

RosettaES: a sampling strategy enabling automated interpretation of difficult cryo-EM maps

Brandon Frenz¹, Alexandra C Walls¹,
Edward H Egelman², David Veessler¹ & Frank DiMaio¹

Accurate atomic modeling of macromolecular structures into cryo-electron microscopy (cryo-EM) maps is a major challenge, as the moderate resolution makes accurate placement of atoms difficult. We present Rosetta enumerative sampling (RosettaES), an automated tool that uses a fragment-based sampling strategy for *de novo* model completion of macromolecular structures from cryo-EM density maps at 3–5-Å resolution. On a benchmark set of nine proteins, RosettaES was able to identify near-native conformations in 85% of segments. RosettaES was also used to determine models for three challenging macromolecular structures.

Accurate atomic models of macromolecular structures are invaluable for understanding the biochemical and cellular processes carried out by proteins. Recent advances in cryo-EM^{1,2} have led to a dramatic expansion in the number of structures that can be studied at high resolution. Though several reconstructions have been achieved at resolutions of 2.5 Å or better³, it is far more common to obtain resolutions between 3–5 Å.

At resolutions worse than 3 Å, model building is challenging and error prone. Semiautomated methods for atomic model building exist, but these methods are laborious and prone to user bias^{4,5}. Automatic model building methods have been developed for X-ray crystallography and work well for cryo-EM data up to 3.0-Å resolution, but such methods have difficulty registering sequence to backbone at lower resolutions^{6–9}. Methods combining homology modeling with rigid-body docking and flexible fitting have also been developed but are limited to cases where a relatively accurate starting model is available^{10–13}. Cryo-EM-specific methods have been developed for structure determination but similarly have difficulty correctly identifying sequence^{14,15}. We previously developed a fully automated *de novo* method¹⁶ that does not have these restrictions; however, its algorithm for model completion required >70% of the protein be placed correctly, which limited its applicability. Here, we describe an approach that overcomes this limitation.

Our approach, RosettaES, further automates, improves upon, and expedites *de novo* model building using 3–5-Å resolution cryo-EM data. Our method uses fragment-based sampling to enumerate a ‘pool’ of possible protein conformations that both possess physically realistic geometry and are consistent with the experimental density data. We use several strategies for pruning this set of solutions during sampling, ensuring the size of this solution pool is reasonable even when building segments with dozens of residues.

The RosettaES algorithm uses a ‘beam search’, in which a fixed-size ensemble of partial structures is maintained throughout sampling (Supplementary Fig. 1; see Online Methods). The method aims to complete a partial model guided by EM density. Missing segments are built iteratively; starting from residues immediately N- or C-terminal to a missing segment, putative solutions are spawned that add one additional residue and sample the conformation of the most recently placed three residues guided by backbone segments from high-resolution structures with similar local sequence. This ensemble is then pruned of models that are: (i) energetically unfavorable, (ii) inconsistent with the data, or (iii) too similar to another solution in the ensemble. Following this pruning, if the ensemble exceeds a predefined maximum size, then models are clustered so that the pool size does not exceed this limit.

For protein structures missing multiple interacting segments, individual segments are sampled and combined in a Monte Carlo assembly algorithm identifying internally consistent combinations; if this fails additional rounds of search are performed in order to force increased diversity¹⁷. Finally, several features increase the stringency of ensemble selection: (i) an explicit penalty on atomic models tracing discontinuous density, (ii) agreement of a model to side-chain density, and (iii) identification of putative strand pairs when growing β strands (Fig. 1).

We benchmarked RosettaES on a set of nine proteins, each with missing segments of between 11 and 160 residues (Supplementary Table 1); we assumed the native structure of all other regions was known. We compared the performance of RosettaES to that of RosettaCM¹⁸ by reporting the high-accuracy version of the global distance test (GDT-HA)¹⁹—a measure that roughly reports the fraction of residues correctly placed—of each completed segment between the two approaches (Fig. 1a). RosettaES outperformed or matched RosettaCM across all residue ranges (Supplementary Table 1) and was able to generate and select accurate solutions up to 111 residues in length, including a 78-residue segment in FrhA (Fig. 1c). We also compared the performance of RosettaES with that of Buccaneer²⁰, which accurately assigned (placed correct sequence within 2.0 Å) 16% of the total residues in the benchmark, while RosettaES accurately assigned 76% (Supplementary Table 2). We also find that the diversity of the final solutions

¹Department of Biochemistry, University of Washington, Seattle, Washington, USA. ²Department of Biochemistry and Molecular Genetics, University of Virginia, Charlottesville, Virginia, USA. Correspondence should be addressed to F.D. (dimaio@uw.edu).

