Published in final edited form as:

J Struct Biol. 2008 August ; 163(2): 163–174. doi:10.1016/j.jsb.2008.05.001.

YUP.SCX: Coaxing Atomic Models into Medium Resolution Electron Density Maps

Robert K.-Z. Tan, Batsal Devkota, and Stephen C. Harvey¹

School of Biology, Georgia Institute of Technology, Atlanta, GA 30332-0230

Abstract

The structures of large macromolecular complexes in different functional states can be determined by cryo-electron microscopy, which yields electron density maps of low to intermediate resolutions. The maps can be combined with high-resolution atomic structures of components of the complex, to produce a model for the complex that is more accurate than the formal resolution of the map. To this end, methods have been developed to dock atomic models into density maps rigidly or flexibly, and to refine a docked model so as to optimize the fit of the atomic model into the map. We have developed a new refinement method called YUP.SCX. The electron density map is converted into a component of the potential energy function to which terms for stereochemical restraints and volume exclusion are added. The potential energy function is then minimized (using simulated annealing) to yield a stereochemically-restrained atomic structure that fits into the electron density map optimally. We used this procedure to construct an atomic model of the 70S ribosome in the pre-accommodation state. Although some atoms are displaced by as much as 33 Å, they divide themselves into nearly rigid fragments along natural boundaries with smooth transitions between the fragments.

Keywords

Electron microscopy; simulated annealing; structural refinement

1. Introduction

Cryo-electron microscopy is an invaluable method for the determination of the structures of large macromolecular complexes (Frank, 1996; Saibil, 2000). The complexes can be captured in different conformations or functional states by chemical, genetic and physical means (Frank, 2001). The electron density maps obtained by this technique are of low to intermediate resolutions. However, it is possible to combine atomic structures, obtained from X-ray crystallography or NMR, with electron density maps of lower resolution, to obtain a structure that is more accurate than the resolution of the maps (Rossmann, 2000; Rossmann et al., 2001). This is analogous with macromolecular crystallography in which the higher resolution data come from small molecules.

In many cases, the atomic structures of the components of the complex are not known. Procedures to detect alpha helices (Jiang et al., 2001) and beta sheets (Kong, et al., 2004) in intermediate-resolution density maps can be used to create tentative atomic models. A

¹ Correspondence: E-mail: steve.harvey@biology.gatech.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

procedure to recognize structural homologues of protein domains in intermediate-resolution maps and to construct quasi-atomic models of the assembly has been demonstrated (Lasker, et al., 2007). Comparative modeling can also be used to create the atomic models. Topf, et al. (2006) enhanced the comparative modeling procedure by also optimizing the fit of the models into the density map of the target sequence and found that the alignment errors decrease dramatically.

Many methods have been developed to dock atomic structures into electron density maps and to refine them (see Fabiola and Chapman, 2005 for a review). Oftentimes, a structure can simply be manually docked in the electron microscopy (EM) map as a starting point for refinement. However, more quantitative methods are required for low to intermediate resolution and symmetrical maps. The EMfit program (Rossmann, 2000; Rossmann et al., 2001) uses a systematic grid search: each rotational and translational parameter is adjusted in turn, in decreasing intervals until convergence. The 3SOM method (Ceulemans and Russell, 2004) maximizes the overlap of the surfaces of the atomic model and the EM map. By considering only a few key vectors that capture local surface features, the complexity of the problem is reduced. Similarly, the Situs (Wriggers et al., 1999) package of programs employs codebook vectors to reduce the combinatorial complexity of the structural comparison. Chacon and Wriggers (2002) have also shown that by also considering a contour-matching criterion in addition to the cross-correlation (of the generated density of the atomic model and the actual density of the map), the contrasts between the potential solutions can be enhanced.

As the resolutions of EM maps improve there is an interest in flexible fitting and refinement methods. Flexible fitting and particularly refinement methods are too slow for interactive use and require a starting structure obtained using one of the search methods described earlier.

Situs can be used for flexible fitting in which the match between the model and the electron density map is optimized using molecular mechanics refinement [Wriggers et al., 2000]. RSRef is a real-space refinement method, originally developed for crystallographic data (Chapman, 1995), and then adapted for EM data (Chen et al., 2001; Fabiola and Chapman, 2005). The central idea is to optimize the structural fit subject to a stereochemical restraint, thus yielding a stereochemically-correct structure that is optimally docked in the electron density map. The atomic model is divided into rigid bodies and several optimization methods can be used. In principle, the refinement can also be done in all-atom basis. Velazquez-Muriel et al. (2006) have devised a method that uses the evolutionarily related structural variability among the domains of the protein superfamily to generate articulated and stereochemically correct models that are then fitted into the density map. The DireX program (Schröder, et al., 2007) generates an ensemble of structures that fit a given density map. The procedure samples conformational space, restrained to an approximate structure by a deformable elastic network, and constrained by the low-resolution density map. This reduces overfitting especially at low resolution. Instead of incorporating the cross-correlation (of the experimental and model densities) directly into the molecular mechanics force field, an approximate stochastic approach is used.

NMFF (Tama et al., 2004) is a suite of programs for the quantitative flexible docking of high-resolution structures into low-resolution electron density maps. An atomic model is deformed along a linear combination of the low-frequency normal modes. An elastic network model of the molecule is used to calculate the normal modes economically. The objective is to maximize the cross-correlation between the density of the map and the simulated density calculated from the model. The direction of the deformation is given by the derivative of the cross-correlation function with respect to the amplitude. The deformation is then carried out iteratively with a maximum displacement of 2 Å. Optimization starts with steepest descent and switches to Newton-Raphson as the solution approaches a local maximum.

In this paper, we describe a method for the refinement of an atomic model in an electron density map of intermediate resolution. The electron density map is converted to a coaxing function that attracts the atoms of the model into the high-density regions of the map. Simplified functions for stereochemical restraint and volume exclusion are added to the coaxing function to form a complete potential energy function. The conformation with the global minimum value of the potential energy function is a stereochemically-correct atomic model that fits optimally into the electron density map. However, because of our approximations, we cannot expect to obtain an exact solution. A global optimization method is used so as to minimize the risk of local minimum traps. This procedure has been implemented as part of YUP (Tan et al., 2006), a general-purpose molecular mechanics program. The program is extremely easy to use and yet provides the user with many customization opportunities.

2. Materials and methods

The refinement procedure is summarized in Figure 1. The required inputs are a target electron density map and the atomic structures of the components that are to be placed in the map. The electron density map is used to construct the *SCX* term (section 2.1) of the YUP.SCX force field. YUP.SCX is a refinement method, not a search procedure and therefore requires the atomic structures to be approximately positioned in the target map. An external search method may be used, or the structure may be positioned interactively using a molecular graphics program. The pre-fitted structure is then used to construct the *GNM* (*Hooke* in section 2.2) and *SSX* (section 2.3) terms to complete the YUP.SCX force field. Starting from the pre-fitted conformation, the model is then refined (section 2.4), using the YUP.SCX force field, to yield the optimized structure.

2.1. The SwissCheeseX (SCX) function

The purpose of the *SwissCheeseX* function is to steer the atoms of the model into the density map. A rigorous implementation would likely start from the electron density or the cross-correlation between the experimental densities (of the map) and the densities calculated for the atomic model. Instead, we use an approximate expression that can be easily coupled with the optimization procedure (see below). Called *SCX* for short, it is an empirical energy function obtained from the density map. It turns the voxels of the map into points in space that attract nearby atoms. The strength of the interaction between a voxel and an atom is scaled according to the density of the voxel and the distance between the voxel and the atom.

The *SCX* function replaces each (prismatic) voxel of the density map with a spherical well of equal or larger volume. If the length of a cubic voxel is L , the well that represents it will have

a minimum or inner radius $r_i = \sqrt[3]{\frac{3}{4\pi}}L$. Each well attracts all the atoms that are within a distance $r_o \geq r_i$. The inner radius is a fixed value, while the maximum or outer radius r_o can be set to any larger value. The wells overlap when $r_o = r_i$, except where the vertices (of the original

voxels) meet. There are no dead spaces as long as $r_o = \frac{\sqrt{3}}{2}L$ or larger. Thus, with the radius

ratio $\frac{r_o}{r_i} = \sqrt[3]{\frac{\sqrt{3}\pi}{2}} \approx 1.396$, the original map is completely covered by the *SCX* function, with the minimum amount of overlap among the wells.

