

ORIGINAL RESEARCH

OPEN ACCESS

Full open access to this and thousands of other papers at <http://www.la-press.com>.

Normal Mode Flexible Fitting of High-Resolution Structures of Biological Molecules Toward SAXS Data

Christian Gorba and Florence Tama

Department of Chemistry and Biochemistry, The University of Arizona, 1041 E. Lowell Street, Tucson, AZ, 85721.
Corresponding author email: ftama@u.arizona.edu

Abstract: We present a method to reconstruct a three-dimensional protein structure from an atomic pair distribution function derived from the scattering intensity profile from SAXS data by flexibly fitting known x-ray structures. This method uses a linear combination of low-frequency normal modes from an elastic network description of the molecule in an iterative manner to deform the structure to conform optimally to the target pair distribution function derived from SAXS data. For computational efficiency, the protein and water molecules included in the protein first hydration shell are coarse-grained. In this paper, we demonstrate the validity of our coarse-graining approach to study SAXS data. Illustrative results of our flexible fitting studies on simulated SAXS data from five different proteins are presented.

Keywords: coarse-grained model, conformational change, modeling, pair distribution function, protein, trusted region method

Bioinformatics and Biology Insights 2010:4 43–54

This article is available from <http://www.la-press.com>.

© the author(s), publisher and licensee Libertas Academica Ltd.

This is an open access article. Unrestricted non-commercial use is permitted provided the original work is properly cited.



Introduction

Dynamical conformational changes of biological molecules are critical for their function. X-ray crystallography has been used to study such conformational changes by providing the structures of distinct conformational states. However, it is often difficult to observe such conformational changes, as it is challenging to trap molecules in a certain conformational state that would crystallize. This phenomenon is particularly true for large biological molecules. Therefore, alternative techniques such as Cryo-Electron Microscopy (cryo-EM), Small Angle X-ray Scattering (SAXS) and Fluorescence Resonance Energy Transfer (FRET) experiments are often used to identify conformational changes of molecules and are becoming a primary tool to obtain information about the dynamics of biological systems.^{1–8}

However, those techniques can only provide low-resolution data and atomic details of the structure cannot be directly obtained. A common approach to infer atomic detail about conformational changes is to construct an atomic model consistent with the low-resolution data, using known X-ray structures.^{9–17}

In this paper, we specifically aim to develop a quantitative tool to extract conformational changes of biological systems from SAXS data. SAXS is a well-established and interesting fundamental tool to study the structure of molecules.^{18,19} Its strength arises from its simpleness, i.e. it does not require special sample preparation and is relatively easy to perform. SAXS can supply a rich perspective on the structure and dynamics of proteins and macromolecular assemblies under a range of solution conditions, in particular physiological conditions that are not accessible to other structural methods. However, SAXS data only provides the overall size of the molecule, thus reconstruction of 3D structures is inherently degenerate though distinct conformational states of the same molecule can be identified if their difference is sufficiently large.

A substantial amount of work has been performed to reconstruct the shape of biological molecules from SAXS data using bead models.^{14,17,20–22} This shape information can be further complemented by X-ray data, where X-ray structures are rigidly docked into the envelope in a similar way as is actually done with cryo-EM data.²³ While current programs provide

overall information about the shape of the biological molecules, they cannot describe conformational changes of the molecules with atomic details. Here we are specifically interested in these conformational changes. For a more explicit interpretation of the conformational transitions observed by SAXS, we propose its integration with already known X-ray structures.

Such fitting of an X-ray structure into low-resolution data of a different conformational state is often done by domain segmentation followed by fitting each domain of the system as an independent rigid body block. These procedures have been useful for the understanding, at a near atomic level of detail, of several conformational changes in major biological systems as obtained from cryo-EM data or SAXS data.^{4,24–29} However, such procedures are subjective because while each domain may move as a rigid unit, large conformational changes often involve tightly coupled motions between domains of the biological system. When domains are fitted as independent rigid body blocks, those correlated motions are not reproduced. Thus, despite the success of these previous studies, the development of more advanced quantitative techniques with data reproducibility is necessary.

We have recently introduced a new algorithm to reconstruct an atomic level structure from a given atomic pair distribution function.³⁰ This new algorithm takes advantage of an already known X-ray structure to predict a different conformation embedded in the atomic pair distribution function. Normal mode theory was used to construct a model by deforming the original X-ray structure to ensure that the new conformation is energetically accessible. The atomic pair distribution function is 1 dimensional (1D) data and it has far less information than 3D structural data obtained from techniques such as cryo-EM. Nevertheless, our protocol is successful in predicting structure from these low-resolution data.

In this paper, we present the extension of this methodology for fitting into experimental data derived from SAXS. Since SAXS experiments on biological molecules are performed in solution, the first layer of ordered water molecules around the protein is also captured in the data.^{31–33} Our approach takes into account the contribution from such a solvation



layer into the fitting. In addition, we employ a coarse grained model for computational efficiency where the protein is represented by its C α atoms and water molecules are treated as dummy atoms. We present illustrative results of our studies on five different proteins with large conformational changes occur and discuss the accuracy of the method as well as the limitations.

Theory

Intensity profile and pair distribution function

SAXS measures the intensity of X-ray scattered at small angles. Information derived from SAXS data is embedded into a 1D profile of the scattering intensity $I(s)$ measured in the experiments. Here, $s = 4\pi\sin(\theta)/\lambda$ is the momentum transfer as a function of the scattering angle 2θ and λ is the X-ray wavelength. The theoretical $I(s)$ can be calculated from a 3D structure of a given macromolecule using coordinates and structure factors of atoms.³¹ Most studies to reconstruct low-resolution models from SAXS experiments use the scattering intensity as the target data. One can also relate the measured scattering intensity to the distribution of electron pair distances within the molecule, which can be directly calculated through a Fourier transform of the scattering intensity curve:¹⁸

$$p(r) = \frac{r}{2\pi^2} \int_0^\infty I(q)q \sin(qr) dq$$

Such $P(r)$ represents a real space approximation of the data. Since it can be readily related to atomic structure, it is more useful when trying to incorporate SAXS and X-ray data. The pair distribution function $P(r)$ derived from $I(s)$ has been used to discuss conformational changes of biological molecules observed by SAXS experiments.^{28,34–37} We will also use $P(r)$ for our scoring function to characterize conformational changes.

