Determining the accurate ensemble of protein atomic structures by combining small angle scattering with molecular simulations based on a two-fold scheme for structural classification

Pan Tan[1‡], Zuyue Fu[2,3‡], Loukas Petridis[4,5], Shuo Qian[4], Delin You[6], Dongqing Wei[6], Jinlai Li[2,7] and Liang Hong[1,2*]

[1]*School of Physics and Astronomy, Shanghai Jiao Tong University, Shanghai 200240, China*

[2]*Institute of Natural Sciences, Shanghai Jiao Tong University, Shanghai 200240, China*

[3] *Zhiyuan College, Shanghai Jiao Tong University, Shanghai 200240, China*

[4]*UT/ORNL Center for Molecular Biophysics, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37830, United States*

[5] *Department of Biochemistry, Cellular & Molecular Biology, The University of Tennessee, Knoxville, Tennessee 37996, United States*

[6]*School of Life Science and Biotechnology, Shanghai Jiao Tong University, Shanghai 200240, China.*

[7] *School of Mathematics, and the MOE Key Laboratory of Scientific and Engineering Computing, Shanghai Jiao Tong University, Shanghai 200240, China.*

ABSTRACT: Atomic-level structural characterization of flexible proteins, such as intrinsic disordered proteins and multi-domain proteins connected by flexible linkers, is of great challenge， as they possess distinct conformations in physiological conditions. Significant efforts have been made to develop integrated approaches by combining small angle neutron/X-ray scattering experiments with molecular simulations to reveal the distinct atomic structures and the corresponding populations of these flexible proteins. A widely used method is to seach for the structural ensemble from conformational candidates pre-generated from molecular simulations, which best fit to the scattering data. This method makes an implicit assumption that protein conformations of similar structures have similar small angle scattering (SAS) profiles. The present work demonstrates, for various protein systems ranging from compact globular proteins and flexible multi-domain proteins through to intrinsic disordered proteins, this ensemble-search method furnishes inaccurate assessment of the structural ensemble of the protein molecules due to the failure of the assumption made. To alleviate this problem, a two-fold-clustering method is developed here, which clusters the simulation-generated protein structures using information on both 3D structure and scattering profiles. As benchmarked by both simulation and experimental results, this new method yields much more accurate populations of protein conformations.

Introduction

A central task in molecular biochemistry and molecular biophysics is to determine the atomic structure of proteins at physiological conditions. Although X-ray crystallography and NMR can furnish high-resolution atomic-level structure of bio-macromolecules, they are limited by either the availability of crystalline samples or the size of the macromolecules. These high-resolution techniques can be complemented by low-resolution ones, such as cryo-electron macroscopy, mass spectroscopy and small angle scattering (SAS)[1]. SAS, either with X-ray or neutron, has the advantage of measuring protein structures in physiological conditions[2–8]. However, SAS is inherently limited because the three-dimensional real-space structural information of a bio-macromolecule is reduced to a one-dimensional scattering profile in reciprocal space, resulting in loss of information and the difficulty of converting the SAS intensity to a 3D structure [1,2,6,9–14]. *Ab initio* methods have been developed to re-construct low-resolution representations of the bio-macromolecules by modeling the experimental SAS data using spatially packed spheres of sub nanometer size[15], which, however, lack the atomic-level, or even secondary-structural-level information.

Recently, there have been significant efforts to develop integrated approaches by combining small-angle scattering experiments with molecular simulations to derive the atomic-level structures of bio-macromolecules [1,2,5,10–12,16–18]. These approaches roughly fall into two categories: one is to search for the protein conformations from existing structural candidates, pre-generated from molecular simulation using standard force fields, which best fits to the experimental SAS data[1,2,10–12,18]; while the other one is to modify the simulation force fields to either drive the simulation towards the protein conformation in better agreement with experiment[16] or to enhance sampling[17]. The former one is more widely adopted as it is intuitively straightforward without advanced molecular simulations, and it is thus the focus of the present work.

It becomes increasingly clear that flexible biomolecules, such as multi-domain proteins linked through flexible linkers and intrinsic disordered proteins, possess multiple conformations in the physiological conditions[1–3]. Revealing the populations of distinct conformations of such protein system in solution is of great importance towards the understanding of the enzymatic mechanism. This inspires the development of the ensemble-based SAS approaches, such as ensemble optimization method[5], basis-set supported ensemble method[1–3], and minimal ensemble search[10]. The ensemble-based SAS approaches present a significant advancement in the interpretation of flexible protein conformations as compared to methods which find a single protein structure fitting best to the experimental SAS data[4].