The total *SCX* energy, E^{SCX} is the sum of the interactions between all the atoms of a model and all the wells of a map. For an atom p at a distance r_{pq} from a well q , the energy is given by a continuous function, defined piecewise as:

$$E_{pq}^{\text{SCX}} = \begin{cases} -E_0\omega_q, & r_{pq} \leq r_i \\ -E_0 \frac{(r_o - r_{pq})^2 (r_o + 2r_{pq} - 3r_i)}{(r_o - r_i)^3} \omega_q, & r_i < r_{pq} < r_o \\ 0, & r_{pq} \geq r_o \end{cases} \quad (1)$$

where ω_q is the normalized density (see section S2 of the Supplement) at the well q and E_0 is the magnitude of the minimum energy that is assigned to the voxel with the maximum density. Equation (1) describes a flat-bottomed well with smoothly curving sides between r_i and r_o . Note that the actual implementation makes use of a slightly more general expression (see section S1 of the Supplement or <http://rumour.biology.gatech.edu/YammpWeb/techdoc/energy/swchx/intro.shtml>).

Thus, an atom is accorded the minimum energy (more negative) from the home voxel (in which it is currently located) and increased energy (less negative) from neighboring wells whose centers are within r_o from the atom. On the other hand, the atom experiences no force (negative of the derivative of the energy with respect to the displacement) from the home well. The neighboring wells exert attractive forces on the atom, with the largest contributions coming from the closest and deepest wells.

Figure 2 shows a sample density distribution along one dimension, and SCX functions that are derived from it at four different radius ratios. As the radius ratio is increased, the SCX function is increasingly smooth and the overall energy is lowered everywhere except at the boundaries.

2.2. The Gaussian Network (Hooke) term

Instead of a conventional molecular mechanics force field, a highly modified Gaussian Network Model (GNM) is used to represent the atomic structure. The GNM is much simpler to set up and the calculations are quicker than if a conventional force field is used. Missing, multi-occupancy and non-standard atoms or residues are handled transparently by a GNM while a conventional force field can be built for such cases only if the appropriate procedures are in place.

The *Hooke* term in the YUP package is used to represent the elastic network:

$$E_{ij}^{\text{Hooke}} = k_{\text{Hooke},ij} (r_{ij} - b_{0,ij})^2, b_{0,ij} < D_{0,ij} \quad (2)$$

$k_{\text{Hooke},ij}$ is a force constant, $b_{0,ij}$ is the distance between atoms i and j when the network is formed, while r_{ij} is the subsequent distance. A link is added to the network only when the initial separation distance is shorter than a cutoff length $D_{0,ij}$.

The GNM or elastic network model was originally devised for normal mode analysis (Tirion, 1996). This idea was adopted as a simplified force field for molecular mechanics calculations (see for example Cui, et al., 2006). Usually, only representative atoms (e.g. alpha carbons or phosphorus) are connected into the elastic network, all such atoms within a single cutoff distance are connected, and only one force constant is used for the entire network.

We use a highly modified GNM. The elastic network is formed from all atoms. An atom pair is added to the network if the separation distance is within the cutoff length, which varies according to atom type. For example, it is 22.5 Å between two phosphorus atoms, 8 Å between two C_α atoms and 4 Å for other pair types that do not have a specific cutoff. The force constant

is assigned according to the distance between the atoms if they are entirely within a residue or they span two consecutive residues of a chain (see below). The largest force constant is assigned to the closest atom pairs and weaker constants are assigned to widely separated atoms. The force constant diminishes quickly as the link length increases. The weakest force constant is also assigned to atom pairs that are not in the same molecule regardless of the distance that separates them. With these modifications, we have a model that is locally stiff but is globally flexible. Compared to the standard model, the modified GNM can easily adopt alternate conformations.

There are a few problems with the GNM. The starting structure is at the global minimum and any conformational change has to result in a higher energy. All links of the same length are of equal strength regardless of atom types. Thus, the refolded structure is likely to have bonds that fall outside the standard range of lengths. This deficiency extends to angles and dihedrals. Finally, it is not always possible to maintain chirality with only distance constraints, as is the case in the GNM. None of these problems is serious if only modest conformational changes are required to fit the density map, but if large changes are required, the refolded structures are not always completely stereochemically-correct. Fortunately, these errors are easily corrected with a few cycles of minimization under a conventional force field.

2.3. The soft sphere (SSX) term

The *SoftSphereX* term from the YUP package, or *SSX* for short, is used to represent the non-bond interactions, i.e., interactions between pairs of atoms that are not linked by the elastic network:

$$E_{ij}^{SSX} = \begin{cases} k_{SSX} (r_{ij} - d_{0,ij})^2 & r_{ij} < d_{0,ij} \\ 0, & r_{ij} \geq d_{0,ij} \end{cases} \quad (3)$$

k_{SSX} is the force constant, r_{ij} is the distance between atoms i and j , and $d_{0,ij}$ is the contact distance. The last is set to the sum of the van der Waals radii (Bondi, 1964) of atoms i and j .

Atom pairs that are part of the elastic network that are either entirely within one residue or which span two consecutive residues, and which have link lengths shorter than 110% of the contact distances ($b_{0,ij} < 1.1d_{0,ij}$), are excluded from the *SSX* term.

2.4. Structural optimization

We now have a potential energy function that contains the *Hooke* term to approximately restrain stereochemistry for the model, the *SSX* term to maintain non-bond separation, and the *SCX* term to coax the model into the target density map:

$$E = \sum_i \sum_{j=i+1}^{N_{\text{atom}}-1} E_{ij}^{\text{Hooke}} + \sum_i \sum_{j=i+1}^{N_{\text{atom}}-1} E_{ij}^{\text{SSX}} + \sum_p \sum_q^{N_{\text{atom}}N_{\text{well}}} E_{pq}^{\text{SCX}} \quad (4)$$

The constant E_0 in Equation (1) is used to control the balance between the *SCX* and the other terms. The structure at the global minimum of the exact function will be the optimal refolding of a stereochemically-correct model into the target map. A gradient-based minimization method could be used but all such methods can only locate a local minimum close to the starting conformation of the model.

We find Simulated Annealing (SA) to be the most effective optimization method. It was introduced in 1983 (Kirkpatrick et al., 1983) as a Monte Carlo-based method. It was then adapted for Molecular Dynamics for the refinement of structures using NMR and crystallographic data (Brunger et al., 1987; Brunger, 1988). For example, the structure factor amplitudes can be incorporated into a function that is then added to the XPLOR potential energy function (Brunger, 1988). A molecular dynamics simulation of the structure is carried out under the influence of the augmented force field. The molecule is heated to a very high temperature (determined empirically) and is allowed to explore the expanded conformation space at the elevated temperature. Thus, the molecule has an opportunity to escape any local minimum in which it may be initially trapped. Then, the molecule is slowly cooled to channel it into a low energy conformation. The desired goal is the global minimum conformation. However, the outcome of the annealing procedure is stochastic and attainment of the global minimum is not guaranteed. The molecular dynamics module from the YUP package was customized to provide a SA function in which both the temperature and the radius ratio can be varied during a molecular dynamics simulation. This fitting method is implemented as an easy-to-use program: `Yup.scx`. We call this procedure YUP.SCX after the name of the program.

Our objective function (Equation 4) is dominated by the SCX function (Equation 1), which is increasingly smooth as the radius ratio is increased. Thus, the maximum temperature is not the driving factor in the refinement procedure. It only has to be high enough for the atoms to move at a decent speed, so that the necessary conformational changes are completed quickly. On the other hand, a very high temperature introduces excessive thermal fluctuations, and a longer cooling cycle is then required to remove them. We found the balance between these competing factors at 10 K. A more important factor in our SA method is the radius ratio r_o/r_i . It can be adjusted to change the range of each well and the smoothness of the objective function.

The simulated annealing procedure is followed by energy minimization with the radius ratio set to 1.0.

