Accurate Flexible Fitting of High-Resolution Protein Structures to Small-Angle X-Ray Scattering Data Using a Coarse-Grained Model with Implicit Hydration Shell

Wenjun Zheng* and Mustafa Tekpinar
Physics Department, University at Buffalo, State University of New York, Buffalo, New York

ABSTRACT Small-angle x-ray scattering (SAXS) is a powerful technique widely used to explore conformational states and transitions of biomolecular assemblies in solution. For accurate model reconstruction from SAXS data, one promising approach is to flexibly fit a known high-resolution protein structure to low-resolution SAXS data by computer simulations. This is a highly challenging task due to low information content in SAXS data. To meet this challenge, we have developed what we believe to be a novel method based on a coarse-grained (one-bead-per-residue) protein representation and a modified form of the elastic network model that allows large-scale conformational changes while maintaining pseudobonds and secondary structures. Our method optimizes a pseudoenergy that combines the modified elastic-network model energy with a SAXS-fitting score and a collision energy that penalizes steric collisions. Our method uses what we consider a new implicit hydration shell model that accounts for the contribution of hydration shell to SAXS data accurately without explicitly adding waters to the system. We have rigorously validated our method using five test cases with simulated SAXS data and three test cases with experimental SAXS data. Our method has successfully generated high-quality structural models with root mean-squared deviation of 1 ~ 3 Å from the target structures.

INTRODUCTION

The biological functions of many biomolecules involve dynamic transitions between different conformational states. Despite recent progress, it remains difficult to capture all conformational states of large biomolecules by high-resolution techniques like x-ray crystallography and NMR. Instead, low-resolution techniques including cryo-electron microscopy (cryo-EM) (1) and small-angle x-ray scattering (SAXS) (2–4) are increasingly used to investigate large biomolecular assemblies. These techniques, however, can only provide nanometer-resolution pictures of biomolecular structures. Computer modeling is needed to deduce atomic details of biomolecular conformations.

SAXS is a well-established technique (2,3,5–7) that measures the orientationally averaged x-ray scattering intensity of biomolecules in solution. It has the following major advantages: first, it does not require special sample preparation and is fairly easy and fast to conduct; second, it can probe the structure and dynamics of biomolecular assemblies under various physiological conditions inaccessible to alternative techniques. However, SAXS data contain much less information than x-ray crystallography and cryo-EM (the SAXS intensity profile is one-dimensional whereas x-ray and cryo-EM data are three-dimensional). Thus, SAXS is often used to measure the overall size and shape of biomolecules. Recently, SAXS data have been integrated with computational algorithms and other structural data to aid protein structure prediction (8–10) and model protein assemblies (4,11–13).

Based on the molecular shape reconstructed from SAXS data (14–18), biomolecular assemblies can be modeled by rigidly fitting known high-resolution structures of individual components into the molecular shape of the assembly (19). However, if there are significant conformational changes during the assembling process, flexible fitting must be used to allow structural changes within each component. The flexible fitting of a high-resolution structure to the low-resolution data of SAXS or cryo-EM is often done by domain segmentation followed by the fitting of each domain as a rigid body (20–27). Such methods depend on a subjective and error-prone partition of a biomolecule into rigid domains and ignore coupled motions between domains that may be functionally important.

The flexibility of biomolecules can be modeled by coarse-grained models using simplified representations of biomolecular structures. A variety of coarse-grained models (28) have been developed to simulate protein conformational dynamics with high efficiency. For example, the elastic network model (ENM) (29–31) represents a protein structure as a network of C_α atoms with neighboring ones connected by springs with a uniform force constant (32). The ENM-based normal mode analysis (NMA) has been widely utilized to flexibly fit high-resolution structures to low-resolution structural data (33–41), or satisfy a few pairwise distance constraints (42,43). In particular, it was successfully used to flexibly fit an x-ray structure to an atomic pair distribution function (PDF) (41). This technique was later improved to take into account a hydration shell around the protein (44,45). Despite great success, the previous ENM-based flexible fitting methods are limited in accuracy because

Submitted July 31, 2011, and accepted for publication November 4, 2011.
*Correspondence: wjzheng@buffalo.edu
Editor: Lois Pollack.

0006-3495/11/12/2981/11 $2.00
doi: 10.1016/j.bpj.2011.11.003

they usually use only a few (2 ~ 3) lowest normal modes solved from ENM (41,44). As shown in one study, the lowest 20 modes only contribute $\leq 50\%$ of the total conformational change in four test cases (myosin, calmodulin, NtrC, and hemoglobin, see Petrone and Pande (46)). Furthermore, it is well known that the lowest normal modes are less accurate for describing small local conformational changes (like rearrangement of helices inside a densely packed region) than large global ones (like domain motions) (31).

Molecular dynamics (MD) simulation is capable of modeling the dynamics of biomolecules at both global and local scales. Recently, several MD-based methods have been introduced for cryo-EM fitting with full flexibility. The common strategy of these methods is to bias the MD simulation toward a conformation that optimally fits the cryo-EM data by using a biasing potential function (47–51). The application of these methods to the fitting of low-resolution structural data, however, has been limited by the high computational cost of running MD simulations for large biomolecular systems. Recent efforts to improve simulation efficiency include a study that combined SAXS data with coarse-grained simulations to characterize the assembly states of a kinase (52).

To achieve both accuracy and efficiency in the flexible fitting of SAXS data, we propose what we believe to be a new coarse-grained method with the following novel features:

First, it is based on a modification of the ENM energy that uses harmonic interactions to maintain pseudobonds and secondary structures, and anharmonic interactions between nonbonded beads to allow nonbonded residues to move apart readily (53). As a result, large global structural changes can be sampled without distorting local structures.

Second, it optimizes a pseudoenergy that combines various terms of the modified ENM energy with a SAXS-fitting score and a collision energy that penalizes steric collisions (see Methods). Unlike previous flexible fitting efforts using the lowest few normal modes (34,35), our approach effectively utilizes all normal modes so that both global and local structural changes can be accurately modeled. Meanwhile, overfitting is controlled by properly using a fitting termination criterion.

Third, it models the protein-surrounding hydration shell implicitly by combining each residue and its nearby "water glob" into a composite glob (see Methods), which accurately accounts for the contribution of hydration shell to experimental SAXS data without explicitly adding waters. In this way, we can maintain a minimal system size and avoid unwanted interference and hindrance to protein motions by water molecules.

Our method is applicable to the fitting of various low-resolution structural data. Recently, it has been successfully used to flexibly fit a high-resolution protein structure into a given cryo-EM map (53). Here we apply it to the flexible fitting of SAXS data, which is considerably more challenging because SAXS data contain much less information than cryo-EM maps and there is additional contribution from hydration shell.

We have validated our method using five test cases with simulated SAXS data (54) and three test cases with experimental SAXS data. This study is, to our knowledge, the first one that attempts to flexibly fit experimental SAXS data directly (rather than using SAXS data as a filter for pregenerated structural models, see Grishaev et al. (55)). This endeavor is much more challenging than fitting simulated SAXS data due to experimental noises and errors. We have shown that our method works well for both simulated and experimental SAXS data, and it compares favorably with an alternative flexible fitting technique based on NMA (34,35).

METHODS

Modified elastic network model

A coarse-grained elastic network model (ENM) is constructed from the atomic coordinates of a protein crystal structure. Each residue is represented by a bead located at the C_α atom.

The original form of the ENM potential energy (32) is

$$E_{\mathrm{ENM}} = \frac{1}{2}\sum_{i<j} C_{ij}\theta\left(R_c - d_{ij,0}\right)\left(d_{ij} - d_{ij,0}\right)^2, \qquad (1)$$

where d_{ij} is the distance between beads i and j, and $d_{ij,0}$ is the value of d_{ij} given by the crystal structure, $\theta(x)$ is the Heaviside function, R_c is the cutoff distance chosen to be 10 Å following our previous study (53), and C_{ij} is the force constant of the spring between beads i and j. C_{ij} can be set to a uniform constant for all residue pairs (30), or two different values for bonded and nonbonded residues (56), or can be allowed to vary as a function of distance $d_{ij,0}$ (57). The unit of C_{ij} can be arbitrarily chosen without changing the modeling results.