The present work mainly tests the basis-set supported ensemble method[1–3], which is more widely adopted[1–3] and schematically illustrated by Fig. 1a. Briefly, a large pool of candidates of protein conformations, pre-generated using a simulation method, e.g., molecular dynamics simulations (MD) or Monte Carlo simulation, is clustered into a limited number of sub ensembles, *i.e.* representative structures, based on structural similarity as quantified by a structural metric. Here, three structural metrics are used: root-mean-squared deviation (RMSD), distance root-mean-square (DRMS)[1] and dihedral distance[19]. Then, SAS profiles are calculated for each sub ensemble and used as a linear basis to fit against the experimental data to obtain the corresponding populations. This method makes an implicit assumption that structurally

similar conformations have similar SAS profiles, which can be invalid as reported by Ref [12,14]. Here, we demonstrate that, for various protein systems, this ensemble-search method can lead to drastic inaccurate assignment of protein structural ensemble because of the failure of the assumption made. To overcome this deficiency, a two-fold-clustering method is developed, which clusters the structural candidates using information on both structures and SAS profiles. The ability of this new method in predicting the accurate populations of protein structures in solution is confirmed by both simulation and experimental tests.
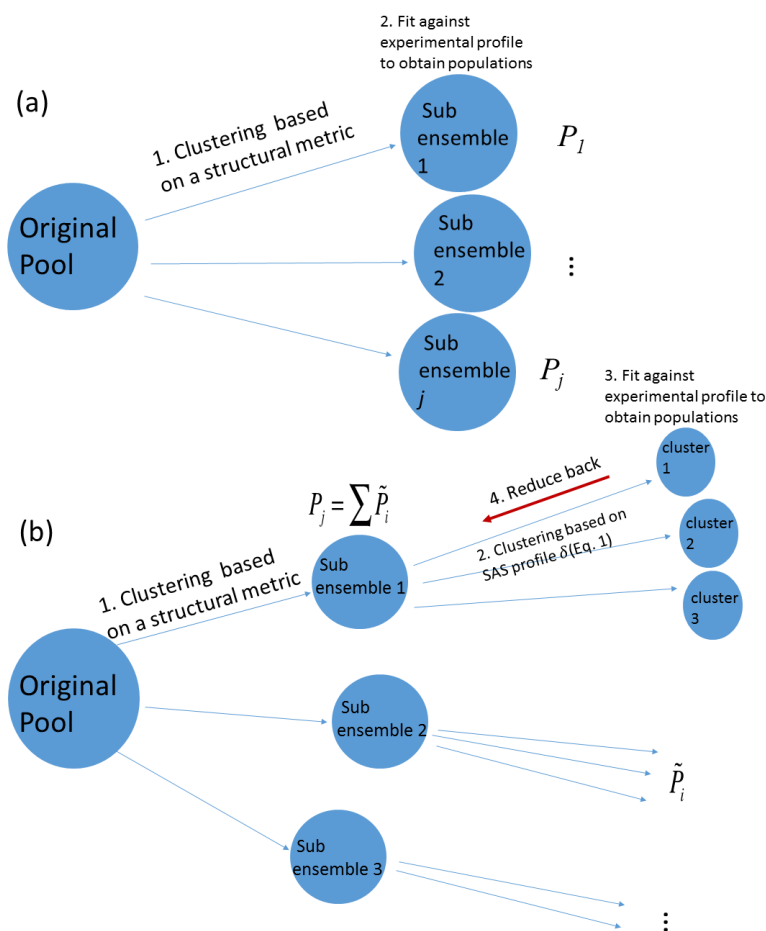
Figure 1, (color online) Schematic illustration for (a) one- and (b) two-fold-clustering methods. Ref.[2] also proposed a two-criteria method using both RMSD and $I(q)$ to cluster simulation-derived protein conformations to reveal the atomic structures of proteins. It is, however, different from the present work. It first clusters the simulation-derived protein conformations to small sub ensembles based on their structural similarity using RMSD, and then merges them to bigger clusters based on similarity in $I(q)$. Finally, these big clusters serve as linear basis to fit against the experimental $I(q)$ to obtain the corresponding populations. Within each of these big clusters, protein conformations can differ significantly despite $I(q)$ is

3

similar. In other words, Ref.2 obtains an ensemble of $I(q)$, from which one can hardly get the information on the populations of different representative structures. In contrast, this is the direct message delivered by the two-fold method proposed in present work.

Method

For simplicity, the basis-set supported ensemble method as shown in Fig. 1a is referred as one-fold-clustering method herein. The two-fold-clustering method introduced here, schematically illustrated in Fig. 1b, contains four steps: 1. The original pool of protein conformations, which are pre-generated by molecular simulations, is clustered into N sub ensembles, using the algorithm developed in Ref[20], based on their structural similarity, as quantified by one of the three structural metrics (RMSD, DMSD or dihedral distance). This step is the same as for the one-fold-clustering method; 2. Protein conformations within each sub ensemble are further divided into several smaller clusters based on their similarity in $I(q)$, quantified by,

$$\delta_{ij} = \frac{1}{L-1}\sum_s \frac{[I_i(q_s)-I_j(q_s)]^2}{\sigma^2(q_s)}, \qquad (1)$$

where, $I_i(q_s)$ and $I_j(q_s)$ are the SAS intensities of the protein conformation i and j, respectively, q_s is the scattering wave vector, L is the total number of the scattering wave vectors, and $\sigma^2(q_s)$ is the variance of $I(q_s)$ among all the conformations. Thus, the protein conformations within each cluster are similar in both 3D structure and $I(q)$. 3. The population of each cluster is obtained by fitting the experimental $I(q)$ to a set of linear basis composed by the scattering profiles calculated from each cluster. 4. The so-obtained populations of the clusters originating from the same sub ensemble are summed together to represent the population of that sub ensemble. Therefore, the two-fold-clustering method yields the same set of sub ensembles (representative structures) as the one-fold-clustering method does, while the corresponding populations could be different.

