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

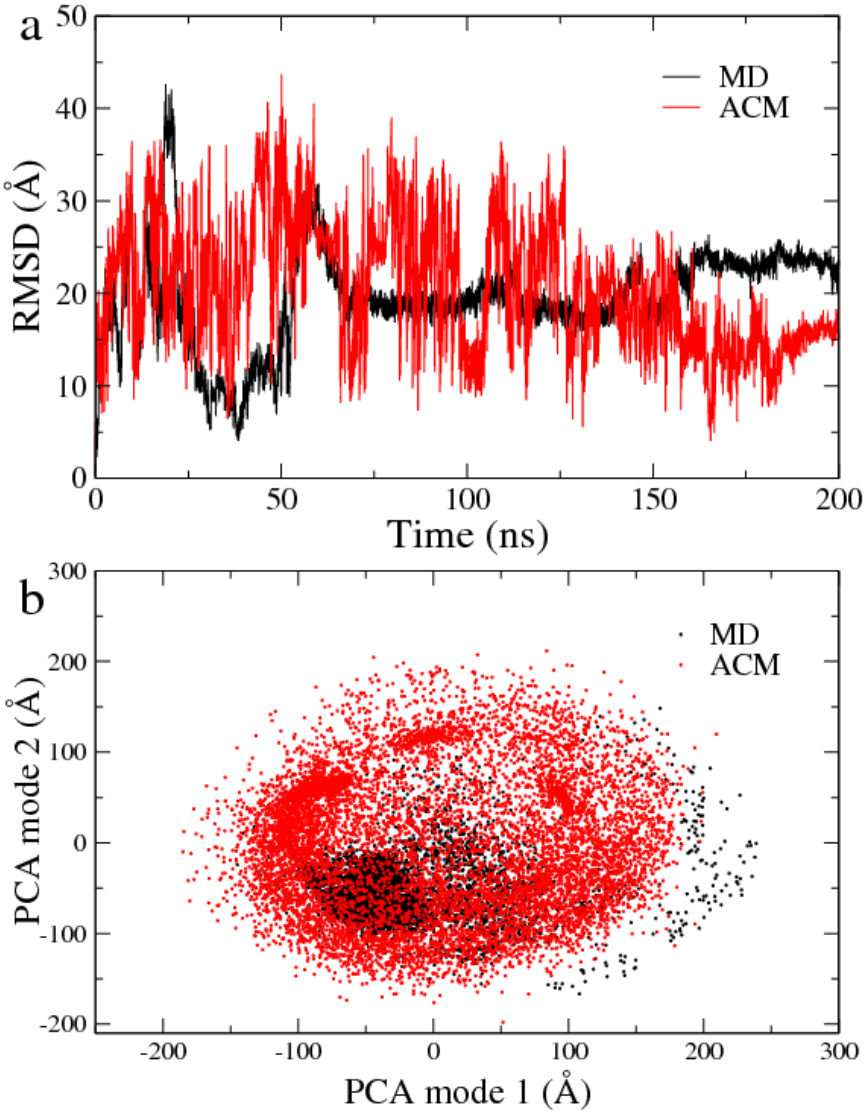

Figure S1. Sampling efficiency of ACM compared to the standard MD. From the same initial structure of FBP21-WWs, both ACM and MD were carried out for 200 ns. (a) Time evolution of the root mean square deviation (RMSD) during the ACM (red) and MD (black) simulations. (b) PCA on the ACM trajectory. The 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.

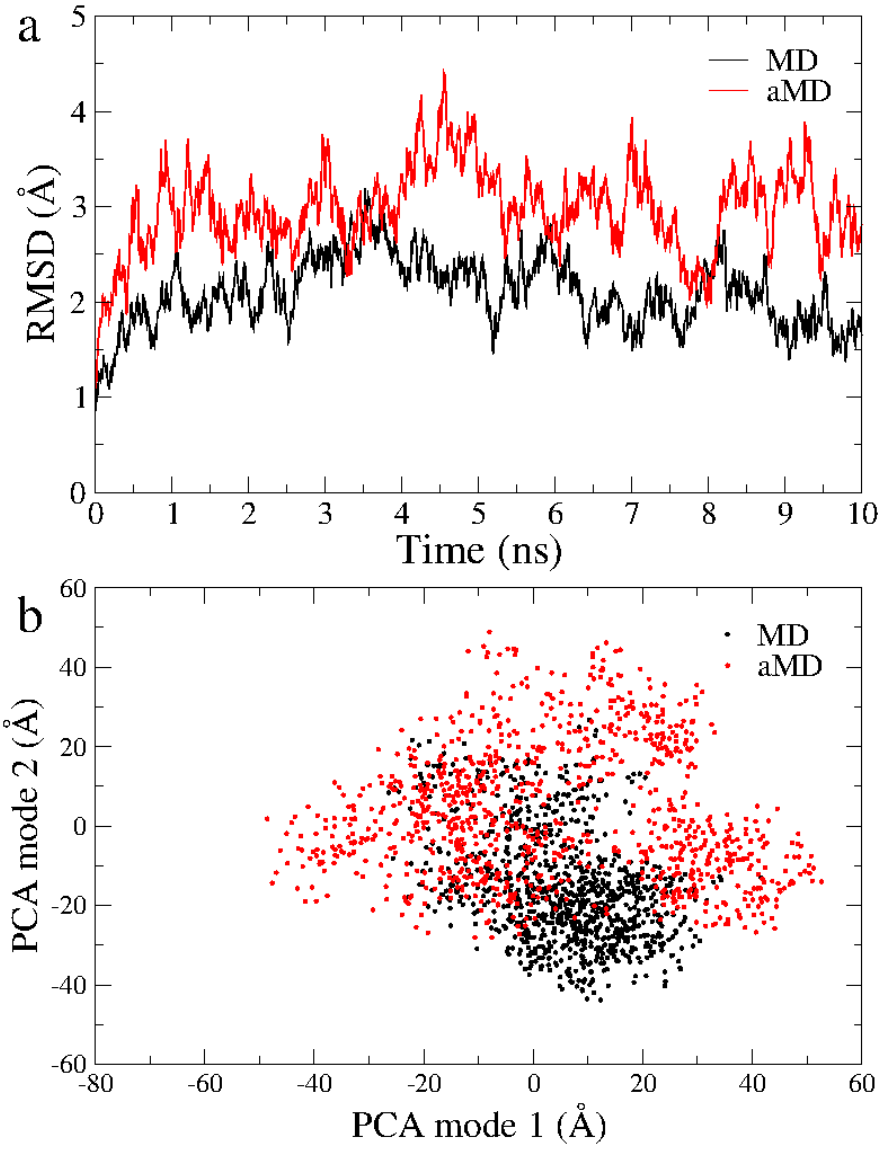

Figure S2. Sampling efficiency of aMD compared to the standard MD. From the crystal structure 2GIS with the ligand removed, both aMD and MD were carried out for 10 ns. (a) Time evolution of the RMSD during the aMD (red) and MD (black) simulations. (b) Projections of the aMD (red) and MD (black) simulations 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.

