well as by the BMBF (ErUM-pro 05K19VTB and 05K22VTA) is gratefully acknowledged. F. Hirschmann acknowledges an ILL student placement in the theory group.

Accepted to J. Chem. Phys. 10.1063/5.0132493

ACCEPTED MANUSCRIPT
The Journal of Chemical Physics
AIP Publishing
This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.
PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0132493

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Frank Hirschmann: Conceptualization (supporting); Data Curation (lead); Formal Analysis (lead); Investigation (lead); Methodology (equal); Software (lead); Validation (lead); Visualization (lead); Writing - original draft (equal). **Hender Lopez**: Conceptualization (equal); Investigation (supporting); Methodology (equal); Project Administration (supporting); Resources (supporting); Software (supporting); Supervision (equal); Validation (supporting); Visualization (supporting); Writing - original draft (equal). **Felix Roosen-Runge**: Conceptualization (supporting); Methodology (supporting); Writing - review & editing (equal). **Tilo Seydel**: Conceptualization (supporting); Funding Acquisition (equal); Writing - review & editing (equal). **Frank Schreiber**: Conceptualization (supporting); Funding Acquisition (equal); Writing - review & editing (equal). **Martin Oettel**: Conceptualization (equal); Formal Analysis (supporting), Funding Acquisition (equal); Investigation (supporting); Methodology (equal); Project Administration (lead); Resources (lead); Software (supporting); Supervision (equal); Validation (supporting); Writing - original draft (equal).

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding authors upon reasonable request.

Appendix A: All-atom simulations of BSA

Molecular dynamics simulations were conducted with the GROMACS software package.[46] The CHARMM27 force field[88] was used and missing hydrogen atoms where added using the GROMACS pdb2gmx[46] tool. The simulations were performed in a cubic box of 12.1 nm length using periodic boundary conditions and a Verlet neighbor list. The BSA molecule (PDB ID: 4F5S)[1] was placed in the simulation box and TIP3P water was added as solvent (55022 water molecules were added). The total net charge of BSA was $-16e$, so 16 Na^+ ions were added to neutralize the charge of the system. In addition, 107 Na^+ and 107 Cl^- ions were added to mimic physiological salt concentration (0.1 mol L^{-1}). The system was first relaxed using a steepest descend energy minimization algorithm with the solvent constrained. Subsequently, the water constraint was released and the system was equilibrated with a leap-frog integrator in the NVT ensemble by velocity rescaling[89] with a coupling time constant of 0.1 ps for 1 ns in total with an integration time step of 2 fs, followed by another 1 ns in the NPT ensemble with the same thermostat and a Parinello-Rahman barostat[90] with a coupling time constant of 2 ps and a reference pressure of 1 bar. Long-range electrostatics were treated with a Particle Mesh Ewald method[91] using a Coulomb cutoff of 1.4 nm. For the van der Waals forces we used a cutoff distance of 1.4 nm. Subsequently, a total of 21 production runs were conducted, each of them lasting 100 ns with an integration time step of 2 fs. The production runs were conducted in the NPT ensemble and the reference temperature of all thermostats was set to 300 K.

Appendix B: CG spring constants from direct Boltzmann inversion

The functional form of the three types of intramolecular potentials (bond, angle, dihedral) are:

$$V(r)_{\text{bond}} = k_{\text{bond}}(r - r_0)^2, \tag{B1}$$

$$V(\theta)_{\text{angle}} = k_{\text{angle}}(\theta - \theta_0)^2 + k_{\text{UB}}(r - r_{0,\text{UB}})^2, \tag{B2}$$

$$V(\phi)_{\text{dihedral}} = k_{\text{dihedral}}(\phi - \phi_0)^2, \tag{B3}$$

where r is the distance between two beads in a bond, θ is the angle between three beads, and ϕ is the dihedral angle in a set of four beads. The spring constants k and the equilibrium values for the distances, $r_0, r_{0,\text{UB}}$ and the angles θ_0 and ϕ_0 are parameters of the CG-model and as mentioned above are obtained from all-atom molecular dynamics simulations of BSA. The process to determine these parameters can be briefly summarized as follows: (i) the motion of the center-of-mass of the six residue groups reported in Tab. I are tracked in a set of trajectories (for details see App. A) and (ii) this information is then used to calculate histograms of the generalized coordinates q, i.e. r, r_{UB}, θ and ϕ. These histograms correspond to the probability density functions $\rho(q)$ of the generalized coordinates and they are related to the interaction potentials $V(q)$ by:[27]

$$V(q) = -k_{\text{B}}T \ln[\rho(q)] + \text{const}, \tag{B4}$$

where T is the temperature and k_{B} is the Boltzmann constant. By assuming harmonic interaction potentials as well as independent and uncorrelated movement of residue groups, the probability distribution ρ of a generalized coordinate q takes a Gaussian form:

$$\rho(q) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(q-\mu)^2}{2\sigma^2}\right), \tag{B5}$$

with μ being the mean value of q and σ^2 its variance. Taking above relation with Eq. (B4) yields:

$$V(q) = \frac{k_B T}{2\sigma^2}(q-\mu)^2 + \text{const}, \tag{B6}$$

with the spring constant given by:

$$k_{\text{spring}} = \frac{k_B T}{2\sigma^2}. \tag{B7}$$

Accepted to J. Chem. Phys. 10.1063/5.0132493

ACCEPTED MANUSCRIPT
The Journal of Chemical Physics
AIP Publishing
This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.
PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0132493

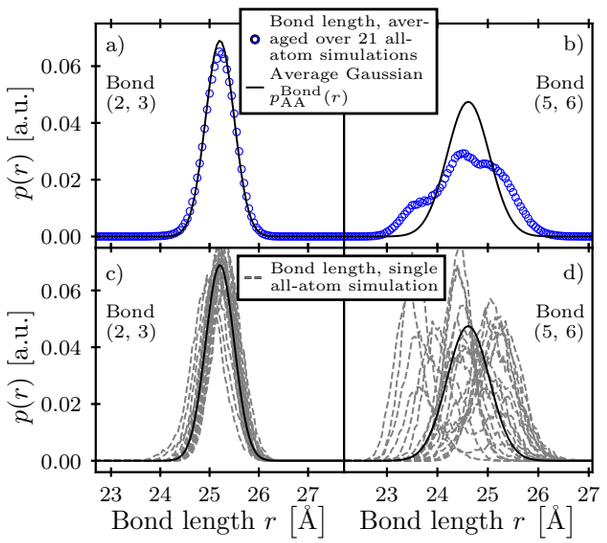


FIG. 13. Bond length histograms obtained for bonds (2,3) (subfigures (a) and (c)) and (5,6) (subfigures (b) and (d)) from all-atom simulations. Blue symbols correspond to the bond length distribution averaged over all trajectories. The grey, dashed lines are histograms obtained from single all-atom trajectories. The black solid lines correspond to the Gaussians $p_{AA}^{Bond\ (2,3)}$ ($p_{AA}^{Bond\ (5,6)}$, respectively), obtained by using the means and variances of the aforementioned histograms.

Multiple independent all-atom simulations (21 in total) yield multiple values for μ and σ^2 for one and the same generalized coordinate. By averaging these means and variances over all conducted simulations, we obtain an averaged Gaussian probability distribution p_{AA}^{j} for each generalized coordinate j, from which the final, averaged means and spring constants are calculated via Eq. (B7).

We illustrate typical results from this procedure in the following. Fig. 13(a) and Fig. 13(b) show the average histograms of distance between two beads in a bond (average over different all-atom trajectories) for the bonds (2,3) and (5,6) in blue circles. For both cases, the corresponding Gaussians $p_{AA}^{Bond\ (2,3)}$ and $p_{AA}^{Bond\ (5,6)}$ (black solid lines), which are obtained by averaging the means and variances of single histograms, are acceptable, but as Fig. 13(b) shows, deviations from a normal distribution are possible. To analyze these deviations, Fig. 13(c) and Fig. 13(d) show the histograms for each trajectory for the bonds (2,3) and (5,6), respectively. For the first case, the distributions for all trajectories show similar mean values and variances, while for the second case the mean values and the variances show a greater dispersion. This result highlights the importance of individually assessing each potential proposed and that adequate sampling is used. The histograms, averages and the parameters obtained for all the proposed potentials can be found in the supplementary material. For the coarse-grained IgG model the spring constants and equilibrium values of Ref. 27 have been used.

Appendix C: Coarse-grained simulation details

To investigate the diffusion and dynamics of both molecules (BSA and IgG) under self-crowding conditions, we perform overdamped Brownian dynamics simulations[92] of the CG models in NVT ensemble using the software package HOOMD-blue.[93,94] The bead diameters σ_{bead}^{i} (i = BSA or IgG) have been determined by calculating the total volume of the proteins based on their partial specific volume $v_{2,i}^{0}$ and by assuming that all CG sites of the protein occupy the same volume. In this way, the bead diameters are given by:

$$\sigma_{bead}^{i} = 2\left(\frac{m_i v_{2,i}^{0}}{N_i}\frac{3}{4\pi}\right)^{1/3}, \tag{C1}$$

where m_i is the molecular weight of the protein, and N_i is the number of CG beads (6 for BSA and 12 for IgG). The values used in the calculations were: $v_{2,BSA}^{0} = (0.733\,mL/g)$[95], $m_{BSA} = 66.4\,kDa$, $v_{2,IgG}^{0} = (0.729\,mL/g)$[96] and $m_{IgG} = (145.5\,kDa)$[27]. With these parameters the obtained diameters are $\sigma_{bead}^{BSA} = 29.52\,Å$ and $\sigma_{bead}^{IgG} = 30.38\,Å$. Notice that the estimated sizes of the beads are in some cases larger than the bond lengths, which causes a slight overlapping of beads. Using a Monte Carlo algorithm we verified that the overlap fraction for both proteins, $V_{BSA}^{overlap}/V_{BSA} \approx 3.2\%$ and $V_{IgG}^{overlap}/V_{IgG} \approx 3.5\%$, is small so this condition has a negligible effect on the diffusion. We use a cubic simulation box of side length $L = (N_{molec}N_i\pi/(6\Phi))^{1/3}\,\sigma_{bead}^{i}$ where Φ is the volume packing fraction. All simulations were performed with periodic boundary conditions and they contained $N_{molec} = 343$ molecules for both BSA and IgG. We assume a temperature of 300 K and for convenience we define the units of energy as $k_BT = 1$, where k_B is the Boltzmann constant.

In the CG simulations, proteins only interact via excluded volume. These intermolecular interactions are modelled by the use of a shifted Weeks-Chandler-Andersen (WCA) potential. For two beads of two different proteins the WCA potential is given by:

$$V_{WCA}(r) = \begin{cases} 4\varepsilon\left[(d/r)^{12} - (d/r)^{6} + \frac{1}{4}\right] & \text{if } r \leqslant 2^{1/6}d \\ 0 & \text{if } r > 2^{1/6}d \end{cases}, \tag{C2}$$

with $d = 2^{-1/6}\sigma_{bead}^{i}$. All beads for the CG models are considered to be of the same size. For the strength of the repulsion we use $\varepsilon = 2k_BT$. No electrostatic nor hydrodynamic interactions have been considered. For systems at infinite dilution, $\Phi = 0$, the intermolecular interactions were disabled and a finite L has been chosen.

Time units are defined by the diffusion coefficients of free beads. Specifically, the simulation time unit is defined based on the Brownian time (τ_B) as follows:

$$\tau_B = \frac{\sigma_{bead}^{2}}{\Gamma k_BT} = \frac{\sigma_{bead}^{2}}{D_{t,0}^{bead}}, \tag{C3}$$

which allows the particle mobility Γ to be absorbed (i.e. it is set to 1 for all CG sites). One τ_B is approximately the time a

Accepted to J. Chem. Phys. 10.1063/5.0132493

ACCEPTED MANUSCRIPT
The Journal of Chemical Physics
AIP Publishing
This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.
PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0132493

free bead (with translational diffusion coefficient $D_{t,0}^{\text{bead}}$) needs to travel its own diameter. A fixed integration time step has been chosen and set to $\Delta t = 2\times10^{-7}...1\times10^{-5}\,\tau_{\text{B}}$, whereas for the rigid topologies a smaller time step has been selected in order to correctly sample steeper potentials.