To test the performance of these two different methods, three proteins with distinct structural features are examined here: an intrinsic disordered protein (IDP, Tau267-312, PDB ID: 2MZ7), a multi-domain protein (mercuric ion reductase, MerA) and a globular single-domain protein (lysozyme). Typical structures of these three systems are displayed in Figs. 2a to c, and the corresponding small angle X-ray scattering (SAXS) profiles, $I(q)$, are presented in Fig. 2d. We employ all-atom molecular dynamics simulation to generate 1000 conformations for each system (see the supporting information for detailed simulation protocols), which are used as the initial pool of protein conformations for cluster analysis.
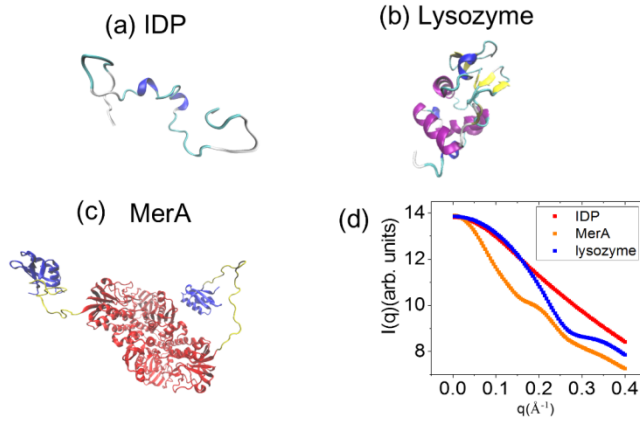
Figure 2, (color online) Snapshots of the simulation systems, (a) IDP (b) lysozyme and (c) MerA. (d) SAXS profiles calculated based on the protein structures displayed in (a), (b) and (c) using software Sassena[21].

Results

Simulation test

In the simulation tests, the averaged I(q) calculated from all snapshots in the first half of the MD trajectories for each system are taken as the target "experimental" data, which will be modeled by the one- and two-fold-clustering methods. As seen in Figs. 3a to c, both methods yield good fits to the "experimental" SAXS data. To quantify the agreement between the fitted and the "experimental" I(q), a score parameter is defined as[4,12]:

$$\chi^2 = \frac{1}{L-1} \sum \frac{\left[\sum P_i * I_i(q) - I_{exp}(q)\right]^2}{\sigma^2(q)}, \qquad (2)$$

where P_i and $I_i(q)$ are the fitted population and small angle scattering profiles for Sub Ensemble i, $I_{exp}(q)$ is the "experimental" SAXS profile, and $\sigma(q)$ is the standard deviation among all $I_i(q)$. The values of χ^2 are displayed in Table I, showing that one-fold method agrees better with the "experimental" results.
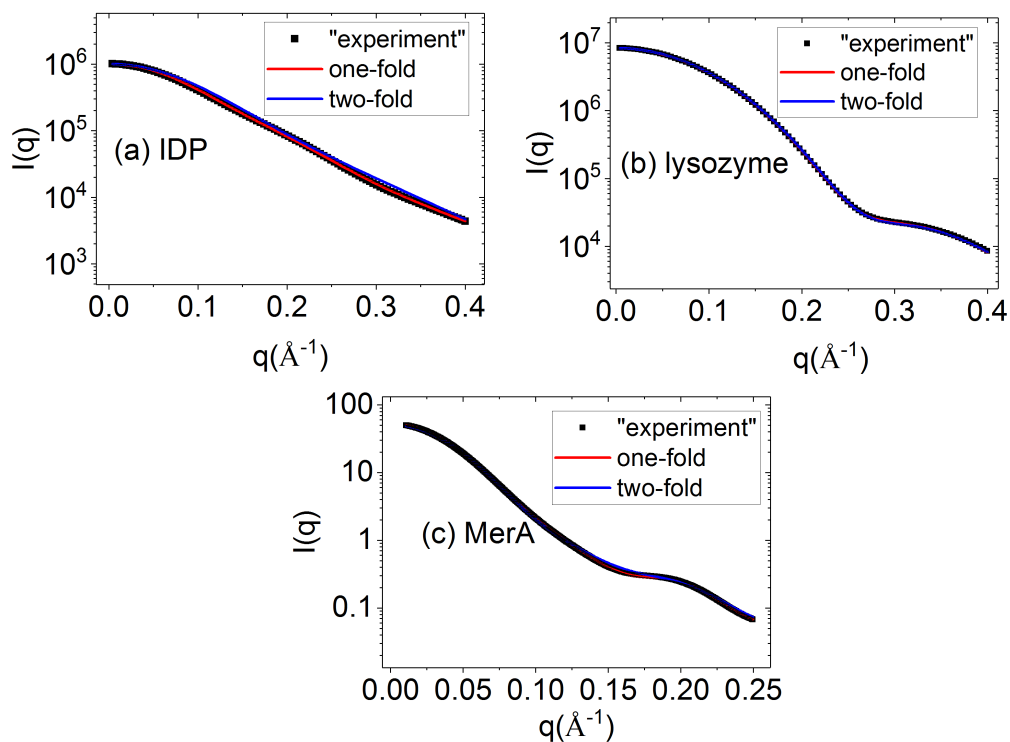
Figure 3, (color online) Comparison of "experimental" SAXS profiles with those derived from one- and two-fold-clustering methods for (a) IDP (b) lysozyme and (c) MerA. Error bars are included, but they are very tiny and comparable to the size of the symbols.

