

Normal Mode-Based Fitting of Atomic Structure into Electron Density Maps: Application to Sarcoplasmic Reticulum Ca-ATPase

Konrad Hinsén,^{*} Nathalie Reuter,^{†‡} Jorge Navaza,[§] David L. Stokes,^{¶||} and Jean-Jacques Lacapère[†]

^{*}Laboratoire Léon Brillouin (CEA-CNRS), 91191 Gif sur Yvette, France; [†]U410 INSERM, Faculté de Médecine Xavier Bichat, 75870 Paris Cédex 18, France; [‡]Computational Biology Unit, Bergen Center for Computational Science, University of Bergen, 5008 Bergen, Norway; [§]Laboratoire de Virologie Moléculaire Structurale, 91198 Gif sur Yvette, France; [¶]Skirball Institute, New York University School of Medicine, New York, New York 10012 USA; and ^{||}New York Structural Biology Center, New York, New York 10027 USA

ABSTRACT A method for the flexible docking of high-resolution atomic structures into lower resolution densities derived from electron microscopy is presented. The atomic structure is deformed by an iterative process using combinations of normal modes to obtain the best fit of the electron microscopical density. The quality of the computed structures has been evaluated by several techniques borrowed from crystallography. Two atomic structures of the SERCA1 Ca-ATPase corresponding to different conformations were used as a starting point to fit the electron density corresponding to a different conformation. The fitted models have been compared to published models obtained by rigid domain docking, and their relation to the known crystallographic structures are explored by normal mode analysis. We find that only a few number of modes contribute significantly to the transition. The associated motions involve almost exclusively rotation and translation of the cytoplasmic domains as well as displacement of cytoplasmic loops. We suggest that the movements of the cytoplasmic domains are driven by the conformational change that occurs between nonphosphorylated and phosphorylated intermediate, the latter being mimicked by the presence of vanadate at the phosphorylation site in the electron microscopy structure.

INTRODUCTION

The spectacular development in the field of cryoelectron microscopy in recent years has made possible the solution low-resolution structures of many proteins and molecular complexes (Saibil, 2000; Chiu et al., 1999; Frank et al., 2002, 2003). To understand the function of these complexes, it is of particular interest to obtain the arrangement of their components at the atomic level. Computational docking and fitting techniques have been developed to assemble the atomic structures of their components (Roseman, 2000; Wriggers and Chacon, 2001; Rossmann, 2000; Chacon and Wriggers, 2002; Navaza et al., 2002; Volkman and Hanein, 1999; Wu et al., 2002). However, in many biological systems, large conformational changes of the proteins occur as part of their function and characterize their different transition states (Saibil, 2000; Unwin, 2003). This lowers the quality of the fit for molecular complexes and even more so for isolated proteins. The challenge thus is to find the best atomic structure fitting into the electron microscopy (EM) densities.

The conformational changes of proteins often involve the relative motion of semirigid structural elements. Such motions are well described by traditional molecular simulation techniques such as molecular dynamics (Karplus and McCammon, 2002; Beckstein et al., 2003), but such calculations are very time consuming and not feasible for large proteins. Normal mode analysis is an interesting alternative, in particular when used with coarse-grained models (Hinsén, 1998; Hinsén et al., 2000; Valadie et al., 2003; Reuter et al.,

2003) that are compatible with the resolution of the electronic densities and that permit the treatment of large systems in a reasonable amount of time.

In this work, we present a new method for flexibly fitting a known structure into an EM density for another conformation of the same protein(s). We demonstrate this method for the example of Ca-ATPase, fitting two known crystallographic structures into the EM density for an unknown conformation. As revealed by atomic structure analysis, (Reuter et al., 2003), this protein undergoes massive domain changes during its catalytic cycle of calcium transport (Green and MacLennan, 2002), thus representing a challenging application. Compared to the manual docking of domains (Xu et al., 2002), it has the advantage of being nearly automatic and of treating the molecule uniformly, with no artificially imposed break points. Compared to the recently developed method of deforming the EM densities directly (Tama et al., 2002; Ming et al., 2002), it has the advantage of providing an atomic structure that can then be studied with standard analysis tools, even though one must be careful not to overinterpret the structure, which is inevitably much less reliable than a structure obtained from high-resolution data by standard refinement techniques.

METHODS

The structure fitting procedure consists of three steps:

1. Find a suitable initial orientation for the known structure relative to the electron density.
2. Deform the initial structure iteratively in such a way that it reflects the electron density better while minimizing the deformation energy at each

Submitted July 30, 2004, and accepted for publication October 21, 2004.

Address reprint requests to J.-J. Lacapère, E-mail: lacapere@bichat.inserm.fr; or K. Hinsén. E-mail: hinsen@llb.saclay.cea.fr

© 2005 by the Biophysical Society

0006-3495/05/02/818/10 \$2.00

doi: 10.1529/biophysj.104.050716

step. A simplified protein model consisting only of the C_α atoms is used at this stage for efficiency.

3. Reconstruct the peptide plane and side-chain atoms to obtain an all-atom model.

These three steps will now be described in detail.

Initial orientation

As with all iterative methods, the quality and even more the efficiency of the structure determination procedure depend on a good starting point. The initial conformation being given, only the position and orientation in space can be chosen for optimization. For the position, the choice is obvious: the center of mass of the protein is moved to the center of the electron density. Finding a good initial orientation is less evident, in particular in the presence of near-symmetries.

We use a heuristic procedure that will find a very good initial orientation in most cases and propose several choices for inspection by the user if the protein is highly symmetric. First, the center of mass and the principal axes of inertia are calculated for the input conformation. The three planes that are perpendicular to the axes of inertia and intersect in the center of mass divide the protein into eight regions, for which we calculate the centers of mass. Then an analogous calculation is performed on the electron density, using, of course, the density itself instead of the mass. We then have eight points that characterize the protein and eight points that characterize the electron density.

The initial orientation for the known structure is then obtained by finding the rotation matrix that, when applied to the protein reference points, minimizes the root mean-square (RMS) distance to the electron density reference points. This is similar in principle to the superposition of two protein conformations, except that in this situation it is not known a priori which of the eight protein points corresponds to which of the eight electron density points. There are 24 possible ways to match up the two point sets that differ among each other only by rotations (for each of the three axes of inertia, there are four 90° rotations around it plus an inversion), which must all be tried to find the one that yields the lowest RMS distance. If several RMS distances are close, then the choice between the corresponding initial orientations will be left to the user.

Modification of the conformation

The physical idea behind the structure deformation process is “pulling” atoms toward places with high electronic densities and away from places with low electronic densities. Since the atoms interact strongly with their neighbors, pulling one of them will not break it out of the molecule, but move the neighboring atoms along to some extent. Each atom is thus subject to the pulling force as well as to the interactions with other atoms.

In the simplified protein model we use at this stage, the protein is represented by N point masses located at the C_α positions and having the masses of the whole residues that they represent. With each conformation \mathbf{R}_i , $i = 1 \dots N$ of the protein, we associate a model electron density

$$d_{\text{model}}(\mathbf{r}; \mathbf{R}_1, \dots, \mathbf{R}_N) = C \sum_{i=1}^N \exp\left(-\frac{(\mathbf{r} - \mathbf{R}_i)^2}{2r_0^2}\right), \quad (1)$$

where C is a normalization coefficient chosen such that the integral over the model electron density is equal to the integral of the experimental electron density $d(\mathbf{r})$. The radius r_0 defines the extent in space of the residue represented by a particle. The value is not very critical; empirically a value of 0.4 nm was found to be suitable.