Modeling water molecules

We have shown that a simple interatomic pair distribution function, $P(r)$, calculated directly from a protein structure is sufficient to predict conformational changes of biological molecules.³⁰ However,

experimental SAXS profiles include the effect of water layers surrounding the protein, while it is not included when $P(r)$ is directly calculated from a protein structure. Therefore, as the next step toward using experimental data, a water shell around the full-atom initial structure is modeled using the program Solvate from Helmut Grubmüller's group. In order to reproduce experimental data, only the first solvent shell of organized water molecules is considered.

Coarse-graining model

In some previous works, several coarse-grained models have been introduced to speed up calculations of theoretical intensity scattering patterns.^{14,21,32,38} One practical approach is to approximate a given molecular structure by a finite number of homogenous spheres of equal radii and scattering density. This approximation is valid within the resolution limit of SAXS.³⁹ Therefore, in the present study only C α atoms are taken into account. Such a coarse-grained model corresponds to about $\sim 1/8$ of the original number of non-hydrogen atoms, because on average each residue consists of 8 non-hydrogen atoms. Consequently, we randomly remove 7/8 of the water molecules that surround the protein so that the density of atoms of the protein and solvent shell are comparable.

Correspondence between coarse-grained model and experimental data

The $P(r)$ derived from the experimental data reflects the all-atom structure of the protein surrounded by a layer of water. Therefore, in order to perform fitting using coarse-grained models, a normalization is necessary, i.e. the $P(r)$ from the coarse-grained model needs to be scaled to match the experimental $P(r)$.

The integral of $P(r)$ is equal to $N(N-1)/2$, where N is the total number of atoms. Theoretically, while the shape of the $P(r)$ is different between two protein conformations, their integrals should remain the same, assuming that the solvation is almost similar for both conformations. Therefore the target experimental data can be scaled so that its integral matches the integral of the initial coarse-grained model.

Pair distribution function

$P(r)$ is approximated by evaluating all interatomic distances in the molecule, including surrounding



water molecules that are represented as dummy atoms. The atomic pair distribution function $P(r)$, denoted g_k in the equations, is represented as a histogram:

$$g_k = \sum_{i < j}^N \int_{r_{k-1}}^{r_k} \delta(r - r_{ij}) dr$$

which is the number of atomic pairs whose distances are between r_{k-1} and r_k . The data is a discrete profile. With such a definition, the derivative of g_k as a function of atomic coordinates cannot be calculated, which is disadvantageous for the optimization process. To make g_k a continuous function of atomic coordinates, a Gaussian kernel is used. Replacing the delta function by a Gaussian, g_k can be rewritten as follows:

$$g_k = \frac{1}{\sqrt{2\pi\sigma^2}} \sum_{i < j}^N \int_{r_{k-1}}^{r_k} e^{-\frac{(r-r_{ij})^2}{2\sigma^2}} dr \quad (1)$$

where r_{ij} is the distance between atoms i and j , and σ is a parameter for the kernel width. g_k is now analytically differentiable as a function of atom coordinates, which is suitable for calculating its derivatives. The parameter σ was calibrated to minimize the error between the exact distance distribution function and the one obtained using eq. 1 while maintaining a smooth profile of the distance distribution function. We have used a value of 2.0 Å throughout our studies.³⁰

Scoring function

Several successful scoring functions have been used to optimize the shape of a molecule toward SAXS data using the scattering intensity.^{14,20,27,38,40} For computational efficiency, the scattering intensity can be replaced by the inter atomic distance distribution function.^{21,41–43}

Here, the scoring function, f , which is used to measure the agreement between the modeled structure and the $P(r)$ profile is the deviation between the two profiles including a weight function:³⁰

$$f = \sum_k w_k (g_k^{\text{model}} - g_k^{\text{tar}})^2$$

where k is the index number of a bin in the profile, g_k^{tar} is the target data and g_k^{model} is the $P(r)$ calculated

from a modeled structure. The weight function, w_k , is introduced to reduce the sensitivity of small deviations at the large distribution regions.³⁰

Methods

Normal mode analysis (NMA) using an elastic network model

NMA has proven to be useful for studies of collective motions of biological systems that may occur on time scales not easily accessible to conventional MD simulations. NMA consists of the decomposition of motions into vibrational modes. Each normal-mode vector represents a collective motion of the molecule. The modes with the lower frequencies represent preferential global motions.⁴⁴

It has been shown that iterative deformations using small steps along a combination of normal modes is effective to provide transitions between two conformational states.⁴⁵ Such a property has been exploited for refinement of low-resolution data.⁴⁶ Thus, in this approach, we use the normal-mode coordinates associated with low-frequency normal modes as the optimization variables in place of Cartesian atomic coordinates. By limiting the deformation of the molecule to the low-frequency normal modes, we ensure that generated structural models are energetically accessible. In addition, by limiting the number of variables, i.e. normal-mode coordinates, to be optimized, we minimize the possibility of over-fitting. We have previously shown that using the 10 lowest frequency normal modes, conformational changes could be described within 2–3 Å RMSD.³⁰

Normal modes were calculated using a simplified representation of the potential energy function, which was first introduced by Tirion.⁴⁷ In this representation, the biological system is described as a three dimensional elastic network based on the equilibrium distribution of atoms. Particles or atoms are used to identify the junctions of the network. These junctions are representative of the mass distribution of the system and are connected together via a simple harmonic restoring force. Studies have shown that C α carbons are sufficient to reproduce mechanical properties of the molecules^{44,48} and are successful in reproducing large and collective motions of macromolecules.⁴⁹

In this study, the C α atoms of the proteins and oxygen atoms of the water molecules were used to identify the junctions of the network. A cut-off



of 8 Å was used. The strength of the potential is a phenomenological constant assumed to be the same for all interacting pairs.