2.5. Test problems

Six problems were devised to test the YUP.SCX fitting procedure. The goal in each case is to flexibly refine a pre-fitted structure to optimally fit into a density map. The steps taken to generate the starting structures and target maps are summarized in Table 1. The four Kinase cases: **K145**, **K149**, **K415** and **K419**, are to be fitted to simulated maps, with known or exact solutions. The extended and compact conformations are manually positioned so that they overlap. Maps are then generated from the extended conformation at two resolutions to be used as targets for the compact conformation (**K145/9**). Similarly, maps are generated from the compact conformation as targets for the extended conformation (**K415/9**). Consequently, the starting structure for **K145/9** is also the exact solution to the **K415/9** cases, and the starting structure for **K415/9** is the exact solution to **K145/9**. **K145/9** are models for fitting problems that require a compact structure to be extended in order to fit the target map; **K415/9** model the opposite conformation change. The two Ribosome cases: **30S** and **50S**, are based on real maps obtained from cryo-electron microscopy measurements. Each of the Ribosome model is manually positioned within the map and then computationally fitted into the target as a single rigid unit.

Simulated maps are generated in the XPLOR/CNS format using the `pdb2sim_map` program in the NMFF suite. The program generates incorrect map indices (first line after the remarks), which we rectify by trial-and-error. For only the map generated for the **K415** test problem, one side was manually padded with zero density layers. This is required by NMFF, which we used as a comparative method. The padding is not necessary for, nor does it have any effect on the YUP.SCX procedure.

The quality of the fitted structures is expressed as two numbers: CCC and RMSD. CCC is the Cross Correlation Coefficient between the maps (simulated or real) of the structures under comparison. RMSD is the Root Mean Square Deviation between the two structures that are being compared. In the Kinase problems, the RMSD is calculated between the fitted structure and the exact solution, since the latter is known. A perfect solution will have a RMSD of 0. In the Ribosome problems, the RMSD is calculated between the fitted structure and the initial structure. All we can say for these cases is that a structure with a larger RMSD has moved further from the initial structure than another structure with a smaller RMSD value.

Cross correlation coefficients (CCC) are calculated using the `corr_coef` program in the NMFF suite.

The RMSD calculator in VMD (Humphrey et al., 1996) is used to align and superimpose structures and to calculate the resulting root mean square deviation (RMSD). RMSD° is the absolute root mean square deviation, calculated without prior alignment and superposition, using the VMD RMSD calculator or a custom script written in Python. All the atoms of the model are used in the superposition and the calculation of RMSD and RMSD° . The distinction between RMSD° and RMSD is important only for Table 9 (see below). Structures that have been correctly and completely fitted into an electron density map are effectively superimposed onto the structure to which it is being compared. Thus, RMSD° and RMSD are interchangeable (except in Table 9). For example, between the initial structure and the best solution to the 50S problem, RMSD° is 4.42 Å and RMSD is 4.28 Å.

Molecular graphics were produced using the UCSF Chimera (Pettersen et al., 2004) package from the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco.

2.6. Evaluating the SCX term

The SCX term is evaluated from a periodically updated list of atom-well pairs that were within the cutoff distance. Please see section S2 of the Supplement for details.

2.7. Establishing the best settings

With the molecular dynamics based SA method, we have to heat the model up to a maximum temperature and then cool it down to a low temperature. Maximum temperatures of 10 K, 100 K and 1000 K were tried and it was quickly obvious that the lowest value is the best one to use. The radius ratio provides for very effective flattening of the potential function and therefore temperature is not an important parameter in the YUP.SCX procedure. The force constants for the Hooke term are listed in Table 2. Other settings are discussed in section S3 of the Supplement.

2.8. Optimum duration for simulated annealing

One of the most important settings for the SA procedure is the duration of the cooling stage. The **K419** problem was solved with different durations achieved by using step sizes of 1 fs, 2.5 fs and 5 fs, with the appropriate number of steps. The CCC and RMSD° of the fitted structures are plotted against duration in Figure 3. The exact solution would have a final CCC of 1 and RMSD° of 0. The highest CCC achieved in these calculations is 0.993; the smallest RMSD° is 1.56 Å.

As expected, the CCC increases and the RMSD° decreases as the duration is increased. However, no further improvement is obtained for durations over about 10 ps. More significantly, nearly the same results can be obtained at any given duration regardless of the step size used. Thus, it is the duration and not the number of steps that matter. We would want

to use the largest step size permissible to reduce the amount of computation. The same trend is found in the **K149** test problem, except that no further improvement in the fitting is obtained once the duration exceeds 5 ps (Figure S1).

We developed a rule-of-thumb for the overall duration of the procedure: it is the number of atoms rounded to the nearest 1000, and multiplied by a user specified number (*stepfactor*). In the **K149** and **K419** problems, the shortest duration that produces the highest CCC are 5 ps and 10 ps respectively. The corresponding *stepfactors* are 0.5 and 1. From limited testing, we find the optimal value for the Ribosome problems to be 0.2 to 0.5. Thus, the default value of *stepfactor* is set at 0.5. There is no physical basis for this formula, but it is a useful starting point and the optimal value for a new system is likely to be in the neighborhood. Note that a *stepfactor* of 2 is used in most of the Kinase runs, because many of them were done before the duration study, and subsequent runs used the same *stepfactor* for consistency.

2.9. Consistency of the simulated annealing procedure

The simulated annealing procedure is not deterministic: the outcome depends on the initial velocity distribution even if the annealing conditions are held constant. For each of the Kinase cases, we carried out multiple fittings from the same starting structures under the same default conditions (except the SA duration which was four times the default, i.e. *stepfactor* = 2). Only the initial velocity distribution varies from one run to the next. The results are then sorted into clusters and are summarized in Table 3.

The results for **K149**, **K415** and **K419** are very consistent. In each case, the fitted structures are clustered tightly together as indicated by the small RMSD° (second column) between the members of each group. This is also reflected in the small deviations in the CCC and the RMSD ° between the fitted and solution structures (rightmost columns). In particular, all 25 fittings of the **K149** case are (consistently) wrong.

The results from the 25 solutions to the **K145** problem can be clustered into four groups with one outlier. One representative (with median CCC) from each group is depicted in Figure 4 (top left) in different shades of blue, along with the exact solution in yellow. These structures differ from each other mainly in the stretch of residues approximately from ASP33 to GLY80. All four structures are consistent in having the residues from LYS50 to VAL59 located away from the exact solution. The four representatives agree remarkably well on the placement of the remaining residues and even the side chains match well with the exact solution. This suggests the existence of at least four closely related local minima for this problem.

For the **K415** and **K419** problems, a single run is likely to result in a structure that has a CCC within 0.001 of the best score selected from 10 runs. For the **K145** and **K149** test problems, any single fitting is likely to produce a result that has a CCC within 0.025 of the best score achieved from 25 runs. These calculations show that, in spite of the stochastic nature of SA, there is a good chance of obtaining a representative solution to a Kinase problem from just one attempt, even for the **K145** problem.

2.10. Some consequences of the approximate force field

Table 4 lists the details of three refinement procedures carried out with the Kinase test cases. In the **Null** procedure, the ideal conformation is minimized to fit the map generated from itself. If the force field is exact, the initial conformation will be at the global minimum and the optimization should converge immediately. As expected, the *SCX* term has a low and negative value indicating good placement of the atoms in the density map, and the *Hooke* term (GNM in Table 4) is zero at the start of the refinement. (Note that the subsequent drop in the *SCX* value is accompanied by a change in the conformation.) However, the *SSX* term has a large

starting value, indicating close contacts. The refinement removes most of the close contacts but at the expense of the other terms. At least for the Kinase test problems, the *SSX* approximation results in a degradation of no more than 0.02 in the CCC and a shift of 0.17 to 0.28 Å away from the exact solution.

In the **Standard** procedure, a starting conformation is refolded to fit into an unrelated target map. The *Hooke* term is constructed from the initial structure, hence the zero GNM value at the start of the calculation. In the **Reverse** procedure, the starting conformation is the exact solution (note the extremely low starting values for the *SCX* term), but the *Hooke* term is calculated from the starting structure used for the same problem in the corresponding **Standard** procedure (hence the very large starting GNM values). Thus, the two procedures start the refinement from opposite ends of the folding pathway. If the force field is exact, the **Reverse** procedure should converge immediately. The two procedures produce comparable results for the **K145** and **K149** cases (see below for a discussion of the **K145** and **K149** problems). Note the very similar CCC and RMSD°, which we confirmed by visualizing the structures together. As noted before, any refolding of the conformation used in the construction of the elastic network causes the value of the *Hooke* term to increase. Thus, the approximations in the *Hooke* and *SSX* terms cause the **Reverse** procedure to miss the exact solution by 1.10 to 1.60 Å. This is below the error that one can expect from fitting a structure into a map of intermediate resolution.