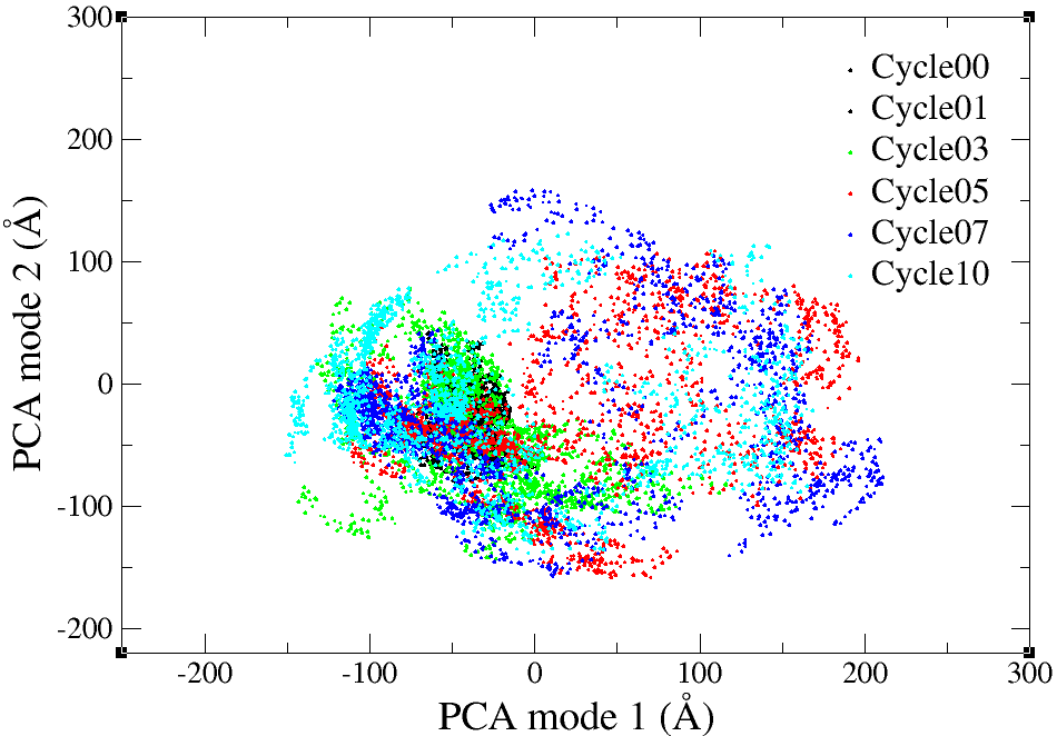

Figure S3. The evolution of the sampled conformational space of FBP21-WWs against the SAXS-ER iteration cycles are shown in projections onto the PCA modes (defined in Figure S1b).

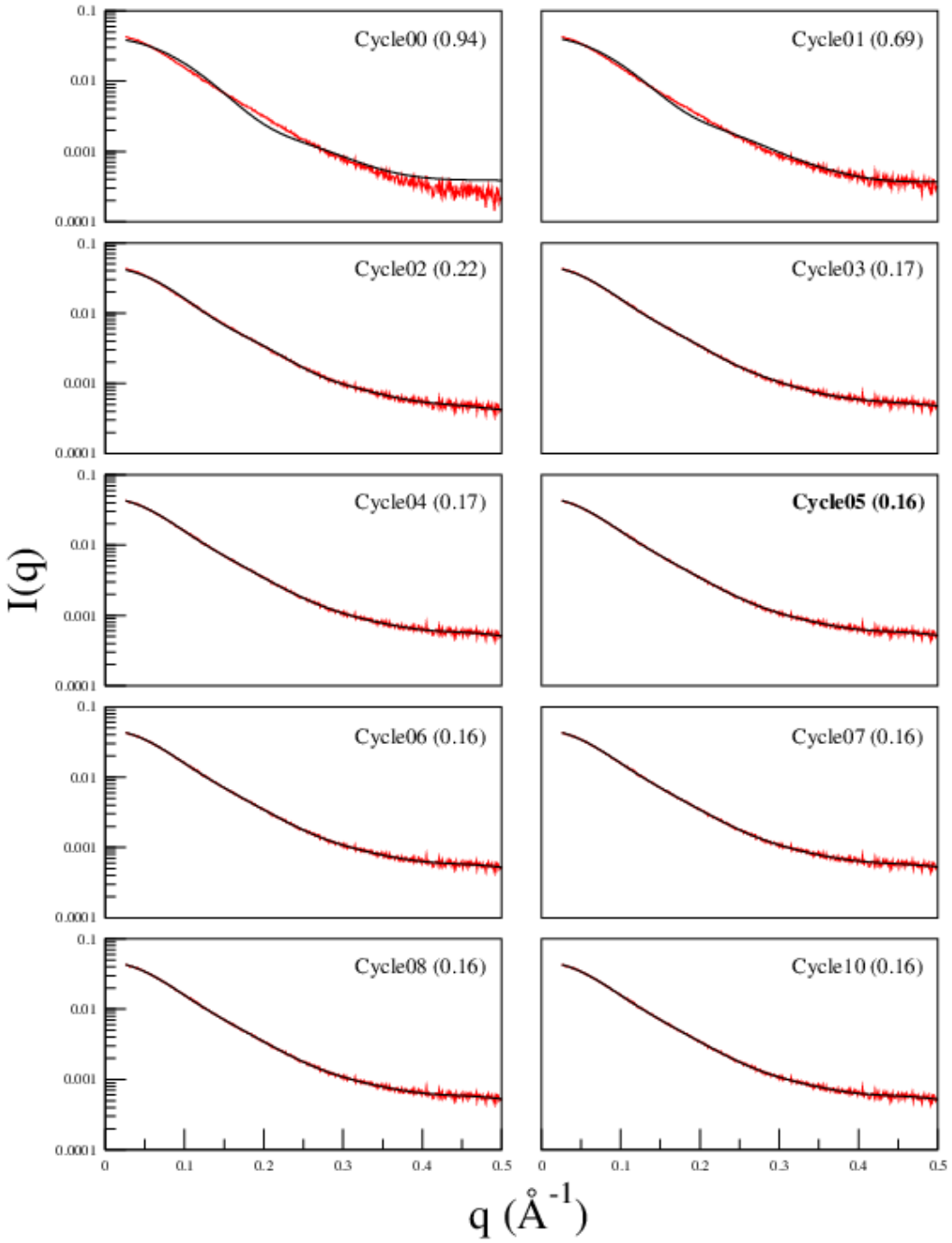

Figure S4. SAXS profiles (black) of the FBP21-WWs ensembles of from the initial to the final SAXS-ER cycles to show how they fit to the experimental SAXS data (red). In each cycle, the χ value between the theoretical and the experimental SAXS profile is given in parenthesis. The final selected ensemble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