To allow nonbonded beads to move apart readily while maintaining pseudobonds and secondary structures, we have modified the ENM energy in Eq. 1 to the form (named mENM energy) (53)

$$\begin{aligned} E_{\mathrm{mENM}} &= E_b + E_{SS} + E_{nb}, \\ E_b &= \frac{1}{2}\sum_{\langle ij\rangle \in P_b} C_b\left(d_{ij} - d_{ij,0}\right)^2, \\ E_{SS} &= \frac{1}{2}\sum_{\langle ij\rangle \in P_{SS}} C_{SS}\left(d_{ij} - d_{ij,0}\right)^2, \\ E_{nb} &= \frac{1}{2}\sum_{\langle ij\rangle \notin P_b \,\&\, \langle \mathrm{ij}\rangle \notin \mathrm{P_{SS}}} C_{nb}\theta\left(R_c - d_{ij,0}\right) \\ &\quad \times \frac{d_{ij,0}^2}{e^2}\left(1 - \frac{d_{ij,0}^e}{d_{ij}^e}\right)^2, \end{aligned} \qquad (2)$$

where E_b is the pseudobonded energy (P_b is the set of pseudobonded bead pairs, the bonded force constant $C_b = 10$), and E_{SS} is the nonbonded energy that maintains the secondary structures of α-helices and β-strands (P_{SS} is the set of C_α atom pairs that are either in an α-helix with a sequential offset ≤ 4, or in a β-strand with a sequential offset ≤ 3; the associated force constant $C_{SS} = 1$). E_{nb} is the remaining nonbonded energy with a new

parameter $e = 6$, corresponding to the Lennard-Jones potential—it has a minimum at $d_{ij,0}$, saturates as d_{ij} goes to infinity, and diverges as d_{ij} approaches zero (the nonbonded force constant $C_{nb} = (4/\ d_{ij,0})^2$, following Yang et al. (57)). Therefore, unlike the harmonic potential in Eq. 1, the mENM energy allows two nonbonded beads to move apart at a finite energy cost.

The mENM energy in Eq. 2 can be expanded near a given conformation X_* to the second order as

$$E_{\mathrm{mENM}}(X) \approx E_{\mathrm{mENM}}(X_*) + \delta X^{\mathrm{T}} G + \frac{1}{2}\delta X^{\mathrm{T}} H \delta X, \qquad (3)$$

where $\delta X = X - X_*$, $G = \nabla E_{\mathrm{mENM}}|_{\mathrm{X=X}^*}$ is the gradient of E_{mENM} at $X = X_*$, and H is the $3N \times 3N$ Hessian matrix comprised of the following 3×3 blocks,

$$H_{ij} = \begin{bmatrix} \left.\frac{\partial^2 E_{\mathrm{mENM}}}{\partial x_{\mathrm{i}} \partial x_{\mathrm{j}}}\right|_{X=X_*} & \left.\frac{\partial^2 E_{\mathrm{mENM}}}{\partial x_{\mathrm{i}} \partial y_{\mathrm{j}}}\right|_{X=X_*} & \left.\frac{\partial^2 E_{\mathrm{mENM}}}{\partial x_{\mathrm{i}} \partial z_{\mathrm{j}}}\right|_{X=X_*} \\ \left.\frac{\partial^2 E_{\mathrm{mENM}}}{\partial y_{\mathrm{i}} \partial x_{\mathrm{j}}}\right|_{X=X_*} & \left.\frac{\partial^2 E_{\mathrm{mENM}}}{\partial y_{\mathrm{i}} \partial y_{\mathrm{j}}}\right|_{X=X_*} & \left.\frac{\partial^2 E_{\mathrm{mENM}}}{\partial y_{\mathrm{i}} \partial z_{\mathrm{j}}}\right|_{X=X_*} \\ \left.\frac{\partial^2 E_{\mathrm{mENM}}}{\partial z_{\mathrm{i}} \partial x_{\mathrm{j}}}\right|_{X=X_*} & \left.\frac{\partial^2 E_{\mathrm{mENM}}}{\partial z_{\mathrm{i}} \partial y_{\mathrm{j}}}\right|_{X=X_*} & \left.\frac{\partial^2 E_{\mathrm{mENM}}}{\partial z_{\mathrm{i}} \partial z_{\mathrm{j}}}\right|_{X=X_*} \end{bmatrix}, \qquad (4)$$

where x_i, y_i, z_i (x_j, y_j, z_j) is the x, y, z component of the coordinate of bead i (j). The gradient and Hessian matrix will be used in the flexible fitting protocol based on the Newton-Raphson algorithm (see below).

Calculation of SAXS profile for a protein surrounded by a hydration shell

SAXS measures the x-ray scattering intensity of proteins in solution as a function of $s = 4\pi\sin(\theta)/\lambda$, where s is the magnitude of scattering vector ($s \in [0, 0.5\ \text{Å}^{-1}]$), 2θ is the scattering angle, and λ is the x-ray wavelength. The SAXS intensity profile $I(s)$ can be calculated from a protein structure using the atomic coordinates and form factors, after including the contribution from a thin shell of water molecules (termed the hydration shell) (54). To reduce computing cost, a coarse-grained, one-bead-per-residue representation was used to calculate $I(s)$ (58),

$$I(s) = \sum_{i=1}^{N+M}\sum_{j=1}^{N+M} F_i(s)F_j(s)\frac{\sin(sd_{ij})}{sd_{ij}}, \qquad (5)$$

where N is the number of residues, M is the number of water molecules, and F_i or F_j represents the coarse-grained form factor of a residue ($1 \le i,j \le N$) or a water molecule ($N < i,j \le N+M$), which have been calculated in Yang et al. (58) by approximating a group of spherical atoms in a residue (or a water molecule) as a "glob". The value d_{ij} is the distance between the C_α atom of residue i (or the oxygen atom of water molecule i) and the C_α atom of residue j (or the oxygen atom of water molecule j). The hydration shell is constructed by submerging a protein in a preequilibrated water box and keeping those water molecules whose oxygen atom is at a minimal distance of 3.5 Å ~ 6.5 Å from the C_α atoms (58).

To further reduce system size and computing cost, we have developed an implicit model of the hydration shell, which combines residue i and its nearby water glob (comprised of those water molecules whose nearest residue is residue i) into a composite glob (see Fig. 1). The coarse-grained form factor of a composite glob is calculated as

$$F'_i(s) = \sqrt{F_i^2(s) + w^2 f_i^2(s) + 2wF_i(s)f_i(s)\frac{\sin(sD_i)}{sD_i}}, \qquad (6)$$

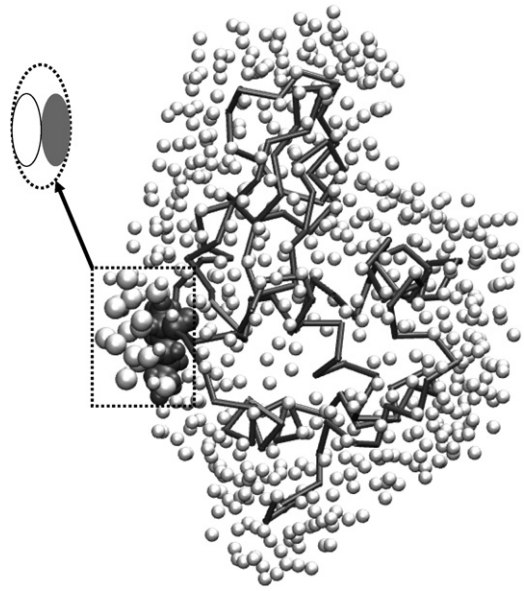

FIGURE 1 Implicit modeling of hydration shell: the protein structure of hen egg white lysozyme (*shaded trace*) is surrounded by an explicit shell of water molecules (*small open spheres*). The atoms of residue 1 are shown (*shaded spheres*). The water glob near residue 1 consists of water molecules (*large open spheres*) whose nearest residue is residue 1. (*Inset*) Schematic. Residue 1 (*shaded ellipse*) and its water glob (*open ellipse*) are combined into a composite glob.

where F_i is the coarse-grained form factor of residue i (58) and f_i is the coarse-grained form factor of the water glob near residue i (termed water glob i). D_i is the distance between the center of electron density distribution in residue i and the geometric center of water glob i. Note that $w = \delta\rho_w/\rho_w$ is the relative contrast of hydration shell, where $\rho_w = 0.334\ e/\text{Å}^3$ is the electron density of pure water, and $\delta\rho_w$ is the contrast of hydration shell that can be adjusted to optimize the fitting of experimental SAXS data (54). The value w usually varies from 0 to 10%.