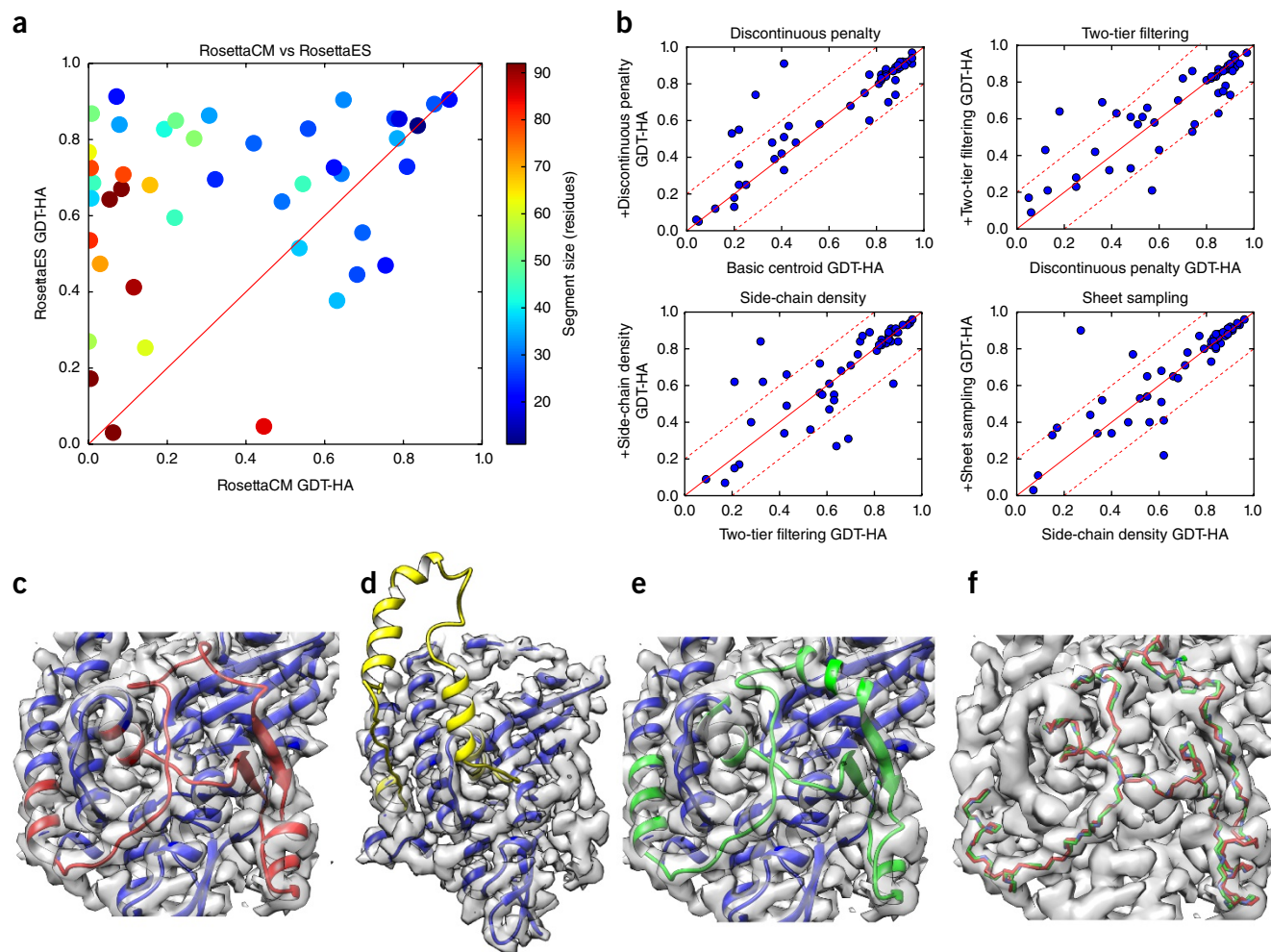


Figure 1 | Accuracy of RosettaES compared with that of RosettaCM. (a) A comparison of RosettaES to RosettaCM reporting GDT-HA over the backbone + C β atoms on a benchmark set of single missing segments extracted from the deposited model. Length of the segment is colored such that longer segments appear a darker red. The x- and y-axes correspond to the GDT-HA of the model compared with the deposited structures under two conditions. The closer the GDT-HA is to 1, the more similar the structures. Values above the solid line indicate a more accurate solution with RosettaES than RosettaCM. (b) The GDT-HAs over the backbone + C β atoms of all the atomic models in the benchmark set as features (discontinuous penalty, two-tier filtering, side-chain density, and sheet sampling) are added. (c) The deposited structure of FrhA (PDB 4ci0) shown in the cryo-EM density map, with the starting model in blue and the removed region in red. Residues 187–265 (highlighted in red) were removed in the benchmark. (d) The top-scoring solution generated by RosettaCM, shown in yellow. (e) The top-scoring solution generated by RosettaES, shown in green. (f) Minimal backbone trace of the deposited model (in red) compared with the one produced by RosettaES (green). The two have a 1.9-Å backbone and C β r.m.s. deviation.

serves as a good metric for assessing the accuracy of the solution set (**Supplementary Fig. 2**).

