Biophysical *Journal*
Computational Tool

Biophysical Society

SAXS-Oriented Ensemble Refinement of Flexible Biomolecules

Peng Cheng,[1] Junhui Peng,[1] and Zhiyong Zhang[1,*]
[1]Hefei National Laboratory for Physical Science at Microscale and School of Life Sciences, University of Science and Technology of China, Hefei, Anhui, People's Republic of China

ABSTRACT The conformational flexibility of a biomolecule may play a crucial role in its biological function. Small-angle x-ray scattering (SAXS) is a very popular technique for characterizing biomolecule flexibility. It can be used to determine a possible structural ensemble of the biomolecule in solution with the aid of a computer simulation. In this article, we present a tool written in Python, which iteratively runs multiple independent enhanced sampling simulations such as amplified collective motions and accelerated molecular dynamics, and an ensemble optimization method to drive the biomolecule toward an ensemble that fits the SAXS data well. The tool has been validated with a protein and an RNA system, i.e., the tandem WW domains of formin-binding protein 21 and the aptamer domain of SAM-1 riboswitch, respectively. These Python scripts are user-friendly and can be easily modified if a different simulation engine is preferred.

INTRODUCTION

Small-angle x-ray scattering (SAXS) has been widely used to obtain low-resolution structural information for large biomolecules in solution (1). Compared to other experimental techniques, SAXS is particularly useful for characterizing the flexibility of biomolecules (2). Many recent studies show the possibility of combining experimental SAXS data and computational simulation to interpret biomolecule dynamics in solution (3,4).

Two strategies, refining-while-sampling and screening-after-sampling, are generally applied to integrate low-resolution SAXS data into the structural modeling of biomolecules (4). For a refining-while-sampling method such as found in Zheng and Tekpinar (5) and Björling et al. (6), a pseudo energy term based on the SAXS data is designed, then a conformation or an ensemble is simulated in optimizing the energy. Such an approach is efficient, but source code must be modified to change the energy function, which may not be an easy job for someone with little programming skill and knowledge about simulation algorithms; furthermore, different research groups prefer different simulation engines. In screening-after-sampling methods such as those found in the literature (7–14), simulations without SAXS restraint are first carried out to generate a large pool of structures that covers the conformational space of the biomolecule. Then, an ensemble containing a small number of conformations selected from the pool is determined to fit the SAXS data. This strategy is particularly suitable for flexible systems. Although there is no need to change the simulation code, adequate sampling of the conformational space is crucial for such an approach—which is a nontrivial issue, especially for very large biomolecules.

To our knowledge, this article presents a new computational tool for the SAXS-oriented ensemble refinement (SAXS-ER) of flexible biomolecules, which has the advantages of both strategies, but not the disadvantages. Our tool consists of cycles of 1) multiple independent enhanced sampling simulation, and 2) selection of an ensemble that contains a certain number of conformations from the combined trajectories, which best reproduce the SAXS data at this stage, to start the next simulation cycle. We designated this approach as an iterative screening-after-sampling strategy, and it may drive the ensemble of the biomolecule to fit the SAXS data better and better until it converges.

In the Materials and Methods, we introduce the biomolecular systems of interest, the SAXS data acquisition and the computational details of SAXS-ER. In the Results and Discussion, we apply this tool to two biomolecules, the tandem WW domains of formin-binding protein 21 (FBP21-WWs) and the ligand-free SAM-1 riboswitch aptamer domain (free SAM-1 aptamer). Finally, we provide concluding remarks.

Submitted September 20, 2016, and accepted for publication February 16, 2017.

*Correspondence: zzyzhang@ustc.edu.cn
Peng Cheng and Junhui Peng contributed equally to this work.

Editor: Jill Trewhella.
http://dx.doi.org/10.1016/j.bpj.2017.02.024

MATERIALS AND METHODS

Biomolecular systems and SAXS data

The formin binding protein 21 (FBP21) is a structural component of the mammalian spliceosomal A/B complex, which plays an important role in pre-RNA splicing (15). The C-terminus of the protein consists of two group-III WW domains (designated "FBP21-WWs"), and a NMR ensemble containing 20 structural models (i.e., PDB: 2JXW) has been resolved (16). Each model has 75 amino acid residues. The individual domains are denoted as "WW1" (residues 6–32) and "WW2" (residues 47–73), respectively, and are structurally well converged. However, a flexible linker (residues 33–46) allows various orientations between the two domains. Because long-range NMR restraints are lacking, we have collected SAXS data for FBP21-WWs (17) to determine the structural ensemble of the protein in solution.

The SAM-1 riboswitch is a RNA element that binds to S-Adenosyl Methionine (SAM) and controls expression of genes for Met and Cys biosynthesis in Gram-positive bacteria (18). The aptamer domain contains 94 nucleotides of the SAM-1 riboswitch in the presence of SAM, and magnesium can form a stable structure (19). However, Stoddard et al. (20) have found that the solution scattering data of the ligand-free SAM-1 aptamer shows an obvious discrepancy with the theoretical curve of the ligand-bound crystal structure, and the former may exist as multiple states in solution. Starting from the crystal structure with the ligand removed, we wanted to construct an ensemble of the free SAM-1 aptamer to reproduce the SAXS data. The SAXS data of the free SAM-1 aptamer was taken from www.bioisis.net with the Bioisis ID: 1SAMRR, and more details can be found in Stoddard et al. (20).

Enhanced sampling techniques

Although an MD simulation is popularly used to generate the conformation of a biomolecule (21), a sampling issue may often arise (22). A flexible biomolecule usually exists in multiple conformational states in solution. In a standard MD simulation on a limited timescale, the biomolecule may be trapped in few local states, so that global conformational transitions are rarely sampled due to the complicated energy landscape of the biomolecule (23). Such inefficient sampling in a MD simulation may not be able to properly interpret the experimental SAXS data. Enhanced sampling techniques (24) can be used to resolve this problem.

We previously developed a sampling method known as amplified collective motions (ACM) (25), which calculates a few low-frequency normal modes for a biomolecule in an elastic network model (26), and couples these modes in a high temperature bath in atomic MD simulations to adequately explore the collective motion of the biomolecule. Accelerated molecular dynamics (aMD) (27) improves the sampling efficiency in reducing the energy barrier separating adjacent conformational states of the biomolecule. The method modifies the potential energy and raises these energy wells below a predetermined threshold value, which may allow the biomolecule to sample its conformational space extensively.

The ACM method implemented in the GROMACS-4.5.5 package (28) was used for FBP21-WWs, and aMD encoded in AMBER14 (29) was used for the free SAM-1 aptamer. All of the simulation details are described in the Supporting Material. The sampling efficiency of ACM and aMD compared to the standard MD are shown in Figs. S1 and S2, respectively, in both the root mean square deviation (RMSD) to the starting structures and the principal component analysis (PCA) (30). A description of how to carry out PCA is presented in the Supporting Material.

Ensemble optimization method

The ensemble optimization method (EOM) (7) was used to select an ensemble containing a small number of conformations of the biomolecule from a large structural pool generated in enhanced sampling to fit the SAXS data. Several programs in the ATSAS-2.4.3-1 package (31) were run sequentially. A theoretical scattering profile was calculated in CRYSOL (32) for each conformation in the pool. A master file combining all of the scattering intensities was created in ONEFILE. The program GAJOE was run twice. The first run created the file listing sizes (R_g and D_{max}) for all of the conformations, and the second run produced the final EOM. The search procedure for the EOM is to minimize the fitting residual between the experimental and calculated SAXS profiles using the genetic algorithm:

$$\chi = \left\{ \frac{1}{K-1} \sum_{i=1}^{K} \left[\frac{\mu I_{cal}(q_i) - I_{exp}(q_i)}{\sigma(q_i)} \right]^2 \right\}^{1/2}, \qquad (1)$$

where K is the number of data points in $I_{exp}(q)$, and $\sigma(q)$ are experimental errors, $I_{cal}(q)$ is the average of the theoretical scattering profiles of these conformations in the ensemble, and μ is a scaling factor. An automatic subtraction of a constant from the experimental data is allowed to facilitate the fitting. The momentum transfer q equals to $4\pi \sin\theta/\lambda$, in which 2θ is the scattering angle and λ is the wavelength. After a certain number of cycles of independent search (the default number is 50), the ensemble with the minimal χ was chosen. The ensemble size was set to a default value of 20 and no repetition was allowed.

Recently, an advanced EOM 2.0 was developed (33), in which the ensemble size can be optimized during the search procedure. We also used this new version of EOM (as implemented in the ATSAS-2.7.2 package) in our SAXS-ER.

SAXS-ER procedure

This tool uses several steps (Fig. 1). Python scripts for running the entire process using either ACM or aMD and corresponding tutorials are freely available under a GNU public license from the website https://github.com/pcheng27/SAXS-ER/tree/v1.1, and the DOI is http://doi.org/10.5281/zenodo.243155.

1) Set up the system starting from an initial structure of the biomolecule, and perform a preliminary enhanced sampling (ACM or aMD) simulation.
2) Run EOM on the structural pool generated in the simulation, and pick the ensemble with the minimal χ (Eq. 1) to the SAXS data. The ensemble size is N_{es}.
3) Starting from the N_{es} conformers selected in EOM, N_{sim} ($\geq N_{es}$) independent ACM/aMD simulations are carried out. Multiple independent short-time simulations may achieve a better sampling than a single long-time simulation (34).
4) Run EOM on the structural pool in combining the N_{sim} trajectories.
5) Repeat steps 3 and 4 for 10 cycles. If both χ and the average R_g ($\langle R_g \rangle$) of the ensemble converge, stop the simulation and pick the final ensemble. Otherwise, run the simulation for another 10 cycles.

RESULTS AND DISCUSSION

FBP21-WWs

We ran a SAXS-ER starting from model 1 in the NMR structures of FBP21-WWs (16). The ensemble size of EOM was $N_{es} = 20$. Each cycle consisted of $N_{sim} = 20$ independent 100-ps ACM simulations, in which the low-frequency collective motions were coupled at 500 K. The evolution of the sampled conformational space of FBP21-WWs with the iteration cycles are shown in projections onto the PCA modes (Fig. S3). The SAXS profiles of the ensembles are

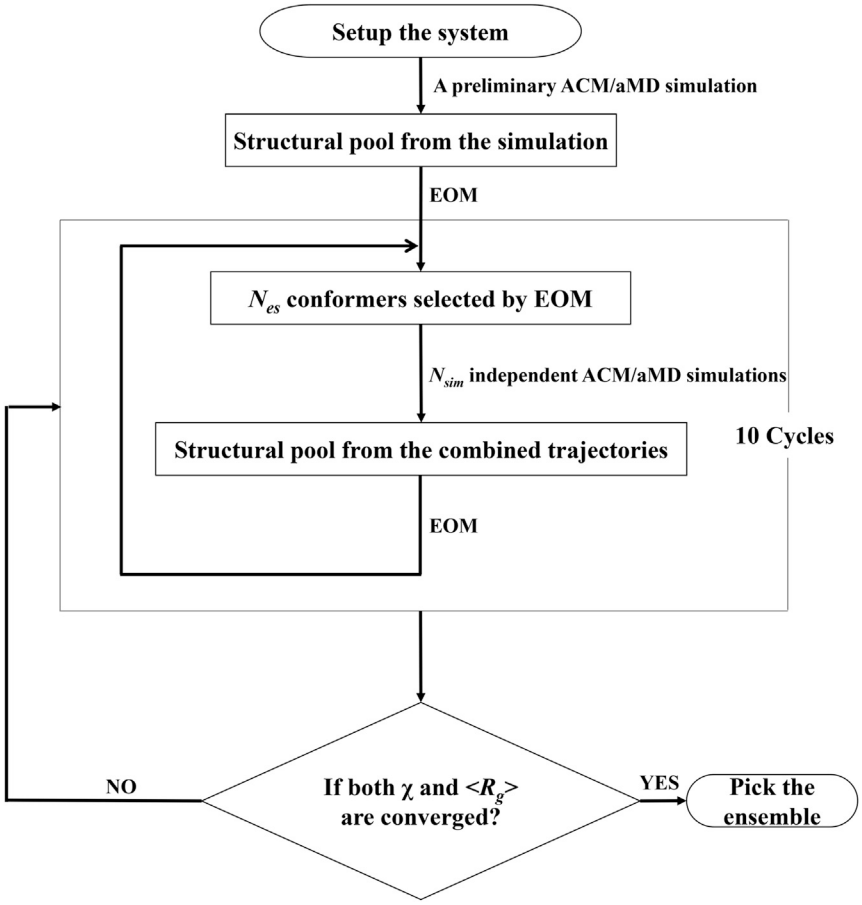

FIGURE 1 General workflow of SAXS-ER.

plotted against cycle number in Fig. S4 to show how they evolve to fit the experimental data. It is found that χ converges very fast in the first three cycles (Fig. 2 *a*, *squares*) while $\langle R_g \rangle$ of the corresponding ensemble (Fig. 2 *a*, *triangles*) starts to saturate at approximately the fifth cycle (Fig. 2 *a*, *red triangle*). Considering the low-resolution nature of SAXS data, we used both metrics as criteria to select the final ensemble, to avoid overfitting. Therefore, the ensemble of 20 conformations at the fifth cycle (Fig. 2 *b*, created by VMD (35)) was selected to reproduce the SAXS profile. The χ-values of individual SAXS curves in the ensemble range from 0.48 to 0.87 (Fig. S5 *a*), but their average profile gives a χ of 0.16 to the experimental data (Fig. 2 *c*), which is significantly less than the value of the initial ensemble (Fig. S5 *b*). This result suggests that the protein should be represented in an ensemble of different conformers. The R_g values of these conformations in the ensemble range from 16.0 to 26.6 Å, with $\langle R_g \rangle$ of 19.9 Å, which is close to the value (19.6 $\pm$ 0.4 Å) estimated from the experimental data. This result indicates that the protein may take either a compact or extended conformation in solution, which is consistent with the EOM result for long MD trajectories with 2 μs (36). However, it should be noted that the total timescale of the SAXS-ER is up to 20 ns (100 ps $\times$ 20 conformers $\times$ 10 cycles). Our tool is efficient because it needs only a relatively short simulation time to obtain an appropriate structural ensemble to fit the SAXS data.