The quantity that we want to minimize in this procedure is

$$\Delta(\mathbf{R}_1, \dots, \mathbf{R}_N) = \sum_{\text{grid points } k} [d(\mathbf{r}) - d_{\text{model}}(\mathbf{r}_k; \mathbf{R}_1, \dots, \mathbf{R}_N)]^2, \quad (2)$$

where the summation runs over all the points of the grid that is defined by the experimental data. We normalize both the experimental and the model electronic densities such that

$$\sum_{\text{grid points } k} d(\mathbf{r}) = 1. \quad (3)$$

As a consequence, $\Delta(\mathbf{R}_1, \dots, \mathbf{R}_N)$ is dimensionless. The negative gradient of this quantity,

$$\mathbf{F}_i = -\frac{\partial \Delta}{\partial \mathbf{R}_i}, \quad (4)$$

is therefore the force we use to pull at the particles. However, we still need to define the interactions between the particles that keep the protein together. It was shown in Hinsen (1998) that a simple harmonic potential together with the simplified protein model described above can reproduce normal modes and domain motions in proteins very well. We therefore use such a harmonic potential as well, which has the form (Hinsen, 1998)

$$U(\mathbf{x}_1, \dots, \mathbf{x}_N) = \sum_{\text{all pairs } \alpha, \beta} U_{\alpha\beta}(\mathbf{x}_\alpha - \mathbf{x}_\beta) \quad (5)$$

with the pair term

$$U_{\alpha\beta}(\mathbf{r}) = k(\mathbf{x}_{\alpha\beta}^{(0)})(|\mathbf{r}| - |\mathbf{x}_{\alpha\beta}^{(0)}|)^2. \quad (6)$$

The displacements $\mathbf{x}_\alpha = \mathbf{R}_\alpha - \mathbf{R}_\alpha^0$ are taken with respect to the input conformation \mathbf{R}_α^0 , and $\mathbf{x}_{\alpha\beta}^{(0)}$ is the pair distance vector $\mathbf{x}_\alpha - \mathbf{x}_\beta = \mathbf{R}_\alpha - \mathbf{R}_\beta$ in that conformation. For the pair force constant $k(r)$, we used the improved version (Hinsen et al., 2000) that was fitted to the Amber 94 all-atom force field (Cornell et al., 1995):

$$k(r) = \begin{cases} 8.6 \times 10^5 \frac{\text{kJ}}{\text{mol nm}} * r - 2.39 \times 10^5 \frac{\text{kJ}}{\text{mol nm}} & \text{for } r < 0.4 \text{ nm} \\ \frac{128 \text{ kJ nm}^4 / \text{mol}}{r} & \text{for } r \geq 0.4 \text{ nm} \end{cases}. \quad (7)$$

The two distance categories are required to describe the substantial difference between nearest-neighbor pairs along the backbone (distances below 0.4 nm) and all other pairs.

In the following, we will write the harmonic potential as

$$U(\mathbf{x}_1, \dots, \mathbf{x}_N) = \frac{1}{2} \mathbf{x} * \mathbf{K} * \mathbf{x}, \quad (8)$$

where \mathbf{x} is a $3N$ -dimensional vector comprising the displacements of all particles, and \mathbf{K} is a $3N \times 3N$ -dimensional force constant matrix. Introducing mass-weighted coordinates $\tilde{\mathbf{x}}$ such that $\tilde{\mathbf{x}}_i = \sqrt{m_i} \mathbf{x}_i$, we can also write

$$U(\tilde{\mathbf{x}}_1, \dots, \tilde{\mathbf{x}}_N) = \frac{1}{2} \tilde{\mathbf{x}} * \tilde{\mathbf{K}} * \tilde{\mathbf{x}} \quad (9)$$

with the mass-weighted force constants $\tilde{\mathbf{K}}$. The eigenvectors $\mathbf{u}^{(k)}$, $k = 1 \dots 3N$, of $\tilde{\mathbf{K}}$ are the normal modes of the protein, and the associated eigenvalues $\lambda^{(k)}$ are the squares of the vibrational frequencies, the first six of which are zero because the global translation and rotation of the protein do not cause a change in energy. The matrix $\tilde{\mathbf{K}}$ can be expressed in terms of its eigenvectors and eigenvalues as

$$\tilde{\mathbf{K}} = \sum_{k=1}^{3N} \lambda^{(k)} \mathbf{u}^{(k)} \mathbf{u}^{(k)}. \quad (10)$$

When an external force is applied to a harmonic system such as the one described above, the particles will move to new positions in which the total

force is zero. The corresponding displacements can be expressed in terms of the normal modes as

$$\tilde{\mathbf{x}} = \sum_{k=7}^{3N} (\lambda^{(k)})^{-1} (\mathbf{u}^{(k)} * \mathbf{F}) \mathbf{u}^{(k)}. \quad (11)$$

Note that the summation starts at 7 to exclude the zero-frequency modes.

In principle, our optimization procedure works as follows:

1. Set \mathbf{R} to the initial conformation.
2. Repeat the following steps until convergence: a), calculate \mathbf{F} , $\mathbf{u}^{(k)}$, and $\lambda^{(k)}$ for the conformation \mathbf{R} ; b), calculate \mathbf{x} from Eq. 11; and c), set $\mathbf{R} \leftarrow \mathbf{R} + a\mathbf{x}$, where a is chosen such that $a\mathbf{x}$ has a predefined length l .

There are, however, a number of details to be taken into account. First, the forces \mathbf{F} contain global translation and rotation components, which are suppressed by Eq. 11. They are, however, important because, in general, global motions must be considered in the optimization process. We include global motion by starting the summation from 1 in Eq. 11 and by setting $\lambda^{(k)} = 0.2\lambda^{(7)}$ for $k < 7$. This is equivalent to adding a very soft harmonic potential for global motions.

Second, it is not efficient to use all $3N$ modes in the procedure. It is clear from Eq. 11 that high-frequency modes contribute only little to the displacement, because their $\lambda^{(k)}$ are large. Their inclusion could even be detrimental, because the higher modes are sensitive to the high-frequency noise that is an inevitable part of experimental data. We are interested in large-scale displacements, which are described by the low-frequency modes. Therefore, we take only a limited number of modes into account. The number of modes critically influences the outcome of the minimization procedure, as will be shown in the next section. It is therefore a parameter whose optimal value must be determined empirically in each application. Since it would be wasteful to calculate $3N$ modes and then throw most of them away, we use the Fourier basis technique described in Hinsen (1998) to calculate only a smaller number of modes, about twice the number of modes that will be used in the fitting procedure.

Third, experience shows that there is a risk of the minimization procedure getting stuck in a shallow local minimum. There are established methods to deal with that problem (e.g., simulated annealing), but a much simpler remedy was found sufficient for this application: adding a random vector to the force obtained from Eq. 4. The random vector is drawn from a Gaussian distribution whose standard deviation is 20% of the norm of the force vector.

We still need to specify the step length parameter l , and there is also one further improvement in the method to discuss. In the initial phase of the procedure, we would like to take only very large-scale motion into account, and make big steps through configuration space. As the optimization progresses, we would like to move more slowly and take more details into account. We do this by introducing a weighting function that emphasizes the lowest frequency modes, i.e., we replace Eq. 11 by

$$\tilde{\mathbf{x}} = \sum_{k=1}^{3N} w(\lambda^{(k)}, \lambda_0)^{-1} (\mathbf{u}^{(k)} \cdot \mathbf{F}) \mathbf{u}^{(k)} \quad (12)$$

with

$$w(\lambda, \lambda_0) = \frac{1}{1 - (\lambda - \lambda_0)^2}. \quad (13)$$

The optimization procedure starts with $l = 0.1$ nm and $\lambda_0 = 0$. As optimization progresses, l is decreased and λ_0 is increased.