Trusted region method (TRM)

The goal of the refinement process is to maximize the scoring function by following the direction of a few low-frequency normal modes. The optimization method used here was originally proposed to find transition states for chemical reactions and relies on the use of the Newton Raphson approach within a trusted region.⁵⁰

In the TRM, one defines a radius within which the quadratic model equation is “trusted” to be a reasonable representation of the function (the simple Newton Raphson method trusts it for the entire space). Then, TRM tries to find the constrained minimum within the trusted region (Δ) using the analytical formula,

$$\max \left\{ f = \frac{1}{2} \mathbf{q}^T \mathbf{H} \mathbf{q} + \mathbf{q}^T \mathbf{g} \text{ such that } \|\mathbf{q}\| < \Delta \right\}$$

where $\mathbf{H} = \partial^2 f / \partial q_i \partial q_m$ is the Hessian and $\mathbf{g} = \{\partial f / \partial q_i\}$ the gradient. The amplitudes $\mathbf{q} = \{q_i\}$ that maximize f within the trusted region, Δ , are obtained using the Lagrange multiplier method. If the original point is sufficiently close to a minimum, then often the minimum lies within the circle of the trusted region and convergence is quite rapid, whereas near the start of a search, the constrained minimum will most likely lie on the boundary. Once the constrained minimum is found, the structure is deformed and a new iteration begins. The process continues until convergence of f is observed. This algorithm is conceptually well suited to the iterative normal mode method, as in both we allow only limited movement. Previously, we have found that a step size of 0.1 Å and displacement along 2 modes was most appropriate in term of conserving structural quality and computational efficiency.³⁰

Simulating SAXS target experimental data

To test our algorithm, we need to use “experimental” data as the fitting target but we also need to know the structure that corresponds to the data. In order to mimic experimental data we use the program suite Crysol/Gnom.³¹ Crysol takes into account details of both the

protein itself and the water molecules surrounding it and generate very accurate scattering intensity profiles from known structures when compared to experimental data.⁵¹ In order to obtain the corresponding $P(r)$, the theoretical SAXS intensity profile of the full-atom target structure obtained from Crysol is taken as input for Gnom which translates it into $P(r)$ which will be used as target data for refinement.

Summary of procedures

The inputs for the refinement protocol are an experimental $P(r)$ of a protein target structure and a high-resolution structure (x-ray or NMR) that the user wants to flexibly fit to the $P(r)$ data. A layer of water molecules is added to the structure, which is further coarse-grained to its $C\alpha$ atoms and 1/8 of the water molecules. The NMA is first performed on this initial coarse-grained structure and the TRM is used to find the optimal structural deformations. Details of the algorithm can be found in Gorba et al.³⁰

Results and Discussion

Test proteins

We have tested our procedure on simulated SAXS data obtained from the program package Crysol/Gnom. We choose proteins (Table 1) that undergo a conformational change and where high-resolution structures of the two states are available.

The purpose of the method presented here is to fit the X-ray structure to raw experimental (in this case simulated) SAXS data of a different conformation of the same molecule. To examine the fitness between the atomic structure and the $P(r)$ from experimental data during the optimization, we examine the *scoring function* between the structure and the target SAXS data. In the following we also examine the Root Mean Square Deviation (RMSD) between the deformed structure and the structure from which the target $P(r)$ was created to monitor the progress of the fitting. Ideally, when the SF reaches a maximum, the RMSD should reach a minimum. However, two different structures with different RMSD could have the same scoring function because SAXS data is 1D and detailed information of the atomic coordinates is lost.

Fit without water inclusion

In order to show that it is critical to take into account water molecules, refinements were first run using

**Table 1.** Protein studied and refinement results with and without first solvent shell around proteins.

Protein	Initial PDB	Target PDB	Number of residues	Initial RMSD (Å)	Final RMSD (Å)	
					Without water	With water
Adenylate kinase	1AKE	4AKE	214	7.1	4.7	2.6
LAO binding protein	2LAO	1LST	238	4.7	\	2.0
Maltodextrin binding protein	1ANF	10MP	370	3.8	\	2.0
Lactoferrin	1LFG	1LFH	691	6.4	\	\
Elongation factor 2	1NOV	1NOU	819	14.4	6.2	5.5

Note: \ indicates that no decrease in the RMSD was observed.

only a coarse-grained model consisting of the C α atoms of the protein, while simulated experimental data obtained from Crysol had the effects of solvent included. The scattering intensity profile was converted to a P(r), which was used as our target data. In most cases, the decrease in the RMSD is relatively small or close to none (Table 1). Compared to our ideal case, in which no influence of water is considered in the target data,³⁰ we can see that the prediction of the conformational changes is limited, and the influence of water molecules needs to be considered.

Solvation protocol

In the following we demonstrate that our approach to include the effects of the solvent shell on P(r) for biological molecules is adequate for the fitting.

The all-atom structure of adenylate kinase (4ake) is first surrounded by several layers of water molecules. Subsequently, only water molecules that are within a certain cut-off distance of any protein atom are kept to represent the first layer of ordered water molecules that is captured by SAXS data (Fig. 1a). After testing several cut-off values, we found that a 3.5 Å cut-off produces a P(r) in good agreement with a simulated experimental one obtained from crysol/gnom³¹ (Fig. 1b).

Our refinement uses a coarse-grained approach. Figure 1c shows adenylate kinase (4ake), which is our initial structure that needs to be fitted to a target-simulated experimental data set generated from a different conformation (1ake). The first layer of water molecules needs to be coarse-grained, as only the C α atoms of the protein are considered. To attain a level of coarse-graining comparable to the protein, only 1/8 of the total number of water molecules are kept for the fitting. Our procedure randomly picks up water molecules, and as is evident from Figure 1c, such water molecules uniformly surround the protein. From this

coarse-grained model, i.e. protein C α model and 1/8 of the first layer of water molecule, the interatomic pair distribution function of our initial structure, P_{ini}(r), is calculated using Eq 1. The target data, P_{exp}(r) for adenylate kinase (1ake), as derived from the simulated experimental scattering intensity profile, is shown in Figure 1d. For the purpose of comparison, the P(r) from the initial structure (4ake) as obtained from Crysol/Gnom is also shown.

The magnitude of P_{ini}(r) and P_{exp}(r) is different since one is obtained from a coarse-grained model and the other one is derived from an all-atom model. Our experimental P_{exp}(r), as simulated using Crysol/Gnom, is scaled so that its integral matches the one observed for P_{ini}(r) (Fig. 1e). The two P(r) are now on the same scale and can be used for estimating the scoring function as well as its derivative. As seen from Figure 1d and 1e, our scaling approach conserves the difference observed between the two P(r) as obtained from Crysol/Gnom. Therefore, this allows us to flexibly fit structures using a coarse-grained model which is computationally more efficient than using an all atom model.