Starting from different conformations, multiple SAXS-ER runs were carried out with different simulation times for each cycle, or different ACM parameters. These ensembles (Fig. S6) are rather similar to the one shown in Fig. 2 *b* in the relative orientation between the two WW domains. We also tried a SAXS-ER using EOM 2.0 (33), in which the ensemble size N_{es} can be optimized (Fig. S7). After the 18th cycle, both χ and $\langle R_g \rangle$ became saturated (Fig. S7 *a*), so we selected the ensemble at this cycle as final. N_{es} varies from 5 to 20 with the iterative cycles (Fig. S7 *b*). The final selected ensemble contains 14 conformers (Fig. S7 *c*), and its back-calculated SAXS profile is in good agreement with the experimental data (Fig. S7 *d*). Despite its different size, the ensemble shows a similar domain orientation with those with $N_{es} = 20$ (Figs. 2 *b* and S6). The latter include more conformers with similar shapes than the former. Overall, the results indicate that the ensembles generated in SAXS-ER are reliable. The two WW domains can be either close or distant to each other in solution, which may facilitate their cooperative binding with different ligands.

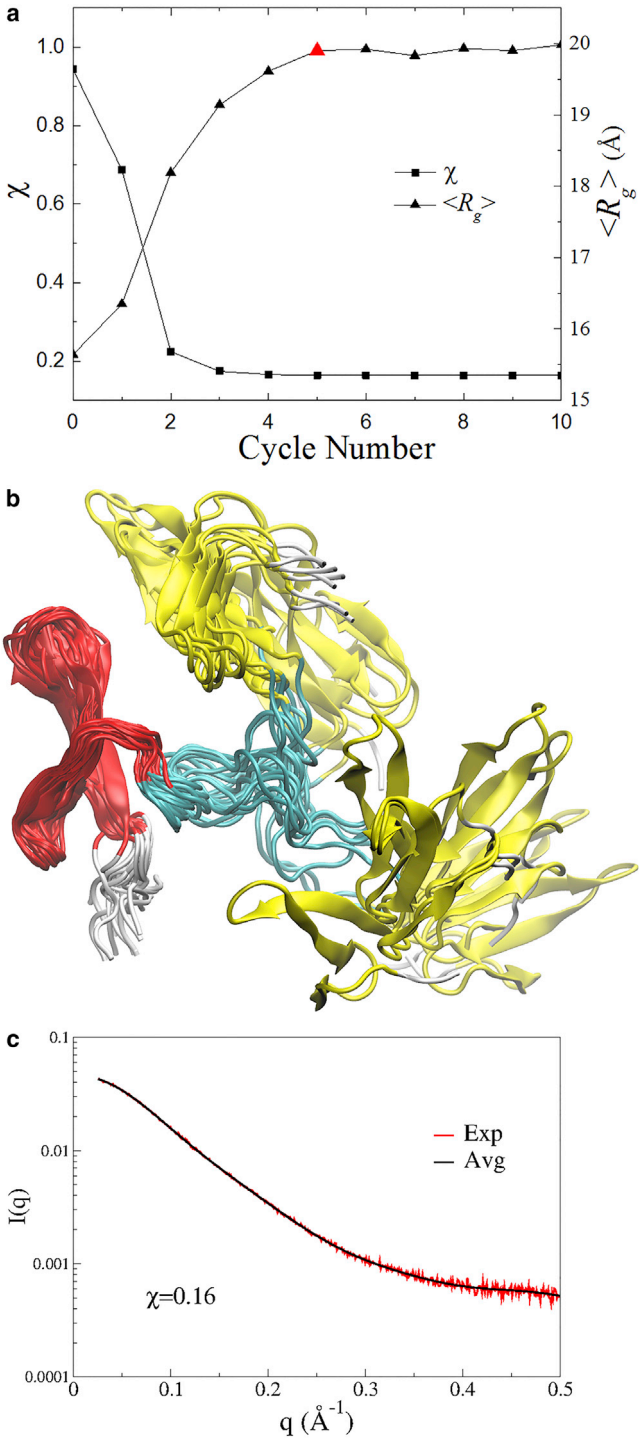

FIGURE 2 The SAXS-ER results for FBP21-WWs. (*a*) The minimal χ and the corresponding $\langle R_g \rangle$ at each cycle. The final selected ensemble at the fifth cycle is indicated by a red triangle. (*b*) The conformers in the final selected ensemble, which are superimposed on the WW1 domain (colored in *red*). The WW2 domain is colored in yellow, the linker is colored in cyan, and the N- (residues 1–5) and C-terminus (residues 74–75) are both colored in gray. (*c*) The back-calculated SAXS profile of the final selected ensemble (*black*) is fitted to the experimental data (*red*). To see this figure in color, go online.

To compare SAXS-ER with other methods, the program Ranch in the ATSAS package was used to generate 10,000 conformers of FBP21-WWs in rigid-body modeling, and then EOM and EOM 2.0 were run to select an ensemble from the pool to reproduce the experimental SAXS data (Fig. S8). We also used the flexible-meccano statistical coil model (37) to generate a pool containing 10,000 conformers, and again ran EOM and EOM 2.0 to select an ensemble (Fig. S9). These ensembles are different than those generated in SAXS-ER (Figs. 2 *b*, S6 and S7), which are clearly indicated in their projections onto the 2D plane of the PCA modes (Fig. S10). Ranch or flexible-meccano only consider relatively simple interactions, so the models generated may freely take many different orientations between the two WW domains (Figs. S8 *a* and S9 *a*). However, SAXS-ER performs all-atom simulations with a well-defined force field, and the sampled conformations would be physically more reasonable than the simple models. Many conformers in the SAXS-ER ensembles are located at a particular region on the plane (Fig. S10, *red circles*), whereas the Ranch/flexible-meccano ensembles consist of diverse conformations with a wide distribution (Fig. S10, *green* and *blue circles*). Due to the low-resolution nature of the SAXS data and the overfitting problem, the ensembles from SAXS-ER and the other two methods have nearly the same χ-values for fitting the experimental data, but the former may present more realistic protein conformations in solution than the latter. Moreover, additional quantitative information about these conformations, such as the energy and the population, could be extracted from the MD simulation.

Free SAM-1 aptamer

A SAXS-ER of the free SAM-1 aptamer was carried out with up to 30 cycles by using a 200-ns MD trajectory to estimate the aMD parameters. The ensemble size of EOM was $N_{es} = 20$. Each cycle consisted of $N_{sim} = 20$ independent 100-ps aMD simulations. The evolution of the sampled conformational space with iteration cycle is shown in projections onto the PCA modes (Fig. S11). The SAXS profiles of the ensembles are plotted against the cycle number in Fig. S12, to show how they evolve to fit the experimental data. The value χ (Fig. 3 *a*, *squares*) converges more slowly than in the SAXS-ER of FBP21-WWs (Fig. 2 *a*, *squares*), from an initial value of 7.37 to a final value of 0.86. The parameter $\langle R_g \rangle$ of the corresponding ensemble increases from the initial 22.2 Å to a final value of 24.7 Å (Fig. 3 *a*, *triangles*), which is consistent with the value (24.8 ± 0.0 Å) estimated from the experimental SAXS data. We selected the ensemble at the 23rd cycle as final because both χ and $\langle R_g \rangle$ were saturated. The 20 conformers in this ensemble were superimposed using the subdomain P4 (Fig. 3 *b*, *blue*). The χ-values of individual SAXS curves in the ensemble range from 1.07 to 2.33 (Fig. S13 *a*), but

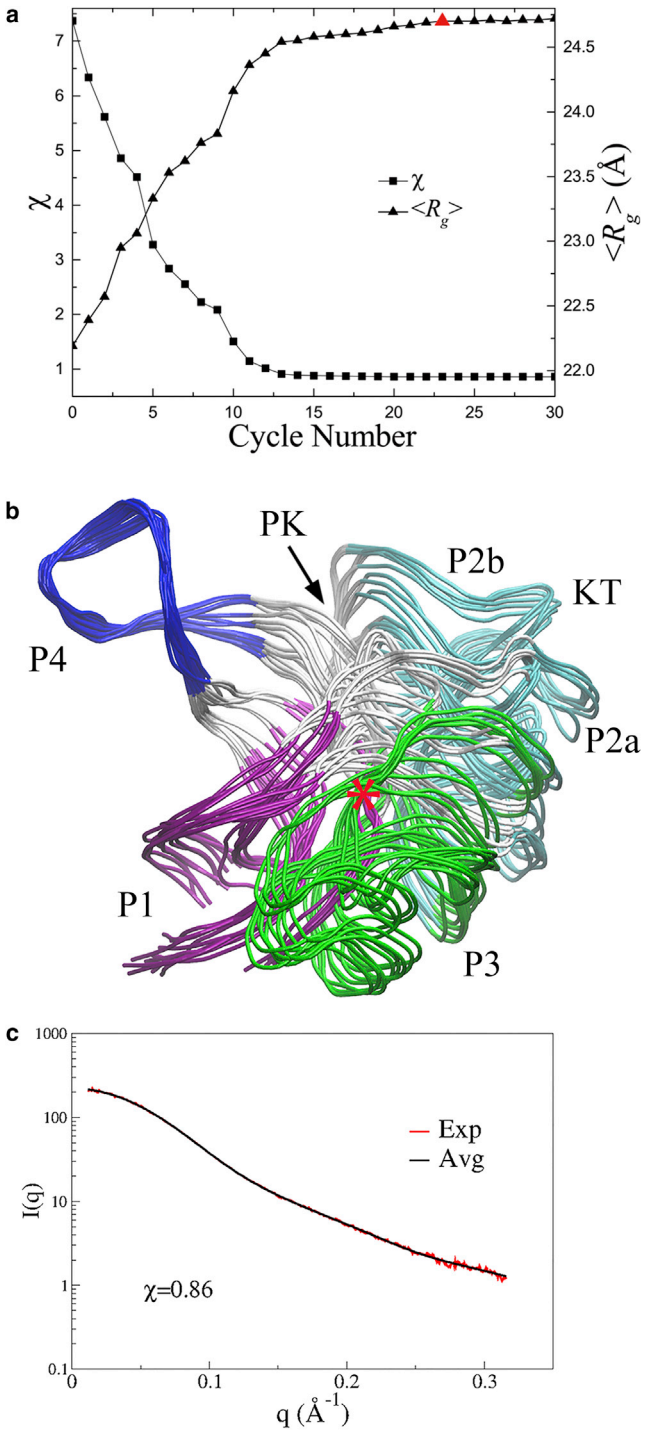

FIGURE 3 The SAXS-ER results for the free SAM-1 aptamer. (*a*) The minimal χ and the corresponding $\langle R_g \rangle$ at each cycle. The final selected ensemble at the 23rd is indicated by a red triangle. (*b*) The conformers in the final ensemble, which are superimposed on the subdomain P4 (nucleotides 69–82, colored in *blue*). The location of SAM is approximated by a red star. (*c*) The back-calculated SAXS profile of the final ensemble (*black*) is fitted to the experimental data (*red*). To see this figure in color, go online.

their average profile gives an χ of 0.86 to the experimental data (Fig. 3 *c*), which is significantly less than the value of the initial ensemble (Fig. S13 *b*). The R_g values of these conformers range from 24.1 to 25.2 Å. The result suggests that the aptamer can assume multiple conformations in solution without the ligand. Orientations of the subdomains P1 (Fig. 3 *b*, *violet*) and P3 (Fig. 3 *b*, *green*), as well as their distances, are variable, which result in the opening/closing of the ligand-binding site. Stoddard et al. (20) generated trajectories of the free aptamer using rigid-body torsion angle MD simulations, and also selected a set of conformations against the SAXS data. Although different simulation algorithms were used, our results show agreement with their ensemble.

The structural ensemble of another SAXS-ER simulation of the free SAM-1 aptamer is shown in Fig. S14. The RMSDs of all of the P atoms of the conformers in the two ensembles (Figs. 3 *b* and S14) were measured. For any conformer in one ensemble, at least one conformer in the other ensemble can be found that has a RMSD $\leq$ 3 Å, which indicates that the two ensembles are similar. A SAXS-ER using EOM 2.0 (33) was also run to optimize the ensemble size (Fig. S15). The ensemble size is five at the most cycles (Fig. S15 *b*), and the final ensemble at the 28th cycle is shown in Fig. S15 *c*. Although N_{es} is significantly smaller, the conformers are still similar to those within the larger (N_{es} = 20) ensembles (Figs. 3 *b* and S14) according to the RMSD calculations. The latter contain more conformers with similar RMSD values than the former. It should be noted that it may not be convenient to use the Ranch or flexible-meccano on the free SAM-1 aptamer because it is not as straightforward as for FBP21-WWs to determine which parts are rigid or flexible. However, this is not an issue for our method.

CONCLUSION

Among the structural modeling methods for biomolecules that integrate SAXS, our SAXS-oriented ensemble refinement tool has the following features: 1) Modification of the complicated simulation code is not required, because this tool uses a screening-after-sampling strategy. 2) Extensive simulations are not necessary to cover the conformational space of the biomolecule. By iteratively running enhanced simulations and EOM, the sampling is efficiently guided by the SAXS data, in a similar manner to refining-while-sampling methods. 3) Although computationally more expensive than other methods that use simple models, SAXS-ER is very suitable for massive parallel computing because it comprises a number of independent simulations. 4) The tool is easy to use and flexible. Any simulation package, sampling method, and ensemble selection algorithm can be chosen by simply changing the Python script.

SUPPORTING MATERIAL

Supporting Material and fifteen figures are available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(17)30238-2.

AUTHOR CONTRIBUTIONS

Z.Z. and J.P. designed the project; P.C. and J.P. wrote the software; P.C. carried out the simulations, and analyzed the data; and all authors contributed to writing the article.

ACKNOWLEDGMENTS

This work is supported by the National Key Basic Research Program of China (under grant No. 2013CB910203), the National Natural Science Foundation of China (under grants No. 31270760 and No. 21573205), the Strategic Priority Research Program of the Chinese Academy of Sciences (under grant No. XDB08030102), and the Supercomputing Center of the University of Science and Technology of China.