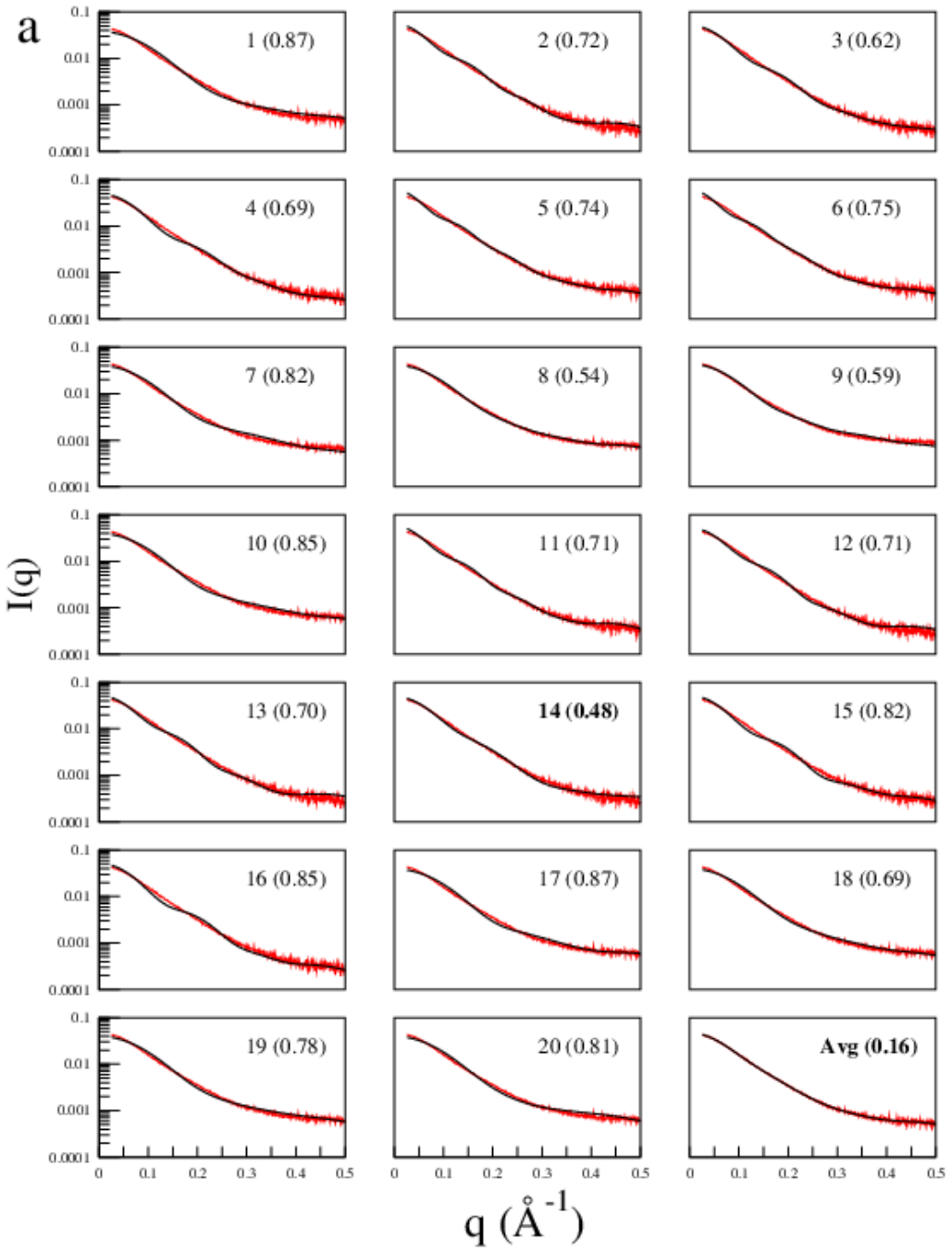

Figure S5a. SAXS curves of individual conformers (black) in the final ensemble of FBP21-WWs (Fig. 2b) and the average profile of the ensemble (bold font), which are all fitted to the experimental data (red). In each panel, the χ value between the theoretical and the experimental SAXS profiles is given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

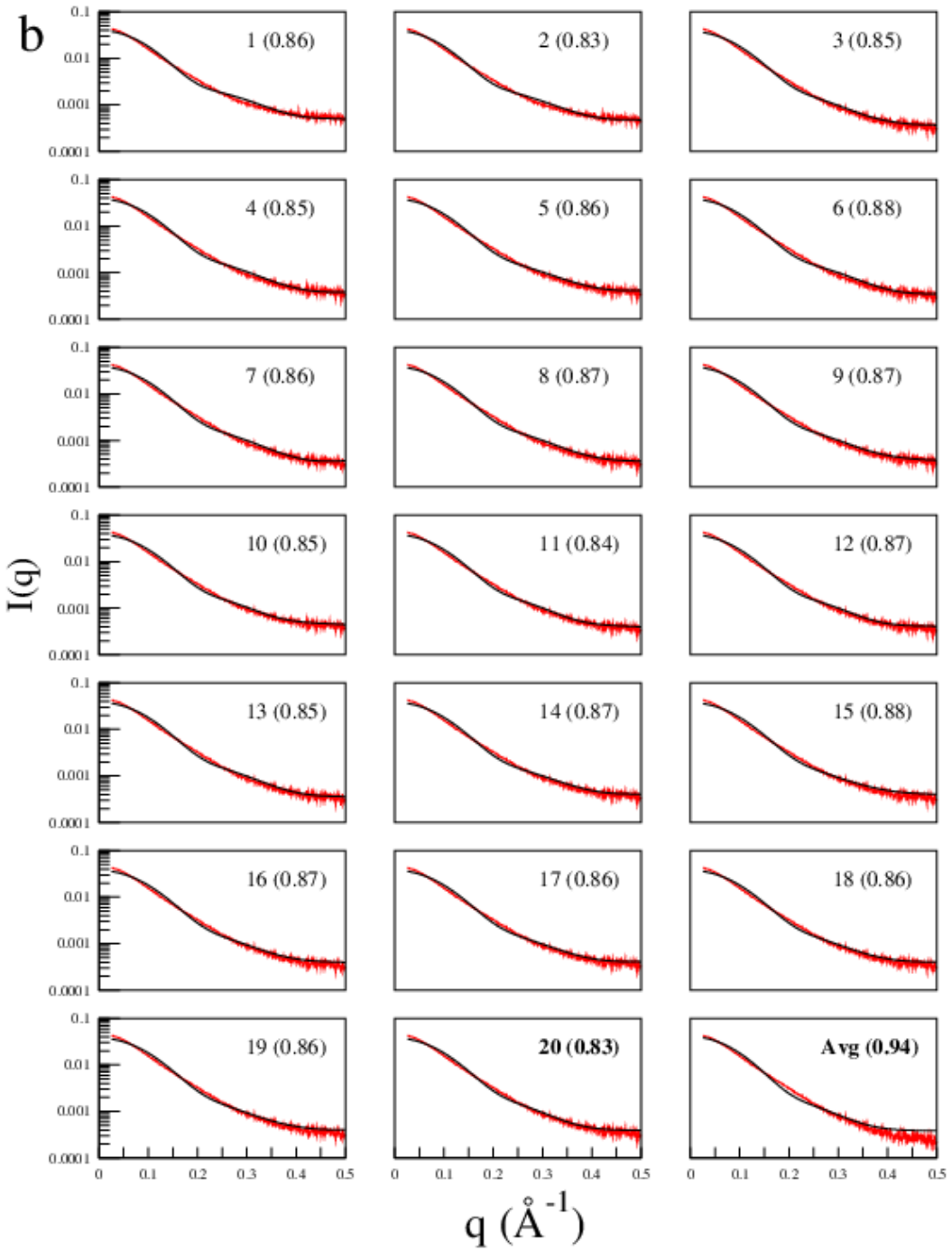

Figure S5b. SAXS curves for the individual conformers (black) in the initial ensemble (Cycle 0) of FBP21-WWs and the average profile of the ensemble (bold font), which are all fitted to the experimental data (red). In each panel, the χ value between the theoretical and the experimental SAXS profile is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