The key to our implicit model is to use Eq. 6 without explicitly adding any water to the system. To this end, we must estimate two key parameters of water glob i—its geometric center position (denoted as P_i) and the average number of water molecules in water glob i (denoted as n_i). To determine these two parameters, we sample the region accessible to the water molecules of water glob i (denoted as R_i) that must satisfy the following two conditions: 1), the distance to the C_α atom of residue i (denoted as d_i) is within the range of [3.5 Å, 6.5 Å] (58); and 2), the distance to the C_α atom of any other residue is $>d_i$. We perform a grid-based exhaustive search for all grid points that satisfy the above two conditions. Then we calculate the volume of R_i (denoted as V_i) and its geometric center position P_i. The search is confined to a cubic box centered at the C_α atom of residue i with edge-length of 13 Å and the grid unit is 0.5 Å. After getting volume V_i, n_i is estimated as $n_i = 0.1 \times \rho_w V_i$. The above two parameters are updated iteratively after protein conformational changes during the fitting process.

Based on the implicit model of hydration shell, the SAXS profile $I'(s)$ can be calculated using a coarse-grained representation of N composite globs (each composed of a residue and its nearby water glob, see Fig. 1) as

$$I'(s) = \sum_{i=1}^{N}\sum_{j=1}^{N} F'_i(s)F'_j(s)\frac{\sin\left(sd'_{ij}\right)}{sd'_{ij}}, \qquad (7)$$

where $F'_i(s)$ and $F'_j(s)$ are calculated in Eq. 6, and d'_{ij} is the distance between the centers of electron density distribution in composite glob i and j.

Modified ENM-based fitting of SAXS profile

We start from two given inputs: an initial high-resolution structure, and a target SAXS profile measured experimentally or simulated from the target structure by the CRYSOL program (54). We optimize a pseudoenergy that is a linear combination of the following components—the nonbonded mENM energy based on the initial structure (weighted by λ), the SAXS-fitting score for the target SAXS profile (weighted by 1-λ, see below), the mENM energy terms maintaining pseudobonds and secondary structures (weighted by 1), and the collision energy (weighted by 1, see below). We keep the energy terms for pseudobonds and secondary structures with weight 1, because we want to preserve the integrity of local structures during flexible fitting.

The weight parameter $\lambda \in [0,1]$ controls the degree of data fitting: at $\lambda \sim 1$, the conformational search is restricted near the initial structure with weak fitting to the target SAXS profile; at $\lambda \sim 0$, the influence of initial structure is very weak with strong fitting to the target SAXS profile. To find an optimal λ between 0 and 1, we minimize the pseudoenergy progressively as λ decreases from 1 to 0 until the root mean-square deviation (RMSD) from the initial structure shows saturation (see below).

The total pseudoenergy function is defined as

$$E_{total} = \lambda E_{nb} + (1-\lambda)E_{SAXS} + E_b + E_{SS} + E_{col}, \quad (8)$$

where E_{nb} is the nonbonded energy based on the initial structure (see Eq. 2), E_b is the pseudobonded energy (see Eq. 2), E_{SS} is the nonbonded energy that maintains the secondary structures of α-helices and β-strands (see Eq. 2), and E_{col} is the collision energy between two nonbonded beads defined as

$$E_{col} = \frac{1}{2}\sum_{\langle ij\rangle \notin P_b \,\&\, \langle ij\rangle \notin P_{SS}} C_{col}\theta\left(R_{col}-d_{ij}\right)\left(d_{ij}-R_{col}\right)^2, \quad (9)$$

where the collision force constant is $C_{col} = 10$ and R_{col} is the minimal distance between nonbonded beads in the initial structure (53,56). The addition of E_{col} penalizes steric collisions between nonbonded beads that are within a distance of R_{col}.

E_{SAXS} is the SAXS-fitting score defined as

$$E_{SAXS} = f_{SAXS}\min_c\left\{\sum_l [cI_m(s_l) - I_t(s_l)]^2\right\}, \quad (10)$$

where the constant prefactor is $f_{SAXS} = 3 \times 10^7$ (a large f_{SAXS} is chosen to enable fast fitting that usually terminates after the first round of minimization with $\lambda = 0.5$), s_l is the scattering vector ranging from 0 to 0.5 Å^{-1}, I_m is the model SAXS profile calculated using Eq. 7, and I_t is the target SAXS profile measured experimentally or simulated by CRYSOL (54). Before the fitting, I_t is rescaled so that $I_t(0) = 1$. During the fitting, I_m is rescaled by a factor c to minimize the SAXS-fitting score in Eq. 10. A similar SAXS-fitting score (χ^2) was widely used in previous studies (for example, see Svergun et al. (54)).

For a series of gradually decreasing λ-values, the SAXS-fitted model is obtained by minimizing the pseudoenergy in Eq. 7. We use the Newton-Raphson algorithm to solve $\nabla E_{total}(\lambda, X_{\min}) = 0$ by using the following iterative procedure:

Step 1. Initialization: set $n = 0$, $\lambda_0 = 0.5$, and $X_0 = X_{\min,0} = X_i$, where X_i represents the bead coordinates of the initial structure.
Step 2. If $n > 0$, decrease λ_n to double the ratio $(1 - \lambda_n)/\lambda_n$.
Step 3. For conformation X_n, calculate the pseudoenergy E_n using Eq. 7, then set $X_{\min,n} = X_n$ if E_n reaches a new low.
Step 4. If E_n fails to be lowered after 20 iterations, stop minimization and go to Step 7.
Step 5. Displace X_n by the following incremental displacement,

$$\begin{aligned}\delta X_n = &-(\lambda_n H_{nb} + (1-\lambda_n)H_{SAXS} + H_b + H_{SS} \\ &+ H_{col})^{-1}(\lambda_n \nabla E_{nb} + (1-\lambda_n)\nabla E_{SAXS} + \nabla E_b \\ &+ \nabla E_{SS} + \nabla E_{col}),\end{aligned} \quad (11)$$

where H_{nb}, H_{SAXS}, H_b, H_{SS}, and H_{col} are the Hessian matrices calculated from E_{nb}, E_{SAXS}, E_b, E_{SS}, and E_{col}, respectively (see Eq. 4).

Step 6. Go to Step 3.
Step 7. Stop if the RMSD to the initial structure saturates (i.e., it increases by <5% from the previous minimization round).
Step 8. Set $n \leftarrow n+1$ and $X_n = X_{\min,n} = X_{\min,n-1}$, then go to Step 2.

To reduce accumulation of structural distortions, we limit the magnitude of each incremental displacement (see Eq. 11) $\leq$0.2 Å in RMSD. This is attained by adding εI (where I is an identity matrix and ε is an adjustable parameter) to the sum of Hessian matrices such that the linear-equation solution in Eq. 11 satisfies this condition. This idea is akin to the trusted region method used in a recent study of flexible fitting to PDF (41).

