Published in final edited form as:

J Phys Chem B. 2013 April 11; 117(14): 3738–3746. doi:10.1021/jp3126297.

Enhanced Sampling and Overfitting Analyses in Structural Refinement of Nucleic Acids into Electron Microscopy Maps

Harish Vashisth[†], Georgios Skiniotis[‡], and Charles L. Brooks III^{*,†}

[†]Department of Chemistry and Biophysics Program, University of Michigan, Ann Arbor, MI

[‡]Life Sciences Institute, Department of Biological Chemistry, and Biophysics Program, University of Michigan, Ann Arbor, MI

Abstract

Flexible-fitting computational algorithms are often useful to interpret low resolution maps of many macromolecular complexes generated by electron microscopy (EM) imaging. One such atomistic simulation technique is molecular dynamics flexible fitting (MDFF), which has been widely applied to generate structural models of large ribonucleoprotein assemblies such as the ribosome. We have previously shown that MDFF simulations of globular proteins are sensitive to the resolution of the target EM map, and the strength of restraints used to preserve the secondary structure elements during fitting (Vashisth et al. *Structure* **2012**, 20, 1453–1462). In this work, we aim to systematically examine the quality of structural models of various nucleic acids obtained via MDFF by varying the map resolution and the strength of structural restraints. We also demonstrate how an enhanced conformational sampling technique for proteins, temperature-accelerated molecular dynamics (TAMD), can be combined with MDFF for the structural refinement of nucleic acids in EM maps. Finally, we also demonstrate application of TAMD-assisted MDFF (TAMDFF) on a RNA/protein complex and suggest that TAMDFF is a viable strategy for enhanced conformational fitting in target maps of ribonucleoprotein complexes.

Keywords

electron microscopy; ribonucleic acid; molecular dynamics flexible fitting; temperature-accelerated molecular dynamics; enhanced sampling

INTRODUCTION

Functional motions of biomolecules are directly related to their structures and the underlying dynamics. A useful experimental technique to probe the macromolecular architecture and conformational dynamics is single-particle electron microscopy (EM).^{1–5} Due in part to the structural flexibility inherent to biomolecules, the EM maps are often low resolution, and require interpretation using high resolution structures of individual components. The computational algorithms that allow fitting of high resolution structural data into low resolution maps range from rigid-body docking^{6,7} to flexible fitting techniques,^{8–16} which use either coarse-grained or all-atom representations of the biomolecules. Especially for low resolution flexible fitting, methods such as normal mode flexible fitting (NMFF)^{8,9} that use a relatively small number of degrees of freedom, have significantly less potential for

*To whom correspondence should be addressed, brookscl@umich.edu, Phone: (734)-647-6682. Fax: (734)-647-1604.

Supporting Information Available

Fitting results from additional simulations and details on subdomain partitions used in TAMDFF simulations appear in the supplemental material. This material is available free of charge via the Internet at <http://pubs.acs.org/>.

overfitting in comparison to all-atom simulation methods such as molecular dynamics flexible fitting (MDFF).^{10–12} However, structural fitting of a number of large biological complexes has been carried out both via NMFF^{17–21} and MDFF.^{10–12,22–37}

While all-atom simulations are not always desirable and computationally feasible, we have previously shown that structural fitting of globular proteins in simulated/experimental maps via MDFF is sensitive to the resolution of the target EM map and the strength of structural restraints applied during the fitting simulation.³⁸ In fact, there is a close interplay between weakening the secondary structure restraints and lowering the map resolution; softer restraints in low resolution maps lead to significant overfitting as indicated by the deviation of fitted-structures from the known crystal conformations. Moreover, functional movements of many biomolecules require large-scale conformational reorganization³⁹ elicited via domain translations/rotations, where methods such as MDFF may be limited in capturing the rotations of structural elements.¹⁰ For MDFF simulations, one way to decrease the computational cost and alleviate the limitations stemming from domain rotations is to combine MDFF with an enhanced sampling simulation technique such as the temperature-accelerated molecular dynamics (TAMD),^{40–43} which has been shown to accelerate conformational search in MDFF simulations of multidomain proteins.³⁸ However, large-scale conformational changes are not unique to proteins, and it is being recognized that nucleic acids also often undergo structural transitions in response to cellular stimuli such as metabolite binding.^{44–47} Given that MDFF is routinely applied to cryo-EM maps of ribonucleoprotein complexes such as the ribosome,^{22,26–29,32,34} it may be useful and informative to test the sensitivity of MDFF simulations to the resolution of EM maps for structural fitting of nucleic acids. Furthermore, demonstrating the usage of enhanced sampling in MDFF simulations of nucleic acids has the potential to extend the capabilities of this method in studying large-scale conformational changes in ribonucleoprotein assemblies.

In this work, we take examples of several different systems to examine the quality of structural models generated via MDFF by varying the resolution of the target EM map, and the strength of structural restraints. Each of these representative systems are experimentally well-characterized, display large-scale conformational changes, and their modest system-sizes are computationally amenable to an all-atom simulation. These five systems are: (i) domain IV of the signal recognition particle RNA from *E. coli* (SRP-RNA);⁴⁸ (ii) RNA from the complex of human U1A protein with the 3′-untranslated region (UTR) of its own messenger RNA (U1A-UTR-RNA);⁴⁹ (iii) the transactivation response element RNA (TAR-RNA) from the human immunodeficiency virus type-1 (HIV-1);^{50,51} (iv) helix 44 RNA (H44-RNA) from the mature small eukaryotic ribosomal subunit (40S);⁵² and (v) complex of U1A protein with UTR-RNA. For each system, we also carried out explicit-solvent MDFF and TAMD-assisted MDFF (TAMDFF) simulations to demonstrate the utility of incorporating enhanced sampling in MDFF simulations of nucleic acids and ribonucleoprotein assemblies.