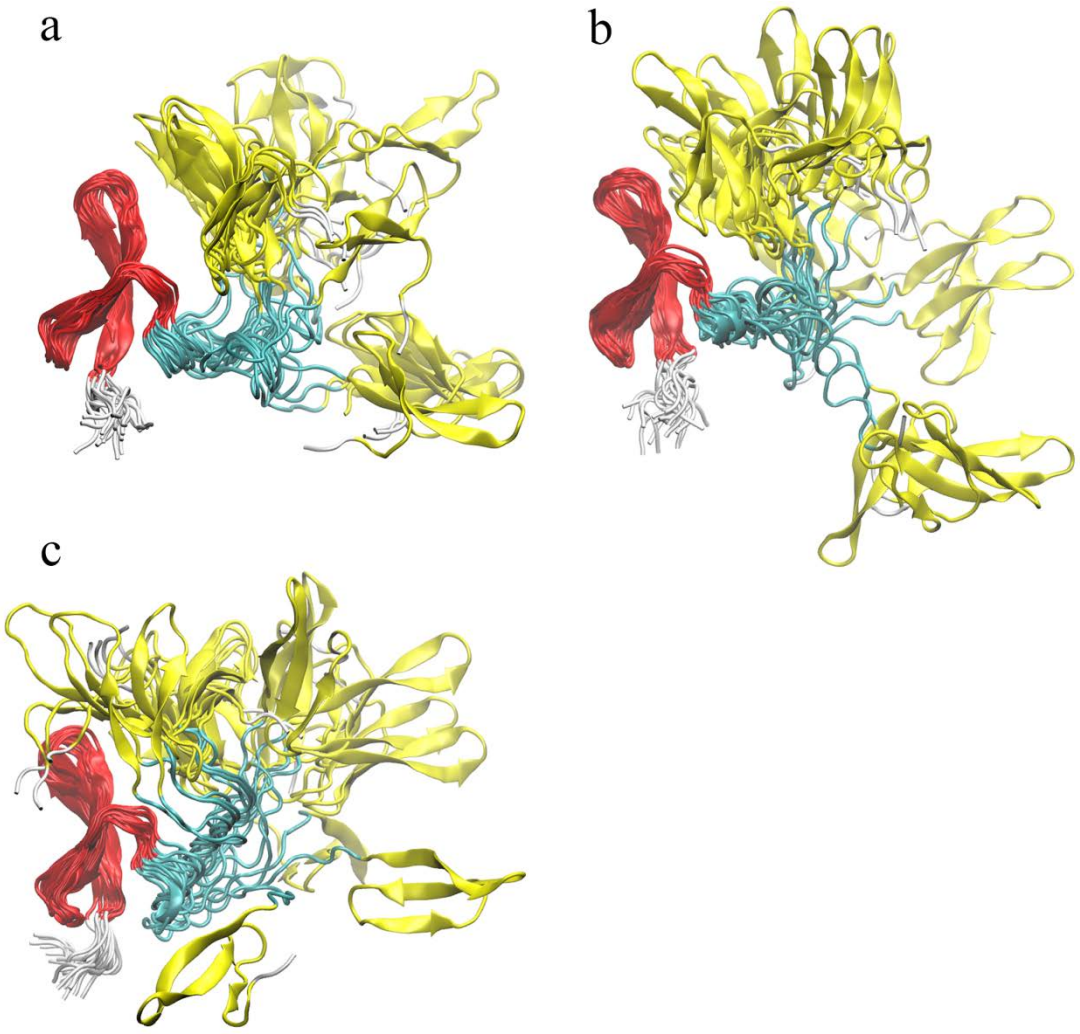

Figure S6. Structural ensembles from other SAXS-ER simulations of FBP21-WWs. (a) The starting conformation was the model 1 of the NMR structures. The ensemble size of EOM was N_{es}=20. Each cycle consisted of N_{sim}=20 independent 200-ps ACM simulations, in which those low-frequency collective motions 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.

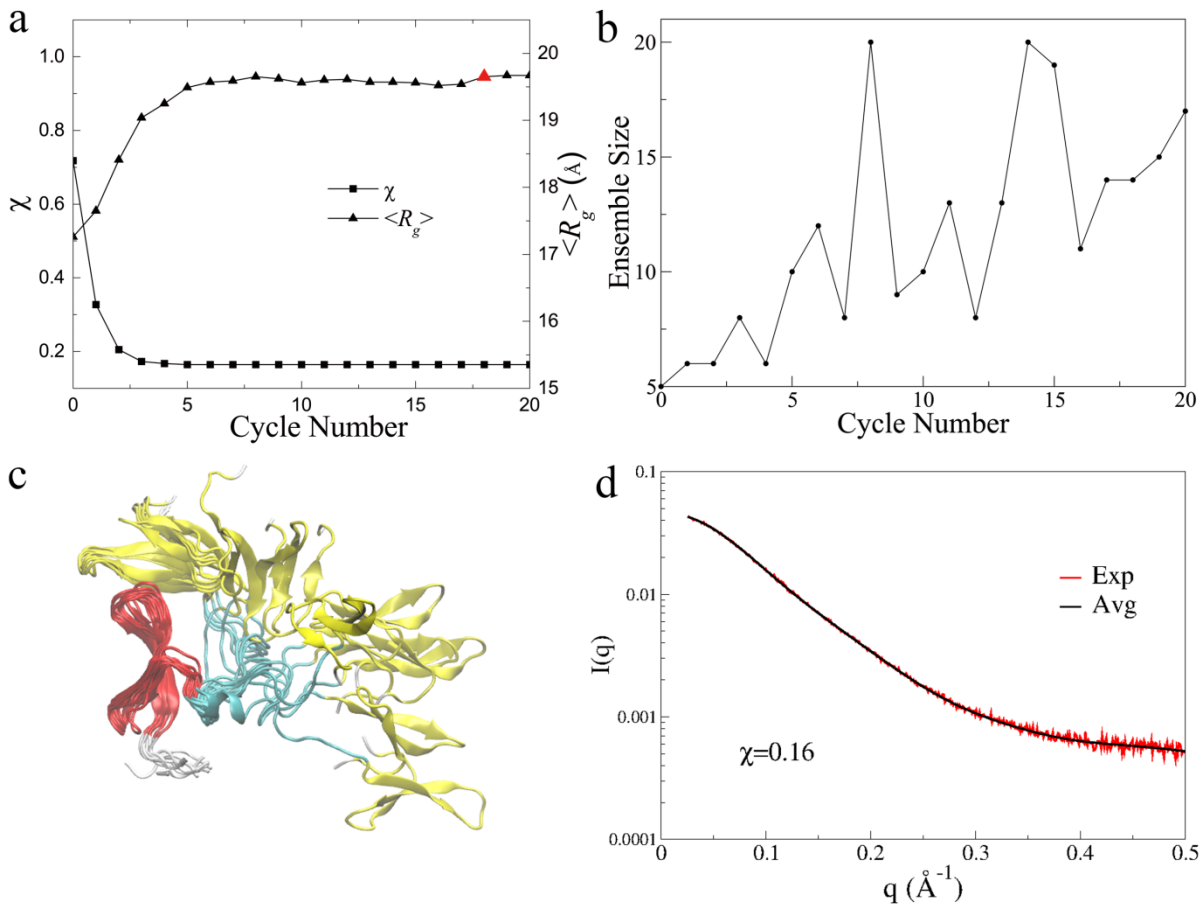

Figure S7. SAXS-ER of FBP21-WWs with EOM 2.0. The starting conformation was the model 1 of the NMR structures. The ensemble size N_{es} at each cycle was optimized in EOM 2.0. Each cycle consisted of N_{sim}=20 independent 100-ps ACM simulations starting from the N_{es} conformers (some of them were 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).