Reconstruction of an all-atom model

The last step in the structure determination consists of recreating an all-atom model from the simplified one particle-per-residue model. This is done by starting from the atom positions in the original input conformation and

displacing them by applying a linear coordinate transformation (i.e., a rigid body translation and rotation) that is obtained from a superposition fit involving the C_α position of the residue under consideration plus its two nearest neighbors. The all-atom conformation obtained in this way is then improved by standard energy minimization using the Amber 94 force field (Cornell et al., 1995).

Preprocessing of the electronic density

In the procedure described above, it was implicitly assumed that the electronic density map describes one protein and nothing else. However, a real experimental electronic density map is usually obtained for a system containing many instances of the same molecule in some regular arrangement, and often in the presence of other molecules that have been added to prepare the system, such as crystallization agents. This difficulty can be resolved by two approaches: extending the fitted model to include the supplementary sources of electronic density, or removing the unwanted parts from the electronic density before performing the fit. We have chosen the second approach, and this section describes how the density has been preprocessed.

The initial isolation of a single molecule was done by hand and is not perfect; little pieces of electronic density representing neighboring molecules remain. There was also a strong contribution (the highest local maximum, in fact) from the decavanadate that was used as a crystallization agent. These pieces can negatively influence the fitting procedure described above by “pulling” the structure toward them.

To avoid such artifacts, we preprocess the experimental electronic density with the aim of removing the excess density. Of course, any algorithmic approach is necessarily heuristic and the quality of its outcome relies on the quality of the initial manual identification of the molecule. However, our preprocessing is adapted to the specific requirements of the fitting procedure. For that application, excess densities are more of a problem than missing densities at the outer parts of the protein, because the latter could only cause an artificial compression of the structure, which, however, is not to be expected in the case of the very dense proteins. We therefore choose a method that reliably removes excess densities at the risk of removing too much.

This method works by building up a subset of the grid points on which the input density values are kept; the grid points not in that subset are assigned a density of zero. We start by adding all those points to the subset whose density value is at least half the highest density in the map. These points form the core of the protein. We then examine another layer of points around the accepted subset. Those that are lower than their already accepted neighbors are accepted as well; those that are higher are discarded. The procedure is iterated until the subset becomes stable. Physically speaking, this algorithm goes “downhill” from the accepted subset but never “uphill”, thus avoiding the inclusion of secondary maxima.

The same basic approach was also used to remove the decavanadate. The algorithm was started with the subset of points corresponding to the highest 10% of values in the density map. The region selected after convergence was considered to represent the decavanadate, and its density values were set to zero. It should be noted that the initial fits were made with the decavanadate, which was only subsequently removed (as discussed in Results and Discussion) to study the effect of this correction on the resulting structure.

Software and experimental data

The software that implements the algorithms described in this section consists of a few modules written in the Python language (van Rossum et al., 1991) and making heavy use of the Molecular Modelling Toolkit (Hinsen, 2000). The URO package (Navaza et al., 2002) was used for evaluating the resulting structures. For preparing the images in this article, we used the molecular visualization programs PyMOL (Delano, 2002) and VMD (Humphrey et al., 1996).

Two crystallographic conformations of Ca-ATPase have been used as starting points for our methods, 1EUL (Toyoshima et al., 2000) and 1IWO (Toyoshima and Nomura, 2002). Two other conformations, obtained by manual fitting into EM densities, have been used for comparison, 1FQU and 1KJU (Xu et al., 2002). The EM density used in our calculations is described in Zhang et al. (1998). Briefly, tubular crystals of Ca-ATPase were obtained from sarcoplasmic reticulum vesicles incubated in the presence of decavanadate and thapsigargin in the absence of calcium. Images of the tubes were obtained by cryoelectron microscopy, and computer analysis of tubes was performed to get density maps. Each of the two molecules comprising the unit cell was masked, aligned, and averaged. The final map was obtained after enforcing the twofold symmetry and truncating data at 8 Å resolution.

RESULTS AND DISCUSSION

The calcium homeostasis in animal, plant, and yeast cells is crucial for their metabolism. It involves many transmembrane calcium flux regulations. In this process, a group of ATP-dependent calcium pumps belonging to the P-type ATPases family (Møller et al., 1996) maintains intracellular calcium concentration at low levels. The calcium ATPases are transmembrane enzymes that actively transport calcium. Among them, the Ca-ATPase from the skeletal muscle sarcoplasmic reticulum is structurally and functionally the best-studied active ion transporter (for reviews, see Lee and East, 2001; Green and MacLennan, 2002; Stokes and Green, 2003; Young and Stokes, 2004). The hydrolysis of ATP provides the required energy to move calcium ions against its electrochemical potential gradient, whereas protons are countertransported. The formation of a stable aspartyl-phosphoryl-enzyme intermediate distinguishes the P-type ATPases from the other ATPases such as the F- and V-type ATPases, which are ATP synthases.

Ca-ATPase forms both two- and three-dimensional crystals. The former have been observed by cryoelectron microscopy (EM), whereas the latter have been analyzed by x-ray diffraction. From the EM images of the tubular crystals obtained in different experimental conditions, several three-dimensional electron densities have been described at various resolutions ranging from 6 to 15 Å (Toyoshima et al., 1993; Zhang et al., 1998; Xu et al., 2002).

More recently, two atomic structures of Ca-ATPase obtained from x-ray diffraction of crystals grown in the presence (Toyoshima et al., 2000) and in the absence (Toyoshima and Nomura, 2002) of calcium have been described. These two different conformations have been obtained at a higher resolution (2.6 and 3.1 Å, respectively) compared to EM. Using the method described in Methods, we fitted these two conformations (Protein Data Bank (PDB) code 1EUL and 1IWO) into an EM density for Ca-ATPase at 8 Å (Zhang et al., 1998).

Among the various parameters of our method, the atomic radius, the strength of the random force, and the step size are not very critical, they have been fixed empirically as described in Methods. The remaining parameter, the number of modes taken into account in the calculation of the

displacements, has to be considered more carefully. A larger number of modes provides the model with more local flexibility, which allows it to adapt better to the EM density, yielding a better fit. On the other hand, one of the reasons for using a limited number of normal modes instead of minimizing Eq. 2 directly is to smooth out noise in the experimental EM density, and this effect is reduced with an increasing number of modes. As a result, the minimization can get stuck in a secondary minimum. This problem could be circumvented by a more elaborate minimization technique, such as simulated annealing.