Table 1 Scoring parameters, χ^2, the cutoffs of RMSD, DRMS and dihedral distance, and the ratio between $\gamma_{one\text{-}fold}$ and $\gamma_{two\text{-}fold}$. The analysis is based on SAXS profiles. To prevent overfitting, the maximum-entropy method[1] is employed for the simulation tests. Detailed procedures can be found in SI.

		RMSD	DRMS	Dihedral distance
lysozyme	χ^2_{one}	0.6183	0.1882	1.0447
	χ^2_{two}	1.3052	0.2751	1.7537
	$\gamma_{one}/\gamma_{two}$	4.2653	3.1623	5.0681
	cutoff	1.6Å	1 Å	0.15
IDP	χ^2_{one}	1.8951	1.2619	0.4810
	χ^2_{two}	3.8908	7.5153	2.2190
	$\gamma_{one}/\gamma_{two}$	4.0954	14.4610	6.9842
	cutoff	12 Å	10 Å	0.52
MerA	χ^2_{one}	0.7662	0.5100	0.5039
	χ^2_{two}	1.3848	1.4313	2.2042
	$\gamma_{one}/\gamma_{two}$	4.1160	5.6464	3.2002
	cutoff	6.5 Å	4 Å	0.16

Although both methods provide seemingly excellent fits to the experimental data (Fig. 3), the resulting populations of sub ensembles differ dramatically (see Figs. 4a to c). For comparison, the true populations of sub ensembles constituting the "experimental" portions of protein conformations, calculated directly from the simulation, are also plotted, and are found to be in much better agreement with those derived from the two-fold-clustering method. (The detailed procedure for how to determine the true populations is presented in Fig. S4 in the supporting information.) To quantify the accuracy of these two methods in predicting the populations of protein conformations, an error parameter, γ, is defined:

$$\gamma = \sqrt{\frac{\Sigma_i^N (P_i - P_i^*)^2}{\Sigma_i^N P_i^{*2}}} \tag{3}$$

where P_i and P_i^* are the fitted and true populations for a given Sub Ensemble i, respectively, and N is the total number of sub ensembles. Thus, γ quantifies the deviation between the fitted and the true populations, $i.e.$, smaller γ indicates better agreement to the true populations. As seen in Figs. 4d, 4e and Table I, $\gamma_{one-fold}$ is many folds larger than $\gamma_{two-fold}$, although $I(q)$ derived from the one-fold method agrees better with the "experimental" results (see Table 1). These findings are independent of which portion of the 1000 MD conformations is used to represent the "experimental" I(q) (see Fig. S1 in the supporting information), independent of the size, compactness, secondary and tertiary structures of the system studied, independent of the choice of the structural metric, and independent of whether SAXS or SANS profiles (see Table S1) are considered.
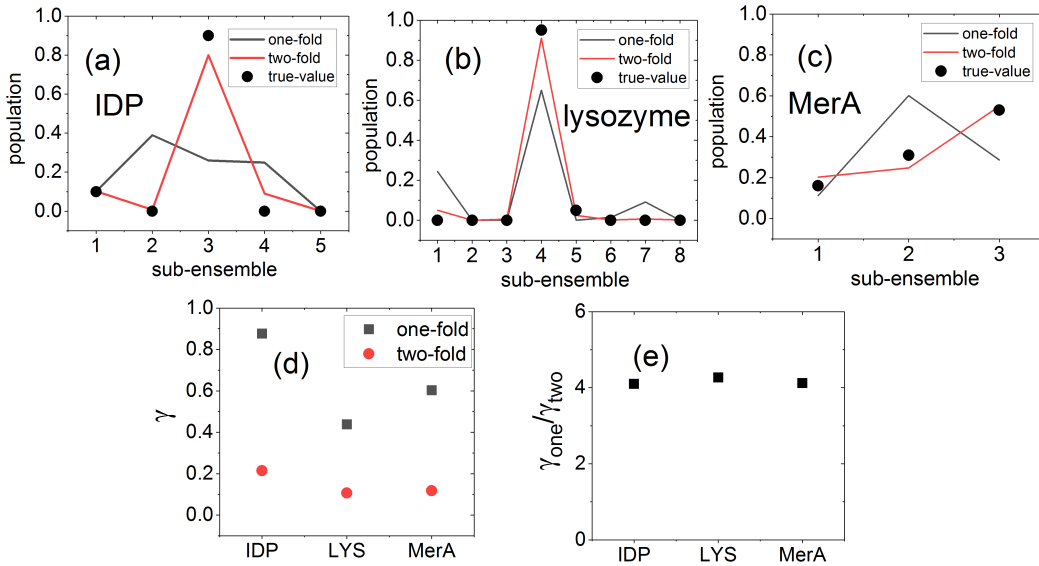

