

Flexible Fitting and Refinement of Atomic Structures Using the Coarse-Grained DDFE Force Field Tailored to 5-10Å Resolution Cryo-TEM Maps

Julio Kovacs¹, Vitold E. Galkin², and Willy Wriggers¹

¹Dept. of Mechanical & Aerospace Engineering and Institute of Biomedical Engineering, Old Dominion University, Norfolk, VA 23529

²Dept. of Physiological Sciences, Eastern Virginia Medical School, Norfolk, VA 23507, USA

In fitting atomic structures into cryo-TEM maps, it often occurs that the map corresponds to a different conformation of the structure. The “Damped-Dynamics Flexible Fitting” methodology (DDFF) [1] is able to model large conformational changes while preserving the covalent geometry. This is accomplished through the use of internal rather than Cartesian coordinates. DDFF scales well with system size, is stable with respect to initial conformations, and provides a natural way to avoid overfitting.

Here, we added two new features to the DDFF approach: (1) Distance constraints (in the present application, to improve the stability of β -sheets); and (2) Automatic, simultaneous determination of kernel width and threshold for generating the simulated map. This is achieved by requiring that the volume at the threshold value, the 95% quantile of the density values, and the integral of the density for the simulated map match the corresponding values for the EM map after proper scaling. We show an application of our method to the refinement of a large multi-chain structure.

The outline of our method [1] is as follows:

- A dynamical system is defined by interatomic forces that are proportional to the velocity difference between atoms within a cutoff distance.
- The system is completely damped by setting the atom masses to 0 and adding a drag term to the equations of motion.
- A force field is defined that pulls the atomic structure toward regions of the EM map that are not yet fully occupied, again up to a cutoff distance.
- The resulting equations of motion are integrated in generalized coordinates \mathbf{q} (positional and dihedral-angle coordinates):

$$[\mathbf{B}(\mathbf{q}) + \mathbf{V}(\mathbf{q})] \cdot \dot{\mathbf{q}} = \mathbf{F}(\mathbf{q}) \quad (1)$$

The trajectory is followed until convergence. Periodically, the side chains are energy-minimized using the SCWRL4 method [2]. This not only compensates for the coarse-grained modeling of side chains (see below) but also allows the trajectory to escape from local energy basins.

In equation (1), matrix $\mathbf{B}(\mathbf{q})$ contains the drag coefficients b_k , and $\mathbf{V}(\mathbf{q})$ contains the damping coefficients C_{ki} . The right-hand side of the equation represents the force field in generalized coordinates; it adds up contributions proportional to $f(\mathbf{p})-g(\mathbf{p})$, where $f(\mathbf{p})$ denotes the EM map and $g(\mathbf{p})$ the simulated map derived from the model at that instant [1]. The system of equations (1) is of first order, linear in the (generalized) velocities, and symmetric. Hence, their integration is simple and efficient and prevents oscillations. However, the naive computations of matrices \mathbf{B} and \mathbf{V} has complexity $O(N^4)$ (where N is the number of atoms). We have developed a recursive method to compute these matrices

with complexity $O(N^2)$, which makes the method feasible even for molecules with many thousands of residues. An additional simplification is the use of a reduced residue model consisting of at most five pseudo-atoms [1].

We applied DDFF to the refinement of cardiac myosin binding protein C bound to the thin filament, using a recent 10 Å resolution cryo-TEM map of this complex (Figure 1). The structure contains 6284 residues. The helical symmetry of the thin filament was not imposed in the refinement. However, distance constraints were imposed between the endpoints of the β -strands to enhance the integrity of the β -sheets. These constraints were imposed by setting the time derivatives of the distances to zero, producing extra linear equations in the generalized velocities.

The width of the Gaussian kernel by which the atomic structure is convolved (at each time step) to yield a simulated density map, as well as the density value used to threshold the simulated map, are simultaneously optimized by equating the volumes of both maps (where the density is higher than the threshold), the integrals of both maps, and the 95% quantiles of the density values of both maps. This procedure eliminates the dependence of the solution on the value that the user inputs for the resolution of the map, which in the earlier approach [1] determined the kernel width. The improvement due to this new optimization is shown in Figure 1 [3].

References:

- [1] J Kovacs, M Yeager and R Abagyan, *Biophysical Journal* **95** (2008), p. 3192.
 [2] G Krivov, M Shapovalov and R Dunbrack, Jr, *Proteins* **77** (2009), p. 778.
 [3] The authors acknowledge funding from the Batten Endowment at ODU and NIH R01GM62968 (to W.W.) and AHA Grant-in-Aid 16GRNT31220040 (to V.E.G.).

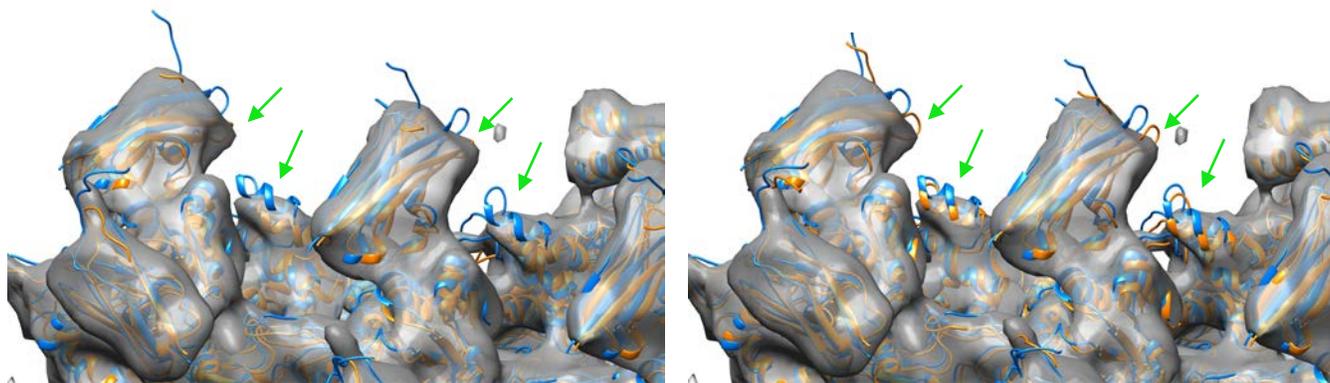


Figure 1. Flexible fitting of cardiac myosin binding protein C to its complex with the thin filament (detail view). The starting structure is shown in blue, and the fitted conformation is shown in orange. *Left:* Fitting obtained with our current approach of optimizing the kernel width and density threshold. The fitting is monitored by the overlap value, which is defined as $1 - (\int |f - g|) / (\int f + \int g)$. In the present application, the refinement improved the overlap from 66% to 78%. *Right:* The fitting obtained with the manual approach [1], which used as width a fixed constant times the resolution of the map input by the user. Note the improvement in the fitting of the helices and β -strands indicated by arrows.