Refinement process

We illustrate some of our results using the LAO binding protein, as we have previously shown that the simulation of its conformational change by iterative NMA was most successful.³⁰ The refinement was performed using parameters that were previously found optimal, i.e. two normal modes for structure displacement with a radius of 0.1 Å for the TRM.³⁰

Figure 2a shows the behavior of the scoring function as a function of iteration step. The scoring function converges slowly toward 0. However, after the RMSD value decreases to a minimum, it starts to increase again (Fig. 2b) even though the scoring

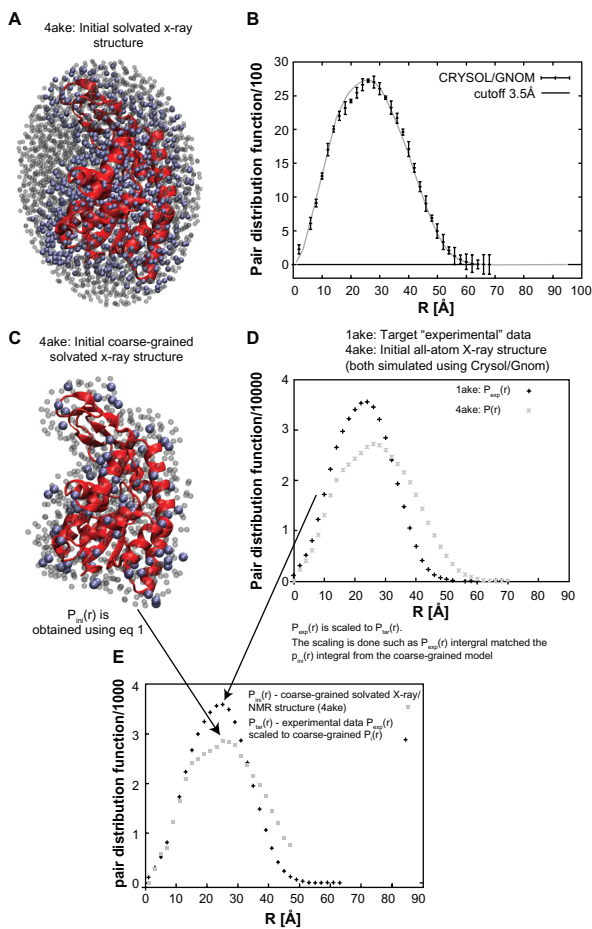


Figure 1. Coarse-graining process and scaling for Adenylate Kinase. (a) Adenylate Kinase surrounded with several layers of water molecules (colored in gray). Only water molecules shown in light blue are considered to estimate the Intensity Scattering profile and consequently derived the $P(r)$. (b) $P(r)$ for the Adenylate Kinase as obtained with eq. 1 or with the program suite Crysol/Gnom. Good agreement between the $P(r)$ is observed. (c) Coarse-graining: adenylate kinase is shown as ribbon, only its C α atoms are considered (not shown) and from the first shell of water molecules surrounding the protein (transparent gray), only 1/8 of these molecules are kept (light blue). (d) Simulated experimental data using all atom structures: $P(r)$ for the open and close forms of the Adenylate Kinase as derived from the Crysol/Gnom program suite. (e) $P(r)$ for the coarse-grained model of the protein/solvent system as calculated using eq. 1, $P_{ini}(r)$. The target $P_{exp}(r)$ is prepared by scaling $P_{exp}(r)$ so that its integral matches the integral of the coarse-grained model $P_{ini}(r)$. Note the good agreement between the scaled $P(r)$ and the $P(r)$ derived from all-atom structure is similar.

function does not improve much. This indicates that the data is over fitted. Due to the low-resolution nature of the SAXS data, the scoring function may continue to increase while additional displacements do not necessarily decrease the RMSD, i.e. a continued increase of the scoring function does not always improve the quality of the model. Therefore, one needs to define a more stringent parameter than the convergence of the scoring function in order to decide whether the refinement has found a reasonable

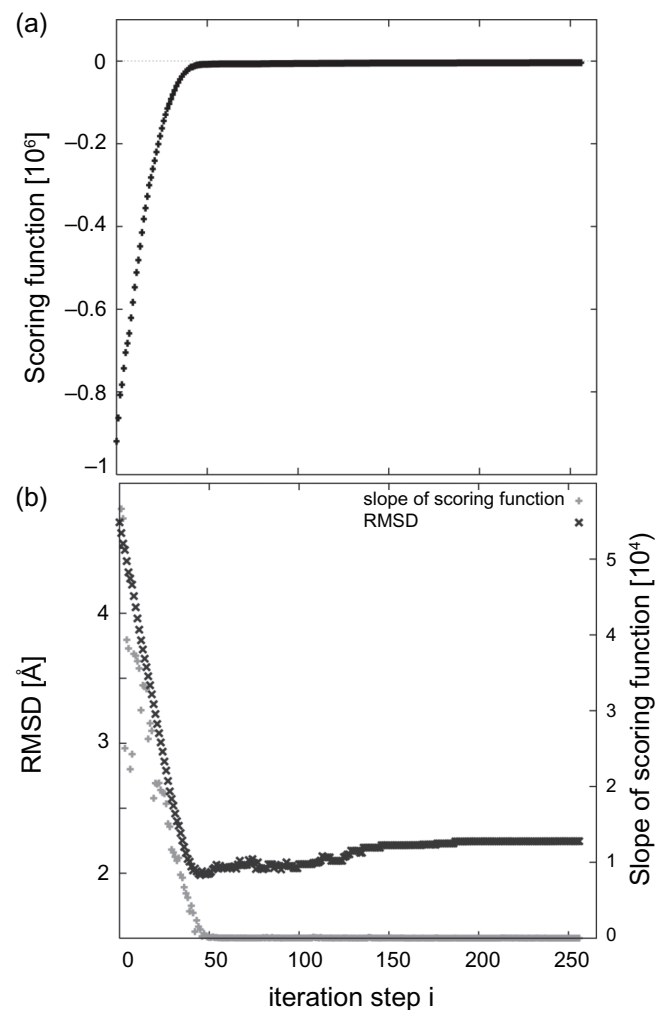


Figure 2. (a) Progress of fitting process measured by fitness scoring function for the LAO binding protein. (b) Progress of RMSD and relative change in the scoring function over the fitting process. The progress of the RMSD is shown in black and the relative change of scoring functions in gray. The relative change reaches zero when the RMSD is close to the minimum.

model, or if the additional displacements observed are only artifacts.