With the Kinase **K145** and **K149** problems, the initial compact structure has to be refolded into maps of an extended structure. This is difficult because the compact structure has a dense elastic network. As a result, many short links have to be stretched to change the conformation. These problems were first solved using a conventional GNM, with a single cutoff $D_{0,ij}$ equal to 3.5 Å for all atom pairs, and a single force constant $k_{Hooke,ij}$ of 29.808 kcal/mol/Å². The SA duration was four times the default. The fit quality is summarized in Table 5 (Method A). The low final CCCs indicate poor fits, although the calculations for the **K145** problem have made more progress towards the target as the smaller RMSD° indicates. Setting a lower force constant or reducing the cutoff could improve the solutions. However, this is at the expense of correct stereochemistry. The highly modified GNM (see above) tries to reconcile the differences. It approximates a conventional force field but retains the simplicity of the elastic network model.

Although better than the standard GNM, the modified model is still imperfect as the comparison between the **Standard** and **Reverse** results for the **K145** problem shows (Table 4). The **Standard** procedure is unable to reach the optimal solution, which is presumably reached by the **Reverse** procedure. However, the **K149** results from the two procedures are comparable. Note that the **Standard** procedure is always able to produce a lower energy conformation than the corresponding **Reverse** procedure, suggesting that the Simulated Annealing procedure is working properly.

2.11. Implementation and availability of YUP.SCX

Implementation details are discussed in section S4 of the Supplement.

An early version of YUP.SCX, known as *VLAT* (unpublished), is available in Yammp 1 (Tan and Harvey, 1993), an older version of YUP, and had been used with some success (Mears et al., 2006; Stagg et al., 2006). YUP, user manuals, initial structures and corrected maps for the Kinase problems, and a tutorial can be found at <http://rumour.biology.gatech.edu/YammpWeb/>.

3. Results

The default parameters and settings listed in **Materials and methods** are used for the fittings and exceptions will be noted.

3.1. Incorrect fitting of K149

The Kinase **K149** problem was solved using the default GNM (highly modified) and default radius ratio (4), using four times the default SA duration. The results are shown in the Method B line of Table 5. The incorrectly refined structure (blue) is shown in Figure 4 (bottom left) within the target map, along with the starting structure (light blue) and the exact solution (yellow). A large stretch of the chain, PRO27 to ASP76 has been moved into the vertical finger of density instead of into the smaller horizontal finger. The location of GLU44 is indicated for all three structures. The default radius ratio of 4 translates to an outer radius of 6.204 Å. Clearly the starting structure is very close to the vertical finger. Thus, at this setting, it is inevitable that the PRO27-ASP76 stretch is drawn towards the vertical finger and not towards the correct target.

3.2. Solutions to the Kinase problems

Representative solutions to the Kinase problems are summarized in Table 6 (*YUP.SCX* lines) and depicted in Figure 4 (middle and right columns).

The representatives for the **K145** (Figure 4, top center), **K415** (bottom center) and **K419** (bottom right) problems are the median CCC structures from the multiple runs discussed previously. The **K149** (top right) structure comes from a single *YUP.SCX* refinement using default settings except for the radius ratio (3 instead of 4) and SA duration (four times the default). At the smaller ratio, the outer radius is 4.653 Å, thus reducing the influence of the vertical finger of density. The **K145** problem was solved successfully at the higher default ratio; the outer radius is only 1.241 Å because of the finer grid of the target map.

For comparison, the Kinase problems were also solved using NMFF (results in the NMFF lines of Table 6). We have experimented with different NMFF settings to optimize its performance. In order to solve the **K415** problem with NMFF, the cutoff distance for the GNM model had to be reduced from 5.5 to 4.5 Å.

It is much easier to refold an extended structure to fit into the map of a compact conformation (**K415** and **K419**) than it is to go in the other direction (**K145** and **K149**). The **K415** and **K419** problems could be solved even with the use of the conventional GNM; the initial conformation is extended and so the elastic network is relatively sparse and therefore easier to deform.

The NMFF solutions to the **K145** and **K149** problems resemble the (incorrect) *YUP.SCX* solutions using the conventional GNM (Method A of Table 5). However, *YUP.SCX* provides much better results with the use of the modified GNM (Table 6). The limitations of the conventional GNM in NMFF can be overcome by doing the refinement in stages, with a freshly built elastic network at each stage. The elastic network gets sparser at each stage and is therefore progressively easier to deform.

The correct *YUP.SCX* solution to the **K145** problem is of fairly good quality, with the CCC reaching above 0.96 while the fitted structure comes within 2.1 Å of the exact solution. The helix is out of registration from about SER41 to ASP51, while the loop from ASP51 to VAL59 is flipped out of the high-density region of the map. The fit for the **K149** case is very similar.

For the solutions to the **K415** and **K419** problems, the CCCs exceed 0.98 and the RMSD°s are lower than 1.6 Å. The most significant deviations are at the residues GLU44 to GLY46. The solutions from NMFF (which uses a conventional GNM) are of nearly the same quality.

NMFF is much slower than YUP.SCX when fitting into the higher-resolution maps (**K145** and **K415**). However, it should be noted that the NMFF procedure was developed for medium- and low-resolution maps, from 10 to 30 Å.

Table 6 also shows summaries of the validation of the stereochemistry carried out using the server at the RCSB web site. The stereochemistry is surprisingly well maintained given the approximate nature of the GNM.

3.3. Optimum radius ratio for the Ribosome problems

The ribosome problems are much larger than the Kinase cases. Thus, it is impractical to solve each problem exhaustively. In these problems, the all-atom crystallographic structure of the non-functional form of the 70S Ribosome is fitted into a cryo-electron density map of the Ribosome that had been trapped in the pre-accommodation stage. As we shall see, the fitting results in large movements of a small number of (mostly surface) components. In calculating the measure of quality, the large but local changes are averaged out over a largely unchanged complex. Thus, the CCC and RMSD° are not very discriminating metrics for these problems.

The ribosome components were fitted at different radius ratios, and the results are summarized in Table 7. (These results closely resemble another set obtained from earlier runs carried out with slightly different settings.) The quality of the fit increases consistently with the radius ratio for the 50S and it is possible that better results could be obtained with even larger ratios. However, the best results for the 30S are obtained at a ratio of 4 and the quality of the results deteriorate with higher ratios.

The inner radius is 1.749 Å for the Ribosome maps. Thus, radius ratios 2.5 and 5.0 in Table 7 correspond to outer radii of 4.373 Å and 8.747 Å respectively. The best and worst structures in each case were compared. The largest deviation for the 30S is 8.81 Å, and there are only 50 atoms that deviate more than 4 Å between the two extreme structures (refined at $r_o/r_i = 2.5$ and 4.0 respectively). All these atoms are in the ribosomal proteins (mostly S2 and S4). The largest deviation between the extreme structures of the 50S is 16.37 Å and there are many more deviations that exceed 4 Å than in the 30S. The deviations between the extreme structures (refined at $r_o/r_i = 2.5$ and 5.0) that exceed 4 Å in the 50S are in the proteins (mainly L9) and the L1 binding domain of the 23S RNA.

Although the changes in the CCC are very small, the trends are consistent. The RMSD° trend is also largely consistent. When local deviations are added to these observations, we get the strong suggestion that the optimal radius ratio for the **30S** problem is no more than 4, but the **50S** problem could benefit from a ratio higher than 5.

3.4. Fitting the 30S and 50S components of the 70S Ribosome

The best solutions are summarized in Table 8 along with the results from parallel attempts to perform these fits using NMFF. The RMSD° is between the initial and the refined structures, since the exact solutions are not known. The structures obtained from YUP.SCX are depicted in the top panels of Figure 5 (30S on the left and 50S on the right). Although, the YUP.SCX procedure is able to improve the CCC from 0.82 to 0.89 in both cases, the final CCC is still below 0.9. Most of the shortfall comes from the large volume of density that has no corresponding atoms. For example, the atomic structure for the 50S does not include the L1 ribosomal protein (Berk, et al., 2006).