MATERIALS AND METHODS Molecular

Dynamics Flexible Fitting

Molecular dynamics flexible fitting (MDFF) is an all-atom simulation method to fit high resolution atomic structures into target EM maps.^{10–12} In addition to the underlying molecular dynamics (MD) force-field, MDFF incorporates an external potential derived from the EM map to steer the fitting of atomic structures into target maps. To preserve the secondary structure elements, maintain correct stereochemistry, or symmetry of the complex, additional restraining potentials need to be applied.^{30,31} Hence, the total potential function for MDFF is:

$$U_{\text{total}} = U_{MD} + U_{EM} + U_{SS} + \dots \quad (1)$$

where U_{MD} is the underlying MD force-field, U_{EM} is the potential derived from the EM map, and U_{SS} is a harmonic restraining potential to preserve secondary structure; additional potential terms can be added if needed. One can also use a scaling factor $\xi > 0$ to uniformly tune the effect of steering forces (arising from the map-derived potential) on the molecular system.

We generated all MD trajectories using NAMD^{53,54} and the CHARMM force-field⁵⁵ with the CMAP correction.⁵⁶ We used VMDv1.9⁵⁷ for system creation, protein rendering, the preparation of simulated maps, and the input files for all MDFF simulations. Initial docking of atomic structures into EM maps was carried out using the SITUS suite.⁶ MDFF simulations reported in Figure 1 and Figures S1–S3 were performed in vacuum, while all other MDFF simulations were performed with explicit-solvent. A 1-fs time-step was used for all MDFF runs, and the MDFF-generated final configurations were further energy-minimized. For MDFF runs of all nucleic acids, a scaling factor of $\xi = 0.1$ kcal/mol was used except for the helix 44 system, where $\xi = 0.2$ kcal/mol was used. In each case, harmonic restraints to preserve the secondary structure elements are applied to the dihedral angles and distances between base pairs (to maintain planarity) with a spring constant $k = 500$ kcal mol⁻¹ rad⁻² (or 500 kcal mol⁻¹ Å⁻²). For runs related to the testing of structural restraints described in Figures S2 and S3, harmonic restraints with $k = 0, 10, 50, 100, 300$, and 500 kcal mol⁻¹ rad⁻² or kcal mol⁻¹ Å⁻² were used. In addition to structural restraints, the correct chirality of all chiral centers was enforced using additional restraints. Initial coordinates (for MDFF fitting) and final coordinates (for target map generation) of each nucleic acid were taken from the following coordinate files: (i) for SRP-RNA, the NMR structure with PDB code 1CQL (model 1, initial; and model 8, final); (ii) for U1A-UTR-RNA, the NMR structure with PDB code 1AUD (model 1, initial; and model 2, final); (iii) for TAR-RNA, the NMR structure with PDB code 1ARJ (model 1, initial) and the NMR structure with PDB code 1LVJ (model 1, final); and (iv) for H44-RNA, the 3 Å resolution crystal structure with PDB code 3U5F (initial) and an experimental map (~18 Å resolution) of pre-40S wild-type eukaryotic subunit (final) from our earlier work.³² The experimental map of pre-40S particle was box-segmented around helix 44 using the software Chimera and filtered to lower resolutions for MDFF fitting as described in Figures S1 and S3. For the U1A protein and UTR-RNA complex, the initial and final coordinates of RNA as well final coordinates of protein were taken from the NMR structure with PDB code 1AUD, while the initial coordinates for protein were taken from the NMR structure with PDB code 1FHT.

Temperature Accelerated Molecular Dynamics

TAMD is an enhanced sampling technique to explore the physical free-energy landscape in a large but finite set of collective variables (CVs), which are the functions of atomic Cartesian coordinates. Maragliano and Vanden-Eijnden originally presented the theoretical basis of TAMD.^{40,41,58} More recently, Abrams and Vanden-Eijnden developed a conformational sampling algorithm for proteins based upon TAMD,⁴² which has been applied successfully to other systems.^{38,43,59,60} Therefore, we simply describe the coupled system of equations governing TAMD, which are:

$$\begin{aligned} m_i \ddot{x}_i &= -\frac{\partial V(x)}{\partial x_i} - \kappa \sum_{j=1}^m [\theta_j^*(x) - \theta_j] \frac{\partial \theta_j^*(x)}{\partial x_i} - \gamma m_i \dot{x}_i + \eta_i(t; \beta) \\ \overline{\gamma m_j} \dot{\theta}_j &= \kappa [\theta_j^*(x) - \theta_j] + \xi_j(t; \beta) \end{aligned} \quad (2)$$

$\theta^*(x) = (\theta_1^*(x), \theta_2^*(x), \dots, \theta_m^*(x))$ are collective variables (CVs) that are functions of the atomic Cartesian coordinates, m_i are the masses of x_i , $V(x)$ is the interatomic MD potential, κ is the “coupling spring-constant”, γ is the Langevin friction coefficient, η is the white noise satisfying fluctuation-dissipation theorem at physical temperature β^{-1} , $\bar{\gamma}$ and $\bar{m}_j$ respectively are fictitious friction and masses of the variables θ_j and ξ is the thermal noise at artificial temperature $\bar{\beta}^{-1}$.