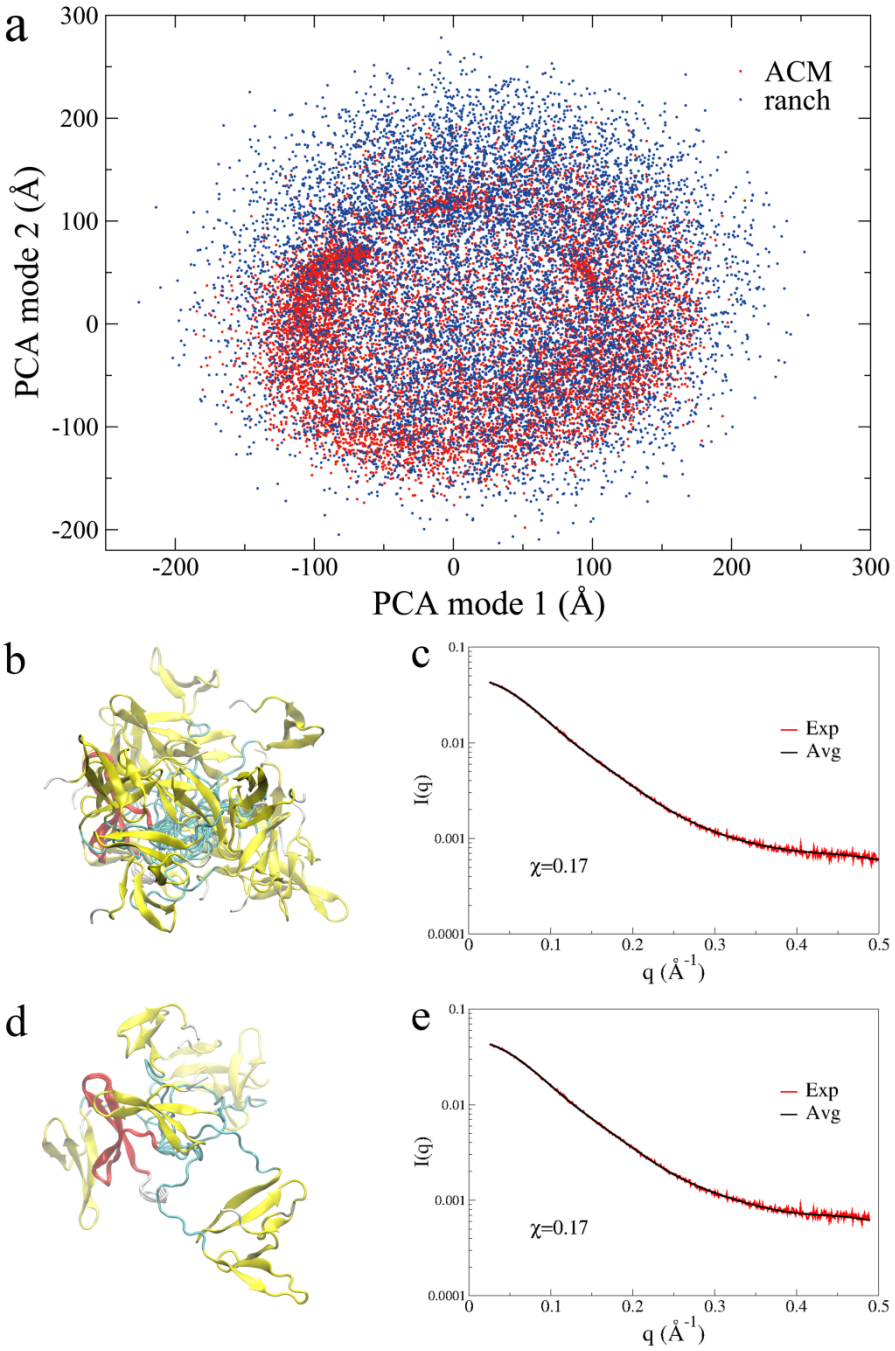

Figure S8. EOM on a structural pool containing 10,000 conformers of FBP21-WWs generated by ranch. (a) Projections of all the conformers (blue) onto the PCA modes defined in Figure S1b. For comparison, projections of the ACM simulation (red) are also shown. (b) The 20 conformers in the ensemble selected 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.

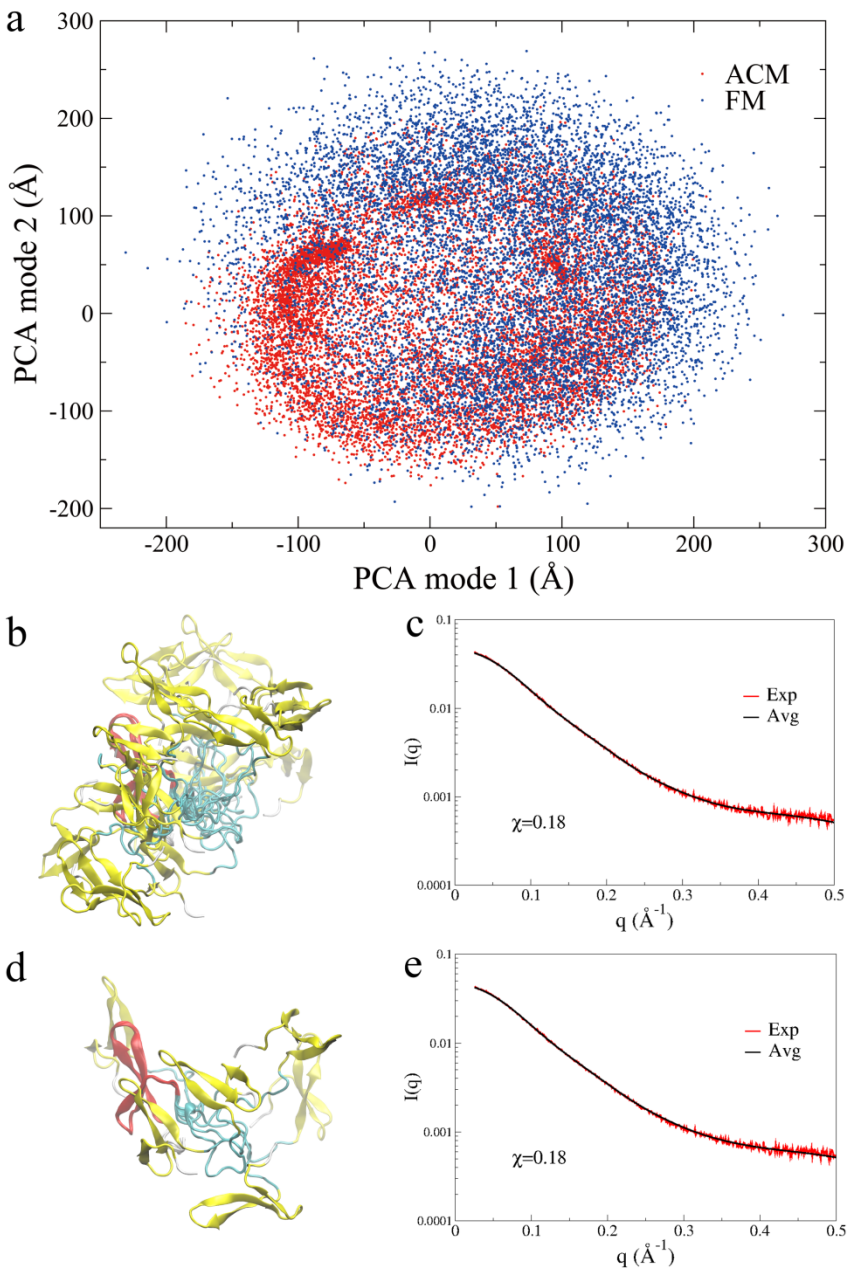

Figure S9. EOM on a structural pool of FBP21-WWs generated in the f*lexible-meccano* statistical coil model. The program produced a large number of linker conformations, and then the WW1 and WW2 domains were "attached" to the linker to obtain 10,000 conformers of the protein with no steric conflicts. (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.