The iterative procedure searches for a minimum defined as the square difference between the experimental EM density and the density generated from the model (Eq. 2). We call this value the fit error. Fig. 1 shows the decrease of the fit error as a function of the number of iterations for the different number of normal modes. It appears that there is no direct correlation between the number of modes used and the minimum reached. Indeed, Fig. 1 A shows that 128 modes give the best result when starting from conformation 1EUL, whereas 322 modes is the optimal value for conformation

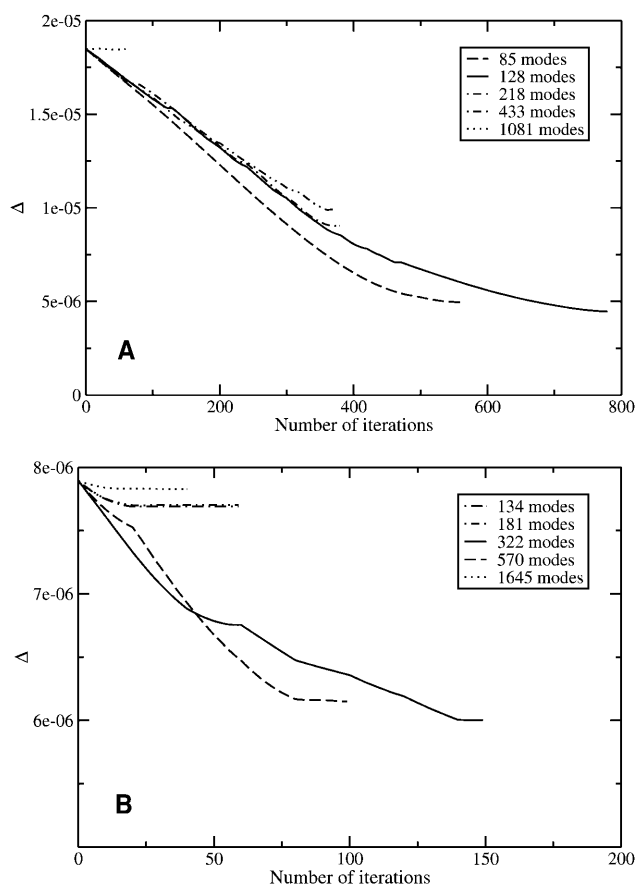


FIGURE 1 Fit error Δ (see Eq. 2) as a function of the number of iterations for different numbers of normal modes. (A) Iteration starting from conformation 1EUL (PDB code) (B) Iteration starting from conformation 1IWO (PDB code).

1IWO. This illustrates the effect explained above: the number of modes should neither be too high nor too low. Therefore one has to test several numbers of modes for each fitting operation. Comparison of Fig. 1, *A* and *B*, shows that the final model structure obtained from conformation 1EUL gives a slightly better fit than the conformation 1IWO even though the initial values are in favor of 1IWO. This suggests that the proximity of the atomic structure used as a starting point to fit the EM density is not the crucial point.

To check the quality of the fit, we used several techniques: i), visual inspection (Fig. 2) and ii), numerical evaluation (Table 1), using the reciprocal space rigid body fitting method described in Navaza et al. (2002). Fig. 2 *A* clearly shows that the superposition of the EM density is much better with the final fitted model (Fit1EUL) than with the initial model (1EUL). The electronic density of the model has been calculated and compared to the experimental one (Fig. 2 *B*). Differences in the outer regions are clearly observed, but Fig. 2, *C* and *D*, show that in the interior, the fit of the EM densities by the model is very good. Indeed, cross sections in two different regions of the protein (trans-

TABLE 1

Model	CC ($\times 100$)	<i>R</i> ($\times 100$)	<i>Q</i>
1EUL	70.0	55.9	27.5
1FQU	88.7	47.3	10.8
1KJU	89.6	47.6	10.0
Fit 1EUL	91.0	45.4	8.7
Fit 1IWO	89.8	45.5	9.8
1IWO	88.5	47.5	11.0

membrane and cytoplasmic region, Fig. 2, *C* and *D*, respectively) show that the densities are well fitted with the model. Similar results have been obtained with the conformation 1IWO as the starting point (Fig. 3).

The quality of the fitted models (Fit1EUL and Fit1IWO) has been evaluated by the reciprocal space rigid body fitting method (Navaza et al., 2002). Table 1 shows the correlation coefficient (CC), the crystallographic *R* factor, and the quadratic misfit *Q* obtained for the fit of both models in the experimental density (cut at 15 Å resolution). We compared them with the initial conformations (1EUL and 1IWO), and

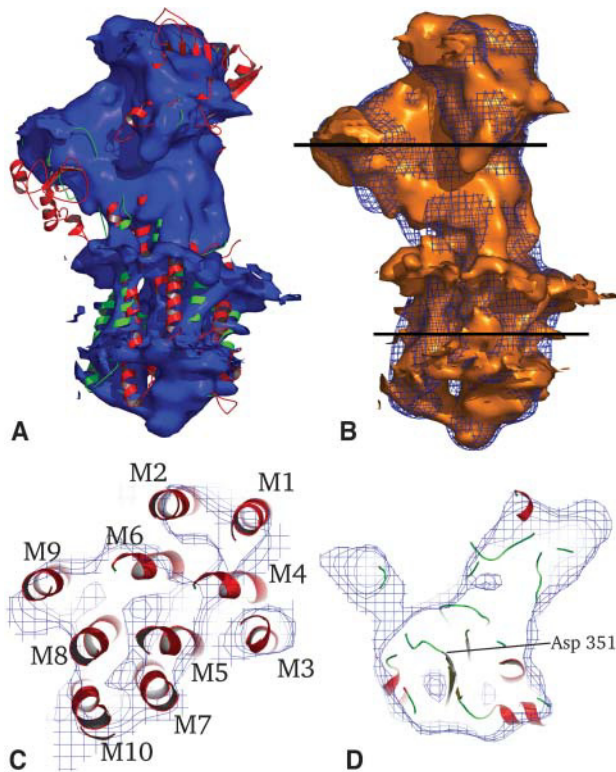


FIGURE 2 Visual inspection of the fitted structures. (A) The density map of Ca-ATPase with the conformations 1EUL (red) and the fit obtained starting from 1EUL (green) (B) Superposition of experimental (orange surface) and model (blue grid) electron densities, both represented by isosurfaces at 40% of the maximum value. The density for the model has been calculated from Eq. 1 for the final conformation. (C and D) Cross sections through the transmembrane and cytoplasmic regions, respectively. The positions of the cross sections are indicated in *B* by the black lines.

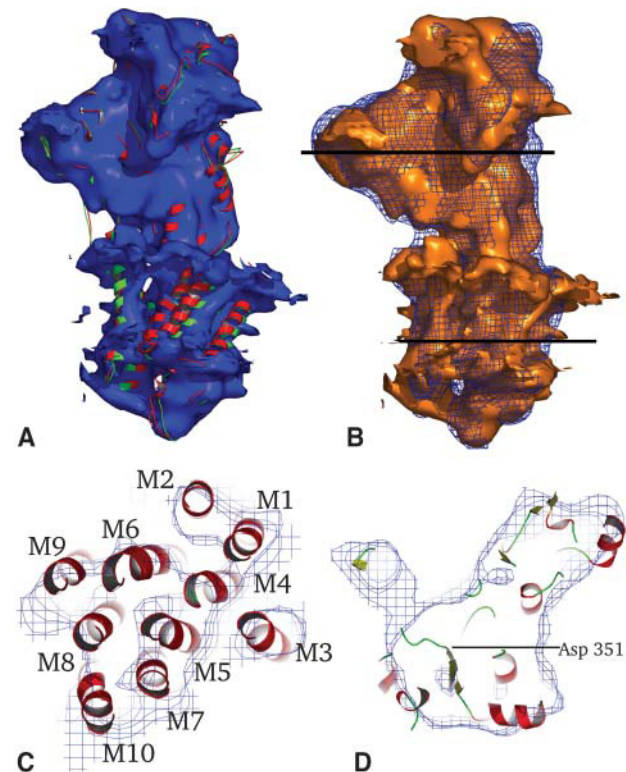


FIGURE 3 Visual inspection of the fitted structures. (A) The density map of Ca-ATPase with the conformations 1IWO (red) and the fit obtained starting from 1IWO (green) (B) Superposition of experimental (orange surface) and model (blue grid) electron densities, both represented by isosurfaces at 40% of the maximum value. The density for the model has been calculated from Eq. 1 for the final conformation. (C and D) Cross sections through the transmembrane and cytoplasmic regions, respectively. The positions of the cross sections are indicated in *B* by the black lines.