Instead of using the convergence of the scoring function as a stop criterion, we monitor the slope or the relative change of the scoring function between two consecutive steps of the refinement procedure. The minimum of the RMSD approximately coincides with the value of the slope approaching zero (Fig. 2b). Thus we use the slope as the stop criterion for refinement, which can be easily applied to real-case refinement in the future.

For the LAO binding protein, the RMSD decreases from 4.7 Å to 2 Å for the best fit. Figure 3a shows the $P(r)$ for the initial, target and best predicted structure. Although the difference in $P(r)$ between initial and

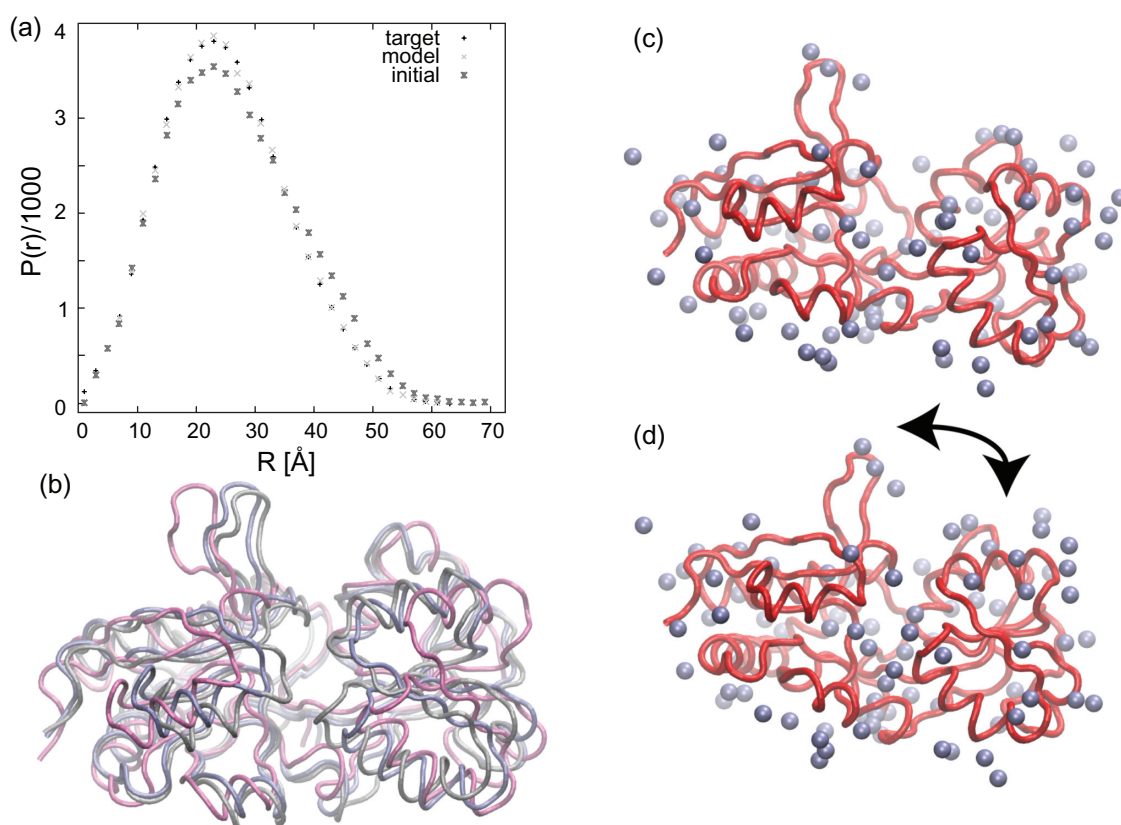


Figure 3. Conformational change of the LAO binding protein obtained from the normal mode flexible fitting. (a) The $P(r)$ of initial (■), predicted (gray x) and target (black +) structures are shown. The $P(r)$ of the predicted structure is in close agreement with the target PDF and the structures are also very similar. (b) The initial structure of LAO binding protein (pink), the modeled structure predicted from the fitting algorithm (iceblue) and the target structure from which $P_{tar}(r)$ was created (silver). (c) and (d) represent the placement of the coarse-grained water molecules in the initial and modeled structures. Arrows indicate the major conformational change of the system.

target is rather small, i.e. information for refinement is limited, the agreement between the target $P(r)$ and the one for the best predicted structure is good. Superposition of the three structures also reveals good agreement between the modeled structure and the target (Fig. 3b). In particular the opening/closing of the two domains is well captured.

Figure 3c and 3d show the behavior of surrounding water molecules during the fitting. The water molecules followed the protein deformations and remain in the vicinity of the protein in the modeled structure. With this method, conformational dynamics of water molecules is not represented, i.e. the water molecules only follow the protein conformational change but cannot rearrange around the protein as such motion cannot be described by low-frequency normal modes. However, we believe that dynamics of water molecules would not significantly alter the overall shape of the protein and its first hydration layer. Such structured water molecules only render the protein surface and volume larger. Therefore, as

an approximation, the conformational dynamics of water can be ignored.

Tests for others proteins

Using the relative change of the scoring function as the convergence parameter, additional fitting processes were performed on several proteins (see Table 1). Except for Lactoferrin, the refinement shows a significant decrease in RMSD. Even when large initial conformational differences are present, such as in the case of the Elongation Factor 2 (~15 Å), the RMSD can be decreased considerably.