The NMFF fit of the 30S terminated not far from the initial structure, because the gradient of the NMFF objective function (CCC) was changing too slowly. Although setting a lower tolerance would allow the procedure to continue, the calculation would take an impractically long time and it was not clear if a better solution would ultimately be found. The NMFF fit of the 50S case terminated on the first iteration, and was not continued for the same reasons. The NMFF procedure can be applied to the 30S problem again, by building a new elastic network from the outcome of the first fitting and continuing from there. Obviously, this is not possible for the 50S. Clearly, the landscape of the NMFF objective function (CCC versus normal modes) is very flat in these problems and it is not practical to use a gradient-following algorithm to solve them. (Since the fitting of the Ribosome components involve the movement of isolated components, and therefore local motions, the NMFF procedure may have to use the higher-frequency normal modes in order to solve these problems.)

Figure 5 (top left) shows the initial and fitted 30S structures within the target map, and five regions with the largest displacements between the initial and final structures are also indicated. Not unexpectedly, the largest displacements are in the peripheral regions of the 30S complex. A list of displacements that exceed 4 Å was generated, and none of these are in the interior of the complex. Large displacements of the 16S RNA occur at the spur (including G86) and the beak (including G1032).

The initial and fitted structures of the 50S are shown in Figure 5 (top right) within the target map. The large displacements are mainly at the top of the structure. The largest displacement is over 33 Å and it occurs in the L1 binding domain (G2159 is indicated). The ribosomal protein L9 was also displaced by over 21 Å (at GLU109). Note the significant unoccupied offshoot in the map at the L1 binding domain. The density is almost certainly of the L1 ribosomal protein, which was not resolved in the crystallographic structure and is not included in our model.

In both the 30S and 50S, most of the RNA and protein atoms are displaced only slightly by the fitting procedure. As shown in the middle of each component in Figure 5 these atoms are initially well positioned within the target map and are not moved much by the refinement procedure.

If the RMSD°s between the atoms of the initial and fitted structures are averaged for each residue and plotted against the residue number, the highly displaced tracts appear as elevated plateaus over a background level of displacement (approximately 3 Å). We examined the RMSD°-residue table to locate the segments in a 337-nucleotide tract of the 23S RNA (that includes the L1 binding domain) that remain rigid through a YUP.SCX fitting. (Note that YUP.SCX does not require manual segmentation of the model.) Once the fragment boundaries have been established, only then did we look at the secondary and 3D structures. This is clearly a subjective procedure.

The characteristics of the fragments that were found are listed in Table 9. We find that some of the fragment boundaries are at the ends of helices (C1902, A2077 and G2110). Most are in internal loops of various kinds (U2076, U2111, U2182, U2183 and G2238). Two boundary bases are in a short helix (C2175 and A2176). For refinement methods that require a manual segmentation of the atomic model, one would naturally choose a base at the end of a helix as a boundary (see for example Gao et al., 2003).

Fragment A is very rigid: the per-residue RMSD° varies over a small range (note the minimum, average and maximum). It is also relatively immobile: the (per atom) RMSD is low and varies little with superposition (note the RMSD° and RMSD). Fragment C is displaced by an average of 23.7 Å, but the fragment from the refined structure can be aligned and superimposed into the initial conformation with a RMSD of just over 2 Å. Thus, this fragment remains more or less a rigid unit even as it is displaced over a large distance. The initial and fitted conformations

for fragment B can be superimposed very closely except for the last few bases. However, these bases can also be superimposed closely at the expense of the leading bases. This shows that the conformation changes necessary to rotate fragment C to fit into the target map is distributed all along fragment B. Except for a few stretches, the path of fragments B to the leading portion of fragment E in the initial and fitted structures can be traced in Figure 5 (bottom panel).

We have also fitted the 30S and the 50S at their optimal radius ratios but with shorter durations. The results are listed in Table 8 in the *YUP.SCX1* column. In each case the resulting CCC is as good as or better than that obtained from the runs of longer duration. There are also fewer stereochemical violations. Thus, a less severe refinement procedure may produce a good fit with fewer stereochemical violations.

A full analysis of these structures and of the Ribosome in the post-accommodation stage is published elsewhere (Devkota et al., to be published).

4. Discussion

The YUP.SCX procedure is successful in solving the Kinase problems as indicated by the high CCC and low RMSD^o of the fitted structures. The Kinase problems were devised to have known or exact solutions. Thus, we are able to see that our fits are correct, not just for the backbone atoms but also for the atoms of many of the side chains. The figures of merit are not as discriminating for the Ribosome problems. We are able to obtain consistent fittings from many different runs carried out with different settings. While most of the complex remains virtually unchanged by the fitting procedure, there are large displacements of some components. When examined visually the displacement pattern is seen to be consistent with optimizing the fit of the model into the density; we cannot find a case where a different conformation change will fit better into the map. Many of the components that are moved are translocated over large distances and angles, which would result in large changes in CCC and RMSD^o if these are calculated individually for each component. However, the complex as a whole is not changed much and consequently, the CCC and RMSD^o are not very discriminating measures of quality for these problems.

Central to the YUP.SCX procedure is the coaxing function *SCX*, which represents each voxel of the density map as an attractive well. The wells can be adjusted by setting the ratio of the outer to inner radii of the well, or the radius ratio for short. The radius ratio controls how much the density map can influence each atom. Naturally, one would use a ratio that is as high as possible. However, the cost of evaluating the *SCX* function is proportional to the number of atoms and the third power of the radius ratio. Thus, the optimal radius ratio for the **50S** problem appears to be larger than 5, but the computations would take an impractically long time. On the other hand, the optimal ratio for the **30S** problem is only 4. The **K149** problem shows the danger of setting a ratio that is so high as to coax some of the atoms into the wrong density region.

Increasing the radius ratio has the effect of smoothing and lowering the *SCX* function everywhere except at the boundaries of the density map, thus confining the atomic model inside the map to an increasing degree. Since the *SCX* term dominates the total potential energy function, the latter is attenuated as well. However, the smoothed energy function will still have turning points to hinder gradient-based optimization methods. The objective function for NMFF is the CCC, which appears to change very slowly in the higher-resolution Kinase problems and to have a very large barrier in the **50S** problem. (We also encountered the local minima problem when we tried to optimize Equation (4) with the steepest descent and conjugate gradient methods.)

The Simulated Annealing method provides a good chance of attaining the global minimum. The smoothing afforded by SCX allows extremely low temperatures to be used, thus keeping thermal fluctuations down. The quality of the results from SA improves with the duration of the annealing (up to a point). The smoothing also provides a shorter path to the solution, which may explain why YUP.SCX is more than 25 times faster than NMFF in the 5 Å Kinase problems. Finally, we find that the stochastic nature of SA to be unproblematic, at least for the Kinase problems. We hypothesize that the smoothing of the energy function eliminates many of the highest minima and the general lowering of the function amplifies the global minimum, making it an easier target.

Some refinement methods (for example, RSRef; Gao et al., 2003) require a highly subjective segmentation of the atomic model into rigid fragments. This is laborious and may require multiple cycles of segmentation and refinement in order to obtain a satisfactory outcome. The results may also be biased by the choice of hinge points. The YUP.SCX procedure does not require segmentation as all atoms are moved simultaneously. Rigid segments may arise naturally as an outcome of the fitting. The transitions between the segments are gradual, which is more realistic than a hard hinge.

The best feature of YUP.SCX is its ease of use. (There are only two important parameters, and the optimal settings for them fall in narrow ranges: 0.2 to 1 for *stepfactor* to set the duration and 3 to 5 for the radius ratio.) One can use the default settings, specify only the initial structure and the target density map, and obtain a refined structure that is nearly as good as one obtained from the use of tuned settings. This simplicity contrasts with the need to segment the model such as in RSRef or to select the codebook vectors for Situs. Yet, YUP.SCX will preserve the rigid segments if that is encoded in the density map. The NMFF method is similarly easy to use because it does not ask the user to segment the model. However, it cannot fit a closed structure into the map of an extended conformation in one step. This is not a problem for YUP.SCX, which overcomes this problem with the use of a highly modified GNM. The biggest advantage that YUP.SCX has over NMFF is the use of a global optimizer to overcome local minimum traps, thus allowing large and complex refinement problems to be solved.

