Figures and figure supplements

Automated cryo-EM structure refinement using correlation-driven molecular dynamics

Maxim Igaev et al
Figure 1. Approach: using correlation-based refinement in MD simulation to steer the atomic positions of a macromolecule such that they optimally fit a cryo-EM map. The molecule is subjected to a global biasing potential $V_{\text{fit}}$ in addition to the MD force field $V_{\text{ff}}$. The forces resulting from $V_{\text{fit}}$ act on every atom to enhance the real-space correlation coefficient $c.c.$ between the cryo-EM density (green) and the density calculated from the current atomic positions (blue). The first step (1) is to generate a simulated density by convoluting the atomic positions with a three-dimensional Gaussian function of width $\sigma$ (Orzechowski and Tama, 2008). The two maps are correlated (2), and the biasing forces are calculated. These forces are then added to the standard MD force field (3), and new atomic positions are evaluated (4). Steps (1–4) are repeated, yielding a structure that correlates better with the cryo-EM map than the starting structure.

DOI: https://doi.org/10.7554/eLife.43542.002
Figure 2. Schematic representation of the proposed continuous refinement protocol: (1) a low temperature optimization phase, where $V_{fi}$ is monotonously increased by increasing the force constant $k$ (columns a–d), followed by (2) simulated annealing (columns e, f). The local effect of the protocol is exemplified in the upper row for a one-dimensional single-atom case. Simulated densities shown in the middle row were generated using the atomic structure of a tubulin dimer (PDB ID: 3JAT; Zhang and Nogales, 2015).

DOI: https://doi.org/10.7554/eLife.43542.003
Figure 2—figure supplement 1. Detailed scheme of the proposed continuous refinement protocol subdivided in five stages. In stage 1 (yellow), the starting structure is subjected to initial equilibration in explicit solvent. Equilibration at $T = 300$ K for 50 ns is needed to drive the starting structure further away from the target state, posing an additional refinement challenge. In real applications, a short equilibration run at $T = 100$ K for 5–10 ns should be sufficient. Stages 2 and 3 (gray and green) include the refinement against one of the half-maps (training map) followed by cross-validation by means of Fourier Shell Correlation (FSC) against the other half-map (validation map). In case no overfitting of the training map is observed, the final structure from the half-map refinement is passed to stage 4 (cyan), where it is refined against the full map to account for high-resolution features not present in both half-maps. Finally, in stage 5 (purple), the average structure from the last 5 ns of the full map refinement is subjected to geometry and goodness-of-fit assessment. Note that the procedure is automated such that stages 1–4 can be run without human intervention. Pausing the refinement to inspect intermediate results or introduce changes to the protocol should be possible at any stage. See Materials and methods for more details.

DOI: https://doi.org/10.7554/eLife.43542.004
Figure 3. Refining a distant starting model into a high-resolution map: rabbit muscle aldolase at 2.6 Å. (a) RMSD (Cα atoms) between the starting and the reference model (5VY5) showing the extent of rearrangements during refinement. (b) Reciprocal-space agreement of the starting (black dashed), the reference (gray) and our refined model (green) with the full map (top) and stereochemical quality for the three models assessed by EMRinger and MolProbity (bottom). (c) Representative secondary structure elements showing local agreement of our model with the full map. (d) Representative region of the protein interior (chain A) showing the closeness of our model and the reference to higher resolution control X-ray structures in terms of RMSD. Some residues are explicitly labeled.

DOI: https://doi.org/10.7554/eLife.43542.005
Figure 3—figure supplement 1. Extension of Figure 3 showing the time evolution of various characteristics during refinement. (a) The simulated map resolution \( \sigma \) and force constant \( k \) were linearly ramped from 0.6 nm and \( 0.5 \times 10^5 \) kJ mol\(^{-1} \) to the target values of 0.2 nm and \( 5 \times 10^5 \) kJ mol\(^{-1} \), respectively. At the bottom, the time evolution of \( c.c. \) with the training and the full map and RMSD to the reference structure are shown. (b) FSC\(_{\text{train}}\) and FSC\(_{\text{val}}\) sampled every five ns. Cross-validation showed no signs of overfitting (black dashed lines).

DOI: https://doi.org/10.7554/eLife.43542.006
Figure 3—figure supplement 2. Extension of Figure 3 showing the comparison of the radii of convergence across different refinement methods for the aldolase system and using the same distant starting structure. RMSD to the reference structure for all methods used is shown in the upper left plot. Overlays between the reference (gray) and the refined structures are depicted as ribbons (colors as in the RMSD plot).