SUPPORTING CITATIONS

References (38–46) appear in the Supporting Material.

REFERENCES

1. Graewert, M. A., and D. I. Svergun. 2013. Impact and progress in small and wide angle x-ray scattering (SAXS and WAXS). *Curr. Opin. Struct. Biol.* 23:748–754.
2. Bernadó, P., and M. Blackledge. 2010. Structural biology: proteins in dynamic equilibrium. *Nature.* 468:1046–1048.
3. Schneidman-Duhovny, D., S. J. Kim, and A. Sali. 2012. Integrative structural modeling with small angle x-ray scattering profiles. *BMC Struct. Biol.* 12:17.
4. Zhang, Y. H., J. H. Peng, and Z. Y. Zhang. 2015. Structural modeling of proteins by integrating small-angle x-ray scattering data. *Chin. Phys. B.* 24:126101.
5. Zheng, W., and M. Tekpinar. 2011. Accurate flexible fitting of high-resolution protein structures to small-angle x-ray scattering data using a coarse-grained model with implicit hydration shell. *Biophys. J.* 101:2981–2991.
6. Björling, A., S. Niebling, …, S. Westenhoff. 2015. Deciphering solution scattering data with experimentally guided molecular dynamics simulations. *J. Chem. Theory Comput.* 11:780–787.
7. Bernadó, P., E. Mylonas, …, D. I. Svergun. 2007. Structural characterization of flexible proteins using small-angle x-ray scattering. *J. Am. Chem. Soc.* 129:5656–5664.
8. Pelikan, M., G. L. Hura, and M. Hammel. 2009. Structure and flexibility within proteins as identified through small angle x-ray scattering. *Gen. Physiol. Biophys.* 28:174–189.
9. Bertini, I., A. Giachetti, …, D. I. Svergun. 2010. Conformational space of flexible biological macromolecules from average data. *J. Am. Chem. Soc.* 132:13553–13558.
10. Yang, S., L. Blachowicz, …, B. Roux. 2010. Multidomain assembled states of Hck tyrosine kinase in solution. *Proc. Natl. Acad. Sci. USA.* 107:15757–15762.
11. Różycki, B., Y. C. Kim, and G. Hummer. 2011. SAXS ensemble refinement of ESCRT-III CHMP3 conformational transitions. *Structure.* 19:109–116.
12. Curtis, J. E., S. Raghunandan, …, S. Krueger. 2012. SASSIE: a program to study intrinsically disordered biological molecules and macromolecular ensembles using experimental scattering restraints. *Comput. Phys. Commun.* 183:382–389.
13. Deshmukh, L., C. D. Schwieters, …, G. M. Clore. 2013. Structure and dynamics of full-length HIV-1 capsid protein in solution. *J. Am. Chem. Soc.* 135:16133–16147.
14. Huang, J. R., L. R. Warner, …, M. Blackledge. 2014. Transient electrostatic interactions dominate the conformational equilibrium sampled by multidomain splicing factor U2AF65: a combined NMR and SAXS study. *J. Am. Chem. Soc.* 136:7068–7076.
15. Bedford, M. T., R. Reed, and P. Leder. 1998. WW domain-mediated interactions reveal a spliceosome-associated protein that binds a third class of proline-rich motif: the proline glycine and methionine-rich motif. *Proc. Natl. Acad. Sci. USA.* 95:10602–10607.
16. Huang, X., M. Beullens, …, Y. Shi. 2009. Structure and function of the two tandem WW domains of the pre-mRNA splicing factor FBP21 (formin-binding protein 21). *J. Biol. Chem.* 284:25375–25387.
17. Wen, B., J. Peng, …, Z. Zhang. 2014. Characterization of protein flexibility using small-angle x-ray scattering and amplified collective motion simulations. *Biophys. J.* 107:956–964.
18. Grundy, F. J., and T. M. Henkin. 1998. The S box regulon: a new global transcription termination control system for methionine and cysteine biosynthesis genes in gram-positive bacteria. *Mol. Microbiol.* 30:737–749.
19. Montange, R. K., and R. T. Batey. 2006. Structure of the *S*-adenosylmethionine riboswitch regulatory mRNA element. *Nature.* 441:1172–1175.
20. Stoddard, C. D., R. K. Montange, …, R. T. Batey. 2010. Free state conformational sampling of the SAM-I riboswitch aptamer domain. *Structure.* 18:787–797.
21. Karplus, M., and J. A. McCammon. 2002. Molecular dynamics simulations of biomolecules. *Nat. Struct. Biol.* 9:646–652.
22. Clarage, J. B., T. Romo, …, G. N. Phillips, Jr. 1995. A sampling problem in molecular dynamics simulations of macromolecules. *Proc. Natl. Acad. Sci. USA.* 92:3288–3292.
23. Onuchic, J. N., Z. Luthey-Schulten, and P. G. Wolynes. 1997. Theory of protein folding: the energy landscape perspective. *Annu. Rev. Phys. Chem.* 48:545–600.
24. Mitsutake, A., Y. Mori, and Y. Okamoto. 2013. Enhanced sampling algorithms. *Methods Mol. Biol.* 924:153–195.
25. Zhang, Z., Y. Shi, and H. Liu. 2003. Molecular dynamics simulations of peptides and proteins with amplified collective motions. *Biophys. J.* 84:3583–3593.
26. Atilgan, A. R., S. R. Durell, …, I. Bahar. 2001. Anisotropy of fluctuation dynamics of proteins with an elastic network model. *Biophys. J.* 80:505–515.
27. Hamelberg, D., J. Mongan, and J. A. McCammon. 2004. Accelerated molecular dynamics: a promising and efficient simulation method for biomolecules. *J. Chem. Phys.* 120:11919–11929.
28. Hess, B., C. Kutzner, …, E. Lindahl. 2008. GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* 4:435–447.
29. Case, D. A., V. Babin, …, P. A. Kollman. 2014. AMBER 14. University of California, San Francisco, CA.
30. Amadei, A., A. B. Linssen, and H. J. C. Berendsen. 1993. Essential dynamics of proteins. *Proteins.* 17:412–425.
31. Petoukhov, M. V., D. Franke, …, D. I. Svergun. 2012. New developments in the ATSAS program package for small-angle scattering data analysis. *J. Appl. Cryst.* 45:342–350.
32. Svergun, D., C. Barberato, and M. H. J. Koch. 1995. CRYSOL—a program to evaluate x-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Cryst.* 28:768–773.
33. Tria, G., H. D. T. Mertens, …, D. I. Svergun. 2015. Advanced ensemble modelling of flexible macromolecules using x-ray solution scattering. *IUCrJ.* 2:207–217.
34. Caves, L. S. D., J. D. Evanseck, and M. Karplus. 1998. Locally accessible conformations of proteins: multiple molecular dynamics simulations of crambin. *Protein Sci.* 7:649–666.
35. Humphrey, W., A. Dalke, and K. Schulten. 1996. VMD: visual molecular dynamics. *J. Mol. Graph.* 14:33–38, 27–28.
36. Zhang, Y., B. Wen, …, Z. Zhang. 2015. Determining structural ensembles of flexible multi-domain proteins using small-angle

x-ray scattering and molecular dynamics simulations. *Protein Cell.* 6:619–623.

37. Ozenne, V., F. Bauer, …, M. Blackledge. 2012. Flexible-meccano: a tool for the generation of explicit ensemble descriptions of intrinsically disordered proteins and their associated experimental observables. *Bioinformatics.* 28:1463–1470.

38. Hockney, R. W., S. P. Goel, and J. W. Eastwood. 1974. Quiet high-resolution computer models of a plasma. *J. Comput. Phys.* 14:148–158.

39. Maier, J. A., C. Martinez, …, C. Simmerling. 2015. ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* 11:3696–3713.

40. Ryckaert, J. P., G. Ciccotti, and H. J. C. Berendsen. 1977. Numerical integration of Cartesian equations of motion of a system with constraints—molecular dynamics of *n*-alkanes. *J. Comput. Phys.* 23: 327–341.

41. Jorgensen, W. L., J. Chandrasekhar, …, M. L. Klein. 1983. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79:926–935.

42. Berendsen, H. J. C., J. P. M. Postma, …, J. R. Haak. 1984. Molecular-dynamics with coupling to an external bath. *J. Chem. Phys.* 81:3684–3690.

43. Essmann, U., L. Perera, …, L. G. Pedersen. 1995. A smooth particle mesh Ewald method. *J. Chem. Phys.* 103:8577–8593.

44. Duan, Y., C. Wu, …, P. Kollman. 2003. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.* 24:1999–2012.

45. Bussi, G., D. Donadio, and M. Parrinello. 2007. Canonical sampling through velocity rescaling. *J. Chem. Phys.* 126:014101.

46. Hess, B. 2008. P-LINCS: a parallel linear constraint solver for molecular simulation. *J. Chem. Theory Comput.* 4:116–122.

Biophysical Journal, Volume *112*

Supplemental Information

SAXS-Oriented Ensemble Refinement of Flexible Biomolecules

Peng Cheng, Junhui Peng, and Zhiyong Zhang

SUPPORTING METHODS

ACM Simulations of FBP21-WWs

The setup procedure for ACM was the same as for the standard MD simulation. Starting from any initial structure of FBP21-WWs, the simulated system was set up with the GROMACS-4.5.5 package (1) and the AMBER03 force field (2). A rhombic dodecahedron box filled with TIP3P waters (3), was used, with a minimum distance between the solute and the box boundary of 1.4 nm. The energy of the system (protein and water) was minimized by the steepest-descent method, until the maximum force on the atoms was <800 kJ mol^{-1} nm^{-1}. Replacing the water molecules at the positions with the most favorable electrostatic potential added in 62 Na^+ and 55 Cl^- to compensate for the net negative charge of the protein and to mimic the salt concentration (300 mM) of the SAXS sample. The final system (protein, water, and ions) was minimized again using the steepest descent followed in the conjugate-gradient method, until the maximum force on the atoms was <50 kJ mol^{-1} nm^{-1}. The simulation was conducted using the leap-frog algorithm (4) with a time step of 2 fs. The initial atomic velocities were generated according to a Maxwell distribution at 310 K, and an equilibration simulation with positional restraints (using a force constant of 1000 kJ mol^{-1} nm^{-2}) was carried out for 100 ps. The production simulation was performed under the constant NPT condition. Each of the three groups (protein, solvent, and ions) was coupled to a thermostat at 310 K using the velocity-rescaling algorithm (5) with a relaxation time of 0.1 ps. The pressure was coupled to 1 bar with a relaxation time of 0.5 ps and a compressibility of 4.5×10^{-5} bar^{-1}. All the bonds in the protein were constrained using the P-LINCS algorithm (6). Twin range cutoff distances for van der Waals interactions were set to 0.9 and 1.4 nm, and the neighbor list was updated every 20 fs. The long-range electrostatic interactions were calculated in the particle mesh Ewald summation (PME) algorithm (7), with an interpolation order of 4 and a tolerance of 10^{-5}.

ACM sampling was begun after the equilibration simulation. Many parameters were the same as those in the standard MD simulation, except that collective motion described in ENM (8) was amplified by coupling them to a high-temperature bath. An ENM was built with CG sites located at the center-of-mass (COM) of residues from an all-atom structure of the protein in the simulation. The potential energy function took the harmonic form:

$$V = \sum_{i,j>i} \frac{1}{2} k_{ij} \Delta r_{ij}^2 \ . \qquad (1)$$

Where Δr_{ij} is the fluctuation of the psuedo bond connecting residues *i* and *j*, with their COM distance r_{ij}. k_{ij} is the spring constant given as:

$$k_{ij} = \begin{cases} 1.0 & r_{ij} \le 0.7\,nm \\ 10^{-2} & 0.7 < r_{ij} \le 1.1\,nm \\ 5\times10^{-4} & 1.1 < r_{ij} \le 1.5\,nm \\ 0 & r_{ij} > 1.5\,nm \end{cases} . \qquad (2)$$

The four-range spring constants described the interactions in the protein from strong to weak. The short cutoff distance, 0.7 nm, defined the first coordination shell, and the long cutoff distance, 1.5 nm, was chosen to avoid unrealistic large-amplitude fluctuations in some residues in particular directions (8). A middle cutoff value of 1.1 nm was set between the short and long cutoff distances. A Hessian matrix of the second derivatives of the overall potential was constructed and then diagonalized to yield a matrix of eigenvectors and corresponding eigenvalues. Each eigenvector with a nonzero eigenvalue is called a normal mode, and the corresponding eigenvalue is proportional to the squared frequency of the motion along the mode. Usually only a few modes with the lowest frequencies are predominate in collective motion of the protein. For FBP21-WWs, we defined an essential subspace using the three slowest modes. At each time step, the velocity of each atom was divided into two components, one projected onto the essential subspace and the remainder. By modifying the weak coupling method (9), the velocity component in the essential subspace was coupled to a high temperature (we tried different values ranging from 500 K to 700 K), whereas the rest of velocity was coupled normally to 310 K. The updated velocity was thus a combination of these two components. During the ACM simulation, the collective modes were updated on the fly by doing ENM calculations every 50 time steps according to the newly generated protein conformation.