Figure 4 shows two different conformations of the Elongation Factor along with the modeled structure (Fig. 4c). In the two known X-ray structures, domains 2 and 5 have different positions. In the modeled structure, domains 2 and 5 are close to each other, which is comparable to their position in the target structure. Although our method only predicts the structure within ~5 Å RMSD (see Table 1), most of the conformational change is accurately described by

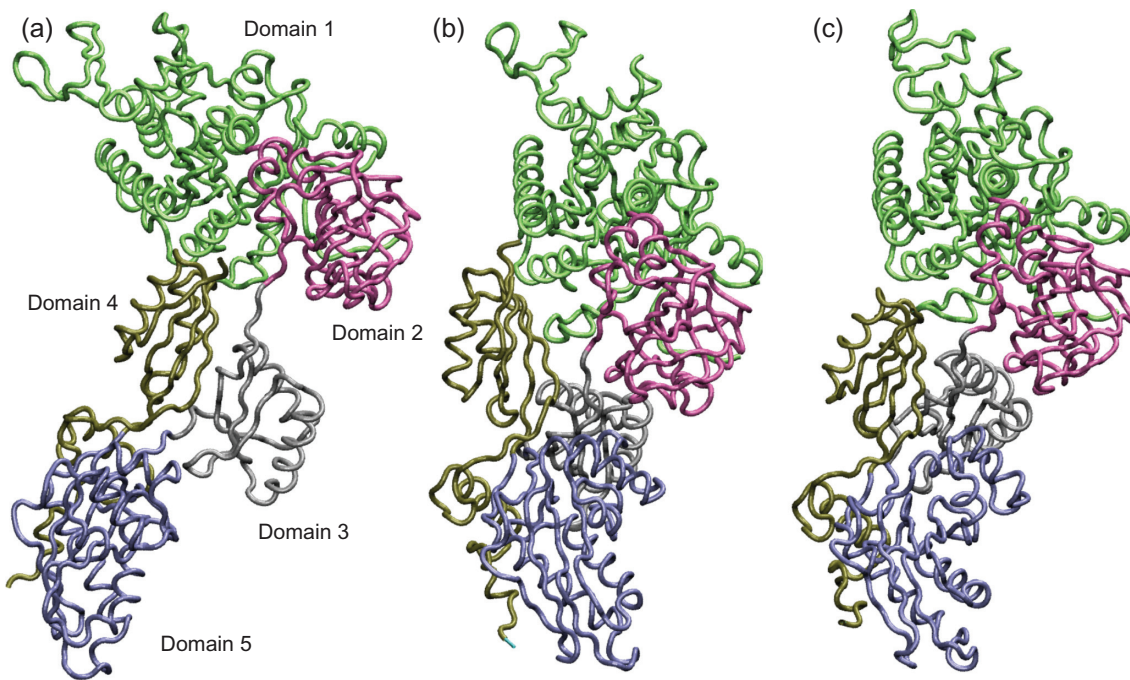


Figure 4. (a) The initial structure of EF2, (b) the target structure from which the simulated experimental data is created and (c) the modeled structure predicted from the fitting algorithm. Domain 2 has a different placement in target and modeled structure compared to the initial structure.

correctly predicting major rearrangements between domains.

Overall, the performance of the prediction depends on the difference between initial and target $P(r)$. Lactoferrin, is a case for which the fitting was unsuccessful. For Lactoferrin, the maxima of both $P(r)$ s are approximately the same and only slight differences in the shape of the $P(r)$ are noticeable (see Figure 5). Consequently, for Lactoferrin as opposed to Adenylate Kinase, for which the initial RMSD is also ~ 7 Å, the information provided by the target $P(r)$ as compared to the initial $P(r)$ is so small that the refinement fails. Normal modes that describe the conformational change are not selected, instead, random modes are chosen leading to a faulty model. In the case of the Maltodextrin binding protein, the difference between initial and target structures measured by RMSD is rather small. In such a case, the difference in the $P(r)$ between the two conformations is also rather minimal, and therefore one would expect that only limited information can be extracted from the refinement. Nonetheless, as illustrated by Figure 6 the overall domain motion of the protein is well captured by our approach. The direction of the displacement obtained is accurate; however, there is not sufficient information from the low-resolution data to fully

reproduce the full conformational change. It indicates that even though our approach may fail to construct a model close to the final structure, insights on the directionality of conformational changes can be reliably obtained.

Advantages over direct fitting methods

The interpretation of conformational changes from SAXS data has been usually performed by domain segmentation, which is followed by fitting each domain as an independent rigid body block in order to match the $P(r)$ or scattering intensity profile.^{4,27–29} Such procedures are subjective because while each domain may move as a rigid unit, large conformational changes often involve tightly coupled motions between domains of the biological system. For example a hinge bending motion is defined as a relative dislocation of one domain with respect to another. These properties are poorly reproduced when each domain is independently fitted as a rigid block, which can lead to misinterpretation of the data.⁵² On the other hand, it has been shown that normal modes can reproduce conformational changes of molecules that have been observed experimentally. With such quantitative methods, concerted motions are naturally taken into account, which reduces error in interpreting low resolution data.

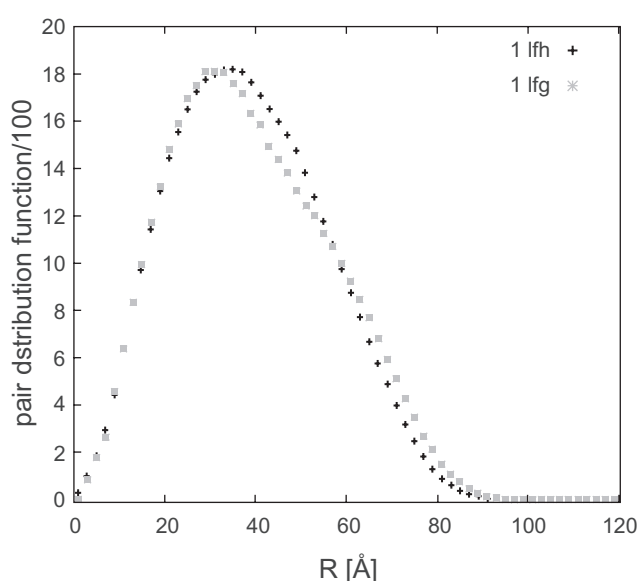


Figure 5. $P(r)$ of the initial (1lfg) and target (1lfh) for Lactoferrin are shown. For this system, only small difference is observed between the initial and target $P(r)$ even though the conformational difference is large, 7 Å RMSD.