both structures obtained by our fitting procedure exhibit much better correlation coefficients. Two atomic models fitting this EM density have been previously described (PDB codes 1FQU and 1KJU). They are included in Table 1, which shows that even though they give good correlation coefficients, they are both worse than our models (Fit1EUL and Fit1IWO). The model 1FQU was generated from the structure 1EUL by rearranging the three well-separated cytoplasmic domains taken as rigid bodies to fit the EM density. The model 1KJU was also generated from the structure 1EUL by breaking the peptides at three hinge points, yielding four domains (Xu et al., 2002). Each domain was then separately fitted into the EM density by manual docking with the main focus on matching the helices into strong columns of density in the map, whereas the three cytoplasmic domains were treated as rigid bodies. Our models have been obtained by correlated displacements of the overall structure following the normal mode prediction—it leads probably to more coherent displacements than those obtained by moving rigid substructures, which then have to be reconnected.

To characterize our fitted structures, we performed several analyses: i), global root mean-square deviation (rmsd) of each structure compared to the others, ii), deformation analysis, and iii), residue displacement analysis. Table 2 shows that all the various conformations obtained from the 1EUL structure (including our fitted one) exhibit a high rmsd (~ 1.2 nm) with respect to 1EUL. For our fitted model (Fit1EUL), a high value is in agreement with large-amplitude movements of the cytoplasmic domains (see Fig. 2 A). Our model shows slightly lower rmsd compared with the structures 1FQU and 1KJU, indicating that the global deformation is less pronounced. Conversely, the model Fit1IWO, obtained from the 1IWO structure, exhibits a much lower rmsd (0.15 nm) with respect to the initial structure (1IWO). This shows that Fit1IWO has been obtained by small amplitude movements.

An analysis of the amplitudes of displacement of the various residues provides further information about the deformations that the conformations have undergone in the fitting process (Figs. 4 and 5). The deformation of 1EUL involves mainly the cytoplasmic domains and the intracellular loops (Fig. 4 A, shown in *red*) and to a lesser extent the transmembrane helices (shown in *green*). The amplitudes of displacement for 1IWO are much smaller (since almost no

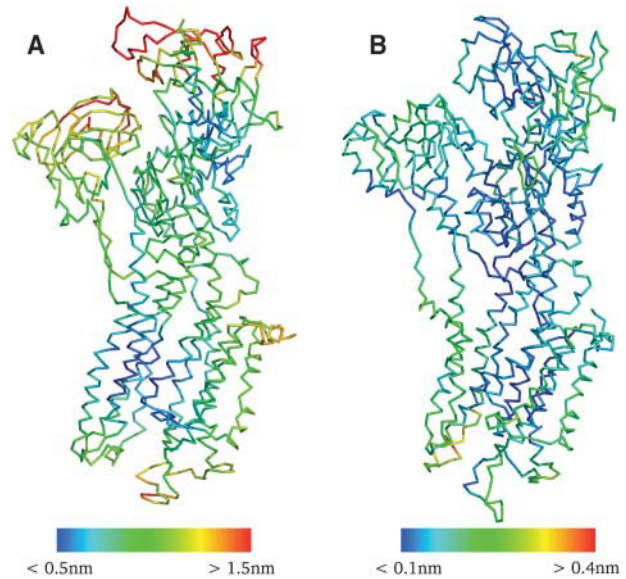


FIGURE 4 Deformation during the fitting procedure for Fit1EUL (A) and Fit1IWO (B). The color indicates by how much a residue has moved relative to the initial conformation during the fitting procedure. Blue stands for small displacements, red for large ones, and green indicates intermediate values.

region is colored *red* in Fig. 4 B). The region containing the phosphorylation site and the adjacent transmembrane helices undergo the strongest displacements in 1IWO (Fig. 4 B, shown in *green*). Fig. 5 shows that the amplitudes of displacement of all the residues of 1EUL during the fitting procedure are far larger than those applied to 1IWO. The smallest values observed for 1EUL correspond to the transmembrane helices. All the residues forming the A domain have been displaced, whereas some residues forming the N domain have not moved much. The highest amplitude of

TABLE 2

Model	1EUL	Fit1EUL	1IWO	Fit1IWO	1FQU	1KJU
1EUL	0	1.13	1.40	1.40	1.38	1.32
Fit1EUL		0	0.77	0.75	0.66	0.67
1IWO			0	0.15	0.57	0.56
Fit1IWO				0	0.53	0.53
1FQU					0	0.56
1KJU						0

The RMS distances (in nm) between the six structures discussed in Results and Discussion, calculated from the C_{α} atoms only.

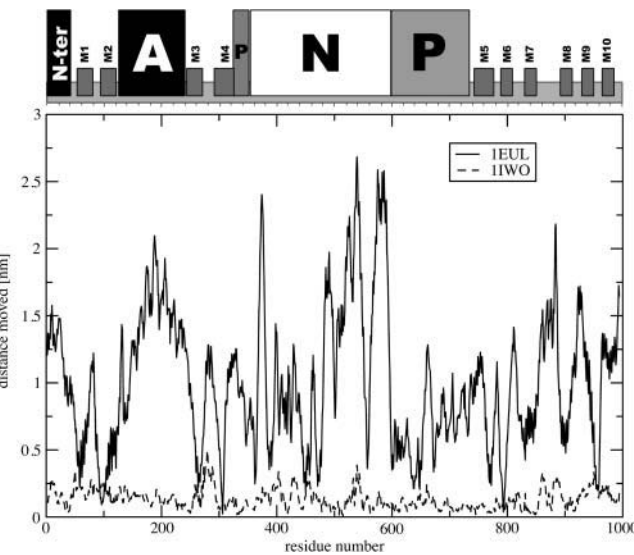


FIGURE 5 Displacement (in nm) of each residue relative to the initial conformations for Fit1EUL and Fit1IWO.

displacement corresponds to the L7-8 loop. This is also true for 1IWO, where similar changes are also observed for the L3-4 loop. The various residues surrounding the phosphorylation site (D351) are displaced as previously observed in Fig. 4.

To characterize the conformational transition between the initial structures and the EM-fitted structures, we have identified the normal modes and thus the movements that contribute most to the two transitions (1EUL to Fit1EUL and 1IWO to Fit1IWO). The full set of normal modes of the initial conformation (1EUL and 1IWO, respectively) was projected onto the normalized vector that describes the difference between final and initial conformations. This vector was constructed from the differences between the positions of the C_α atoms after optimal superimposition of the final onto the initial conformation. Fig. 6 shows the squared overlap (squared dot product) in plots A and C, and the cumulative squared overlap in plots B and D.

Twenty modes already account for 86% of the difference between the 1EUL and Fit1EUL structures, but for only 65% of the difference between the 1IWO and Fit1IWO structures. Only 350 modes are necessary to represent 97% of the difference between 1EUL and Fit1EUL, whereas 1000

modes are needed to represent 97% of the difference between 1IWO and Fit1IWO. This shows that the low-energy modes (which describe large-amplitude motions) are the most involved in the deformation for both initial structures (1EUL and 1IWO). However, the open structure (1EUL) requires much fewer modes to generate a good fit (Fig. 1 A), whereas the closed structure (1IWO) needs more modes (Fig. 1 B). This phenomenon is also observed when comparing the forward and backward transition between 1EUL and 1IWO (also shown in Fig. 6, B and D, as *dotted lines*).