RESULTS

Test cases and modeling evaluation

We have developed what we consider a novel method that flexibly fits a high-resolution protein structure to a given SAXS profile using a modified form of ENM (i.e., mENM) and an implicit model of hydration shell (see Methods). This method is validated using five simulated test cases from previous studies (41,44): lysine/arginine/ornithine (LAO) binding protein (PDB code: 1lst, 2lao), adenylate kinase (PDB code: 1ake, 4ake), maltodextrin-binding protein (PDB code: 1omp, 1anf), lactoferrin (PDB code: 1lfh, 1lfg), and elongation factor 2 (PDB code: 1n0v, 1n0u). Each case consists of two high-resolution x-ray structures—a target structure used to simulate the target SAXS profile for fitting and an initial structure as the starting model for fitting. The target SAXS profile is simulated using CRYSOL, which is a widely used program that gives fast and accurate simulation of SAXS profiles from known protein structures by taking into account contributions from both the protein and its surrounding hydration shell (54). The five test cases sample a wide range of protein size (214 ~ 819 residues) and conformational change (3.8 Å ~ 14.4 Å in RMSD). We have also assessed our method using three more test cases with experimental SAXS data (see below).

Our method iteratively displaces the initial structure to optimize a pseudoenergy that combines several energy terms that maintain the initial tertiary structure, pseudobonds, secondary structures, and penalize steric collisions, together with an SAXS-fitting score (see Methods). We define the SAXS-fitting score as a sum of squared deviations between a model SAXS profile and a target one (see Eq. 10 in Methods), which is similar to the χ^2 score used in previous SAXS-fitting studies (54). An alternative score, which is based on the PDF derived from a SAXS profile,

was also used in previous studies (41,44). We choose to directly fit a SAXS profile instead of a PDF to avoid the additional uncertainty of constructing a PDF from a SAXS profile (59). Neither score alone is sufficient for model reconstruction because of the low information content of SAXS data—many different structures may have similarly low SAXS-fitting scores. Therefore, the SAXS-fitting score must be properly combined with other scores to achieve reliable structural modeling.

To unambiguously assess the quality of SAXS-fitted models, we compute the C_α-based RMSD between the final model and the target structure (denoted as $RMSD_t$). It is then compared with the C_α-based RMSD between the initial structure and the target structure (denoted as $RMSD_0$). A significant decrease from $RMSD_0$ to $RMSD_t$ indicates a successful flexible fitting. We also compute the C_α-based RMSD between the fitted models and the initial structure (denoted as $RMSD_i$) to assess the extent of conformational change during flexible fitting. The saturation of $RMSD_i$ is used as the criterion to terminate the fitting (see below). The results are summarized in Table 1.

Implicit model accurately accounts for hydration shell contribution to SAXS data

SAXS experiments measure the x-ray scattering intensity of a protein surrounded by a thin shell of solvent after subtracting the bulk solvent background (54). The presence of hydration shell is attributed to protein-solvent interactions, which result in higher solvent density near protein surface than bulk solvent (60). The explicit incorporation of a hydration shell was previously found to improve the fitting of experimental SAXS data (55,58).

The flexible fitting of SAXS data requires rapid calculation of SAXS profiles for many structural models. To enable efficient calculation of SAXS profiles, a coarse-grained model of protein-solvent system was proposed (58). In this simplified model, a 3 Å-thick shell of water molecules is added explicitly, and each water molecule (or residue) is approximated by a coarse-grained glob (58). Despite such simplification, the addition of explicit solvents has two disadvantages: first, it increases the system size significantly, particularly for small proteins; second, it introduces undesirable dependence of SAXS profiles on solvent degrees of freedom (i.e., coordinates of water molecules), so the calculated SAXS profiles need to be averaged over solvent degrees of freedom, which further increases computing cost. To circumvent these problems, we introduce what we believe to be a new implicit model of hydration shell that combines each residue (say, residue i) and its nearby water glob (comprised of water molecules whose nearest residue is residue i) into a composite glob (see Methods and Fig. 1). Then we calculate the coarse-grained form factor for each composite glob (see Eq. 6 of Methods), where the contribution of water glob is weighted by $w = \delta\rho_w/\rho_w$ (relative contrast of hydration shell). The water glob near residue i is characterized by two key parameters (the average number of water molecules n_i and the geometric center position P_i), which are not sensitive to solvent degrees of freedom, and can be estimated efficiently without explicitly adding solvents (see Methods).

Next, we will assess the accuracy of our implicit model in describing an explicitly constructed hydration shell. To this end, we have generated a hydration shell using the Fast-SAXS program (58), which submerges a protein in a water box and keeps those water molecules whose minimal distance to the C_α atoms is within the range [3.5 Å, 6.5 Å]. Then we regroup the water molecules into individual water globs (one water glob per residue)—the water glob near residue i consists of those water molecules whose nearest residue is residue i. Then we count the number of water molecules (denoted as n'_i) in each water glob, and calculate the geometric center position of each water glob (denoted as P'_i). Indeed, the agreement between (n'_i, P'_i) and (n_i, P_i) estimated by our implicit model is very good. The cross-correlation coefficient between n_i and n'_i of N residues is 0.97 ~ 0.98 for the five test cases. To compare P_i and P'_i, we calculate the overlap (or generalized cosine) between two $3N$-dimensional vectors, each comprised of N vectors pointing from the C_α atom i to P_i or P'_i. The overlap is 0.92 ~ 0.95 for the five test cases. Therefore, our implicit model indeed accurately describes water globs without explicitly adding waters to the system.

Next, we will evaluate the accuracy of our coarse-grained calculation of SAXS profiles based on our implicit model of hydration shell (see Eq. 7 of Methods). For $w = \delta\rho_w/\rho_w$ varying from 0 to 10%, we have calculated SAXS profiles from a crystal structure of hen egg white lysozyme (PDB code: 6lyz), and compared with the experimental SAXS data obtained from the CRYSOL program (54). The calculated and experimental SAXS profiles fit optimally when $w = \delta\rho_w/\rho_w = 3\%$ (χ ~ 0.46, see Fig. 2). The best fitting of experimental SAXS data by CRYSOL is also achieved when $w = \delta\rho_w/\rho_w = 3\%$ (χ ~ 0.45, see Fig. 2). The good agreement between our coarse-grained calculation and CRYSOL strongly supports its validity in fitting SAXS data. Throughout this study, we take $w = \delta\rho_w/\rho_w = 3\%$ by default to simulate target SAXS profiles with CRYSOL and running flexible fitting with mENM.

Consideration of implicit hydration shell improves flexible fitting

In a previous study, the consideration of hydration shell was found to improve the NMA-based flexible fitting of PDF (44). To evaluate the importance of implicit hydration shell to flexible fitting, we have rerun the fitting with $w = \delta\rho_w/\rho_w = 0$ (i.e., without hydration shell), and then compared with the result of $w = \delta\rho_w/\rho_w = 3\%$. Indeed, the