DOI: https://doi.org/10.7554/eLife.43542.007
Figure 4. Refinement of a curved tubulin dimer into a map of the straight, microtubule-like state. (a) RMSD (Cα atoms) between the starting and the control model (6DPV) showing the extent of rearrangements between the solution (curved) and the microtubule-like (straight) tubulin conformation. The α-subunits of both models were aligned for the RMSD calculation. (b) Reciprocal-space agreement of the starting (black dashed), the control (gray and dark gray) and our model (cyan) with the full map (top) and stereochemical quality for the four models assessed by EMRinger and MolProbity (bottom). (c) Representative secondary structure elements showing local agreement of our model with the full map. (d) Same as in c but showing the closeness of our model to higher resolution control structures in terms of RMSD. Some residues are explicitly labeled.

DOI: https://doi.org/10.7554/eLife.43542.008
Figure 4—figure supplement 1. Extension of Figure 4 showing the comparison of the radii of convergence across different refinement methods for the tubulin system and using the same distant starting structure. RMSD to the reference structure for all methods used is shown in the upper left plot. Overlays between the control (6DPV, gray) and the refined structures are depicted as ribbons (colors as in the RMSD plot).

DOI: https://doi.org/10.7554/eLife.43542.009
Figure 5. Refinement of a poor structure of TRPV1 in a distant conformation into a map with highly heterogeneous local resolution. (a) RMSD (Cα atoms) between the starting and the reference model (3J5P) showing the extent of rearrangements the TRPV1 structure undergoes during refinement. (b) Reciprocal-space agreement of the starting (black dashed), the reference (gray) and our model (purple) with the full map (top) and stereochemical quality of the models assessed by EMRinger and MolProbity (bottom). Note that the starting model was derived from the deposited structure by subjecting the latter to MD at $T = 300$ K. (c) Representative secondary structure elements showing local agreement between of our model with the full map.

DOI: https://doi.org/10.7554/eLife.43542.010
Figure 5—figure supplement 1. Extension of Figure 5 showing the comparison of the radii of convergence across different refinement methods for the TRPV1 system and using the same distant starting structure. RMSD to the reference structure for all methods used is shown in the upper left plot. Overlays between the reference (3JSP, gray) and the refined structures are depicted as ribbons (colors as in the RMSD plot).

DOI: https://doi.org/10.7554/eLife.43542.011
Figure 5—figure supplement 2. Convergence of the TRPV1 refinement both in the higher resolution TM region and in the lower resolution ARD region assessed by means of three independent refinement runs using different but similarly distant starting structures. Only one TRPV1 monomer and no side chains are shown for clarity.

DOI: https://doi.org/10.7554/eLife.43542.012
Figure 6. Comparison of our TRPV1 model with those previously refined using Rosetta (a, b) and ReMDFF (c, d). Overlays of our model (pink and violet ribbon) with the Rosetta (left, gray ribbon) and ReMDFF (right, gray ribbon) models are shown in (a) and (c), respectively. Reciprocal-space agreement with the full map (top) and stereochemical quality for the four models assessed by EMRinger and MolProbity (bottom) are shown in (b) for the Rosetta model and in (d) for the ReMDFF model.
DOI: https://doi.org/10.7554/eLife.43542.013
Figure 7. Refinement of a substrate-free NSF complex in a distant conformation into a medium-resolution map at 3.9 Å. (a) RMSD (Cα atoms) between the starting and the reference model (6MDO) showing the extent of rearrangements the NSF structure undergoes during refinement. (b) Reciprocal-space agreement with the full map for the starting (black dashed), the reference (gray) and the set of final models (red gradient) refined using a wide range of target force constants (top) and stereochemical quality assessed by EMRinger and MolProbity (bottom). (c) Representative secondary structure elements showing local agreement of the model refined at $k = 5 \times 10^5$ kJ mol$^{-1}$ with the map. (d) ATP binding pocket of the D2 domain (chain A) showing the closeness of our model and the reference to higher-resolution control X-ray structures in terms of RMSD. Some residues are explicitly labeled.

DOI: https://doi.org/10.7554/eLife.43542.014
Figure 7—figure supplement 1. Extension of Figure 7 showing the comparison of the radii of convergence across different refinement methods for the NSF system and using the same distant starting structure. RMSD to the reference structure for all methods used is shown in the upper left plot. Overlays between the reference (6MDO, gray) and the refined structures are depicted as ribbons (colors as in the RMSD plot).

DOI: https://doi.org/10.7554/eLife.43542.015
Figure 8. Refinement of a distant nucleosome structure into a medium-resolution map of the canonical nucleosome state. (a) RMSD (DNA and protein backbone) between the starting and the reference model (6ESF) showing the extent of rearrangements during the refinement. (b) Reciprocal-space agreement of the starting (black dashed), the reference (gray) and our model (yellow) with the full map (top) and stereochemical quality assessed by EMRinger and MolProbity (bottom). (c) Representative secondary structure elements showing local agreement of our model with the full map. (d) Representative region next to the dyad DNA showing the closeness of our model to higher-resolution control structures in terms of RMSD (only protein non-hydrogen atoms). Some residues are explicitly labeled.