The mode counts given here should not be confused with the number of modes required during the fitting procedure. It is a general feature of physical models that a good but approximative description of a phenomenon can be achieved with a less detailed model than a reconstruction or simulation of the same phenomenon. Twenty modes may give a good description (86%) of the structural change from 1EUL to Fit1EUL, but the remaining 14% must still come from somewhere to make the change happen. It must also be stressed that the mode counts given here refer to a single-step description of the structural change, whereas the fitting procedure is iterative and requires repeated normal mode calculations.

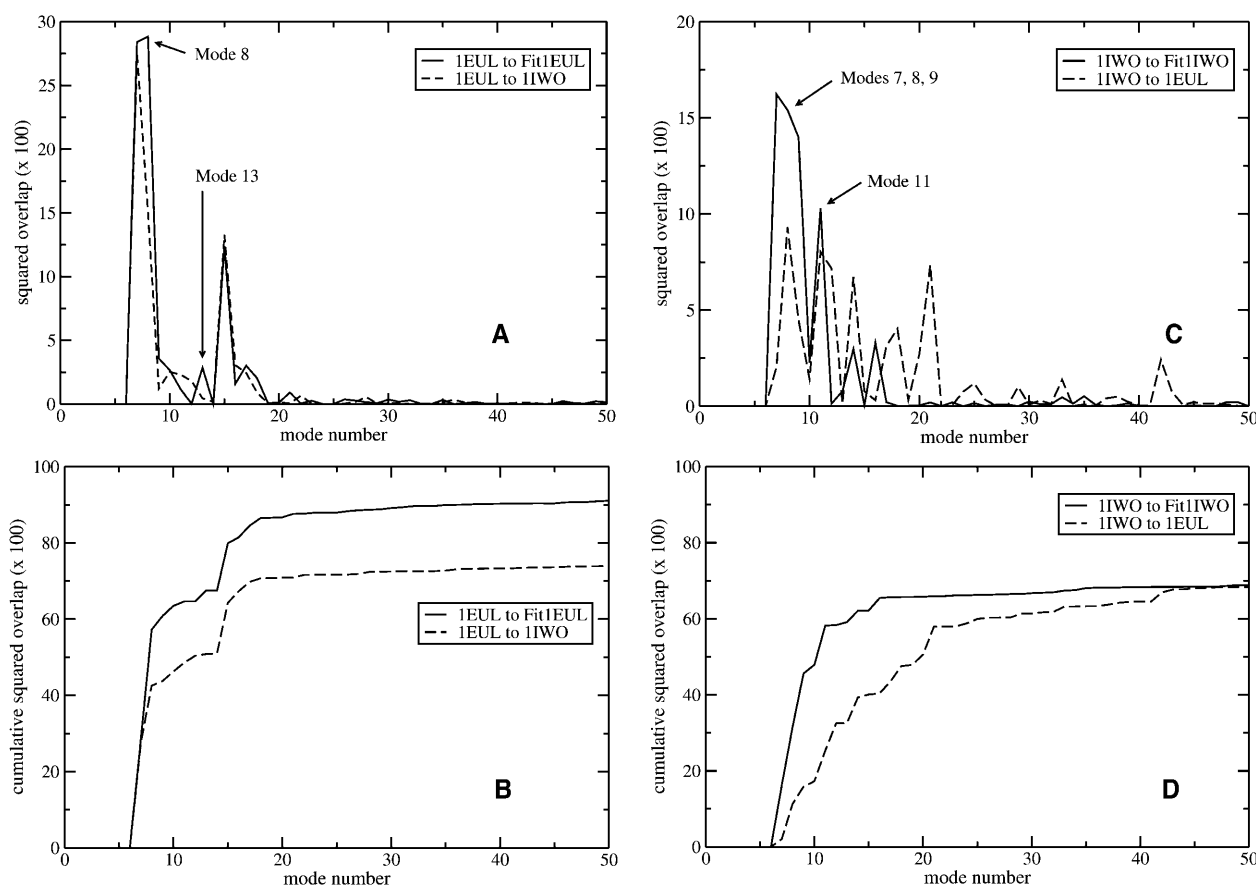


FIGURE 6 Squared overlap (A and C) and cumulative squared overlap (B and D) of the distance vector between initial and final conformations with the normal modes calculated for the initial conformations, 1EUL (A and B) and 1IWO (C and D). Only the first 50 modes are shown.

Previous manual fitting of EM density starting from the 1EUL conformation (Toyoshima et al., 2000; Xu et al., 2002) involved mostly movements of the cytoplasmic residues. This agrees with our normal mode-based fitting technique, which emphasizes the slow modes that represent movements of the cytoplasmic domains. Considering the transition between 1EUL and Fit1EUL, modes 7 and 8, the two slowest nonzero modes, have a similar contribution of 28% each (Fig. 6 A). The next modes, 9–14, do not make important contributions, but mode 15 contributes up to 12% of the difference vector. Thus, in total, three modes (7, 8, and 15) make up for 69% of the difference vector. Looking at the transition between 1IWO and Fit1IWO, modes 7, 8, 9, and 11 have contributions above 10% (Fig. 6 C). Thus, in total, four modes (7, 8, 9, and 11) make up for 55% of the difference vector. The 1IWO structure is very close to the EM density (see Table 1), so the deformations required to obtain an even better fit involve mainly local high energy deformations (i.e., higher mode numbers). In other words, the normal mode-based improvement of the EM density fit requires a larger number of modes for an already close structure than for a more distinct structure. This reflects the general feature of normal modes that low-energy modes correspond to large-amplitude motions.

For the transition between 1EUL and Fit1EUL, modes 7 and 8 are characterized by large movements of the *N* domain, whereas mode 15 shows movements of the luminal loops, and in particular L7-8 (Reuter et al., 2003). Interestingly, mode 8, which is more important than in the transition between 1EUL and 1IWO, is characterized by a rotation of the *N* domain, suggesting a difference between 1IWO and the EM density. This is confirmed by the contribution of mode 11 for the transition between 1IWO and Fit1IWO, which is characterized by both rotation and translation of the *N* domain (Fig. 7). Another difference between the EM density and 1IWO arises from the contribution of mode 13 in

the transition from 1EUL toward Fit1EUL, which is not observed for the one from 1EUL toward 1IWO (Fig. 6 A). The movement associated with this mode is a rotation of domains *A* and *N* around axes parallel to the membrane (Reuter et al., 2003). This movement of domains *A* and *N* is confirmed by the contribution of mode 11 and to lesser extent of modes 14 and 16 in the transition between 1IWO and Fit1IWO (Fig. 6 C). Fig. 7 shows that movements associated with these modes are described by a rotation of domains *A* and *N* along axes parallel to the membrane (axis 1 and 2). Interestingly, mode 11 also shows concerted movements of luminal loops (L3-4 and L7-8), whereas mode 14 shows concerted movements of domains *A* and *N*. Mode 16 involves mostly concerted displacement of L3-4 and L7-8.