In the SAXS-ER of FBP211-WWs, the preliminary ACM simulation was run for 100 ps, and the independent simulations in each cycle were either 100 or 200 ps.

aMD simulations of the free SAM-1 aptamer

The simulations were performed using the AMBER14 package (10). The initial structure was taken from the crystal structure of the bound SAM-1 aptamer (PDB entry 2GIS) (11), with the coordinates of the RNA (94 nucleotides), two Mg^{2+}, and crystal waters retained. The simulated system was built in the tleap module using the ff14SB force field (12). The structure was solvated in a truncated octahedral box that extended 20 Å from the solute surface, using the TIP3P water model (3). Three more Mg^{2+}, 98 Na^{+}, and 19 Cl^{-} were added in the box to neutralize the system and also to mimic the salt concentration (7.6 mM $MgCl_2$ and 150 mM NaCl) in the SAXS sample. Therefore, the total number of atoms was 99510. The waters and ions were initially minimized for 1000 steps using the steepest descent method for the first 500 steps and then the conjugate gradient algorithm for the last 500 steps, with the position of RNA fixed (force constant was 500 kcal mol^{-1} $Å^{-2}$). The second energy minimization of the entire system was conducted for 2500 steps, using the steepest descent method in the first 1000 steps and then the conjugate gradient algorithm for the last 1500 steps. After that, a heat-up MD was run at a constant volume. The system was heated from 0 to 300 K for 100 ps with a weak restraint of 10 kcal mol^{-1} $Å^{-2}$ on the solute. Then, free MD simulations were carried out under the NPT condition. Langevin dynamics were used to control the temperature with a collision frequency of 1.0 ps^{-1}. Isotropic position scaling was used to maintain the pressure at 1 bar with a relaxation time of 1.0 ps. All of the bonds involving hydrogen atoms were constrained using the SHAKE algorithm (13), and the time step was set to 2 fs. The long-range electrostatic interactions were calculated in PME (7) with a 10 Å cutoff for the range-limited non-bonded interactions.

aMD introduces a boost potential, $\Delta V(r)$ to the original potential energy $V(r)$ when the latter is below a threshold energy E,

$$\Delta V(r) = \begin{cases} 0 & V(r) \geq E \\ \dfrac{\left(E - V(r)\right)^2}{\alpha + \left(E - V(r)\right)} & V(r) < E \end{cases}. \qquad (3)$$

Where α is a factor that tunes the depth of the modified energy basins. Boosting potentials were applied to both the total potential and the individual dihedral energy term. A standard MD simulation with a total of 200 ns was performed, and we used different trajectory lengths to estimate the aMD parameters. For example, for the 200-ns MD trajectory, the average total potential energy was -342270 kcal mol^{-1} and the average dihedral energy was 2320 kcal mol^{-1}. The free SAM-1 aptamer had 94 nucleotides and the simulated system consisted of 99510 atoms. The following parameters were set based on the above information:
E(tot) = -342270 kcal mol^{-1} + (0.2 kcal mol^{-1} $atom^{-1}$ × 99510 atoms) = -322368 kcal mol^{-1}
α(tot) = (0.2 kcal mol^{-1} $atom^{-1}$ × 99510 atoms) = 19902 kcal mol^{-1}
E(dih) = 2320 kcal mol^{-1} + (3.5 kcal mol^{-1} $residue^{-1}$ × 94 residues) = 2649 kcal mol^{-1}
α(dih) = 0.2 × (3.5 kcal mol^{-1} $residues^{-1}$ × 94 residues) = 66 kcal mol^{-1}
The other aMD parameters were the same as those in the standard MD simulation.

In the SAXS-ER of the free SAM-1 aptamer, the preliminary aMD simulation was run for 100 ps, and all of the independent simulations at each cycle were also 100 ps.

Principal component analysis

PCA on a simulated trajectory, also called essential dynamics analysis (14), allows one to extract global collective motions of the biomolecule from local fluctuations. PCA consists of the following steps. (1) One needs to choose which subset of atoms of the biomolecule are used for analysis, such as C_α atoms in the protein. (2) All the conformations in the trajectory are superimposed on a reference structure to eliminate overall translational and rotational motions of the system. (3) With the selected subset of N atoms, a covariance matrix of positional fluctuation is constructed. (4) The covariance matrix is diagonalized to yield $3N$-6 eigenvectors (PCA modes) with non-zero eigenvalues (mean square fluctuations of the modes). Generally, only a small number of the PCA modes with the largest eigenvalues (termed as essential modes) represent collective motions of the biomolecule.

For the trajectories of FBP21-WWs generated by the GROMACS package, PCA were carried out using the programs g_covar and g_anaeig sequentially. For the trajectories of the free SAM-1 aptamer generated by the AMBER package, PCA were performed using CPPTRAJ.

SUPPORTING REFERENCES

1. Hess, B., C. Kutzner, D. van der Spoel, and E. Lindahl. 2008. GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. J. Chem. Theory Comput. 4:435-447.
2. Duan, Y., C. Wu, S. Chowdhury, M. C. Lee, G. M. Xiong, W. Zhang, R. Yang, P. Cieplak,

R. Luo, T. Lee, J. Caldwell, J. M. Wang, and P. Kollman. 2003. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. J. Comput. Chem. 24:1999-2012.

3. Jorgensen, W. L., J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein. 1983. Comparison of Simple Potential Functions for Simulating Liquid Water. J. Chem. Phys. 79:926-935.
4. Hockney, R. W., S. P. Goel, and J. W. Eastwood. 1974. Quiet High-Resolution Computer Models of a Plasma. J. Comput. Phys. 14:148-158.
5. Bussi, G., D. Donadio, and M. Parrinello. 2007. Canonical sampling through velocity rescaling. J. Chem. Phys. 126.
6. Hess, B. 2008. P-LINCS: A parallel linear constraint solver for molecular simulation. J. Chem. Theory Comput. 4:116-122.
7. Essmann, U., L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen. 1995. A Smooth Particle Mesh Ewald Method. J. Chem. Phys. 103:8577-8593.
8. Atilgan, A. R., S. R. Durell, R. L. Jernigan, M. C. Demirel, O. Keskin, and I. Bahar. 2001. Anisotropy of fluctuation dynamics of proteins with an elastic network model. Biophys. J. 80:505-515.
9. Berendsen, H. J. C., J. P. M. Postma, W. F. Vangunsteren, A. Dinola, and J. R. Haak. 1984. Molecular-Dynamics with Coupling to an External Bath. J. Chem. Phys. 81:3684-3690.
10. Case, D. A., V. Babin, J. T. Berryman, R. M. Betz, Q. Cai, D. S. Cerutti, I. Cheatham, T.E., T. A. Darden, R. E. Duke, H. Gohlke, A. W. Goetz, S. Gusarov, N. Homeyer, P. Janowski, J. Kaus, I. Kolossváry, A. Kovalenko, T. S. Lee, S. LeGrand, T. Luchko, R. Luo, B. Madej, K. M. Merz, F. Paesani, D. R. Roe, A. Roitberg, C. Sagui, R. Salomon-Ferrer, G. Seabra, C. L. Simmerling, W. Smith, J. Swails, R. C. Walker, J. Wang, R. M. Wolf, X. Wu, and P. A. Kollman. 2014. AMBER 14, University of California, San Francisco.
11. Montange, R. K., and R. T. Batey. 2006. Structure of the S-adenosylmethionine riboswitch regulatory mRNA element. Nature 441:1172-1175.
12. Maier, J. A., C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser, and C. Simmerling. 2015. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. J. Chem. Theory Comput. 11:3696-3713.
13. Ryckaert, J. P., G. Ciccotti, and H. J. C. Berendsen. 1977. Numerical-Integration of Cartesian Equations of Motion of a System with Constraints - Molecular-Dynamics of N-Alkanes. J. Comput. Phys. 23:327-341.
14. Amadei, A., A. B. M. Linnsen, and H. J. C. Berendsen. 1993. Essential dynamics of proteins. Proteins 17:412-425.

SUPPORTING FIGURES

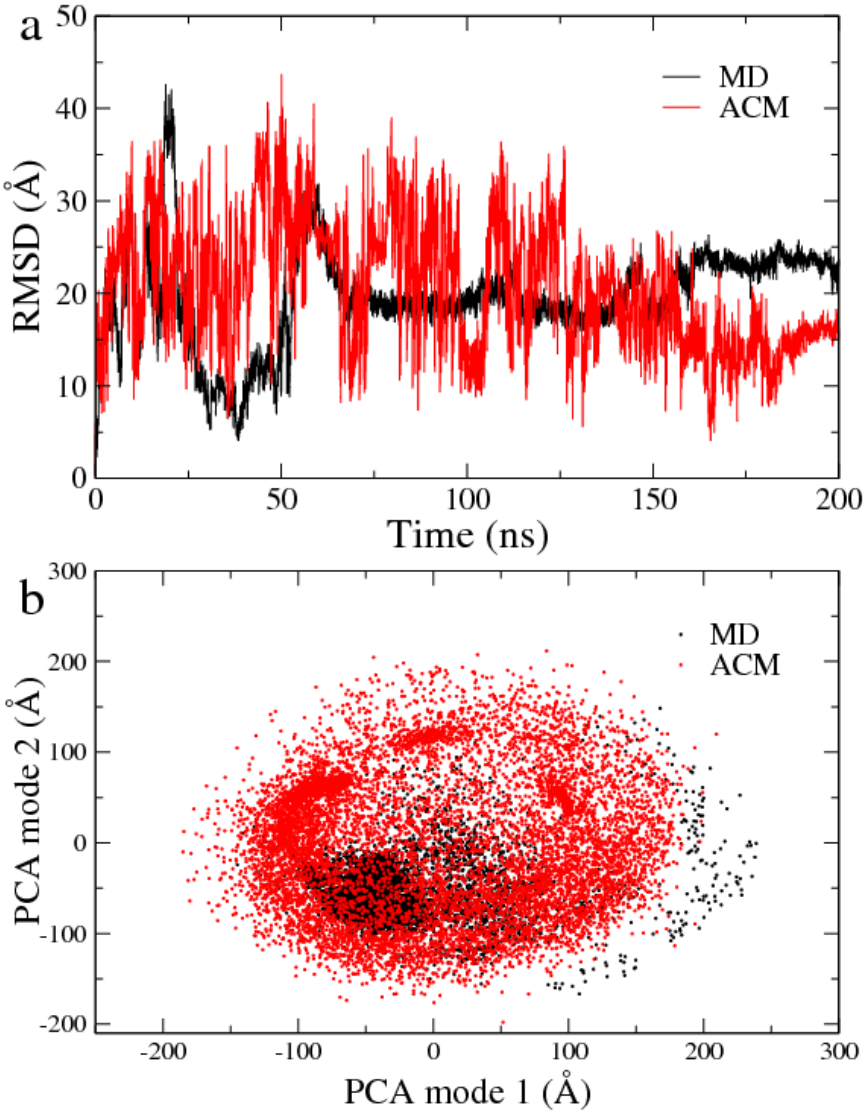

Figure S1. Sampling efficiency of ACM compared to the standard MD. From the same initial structure of FBP21-WWs, both ACM and MD were carried out for 200 ns. (a) Time evolution of the root mean square deviation (RMSD) during the ACM (red) and MD (black) simulations. (b) PCA on the ACM trajectory. The ACM (red) and MD (black) simulations are projected onto the plane defined by the first and the second PCA modes. In both the RMSD calculation and PCA, the 54 C_α atoms in the two WW domains were used. All of the frames were superimposed on the initial structure using the 27 C_α atoms in the WW1 domain. Therefore, both RMSD and PCA results describe relative domain motions in the protein.

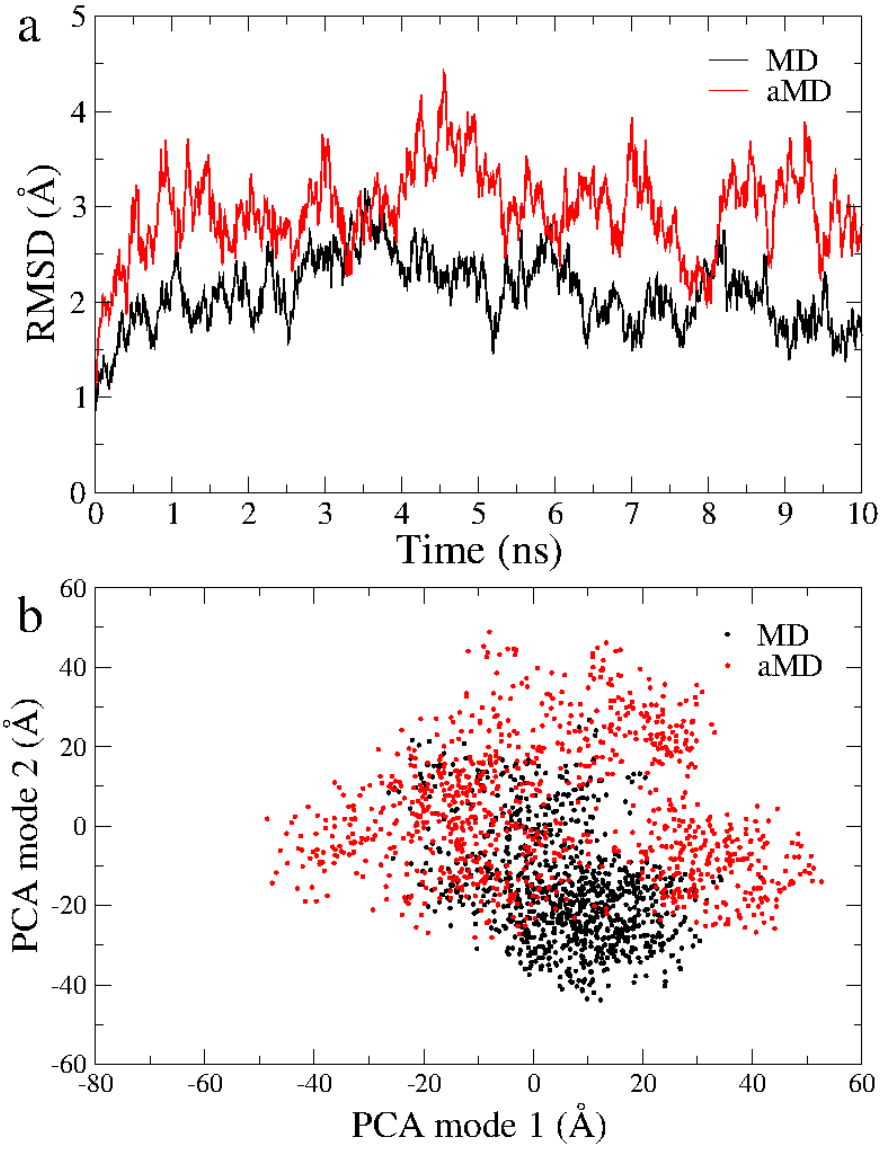

Figure S2. Sampling efficiency of aMD compared to the standard MD. From the crystal structure 2GIS with the ligand removed, both aMD and MD were carried out for 10 ns. (a) Time evolution of the RMSD during the aMD (red) and MD (black) simulations. (b) Projections of the aMD (red) and MD (black) simulations onto the first and the second PCA modes calculated from the aMD trajectory. In both the RMSD calculation and PCA, all the P, O3', O5', C3', C4', C5' atoms in the RNA were used. All of the frames were superimposed on the initial structure using the same atoms.