RosettaES rebuilds multiple interacting segments by independently sampling and combining, iterating the process as necessary until a set of nonclashing solutions is found. Resulting nonclashing models are then refined and ranked using Rosetta. Using the benchmark set described above, we ran the full assembly process—in this case not assuming other segments were known a priori—with models ranging from 20–80% complete. In four cases (FrhA, FrhB, FrhG, and TMV), the final backbone r.m.s. deviation (C α atoms only) was ≤ 2 Å from the deposited model; in one case (TRPV1), RosettaES produced a nonclashing atomic model about 3.6 Å from the deposited model. In the remaining cases (BPP1, VP6, STIV, and T20S), which all contained large β sheets, RosettaES produced a solution matching the deposited model for most (but not all) missing segments (**Supplementary Table 3**).

We also had previously applied RosettaES in determining the structures of three very challenging proteins. The first included domains C and D of the mouse hepatitis virus (MHV) spike protein at 4.0-Å resolution²¹. The MHV spike protein is a homotrimer comprising ~1,300 residues per protomer, the bulk of which was solved using homology modeling and *de novo* fragment docking. However, only 30 of the 180 residues comprising domains C and D could be assigned using *de novo* fragment docking. Starting from this 30-residue fragment, we completed an atomic structure using RosettaES and tracing in parallel. The top-scoring RosettaES models were visually selected and refined^{22–24}. RosettaES was carried out in multiple steps; the 154-residue C terminus of domain D was poorly converged when run to completion, so it was instead broken into two separate runs of 125 and 29 residues, where the best model after refinement of the first run served as a seed for the second. Comparing this model with a manually traced model