It is safe to predict the need to fit ever larger molecular complexes into higher-resolution maps. High-resolution maps have finer grids and therefore there are more voxels for a given volume. Large molecular dimensions and finer grids compound into very large density maps. Thus, we will have to address the need for more computer memory and speed. The memory requirement can be reduced somewhat by raising the threshold (see section S2 of the Supplement). For example, about 20% of the lowest densities can be excised from the Ribosome maps without affecting the outcome of the YUP.SCX procedure. Contemporary computers are now equipped with 64-bit processors, which can address enough memory to accommodate future maps. We also plan to parallelize the YUP package to take advantage of modern computers that are equipped with multiple and multi-core processors. (Even personal computers today can have as many as eight cores.) Thus, YUP.SCX can keep up with the demands of ever-larger refinement problems for the foreseeable future.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work is supported by a grant from the National Institutes of Health: RR12255 (C.L. Brooks III, P.I.)

References

- Berk V, Zhang W, Pai RD, Cate JHD. Structural basis for mRNA and tRNA positioning on the ribosome. *Proc Natl Acad Sci USA* 2006;103:15830–15834. [PubMed: 17038497]
- Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank. *Nucleic Acids Research* 2000;28:235–242. [PubMed: 10592235]
- Bondi A. van der Waals Volumes and Radii. *J Phys Chem* 1964;68:441–451.
- Brunger AT, Kuriyan J, Karplus M. Crystallographic R factor refinement by molecular dynamics. *Science* 1987;235:458–460. [PubMed: 17810339]
- Brunger AT. Crystallographic refinement by simulated annealing: Application to a 2.8 Å resolution structure of aspartate aminotransferase. *J Mol Biol* 1988;203:803–816. [PubMed: 3062181]
- Ceulemans H, Russell RB. Fast fitting of atomic structures to low-resolution electron density maps by surface overlap maximization. *J Mol Biol* 2004;338:783–793. [PubMed: 15099745]
- Chacon P, Wriggers W. Multi-resolution contour-based fitting of macromolecular structures. *J Mol Biol* 2002;317:375–384. [PubMed: 11922671]
- Chapman MS. Restrained real-space macro- molecular atomic refinement using a new resolution-dependent electron-density function. *Acta Crystallogr A* 1995;51:69–80.
- Chen LF, Blanc E, Chapman MS, Taylor KA. Real space refinement of acto-myosin structures from sectioned muscle. *J Struct Biol* 2001;133:221–232. [PubMed: 11472093]
- Cui Q, Tan RKZ, Harvey SC, Case DA. Low-Resolution Molecular Dynamics Simulations of the 30S Ribosomal Subunit. *Multiscale Modeling and Simul* 2006;5:1248–1263.
- Devkota, B.; Caulfield, TR.; Tan, RKZ.; Harvey, SC. The Structure of the E. coli Ribosome Before and After Accommodation: Implications for Proofreading. to be published
- Fabiola F, Chapman MS. Fitting of high-resolution structures into electron microscopy reconstruction images. *Structure* 2005;13:389–400. [PubMed: 15766540]
- Frank, J. Three-dimensional electron microscopy of macromolecular assemblies. Academic Press; San Diego: 1996.
- Frank J, Radermacher M, Penczek P, Zhu J, Li Y, Ladjadj M, Leith A. SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. *J Struct Biol* 1996;116:190–199. [PubMed: 8742743]
- Frank J. Ribosomal dynamics explored by cryo-electron microscopy. *Methods* 2001;25:309–315. [PubMed: 11860285]
- Gao HX, Sengupta J, Valle M, Korostelev A, Eswar N, Stagg SM, Van Roey P, Agrawal RK, Harvey SC, Sali A, Chapman MS, Frank J. Study of the Structural Dynamics of the E. coli 70S Ribosome Using Real-Space Refinement. *Cell* 2003;113:789–801. [PubMed: 12809609]
- Goddard TD, Huang CC, Ferrin TE. Visualizing density maps with UCSF Chimera. *J Struct Biol* 2007;157:281–287. [PubMed: 16963278]
- Humphrey W, Dalke A, Schulten K. VMD - Visual Molecular Dynamics. *J Molec Graphics* 1996;14:33–38.
- Jiang W, Baker ML, Ludtke SJ, Chiu W. Bridging the information gap: computational tools for intermediate resolution structure interpretation. *J Mol Biol* 2001;308:1033–1044. [PubMed: 11352589]
- Kirkpatrick S, Gelatt CD Jr, Vecchi MP. Optimization by Simulated Annealing. *Science* 1983;220:671–680. [PubMed: 17813860]
- Kong Y, Zhang X, Baker TS, Ma J. A Structural-informatics Approach for Tracing β -Sheets: Building Pseudo-C α Traces for β -Strands in Intermediate-resolution Density Maps. *J Mol Biol* 2004;339:117–130. [PubMed: 15123425]
- Lasker K, Dror O, Shatsky M, Nussinov R, Wolfson HJ. EMatch: Discovery of High Resolution Structural Homologues of Protein Domains in Intermediate Resolution Cryo-EM Maps. *IEEE/ACM Trans Comput Biol Bioinformatics* 2007;4:28–39.
- Mears JA, Sharma MR, Gutell RR, McCook AS, Richardson PE, Caulfield TR, Agrawal RK, Harvey SC. A Structural Model for the Large Subunit of the Mammalian Mitochondrial Ribosome. *J Mol Biol* 2006;358:193–212. [PubMed: 16510155]

- Muller CW, Schulz GE. Structure of the complex between adenylate kinase from *Escherichia coli* and the inhibitor Ap5A refined at 1.9 Å resolution. A model for a catalytic transition state. *J Mol Biol* 1992;224:159–177. [PubMed: 1548697]
- Muller CW, Schlauderer GJ, Reinstein J, Schulz GE. Adenylate kinase motions during catalysis: an energetic counterweight balancing substrate binding. *Structure* 1996;4:147–156. [PubMed: 8805521]
- Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. UCSF Chimera - A Visualization System for Exploratory Research and Analysis. *J Comput Chem* 2004;25:1605–1612. [PubMed: 15264254]
- Rossmann MG. Fitting atomic models into electron-microscopy maps. *Acta Crystallogr D, Biol Crystallogr* 2000;56:1341–1349. [PubMed: 10998631]
- Rossmann MG, Bernal R, Pletnev SV. Combining electron microscopic with X-ray crystallographic structures. *J Struct Biol* 2001;136:190–200. [PubMed: 12051899]
- Rossmann MG, Morais MC, Leiman PG, Zhang W. Combining X-ray crystallography and electron microscopy. *Structure* 2005;13:355–362. [PubMed: 15766536]
- Saibil HR. Conformational changes studied by cryo-electron microscopy. *Nat Struct Biol* 2000;7:711–714. [PubMed: 10966635]
- Schröder GF, Brunger AT, Levitt M. Combining Efficient Conformational Sampling with a Deformable Elastic Network Model Facilitates Structure Refinement at Low Resolution. *Structure* 2007;15:1630–1641. [PubMed: 18073112]
- Stagg SM, Lander GC, Pulokas J, Fellmann D, Cheng A, Quispe JD, Mallick SP, Avila RM, Carragher B, Potter CS. Automated cryoEM data acquisition and analysis of 284742 particles of GroEL. *J Struct Biol* 2006;155:470–481. [PubMed: 16762565]
- Tama F, Miyashita O, Brooks CL III. Flexible Multi-scale Fitting of Atomic Structures into Low-resolution Electron Density Maps with Elastic Network Normal Mode Analysis. *J Mol Biol* 2004;337:985–999. [PubMed: 15033365]
- Tan RKZ, Harvey SC. Yampmp: Development of a Molecular Mechanics Program Using the Modular Programming Method. *J Comput Chem* 1993;14:455–470.
- Tan RKZ, Petrov AS, Harvey SC. YUP: A Molecular Simulation Program for Coarse-Grained and Multiscaled Models. *J Chem Theory Comput* 2006;2:529–540.
- Tirion MM. Large Amplitude Elastic Motions in Proteins from a Single-Parameter, Atomic Analysis. *Phys Rev Lett* 1996;77:1905–1908. [PubMed: 10063201]
- Topf M, Baker ML, Marti-Renom MA, Chiu W, Sali A. Refinement of Protein Structures by Iterative Comparative Modeling and CryoEM Density Fitting. *J Mol Biol* 2006;357:1655–1668. [PubMed: 16490207]
- Valle M, Zavialov A, Li W, Stagg SM, Sengupta J, Nielen RC, Nissen P, Harvey SC, Ehrenburgh M, Frank J. Incorporation of aminoacyl-tRNA into the ribosome as seen by cryo-electron microscopy. *Nat Struct Biol* 2003;10:899–906. [PubMed: 14566331]
- Velazquez-Muriel JA, Valle M, Santamaría-Pang A, Kakadiaris IA, Carazo JM. Flexible Fitting in 3D-EM Guided by the Structural Variability of Protein Superfamilies. *Structure* 2006;14:1115–1126. [PubMed: 16843893]
- Wriggers W, Milligan RA, McCammon JA. Situs: a package for docking crystal structures into low-resolution maps from electron microscopy. *J Struct Biol* 1999;125:185–195. [PubMed: 10222274]
- Wriggers W, Agrawal RK, Drew DL, McCammon A, Frank J. Domain motions of EF-G bound to the 70S ribosome: insights from a hand-shaking between multi-resolution structures. *Biophys J* 2000;79:1670–1678. [PubMed: 10969026]

Figure 1.