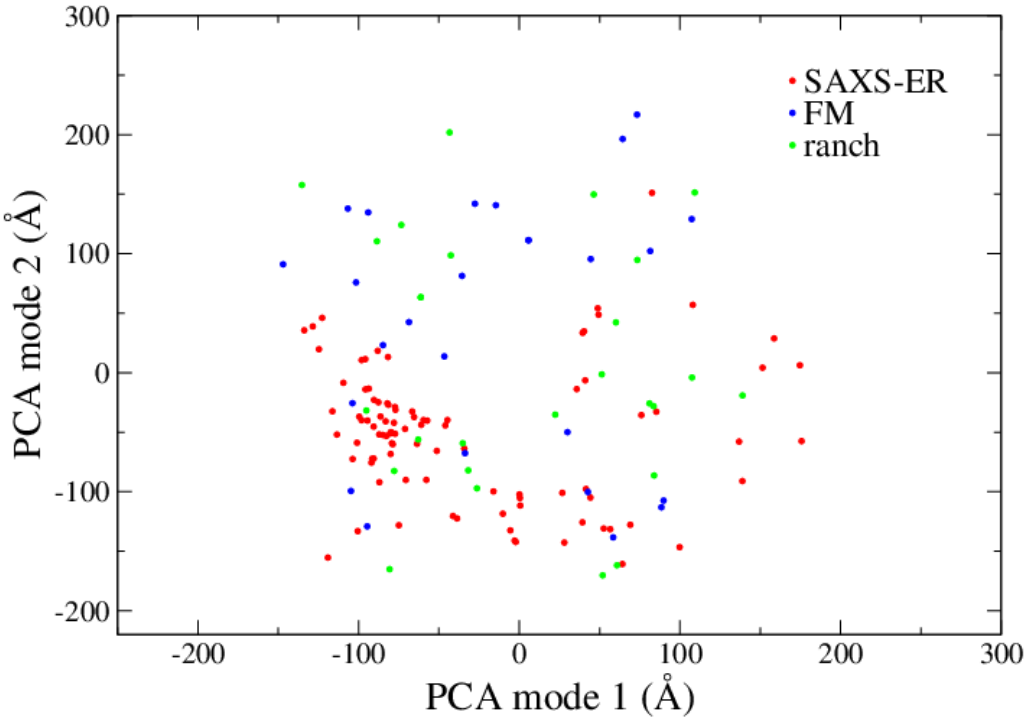

Figure S10. Projections of the conformers in all the SAXS-ER ensembles (Fig. 2b, S6 and S7c) onto the PCA modes defined in Figure S1. For comparison, projections of the ranch (Fig. S8b and S8d) and *flexible-meccano* (Fig. S9b and S9d) ensembles are also shown.

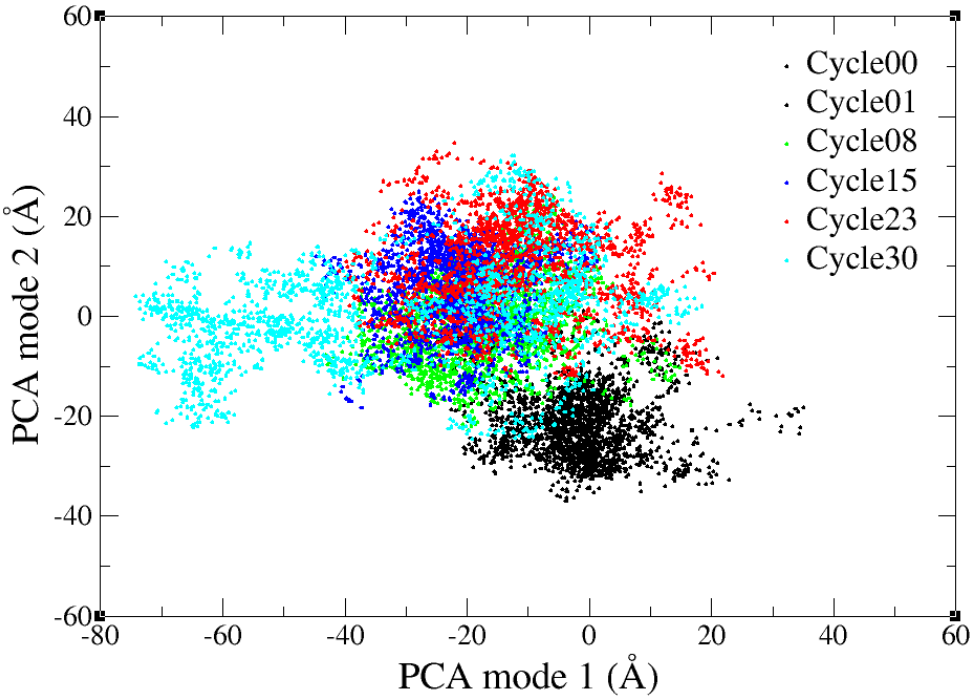

Figure S11. The evolution of sampled conformational space of the free SAM-1 aptamer with the iteration cycles are shown in projections onto the PCA modes defined in Figure S2b.