TABLE 1 Results of SAXS-based flexible fitting

Target PDB	Initial PDB	No. of residues	$RMSD_0$ (Å)	w	Selected*, min-end†, min-all‡ $RMSD_t$ (Å)	Final SAXS-fitting score
1ake	4ake	214	7.1	0	2.9 2.9 2.9	2.27e-01
				0.01	2.8 2.8 2.6	3.39e-01
				0.02	2.1 2.1 1.8	1.12e-01
				0.03	2.0 2.0 2.0	1.90e-01
				0.04	3.8 3.8 3.0	1.49e+00
				0.05	3.3 3.3 3.1	1.06e+00
				0.06	3.5 3.5 3.4	1.74e+00
				0.07	4.1 4.1 3.8	2.72e+00
				0.08	6.5 5.7 4.7	4.15e-01
				0.09	6.3 6.3 4.7	5.79e+00
				0.1	7.2 6.4 4.8	4.28e+00
1lst	2lao	238	4.7	0	1.9 1.8 1.8	7.92e-01
				0.01	1.3 1.3 1.2	2.65e-01
				0.02	1.1 1.1 0.7	1.34e-01
				0.03	1.2 1.2 0.9	4.43e-01
				0.04	5.5 3.7 3.2	4.50e-01
				0.05	2.8 2.8 2.7	2.13e+00
				0.06	4.8 4.3 4.2	7.63e-01
				0.07	7.2 7.2 4.2	3.86e+00
				0.08	4.6 4.6 4.2	2.57e+01
				0.09	7.4 7.4 4.2	3.12e+00
				0.1	7.6 5.7 4.3	2.59e+00
1omp	1anf	380	3.8	0	2.2 2.0 2.0	1.00e+00
				0.01	1.4 1.4 1.4	1.01e+00
				0.02	1.2 1.1 1.1	6.06e-01
				0.03	1.1 0.8 0.8	4.24e-01
				0.04	1.3 1.1 1.1	1.09e+00
				0.05	1.5 1.5 1.5	3.13e+00
				0.06	2.1 2.1 1.7	6.22e+00
				0.07	3.9 3.9 3.1	1.01e+01
				0.08	4.0 4.0 3.4	1.36e+01
				0.09	4.3 4.0 3.5	9.68e+00
				0.1	4.5 4.5 3.5	2.24e+01
1lfh	1lfg	691	6.4	0	5.9 4.7 4.7	1.07e+01
				0.01	1.6 1.6 1.6	6.33e+00
				0.02	1.6 1.6 1.6	2.64e+00
				0.03	2.1 2.1 1.8	1.29e+00
				0.04	3.2 3.2 2.4	1.73e+00
				0.05	3.1 3.1 2.6	3.03e+00
				0.06	4.3 4.2 4.2	9.59e+00
				0.07	4.9 4.8 4.8	1.98e+00
				0.08	5.3 5.1 5.1	6.86e+00
				0.09	4.7 4.7 4.6	3.11e+01
				0.1	6.4 6.4 5.9	4.49e+01
1n0v	1n0u	819	14.4	0	10.2 10.2 10.1	1.78e+01
				0.01	11.4 11.4 10.5	1.31e+01
				0.02	10.6 10.6 10.0	7.96e+00
				0.03	3.0 3.0 3.0	4.38e+00
				0.04	3.9 3.9 3.9	1.83e+00
				0.05	4.0 4.0 3.9	1.95e+00
				0.06	6.7 6.7 6.7	1.27e+00
				0.07	8.7 8.7 8.2	3.04e+00
				0.08	8.8 8.2 8.2	1.63e+00
				0.09	8.6 8.4 8.4	5.35e+00
				0.1	6.7 6.7 6.6	8.81e+00
2ko1 model 14	2ko1 model 16	176	3.2	0	2.1 2.1 2.1	1.57e+00
				0.01	2.0 2.0 2.0	1.29e+00
				0.02	2.0 2.0 1.9	1.08e+00
				0.03	1.9 1.9 1.9	7.96e-01

Table 1. Continued

Target PDB	Initial PDB	No. of residues	$RMSD_0$ (Å)	w	Selected*, min-end†, min-all‡ $RMSD_t$ (Å)	Final SAXS-fitting score
				0.04	1.8 1.8 1.8	7.21e-01
				0.05	1.7 1.7 1.6	6.65e-01
				0.06	1.7 1.7 1.6	6.29e-01
				0.07	1.6 1.6 1.6	5.35e-01
				0.08	1.6 1.6 1.5	5.42e-01
				0.09	1.6 1.6 1.6	5.35e-01
				0.1	1.6 1.6 1.5	5.44e-01
1crc		104	6.0	0	3.3 3.2 3.2	5.20e-03
				0.01	3.5 3.2 3.2	1.83e-03
				0.02	3.4 3.2 3.2	1.90e-02
				0.03	3.7 3.2 3.2	5.63e-03
				0.04	4.2 3.2 3.2	3.61e-03
				0.05	4.5 3.3 3.2	4.08e-03
				0.06	4.4 3.2 3.2	1.10e-02
				0.07	4.8 3.3 3.2	1.16e-02
				0.08	5.3 3.2 3.2	5.72e-03
				0.09	5.3 3.2 3.2	1.12e-02
				0.1	5.1 3.3 3.2	3.34e-02

See Table S1 in the Supporting Material for the results of radius of gyration.
*$RMSD_t$ (relative to target structure) for the model selected based on the termination criterion.
†Minimal $RMSD_t$ (relative to target structure) of the end models generated by minimization at various λ-values.
‡Minimal $RMSD_t$ (relative to target structure) of all the models sampled during minimization at various λ-values.

consideration of hydration shell results in better final models with lower $RMSD_t$ in all five cases (see Table 1). The greatest improvement is attained for the two biggest proteins: the final $RMSD_t$ is lowered from 5.9 Å to 2.1 Å for lactoferrin, and from 10.2 Å to 3 Å for elongation factor 2. We have also tried flexible fitting with other values of $w = \delta\rho_w/\rho_w$ between 0 and 10%. The result worsens as $w = \delta\rho_w/\rho_w$ deviates further from 3% (see Table 1). Therefore, it is important to choose $w = \delta\rho_w/\rho_w$ correctly when flexibly fitting SAXS data. In what follows, we will further discuss how to perform flexible fitting without knowing the correct $w = \delta\rho_w/\rho_w$.

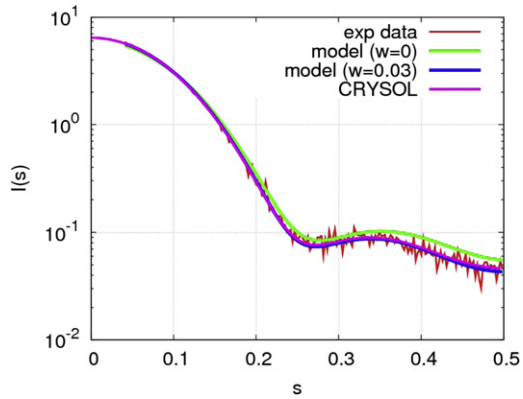

FIGURE 2 Fitting of experimental SAXS profile of hen egg white lysozyme. (*Red*) Experimental data; (*purple*) CRYSOL-fitted SAXS profile; (*blue* and *green*) SAXS profiles generated by our coarse-grained model with implicit hydration shell ($w = \delta\rho_w/\rho_w = 3\%$) and without hydration shell ($w = \delta\rho_w/\rho_w = 0$), respectively. The magnitude of scattering vector s is in units of $Å^{-1}$.

Flexible fitting without knowing the relative contrast of hydration shell

In our test of flexible fitting so far, it is assumed that we know $w = \delta\rho_w/\rho_w = 3\%$, which is used to simulate the target SAXS profiles by CRYSOL. In practice, $w = \delta\rho_w/\rho_w$ varies from protein to protein, and it is unknown for experimental SAXS data.

How do we perform flexible fitting without knowing $w = \delta\rho_w/\rho_w$?

To meet this challenge, we have repeated the flexible fitting for a range of $w = \delta\rho_w/\rho_w$ from 0 to 10% (with increment of 1%), and collected 11 final models together with their final SAXS-fitting scores (see Table 1). Intuitively, a mismatch in $w = \delta\rho_w/\rho_w$ would cause poor fitting of the target SAXS data. Indeed, for all except one case, the lowest final SAXS-fitting score is attained at $w = \delta\rho_w/\rho_w = 2\%$ or 3% (see Table 1). Therefore, one can select the model with the lowest final SAXS-fitting score, which probably corresponds to a roughly correct $w = \delta\rho_w/\rho_w$. The same strategy will be used to test our method on experimental SAXS data (see below). We note that the fitted w value may reflect both contributions from hydration shell and the fluctuation of surface residues, which may lead to large variations in w value for different proteins (see below).