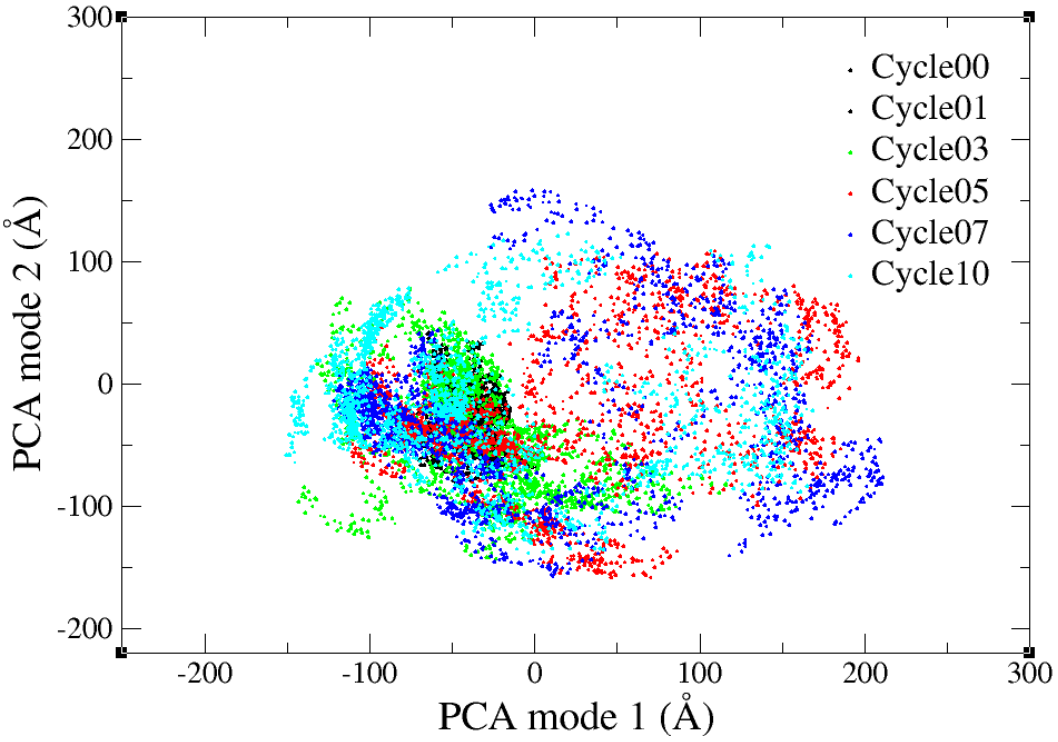

Figure S3. The evolution of the sampled conformational space of FBP21-WWs against the SAXS-ER iteration cycles are shown in projections onto the PCA modes (defined in Figure S1b).

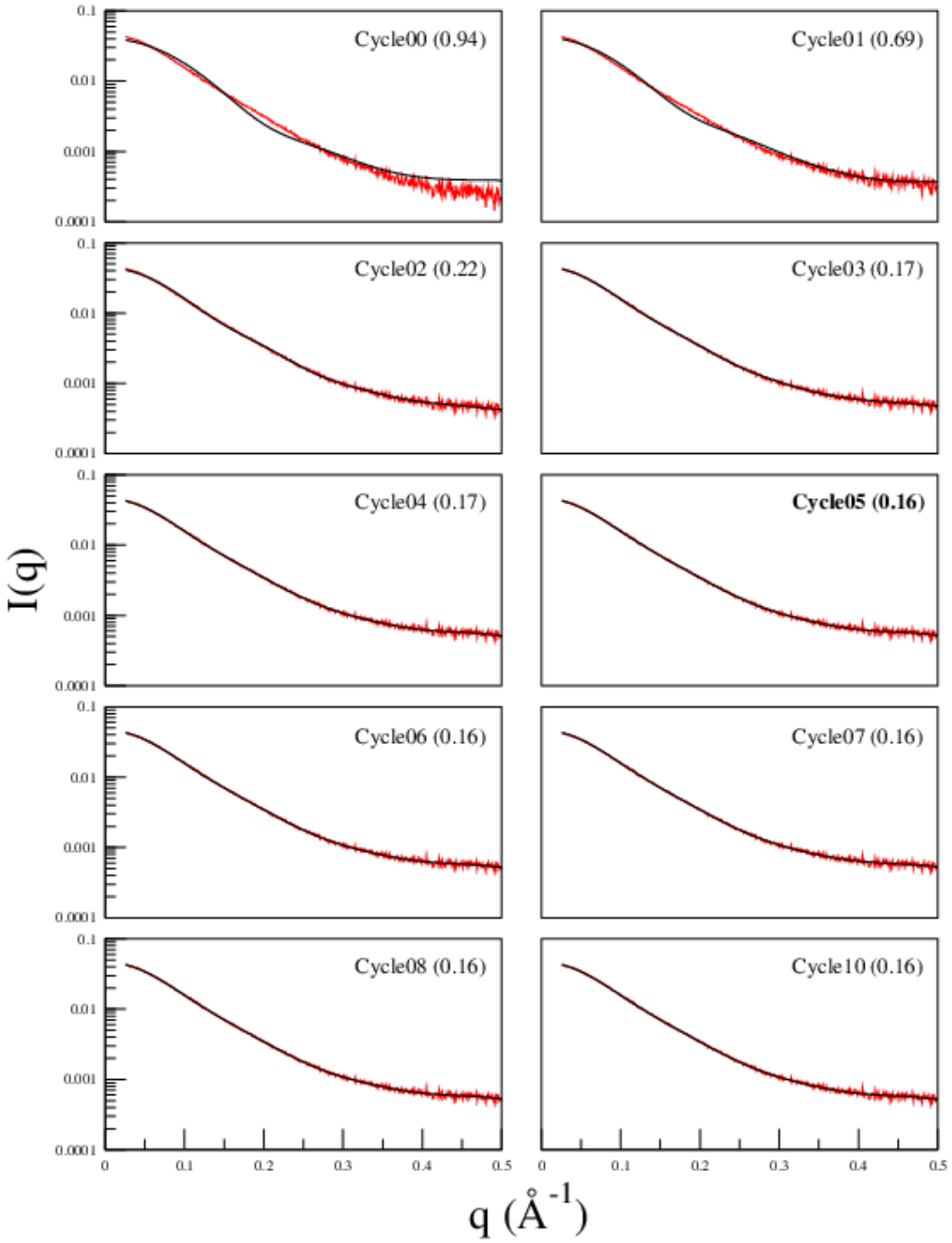

Figure S4. SAXS profiles (black) of the FBP21-WWs ensembles of from the initial to the final SAXS-ER cycles to show how they fit to the experimental SAXS data (red). In each cycle, the χ value between the theoretical and the experimental SAXS profile is given in parenthesis. The final selected ensemble is at the 5^{th} cycle (bold font).

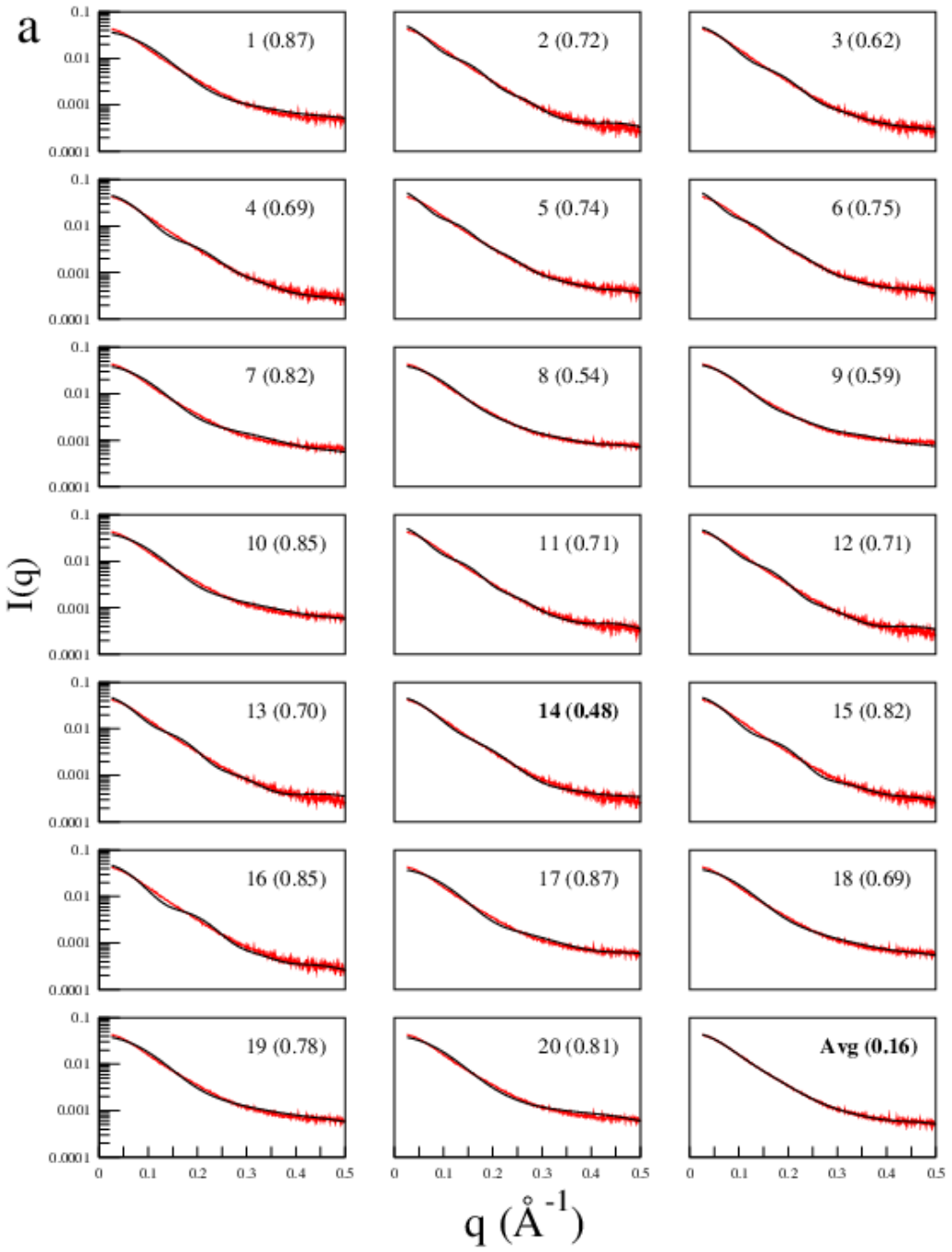

Figure S5a. SAXS curves of individual conformers (black) in the final ensemble of FBP21-WWs (Fig. 2b) and the average profile of the ensemble (bold font), which are all fitted to the experimental data (red). In each panel, the χ value between the theoretical and the experimental SAXS profiles is given in parenthesis. The single conformation with the best fitting SAXS curve (the smallest χ) is also highlighted in bold font.

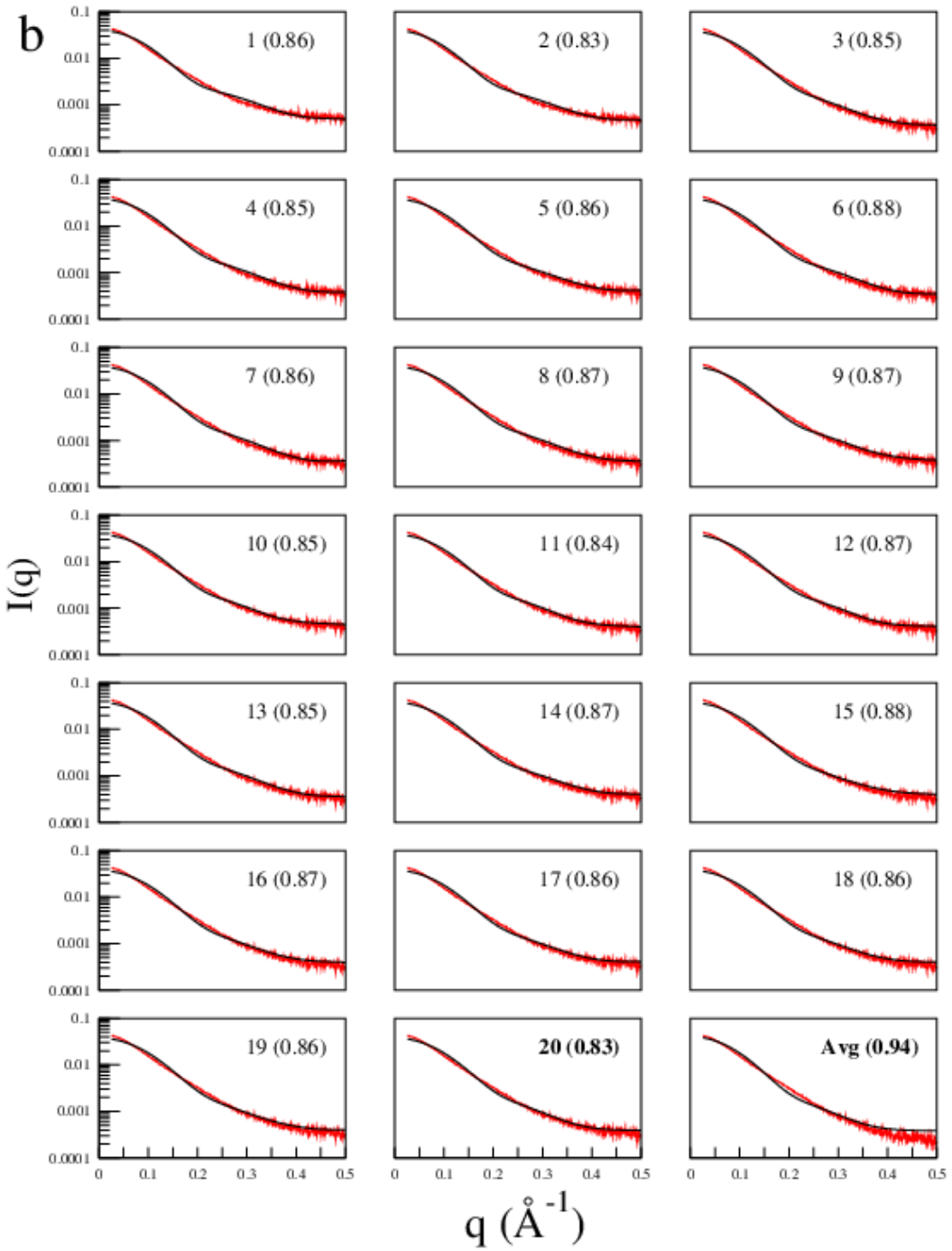

Figure S5b. SAXS curves for the individual conformers (black) in the initial ensemble (Cycle 0) of FBP21-WWs and the average profile of the ensemble (bold font), which are all fitted to the experimental data (red). In each panel, the χ value between the theoretical and the experimental SAXS profile is given in parenthesis. The single conformation with the best fitting SAXS curve (the smallest χ) is also highlighted in bold font.