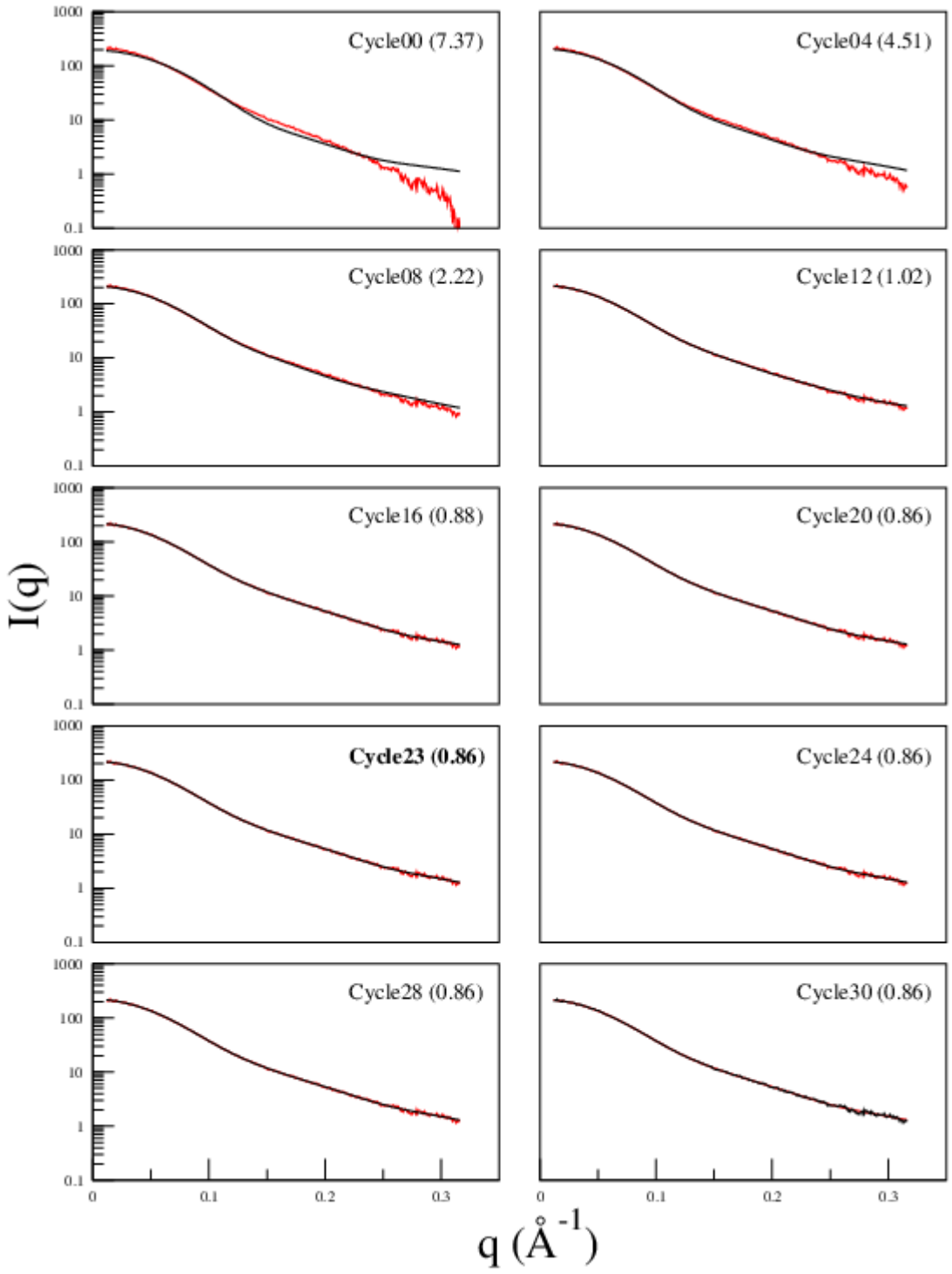

Figure S12. SAXS profiles (black) of the ensembles of the free SAM-1 aptamer from the initial to the final cycles of SAXS-ER, to show how they fit to the experimental SAXS data (red). In each cycle, the χ value between the theoretical and the experimental SAXS profiles is given in parenthesis. The final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

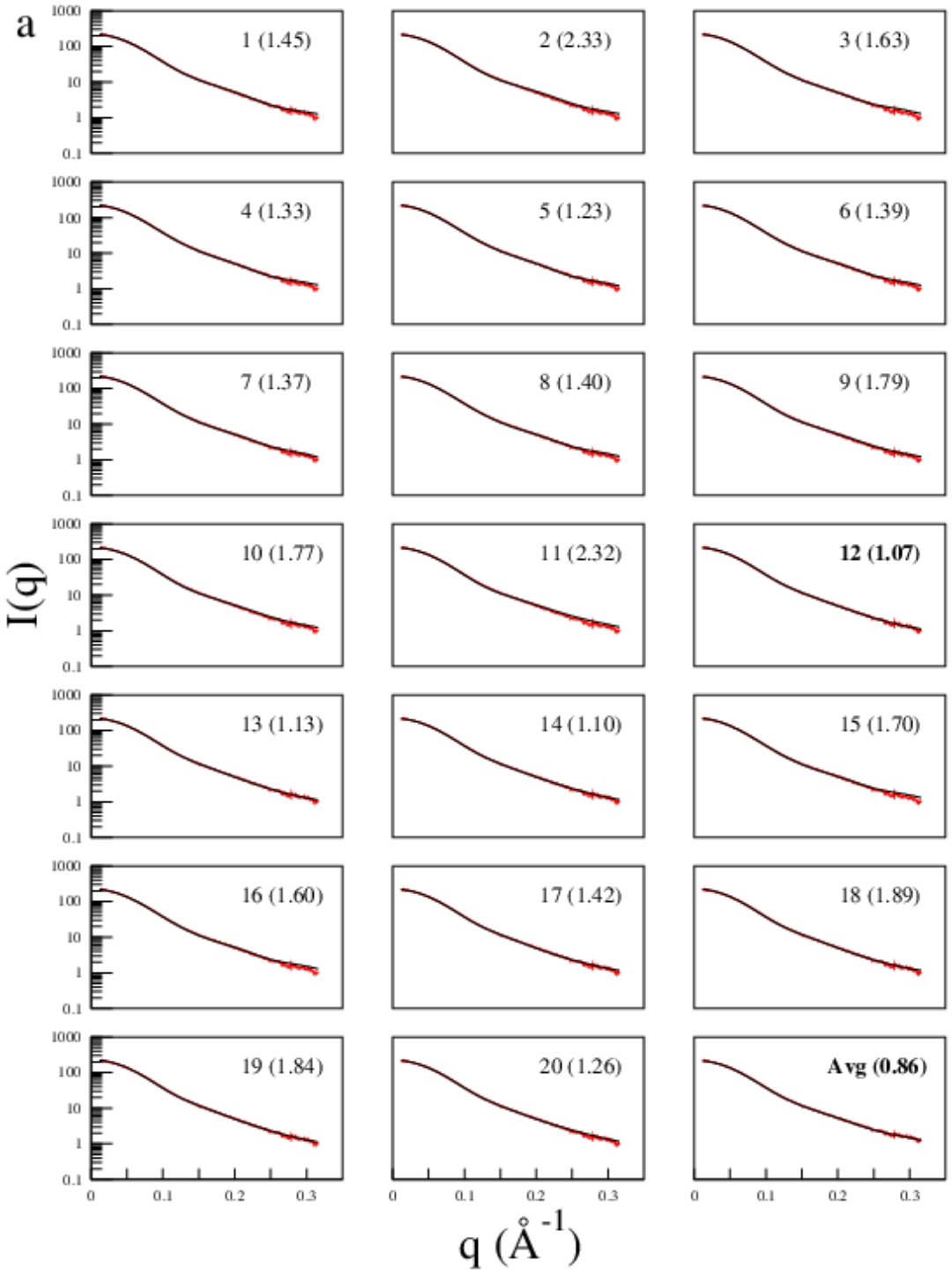

Figure S13a. SAXS curves of individual conformers (black) in the final ensemble of the free SAM-1 aptamer (Fig. 3b) and the average profile of the ensemble (bold font), which are all fitted to the experimental data (red). In each panel, the χ value between the theoretical and the experimental SAXS profiles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