The biological significance of these findings is that the differences in the observed movements of the cytoplasmic domains between the transition from 1EUL to 1IWO on one hand and the fits from either 1EUL or 1IWO to the experimental EM density on the other hand suggest that the EM density represents a conformation that is distinct from both 1EUL and 1IWO, i.e., from structures both with and without calcium. Electron microscopy has solved an analog of the calcium-free enzyme stabilized by vanadate, which is therefore closer to 1IWO (without calcium) than 1EUL (with calcium). Moreover, both 1IWO and the electron density have been obtained in the presence of thapsigargin, an inhibitor that locks into the transmembrane domain in a cavity surrounded by helices M3, M5, and M7, reinforcing the proximity of these two structures. Comparing our fitted models (Fit1EUL and Fit1IWO) to the EM density, we looked for the presence of specific density corresponding to thapsigargin in the EM data, but we could not find any in the region surrounded by M3, M5, and M7 (Toyoshima and Nomura, 2002). The presence of vanadate in the crystallization buffer of EM experiments has two consequences: i), orthovanadate, which mimics the phosphate (Bennett and

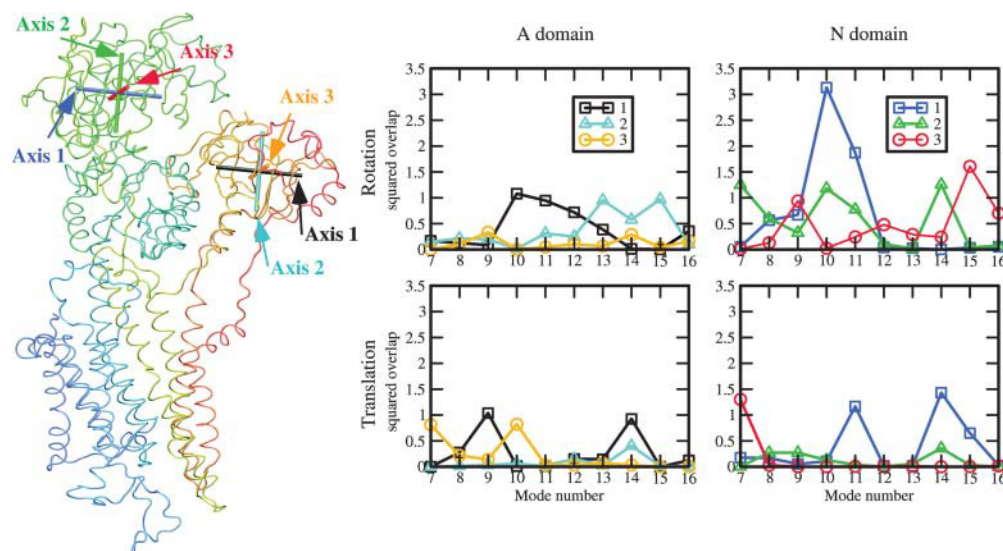


FIGURE 7 Movements of the cytoplasmic domains of the 1IWO structure. (Left) Orthogonal axes used to characterize the movements of domains *A*-actuator and *N*-nucleotide and (right) overlap between the modes and the displacement vectors describing translation along and rotation about these axes.

Dupont, 1982), is probably bound to or close to the phosphorylation site (Asp-351), making this crystallized intermediate closer to a phosphorylated form than to the nonphosphorylated calcium-free enzyme; and ii), decavanadate, which is generally in equilibrium with orthovanadate, is the active agent of crystallization. Two sites of interaction have been identified, a first one mediating an intermolecular contact making the rows of dimers observed within the crystal and a second one mediating intermolecular contacts between the cytoplasmic domains (Stokes and Wagenknecht, 2000). Structures obtained by rigidly docking the domains of 1EUL suggest that this second decavanadate is close to Arg-489, Lys-492, and Arg-678 (Toyoshima et al., 2000). The 1KJU model (Xu et al., 2002) obtained by fitting the density map as described above confirms that decavanadate mediates intermolecular contacts between the cytoplasmic domains, but decavanadate was absent for FITC-labeled ATPase. Our fitted models (Fit1EUL and Fit1IWO) both show that decavanadate is located in a groove. However, whereas in Fit1IWO the side chains of residues Arg-489, Lys-492, and Arg-678 are within or very close to the pronounced density maximum caused by the decavanadate, these residues are further apart in Fit1EUL. Therefore, we repeated the fitting process with an EM density from which this highest maximum has been removed as described in the section, "Preprocessing of the electronic density". In the modified fits, residues Arg-489, Lys-492, and Arg-678 are further apart than in the initial ones. The difference is small for Fit1IWO (a displacement of 1 Å for the C α atoms relative to the initial fit), but more pronounced for Fit1EUL (the C α atoms moved by 2–6 Å). A possible conclusion is that the presence of the electronic density caused by the decavanadate does not affect the structures resulting from our fitting process by much if the initial structure is already close to the target density. Another possible explanation for our findings is the general property of normal mode-based techniques that assign much lower energies to the closing of an open structure than to the corresponding opening of a closed structure. The decavanadate density could thus more easily cause the (open) 1EUL structure to close too much than open up the (closed) 1IWO structure. The limited resolution of the EM data and our fitting method may not permit resolving this question. However, one can safely conclude that the rotational and translational movements we see by normal mode analysis are not caused by the presence of decavanadate, but are a real evidence for conformational changes that occur between nonphosphorylated (1IWO) and pseudophosphorylated (Fit1IWO) conformations of Ca-ATPase.

Delarue and Dumas (2004) have very recently published a fitting technique that is in some respects similar to ours. The main differences are a quality measure defined in reciprocal space and an optimization in normal mode coordinates, using the normal modes of the initial structure. They tested their method on structure factors of known structures, whose resolution was computationally reduced, for targets up to 0.38

nm away from the initial conformation. Suhre and Sanejouand (2004) have used a similar approach for molecular replacement in crystallographical structure determination, i.e., with higher resolution data, and Wu and Ma (2004) have developed a related scheme for use with fiber diffraction data. We expect that our multi-step iterative procedure will perform better for real EM data with its inevitable noise, and for larger deformations such as seen in our example, Ca-ATPase, in particular in the presence of large domain rotations, as rotational motion is not very well approximated by a linear extrapolation of its infinitesimal displacement vectors.

During the final stages of the preparation of this article, we found that Tama et al. (2004) published a fitting method that is similar to ours in principle, though different in many details. However, like Delarue and Dumas, they tested it only on electron densities that were artificially constructed from known atomic structures, thus avoiding the difficulties associated with real experimental densities. One of these difficulties was discussed in the section, "Preprocessing of the electronic density". Another one lies in the necessarily imperfect overlap between a model density that is a sum of Gaussians and an experimental density, which needs to be taken into account in the choice of a minimization algorithm.

CONCLUSIONS

We have presented a general method for fitting protein structures into low-resolution electronic density map obtained from electron microscopy. For isolated proteins, such as our example case, it is simple to apply and has only one essential parameter whose optimal value must be determined empirically. Extensions with more parameters yielding more flexibility and better fits are, of course, possible. For example, one could assign different weights to specific parts of the protein to improve the fit in critical regions. Another possible extension of the method is periodic arrangements of proteins by using periodic boundary conditions.

The use of normal modes, which makes it possible to fit into a low-resolution electronic density, also causes a few limitations. Normal modes describe in principle infinitesimal displacements, which are then often extrapolated linearly, which is also what we do in our algorithm. This works well for most large-amplitude motions, e.g., domain motions, but not so well for rearrangements inside densely packed regions of the protein, such as rearrangements of helices. Such motions are described by modes of much higher energy.

D.L.S. was funded by National Institutes of Health grant R01 GM056960.