TAMD equations describe the motion of $x(t)$ and $\theta(t)$ over the following extended potential:

$$U_\kappa(x, \theta) = V(x) + \frac{\kappa}{2} \sum_{j=1}^m [\theta_j^*(x) - \theta_j]^2 \quad (3)$$

By choosing κ such that $\theta^*(x(t)) \approx \theta(t)$ and the fictitious friction coefficient $\bar{\gamma}$ such that θ moves slower than x , one can generate a trajectory $\theta(t)$ which moves at the artificial temperature $\bar{\beta}^{-1}$ on the free energy landscape computed at the physical temperature β^{-1} . Therefore, we have chosen a TAMD friction ($\bar{\gamma}$) of 50 ps⁻¹ and a spring constant (κ) of 100 kcal mol⁻¹ Å⁻² in this work, which have been used previously.⁴² Each nucleic acid was divided into multiple subdomains (see Figure S4 and Tables S1–S4 for the identity of residues in each subdomain) and the Cartesian coordinates of the center of mass of each subdomain are CVs. Explicitly-solvated and ionized systems were prepared for TAMDFF simulations of each nucleic acid, and the final size of each system is described in the corresponding figure caption. The parameters for the MDFF part of TAMDFF simulations are described above. For all nucleic acid systems except H44 RNA, we carried out the TAMD part of TAMDFF simulations at a fictitious thermal energy $\bar{\beta}^{-1} = 7$ kcal/mol, where $\bar{\beta} = 1/k_B \bar{T}$, k_B is Boltzmann’s constant, and $\bar{T}$ is the fictitious temperature. For H44 RNA system, we used fictitious thermal energies of $\bar{\beta}^{-1} = 9$ or 10 kcal/mol. For each system, we carried out four TAMDFF simulations (five for H44-RNA) and one MDFF simulation for comparison. The correlation coefficients from these simulations are reported in Figures 2–5. For the complex of U1A protein with UTR-RNA, we used a fictitious thermal energy of $\bar{\beta}^{-1} = 5$ kcal/mol for the fitting of protein, while RNA-fitting was carried out with parameters similar to U1A-UTR-RNA (*vide supra*). The results for this system are shown in Figure 6.

RESULTS

Effect of map resolution and structural restraints on MDFF fitting

To understand the effect of changing map resolution on MDFF refinement of nucleic acids, we carried out MDFF simulations of four different systems (SRP-RNA, U1A-UTR-RNA, TAR-RNA, and H44-RNA; see introduction for details) with significantly large structural restraints ($k = 500$ kcal mol⁻¹ rad⁻² or kcal mol⁻¹ Å⁻²). For MDFF simulations of SRP-RNA, U1A-UTR-RNA, and TAR-RNA, simulated target maps at 5 Å, 10 Å, 15 Å, 20 Å, and 25 Å resolution were used, while for H44-RNA, an experimental target map at 18 Å resolution was filtered to 20 Å, 22 Å, and 25 Å resolution. Five independent MDFF simulations for each nucleic acid were carried out at each map resolution with a total of 95 MDFF simulations. Although we measured the root-mean-square-deviation (RMSD; backbone P-atoms) both from the known starting and final structures in MDFF simulations, the quality of MDFF-generated models was judged based only upon the RMSD from the known target conformation. An RMSD of 0.0 Å from the final structure will indicate perfect overlap between the MDFF-generated model and the known target structure of each nucleic acid, and a positive RMSD will indicate deviation from the known conformation. For systems where simulated maps were used, we present these RMSD traces in Figure 1A for

each map resolution along with average (over five MDFF runs) RMSD from the known target conformation (numbers in italics in lower RMSD panels in Figure 1A).

In general, we find that average achievable RMSD from the known target structure of each nucleic acid decreases with decreasing map resolution suggesting that the quality of MDFF-generated models (Figure 1B) deteriorates as the resolution of the target EM map decreases despite employing significantly large structural restraints and lower steering forces. For SRP-RNA, the initial RMSD from the known target structure is ~6 Å, and final average RMSDs of MDFF-generated conformations for each map resolution are the following: 1.47 (5 Å), 1.93 (10 Å), 2.43 (15 Å), 2.92 (20 Å), 4.86 (25 Å). These data suggest that although significant (~76%) target conformational change could be achieved for SRP-RNA at 5 Å resolution, only limited (~19%) target conformational change is achieved at 25 Å resolution. The RMSD trends for MDFF-generated conformations of U1A-UTR-RNA and TAR-RNA at different map resolutions are similar, where conformational changes are achieved to the following extent at the highest (5 Å) and the lowest (25 Å) map resolution: ~82%/6% (5/25 Å; U1A-UTR-RNA) and ~60%/5% (5/25 Å; TAR-RNA). We note that the conformational change in TAR-RNA is observed to a relatively low-degree (60%) even at 5 Å resolution, because MDFF is unable to capture the change in the conformation of a specific bulge region and an apical loop, which requires incorporating enhanced sampling in the MDFF runs (*vide infra*). The last system we considered here is the conformational change in the H44-RNA of the eukaryotic ribosome on maturation, where we used an experimental EM map of an immature conformation of the smaller eukaryotic ribosomal subunit (pre-40S) from our earlier work.³² We performed MDFF runs at 18 Å, 20 Å, 22 Å, and 25 Å resolution. These data are shown in Figure S1. The initial RMSD from the target structure considered is ~14 Å, and the final RMSD at 20 Å, 22 Å, and 25 Å resolution, respectively, are: 3.47 Å, 3.73 Å, and 4.81 Å. Similar to the other cases described above, MDFF refinement of H44-RNA in experimental maps with decreasing resolution gradually leads to conformations that deviate from the target structure. We also note that the final conformation of H44-RNA represented by the EM map has not yet been experimentally characterized, and hence the RMSDs from the target conformation in EM maps filtered to 20 Å to 25 Å resolution (Figure S1) were computed by considering as a basis the structure with the highest correlation-coefficient (among five runs) obtained from MDFF refinement in the 18 Å map.

Additionally, we analyzed the interplay between weakening the structural restraints and lowering the target EM map resolution. We carried out MDFF simulations for each nucleic acid at each resolution with harmonic restraining spring constants with $k = 0, 10, 50, 100, 300, \text{ and } 500 \text{ kcal mol}^{-1} \text{ rad}^{-2}$ or $\text{kcal mol}^{-1} \text{ Å}^{-2}$. The RMSD traces from the initial and final structures of each nucleic acid from these MDFF runs are shown in Figures S2 and S3. We observe from these RMSD trends that MDFF leads to significant overfitting of structures on lowering k -values because MDFF-generated conformations diverge from the known target conformations of each nucleic acid. This effect is directly coupled to the resolution of the EM map because the fitting via MDFF worsens if lower restraints are used in a low resolution map. This suggests that although MDFF drives the initial structures to fit in the target maps, the presence of less information in low resolution maps (and hence the EM-map-derived potential used in MDFF) makes it difficult for MDFF to locate the actual target conformation(s).