A possible approach[97] to approximately convert simulation time to a realistic time scale, is to equate the experimentally measured translational diffusion constant $D_{t,0}^{\text{exp}}$ of the protein at dilute conditions to the one present in simulations, $D_{t,0}^{\text{sim}} = D_{t,0}^{\text{bead}}/N = \sigma_{\text{bead}}^2/\tau_{\text{B}}N$, where N is the number of beads in the molecule (Rouse model[66]). By taking $D_{t,0}^{\text{exp,BSA}} \simeq (6.1\times10^{-11}\,\text{m}^2/\text{s})$[98] and $D_{t,0}^{\text{exp,IgG}} \simeq (4\times10^{-11}\,\text{m}^2/\text{s})$[99] one derives $\tau_{\text{B}}^{\text{BSA}} \simeq 24\,\text{ns}$ and $\tau_{\text{B}}^{\text{IgG}} \simeq 19\,\text{ns}$.

For each state point 6 - 25 independent systems have been equilibrated and subsequent production runs lasted 200 - 1000 τ_{B}. For dense systems in equilibration the simulation box has been scaled down from a larger box to the desired packing fraction to avoid incompatible starting configurations. Error bars in graphs show the sample standard deviation of ensemble averaged values of independent systems.

Appendix D: Selection of an optimal topology for the CG model

To judge the quality of the CG models specified in the main text, we introduce a quantity $S \in [0,1]$, which acts as a metric of comparison between the averaged center-of-mass movement of the six residue groups in all-atom and bead movement in coarse-grained simulations. More precisely, it is the common area of two corresponding discretized and normalized coordinate probability distributions $p_{\text{AA}}^j(q^{(j)})$ and $p_{\text{CG}}^j(q^{(j)})$ of a generalized coordinate $q^{(j)}$ with the $q^{(j)}$-axis (j enumerates the generalized coordinate):

$$S_j = \sum_{i=1}^{N_{\text{bins}}} \min\left(p_{\text{AA}}^j(q_i^{(j)}), p_{\text{CG}}^j(q_i^{(j)})\right). \tag{D1}$$

Here, p_{CG}^j and p_{AA}^j were discretized on the same grid for bonds ($\Delta r = 0.05\,\text{Å}$, $r \in [10\,\text{Å}, 160\,\text{Å}]$), angles ($\Delta\theta = 0.3°$, $\theta \in [0°, 180°]$) and dihedrals ($\Delta\phi = 0.3°$, $\phi \in [-180°, 180°]$). Randomly generated topologies yielding bond lengths outside the grid interval were considered unstable and thus omitted. The p_{CG}^j were recorded in CG simulations of non-interacting molecules (see below) and p_{AA}^j were obtained by mapping the calculated average Gaussians (from all-atom simulations) of Eq. (B5) onto the grid. The probability distributions were then normalized.

For illustration, see Fig. 14 for an exemplary S for $j = \text{Bond}(1,3)$. Further, we calculate the mean $\bar{S}$ of all generalized coordinates of the same type:

$$\bar{S}_{\text{type}} = \frac{1}{N_{\text{type}}}\sum_{j=1}^{N_{\text{type}}} S_j, \tag{D2}$$

with type $\in$ [Bonds, UB-Bonds, Angles, Dihedrals], as well as their collective mean $\bar{\bar{S}}$:

$$\bar{\bar{S}} = \sum_{\text{type}} N_{\text{type}}\bar{S}_{\text{type}} \Big/ \sum_{\text{type}} N_{\text{type}}. \tag{D3}$$

Hereby, the calculation of $\bar{S}$ and $\bar{\bar{S}}$ always includes the full set of all recorded generalized coordinate distributions of the FULL topology, as specified in Tab. II. In the generation of random topologies described in the text, only interaction potentials are disabled, but the corresponding generalized coordinate is not taken out of the calculation of $\bar{S}$.

The other measure for an optimal model is based on the commonly used root-mean-square deviation (RMSD, see Eq. (1)), of i) the positions of coarse-grained beads for CG simulations and ii) the centers-of-mass of the six residue groups defined in Tab. I for all-atom simulations. The optimal trans- and rotational superposition of the structure at time t to itself at $t = 0$ has been conducted via a quaternion based algorithm[100,101] implemented in the molecular dynamics software package MDAnalysis,[102,103] which we also use to process simulation trajectories.

The average $\langle\cdot\rangle$ in Eq. (1) is an ensemble average, which, in all-atom simulations, comes from averaging over 21 independent trajectories of well equilibrated BSA molecules which have different starting positions for the centers-of-mass of the six residue groups. The first 10 ns of these trajectories have been omitted and the RMSD was sampled every 0.25 ns. See App. A for details on all-atom simulations.

In CG simulations, it is realized as follows: For each of the 4735 individual coarse-grained models, four independent systems, each containing 27 non-interacting ($\Phi = 0$) molecules of the respective model, are placed in a simulation box and, after equilibration, the RMSD(t) for each individual molecule is calculated. These RMSDs are then averaged. In this exact same manner all coarse-grained probability histograms $p_{\text{CG}}(q)$ of Eq. (D1) are calculated. Production runs with a time step of $\Delta t = 10^{-6}\,\tau_{\text{B}}$ lasted 10 τ_{BD}, whereby observables have been sampled every 5000 time steps.

[1] A. Bujacz, "Structures of bovine, equine and leporine serum albumin," Acta Crystallographica Section D Biological Crystallography **68**, 1278–1289 (2012).
[2] A. G. Ribeiro, J. E. F. Alves, J. C. S. Soares, K. L. dos Santos, Í. T. T. Jacob, C. J. da Silva Ferreira, J. C. dos Santos, R. D. S. de Azevedo, S. M. V. de Almeida, and M. d. C. A. de Lima, "Albumin roles in developing anticancer compounds," Medicinal Chemistry Research **30**, 1469–1495 (2021).
[3] M. Roche, P. Rondeau, N. R. Singh, E. Tarnus, and E. Bourdon, "The antioxidant properties of serum albumin," FEBS letters **582**, 1783–1787 (2008).
[4] T. Peters Jr, "Serum albumin: recent progress in the understanding of its structure and biosynthesis." Clinical chemistry **23**, 5–12 (1977).
[5] A. A. Spector, "Fatty acid binding to plasma albumin," Journal of Lipid Research **16**, 165–179 (1975).
[6] A. N. Beris, J. S. Horner, S. Jariwala, M. Armstrong, and N. J. Wagner, "Recent advances in blood rheology: A review," Soft Matter (2021).
[7] M. Roos, M. Ott, M. Hofmann, S. Link, E. Roössler, J. Balbach, A. Krushelnitsky, and K. Saalwächter, "Coupling and decoupling of rotational and translational diffusion of proteins under crowding conditions," Journal of the American Chemical Society **138**, 10365–10372 (2016).
[8] S. Amin, G. V. Barnett, J. A. Pathak, C. J. Roberts, and P. S. Sarangapani, "Protein aggregation, particle formation, characterization & rheology," Current Opinion in Colloid & Interface Science **19**, 438–449 (2014).

Accepted to J. Chem. Phys. 10.1063/5.0132493

ACCEPTED MANUSCRIPT
The Journal of Chemical Physics
AIP Publishing
This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.
PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0132493

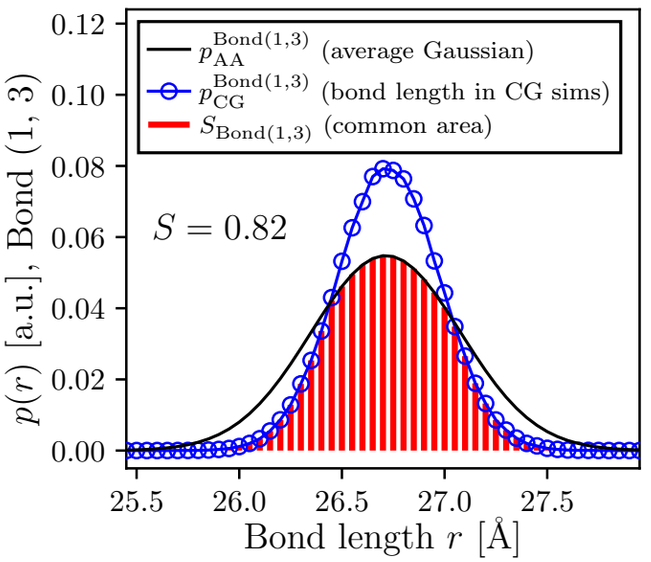


FIG. 14. Schematic calculation of the quantity S (Eq. (D1)) for bond (1,3). The black curve is the averaged Gaussian probability distribution $p_{\mathrm{AA}}(r)$ for the bond, obtained via direct Boltzmann inversion of all-atom trajectories (see App. C). The blue line is the probability histogram $p_{\mathrm{CG}}(r)$ of the measured bond length in the coarse-grained simulations. The common overlap-area S between the two distributions is depicted in red.