DOI: https://doi.org/10.7554/eLife.43542.016
Figure 8—figure supplement 1. Extension of Figure 8b showing the reciprocal-space agreement and stereochemical quality for nucleosome models independently refined using force constants ranging from 2 to 4.5 \( \times 10^5 \) kJ mol\(^{-1}\). The structure shown in Figure 8 was refined using \( k = 3 \times 10^5 \) kJ mol\(^{-1}\).

DOI: https://doi.org/10.7554/eLife.43542.017
Figure 8—figure supplement 2. Extension of Figure 8 showing the comparison of the radii of convergence across different refinement methods for the nucleosome system and using the same distant starting structure. RMSD to the reference structure for all methods used is shown in the upper left plot. Overlays between the reference (6ESF, gray) and the refined structures are depicted as ribbons (colors as in the RMSD plot).

DOI: https://doi.org/10.7554/eLife.43542.018
**Figure 9.** Refinement of a ribosome complex in the CR state into a 3.4 Å map of the GA state. (a) RMSD (RNA and protein backbone) between the starting (CR) and the final (GA) model showing the extent of rearrangements during the refinement. (b) Reciprocal-space agreement of the starting (black dashed), the reference (gray) and our model (orange) with the full map (top) and stereochemical quality for the three models assessed by EMRinger and MolProbity (bottom). (c) Representative regions showing local agreement between our model and the map (the codon-anticodon region and a RNA stem-loop in the 30S subunit). (d, e) Representative regions in the ribosomal exit tunnel (constriction site formed by L4 and L22 protein chains is shown) and in the SelB-mRNA contact interface, both demonstrating the closeness of our model to higher resolution control structures in terms of RMSD. Some protein residues are explicitly labeled.

DOI: https://doi.org/10.7554/eLife.43542.019
Figure 9—figure supplement 1. Extension of Figure 9 showing the comparison of the radii of convergence across different refinement methods for the ribosome system and using the same distant starting structure. RMSD to the reference structure for all methods used is shown in the upper left plot. Overlays between the reference (SLZD, gray) and the refined structures are depicted as ribbons (colors as in the RMSD plot).

DOI: https://doi.org/10.7554/eLife.43542.020
Figure 10. Refinement of a CorA magnesium transporter in the symmetric closed state into a low-resolution 7.1 Å map of the asymmetric open state. (a) RMSD (Ca atoms) between the starting (closed) and the reference (open) model showing the extent of rearrangements in the cytosolic part of the channel during refinement. (b) Reciprocal-space agreement of the starting (black dashed), the reference (gray) and our model refined with $k = 1.0 \times 10^5$ kJ mol$^{-1}$ (sea green) with the full map (top) and stereochemical quality for the three models assessed by EMRinger and MolProbity (bottom). The FSC curves were calculated using the backbone atoms only, whereas the full-atom models were used for geometry analysis. No EMRinger scores were calculated. (c) Same as in (b) but for the two MDFF models refined with (blue) and without (dark blue) secondary structure, chirality and cis peptide bond restraints. (d) Overlay of our full-atom model (see green) with full map. DOI: https://doi.org/10.7554/eLife.43542.022
Figure 10—figure supplement 1. Extension of Figure 10 showing map-model agreement vs. rotamer or Ramachandran outliers for all CDMD and MDFF refinements. Average FSC values ($FSC_{avg}$) were calculated as described in Materials and methods. Each dot represents the result of a single independent refinement. MDFF force constants ranged from 0.05 to 0.5 (see Materials and methods for the MDFF protocol). Arrows indicate the trend in $FSC_{avg}$, rotamer and Ramachandran outliers as the force constant increases.

DOI: https://doi.org/10.7554/eLife.43542.023
Figure 10—figure supplement 2. Extension of Figure 10 showing the structure of the gating pore. The pore transmembrane helices (residues 281–312) are shown for clarity.

DOI: https://doi.org/10.7554/eLife.43542.024
Figure 10—figure supplement 3. Extension of Figure 10 showing how the pore radius changes along the nonlinear gating pathway shown in Figure 10—figure supplement 2 (bottom). Residues facing the gating pathway are shown as dots color-coded by their relative hydrophobicities. DOI: https://doi.org/10.7554/eLife.43542.025
Figure 11. Per-residue outlier analysis. Per-residue rotamer and Ramachandran outlier propensities calculated from the non-ribosome structures (left panels; CHARMM22*/CHARMM36m force fields were used \cite{Piana2011, Huang2017} and the ribosome structure (right panels; AMBER12sb was used; \cite{Lindorff-Larsen2010}) after pre-equilibration with the force field (gray dots) and after the actual refinement with our approach (blue and red dots, respectively).

DOI: https://doi.org/10.7554/eLife.43542.026