Scheme for the fitting procedure described in this paper, showing data or files (rectangles), procedures (rounded rectangles) and data flow (arrows). Pre-fitting of the atomic model into the target map is accomplished with programs that are not part of YUP.SCX (dashed outline). The components that make up YUP.SCX are shaded.

Figure 2.

A sample density distribution (positive values), normalized to a 0...1 scale, along one dimension r , and the functions $E^{SCX}(r; r_o)$ (Equation 1) that are derived from the density distribution, with the radius ratio $r_o : r_i$ set to 1.396, 2.5, 3.5 and 5.0 (negative values, light gray to black). The density map contains 17 voxels, each 1.5 Å wide.

Figure 3.

The CCC and RMSD° between the fitted structure and the exact solution for the **K419** test case, obtained from procedures carried out at different molecular dynamics step sizes and durations. Step sizes are 1 fs (square), 2.5 fs (circle) and 5 fs (triangle). The CCC is plotted with filled symbols (left ordinate axis), the RMSD° with open symbols (right ordinate axis). Except for step size, duration and initial velocity distribution, each run is carried out from the same starting conformation and under the same conditions. Figure S1 shows a similar plot for the **K149** test problem.

Figure 4.

Top left: representative structures from the four groups of structures obtained from multiple fittings of the **K145** test case are shown in different shades of blue, and the exact solution is rendered in yellow. Bottom left: An incorrect folding of the **K149** test case, using the default radius ratio of 4. The starting structure is rendered in light blue, the exact solution in yellow and the fitted structure in blue. Center and Right: Representative structures obtained from multiple YUP.SCX fits for the **K145** (top center), **K415** (bottom center) and **K419** (bottom right) test problems. The **K149** (top right) structure was obtained from a single fit, using a radius ratio of 3. The fitted structures are shown in blue, while the exact solutions are rendered in yellow. All the panels are at the same scale, and in the same orientation, selected to show the maximum deviation between the fitted structures and the exact solutions. Note that the exact solution in one row is the starting structure for the other. The boundary of each map was slightly adjusted from the default, to barely overlap the backbones of the exact solutions.

Figure 5.

Top left: cytoplasmic view of the small subunit, the initial structure (green, backbone trace only) and the best fit obtained from the YUP.SCX procedure (red) for the **30S**. Top right: intersubunit view of the large subunit, the initial structure (green, backbone trace only) and the best fit obtained from the YUP.SCX procedure (red) for the **50S** test structure. Bottom: expanded view of the L1 binding domain region in a different orientation from the above. The approximate positions of selected bases in the initial and fitted structures are marked by lowercase letters: **a:** A2077, **b:** U2085, **c:** A2095, **d:** C2103, **e:** U2111, **f:** G2116, **g:** G2133, **h:** G2144, **i:** G2168, **j:** A2176, **k:** A2183, **l:** G2186 and **m:** A2199. The color of each label matches the color of the indicated strand; black letters are used when the locations in the two

strands are very close (*a* to *d* and *k* to *m*). Some of the landmarks are at the start of the fragments listed in Table 9, *a*: fragment B, *e*: C, *j*: D and *k*: E. A larger version of this panel is available as Figure S2 of the Supplement.

Table 1

Preparation of the test problems

Case	30S	50S	K145	K149	K415	K419
Initial Structure						
Procedure	1	Separate the crystal structure of the 70S Ribosome (Berk et al., 2006) into 30S and 50S components.				
	2	Model-build the missing residues (Devkota et al., to be published).				
	3	Rigid fit of each component into the target maps (Goddard et al., 2007).				
Structure	212P	212V	1AKE_A (compact)		4AKE_A (extended)	
Size	51,369 atoms	91,014 atoms	1,656 atoms			
Target Map						
Procedure	Use SPIDER (Frank et al., 1996) to separate the cryo-EM map of the pre-accommodation state 70S (Valle et al., 2003; EMD-1055 ⁷) into 30S, 50S and tRNA components.					
Resolution	9 Å	9 Å	5 Å	9 Å	5 Å	9 Å
Grid Width	2.82 Å	2.82 Å	0.5 Å	2.5 Å	0.5 Å	2.5 Å

⁷ record locator in the Electron Microscopy Database (e.g. use the search tool at <http://www.ebi.ac.uk/msd-srv/emsearch/>)

Table 2

Force constants for the highly modified GNM

Residue Distance ²	Link Length [$\text{\AA}$]	k_{Hooke} [kcal/mol/ $\text{\AA}^2$]
0...1	0...2	59.616
	2...4	29.808
	4...8	14.904
	>8	0.0059616
>1	>0	0.0059616

² Number of consecutive residues between the atoms if they are within the same polymer chain; otherwise taken as >1 if the atoms are in different molecules.

Table 3
Results from multiple refinements of the Kinase problems³

Case	RMSD Range ⁴ [Å]	Pop. ⁵	Profiles ⁶		RMSD • [Å] ⁷
			CCC		
K145 ⁸	—	1	0.940		3.18
	0.05...0.21	7	0.949	0.951	2.60
	0.17...0.52	3	0.953	0.953	2.45
	0.06...0.46	5	0.961	0.961	2.11
	0.04...0.18	9	0.964	0.965	1.85
K149	0.23...1.15	25	0.913	0.917	9.14
K415	0.02...0.13	10	0.980	0.981	1.10
K419	0.05...1.16	10	0.991	0.992	1.57

³ Clustering of the structures was done manually (with the aid of a spreadsheet and graphing program). The RMSD[°] was calculated for all pairs of structures, sorted, and graphed. An abrupt jump in RMSD[°] indicates the start of each sub-group. Taking each sub-list, and considering each pair, a structure is a member of the sub-group if it appears more than once in the sub-list. Otherwise, the corresponding RMSD[°] is the distance between members from two different subgroups.

⁴ within the members of each sub-group (**K145**) or group (other cases)

⁵ population of the group or sub-group

⁶ Listed are the CCC and the corresponding RMSD[°] of the structures with the minimum, median and maximum CCC. In the case of groups or sub-groups with an even number of members, the member with the CCC closest to the mean value replaces the median.

⁷ between the refined structure and the exact solution

⁸ The RMSDs between the sub-groups are in the range 0.67 ...2.24 Å.