Figure 4, (color online) Simulation test on the performance of one-fold and two-fold-clustering methods. Comparison of the true populations of sub ensembles constituting the "experimental" portion of protein conformations with the populations derived from one-fold and two-fold methods by modeling the "experimental" $I(q)$ for (a) IDP, (b) lysozyme and (c) MerA. (d) Relative errors, γ, defined in Eq. (3) resulting from one- and two-fold methods and (e) the ratio of $\gamma_{one}/\gamma_{two}$ for different systems. A Monte Carlo procedure is used to derive the populations of sub ensembles using the one- and two-fold methods, details of which can be found in supporting information. All the results presented in this figure are obtained by using RMSD as the structural metric. Using other structural metrics will furnish qualitative similar results (see

Table I).

The improved performance of the two-fold approach arises from not assuming that the structurally proximate protein conformations possess similar $I(q)$, an assumption which can be invalid. This is evident by Figs. S2 and S3, which show no strong correlation between the structural difference (RMSD, DMSD or dihedral distance) and δ (difference between $I(q)$, Eq. (1)) for every two conformations of a given protein system. Similar findings have been also reported by others[12,14] and should be generally applicable to bio-macromolecules as the test is conducted herein over a wide range of systems, and independent on whether small angle X-ray or neutron scattering is considered (see Figs. S2 and S3 in the supporting information).

Experimental test

In addition to the simulation test, we also examined the performance of the two methods using the experimental SAXS data of MerA[4]. They furnish fits of similar quality to the experimental $I(q)$ (Fig. 5a), but yield vastly different populations of representative structures, i.e., sub ensembles (see Fig. 5b). The two-fold method predicts only one dominant (86%) protein structure, while the one-fold method furnishes two highly populated structures (57% and 43%). More importantly, the one dominant protein structure identified by the two-fold method does not agree with any of the two structures revealed by the one-fold method.
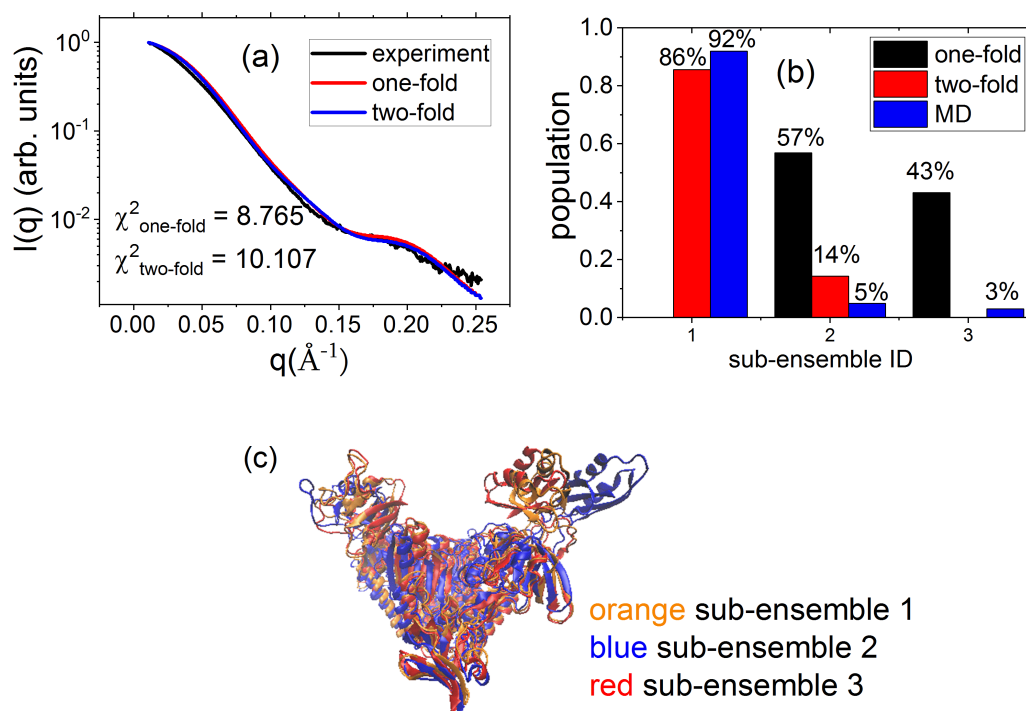