Termination criterion for SAXS-based flexible fitting based on saturation of $RMSD_t$

As shown by previous studies (41,44), the fitting of low-resolution structural data like SAXS is susceptible to

overfitting unless it can be terminated before the onset of overfitting. We use the LAO-binding protein as an example to illustrate the overfitting problem. During the first round of minimization at $\lambda = 0.5$, $RMSD_t$ initially decreases from 4.7 Å to 0.9 Å and then increases to 1.2 Å; it continues to increase during the second round of minimization at $\lambda = 0.33$ (see Fig. 3 *c*). Meanwhile, the SAXS-fitting score continues to decrease during both rounds of minimization, even after $RMSD_t$ reaches the minimum (see Fig. 3 *b*). Therefore, the continuation of fitting for $\lambda < 0.5$ will result in overfitting. We note that the minimal $RMSD_t$ is reached approximately when $RMSD_i$ starts to saturate (i.e., $RMSD_i$ does not increase by >5% from previous minimization, see Fig. 3 *c*). Therefore, we use $RMSD_i$ saturation as the criterion to terminate SAXS-based flexible fitting. Indeed, by using this criterion, the fitting for LAO-binding protein is terminated at $\lambda = 0.5$. For the other four cases, the use of this criterion often results in the selection of final models with $RMSD_t$ close to the minimal $RMSD_t$ (see Table 1). Therefore, this termination criterion seems to work well in controlling overfitting. An alternative termination criterion based on the relative change in PDF-fitting score was used in a previous study (41).

Comparison with an alternative method

We have compared the performance of our method with an alternative NMA-based PDF-fitting method (44) using the same five test cases. In all cases (adenylate kinase, LAO-binding protein, maltodextrin-binding protein, elongation factor 2, lactoferrin), our method (with final $RMSD_t$ = 2 Å, 1.2 Å, 1.1 Å, 3 Å, 2.1 Å, see Table 1) performs better than the alternative method (with $RMSD_t$ = 2.6 Å, 2 Å, 2 Å, 5.5 Å, >6.4 Å) (44). Notably, for lactoferrin, the alternative method failed to generate a fitted model with $RMSD_t < RMSD_0$, whereas our method has obtained a fitted model with $RMSD_t$ ~ 2.1 Å.

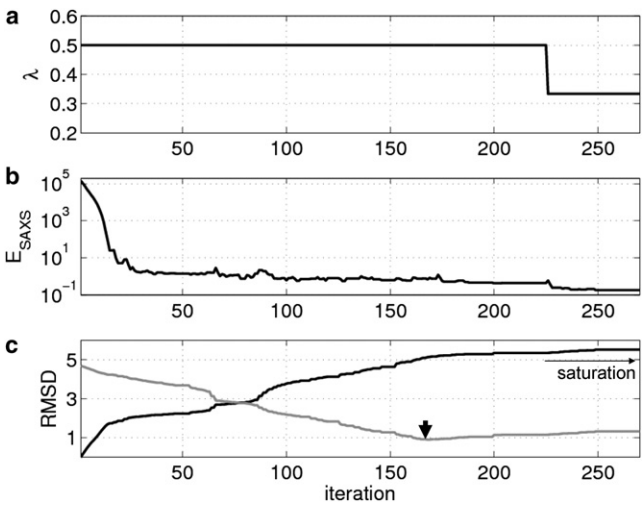

FIGURE 3 Result of flexible fitting for LAO-binding protein: (*a*) weight parameter λ; (*b*) SAXS-fitting score E_{SAXS}; and (*c*) $RMSD_t$ (*gray*) and $RMSD_i$ (*black*) are shown as a function of iteration step. (*Arrow*) The minimal $RMSD_t$.

Flexible fitting of experimental SAXS data

Thanks to experimental noise and other errors, it is much more challenging to perform flexible fitting to experimental SAXS data than to simulated data. To meet this great challenge, we have tested our method on the following two cases:

In the first case, we have used the SAXS data collected for the Act domain of GTP pyrophosphokinase. It is a homo-dimer with 176 residues (see Fig. 4 *f*), and its structure has been solved by NMR (PDB code: 2ko1, with 20 models). Among all NMR models, the greatest RMSD is found between models No. 14 and 16 ($RMSD_0$ ~ 3.2 Å), and the former fits better to the SAXS data, so it is chosen as the target structure, and the latter is chosen as the initial structure. Then we run mENM to fit the experimental SAXS data starting from the initial structure. We have used $w = \delta\rho_w/\rho_w = 7\%$, which yields the lowest SAXS-fitting score (see Table 1). We note that the relatively high w value is likely due to the high flexibility of this protein, which effectively increases its volume in a way similar to the effect of a hydration shell. Indeed, in the NMR structures of this protein, the C-terminus regions are very floppy. The final fitted model has $RMSD_t$ of 1.6 Å. Structural comparison indicates that the model aligns well with the target structure, especially in the C-terminus region that differs most between the initial and target structure (see Fig. 4 *f*). The fitting moderately reduces the difference in SAXS profiles (see Fig. S1 *f* in the Supporting Material).

The second case is cytochrome *c* with 105 residues. We have used the SAXS data collected by Grishaev et al. (55). A crystal structure (PDB code: 1crc) is used as the target structure. To generate an initial model, we have used the Flexweb server (http://flexweb.asu.edu/software/first/first_online_newsim/) to generate 500 conformations starting from 1crc (with energy cutoff of −3 kcal/mol). We keep the final model of the simulation with $RMSD_0$ ~ 6 Å. Then we run mENM to fit the experimental SAXS data starting from the initial model. We have used $w = \delta\rho_w/\rho_w = 1\%$, which yields the lowest SAXS-fitting score (see Table 1). The fitting results in a final fitted model with $RMSD_t$ of 3.5 Å. Structural comparison indicates that the model aligns reasonably well with the target structure (see Fig. 4 *g*). The fitting is found to greatly reduce the difference in SAXS profiles (see Fig. S1 *g*).

To further test our method, we have applied it to a third experimental test case of myosin V that involves a large conformational change between two biochemical states (see Supporting Material).

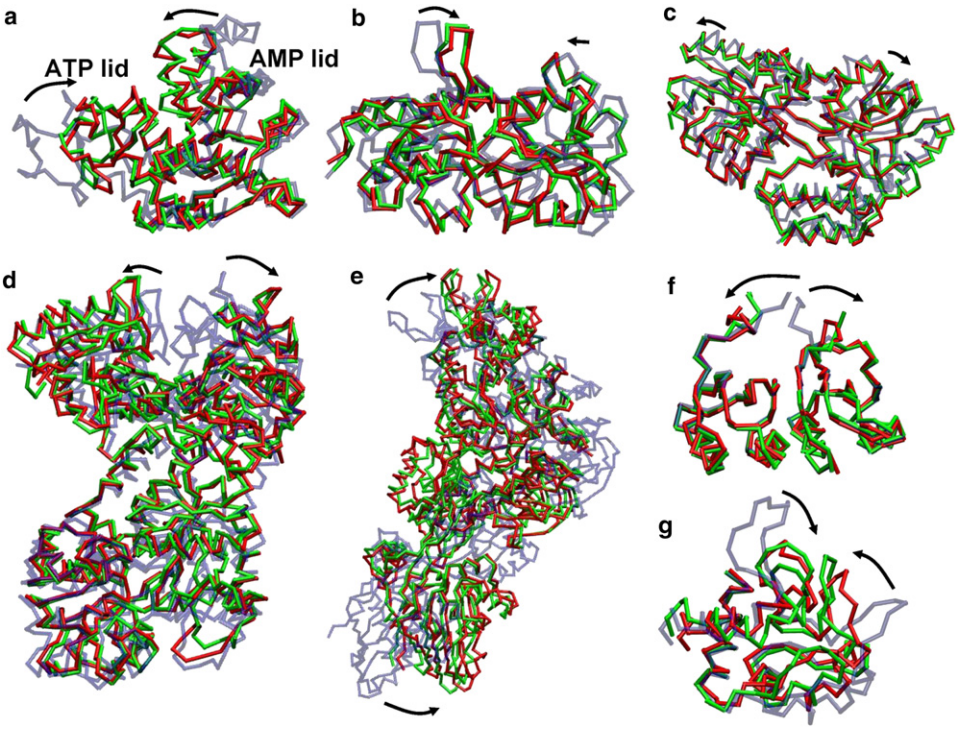

FIGURE 4 Structural comparison among initial structure, target structure, and fitted model for: (*a*) adenylate kinase, (*b*) LAO-binding protein, (*c*) maltodextrin-binding protein, (*d*) lactoferrin, (*e*) elongation factor 2, (*f*) the Act domain of GTP pyrophosphokinase, and (*g*) cytochrome *c*. (*Backbone traces*) Initial structure (*transparent-shaded*), target structure (*dark-shaded*), and fitted model (*light-shaded*). (*Arrows*) Domain motions from the initial structures to the target structures. One can see that the fitted models are very similar to the target structures (i.e., the *dark-shaded* and *light-shaded traces* follow each other closely), which supports the finding of small $RMSD_t$ between the fitted models and target structures (see Table 1). For a comparison of SAXS profiles based on the initial structures, fitted models, and target structures, see Fig. S1 of the Supporting Material.