Table 4
Energies and measures of quality^o at the start and finish of three folding pathways¹⁰

Case	K145		K149		K415		K419	
Radius ratio	4.0		3.0		4.0		4.0	
Position	Start	Finish	Start	Finish	Start	Finish	Start	Finish
Null ¹¹ : energy minimization, from ideal conformation								
SCX	-669.8	-687.9	-665.8	-679.5	-625.7	-610.1	-691.8	-694.9
SSX	3182	1.998	3182	1.598	1413	0.6477	1413	0.5391
GNM	0	35.93	0	31.46	0	13.54	0	11.87
PE	2512	-650	2516	-646.5	787.7	-595.9	721.7	-682.5
CCC	1.000	0.998	1.000	0.999	1.000	0.999	1.000	1.000
RMSD ^o	0.00	0.23	0.00	0.28	0.00	0.17	0.00	0.20
Standard ¹² : simulated annealing, from opposite conformation								
SCX	-5480	-610.4	-4012	-651.7	-4784	-559.2	-8083	-675.1
SSX	1413	1.461	1413	0.7954	3182	2.252	3182	2.763
GNM	0	450.1	0	222	0	111.3	0	107
PE	-4066	-158.8	-2599	-428.9	-1602	-445.6	-4901	-565.4
CCC	0.5626	0.961	0.729	0.971	0.563	0.980	0.729	0.992
RMSD ^o	11.15	2.11	11.15	2.87	11.15	1.11	11.15	1.57
Reverse ¹³ : simulated annealing, from ideal conformation, GNM from opposite conformation								
SCX	-12470	-619.7	-5833	-651.5	-10840	-555.6	-11530	-654.6
SSX	8124	2.324	8124	2.056	2622	4.076	2622	3.637
GNM	27250	771.4	27250	445.3	22170	452	22170	407
PE	22900	153.9	29540	-204.1	13950	-99.6	13260	-243.9
CCC	0.999	0.974	1.000	0.974	1.000	0.979	1.000	0.993
RMSD ^o	0.00	1.19	0.00	2.47	0.00	1.10	0.00	1.60

⁹ SCX = value of the SCX term [Kcal/mol]; the scores at the *Start* for the **Standard** and **Reverse** folding pathways are calculated at the radius ratio listed for each case (hence the values increase contrary to expectation), and calculated at a radius ratio of 1.0 everywhere else. SSX = value of the SSX term [Kcal/mol]. GNM = value of the *Hooke* term [Kcal/mol]. PE = potential energy [Kcal/mol], sum of the three previous terms. CCC = cross-correlation coefficient. $RMSD^\circ$ = root mean square deviation [$^\circ$] from the exact solution. The *Start* and *Finish* columns show these values at the start and finish of each folding path.

¹⁰ Multiple foldings were carried out for the **Reverse** and **Standard** pathways and the median (or the structure closest to the mean in the case of even-numbered populations) result is presented here. The **Null** folding consists of a single calculation per problem, since the outcome is deterministic for the method used.

¹¹ Energy minimization, starting from the ideal conformation, with the radius ratio held at 1.0 throughout. (Simulated annealing yielded similar results.)

¹² Standard simulated annealing as described in the text, starting from the opposite conformation, i.e., for the **K14x** problem, start from the ideal solution to the **K41x** problem, and vice versa.

¹³ Simulated annealing, starting with the ideal conformation, but using the GNM constructed from the opposite conformation.

Table 5

Incorrect solutions to the **K145** and **K149** problems

Method ¹⁴	Case	CCC		RMSD° [Å]		Runtime ¹⁵ [min]
		Initial	Final	Initial	Final	
A ¹⁶	K145	0.563	0.851	11.15	4.62	10.19
	K149	0.729	0.863	11.15	9.19	10.30
B ¹⁷	K149	0.729	0.915	11.15	9.33	11.19

¹⁴ using default YUP.SCX settings, except duration (four times default), and where noted

¹⁵ on a 600 MHz SGI R14000 processor

¹⁶ using a conventional GNM model with a single cutoff distance (3.5Å) and force constant (29.808 kcal/mol/Å²)

¹⁷ using the default radius ratio of 4.0

Table 6
Representative solutions to the Kinase test problems¹⁸.

Case	<i>K145</i>	<i>K149</i> ¹⁹	<i>K415</i> ²⁰	<i>K419</i>
<i>CCC</i>				
<i>Initial</i>	0.563	0.729	0.563	0.729
<i>YUP.SCX</i>	0.961	0.971	0.980	0.992
<i>NMFF</i>	0.816	0.886	0.954	0.990
<i>RMSD° [Å]</i>				
<i>Initial</i>	11.15	11.15	11.15	11.15
<i>YUP.SCX</i>	2.11	2.87	1.11	1.57
<i>NMFF</i>	6.30	7.88	1.61	1.80
<i>Runtime</i> ²¹ [min]				
<i>YUP.SCX</i>	11.24	6.07	11.82	11.35
<i>NMFF</i>	301.63	13.17	405.23	17.64
<i>Validation</i> ²²				
<i>Initial</i>	0,0,0,0	0,0,0,0	0,1,5,0	0,1,5,0
<i>YUP.SCX</i>	0,1,0,1	0,0,0,0	0,1,1,0	0,1,1,0
<i>NMFF</i>	0,3,1,0	0,4,0,0	0,1,4,0	0,5,7,0

¹⁸ Using YUP.SCX with default settings, except for duration, which is always at four times the default, and except where noted. Compared with NMFF with default settings, except where noted.

¹⁹ YUP.SCX was applied with a radius ratio of 3.0 instead of the default value of 4.0.

²⁰ NMFF was applied using a GNM cutoff distance of 4.5 instead of 5.5 Å used for the other cases.

²¹ on a 600 MHz SGI R14000 processor

²² Summary of the report from the Validation Tool Server at the RCSB PDB (<http://www.pdb.org>; Berman et al, 2000); the four numbers are: the number of close contacts in the same asymmetrical unit, the number of bonds that deviate from dictionary values by more than six times the root mean square error, the number of angles that deviate by more than six times RMSD, and the number of inverted chiral centers.

Table 7
The effect of the radius ratio on the Ribosome results²³

Case	30S			50S		
	r_o/r_i	CCC	RMSD*	Runtime	CCC	RMSD* Runtime
2.5	3.0	0.890	3.68	0.70 D	0.881	3.69
		0.891	3.71	0.78 D	0.884	4.04
3.5	4.0 ²⁴	0.891	3.75	0.86 D	0.885	4.27
		0.892	3.76	0.96 D	0.885	4.35
4.5	5.0	0.892	3.75	0.94 D	0.886	4.41
		0.891	3.77	1.26 D	0.886	4.42

²³Calculations were done on three different computer platforms and speed ratios were determined from these and numerous other overlapping runs. Runtimes are expressed as a multiple of D CPU-days where $D = 1.0$ for 2 GHz PowerPC G5, 3.08 for 600 MHz SGI R14000, 0.92 for 3.2 GHz Pentium Xeon.

²⁴The two calculations for the 30S were carried out on different computer platforms

Table 8
Results from the refinement of the Ribosome test cases.

	Initial	YUP.SCX ²⁵	NMFF	YUP.SCX1 ²⁶
<i>30S</i>				
CCC	0.818	0.892	0.839	0.892
RMSD ²⁷ [Å]	0.0	3.75	1.62	3.65
Runtime ²⁸ [hr]	—	70	6.45	55
Validation ²⁹	3,53,58,0	0,216,263,60	3,182,64,0	0,156,199,43
Min/Validation ³⁰	—	0,0,31,51	—	0,0,34,37
<i>50S</i>				
CCC	0.816	0.886	0.816	0.892
RMSD ²⁷ [Å]	0.0	4.42	0.06	4.09
Runtime [hr]	—	361	— ³¹	258
Validation	7,2,8,2	0,66,152,149	—	0,43,140,108
Min/Validation	—	0,3,72,128	—	0,0,66,90

²⁵With *stepfactor* = 0.5 (default) and radius ratio = 4.0 (*30S*) and 5.0 (*50S*)

²⁶Radius ratios as used in YUP.SCX, with energy minimization before the simulated annealing, and with *stepfactor* = 0.25 (*30S*) and 0.2 (*50S*).

²⁷between the initial and final structures

²⁸Calculations were done on three different computer platforms. All timings have been normalized to a 600 MHz SGI R14000 processor.

²⁹Summary of the validation report: number of close contacts, number of bond lengths outside six standard deviations, number of angles outside six standard deviations and number of chiral inversions.

³⁰Summary of validation carried out after 1000 steps of energy minimization in NAMD2 using the Charmm force field.

³¹This calculation terminated prematurely, at the first cycle.

Table 9
Looking for rigid segments in a YUP.SCX fitted structure

23S RNA Fragment ³²	<Per Residue RMSD> ³³			RMSD ³⁴	RMSD ³⁵
	Minimum	Average	Maximum		
A	C1902...U2076	0.54	1.19	2.16	1.01
B	A2077...G2110	1.72	3.39	16.08	2.81
C	U2111...C2175	13.09	23.19	31.39	2.15
D	A2176...U2182	3.52	7.34	13.70	0.91
E	A2183...G2238	1.89	2.42	3.25	1.03

³²The path of the fragments from B to the middle of D can be traced in Figure 5 (bottom panel).
³³The RMSD° averaged over each residue was plotted against the residue number and the profile was used to tease out the fragments.
³⁴between the absolute initial and fitted positions of all atoms of each fragment
³⁵after alignment and superimposition of all atoms of each fragment