Figure 5 (color online) Experimental test on the performance of one- and two-fold-clustering methods. (a) Comparisons of $I(q)$ of MerA obtained from the SAXS experiment[4] with that derived from one- (RMSD) and two-fold methods. (b) Populations of sub ensembles derived from one-fold (black), two-fold methods (red) and directly from the MD trajectory (blue). (c) Structural representatives for Sub Ensemble 1 (orange), Sub Ensemble 2 (blue) and Sub Ensemble 3 (red) obtained through cluster analysis on the simulation-generated protein
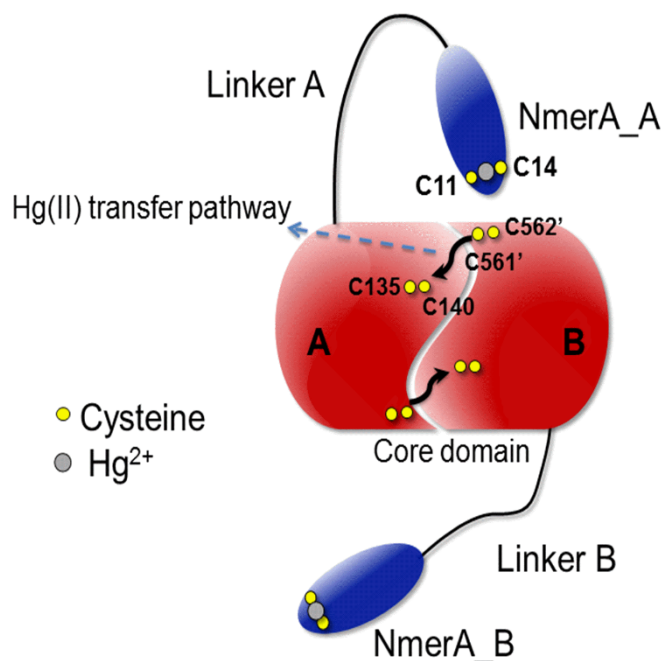
conformations.

Figure 6: A schematic picture of MerA where the crucial residues for the enzymatic function are labeled.

MerA possesses metallochaperone-like N-terminal domains (NMerA), which are tethered to the homodimeric catalytic core by ~ 30 residue linkers [22,23] (Fig. 6). NMerA contains a pair of cysteine residues in a GMTCXXC sequence motif that is conserved in those soft metal ion trafficking proteins sharing the common $\beta\alpha\beta\beta\alpha\beta$ structural fold[24]. MerA binding and processing of Hg(II) is schematically illustrated in Fig. 6. Hg(II) binds first to a pair of cysteines (C11 and C14) in NMerA, these being the most solvent-accessible cysteines in the protein [24,25]. The ion is then transferred to another pair of cysteines (C561' and C562') located on the mobile C-terminal segment of the core. The mobile segment then moves the C-terminal Hg(II)-bound cysteines from the surface to the interior of the protein, from which the ions is further delivered to the buried, active-site cysteine pair (C135 and C140) where the reduction occurs[26,27]. *In vivo* experiments showed that the presence of NMerA domains significantly enhances cell survival in the presence of Hg(II)[25], in which the high-affinity Hg(II)-chelating NMerA domains are essential for rapid acquisition, localization and transfer of Hg(II) to the core for reduction and detoxification[25]. Ref.[28], by combining small angle scattering, coarse-grained simulation and all-atom molecular dynamics simulation, revealed that MerA adopts a highly compact structure in solution, where the NMerA domains are leashed by the linkers to bound strongly with the core domain, being close to the C-terminal Hg(II)-binding cysteines for rapid delivery of mercury ions. It is further confirmed by neutron spin echo measurements, which showed no appreciable inter-domain dynamics being present between the NMerA and core domains[28].

As shown by Table 2, Sub Ensemble 2 (Blue in Fig. 5c) is a rather extended conformation, in which one of its NMerA domains is detached significantly from the catalytic core (marked in

red in Table 2), thus contradicting the findings of Ref.[28]. This unlikely conformation is highly populated (57%) as revealed by the one-fold method, but weights little (10%) as predicted by the two-fold method. Moreover, the most dominant structure identified by the two-fold method, i.e., Sub Ensemble 1 (Orange in Fig. 5c), has on-average the shortest distances between C_{11} and C_{561}' and between C_{11}' and C_{561}, *i.e.*, characteristic distances measuring the capability of the linker to leash the NMerA domain around the core C-terminal Hg(II)-binding cysteines, which is in better agreement with the finding of Ref.[28].