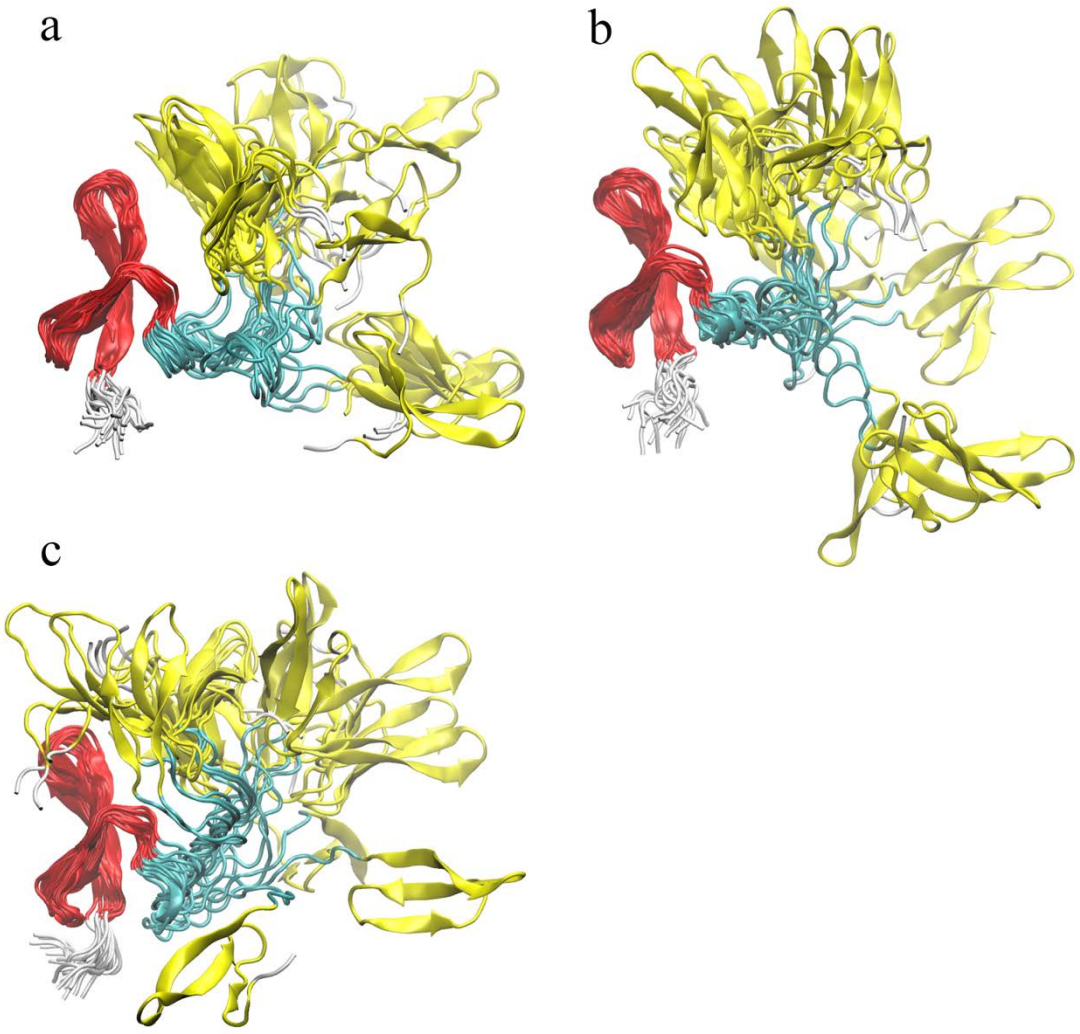

Figure S6. Structural ensembles from other SAXS-ER simulations of FBP21-WWs. (a) The starting conformation was the model 1 of the NMR structures. The ensemble size of EOM was N_{es}=20. Each cycle consisted of N_{sim}=20 independent 200-ps ACM simulations, in which those low-frequency collective motions were coupled at 500 K. (b) The starting conformation was the model 1 of the NMR structures. The ensemble size of EOM was N_{es}=20. Each cycle consisted of N_{sim}=20 independent 100-ps ACM simulations, in which those low-frequency collective motions were coupled at 580 K. (c) The starting conformation was an extended one. The ensemble size of EOM was N_{es}=20. Each cycle consisted of N_{sim}=20 independent 100-ps ACM simulations, in which those low-frequency collective motions were coupled at 650 K. All the structures are superimposed on the WW1 domain, and the coloring is the same as that in Figure 2b.

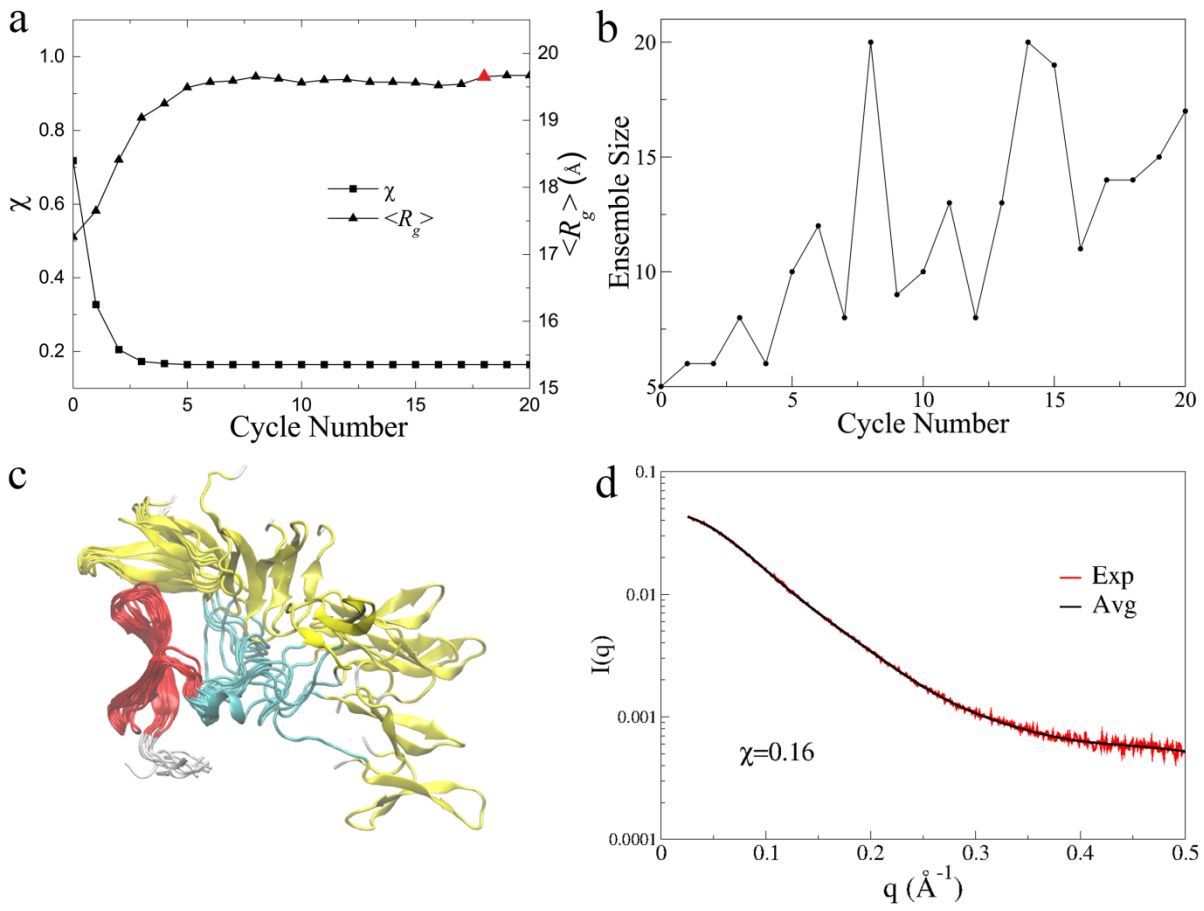

Figure S7. SAXS-ER of FBP21-WWs with EOM 2.0. The starting conformation was the model 1 of the NMR structures. The ensemble size N_{es} at each cycle was optimized in EOM 2.0. Each cycle consisted of N_{sim}=20 independent 100-ps ACM simulations starting from the N_{es} conformers (some of them were used more than once), in which the low-frequency collective motions were coupled at 500 K. (a) The minimal χ and the corresponding $<R_g>$ at each cycle. The final ensemble at the 18th cycle is indicated by a red triangle. (b) The ensemble size at each cycle. (c) The conformers in the final ensemble, which are superimposed on the WW1 domain. The coloring is the same as that in Figure 2b. (d) The back-calculated SAXS profile of the final ensemble (black) is fitted to the experimental data (red).

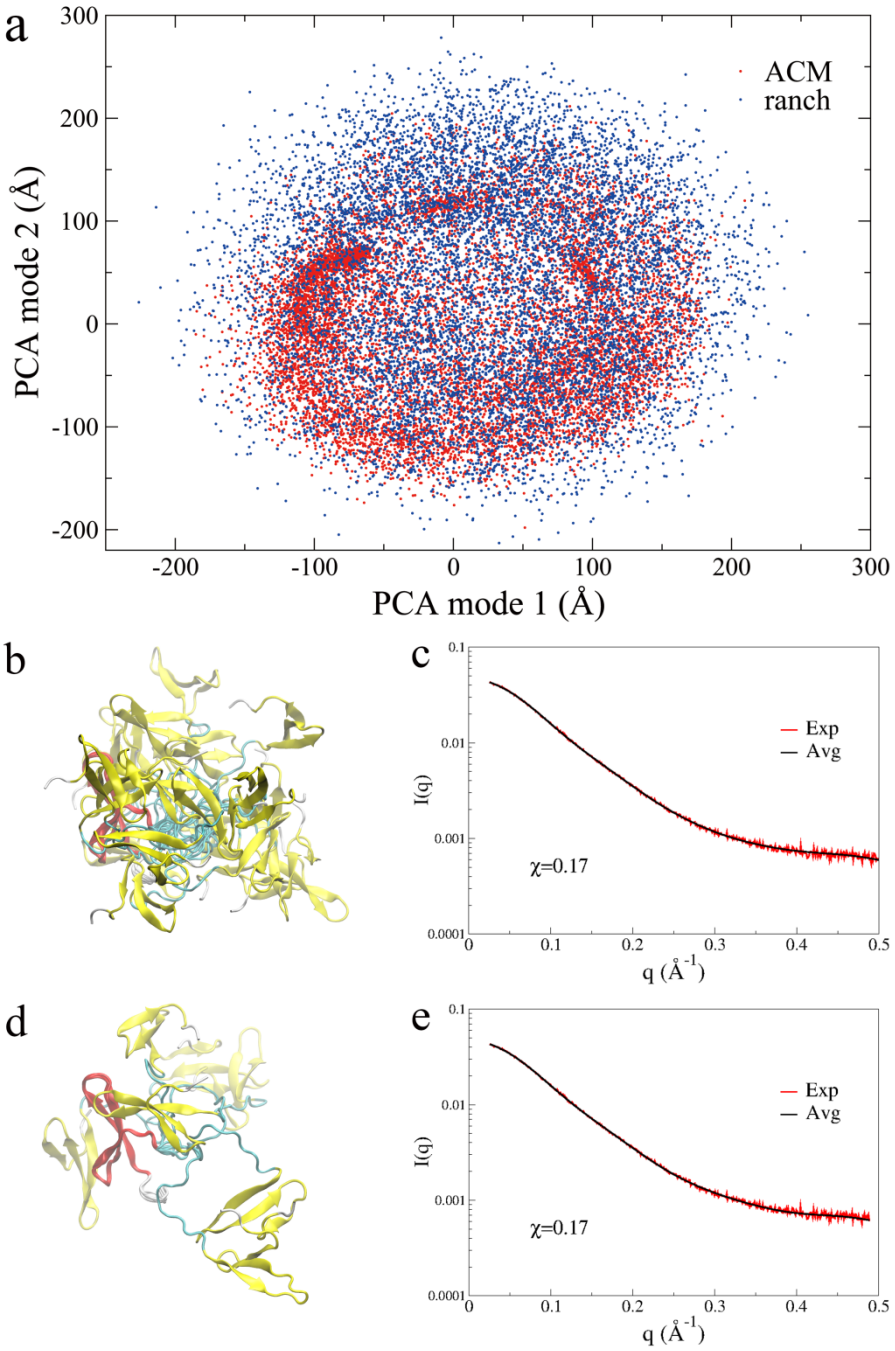

Figure S8. EOM on a structural pool containing 10,000 conformers of FBP21-WWs generated by ranch. (a) Projections of all the conformers (blue) onto the PCA modes defined in Figure S1b. For comparison, projections of the ACM simulation (red) are also shown. (b) The 20 conformers in the ensemble selected by EOM, and (c) the back-calculated SAXS profile of the ensemble (black) is fitted to the experimental data (red). (d) The seven conformers in the ensemble selected by EOM 2.0, and (e) the back-calculated SAXS profile of the ensemble is fitted to the experimental data.