[9]S. Loveday, L. K. Creamer, H. Singh, and M. Rao, "Phase and rheological behavior of high-concentration colloidal hard-sphere and protein dispersions," Journal of food science **72**, R101–R107 (2007).
[10]M. Tehei, D. Madern, C. Pfister, and G. Zaccai, "Fast dynamics of halophilic malate dehydrogenase and bsa measured by neutron scattering under various solvent conditions influencing protein stability," Proceedings of the National Academy of Sciences **98**, 14356–14361 (2001).
[11]M. Grimaldo, F. Roosen-Runge, F. Zhang, F. Schreiber, and T. Seydel, "Dynamics of proteins in solution," Quart. Rev. Biophys. **52** (2019), 10.1017/S0033583519000027.
[12]F. Zhang, M. W. A. Skoda, R. M. J. Jacobs, R. A. Martin, C. M. Martin, and F. Schreiber, "Protein interactions studied by SAXS: effect of ionic strength and protein concentration for BSA in aqueous solutions," The Journal of Physical Chemistry B **111**, 251–259 (2007).
[13]G. Pellicane and M. Cavero, "Theoretical study of interactions of BSA protein in a NaCl aqueous solution," The Journal of Chemical Physics **138**, 115103 (2013).
[14]F. Zhang, M. W. A. Skoda, R. M. J. Jacobs, S. Zorn, R. A. Martin, C. M. Martin, G. F. Clark, S. Weggler, A. Hildebrandt, O. Kohlbacher, and F. Schreiber, "Reentrant condensation of proteins in solution induced by multivalent counterions," Physical Review Letters **101** (2008), 10.1103/physrevlett.101.148101.
[15]M. Grimaldo, F. Roosen-Runge, M. Hennig, F. Zanini, F. Zhang, M. Zamponi, N. Jalarvo, F. Schreiber, and T. Seydel, "Salt-induced universal slowing down of the short-time self-diffusion of a globular protein in aqueous solution," The Journal of Physical Chemistry Letters **6**, 2577–2582 (2015).
[16]O. Matsarskaia, F. Roosen-Runge, G. Lotze, J. Möller, A. Mariani, F. Zhang, and F. Schreiber, "Tuning phase transitions of aqueous protein solutions by multivalent cations," Physical Chemistry Chemical Physics **20**, 27214–27225 (2018).
[17]M. K. Braun, A. Sauter, O. Matsarskaia, M. Wolf, F. Roosen-Runge, M. Sztucki, R. Roth, F. Zhang, and F. Schreiber, "Reentrant phase behavior in protein solutions induced by multivalent salts: Strong effect of anions $Cl^{\check{}}$ versus $NO3^{-}$," The Journal of Physical Chemistry B **122**, 11978–11985 (2018).
[18]C. Beck, M. Grimaldo, M. K. Braun, L. Bühl, O. Matsarskaia, N. H. Jalarvo, F. Zhang, F. Roosen-Runge, F. Schreiber, and T. Seydel, "Temperature and salt controlled tuning of protein clusters," Soft Matter **17**, 8506–8516 (2021).
[19]N. Muramatsu and A. P. Minton, "Tracer diffusion of globular proteins in concentrated protein solutions." Proceedings of the National Academy of Sciences **85**, 2984–2988 (1988).
[20]M. Tokuyama, T. Moriki, and Y. Kimura, "Self-diffusion of biomolecules in solution," Physical Review E **83** (2011), 10.1103/physreve.83.051402.
[21]J. C. Dyre, "Simple liquids' quasiuniversality and the hard-sphere paradigm," Journal of Physics: Condensed Matter **28**, 323001 (2016).
[22]F. Roosen-Runge, M. Hennig, T. Seydel, F. Zhang, M. W. Skoda, S. Zorn, R. M. Jacobs, M. Maccarini, P. Fouquet, and F. Schreiber, "Protein diffusion in crowded electrolyte solutions," Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics **1804**, 68–75 (2010).
[23]F. Roosen-Runge, M. Hennig, F. Zhang, R. M. J. Jacobs, M. Sztucki, H. Schober, T. Seydel, and F. Schreiber, "Protein self-diffusion in crowded solutions," Proceedings of the National Academy of Sciences **108**, 11815–11820 (2011).
[24]J. Balbo, P. Mereghetti, D.-P. Herten, and R. C. Wade, "The shape of protein crowders is a major determinant of protein diffusion," Biophysical Journal **104**, 1576–1584 (2013).
[25]F. G. Díaz, A. Iniesta, and J. G. D. L. Torre, "Hydrodynamic study of flexibility in immunoglobulin igG1 using brownian dynamics and the monte carlo simulations of a simple model," Biopolymers **30**, 547–554 (1990).
[26]F. Díaz, J. Cascales, and J. de la Torre, "Bead-model calculation of scattering diagrams: Brownian dynamics study of flexibility in immunoglobulin IgG1," Journal of Biochemical and Biophysical Methods **26**, 261–271 (1993).
[27]A. Chaudhri, I. E. Zarraga, T. J. Kamerzell, J. P. Brandt, T. W. Patapoff, S. J. Shire, and G. A. Voth, "Coarse-grained modeling of the self-association of therapeutic monoclonal antibodies," The Journal of Physical Chemistry B **116**, 8045–8057 (2012).
[28]A. Chaudhri, I. E. Zarraga, S. Yadav, T. W. Patapoff, S. J. Shire, and G. A. Voth, "The role of amino acid sequence in the self-association of therapeutic monoclonal antibodies: Insights from coarse-grained modeling," The Journal of Physical Chemistry B **117**, 1269–1279 (2013).
[29]K. P. Johnston, J. A. Maynard, T. M. Truskett, A. U. Borwankar, M. A. Miller, B. K. Wilson, A. K. Dinin, T. A. Khan, and K. J. Kaczorowski, "Concentrated dispersions of equilibrium protein nanoclusters that reversibly dissociate into active monomers," ACS nano **6**, 1357–1369 (2012).
[30]P. M. Buck, A. Chaudhri, S. Kumar, and S. K. Singh, "Highly viscous antibody solutions are a consequence of network formation caused by domain–domain electrostatic complementarities: Insights from coarse-grained simulations," Molecular Pharmaceutics **12**, 127–139 (2014).
[31]C. Calero-Rubio, A. Saluja, and C. J. Roberts, "Coarse-grained antibody models for "weak" protein–protein interactions from low to high concentrations," The Journal of Physical Chemistry B **120**, 6592–6605 (2016).
[32]G. Sun, Y. Wang, A. Lomakin, G. B. Benedek, H. E. Stanley, L. Xu, and S. V. Buldyrev, "The phase behavior study of human antibody solution using multi-scale modeling," The Journal of Chemical Physics **145**, 194901 (2016).
[33]G. Wang, Z. Varga, J. Hofmann, I. E. Zarraga, and J. W. Swan, "Structure and relaxation in solutions of monoclonal antibodies," The Journal of Physical Chemistry B **122**, 2867–2880 (2018).
[34]N. Skar-Gislinge, M. Ronti, T. Garting, C. Rischel, P. Schurtenberger, E. Zaccarelli, and A. Stradner, "A colloid approach to self-assembling antibodies," Molecular Pharmaceutics **16**, 2394–2404 (2019).
[35]A. Chowdhury, G. Guruprasad, A. T. Chen, C. A. Karouta, M. A. Blanco, T. M. Truskett, and K. P. Johnston, "Protein-protein interactions, clustering, and rheology for bovine IgG up to high concentrations characterized by small angle x-ray scattering and molecular dynamics simulations," Journal of Pharmaceutical Sciences **109**, 696–708 (2020).
[36]S. Izadi, T. W. Patapoff, and B. T. Walters, "Multiscale coarse-grained approach to investigate self-association of antibodies," Biophysical Journal **118**, 2741–2754 (2020).
[37]H. Shahfar, J. K. Forder, and C. J. Roberts, "Toward a suite of coarse-grained models for molecular simulation of monoclonal antibodies and therapeutic proteins," The Journal of Physical Chemistry B **125**, 3574–

Accepted to J. Chem. Phys. 10.1063/5.0132493

ACCEPTED MANUSCRIPT
The Journal of Chemical Physics
AIP Publishing
This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.
PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0132493

3588 (2021).
[38]Y. Zhai, N. S. Martys, W. L. George, J. E. Curtis, J. Nayem, Y. Z, and Y. Liu, "Intermediate scattering functions of a rigid body monoclonal antibody protein in solution studied by dissipative particle dynamic simulation," Structural Dynamics **8**, 024102 (2021).
[39]P.-K. Lai, J. W. Swan, and B. L. Trout, "Calculation of therapeutic antibody viscosity with coarse-grained models, hydrodynamic calculations and machine learning-based parameters," mAbs **13** (2021), 10.1080/19420862.2021.1907882.
[40]S. Pusara, P. Yamin, W. Wenzel, M. Krstić, and M. Kozlowska, "A coarse-grained xDLVO model for colloidal protein–protein interactions," Physical Chemistry Chemical Physics **23**, 12780–12794 (2021).
[41]S. Pandit, S. Kundu, and V. K. Aswal, "Effect of monovalent salts on molecular interactions of globular protein (BSA) above its isoelectric point," Chemical Physics Letters **804**, 139916 (2022).
[42]I. Sillitoe, N. Bordin, N. Dawson, V. P. Waman, P. Ashford, H. M. Scholes, C. S. M. Pang, L. Woodridge, C. Rauer, N. Sen, M. Abbasian, S. L. Cornu, S. D. Lam, K. Berka, I. H. Varekova, R. Svobodova, J. Lees, and C. A. Orengo, "CATH: increased structural coverage of functional space," Nucleic Acids Research **49**, D266–D273 (2020).
[43]S. J. Marrink, H. J. Risselada, S. Yefimov, D. P. Tieleman, and A. H. de Vries, "The MARTINI force field: coarse grained model for biomolecular simulations," The Journal of Physical Chemistry B **111**, 7812–7824 (2007).
[44]P. C. T. Souza, R. Alessandri, J. Barnoud, S. Thallmair, I. Faustino, F. Grünewald, I. Patmanidis, H. Abdizadeh, B. M. H. Bruininks, T. A. Wassenaar, P. C. Kroon, J. Melcr, V. Nieto, V. Corradi, H. M. Khan, J. Domański, M. Javanainen, H. Martinez-Seara, N. Reuter, R. B. Best, I. Vattulainen, L. Monticelli, X. Periole, D. P. Tieleman, A. H. de Vries, and S. J. Marrink, "Martini 3: a general purpose force field for coarse-grained molecular dynamics," Nature Methods **18**, 382–388 (2021).
[45]A. Arkhipov, P. L. Freddolino, and K. Schulten, "Stability and dynamics of virus capsids described by coarse-grained modeling," Structure **14**, 1767–1777 (2006).
[46]M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, and E. Lindahl, "GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers," SoftwareX **1-2**, 19–25 (2015).
[47]W. G. Noid, "Perspective: Coarse-grained models for biomolecular systems," The Journal of Chemical Physics **139**, 090901 (2013).
[48]E. Brini, E. A. Algaer, P. Ganguly, C. Li, F. Rodríguez-Ropero, and N. F. A. van der Vegt, "Systematic coarse-graining methods for soft matter simulations – a review," Soft Matter **9**, 2108–2119 (2013).
[49]M. G. Saunders and G. A. Voth, "Coarse-graining methods for computational biology," Annual Review of Biophysics **42**, 73–93 (2013).
[50]R. Potestio, C. Peter, and K. Kremer, "Computer simulations of soft matter: Linking the scales," Entropy **16**, 4199–4245 (2014).
[51]S. Kmiecik, D. Gront, M. Kolinski, L. Wieteska, A. E. Dawid, and A. Kolinski, "Coarse-grained protein models and their applications," Chemical Reviews **116**, 7898–7936 (2016).
[52]S. Y. Joshi and S. A. Deshmukh, "A review of advancements in coarse-grained molecular dynamics simulations," Molecular Simulation **47**, 786–803 (2020).
[53]M. M. Tirion, "Large amplitude elastic motions in proteins from a single-parameter, atomic analysis," Physical Review Letters **77**, 1905–1908 (1996).
[54]I. Bahar, T. R. Lezon, L.-W. Yang, and E. Eyal, "Global dynamics of proteins: Bridging between structure and function," Annual Review of Biophysics **39**, 23–42 (2010).
[55]K. Hinsen, "Analysis of domain motions by approximate normal mode calculations," Proteins: Structure, Function, and Genetics **33**, 417–429 (1998).
[56]E. Lyman, J. Pfaendtner, and G. A. Voth, "Systematic multiscale parameterization of heterogeneous elastic network models of proteins," Biophysical Journal **95**, 4183–4192 (2008).
[57]S. Riniker, J. R. Allison, and W. F. van Gunsteren, "On developing coarse-grained models for biomolecular simulation: a review," Physical Chemistry Chemical Physics **14**, 12423 (2012).
[58]A. Amadei, A. B. M. Linssen, and H. J. C. Berendsen, "Essential dynamics of proteins," Proteins: Structure, Function, and Genetics **17**, 412–425 (1993).
[59]J. P. Brandt, T. W. Patapoff, and S. R. Aragon, "Construction, MD simulation, and hydrodynamic validation of an all-atom model of a monoclonal IgG antibody," Biophysical Journal **99**, 905–913 (2010).
[60]J. K. G. Dhont, *An Introduction to Dynamics of Colloids* (Elsevier, 1996).
[61]G. Mie, "Zur kinetischen Theorie der einatomigen Körper," Annalen der Physik **316**, 657–697 (1903).
[62]J. A. Barker and D. Henderson, "Perturbation theory and equation of state for fluids. II. a successful theory of liquids," The Journal of Chemical Physics **47**, 4714–4721 (1967).
[63]T. Ando and J. Skolnick, "Crowding and hydrodynamic interactions likely dominate in vivo macromolecular motion," Proceedings of the National Academy of Sciences **107**, 18457–18462 (2010).
[64]M. Grimaldo, F. Roosen-Runge, F. Zhang, T. Seydel, and F. Schreiber, "Diffusion and dynamics of γ-globulin in crowded aqueous solutions," The Journal of Physical Chemistry B **118**, 7203–7209 (2014), pMID: 24871685.
[65]I. V. Nesmelova, V. D. Skirda, and V. D. Fedotov, "Generalized concentration dependence of globular protein self-diffusion coefficients in aqueous solutions," Biopolymers **63**, 132–140 (2002).
[66]M. Rubinstein and R. Colby, *Polymer Physics* (Oxford University Press, 2003).
[67]A. J. Banchio, J. Bergenholtz, and G. Nägele, "Rheology and dynamics of colloidal suspensions," Physical Review Letters **82**, 1792–1795 (1999).
[68]M. Roos, M. Ott, M. Hofmann, S. Link, E. Rössler, J. Balbach, A. Krushelnitsky, and K. Saalwächter, "Coupling and decoupling of rotational and translational diffusion of proteins under crowding conditions," Journal of the American Chemical Society **138**, 10365–10372 (2016).
[69]A. van Blaaderen, J. Peetermans, G. Maret, and J. K. G. Dhont, "Long-time self-diffusion of spherical colloidal particles measured with fluorescence recovery after photobleaching," The Journal of Chemical Physics **96**, 4591–4603 (1992).
[70]M. Medina-Noyola, "Long-time self-diffusion in concentrated colloidal dispersions," Physical Review Letters **60**, 2705–2708 (1988).
[71]J. A. Leegwater and G. Szamel, "Dynamical properties of hard-sphere suspensions," Physical Review A **46**, 4999–5011 (1992).
[72]M. Tokuyama and I. Oppenheim, "Dynamics of hard-sphere suspensions," Physical Review E **50**, R16–R19 (1994).
[73]M. Grimaldo, F. Roosen-Runge, N. Jalarvo, M. Zamponi, F. Zanini, M. Hennig, F. Zhang, F. Schreiber, and T. Seydel, "High-resolution neutron spectroscopy on protein solution samples," EPJ Web of Conferences **83**, 02005 (2015).
[74]L. D. Favro, "Theory of the rotational brownian motion of a free rigid body," Physical Review **119**, 53–62 (1960).
[75]M. Linke, J. Köfinger, and G. Hummer, "Fully anisotropic rotational diffusion tensor from molecular dynamics simulations," The Journal of Physical Chemistry B **122**, 5630–5639 (2018).
[76]N. Haridasan, S. K. Kannam, S. Mogurampelly, and S. P. Sathian, "Rotational diffusion of proteins in nanochannels," The Journal of Physical Chemistry B **123**, 4825–4832 (2019), pMID: 31117604.
[77]J. Wei, Y. Liu, and F. Song, "Coarse-grained simulation of the translational and rotational diffusion of globular proteins by dissipative particle dynamics," The Journal of Chemical Physics **153**, 234902 (2020).
[78]P. E. Smith and W. F. van Gunsteren, "Translational and rotational diffusion of proteins," Journal of Molecular Biology **236**, 629–636 (1994).
[79]A. Girelli, C. Beck, F. Bäuerle, O. Matsarskaia, R. Maier, F. Zhang, B. Wu, C. Lang, O. Czakkel, T. Seydel, F. Schreiber, and F. Roosen-Runge, "Molecular flexibility of antibodies preserved even in the dense phase after macroscopic phase separation," Molecular Pharmaceutics **18**, 4162–4169 (2021), pMID: 34637319.
[80]F. Zhang, F. Roosen-Runge, M. W. A. Skoda, R. M. J. Jacobs, M. Wolf, P. Callow, H. Frielinghaus, V. Pipich, S. Prévost, and F. Schreiber, "Hydration and interactions in protein solutions containing concentrated electrolytes studied by small-angle scattering," Physical Chemistry Chemical Physics **14**, 2483 (2012).
[81]S. Qin and H.-X. Zhou, "Calculation of second virial coefficients of atomistic proteins using fast fourier transform," The Journal of Physical Chemistry B **123**, 8203–8215 (2019).
[82]M. R. Fries, D. Stopper, M. K. Braun, A. Hinderhofer, F. Zhang, R. M. J. Jacobs, M. W. A. Skoda, H. Hansen-Goos, R. Roth, and