Conclusions

We have presented a new method to predict conformational changes of biological molecules as observed from SAXS experiments. This method takes advantage of an already known X-ray structure to predict a different conformation embedded in the pair distribution function derived from SAXS experiments. Normal mode theory is used to construct a model by deforming an original X-ray structure to ensure that

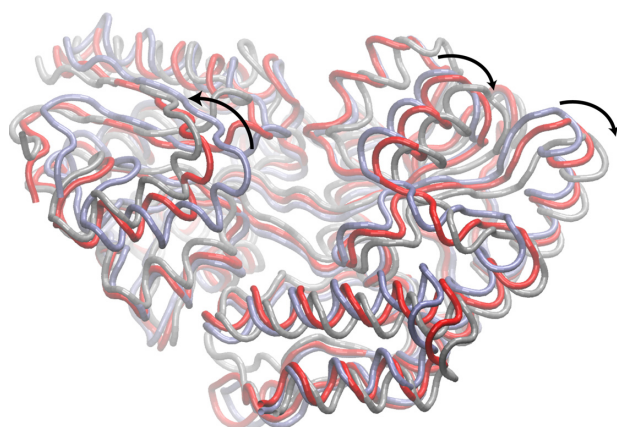


Figure 6. The initial structure of the maltodextrin binding protein (blue), the structure from which the simulated experimental data is created (silver), and the modeled structure predicted from the fitting algorithm (red) are shown. For this system, since the difference between the initial and target PDFs is small, the modeled structure does not exactly agree with the target structure; however, trends of the conformational change are predicted.

the new conformation is energetically accessible. As a consequence, concerted motions occurring during conformational changes are naturally taken into account to provide a more accurate description than rigid-body approaches. Water molecules surrounding proteins are included in the model to accurately represent the SAXS experimental data. The protein and water molecules are coarse-grained to speed up the calculations. Our approach allows prediction of the directions of conformational changes and the construction of structures within 3 Å of the target data when sufficiently large difference in the $P(r)$ are observed. In terms of accuracy, our method is close to what can be observed for structures obtained from homology modeling. In cases where two conformations cannot be well distinguished by their SAXS profiles, characteristics of the conformational change can still be extracted from our approach.

Disclosures

This manuscript has been read and approved by all authors. This paper is unique and not under consideration by any other publication and has not been published elsewhere. The authors and peer reviewers report no conflicts of interest. The authors confirm that they have permission to reproduce any copyrighted material.

Acknowledgements

We thank Dr. Osamu Miyashita for helpful discussions about this project. Financial support from the National Science Foundation grant 0744732 (Molecular Cellular and Biosciences) is greatly appreciated.

References

1. Saibil HR. Conformational changes studied by cryo-electron microscopy. *Nat Struct Biol.* 2000;7:711–4.
2. Heller WT, Vigil D, Brown S, Blumenthal DK, Taylor SS, Trewella JC. Subunits Binding to the Protein Kinase A RI{alpha} Dimer Induce a Large Conformational Change. *J Biol Chem.* 2004;279:19084–90.
3. Priddy TS, Macdonald BA, Heller WT, Nadeau OW, Trewella J, Carlson GM. Ca²⁺-induced structural changes in phosphorylase kinase detected by small-angle X-ray scattering. *New Journal.* 2005;14:1039–48.
4. Vestergaard B, Sanyal S, Roessle M, et al. The SAXS solution structure of RF1 differs from its crystal structure and is similar to its ribosome bound cryo-EM structure. *Mol Cell.* 2005;20:929–38.
5. Ermolenko DN, Majumdar ZK, Hickerson RP, Spiegel PC, Clegg RM, Noller HF. Observation of Intersubunit Movement of the Ribosome in Solution Using FRET. *J Mol Biol.* 2007;370:530–40.