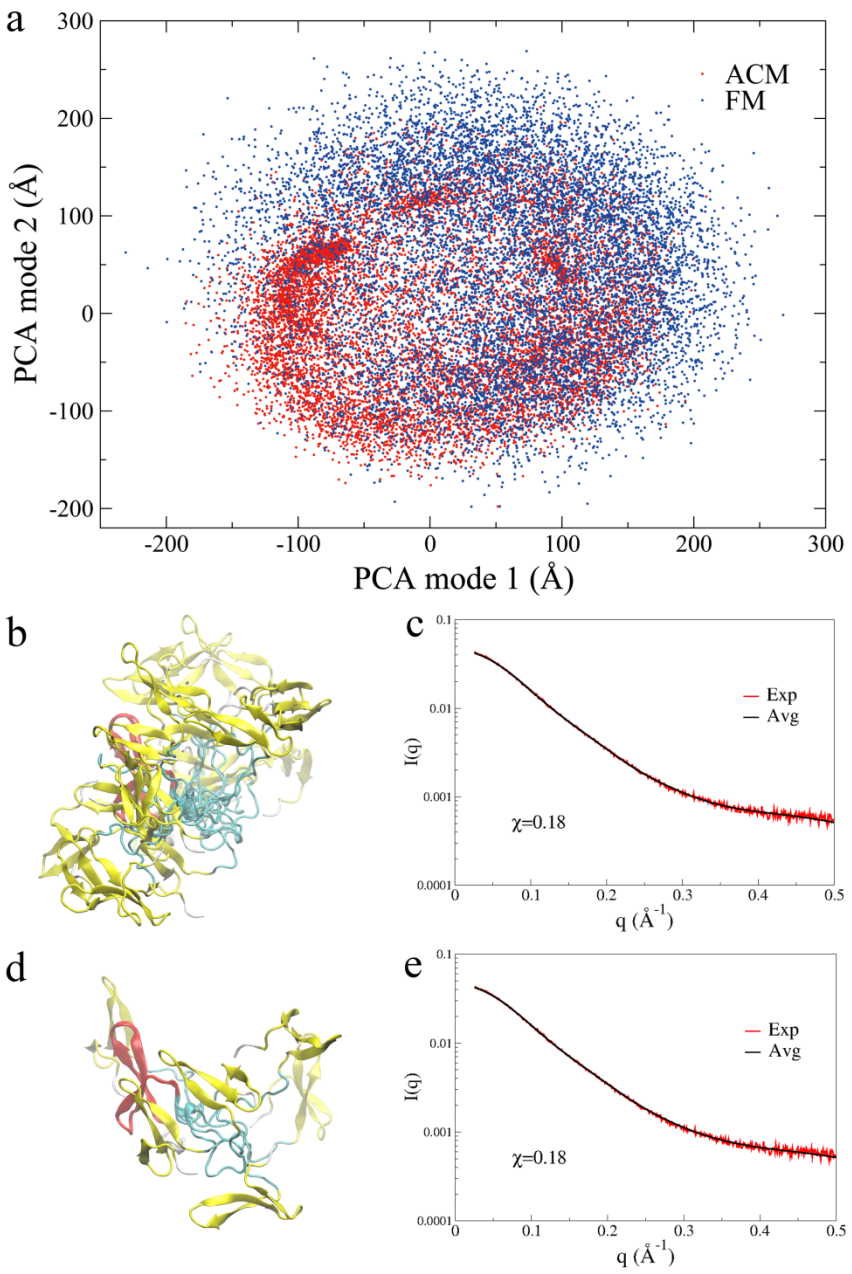

Figure S9. EOM on a structural pool of FBP21-WWs generated in the f*lexible-meccano* statistical coil model. The program produced a large number of linker conformations, and then the WW1 and WW2 domains were "attached" to the linker to obtain 10,000 conformers of the protein with no steric conflicts. (a) Projections of all the conformers (blue) onto the PCA modes is defined in Figure S1b. For comparison, projections of the ACM simulation (red) are also shown. (b) The 20 conformers in the ensemble selected by EOM, and (c) the back-calculated SAXS profile of the ensemble (black) is fitted to the experimental data (red). (d) The six conformers in the ensemble selected by EOM 2.0, and (e) the back-calculated SAXS profile of the ensemble is fitted to the experimental data.

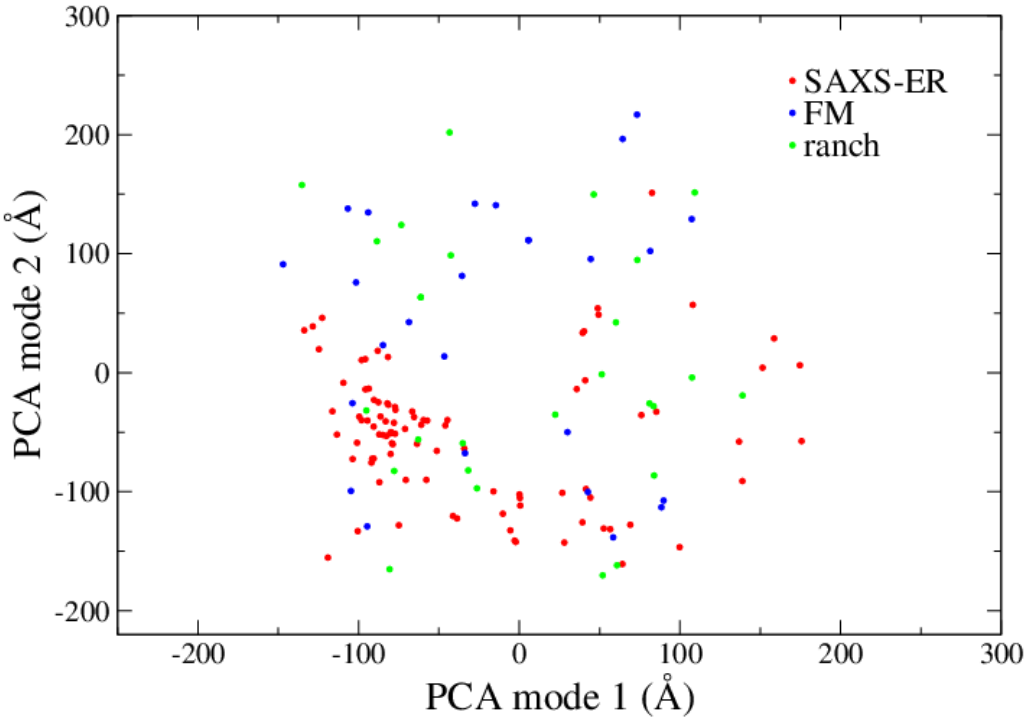

Figure S10. Projections of the conformers in all the SAXS-ER ensembles (Fig. 2b, S6 and S7c) onto the PCA modes defined in Figure S1. For comparison, projections of the ranch (Fig. S8b and S8d) and *flexible-meccano* (Fig. S9b and S9d) ensembles are also shown.

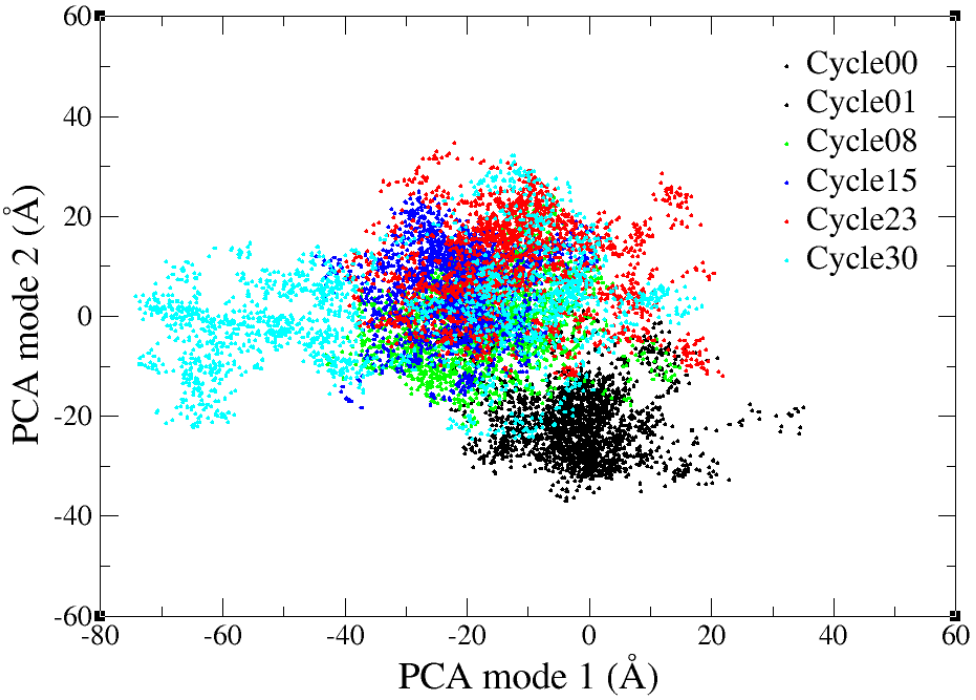

Figure S11. The evolution of sampled conformational space of the free SAM-1 aptamer with the iteration cycles are shown in projections onto the PCA modes defined in Figure S2b.

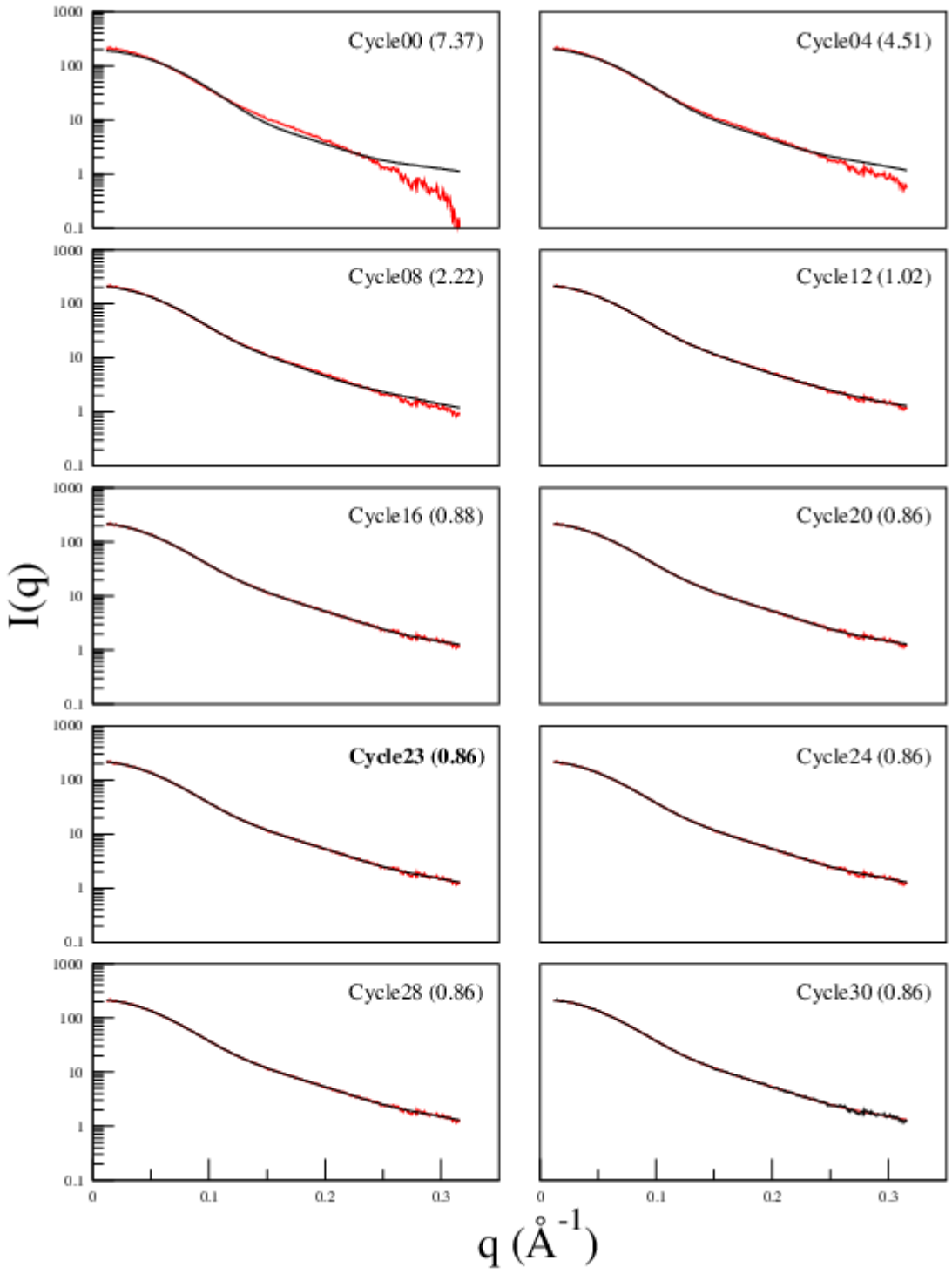

Figure S12. SAXS profiles (black) of the ensembles of the free SAM-1 aptamer from the initial to the final cycles of SAXS-ER, to show how they fit to the experimental SAXS data (red). In each cycle, the χ value between the theoretical and the experimental SAXS profiles is given in parenthesis. The final ensemble is at the 23rd cycle (bold font).