Accepted to J. Chem. Phys. 10.1063/5.0132493

ACCEPTED MANUSCRIPT
The Journal of Chemical Physics
AIP Publishing
This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.
PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0132493

F. Schreiber, "Multivalent-ion-activated protein adsorption reflecting bulk reentrant behavior," Physical Review Letters **119** (2017), 10.1103/physrevlett.119.228001.

[83] M. R. Fries, D. Stopper, M. W. A. Skoda, M. Blum, C. Kertzscher, A. Hinderhofer, F. Zhang, R. M. J. Jacobs, R. Roth, and F. Schreiber, "Enhanced protein adsorption upon bulk phase separation," Scientific Reports **10** (2020), 10.1038/s41598-020-66562-0.

[84] N. Begam, O. Matsarskaia, M. Sztucki, F. Zhang, and F. Schreiber, "Unification of lower and upper critical solution temperature phase behavior of globular protein solutions in the presence of multivalent cations," Soft Matter **16**, 2128–2134 (2020).

[85] F. Roosen-Runge, F. Zhang, F. Schreiber, and R. Roth, "Ion-activated attractive patches as a mechanism for controlled protein interactions," Scientific Reports **4** (2014), 10.1038/srep07016.

[86] S. Bucciarelli, J. S. Myung, B. Farago, S. Das, G. A. Vliegenthart, O. Holderer, R. G. Winkler, P. Schurtenberger, G. Gompper, and A. Stradner, "Dramatic influence of patchy attractions on short-time protein diffusion under crowded conditions," Science Advances **2** (2016), 10.1126/sciadv.1601432.

[87] C. Beck, M. Grimaldo, H. Lopez, S. D. Vela, B. Sohmen, F. Zhang, M. Oettel, J.-L. Barrat, F. Roosen-Runge, F. Schreiber, and T. Seydel, "Short-time transport properties of bidisperse suspensions of immunoglobulins and serum albumins consistent with a colloid physics picture," The Journal of Physical Chemistry B **126**, 7400–7408 (2022).

[88] P. Bjelkmar, P. Larsson, M. A. Cuendet, B. Hess, and E. Lindahl, "Implementation of the CHARMM force field in GROMACS: Analysis of protein stability effects from correction maps, virtual interaction sites, and water models," Journal of Chemical Theory and Computation **6**, 459–466 (2010).

[89] G. Bussi, D. Donadio, and M. Parrinello, "Canonical sampling through velocity rescaling," The Journal of Chemical Physics **126**, 014101 (2007).

[90] M. Parrinello and A. Rahman, "Polymorphic transitions in single crystals: A new molecular dynamics method," Journal of Applied Physics **52**, 7182–7190 (1981).

[91] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen, "A smooth particle mesh ewald method," The Journal of Chemical Physics **103**, 8577–8593 (1995).

[92] M. P. Allen and D. J. Tildesley, *Computer simulation of liquids* (Oxford University Press, Oxford, 1987).

[93] J. A. Anderson, C. D. Lorenz, and A. Travesset, "General purpose molecular dynamics simulations fully implemented on graphics processing units," Journal of Computational Physics **227**, 5342–5359 (2008).

[94] J. Glaser, T. D. Nguyen, J. A. Anderson, P. Lui, F. Spiga, J. A. Millan, D. C. Morse, and S. C. Glotzer, "Strong scaling of general-purpose molecular dynamics simulations on GPUs," Computer Physics Communications **192**, 97–107 (2015).

[95] F. W. Putnam, *The Plasma Proteins (Second Edition)* (Academic Press, 1977).

[96] L. E. Rayner, G. K. Hui, J. Gor, R. K. Heenan, P. A. Dalby, and S. J. Perkins, "The solution structures of two human IgG1 antibodies show conformational stability and accommodate their C1q and Fcγr ligands," Journal of Biological Chemistry **290**, 8420–8438 (2015).

[97] A. Zaccone, I. Terentjev, L. D. Michele, and E. M. Terentjev, "Fragmentation and depolymerization of non-covalently bonded filaments," The Journal of Chemical Physics **142**, 114905 (2015).

[98] A. K. Gaigalas, J. B. Hubbard, M. McCurley, and S. Woo, "Diffusion of bovine serum albumin in aqueous solutions," The Journal of Physical Chemistry **96**, 2355–2359 (1992).

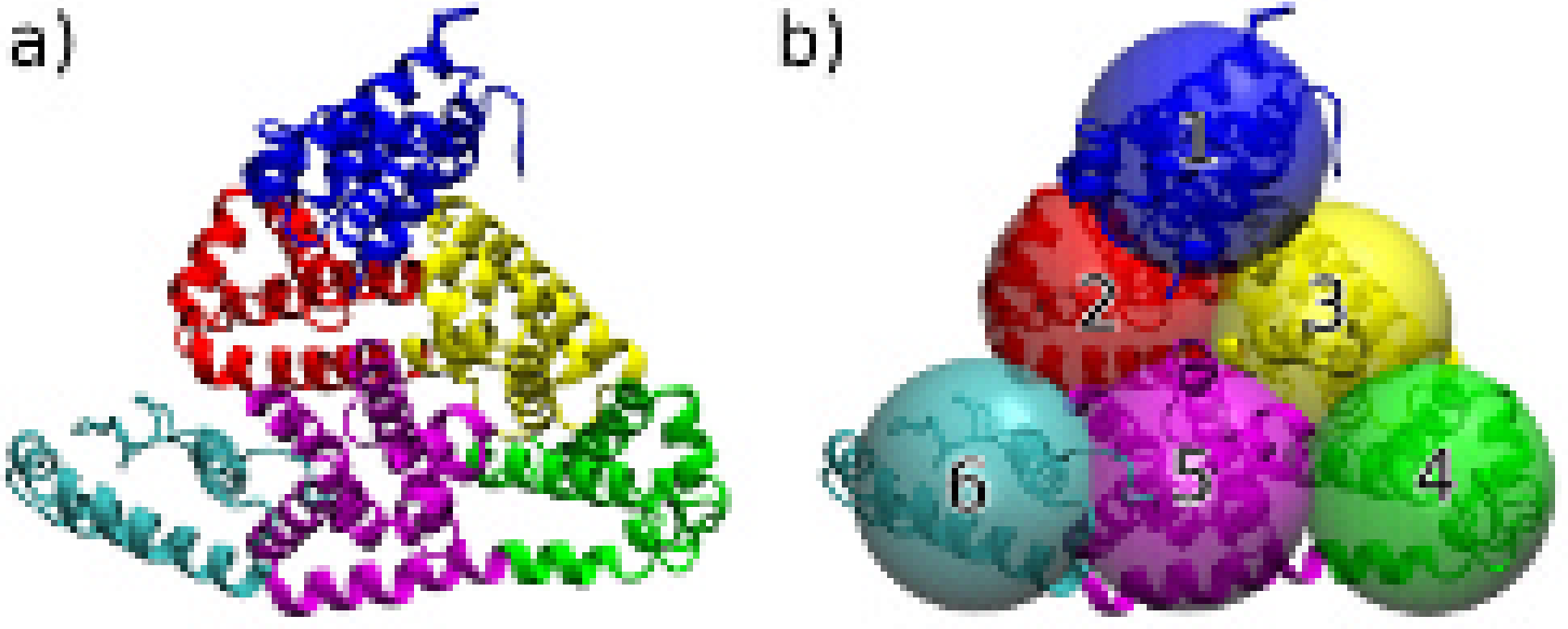
[99] D. Berk, F. Yuan, M. Leunig, and R. Jain, "Fluorescence photobleaching with spatial fourier analysis: measurement of diffusion in light-scattering media," Biophysical Journal **65**, 2428–2436 (1993).

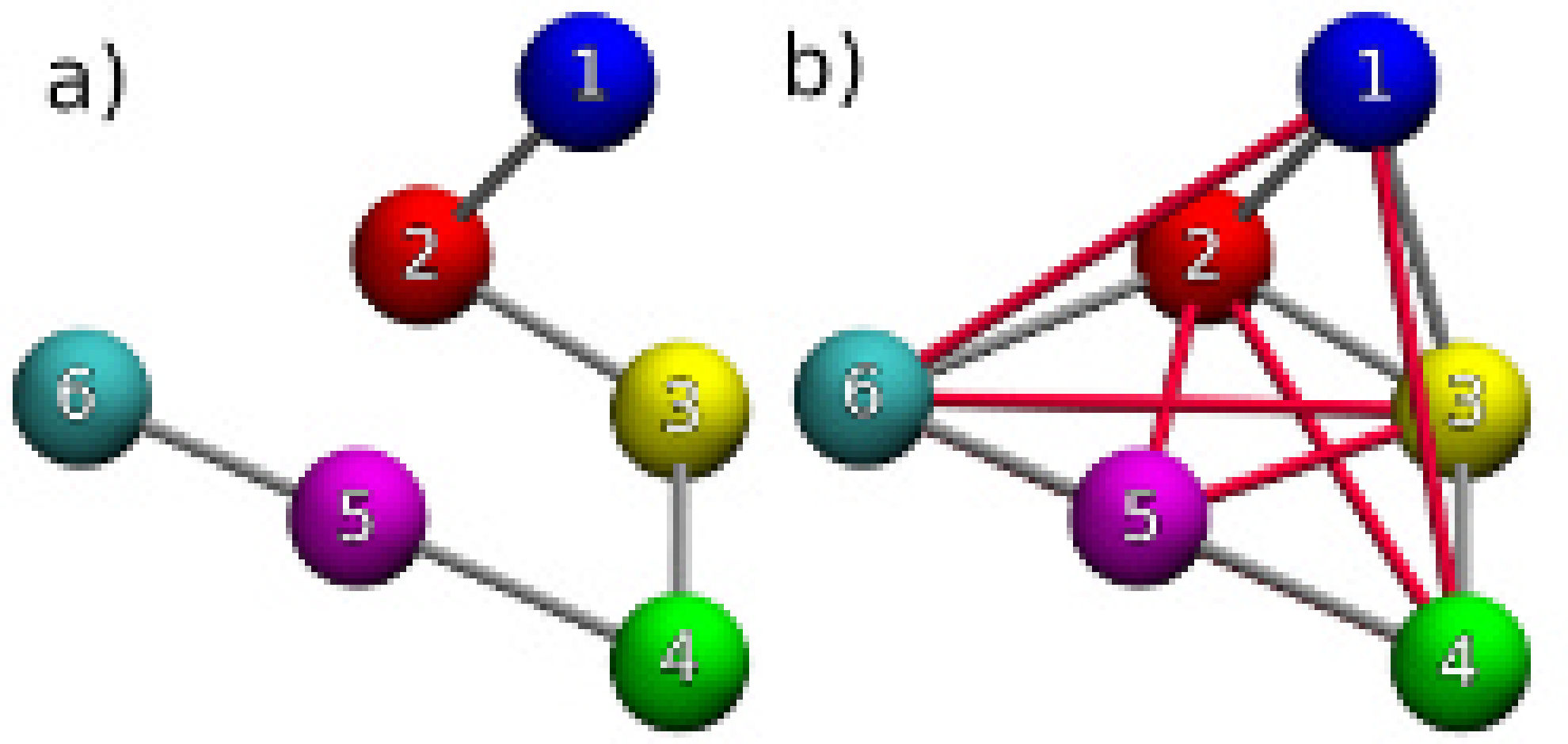
[100] D. L. Theobald, "Rapid calculation of RMSDs using a quaternion-based characteristic polynomial," Acta Crystallographica Section A Foundations of Crystallography **61**, 478–480 (2005).