6. Hickerson R, Majumdar ZK, Baucom A, Clegg RM, Noller HF. Measurement of internal movements within the 30 S ribosomal subunit using Forster resonance energy transfer. *J Mol Biol.* 2005;354:459–72.
7. Hammel M, Fierober HP, Czjzek M, et al. Structural basis of cellulose efficiency explored by small angle X-ray scattering. *J Biol Chem.* 2005;280:38562–8.
8. Aramayo R, Merigoux C, Larquet E, et al. Divalent ion-dependent swelling of tomato bushy stunt virus: A multi-approach study. *Biochim Biophys Acta.* 2005;1724:345–54.
9. Fabiola F, Chapman MS. Fitting of high-resolution structures into electron microscopy reconstruction images. *Structure.* 2005;13:389–400.
10. Wriggers W, Birmanns S. Using Situs for flexible and rigid-body fitting of multi-resolution single-molecule data. *J Struct Biol.* 2001;133:193–202.
11. Petoukhov MV, Svergun DI. Global rigid body modeling of macromolecular complexes against small-angle scattering data. *Biophys J.* 2005;89:1237–50.
12. Zheng W, Doniach S. Protein structure prediction constrained by solution X-ray scattering data and structural homology identification. *J Mol Biol.* 2002;316:173–87.
13. Wu YH, Tian X, Lu MY, Chen MZ, Wang QH, Ma JP. Folding of small helical proteins assisted by small-angle X-ray scattering profiles. *Structure.* 2005;13:1587–97.
14. Walther D, Cohen FE, Doniach S. Reconstruction of low-resolution three-dimensional density maps from one-dimensional small-angle X-ray solution scattering data for biomolecules. *J Appl Cryst.* 2000;33:350–63.
15. Topf M, Sali A. Combining electron microscopy and comparative protein structure modeling. *Curr Opin Struct Biol.* 2005;15:578–85.
16. Mears JA, Sharma MR, Gutell RR, et al. A structural model for the large subunit of the mammalian mitochondrial ribosome. *J Mol Biol.* 2006;358:193–212.
17. Chacon P, Diaz JF, Moran F, Andreu JM. Reconstruction of protein form with X-ray solution scattering and a genetic algorithm. *J Mol Biol.* 2000;299:1289–302.
18. Koch MH, Vachette P, Svergun DI. Small-angle scattering: a view on the properties, structures and structural changes of biological macromolecules in solution. *Q Rev Biophys.* 2003;36:147–227.
19. Putnam CD, Hammel M, Hura GL, Tainer JA. X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Q Rev Biophys.* 2007;40:191–285.
20. Svergun DI. Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys J.* 1999;76:2879–86.
21. Svergun DI, Petoukhov MV, Koch MH. Determination of domain structure of proteins from X-ray solution scattering. *Biophys J.* 2001;80:2946–53.
22. Svergun DI, Koch MH. Advances in structure analysis using small-angle scattering in solution. *Curr Opin Struct Biol.* 2002;12:654–60.
23. Wriggers W, Chacon P. Using Situs for the registration of protein structures with low-resolution bead models from X-ray solution scattering. *J Appl Cryst.* 2001;34:773–6.
24. Volkman N, Hanein D, Ouyang G, Trybus KM, DeRosier DJ, Lowey S. Evidence for cleft closure in actomyosin upon ADP release. *Nat Struct Biol.* 2000;7:1147–55.
25. Wendt T, Taylor D, Trybus KM, Taylor K. 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 U S A.* 2001;98:4361–6.
26. Rawat UBS, Zavialov AV, Sengupta J, et al. A cryo-electron microscopic study of ribosome-bound termination factor RF2. *Nature.* 2003;421:87–90.
27. Arai R, Wriggers W, Nishikawa Y, Nagamune T, Fujisawa T. Conformations of variably linked chimeric proteins evaluated by synchrotron X-ray small-angle scattering. *Proteins.* 2004;57:829–38.
28. Shiozawa K, Konarev PV, Neufeld C, Wilmanns M, Svergun DI. Solution structure of human Pex5.Pex14.PTS1 protein complexes obtained by small angle X-ray scattering. *J Biol Chem.* 2009;284:25334–42.
29. Bernadó P, Pérez Y, Blobel J, Fernández-Recio J, Svergun DI, Pons M. Structural characterization of unphosphorylated STAT5a oligomerization equilibrium in solution by small-angle X-ray scattering. *Protein Sci.* 2009;18:716–26.
30. Gorba C, Miyashita O, Tama F. Normal-mode flexible fitting of high-resolution structure of biological molecules toward one-dimensional low-resolution data. *Biophys J.* 2008;94:1589–99.
31. Svergun D, Barberato C, Koch MHJ. CRYSOLE—A program to evaluate x-ray solution scattering of biological macromolecules from atomic coordinates. *Journal of Applied Crystallography.* 1995;28:768–73.
32. Yang S, Park S, Makowski L, Roux B. 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.* 2009;96:4449–63.
33. Oroguchi T, Hashimoto H, Shimizu T, Sato M, Ikeguchi M. Intrinsic Dynamics of Restriction Endonuclease EcoO109I Studied by Molecular Dynamics Simulations and X-Ray Scattering Data Analysis. *Biophys J.* 2009;96:2808–22.
34. Zhao J, Hoye E, Boylan S, Walsh DA, Trehwella J. Quaternary structures of a catalytic subunit-regulatory subunit dimeric complex and the holoenzyme of the cAMP-dependent protein kinase by neutron contrast variation. *J Biol Chem.* 1998;273:30448–59.
35. Krueger JK, Gallagher SC, Zhi G, et al. Activation of myosin light chain kinase requires translocation of bound calmodulin. *J Biol Chem.* 2001;276:4535–8.
36. Akiyama S, Takahashi S, Kimura T, et al. Conformational landscape of cytochrome c folding studied by microsecond-resolved small-angle x-ray scattering. *Proc Natl Acad Sci U S A.* 2002;99:1329–34.
37. Uzawa T, Akiyama S, Kimura T, et al. Collapse and search dynamics of apomyoglobin folding revealed by submillisecond observations of alpha-helical content and compactness. *Proc Natl Acad Sci U S A.* 2004;101:1171–6.
38. Chacon P, Moran F, Diaz JF, Pantos E, Andreu JM. Low-resolution structures of proteins in solution retrieved from X-ray scattering with a genetic algorithm. *Biophys J.* 1998;74:2760–75.
39. Pantos E, vanGarderen HF, Hilbers PAJ, Beelen TPM, vanSanten RA. Simulation of small-angle scattering from large assemblies of multi-type scatterer particles. *J Mol Struct.* 1996;383:303–8.
40. Grishaev A, Wu J, Trehwella J, Bax A. Refinement of multidomain protein structures by combination of solution small-angle X-ray scattering and NMR data. *J Am Chem Soc.* 2005;127:16621–8.
41. Zheng WJ, Doniach S. A comparative study of motor-protein motions by using a simple elastic-network model. *Proc Natl Acad Sci U S A.* 2003;100:13253–8.
42. Zheng W, Doniach S. Fold recognition aided by constraints from small angle X-ray scattering data. *Protein Eng Des Sel.* 2005;18:209–19.
43. Petoukhov MV, Svergun DI. New methods for domain structure determination of proteins from solution scattering data. *J Appl Cryst.* 2003;36:540–4.
44. Tama F, Sanejouand YH. Conformational change of proteins arising from normal mode calculations. *Protein Eng.* 2001;14:1–6.
45. Miyashita O, Onuchic JN, Wolynes PG. Nonlinear elasticity, proteinquakes, and the energy landscapes of functional transitions in proteins. *Proc Natl Acad Sci U S A.* 2003;100:12570–5.
46. Tama F, Miyashita O, Brooks III CL. Flexible multi-scale fitting of atomic structures into low-resolution electron density maps with elastic network normal mode analysis. *J Mol Biol.* 2004;337:985–99.
47. Tirion MM. Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. *Phys Rev Lett.* 1996;77:1905–8.
48. Bahar I, Atilgan AR, Erman B. Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. *Fold and Design.* 1997;2:173–81.



49. Bahar I, Rader AJ. Coarse-grained normal mode analysis in structural biology. *Curr Opin Struct Biol.* 2005;15:586–92.
50. Cerjan CJ, Miller WH. On finding transition states. *J Chem Phys.* 1981;75:2800–6.
51. Petoukhov MV, Svergun DI. Analysis of X-ray and neutron scattering from biomacromolecular solutions. *Curr Opin Struct Biol.* 2007;17:562–71.
52. Grubisic I, Shokhirev MN, Orzechowski M, Miyashita O, Tama F. Biased coarse-grained molecular dynamics simulation approach for flexible fitting of X-ray structure into cryo electron microscopy maps. *J Struct Biol.* 2010;169:95–105.

Publish with Libertas Academica and every scientist working in your field can read your article

“I would like to say that this is the most author-friendly editing process I have experienced in over 150 publications. Thank you most sincerely.”

“The communication between your staff and me has been terrific. Whenever progress is made with the manuscript, I receive notice. Quite honestly, I’ve never had such complete communication with a journal.”

“LA is different, and hopefully represents a kind of scientific publication machinery that removes the hurdles from free flow of scientific thought.”

Your paper will be:

- Available to your entire community free of charge
- Fairly and quickly peer reviewed
- Yours! You retain copyright

<http://www.la-press.com>