Table 2 Structural characteristics of MerA conformations revealed by one- and two-fold-clustering methods by modeling the experimental SAXS data[4]. The values inside brackets represent one standard deviation.

	sub-ensemble ID	population	Minimum distance between each of the two NmerAs and the core domain (Å)	$C_{11}\text{-}C_{516}'$ / $C_{11}'\text{-}C_{516}$ (Å)
one-fold	2	57%	2.2($\pm$0.2)/7.2($\pm$2.9)	58($\pm$1.0)/71($\pm$3.0)
	3	43%	3.63($\pm$3.1)/1.81($\pm$0.1)	52($\pm$3.3)/71($\pm$3.0)
two-fold	1	86%	3.0($\pm$2.7)/1.78($\pm$0.9)	51($\pm$3.4)/68($\pm$4.0)
	2	14%	2.2($\pm$0.2)/7.2($\pm$2.9)	58($\pm$1.0)/71($\pm$3.0)

We also compared the populations of sub ensembles of MerA structures derived by one-fold and two-fold methods with those obtained directly from MD. As shown in Fig. 5b, the results obtained using the two-fold method is in much better agreement with the MD-derived ones. This again supports the two-fold method as the MD simulation is based on rational force fields and its results are in agreement with experimental data[28]. Moreover, as found in Ref[28], there is very strong electro-static attraction between NmerA domain and the catalytic core, that will drive the NmerA to attach to the surface of the catalytic core, and thus to perform a highly retarded diffusive motion on the core. Therefore, Sub ensemble 2, where the NmerA is detached significantly (~ 7 Å) away from the catalytic core, is unlikely to be highly populated, disfavoring the one-fold method.

In summary, the experimental test also favors the protein conformational ensemble revealed by the two-fold clustering method.

Conclusion

By studying three proteins with different flexibility - an intrinsic disordered protein, a compact single-domain globular protein and a multi-domain protein, the present work demonstrates that the basis-supported ensemble method can give rise to dramatically inaccurate assessment of the populations of protein conformations in solution when modeling

the small angle scattering experimental data. This arises from the failure of the assumption that structurally similar protein conformations have similar small angle scattering profiles. Here, a two-fold clustering method is developed, which classifies the simulation-derived protein conformations based on similarity on both 3D structure and scattering profiles. As confirmed by both simulation and experimental results, this new method provides exact the same set of representative protein structures as the basis-supported ensemble method does, but much more accurate populations.

Corresponding Author:

[*]hongl3liang@sjtu.edu.cn

Author Contributions:

[‡]These authors contributing equally to this work.

REFERENCES:

(1)   Róycki, B.; Kim, Y. C.; Hummer, G. SAXS Ensemble Refinement of ESCRT-III CHMP3 Conformational Transitions. *Structure* **2011**, *19* (1), 109–116.

(2)   Yang, S.; Blachowicz, L.; Makowski, L.; Roux, B. Multidomain Assembled States of Hck Tyrosine Kinase in Solution. *Proc. Natl. Acad. Sci.* **2010**, *107* (36), 15757–15762.

(3)   Boura, E.; Rozycki, B.; Herrick, D. Z.; Chung, H. S.; Vecer, J.; Eaton, W. A.; Cafiso, D. S.; Hummer, G.; Hurley, J. H. Solution Structure of the ESCRT-I Complex by Small-Angle X-Ray Scattering, EPR, and FRET Spectroscopy. *Proc. Natl. Acad. Sci.* **2011**, *108* (23), 9437–9442.

(4)   Johs, A.; Harwood, I. M.; Parks, J. M.; Nauss, R. E.; Smith, J. C.; Liang, L.; Miller, S. M. Structural Characterization of Intramolecular Hg 2+ Transfer between Flexibly Linked Domains of Mercuric Ion Reductase. *J. Mol. Biol.* **2011**, *413* (3), 639–656.

(5)   Bernadó, P.; Mylonas, E.; Petoukhov, M. V.; Blackledge, M.; Svergun, D. I. Structural Characterization of Flexible Proteins Using Small-Angle X-Ray Scattering. *J. Am. Chem. Soc.* **2007**, *129* (17), 5656–5664.

(6)   Trewhella, J. Small-Angle Scattering and 3D Structure Interpretation. *Curr. Opin. Struct. Biol.* **2016**, *40*, 1–7.

(7)   Graewert, M. A.; Svergun, D. I. Impact and Progress in Small and Wide Angle X-Ray Scattering (SAXS and WAXS). *Curr. Opin. Struct. Biol.* **2013**, *23* (5), 748–754.

(8)   Rambo, R. P.; Tainer, J. A.; Division, S.; Source, A. L.; Berkeley, L.; Biology, C.; Jolla, L. Accurate Assessment of Mass, Models and Resolution by Small- Angle Scattering. *Nature* **2013**, *496* (7446), 477–481.

(9)   Svergun, D. I. Small-Angle Scattering Studies of Macromolecular Solutions. *J. Appl. Crystallogr.* **2007**, *40* (s1), s10–s17.

(10)   Pelikan, M.; Hura, G. L.; Hammel, M. Structure and Flexibility within Proteins as Identified through Small Angle X-Ray Scattering. *Gen. Physiol. Biophys.* **2009**, *28* (2), 174–189.

(11)   Förster, F.; Webb, B.; Krukenberg, K. A.; Tsuruta, H.; Agard, D. A.; Sali, A. Integration of Small-Angle X-Ray Scattering Data into Structural Modeling of Proteins and Their Assemblies. *J. Mol. Biol.* **2008**, *382* (4), 1089–1106.