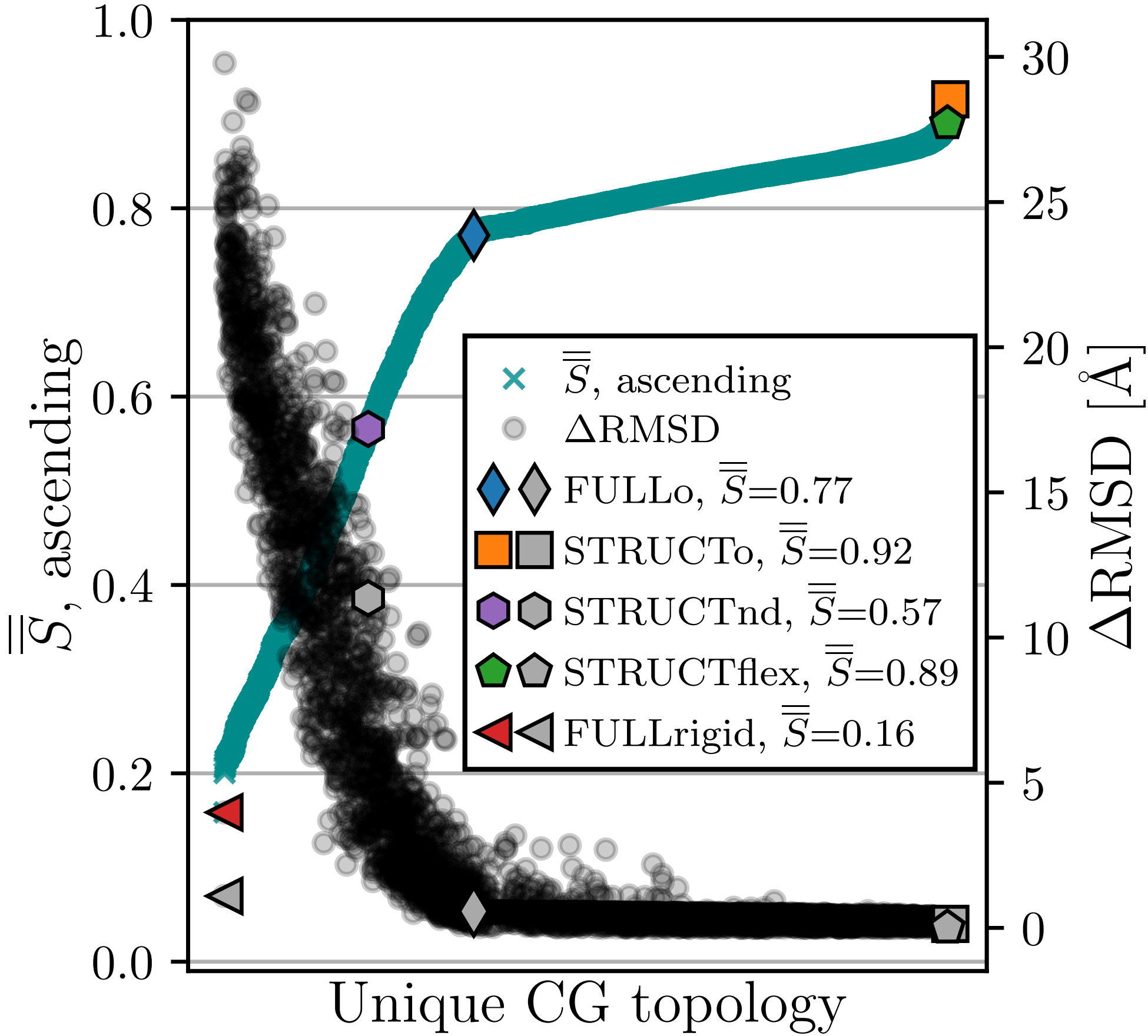
[101] P. Liu, D. K. Agrafiotis, and D. L. Theobald, "Fast determination of the optimal rotational matrix for macromolecular superpositions," Journal of Computational Chemistry , 1561–1563 (2009).

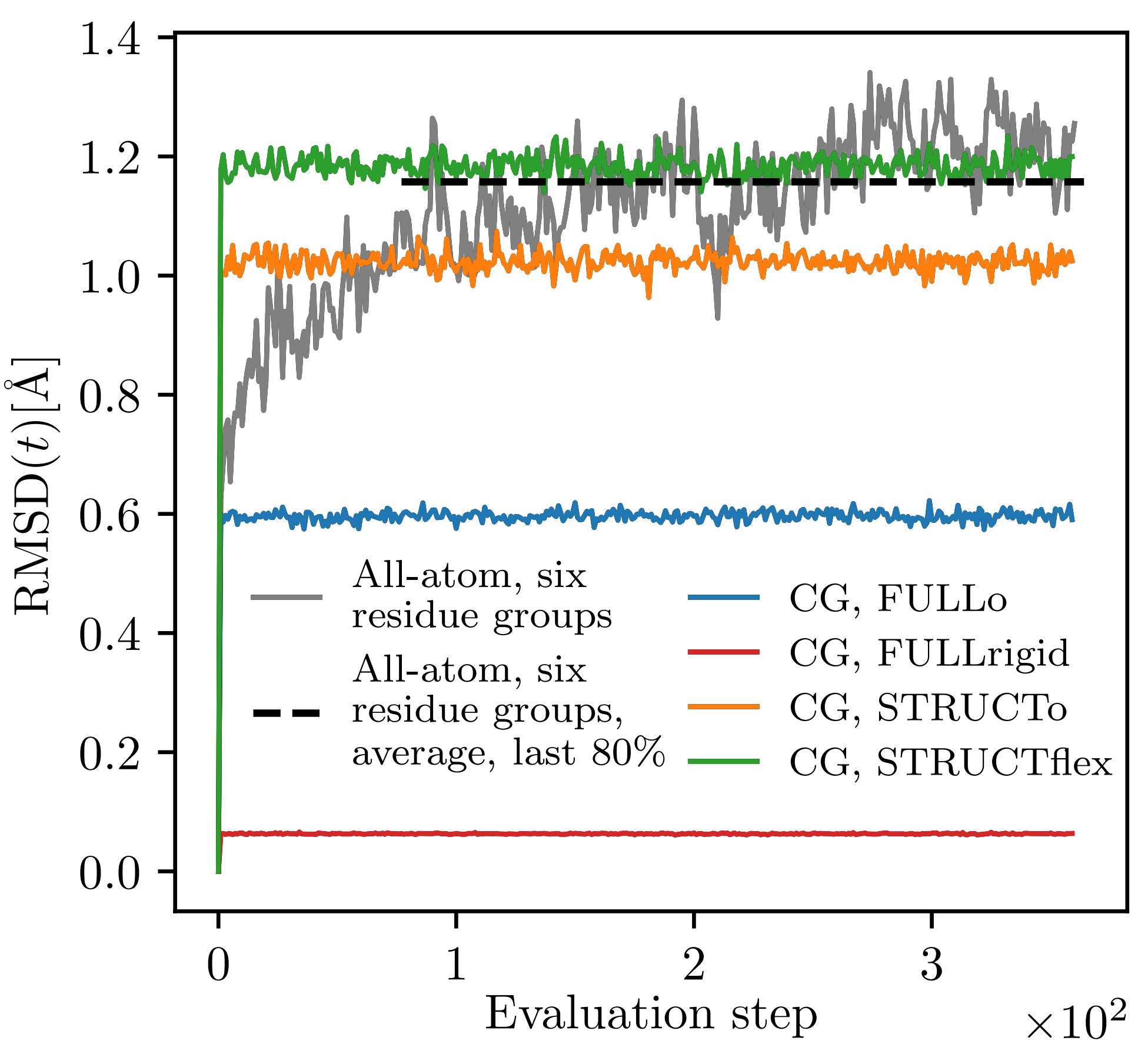
[102] R. Gowers, M. Linke, J. Barnoud, T. Reddy, M. Melo, S. Seyler, J. Domański, D. Dotson, S. Buchoux, I. Kenney, and O. Beckstein, "MDAnalysis: A python package for the rapid analysis of molecular dynamics simulations," in *Proceedings of the 15th Python in Science Conference* (SciPy, 2016).

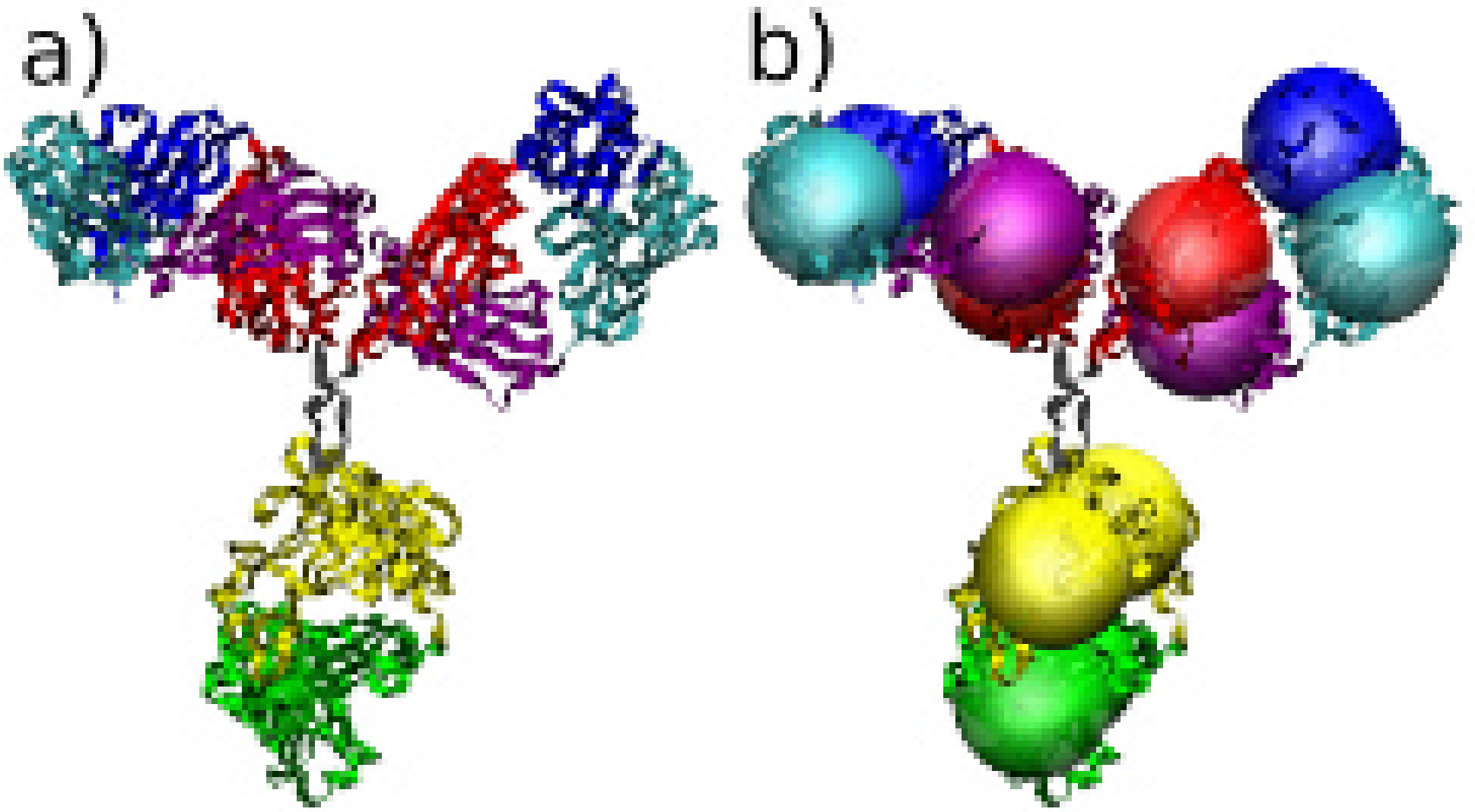
[103] N. Michaud-Agrawal, E. J. Denning, T. B. Woolf, and O. Beckstein, "MDAnalysis: A toolkit for the analysis of molecular dynamics simulations," Journal of Computational Chemistry **32**, 2319–2327 (2011).