MDFF versus TAMDDF fitting of all systems

In this section, we demonstrate the utility of combining enhanced sampling methods with MDFF in a single atomistic simulation for structural refinement of nucleic acids. If the conformational search in MDFF simulations can be accelerated with the help of enhanced

sampling, it may be useful: (a) in the study of significantly large-scale conformational changes; and (b) in decreasing the computational cost of explicit-solvent MDFF simulations. We have previously shown that TAMD,^{40,41,58} an enhanced sampling method, can be successfully combined with MDFF for the refinement of multidomain proteins in EM maps.³⁸ To extend the use of TAMD for the study of nucleic acids, we prepared explicit-solvent systems of all four nucleic acids (SRP-RNA, U1A-UTR-RNA, TAR-RNA, and H44-RNA). These solvated and ionized systems range from 40,000 atoms (TAR-RNA) to 200,000 atoms (H44-RNA), involve conformational changes in various types of structural elements (bulges, loops, and helices), and to a varied degree, as indicated by the RMSD changes (5 to 12 Å) (Figure 2–Figure 5). For each system, we performed four independent explicit-solvent TAMDFF simulations starting with their known initial conformations, which are fitted into the target maps at 5 Å resolution (18 Å experimental map for H44-RNA). We show the RMSD traces from the TAMDFF simulations and a typical MDFF run along with the initial/final correlation coefficients in Figure 2C, Figure 3C, Figure 4B, and Figure 5B. In general, we observed consistently in all TAMDFF runs of each system that the convergence to the final conformation encoded in the map is relatively faster than the corresponding MDFF run as indicated by the change in the slope of the RMSD traces. Furthermore, the final correlation coefficients from TAMDFF runs and MDFF runs, and the overlay of final conformations (Figure 2D, Figure 3D, Figure 4C, and Figure 5C) indicate that the quality of structures generated by TAMDFF is largely similar to the MDFF runs (except for TAR-RNA as discussed below).

The conformational changes in SRP-RNA and U1A-UTR-RNA involve opening of their terminal helices by ~6 Å in each case, which could be captured both via MDFF and TAMDFF simulations, albeit 1.5 times slower in MDFF runs than in TAMDFF runs. The initial/final correlation-coefficients from the MDFF run, and the average (over four runs) initial/final correlation-coefficients from TAMDFF runs are comparable: 0.724/0.950 (MDFF) and 0.724/0.934 (TAMDFF) for SRP-RNA, and 0.474/0.963 (MDFF) and 0.471/0.950 (TAMDFF) for U1A-UTR-RNA.

The case of TAR-RNA is slightly different than the other two systems described above, because the conformational change in addition to opening of helices also involves changes in the conformations of the apical loop of the lower helix and a bulge region at the junction of both helices (Figure 4A). The RMSD traces (Figure 4B) indicate that the MDFF run is not only significantly slower than TAMDFF runs in capturing all conformational changes in TAR-RNA, but also unable to generate the necessary conformational changes in the bulge region and in the apical loop on the same timescale (see dotted circles in the overlay of final conformations in Figure 4C). In such cases, enhanced sampling methods such as TAMD in combination with MDFF may be useful in overcoming the underlying free-energy barriers, which are difficult to surmount via MDFF alone on similar timescales. As a result, better RMSDs from the known target conformation are achieved in TAMDFF runs in comparison to the MDFF run of TAR-RNA (lower RMSD panels in Figure 4B).

Finally, the largest nucleic acid system considered here is H44-RNA from the mature small subunit of the eukaryotic ribosome. We have previously shown³² that H44 in the immature small ribosomal subunit (pre-40S) is bent outwards in comparison to its conformation in the mature 40S subunit (Figure 5A). Due to a large overall conformational change (over 12 Å RMSD) and the use of a low resolution (18 Å) experimental EM map, the H44-RNA is a more realistic example to see the utility of using enhanced sampling in MDFF runs. The RMSD (with respect to the initial crystal conformation) traces from a typical MDFF run and five independent TAMDFF runs are shown in Figure 5B along with the initial/final correlation coefficients. These data indicate that TAMDFF runs are marginally faster than the MDFF run in the first ~0.35 ns after which the convergence of MDFF run is significantly

slower than the TAMDFF runs. We also note that we extended the MDFF run beyond 2 ns (data not shown), and observed that it takes ~4 ns for MDFF to achieve ~11–12 Å RMSD, which could be achieved in less than ~1 ns via TAMDFF. The overlay of final map-docked conformations shows significant similarity in MDFF and TAMDFF generated conformations (Figure 5C), the only difference is that these conformations could be achieved in shorter simulation time due to accelerated conformational search in TAMDFF runs.

To further demonstrate the usefulness of incorporating enhanced sampling in MDFF simulations of ribonucleoprotein complexes, we apply both MDFF and TAMDFF simulations in explicit-solvent to a complex of the U1A protein with UTR-RNA. In this system, both RNA and the associated protein undergo conformational changes in various structural elements (Figure 6A). During both RNA- and protein-fitting steps, the RMSD traces (Figure 6B) indicate that the convergence of TAMDFF simulations to final structures is relatively faster than MDFF simulations, and the quality of final structures generated by MDFF and TAMDFF strategies are highly similar (Figure 6C). Taken together, these data reinforce the conclusion that TAMDFF can be applied to proteins, nucleic acids, and their assemblies for their faster and improved (in some cases) structural refinement in EM maps.