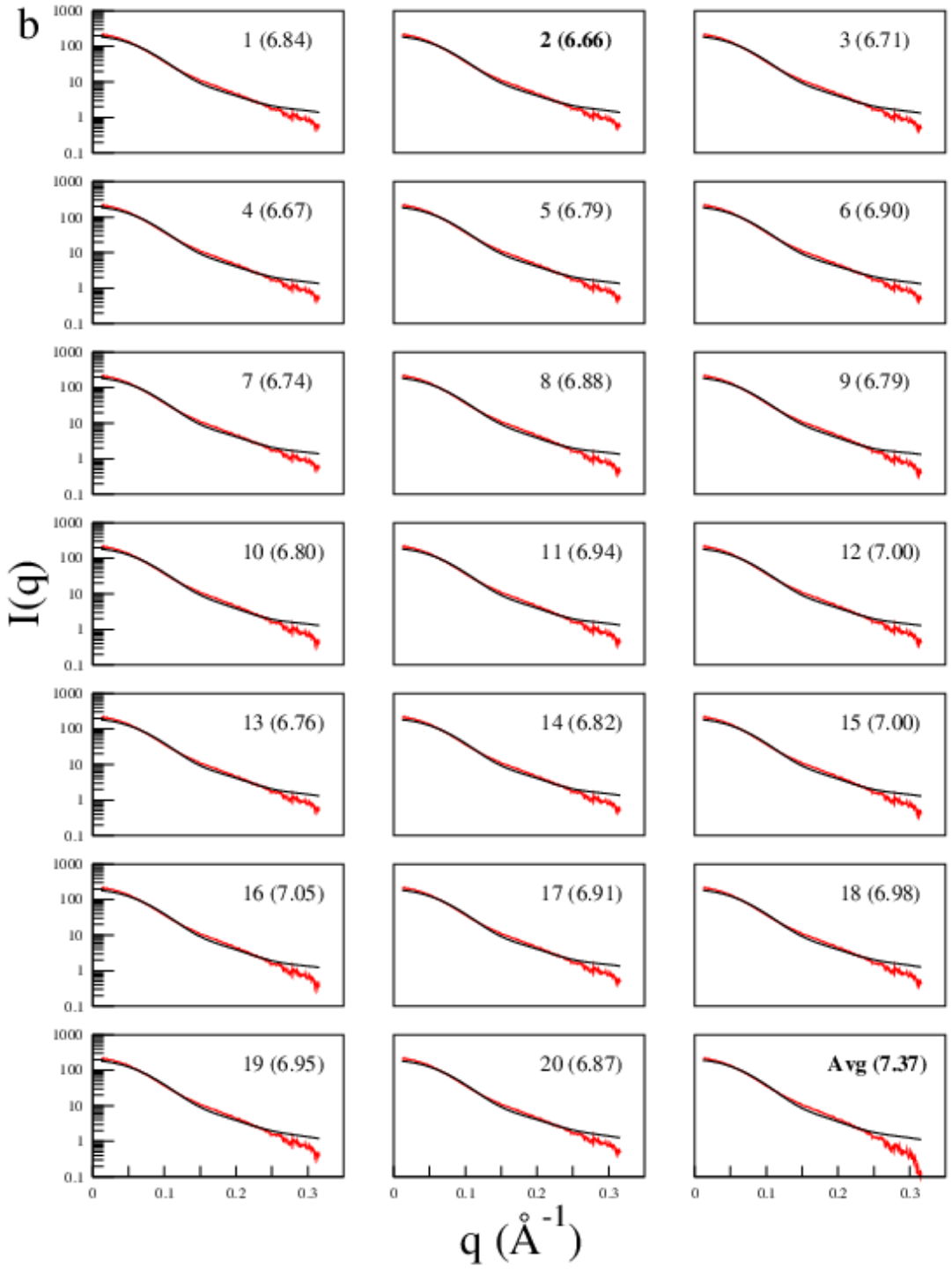

Figure S13b. SAXS curves of individual conformers (black) in the initial ensemble (Cycle 0) of the free SAM-1 aptamer and the average profile of the ensemble (bold font), which are all fitted to the experimental data (red). In each panel, the χ value between the theoretical and the experimental SAXS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

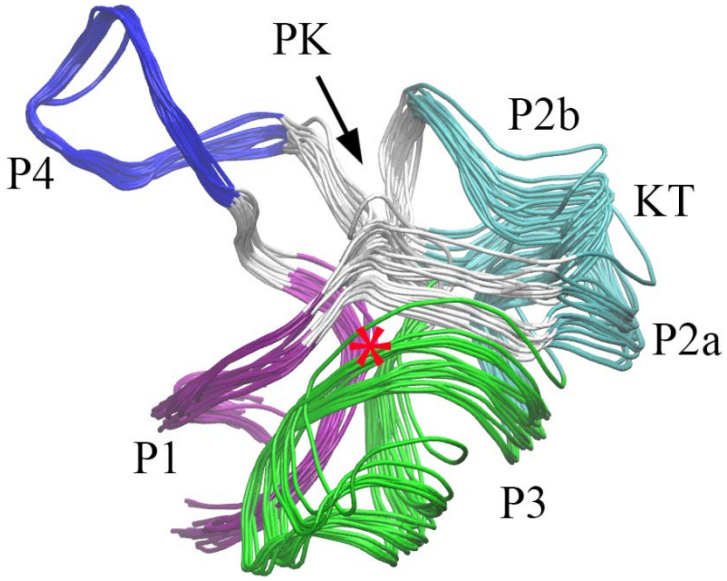

Figure S14. Structural ensembles from another SAXS-ER of the free SAM-1 aptamer. The ensemble size of EOM was N_{es}=20. Each cycle consisted of N_{sim}=20 independent 100-ps aMD simulations. The aMD parameters were estimated from a 10-ns MD trajectory. All the structures are superimposed on the subdomain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

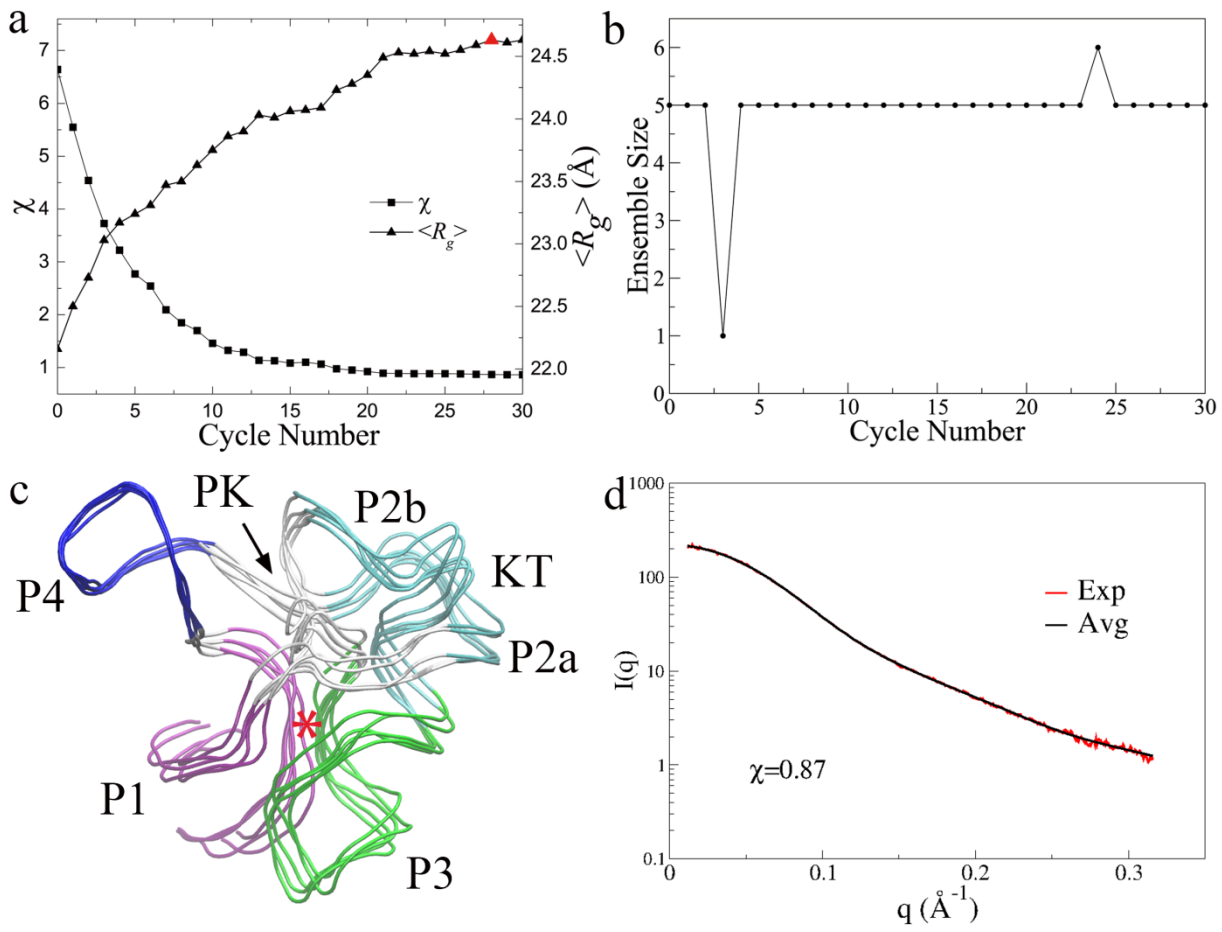

Figure S15. SAXS-ER of the free SAM-1 aptamer with EOM 2.0. The ensemble size N_{es} at each cycle was optimized in EOM 2.0. Each cycle consisted of N_{sim}=20 independent 100-ps aMD simulations starting from the N_{es} conformers (each of them was used more than once). Those aMD parameters were estimated 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).