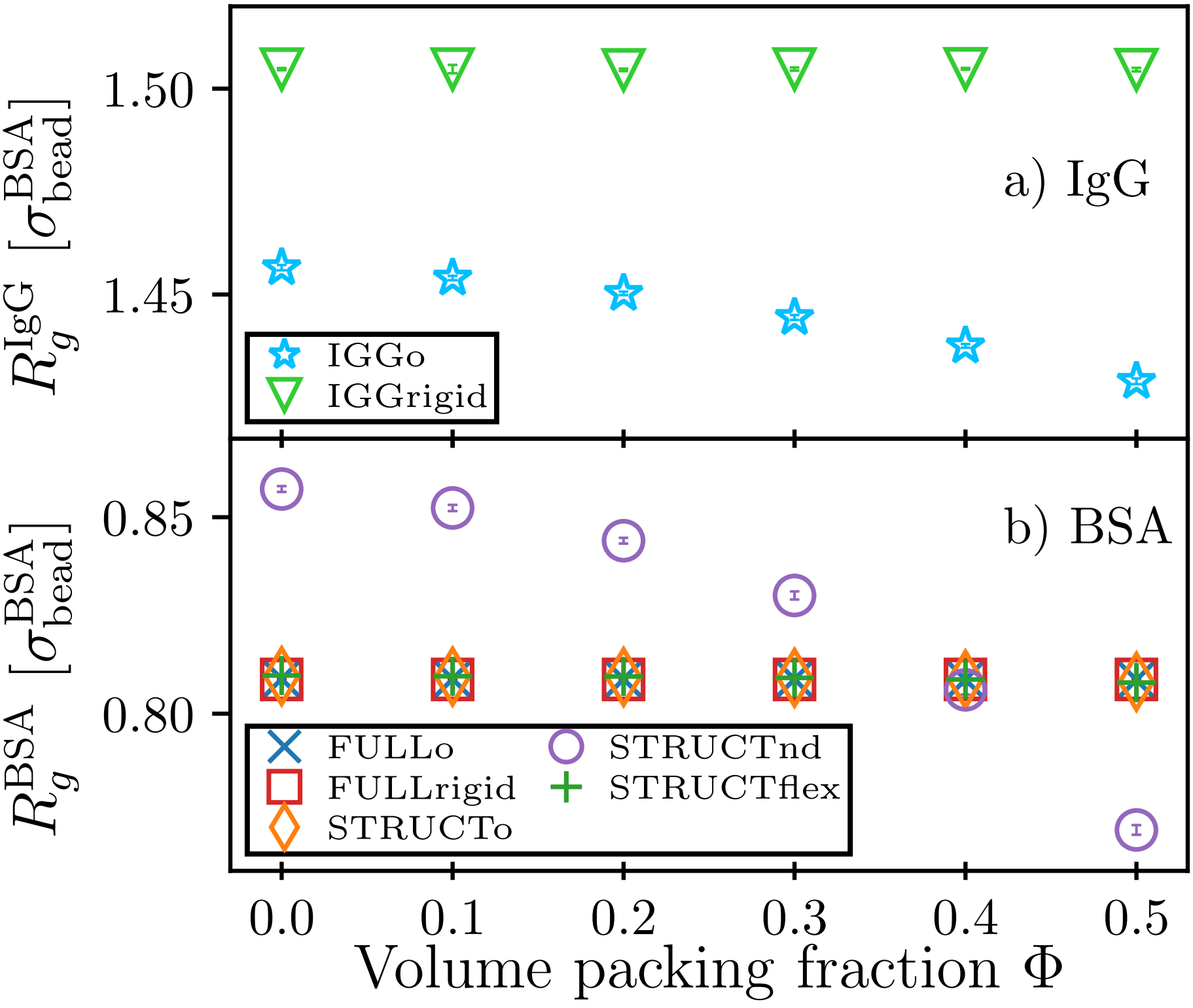

a)
b)
1
2
3
4
5
6

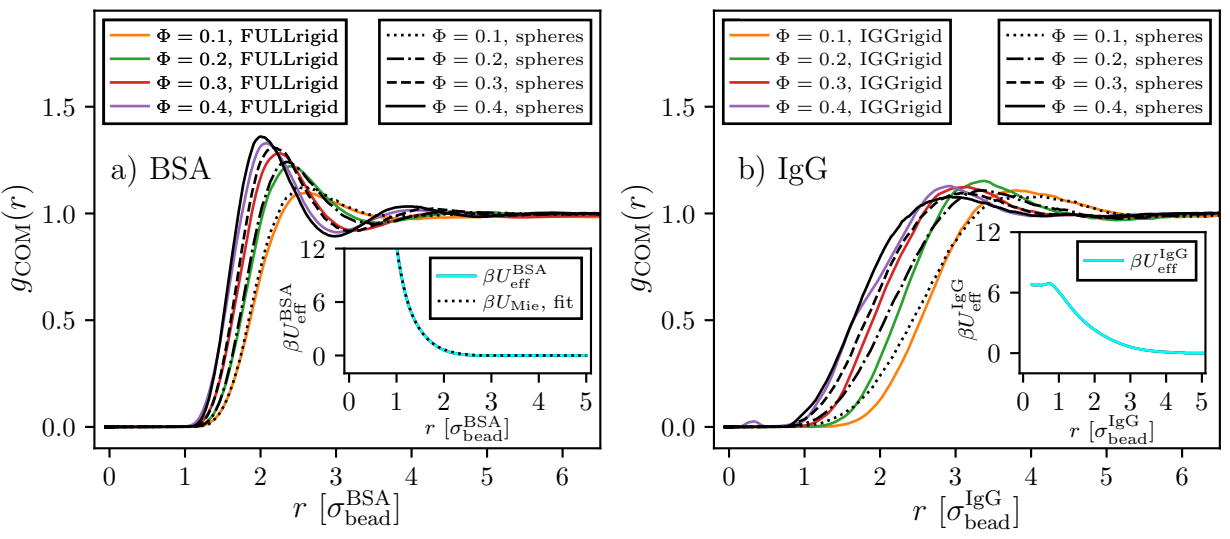

a)
1
2
3
4
5
6
b)
1
2
3
4
5
6

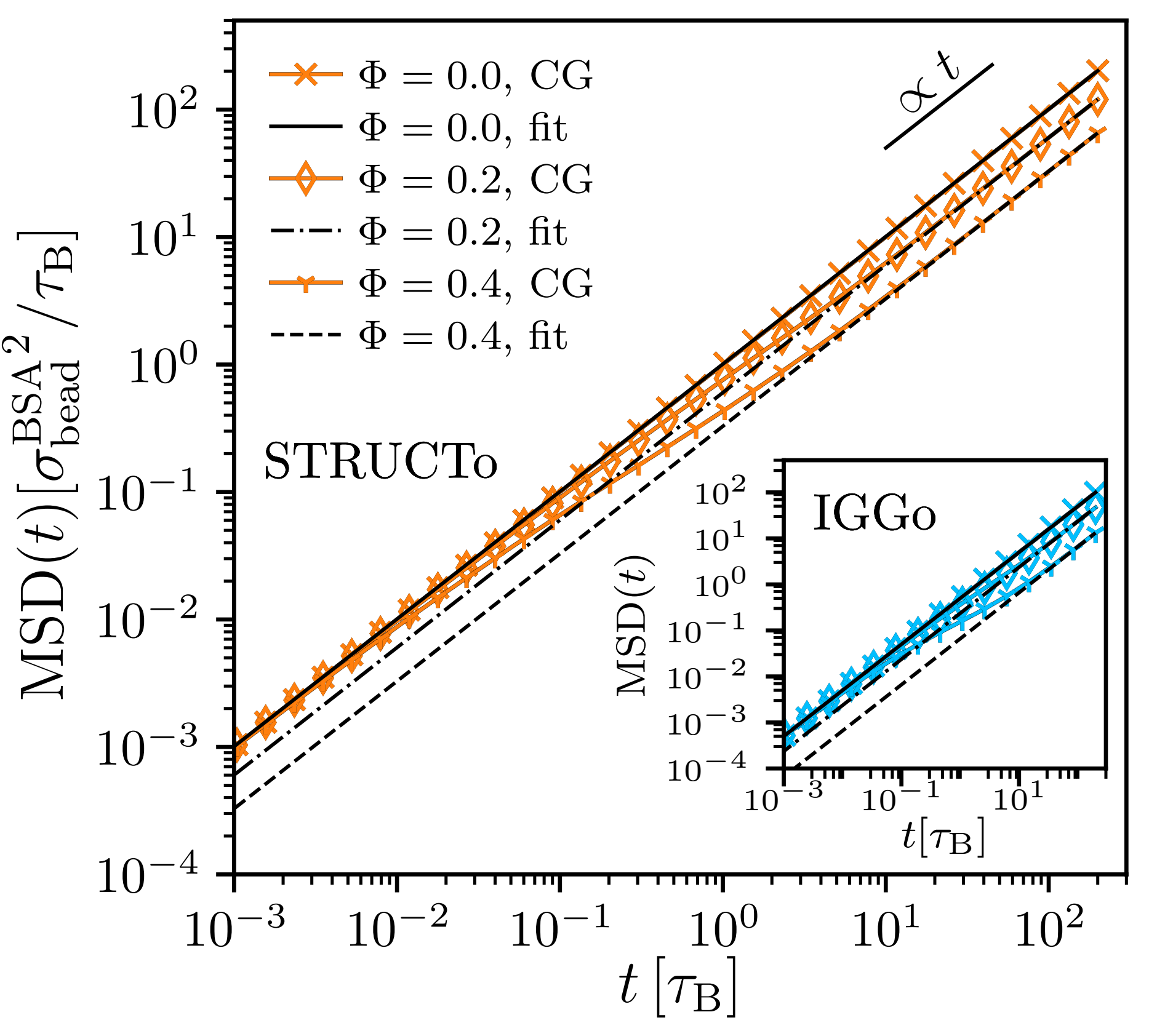

1.0
0.8
0.6
0.4
0.2
0.0
$\overline{\overline{S}}$, ascending
30
25
20
15
10
5
0
ΔRMSD [Å]
Unique CG topology
$\overline{\overline{S}}$, ascending
ΔRMSD
FULLo, $\overline{\overline{S}}$=0.77
STRUCTo, $\overline{\overline{S}}$=0.92
STRUCTnd, $\overline{\overline{S}}$=0.57
STRUCTflex, $\overline{\overline{S}}$=0.89
FULLrigid, $\overline{\overline{S}}$=0.16