DISCUSSION

In this work, we have attempted to understand the effect of varying map resolution and the strength of structural restraints for the MDFF fitting of nucleic acids. A key finding is that although MDFF is able to improve the initial fit of a nucleic acid in the EM map even at 25 Å resolution, the quality of final MDFF-generated models decreases with decreasing map resolution. Furthermore, we find in MDFF runs that the resolution of the target EM map is coupled to the strength of structural restraints used, because weakening of these restraints worsens MDFF fitting even at relatively higher resolution of the EM map. Hence, we suggest that weaker restraints in low resolution MDFF fitting may lead to poorer quality in the final models. Although we have considered nucleic acid systems of different sizes, shapes, and having conformational changes of varied magnitude, the extent of overfitting observed is different in different cases. This also suggests that the application of MDFF to low resolution EM data can be dependent upon the shape and size of the nucleic acid under consideration, and the distribution of flexible elements (bulges and internal/apical loops) in the structure, because flexible structural elements cannot be restrained according to secondary structure restraints; such restraints are used to maintain stereochemically correct base-pairing and the geometry of helices. Therefore, we recommend that higher structural restraints should be used in MDFF runs to prevent overfitting of structural elements.

MDFF simulations of nucleic acids and ribonucleoprotein assemblies like the ribosome are often carried out in explicit-solvent,³² which may be computationally demanding for such large systems. Although one way to decrease the computational expense is to use the implicit-solvent in MDFF runs,⁶¹ one can also use enhanced sampling methods in combination with MDFF or alternative fitting techniques that use enhanced sampling,^{14,15} which have the advantage of accelerating conformational search in a single atomistic simulation, leading to significantly shorter trajectories. An additional positive aspect of enhanced sampling is that it may help realize conformational changes that occur at slower timescale and are difficult to capture via MDFF due to large orientational changes in structural elements.^{10,38} For structural refinement of nucleic acids, we therefore combined MDFF with a relatively new enhanced sampling method, TAMD,^{40,41,58} which we demonstrated earlier to accelerate conformational search in MDFF refinement of proteins as well.³⁸ By performing MDFF and TAMDFF simulations of four nucleic acids and a ribonucleoprotein complex in simulated/experimental maps, we showed here that TAMDFF

simulations can help generate final conformations (encoded in the EM map) in relatively less simulation time, and resulting in final models that are significantly similar to MDFF runs. In some cases such as the TAR-RNA, TAMDFF runs prove advantageous to MDFF in achieving conformational changes in specific loops and bulges. Given that TAMDFF can be applied to proteins,³⁸ nucleic acids, and a ribonucleoprotein complex (as shown here), it may also be a useful tool in the structural refinement of other large ribonucleoprotein assemblies, such as the ribosome.

CONCLUSION

In this work, we have carried out viaMDFF the structural refinement of four nucleic acid systems in target simulated/experimental EM maps by varying the resolution of the map as well as the strength of structural restraints. We find that MDFF simulations are sensitive to the resolution of the EM map and the structural restraints used to preserve secondary structure elements, because weakening of such restraints leads to structures diverging from the known final conformations, thereby causing overfitting. We think that our current work on nucleic acids and previous work³⁸ on proteins may serve as a guide to choose appropriate structural restraints for the fitting of atomic structures in low resolution EM data via MDFF. For all nucleic acid systems and a ribonucleoprotein complex, we also carry out enhanced-sampling-assisted MDFF runs to show that MDFF simulations can be accelerated further with such strategies. Finally, we suggest that TAMDFF simulations may be useful in the structural refinement of proteins, nucleic acids, and ribonucleoprotein complexes. We also believe that this work reports the first application of TAMD with MDFF for the structural refinement of nucleic acids.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by the NIH-supported resource Multiscale Modeling Tools for Structural Biology (grant RR012255 to C.L.B.), the NSF funded Center for Theoretical Biological Physics (grant PHY0216576 to C.L.B.), and the University of Michigan Biological Sciences Scholars Program (G.S.); G.S. is a Pew Scholar of Biomedical Sciences.

References

1. Frank J. Classification of Macromolecular Assemblies Studied as Single Particles. *Quart. Rev. Biophys.* 1990; 23:281–329.
2. Frank J. Single-particle Imaging of Macromolecules by Cryo-electron Microscopy. *Annu. Rev. Biophys. Biomol. Struct.* 2002; 31:303–319. [PubMed: 11988472]
3. Mitra K, Frank J. Ribosome Dynamics: Insights from Atomic Structure Modeling into Cryo-electron Microscopy Maps. *Annu. Rev. Biophys.* 2006; 35:299–317.
4. Frank J. Single-particle Reconstruction of Biological Macromolecules in Electron Microscopy-30 Years. *Quart. Rev. Biophys.* 2009; 42:139–158.
5. Frank J, Gonzalez RL Jr. Structure and Dynamics of a Processive Brownian Motor: the Translating Ribosome. *Annu. Rev. of Biochem.* 2010; 79:381–412. [PubMed: 20235828]
6. 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]
7. Wriggers W, Chacón P. Modeling Tricks and Fitting Techniques for Multiresolution Structures. *Structure.* 2001; 9:779–788. [PubMed: 11566128]