(12)   Vandavasi, V. G.; Putnam, D. K.; Zhang, Q.; Petridis, L.; Heller, W. T.; Nixon, B. T.; Haigler, C. H.; Kalluri, U.; Coates, L.; Langan, P.; Smith, J. C.; Meiler, J.; O'Neill, H. A Structural Study of CESA1 Catalytic Domain of Arabidopsis Cellulose Synthesis Complex: Evidence for CESA Trimers. *Plant Physiol.* **2016**, *170* (1), 123–135.

(13)   Panjkovich, A.; Svergun, D. I. Deciphering Conformational Transitions of Proteins by Small Angle X-Ray Scattering and Normal Mode Analysis. *Phys. Chem. Chem. Phys.* **2016**, *18* (8), 5707–5719.

(14)   Kim, H. S.; Gabel, F. Uniqueness of Models from Small-Angle Scattering Data: The Impact of a Hydration Shell and Complementary NMR Restraints. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2015**, *71* (Pt 1), 57–66.

(15)   Svergun, D.; Barberato, C.; Koch, M. H. CRYSOL - A Program to Evaluate X-Ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. *J. Appl. Crystallogr.* **1995**, *28* (6), 768–773.

(16)   Björling, A.; Niebling, S.; Marcellini, M.; Van Der Spoel, D.; Westenhoff, S. Deciphering Solution Scattering Data with Experimentally Guided Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **2015**, *11* (2), 780–787.

(17)   Kimanius, D.; Pettersson, I.; Schluckebier, G.; Lindahl, E.; Andersson, M. SAXS-Guided Metadynamics. *J. Chem. Theory Comput.* **2015**, *11* (7), 3491–3498.

(18)   Naganathan, A. N.; Orozco, M. The Native Ensemble and Folding of a Protein Molten-Globule: Functional Consequence of Downhill Folding. *J. Am. Chem. Soc.* **2011**, *133* (31), 12154–12161.

(19)   Cossio, P.; Laio, A.; Pietrucci, F. Which Similarity Measure Is Better for Analyzing Protein Structures in a Molecular Dynamics Trajectory? *Phys. Chem. Chem. Phys.* **2011**, *13* (22), 10421–10425.

(20)   Daura, X.; Gademann, K.; Jaun, B.; Seebach, D.; van Gunsteren, W. F.; Mark, A. E. Peptide Folding: When Simulation Meets Experiment. *Angew. Chemie Int. Ed.* **1999**, *38* (1–2), 236–240.

(21)   Lindner, B.; Smith, J. C. Sassena - X-Ray and Neutron Scattering Calculated from Molecular Dynamics Trajectories Using Massively Parallel Computers. *Comput. Phys. Commun.* **2012**, *183* (7), 1491–1501.

(22)   Barkay, T.; Miller, S. M.; Summers, A. O. Bacterial Mercury Resistance from Atoms to Ecosystems. *FEMS Microbiol. Rev.* **2003**, *27* (2–3), 355–384.

(23)   Barkay, T.; Kritee, K.; Boyd, E.; Geesey, G. A Thermophilic Bacterial Origin and Subsequent Constraints by Redox, Light and Salinity on the Evolution of the Microbial Mercuric Reductase. *Environ. Microbiol.* **2010**, *12* (11), 2904–2917.

(24)   Ledwidge, R.; Hong, B.; Dötsch, V.; Miller, S. M. NmerA of Tn 501 Mercuric Ion Reductase: Structural Modulation of the P K a Values of the Metal Binding Cysteine Thiols. *Biochemistry* **2010**, *49* (41), 8988–8998.

(25)   Ledwidge, R.; Patel, B.; Dong, A.; Fiedler, D.; Falkowski, M.; Zelikova, J.; Summers, A. O.; Pai, E. F.; Miller, S. M. NmerA, the Metal Binding Domain of Mercuric Ion Reductase, Removes Hg 2+ from Proteins, Delivers It to the Catalytic Core, and Protects Cells under Glutathione-Depleted Conditions. *Biochemistry* **2005**, *44* (34), 11402–11416.

(26)   Engst, S.; Miller, S. M. Rapid Reduction of Hg(II) by Mercuric Ion Reductase Does Not Require the Conserved C-Terminal Cysteine Pair Using HgBr2 as the Substrate. *Biochemistry* **1998**, *37* (33), 11496–11507.

(27)   Engst, S.; Miller, S. M. Alternative Routes for Entry of HgX2 into the Active Site of Mercuric Ion Reductase
         Depend on the Nature of the X Ligands. *Biochemistry* **1999**, *38* (12), 3519–3529.

(28)   Hong, L.; Sharp, M. A.; Poblete, S.; Biehl, R.; Zamponi, M.; Szekely, N.; Appavou, M.-S. S.; Winkler, R. G.;
         Nauss, R. E.; Johs, A.; Parks, J. M.; Yi, Z.; Cheng, X.; Liang, L.; Ohl, M.; Miller, S. M.; Richter, D.; Gompper, G.;
         Smith, J. C. Structure and Dynamics of a Compact State of a Multidomain Protein, the Mercuric Ion Reductase.
         Biophys. J. **2014**, *107* (2), 393–400.