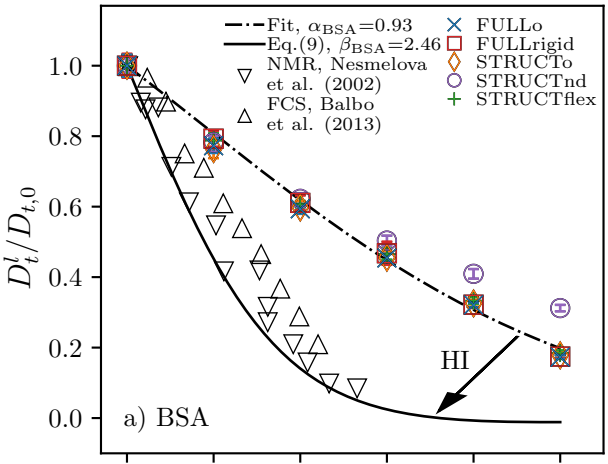

1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
RMSD(t)[Å]
0
1
2
3
Evaluation step
$\times 10^2$
All-atom, six residue groups
All-atom, six residue groups, average, last 80%
CG, FULLo
CG, FULLrigid
CG, STRUCTo
CG, STRUCTflex

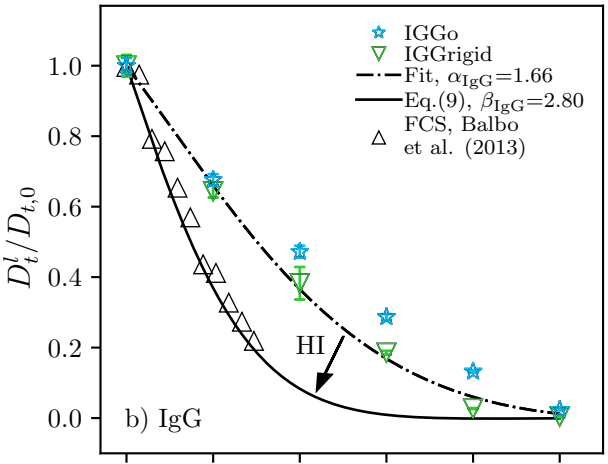

a)
b)

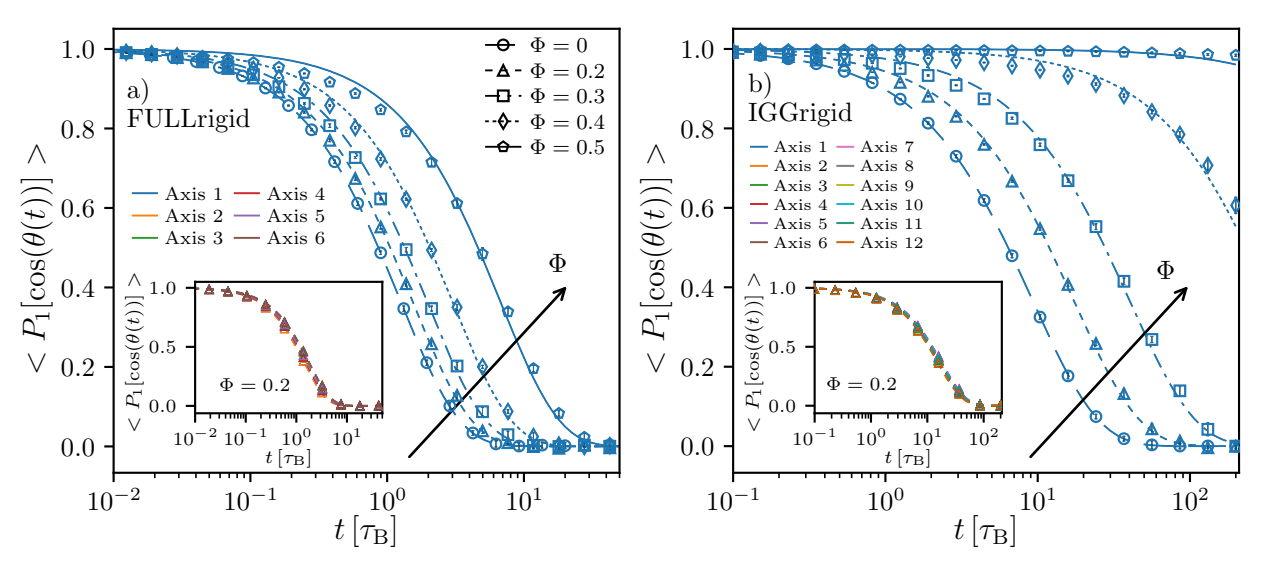

1.50
1.45
R_g^{IgG} $[\sigma_{\mathrm{bead}}^{\mathrm{BSA}}]$
a) IgG
IGGo
IGGrigid
0.85
0.80
R_g^{BSA} $[\sigma_{\mathrm{bead}}^{\mathrm{BSA}}]$
b) BSA
FULLo
FULLrigid
STRUCTo
STRUCTnd
STRUCTflex
0.0
0.1
0.2
0.3
0.4
0.5
Volume packing fraction Φ

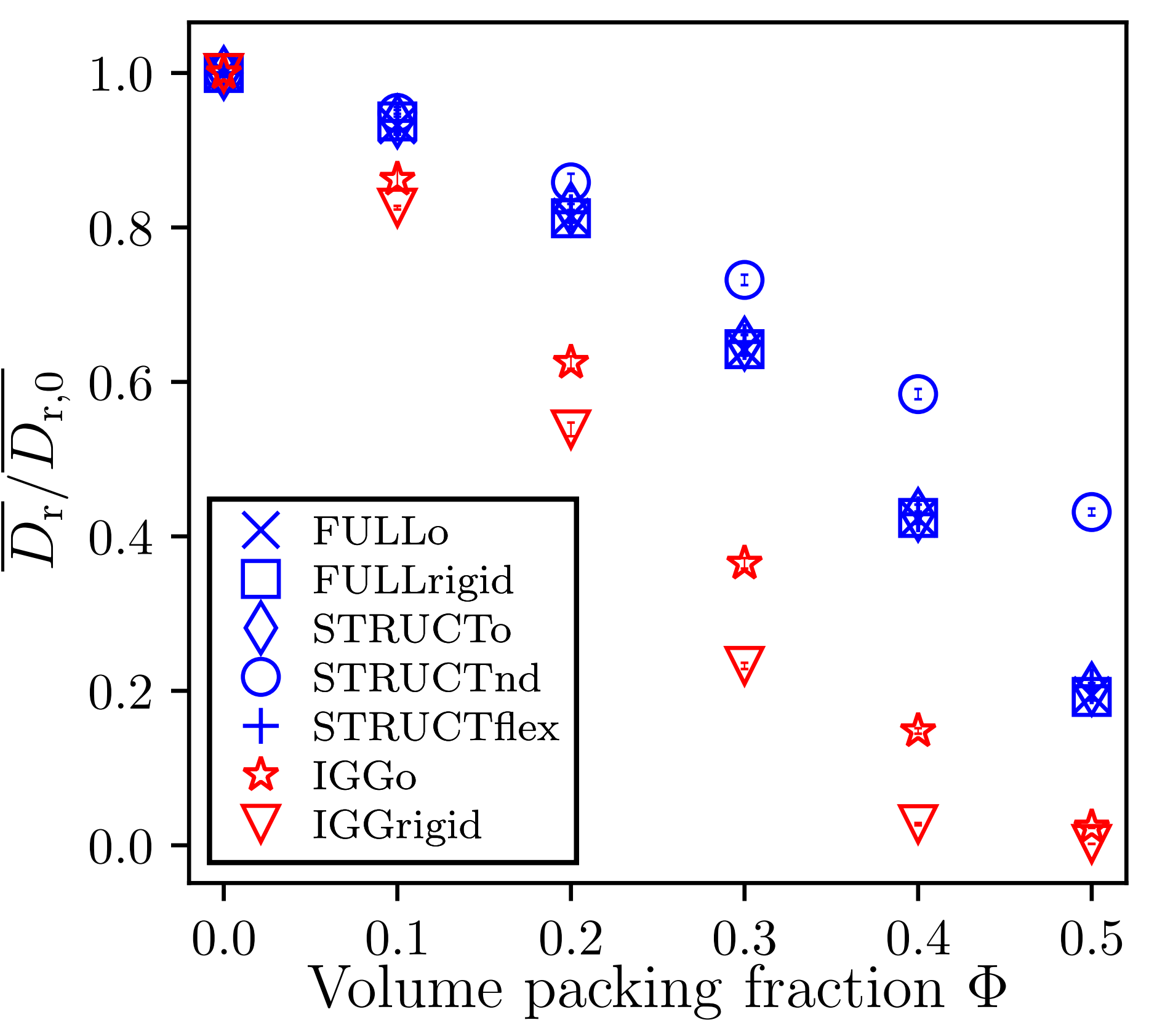

Φ = 0.1, FULLrigid
Φ = 0.2, FULLrigid
Φ = 0.3, FULLrigid
Φ = 0.4, FULLrigid
Φ = 0.1, spheres
Φ = 0.2, spheres
Φ = 0.3, spheres
Φ = 0.4, spheres
a) BSA
$g_{\mathrm{COM}}(r)$
$r\ [\sigma_{\mathrm{bead}}^{\mathrm{BSA}}]$
$\beta U_{\mathrm{eff}}^{\mathrm{BSA}}$
$\beta U_{\mathrm{eff}}^{\mathrm{BSA}}$
βU_{Mie}, fit
$r\ [\sigma_{\mathrm{bead}}^{\mathrm{BSA}}]$
Φ = 0.1, IGGrigid
Φ = 0.2, IGGrigid
Φ = 0.3, IGGrigid
Φ = 0.4, IGGrigid
Φ = 0.1, spheres
Φ = 0.2, spheres
Φ = 0.3, spheres
Φ = 0.4, spheres
b) IgG
$g_{\mathrm{COM}}(r)$
$r\ [\sigma_{\mathrm{bead}}^{\mathrm{IgG}}]$
$\beta U_{\mathrm{eff}}^{\mathrm{IgG}}$
$\beta U_{\mathrm{eff}}^{\mathrm{IgG}}$
$r\ [\sigma_{\mathrm{bead}}^{\mathrm{IgG}}]$