8. 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]
9. Tama F, Miyashita O, Brooks CL III. Normal Mode Based Flexible Fitting of High-Resolution Structure into Low-Resolution Experimental Data from Cryo-EM. *J. Struct. Biol.* 2004; 147:315–326. [PubMed: 15450300]
10. Trabuco LG, Villa E, Mitra K, Frank J, Schulten K. Flexible Fitting of Atomic Structures into Electron Microscopy Maps Using Molecular Dynamics. *Structure.* 2008; 16:673–683. [PubMed: 18462672]
11. Trabuco LG, Villa E, Schreiner E, Harrison CB, Schulten K. Molecular dynamics Flexible Fitting: A Practical Guide to Combine Cryo-electron Microscopy and X-ray Crystallography. *Methods.* 2009; 49:174–180. [PubMed: 19398010]
12. Trabuco LG, Schreiner E, Gumbart J, Hsin J, Villa E, Schulten K. Applications of the Molecular Dynamics Flexible Fitting Method. *J. Struct. Biol.* 2011; 173:420–427. [PubMed: 20932910]
13. Whitford PC, Geggier P, Altman RB, Blanchard SC, Onuchic JN, Sanbonmatsu KY. Accommodation of Aminoacyl-tRNA into the Ribosome Involves Reversible Excursions Along Multiple Pathways. *RNA.* 2010; 16:1196–1204. [PubMed: 20427512]
14. Whitford PC, Ahmed A, Yu Y, Hennelly SP, Tama F, Spahn CMT, Onuchic JN, Sanbonmatsu KY. Excited States of Ribosome Translocation Revealed Through Integrative Molecular Modeling. *Proc. Natl. Acad. Sci. USA.* 2011; 108:18943–18948. [PubMed: 22080606]
15. Ratje AH, Loerke J, Mikolajka A, Bruenner M, Hildebrand PW, Starosta AL, Doenhoefer A, Connell SR, Fucini P, Mielke T, et al. Head Swivel on the Ribosome Facilitates Translocation by Means of Intra-subunit tRNA Hybrid Sites. *Nature.* 2010; 468:713–716. [PubMed: 21124459]
16. Tjioe E, Lasker K, Webb B, Wolfson HJ, Sali A. MultiFit: A Web Server for Fitting Multiple Protein Structures into Their Electron Microscopy Density Map. *Nucleic Acids Res.* 2011; 39:W167–W170. [PubMed: 21715383]
17. Tama F, Brooks CL III. The Mechanism and Pathway of pH Induced Swelling in Cowpea Chlorotic Mottle Virus. *J. Mol. Biol.* 2002; 318:733–747. [PubMed: 12054819]
18. Tama F, Wriggers W, Brooks CL III. Exploring Global Distortions of Biological Macro-molecules and Assemblies from Low-Resolution Structural Information and Elastic Network Theory. *J. Mol. Biol.* 2002; 321:297–305. [PubMed: 12144786]
19. Tama F, Valle M, Frank J, Brooks CL III. Dynamic Reorganization of the Functionally Active Ribosome Explored by Normal Mode Analysis and Cryo-electron Microscopy. *Proc. Natl. Acad. Sci. USA.* 2003; 100:9319–9323. [PubMed: 12878726]
20. Tama F, Brooks CL III. Diversity and Identity of Mechanical Properties of Icosahedral Viral Capsids Studied with Elastic Network Normal Mode Analysis. *J. Mol. Biol.* 2005; 345:299–314. [PubMed: 15571723]
21. Tama F, Brooks CL III. Symmetry, Form, and Shape: Guiding Principles for Robustness in Macromolecular Machines. *Ann. Rev. Biophys.* 2006; 35:115–133.
22. Gumbart J, Trabuco LG, Schreiner E, Villa E, Schulten K. Regulation of the Protein-Conducting Channel by a Bound Ribosome. *Structure.* 2009; 17:1453–1464. [PubMed: 19913480]
23. Hsin J, Gumbart J, Trabuco LG, Villa E, Qian P, Hunter CN, Schulten K. Protein-Induced Membrane Curvature Investigated Through Molecular Dynamics Flexible Fitting. *Biophys. J.* 2009; 97:321–329. [PubMed: 19580770]
24. Li W, Villa E, Frank J. The Mechanism of aa-tRNA Entry into the Ribosome. *J. Biomol. Struct. Dynam.* 2009; 26:793–794.
25. Sener M, Hsin J, Trabuco LG, Villa E, Qian P, Hunter CN, Schulten K. Structural Model and Excitonic Properties of the Dimeric RC-LH1-Pufx Complex From *Rhodobacter Sphaeroides*. *Chem. Phys.* 2009; 357:188–197. [PubMed: 20161332]
26. Villa E, Sengupta J, Trabuco LG, LeBarron J, Baxter WT, Shaikh TR, Grassucci RA, Nissen P, Ehrenberg M, Schulten K, et al. Ribosome-Induced Changes in Elongation Factor Tu Conformation Control GTP Hydrolysis. *Proc. Natl. Acad. Sci. USA.* 2009; 106:1063–1068. [PubMed: 19122150]