DISCUSSION AND CONCLUSION

We have developed, to our knowledge, a new computational method to flexibly fit a given protein structure to a SAXS profile using a coarse-grained model. Our method uses a modified form of ENM that allows large-scale conformational changes while maintaining pseudobonds and secondary structures and avoiding residue collisions. In addition, we model the hydration shell implicitly by combining each residue and its nearby water glob into a composite glob, which allows accurate and efficient calculation of SAXS profiles.

Compared with the explicit modeling of hydration shell (44), our implicit model has the following advantages: 1), It does not increase the system size and is therefore computationally cheaper. 2), It avoids possible interference with conformational changes caused by adding waters into the interdomain cleft, which may prevent domain closing. 3), It allows the implicit hydration shell to expand or shrink (with constant thickness) as a protein undergoes large conformational changes, which is not possible if a fixed number of explicit waters is added.

In principle, the flexible fitting of a high-resolution structure to low-resolution data like SAXS profiles is prone to overfitting, because the many degrees of freedom involved in fitting are generally insufficiently constrained by the given data. In our method, we take the following measures to control overfitting: 1), The total number of degrees of freedom is reduced by using a coarse-grained model and maintaining the pseudobonds and the secondary structures. 2), The flexible fitting is timely terminated when $RMSD_i$ (relative to the initial structure) starts to saturate. (This termination criterion seems to work well for our test cases (see Table 1).) We have also visually inspected the final models and did not find serious structural distortion indicative of overfitting (see Fig. 4).

The quality of SAXS-based flexible fitting ultimately depends on the level of degeneracy of SAXS data (i.e., the abundance of alternative conformations with low SAXS-fitting score and their accessibility from the initial structure). The higher the degeneracy, the harder it is to construct the correct model from SAXS data. To improve the sampling of conformations compatible with the given SAXS profile, one can start flexible fitting from multiple initial structures (for example, different x-ray/NMR structures or those obtained from extensive MD simulations). The selection of correct model from the ensemble of fitted models may require additional information and processing. Indeed, given the limitation of SAXS data, it is highly desirable to integrate them with other structural information and modeling techniques to achieve optimal performance (61).

Our SAXS-based flexible fitting method complements alternative methods that optimize the fitting of structural data (SAXS or cryo-EM) using Monte Carlo sampling (11,12), MD simulation (48–50,62), and coarse-grained simulation (52). Our method is based on the optimization of a pseudoenergy that is computationally fast but susceptible to trapping at local minima, whereas the alternative methods like Monte Carlo and MD are computationally more expensive but also capable of more extensive sampling.

Our flexible fitting method assumes the target SAXS data are based on a single target conformation that is distinct from the initial structure. In reality, experimentally measured SAXS data are a weighted average of multiple SAXS profiles based on an ensemble of conformations. To explore such an ensemble-effect, we have considered a situation where the SAXS data to be fitted are a weighted average of two SAXS profiles—one based on a known initial structure and the other based on an unknown target structure. We have found that our flexible fitting method works well as long as the unknown target structure dominates the SAXS data or the weight of the target structure is known (see Supporting Material). Future study will be needed to treat the ensemble aspect of SAXS data for flexible fitting in a comprehensive way.

SUPPORTING MATERIAL

Additional information with two figures, two tables, and reference (63) is available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(11)01313-0.

We thank Drs. Edward Snell and Ad Bax for sharing the SAXS data.

We also acknowledge funding support from the American Heart Association (grant No. 0835292N) and the National Science Foundation (grant No. 0952736).

REFERENCES

1. Saibil, H. R. 2000. Conformational changes studied by cryo-electron microscopy. *Nat. Struct. Biol.* 7:711–714.
2. Koch, M. H., P. Vachette, and D. I. Svergun. 2003. Small-angle scattering: a view on the properties, structures and structural changes of biological macromolecules in solution. *Q. Rev. Biophys.* 36:147–227.
3. Putnam, C. D., M. Hammel, …, J. A. Tainer. 2007. X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Q. Rev. Biophys.* 40:191–285.
4. Mertens, H. D. T., and D. I. Svergun. 2010. Structural characterization of proteins and complexes using small-angle x-ray solution scattering. *J. Struct. Biol.* 172:128–141.
5. Lipfert, J., and S. Doniach. 2007. Small-angle x-ray scattering from RNA, proteins, and protein complexes. *Annu. Rev. Biophys. Biomol. Struct.* 36:307–327.
6. Tsuruta, H., and T. C. Irving. 2008. Experimental approaches for solution x-ray scattering and fiber diffraction. *Curr. Opin. Struct. Biol.* 18:601–608.
7. Jacques, D. A., and J. Trewhella. 2010. Small-angle scattering for structural biology—expanding the frontier while avoiding the pitfalls. *Protein Sci.* 19:642–657.
8. Zheng, W. J., and S. Doniach. 2002. Protein structure prediction constrained by solution x-ray scattering data and structural homology identification. *J. Mol. Biol.* 316:173–187.
9. Zheng, W. J., and S. Doniach. 2005. Fold recognition aided by constraints from small angle x-ray scattering data. *Protein Eng. Des. Sel.* 18:209–219.
10. Wu, Y. H., X. Tian, …, J. Ma. 2005. Folding of small helical proteins assisted by small-angle x-ray scattering profiles. *Structure.* 13:1587–1597.
11. Förster, F., B. Webb, …, A. Sali. 2008. Integration of small-angle x-ray scattering data into structural modeling of proteins and their assemblies. *J. Mol. Biol.* 382:1089–1106.
12. Petoukhov, M. V., and D. I. Svergun. 2005. Global rigid body modeling of macromolecular complexes against small-angle scattering data. *Biophys. J.* 89:1237–1250.
13. Schneidman-Duhovny, D., M. Hammel, and A. Sali. 2010. Macromolecular docking restrained by a small angle x-ray scattering profile. *J. Struct. Biol.* 173:461–471.
14. Chacón, P., J. F. Díaz, …, J. M. Andreu. 2000. Reconstruction of protein form with x-ray solution scattering and a genetic algorithm. *J. Mol. Biol.* 299:1289–1302.
15. Walther, D., F. E. Cohen, and S. Doniach. 2000. Reconstruction of low-resolution three-dimensional density maps from one-dimensional small-angle x-ray solution scattering data for biomolecules. *J. Appl. Cryst.* 33:350–363.
16. Svergun, D. I. 1999. Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys. J.* 77:2879–2886.
17. Svergun, D. I., M. V. Petoukhov, and M. H. Koch. 2001. Determination of domain structure of proteins from x-ray solution scattering. *Biophys. J.* 80:2946–2953.
18. Svergun, D. I., and M. H. J. Koch. 2002. Advances in structure analysis using small-angle scattering in solution. *Curr. Opin. Struct. Biol.* 12:654–660.
19. Wriggers, W., and P. Chacon. 2001. Using Situs for the registration of protein structures with low-resolution bead models from x-ray solution scattering. *J. Appl. Cryst.* 34:773–776.
20. Volkmann, N., D. Hanein, …, S. Lowey. 2000. Evidence for cleft closure in actomyosin upon ADP release. *Nat. Struct. Biol.* 7:1147–1155.
21. Wendt, T., D. Taylor, …, K. Taylor. 2001. Three-dimensional image reconstruction of dephosphorylated smooth muscle heavy meromyosin reveals asymmetry in the interaction between myosin heads and placement of subfragment 2. *Proc. Natl. Acad. Sci. USA.* 98:4361–4366.
22. Rawat, U. B. S., A. V. Zavialov, …, J. Frank. 2003. A cryo-electron microscopic study of ribosome-bound termination factor RF2. *Nature.* 421:87–90.
23. Shiozawa, K., P. V. Konarev, …, D. I. Svergun. 2009. Solution structure of human Pex5·Pex14·PTS1 protein complexes obtained by small angle x-ray scattering. *J. Biol. Chem.* 284:25334–25342.
24. Arai, R., W. Wriggers, …, T. Fujisawa. 2004. Conformations of variably linked chimeric proteins evaluated by synchrotron x-ray small-angle scattering. *Proteins.* 57:829–838.
25. 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.
26. Bernadó, P., Y. Pérez, …, M. Pons. 2008. Structural characterization of the active and inactive states of Src kinase in solution by small-angle x-ray scattering. *J. Mol. Biol.* 376:492–505.
27. Bernadó, P., Y. Pérez, …, M. Pons. 2009. Structural characterization of unphosphorylated STAT5a oligomerization equilibrium in solution by small-angle x-ray scattering. *Protein Sci.* 18:716–726.
28. Tozzini, V. 2005. Coarse-grained models for proteins. *Curr. Opin. Struct. Biol.* 15:144–150.
29. Hinsen, K. 1998. Analysis of domain motions by approximate normal mode calculations. *Proteins.* 33:417–429.
30. 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.
31. Tama, F., and Y. H. Sanejouand. 2001. Conformational change of proteins arising from normal mode calculations. *Protein Eng.* 14:1–6.
32. Tirion, M. M. 1996. Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. *Phys. Rev. Lett.* 77:1905–1908.