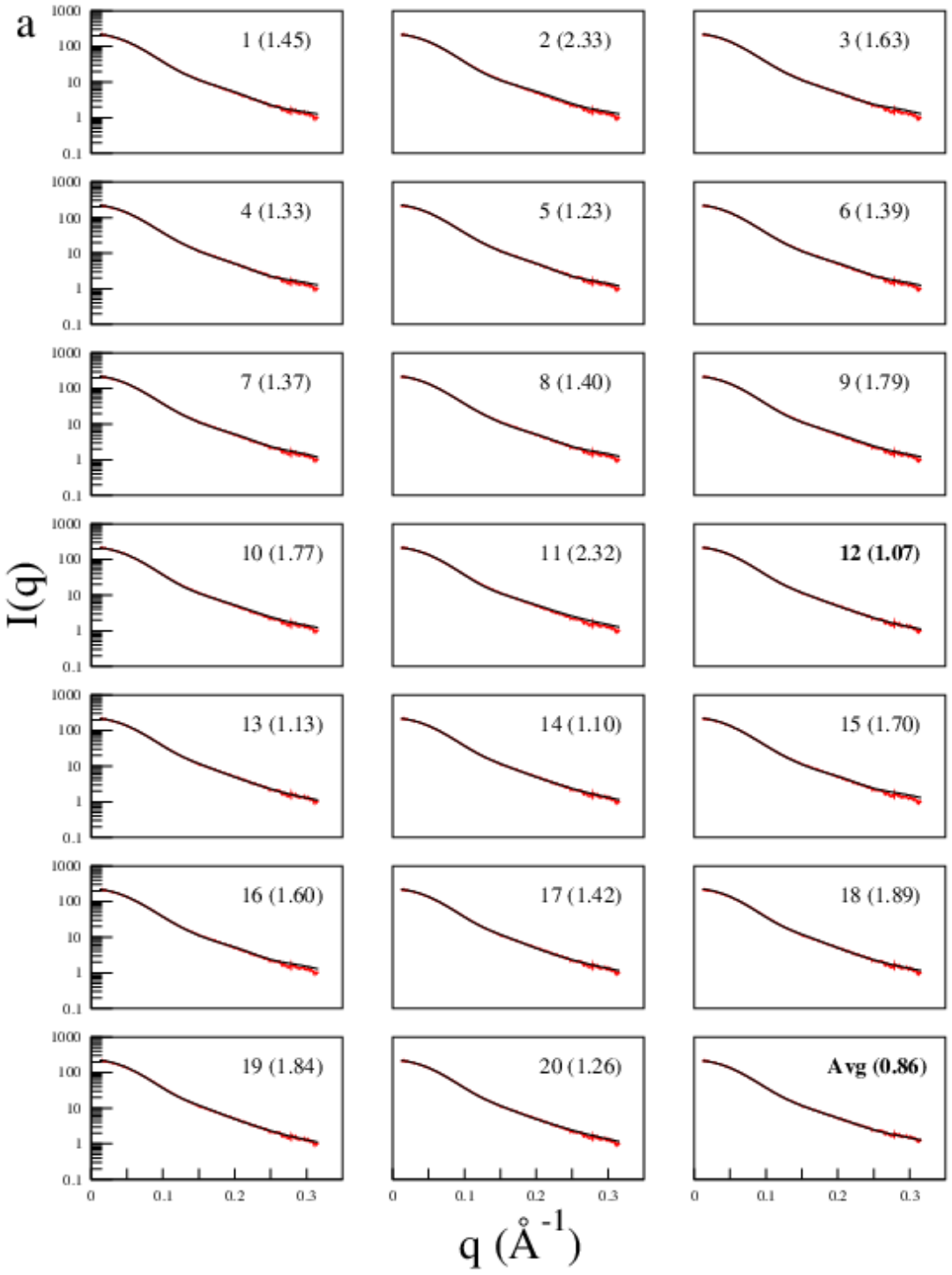

Figure S13a. SAXS curves of individual conformers (black) in the final ensemble of the free SAM-1 aptamer (Fig. 3b) and the average profile of the ensemble (bold font), which are all fitted to the experimental data (red). In each panel, the χ value between the theoretical and the experimental SAXS profiles is given in parenthesis. The single conformation with the best fitting SAXS curve (the smallest χ) is also highlighted in bold font.

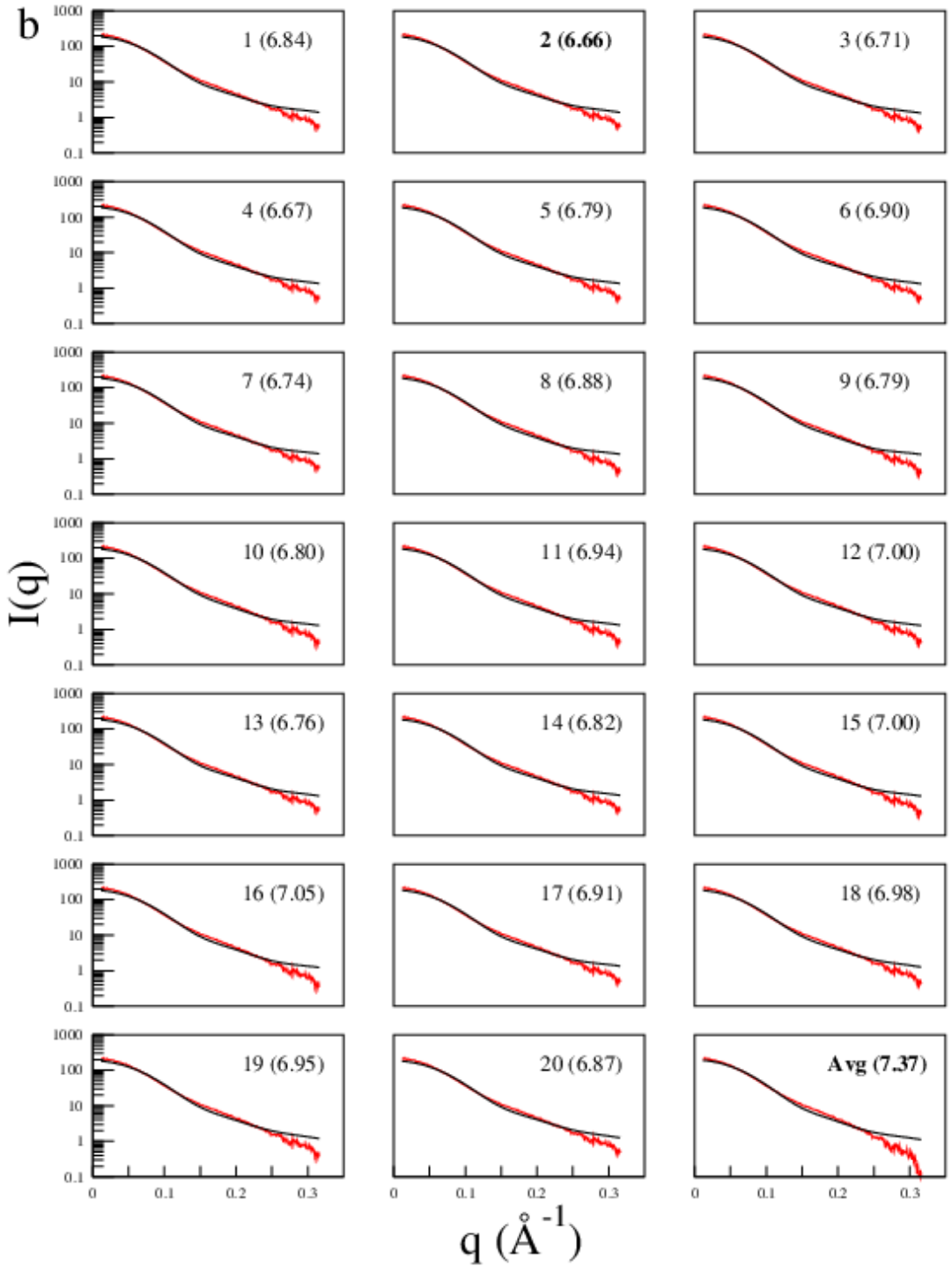

Figure S13b. SAXS curves of individual conformers (black) in the initial ensemble (Cycle 0) of the free SAM-1 aptamer and the average profile of the ensemble (bold font), which are all fitted to the experimental data (red). In each panel, the χ value between the theoretical and the experimental SAXS profiles is given in parenthesis. The single conformation with the best fitting SAXS curve (the smallest χ) is also highlighted in bold font.

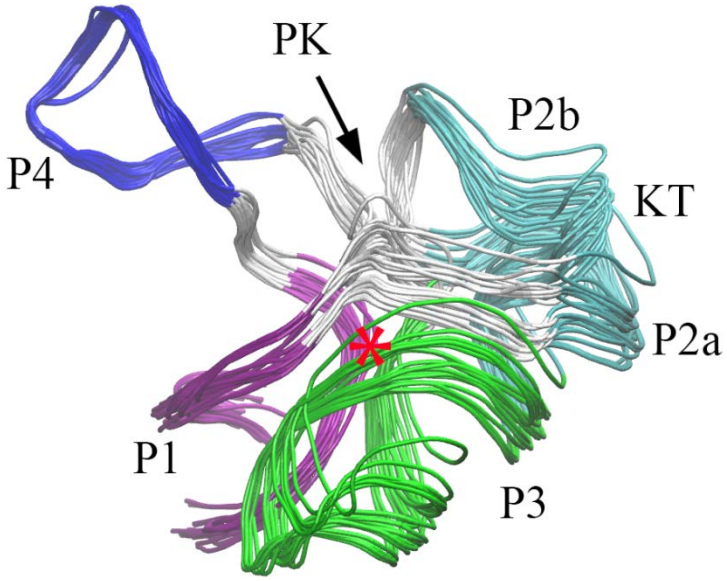

Figure S14. Structural ensembles from another SAXS-ER of the free SAM-1 aptamer. The ensemble size of EOM was N_{es}=20. Each cycle consisted of N_{sim}=20 independent 100-ps aMD simulations. The aMD parameters were estimated from a 10-ns MD trajectory. All the structures are superimposed on the subdomain P4, and the coloring is the same as that in Figure 3b. The location of SAM is approximated by a red star.

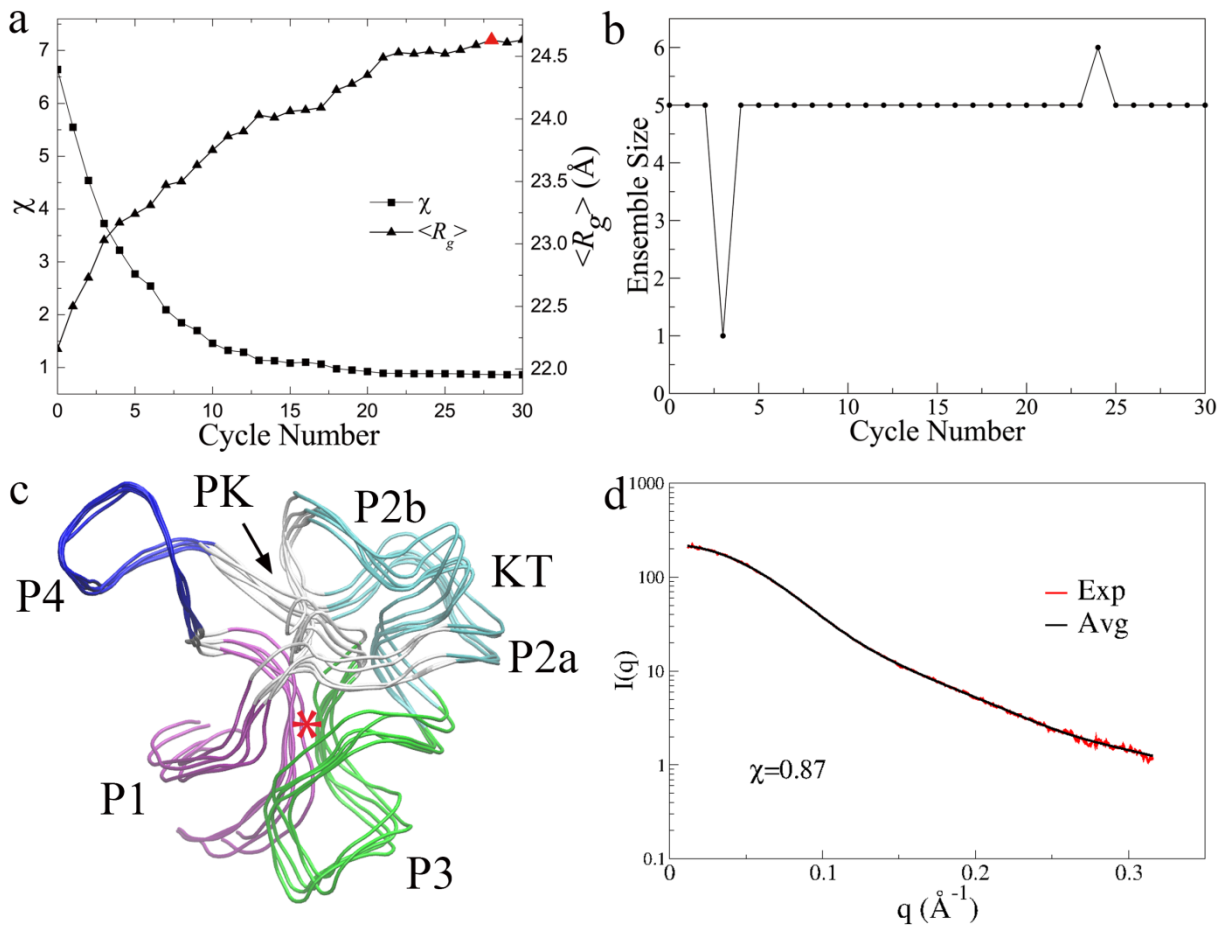

Figure S15. SAXS-ER of the free SAM-1 aptamer with EOM 2.0. The ensemble size N_{es} at each cycle was optimized in EOM 2.0. Each cycle consisted of N_{sim}=20 independent 100-ps aMD simulations starting from the N_{es} conformers (each of them was used more than once). Those aMD parameters were estimated from a 200-ns MD simulation. (a) The minimal χ and the corresponding <R_g> at each cycle. The final ensemble at the 28th cycle is indicated by a red triangle. (b) The ensemble size at each cycle. (c) The conformers in the final ensemble, which are superimposed on the subdomain P4. The coloring is the same as that in Figure 3b. (d) The back-calculated SAXS profile of the final ensemble (black) is fitted to the experimental data (red).