27. Trabuco LG, Schreiner E, Eargle J, Cornish P, Ha T, Luthey-Schulten Z, Schulten K. The Role of L1 stalk-tRNA Interaction in the Ribosome Elongation Cycle. *J. Mol. Biol.* 2010; 402:741–760. [PubMed: 20691699]
28. Armache J-P, Jarasch A, Anger AM, Villa E, Becker T, Bhushan S, Jossinet F, Habeck M, Dindar G, Franckenberg S, et al. Cryo-EM Structure and rRNA Model of a Translating Eukaryotic 80S Ribosome at 5.5 Å Resolution. *Proc. Natl. Acad. Sci. USA.* 2010; 107:19748–19753. [PubMed: 20980660]
29. Armache J-P, Jarasch A, Anger AM, Villa E, Becker T, Bhushan S, Jossinet F, Habeck M, Dindar G, Franckenberg S, et al. Localization of Eukaryote-Specific Ribosomal Proteins in a 5.5 Å Cryo-EM Map of the 80S Eukaryotic Ribosome. *Proc. Natl. Acad. Sci. USA.* 2010; 107:19754–19759. [PubMed: 20974910]
30. Chan KY, Gumbart J, McGreevy R, Watermeyer JM, Sewell BT, Schulten K. Symmetry-
Restrained Flexible Fitting for Symmetric EM maps. *Structure.* 2011; 19:1211–1218. [PubMed: 21893283]
31. Schreiner E, Trabuco LG, Freddolino PL, Schulten K. Stereochemical Errors and Their Implications for Molecular Dynamics Simulations. *BMC Bioinform.* 2011; 12:190.
32. Strunk BS, Loucks CR, Su M, Vashisth H, Cheng S, Schilling J, Brooks CL III, Karbstein K, Skiniotis G. Ribosome Assembly Factors Prevent Premature Translation Initiation by 40S Assembly Intermediates. *Science.* 2011; 333:1449–1453. [PubMed: 21835981]
33. Li W, Trabuco LG, Schulten K, Frank J. Molecular Dynamics of EF-G During Translocation. *Proteins.* 2011; 79:1478–1486. [PubMed: 21365677]
34. Frauenfeld J, Gumbart J, van der Sluis EO, Funes S, Gartmann M, Beatrix B, Mielke T, Berninghausen O, Becker T, Schulten K, et al. Cryo-EM Structure of the Ribosome-SecYE Complex in the Membrane Environment. *Nat. Struct. Mol. Biol.* 2011; 18:614–621. [PubMed: 21499241]
35. Agirrezabala X, Schreiner E, Trabuco LG, Lei J, Ortiz-Meoz RF, Schulten K, Green R, Frank J. Structural Insights into Cognate Versus Near-Cognate Discrimination During Decoding. *EMBO J.* 2011; 30:1497–1507. [PubMed: 21378755]
36. Agirrezabala X, Liao HY, Schreiner E, Fu J, Ortiz-Meoz RF, Schulten K, Green R, Frank J. Structural Characterization of mRNA-tRNA Translocation Intermediates. *Proc. Natl. Acad. Sci. USA.* 2012; 109:6094–6099. [PubMed: 22467828]
37. Gumbart J, Schreiner E, Wilson DN, Beckmann R, Schulten K. Mechanisms of SecM-Mediated Stalling in the Ribosome. *Biophys. J.* 2012; 103:331–341. [PubMed: 22853911]
38. Vashisth H, Skiniotis G, Brooks CL III. Using Enhanced Sampling and Structural Restraints to Refine Atomic Structures into Low-Resolution Electron Microscopy Maps. *Structure.* 2012; 20:1453–1462. [PubMed: 22958641]
39. Rasmussen SGF, DeVree BT, Zou Y, Kruse AC, Chung KY, Kobilka TS, Thian FS, Chae PS, Pardon E, Calinski D, et al. Crystal Structure of the β_2 Adrenergic Receptor-Gs Protein Complex. *Nature.* 2011; 477:549–555. [PubMed: 21772288]
40. Maragliano L, Vanden-Eijnden E. A Temperature Accelerated Method for Sampling Free Energy and Determining Reaction Pathways in Rare Events Simulations. *Chem. Phys. Lett.* 2006; 426:168–175.
41. Maragliano L, Vanden-Eijnden E. Single-Sweep Methods for Free Energy Calculations. *J. Chem. Phys.* 2008; 128:184110. [PubMed: 18532802]
42. Abrams CF, Vanden-Eijnden E. Large-scale Conformational Sampling of Proteins Using Temperature-Accelerated Molecular Dynamics. *Proc. Natl. Acad. Sci. USA.* 2010; 107:4961–4966. [PubMed: 20194785]
43. Vashisth H, Maragliano L, Abrams CF. “DFG-flip” in the Insulin Receptor Kinase is Facilitated by a Helical Intermediate State of the Activation Loop. *Biophys. J.* 2012; 102:1979–1987. [PubMed: 22768955]
44. Al-Hashimi HM, Walter NG. RNA Dynamics: It is About Time. *Curr. Op. Struct. Biol.* 2008; 18:321–329.
45. Zhang Q, Stelzer AC, Fisher CK, Al-Hashimi HM. Visualizing Spatially Correlated Dynamics that Directs RNA Conformational Transitions. *Nature.* 2007; 450:1263–1267. [PubMed: 18097416]

46. Williamson JR. Induced Fit in RNA-protein Recognition. *Nat. Struct. Biol.* 2000; 7:834–837. [PubMed: 11017187]
47. Casiano-Negroni A, Sun X, Al-Hashimi HM. Probing Na⁺-Induced Changes in the HIV-1 TAR Conformational Dynamics Using NMR Residual Dipolar Couplings: New Insights into the Role of Counterions and Electrostatic Interactions in Adaptive Recognition. *Biochemistry.* 2007; 46:6525–6535. [PubMed: 17488097]
48. Schmitz U, Behrens S, Freymann DM, Keenan RJ, Lukavsky P, Walter P, James TL. Structure of the Phylogenetically Most Conserved Domain of SRP RNA. *RNA.* 1999; 5:1419–1429. [PubMed: 10580470]
49. Allain FHT, Howe PWA, Neuhaus D, Varani G. Structural Basis of the RNA-Binding Specificity of Human U1A Protein. *EMBO J.* 1997; 16:5764–5772. [PubMed: 9312034]
50. Aboulela F, Karn J, Varani G. The Structure of the Human-Immunodeficiency-Virus Type-1 TAR RNA Reveals Principles of RNA Recognition by TAT Protein. *J. Mol. Biol.* 1995; 253:313–332. [PubMed: 7563092]
51. Du ZH, Lind KE, James TL. Structure of TAR RNA Complexed with a Tat-TAR Interaction Nanomolar Inhibitor that was Identified by Computational Screening. *Chem. and Biol.* 2002; 9:707–712. [PubMed: 12079782]
52. Ben-Shem A, de Loubresse NG, Melnikov S, Jenner L, Yusupova G, Yusupov M. The Structure of the Eukaryotic Ribosome at 3.0 Å Resolution. *Science.* 2011; 334:1524–1529. [PubMed: 22096102]
53. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kalé L, Schulten K. Scalable Molecular Dynamics with NAMD. *J. Comput. Chem.* 2005; 26:1781–1802. [PubMed: 16222654]
54. Kalé L, Skeel R, Bhandarkar M, Brunner R, Gursoy A, Krawetz N, Phillips J, Shinozaki A, Varadarajan K, Schulten K. NAMD2: Greater Scalability for Parallel Molecular Dynamics. *J. Comput. Phys.* 1999; 151:283–312.
55. MacKerell AD Jr. Bashford D, Bellott M, Dunbrack RL Jr. Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, et al. All-atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B.* 1998; 102:3586–3616.
56. MacKerell AD Jr. Feig M, Brooks C. L., III, Extending the Treatment of Backbone Energetics in Protein Force Fields: Limitations of Gas-phase Quantum Mechanics in Reproducing Protein Conformational Distributions in Molecular Dynamics Simulations. *J. Comput. Chem.* 2004; 25:1400–1415. [PubMed: 15185334]
57. Humphrey W, Dalke A, Schulten K. VMD - Visual Molecular Dynamics. *J. Mol. Graph.* 1996; 14:33–38. [PubMed: 8744570]
58. Vanden-Eijnden E. Some Recent Techniques for Free Energy Calculations. *J. Comput. Chem.* 2009; 30:1737–1747. [PubMed: 19504587]
59. Maragliano L, Cottone G, Ciccotti G, Vanden-Eijnden E. Mapping the Network of Pathways of CO Diffusion in Myoglobin. *J. Am. Chem. Soc.* 2010; 132:1010–1017. [PubMed: 20039718]
60. Vashisth H, Brooks CL III. Conformational Sampling of Maltose-Transporter Components in Cartesian collective Variables is Governed by the Low-frequency Normal Modes. *J. Phys. Chem. Lett.* 2012; 3:3379–3384. [PubMed: 23185650]
61. Tanner DE, Chan KY, Phillips JC, Schulten K. Parallel Generalized Born Implicit Solvent Calculations with NAMD. *J. Chem. Theory Comput.* 2011; 7:3635–3642. [PubMed: 22121340]