REFERENCES

- Beckstein, O., P. C. Biggin, P. Bond, J. N. Bright, C. Domene, A. Grottesi, J. Holyoate, and M. S. P. Sansom. 2003. Ion channel gating: insight via molecular simulations. *FEBS Lett.* 555:85–90.

- Bennett, N., and Y. Dupont. 1982. Vanadate inhibition of the Ca^{2+} -dependent conformational change of the sarcoplasmic reticulum Ca^{2+} -ATPase. *FEBS Lett.* 139:237–240.
- Chacon, P., and W. Wriggers. 2002. Multi-resolution contour-based fitting of macromolecular structures. *J. Mol. Biol.* 317:375–384.
- Chiu, W., A. McGough, M. B. Sherman, and M. F. Schmidt. 1999. High-resolution electron cryomicroscopy of macromolecular assemblies. *Trends Cell Biol.* 9:154–159.
- Cornell, W. D., P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, Jr., D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman. 1995. A second generation force field for the simulation of proteins and nucleic acids. *J. Am. Chem. Soc.* 117:5179–5197.
- Delano, W. L. 2002. The PyMOL Molecular Graphics System. DeLano Scientific, San Carlos, CA. <http://www.pymol.org/>
- 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.
- Frank, J. 2003. Electron microscopy of functional ribosome complexes. *Biopolymers.* 68:223–233.
- Frank, J., T. Wagenknecht, B. F. McEwen, M. Marko, C. E. Hsieh, and C. A. Mannella. 2002. Three-dimensional imaging of biological complexity. *J. Struct. Biol.* 138:85–91.
- Green, N. M., and D. H. MacLennan. 2002. Calcium calisthenics. *Nature.* 418:598–599.
- Hinsen, K. 1998. Analysis of domain motions by approximate normal mode calculations. *Proteins.* 33:417–429.
- Hinsen, K. 2000. The molecular modeling toolkit: a new approach to molecular simulations. *J. Comput. Chem.* 21:79–85. <http://dirac.cnrs-orleans.fr/MMTK/>
- Hinsen, K., A.-J. Petrescu, S. Dellerue, M.-C. Bellissent-Funel, and G. R. Kneller. 2000. Harmonicity in slow protein dynamics. *Chem. Phys.* 261:25–37.
- Humphrey, W., A. Dalke, and K. Schulten. 1996. VMD: Visual molecular dynamics. *J. Mol. Graph.* 14:33–38. <http://www.ks.uiuc.edu/Research/vmd/>
- Karplus, M., and J. A. McCammon. 2002. Molecular dynamics simulations of biomolecules. *Nat. Struct. Biol.* 9:646–652.
- Lee, A. G., and J. M. East. 2001. What the structure of a calcium pump tells us about its mechanism. *Biochem. J.* 356:665–683.
- Ming, D., Y. Kong, M. A. Lambert, Z. Huang, and J. Ma. 2002. How to describe protein motion without amino acid sequence and atomic coordinates. *Proc. Natl. Acad. Sci. USA.* 99:8620–8625.
- Møller, J. V., B. Jull, and M. le Maire. 1996. Structural organization, ion transport, and energy transduction of P-type ATPases. *Biochim. Biophys. Acta.* 1286:1–51.
- Navaza, J., J. Lepault, F. A. Rey, C. Alvarez-Rua, and J. Borge. 2002. On the fitting of model electron densities into EM reconstructions: a reciprocal-space formulation. *Acta Crystallogr. D Biol. Crystallogr.* 58:1820–1825.
- Reuter, N., K. Hinsin, and J.-J. Lacapère. 2003. Transconformations of the SERCA1 Ca-ATPase: a normal mode study. *Biophys. J.* 85:2186–2197.
- Roseman, A. M. 2000. Docking structures of domains into maps from cryo-electron microscopy using local correlation. *Acta Crystallogr. D Biol. Crystallogr.* 56:1332–1340.
- Rossmann, M. G. 2000. Fitting atomic models into electron-microscopy maps. *Acta Crystallogr. D Biol. Crystallogr.* 56:1341–1349.
- Saibil, H. R. 2000. Conformational changes studied by cryo-electron microscopy. *Nat. Struct. Biol.* 7:711–714.
- Stokes, D. L., and N. M. Green. 2003. Structure and function of the calcium pump. *Annu. Rev. Biophys. Biomolec. Struct.* 19:445–468.
- Stokes, D. L., and T. Wagenknecht. 2000. Calcium transport across the sarcoplasmic reticulum: structure and function of Ca^{2+} -ATPase and the ryanodine receptor. *Eur. J. Biochem.* 267:5274–5279.
- Suhre, K., and Y.-H. Sanejouand. 2004. On the potential of normal-mode analysis for solving difficult molecular replacement problems. *Acta Crystallogr. D Biol. Crystallogr.* 60:769–799.
- Tama, F., O. Miyashita, and C. L. Brooks. 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.
- Tama, F., W. Wriggers, and C. L. Brooks. 2002. Exploring global distortions of biological macromolecules and assemblies from low-resolution structural information and elastic network theory. *J. Mol. Biol.* 321:297–305.
- Toyoshima, C., M. Nakasako, H. Nomura, and H. Ogawa. 2000. Crystal structure of the calcium pump of sarcoplasmic reticulum at 2.6 Å resolution. *Nature.* 405:647–655.
- Toyoshima, C., and H. Nomura. 2002. Structural changes in the calcium pump accompanying the dissociation of calcium. *Nature.* 418:605–611.
- Toyoshima, C., H. Sasabe, and D. L. Stokes. 1993. Three-dimensional cryo-electron microscopy of the calcium ion pump in the sarcoplasmic reticulum membrane. *Nature.* 362:467–471.
- Unwin, N. 2003. Structure and action of the nicotinic acetylcholine receptor explored by electron microscopy. *FEBS Lett.* 555:91–95.
- Valadie, H., J.-J. Lacapère, Y.-H. Sanejouand, and C. Etchebest. 2003. Dynamical properties of the MscL of *Escherichia coli*: a normal mode analysis. *J. Mol. Biol.* 332:657–674.
- van Rossum, G., and Python Software Foundation. 1991. Python programming language. <http://www.python.org/>
- Volkman, N., and D. Hanein. 1999. Quantitative Fitting of Atomic Models into Observed Densities Derived by Electron Microscopy. *J. Struct. Biol.* 125:176–184.
- Wriggers, W., and P. Chacon. 2001. Modeling tricks and fitting techniques for multiresolution structures. *Structure (Camb).* 9:779–788.
- Wu, X., J. L. S. Milne, M. J. Borgnia, A. V. Rostapshov, S. Subramaniam, and B. R. Brooks. 2002. A core-weighted fitting method for docking atomic structures into low-resolution maps: application to cryo-electron microscopy. *J. Struct. Biol.* 141:63–76.
- Wu, Y., and J. Ma. 2004. Refinement of F-actin model against fiber diffraction data by long-range normal modes. *Biophys. J.* 86:116–124.
- Xu, C., W. J. Rice, W. He, and D. L. Stokes. 2002. A structural model for the catalytic cycle of Ca-ATPase. *J. Mol. Biol.* 316:201–211.
- Young, H. S., and D. L. Stokes. 2004. The mechanics of calcium transport. *J. Membr. Biol.* 198:55–63.
- Zhang, P., C. Toyoshima, K. Yonekura, N. M. Green, and D. L. Stokes. 1998. Structure of the calcium pump from sarcoplasmic reticulum at 8 Å resolution. *Nature.* 392:835–839.