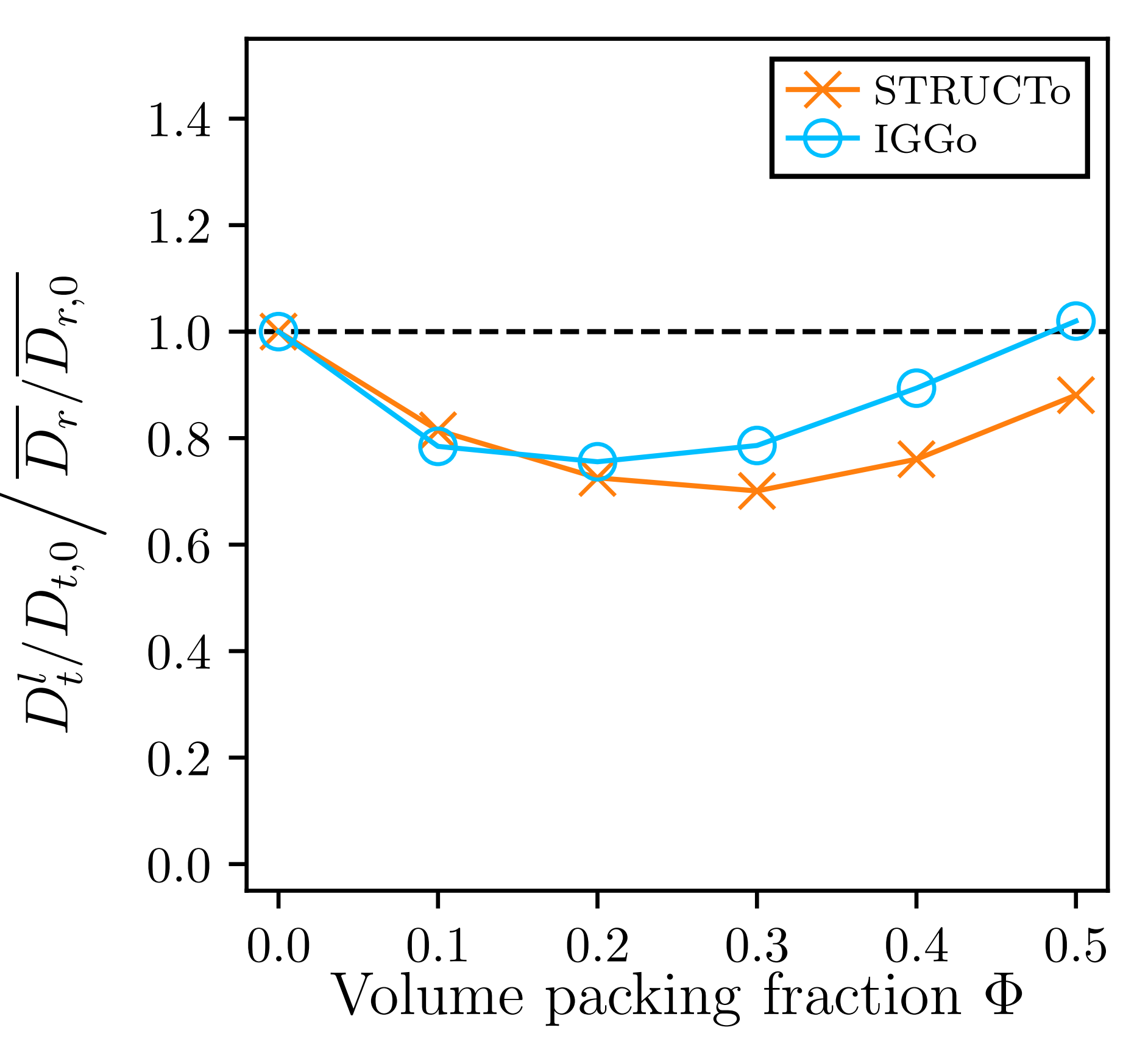

$\Phi = 0.0$, CG
$\Phi = 0.0$, fit
$\Phi = 0.2$, CG
$\Phi = 0.2$, fit
$\Phi = 0.4$, CG
$\Phi = 0.4$, fit
$\propto t$
STRUCTo
IGGo
$\mathrm{MSD}(t)\,[\sigma_{\mathrm{bead}}^{\mathrm{BSA}\,2}/\tau_{\mathrm{B}}]$
$t\,[\tau_{\mathrm{B}}]$
10^{-4}
10^{-3}
10^{-2}
10^{-1}
10^{0}
10^{1}
10^{2}

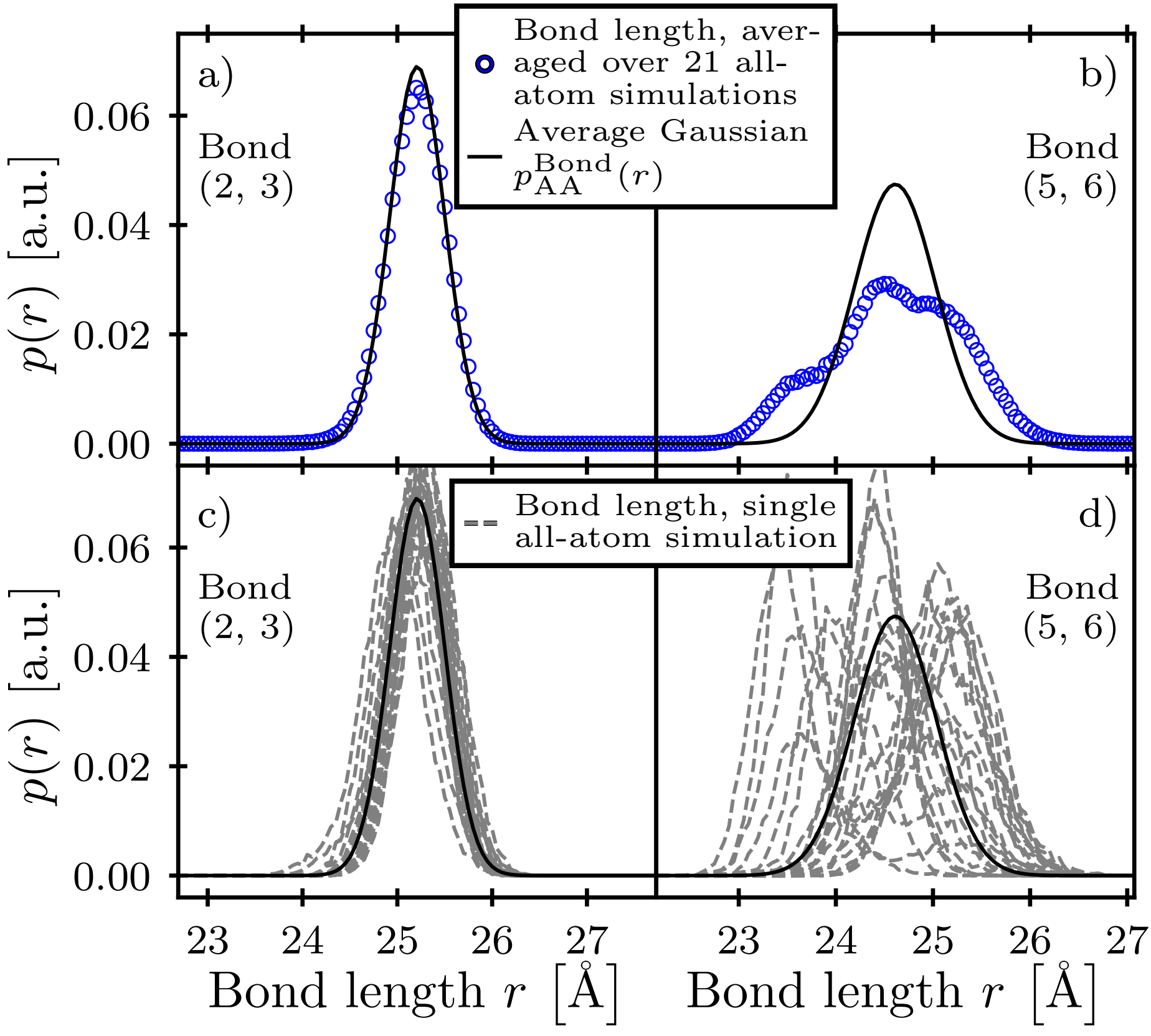

Fit, α_{BSA}=0.93
Eq.(9), β_{BSA}=2.46
NMR, Nesmelova et al. (2002)
FCS, Balbo et al. (2013)
FULLo
FULLrigid
STRUCTo
STRUCTnd
STRUCTflex
HI
a) BSA
$D_t^l/D_{t,0}$
Volume packing fraction Φ

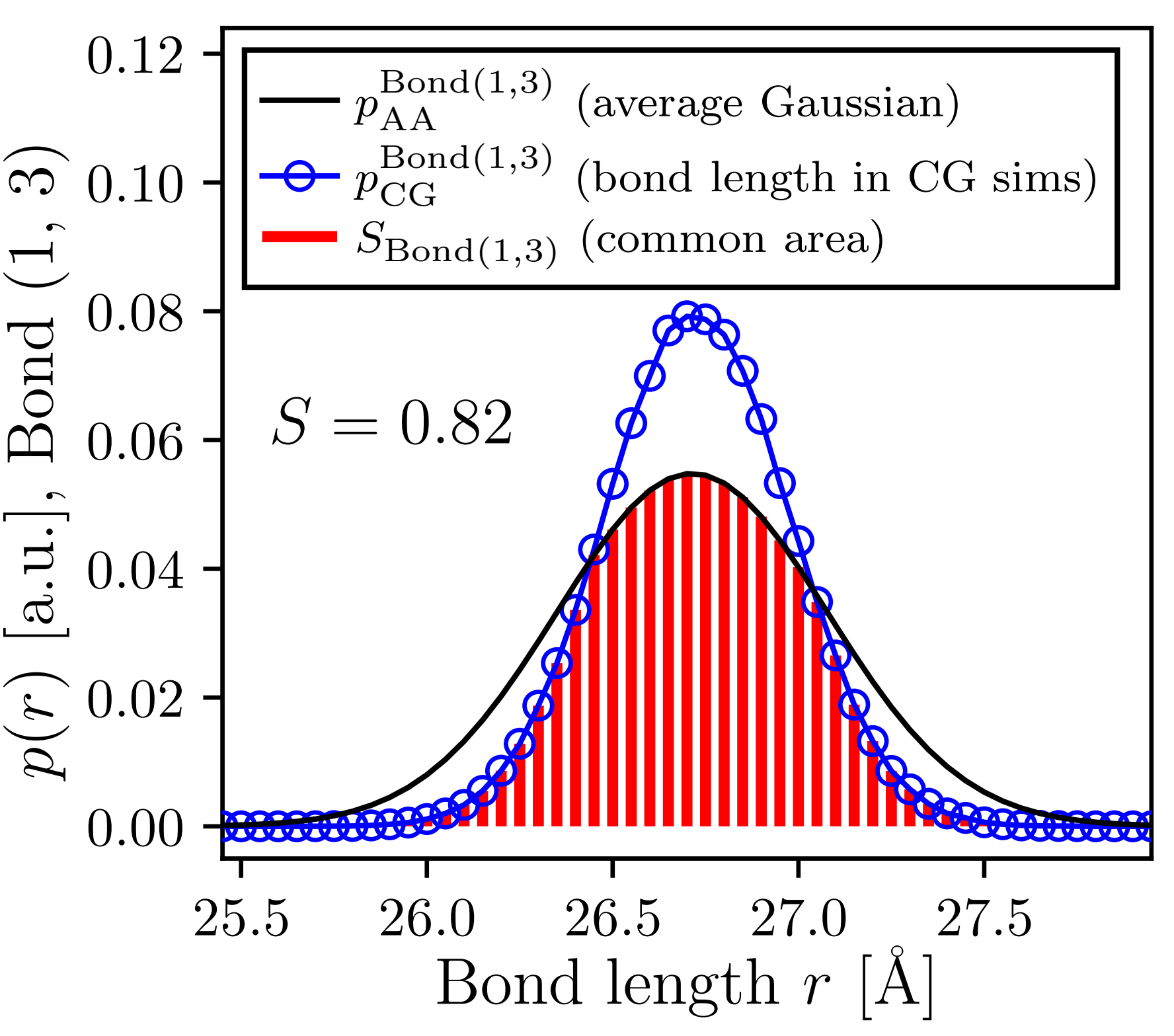

IGGo
IGGrigid
Fit, α_{IgG}=1.66
Eq.(9), β_{IgG}=2.80
FCS, Balbo et al. (2013)
HI
b) IgG
$D_t^l/D_{t,0}$
Volume packing fraction Φ


a)
FULLrigid
b)
IGGrigid
$< P_1[\cos(\theta(t))] >$
$t\ [\tau_B]$
$\Phi = 0$
$\Phi = 0.2$
$\Phi = 0.3$
$\Phi = 0.4$
$\Phi = 0.5$
Axis 1
Axis 2
Axis 3
Axis 4
Axis 5
Axis 6
Axis 7
Axis 8
Axis 9
Axis 10
Axis 11
Axis 12
Φ


1.0
0.8
0.6
0.4
0.2
0.0
$\overline{D_r}/\overline{D_{r,0}}$
FULLo
FULLrigid
STRUCTo
STRUCTnd
STRUCTflex
IGGo
IGGrigid
0.0
0.1
0.2
0.3
0.4
0.5
Volume packing fraction Φ


STRUCTo
IGGo
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.1
0.2
0.3
0.4
0.5
Volume packing fraction Φ
$D_t^l / D_{t,0} \Big/ \overline{D_r} / \overline{D_{r,0}}$


Bond length, averaged over 21 all-atom simulations
Average Gaussian $p_{\mathrm{AA}}^{\mathrm{Bond}}(r)$
Bond length, single all-atom simulation
a)
Bond (2, 3)
b)
Bond (5, 6)
c)
Bond (2, 3)
d)
Bond (5, 6)
$p(r)$ [a.u.]
0.00
0.02
0.04
0.06
23
24
25
26
27
Bond length r [Å]


$p_{\mathrm{AA}}^{\mathrm{Bond}(1,3)}$ (average Gaussian)
$p_{\mathrm{CG}}^{\mathrm{Bond}(1,3)}$ (bond length in CG sims)
$S_{\mathrm{Bond}(1,3)}$ (common area)
$S = 0.82$
$p(r)$ [a.u.], Bond (1, 3)
0.12
0.10
0.08
0.06
0.04
0.02
0.00
25.5
26.0
26.5
27.0
27.5
Bond length r [Å]