Figure 1.

(A) Root-mean-squared-deviation (RMSD; backbone P-atoms) as a function of simulation time (ns) during MDFF fitting of each RNA into five maps of different resolutions (5Å, 10Å, 15Å, 20 Å, and 25 Å). Top panels for each RNA show the RMSD from the known initial structures and bottom panels show the RMSD from the known final structures. Different colors for traces indicate five independent MDFF runs for each map. The numbers in italics on bottom RMSD panels for each protein represent the final average (from five runs) RMSD with respect to the known target structures. (B) Cartoon representations of two different views of the overlay of final conformations generated via MDFF fitting of each RNA at each map resolution is shown. Out of five independent runs for each map resolution, only the structure with the highest correlation-coefficient is rendered. Black cartoons correspond to the known target conformations which are used as a basis to judge the quality of all fitting simulations, while cartoons of other colors correspond to the fitting with the highest correlation-coefficient for each map resolution of each RNA.

Figure 2.

MDFF vs. TAMDFF fitting of SRP-RNA in explicit-solvent. (A) Initial (black cartoon) and final (white cartoon) atomic structures of SRP-RNA. (B) Schematic representation of the simulation domain (52698 atoms) of solvated SRP-RNA: starting docked closed-conformation of SRP-RNA (black cartoon), 5 Å resolution target map (cyan surface), water molecules (wireframe), and Mg²⁺ ions (green spheres). (C) Top and bottom panels respectively show the backbone (P-atoms) RMSD traces from the known initial and final conformations of SRP-RNA. The black trace is from an MDFF simulation, while the traces of other color are from four independent TAMDFF simulations. Initial/final correlation-

coefficients for all five simulations are shown in the bottom panel. **(D)** Map-docked cartoon representations of two different views of the overlay of final conformations generated via MDFF and TAMDFF simulations are shown. Cartoon colors are same as the RMSD traces in panel C.

Figure 3. MDFF vs. TAMDFF fitting of U1A-UTR-RNA in explicit-solvent. The solvated and ionized simulation domain of U1A-UTR-RNA has 48622 atoms (panel **B**). Coloring and labeling details are same as Figure 2.

Figure 4. MDFF vs. TAMDFF fitting of TAR-RNA in explicit-solvent. (A) Initial (black cartoon) and final (white cartoon) atomic structures of TAR-RNA. (B) The RMSD traces from MDFF and TAMDFF simulations are shown. See the caption of Figure 2C for coloring and labeling details. (C) Map-docked cartoon representations of two different views of the overlay of final conformations generated via MDFF and TAMDFF simulations are shown. Representations of bases are omitted for clarity. Dotted circles highlight the differences in the final conformations generated via MDFF (black cartoon) and TAMDFF (cartoons in colors other than black) simulations. The solvated and ionized simulation system (not shown) of TAR-RNA has 40622 atoms.

Figure 5.

Explicit-solvent MDFF vs. TAMDDFF fitting of helix-44 (H44) from the mature small (40S) eukaryotic ribosomal subunit⁵² into an experimental map of a pre-40S maturation intermediate.³² **(A)** Schematic representation of the simulation domain (208390 atoms) of solvated H44 as viewed along the z-axis: starting docked conformation of H44 (black cartoon), ~18 Å resolution target map (cyan surface), water molecules (wireframe), and Mg²⁺ ions (green spheres). The additional globular blobs of density near H44 are from some accessory factor proteins (not modeled here). **(B)** The backbone (P-atoms) RMSD traces from the known initial crystal conformation of H44 (PDB code 3U5F). The black trace is

from an MDFF simulation, while the traces of other color are from five independent TAMDFF simulations. Initial/final correlation-coefficients for all six simulations are also shown. Inset highlights the RMSD traces in early parts of MDFF and TAMDFF simulations. (C) Map-docked cartoon representations are shown for two different views of the overlay of final conformations generated via MDFF and TAMDFF simulations.

Figure 6. MDFF vs. TAMDFF fitting of the complex of U1A protein with UTR-RNA in explicit-solvent. (A) Initial (black cartoon) and final (white cartoon) atomic structures. (B) The RMSD traces for RNA- and protein-fitting steps from MDFF and TAMDFF simulations are shown. Top panels show the RMSD from the initial structure and bottom panels show the RMSD from the final structure. (C) Map-docked cartoon representations of two different views of the overlay of final conformations generated via MDFF (black cartoon) and TAMDFF (red cartoon) simulations are shown. The solvated and ionized simulation system of this complex has ~95700 atoms.