33. Delarue, M., and P. Dumas. 2004. On the use of low-frequency normal modes to enforce collective movements in refining macromolecular structural models. *Proc. Natl. Acad. Sci. USA.* 101:6957–6962.

34. Tama, F., O. Miyashita, and C. L. Brooks, III. 2004. Flexible multi-scale fitting of atomic structures into low-resolution electron density maps with elastic network normal mode analysis. *J. Mol. Biol.* 337:985–999.

35. Tama, F., O. Miyashita, and C. L. Brooks, III. 2004. NMFF: flexible high-resolution annotation of low-resolution experimental data from cryo-EM maps using normal mode analysis. *J. Struct. Biol.* 147: 315–326.

36. Hinsen, K., N. Reuter, …, J. J. Lacapère. 2005. Normal mode-based fitting of atomic structure into electron density maps: application to sarcoplasmic reticulum Ca-ATPase. *Biophys. J.* 88:818–827.

37. Suhre, K., J. Navaza, and Y. H. Sanejouand. 2006. NORMA: a tool for flexible fitting of high-resolution protein structures into low-resolution electron-microscopy-derived density maps. *Acta Crystallogr. D Biol. Crystallogr.* 62:1098–1100.

38. Mitra, K., C. Schaffitzel, …, J. Frank. 2005. Structure of the *E. coli* protein-conducting channel bound to a translating ribosome. *Nature.* 438:318–324.

39. Falke, S., F. Tama, …, M. T. Fisher. 2005. The 13 Ångstroms structure of a chaperonin GroEL-protein substrate complex by cryo-electron microscopy. *J. Mol. Biol.* 348:219–230.

40. Tama, F., G. Ren, …, A. K. Mitra. 2006. Model of the toxic complex of anthrax: responsive conformational changes in both the lethal factor and the protective antigen heptamer. *Protein Sci.* 15:2190–2200.

41. Gorba, C., O. Miyashita, and F. Tama. 2008. Normal-mode flexible fitting of high-resolution structure of biological molecules toward one-dimensional low-resolution data. *Biophys. J.* 94:1589–1599.

42. Zheng, W., and B. R. Brooks. 2005. Normal-modes-based prediction of protein conformational changes guided by distance constraints. *Biophys. J.* 88:3109–3117.

43. Zheng, W., and B. R. Brooks. 2006. Modeling protein conformational changes by iterative fitting of distance constraints using reoriented normal modes. *Biophys. J.* 90:4327–4336.

44. Gorba, C., and F. Tama. 2010. Normal mode flexible fitting of high-resolution structures of biological molecules toward SAXS data. *Bioinform. Biol. Insights.* 4:43–54.

45. Miyashita, O., C. Gorba, and F. Tama. 2011. Structure modeling from small angle x-ray scattering data with elastic network normal mode analysis. *J. Struct. Biol.* 173:451–460.

46. Petrone, P., and V. S. Pande. 2006. Can conformational change be described by only a few normal modes? *Biophys. J.* 90:1583–1593.

47. Noda, K., M. Nakamura, …, T. Yasunaga. 2006. Atomic model construction of protein complexes from electron micrographs and visualization of their 3D structure using a virtual reality system. *J. Plasma Phys.* 72:1037–1040.

48. Trabuco, L. G., E. Villa, …, K. Schulten. 2008. Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics. *Structure.* 16:673–683.

49. Trabuco, L. G., E. Villa, …, K. Schulten. 2009. Molecular dynamics flexible fitting: a practical guide to combine cryo-electron microscopy and x-ray crystallography. *Methods.* 49:174–180.

50. Orzechowski, M., and F. Tama. 2008. Flexible fitting of high-resolution x-ray structures into cryoelectron microscopy maps using biased molecular dynamics simulations. *Biophys. J.* 95:5692–5705.

51. Caulfield, T. R., and S. C. Harvey. 2007. Conformational fitting of atomic models to cryogenic-electron microscopy maps using Maxwell's demon molecular dynamics. 2007 Biophysical Society Meeting Abstracts.

52. Yang, S. C., L. Blachowicz, …, B. Roux. 2010. Multidomain assembled states of Hck tyrosine kinase in solution. *Proc. Natl. Acad. Sci. USA.* 107:15757–15762.

53. Zheng, W. J. 2011. Accurate flexible fitting of high-resolution protein structures into cryo-electron microscopy maps using coarse-grained pseudo-energy minimization. *Biophys. J.* 100:478–488.

54. 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.

55. Grishaev, A., L. Guo, …, A. Bax. 2010. Improved fitting of solution x-ray scattering data to macromolecular structures and structural ensembles by explicit water modeling. *J. Am. Chem. Soc.* 132:15484–15486.

56. Tekpinar, M., and W. J. Zheng. 2010. Predicting order of conformational changes during protein conformational transitions using an interpolated elastic network model. *Proteins.* 78:2469–2481.

57. Yang, L., G. Song, and R. L. Jernigan. 2009. Protein elastic network models and the ranges of cooperativity. *Proc. Natl. Acad. Sci. USA.* 106:12347–12352.

58. Yang, S., S. Park, …, B. Roux. 2009. A rapid coarse residue-based computational method for x-ray solution scattering characterization of protein folds and multiple conformational states of large protein complexes. *Biophys. J.* 96:4449–4463.

59. Svergun, D. I. 1992. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Cryst.* 25:495–503.

60. Svergun, D. I., S. Richard, …, G. Zaccai. 1998. Protein hydration in solution: experimental observation by x-ray and neutron scattering. *Proc. Natl. Acad. Sci. USA.* 95:2267–2272.

61. Alber, F., F. Förster, …, A. Sali. 2008. Integrating diverse data for structure determination of macromolecular assemblies. *Annu. Rev. Biochem.* 77:443–477.

62. Grubisic, I., M. N. Shokhirev, …, F. Tama. 2010. Biased coarse-grained molecular dynamics simulation approach for flexible fitting of x-ray structure into cryo electron microscopy maps. *J. Struct. Biol.* 169:95–105.

63. Sugimoto, Y., O. Sato, …, K. Wakabayashi. 2009. Reverse conformational changes of the light chain-binding domain of myosin V and VI processive motor heads during and after hydrolysis of ATP by small-angle x-ray solution scattering. *J. Mol. Biol.* 392:420–435.