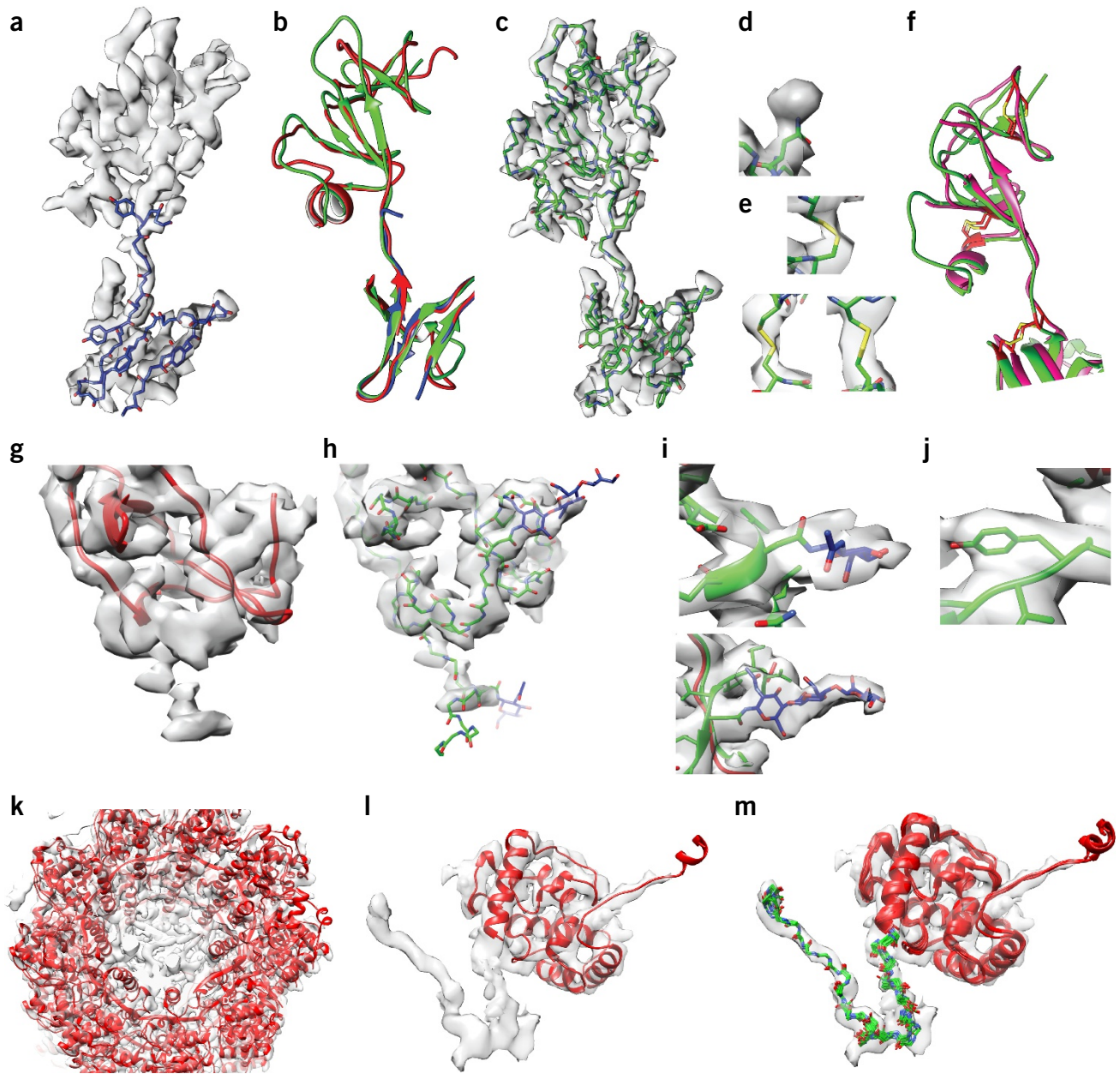


Figure 2 | RosettaES enables structure determination in challenging cases. (a–f) Building domains C and D of the MHV coronavirus spike with RosettaES. (a) A 30-residue segment of MHV domain C was placed by Rosetta *de novo* fragment docking¹⁴. (b) The model completed by RosettaES (green) and a hand-traced model (red). (c) The RosettaES-generated model shown in the cryo-EM density map. Large aromatic residues are shown as sticks. (d) A tube of density shown at the putatively glycosylated ASN 657 side chain. (e) Correct positioning of cysteines to form three unique disulfide bonds. (f) A recently determined structure of HKU1 spike protein (magenta, PDB 5i08) matches the topology obtained by RosettaCM (green; cysteines are highlighted in red). (g) Density for the C-terminal tail of HCoV-NL63. Docked in red is the partial model of HCoV-NL63 built using our structure of MHV as a template. (h) The final structure, after completion with RosettaES, attachment of glycans (shown in blue), and refinement. (i) Placement of Asp 1201 and Asp 1218 (green) and glycans (blue) in the density map. (j) Placement of tyrosine 1227 in the density map. (k) A symmetric homology model of the papaya mosaic virus (PDB 4DOX) docked into the reconstruction of the bamboo mosaic virus. The C termini in the core are missing from the model. (l) A close-up view of the asymmetric unit with the homologous structure shown in red. (m) The top-scoring models produced by RosettaES, shown in green, placed into the density map.

revealed significant topological differences between the two. The RosettaES model displays several key features that suggest it is correct: three disulfide bonds accounted for by the density and the observation of extra density corresponding to a glycosylated asparagine residue. The recent determination of the orthologous HCoV-NL63 (ref. 25) and HKU1 (ref. 26) spikes confirm our model (Fig. 2a–f; PDB 3jcl).

The HCoV-NL63 spike protein has a similar architecture to that of the MHV spike²⁵. In this 3.4-Å-resolution map it was possible to resolve residues in the C termini of the HCoV-NL63 spike protomers that are invisible in the reconstruction of MHV and difficult to place manually. RosettaES assigned residues 1,197 to 1,224; and it explained the density better than a hand-traced model (data not shown). The accuracy of the atomic model we

obtained is supported by several of its key features, including the position of two asparagines with corresponding glycan density and good agreement of several hydrophobic residues into density (Fig. 2g–j).

A cryo-EM reconstruction of the flexible filamentous bamboo mosaic virus²⁷ was consistent with a crystal structure of the homologous papaya mosaic virus but featured additional density corresponding to the termini (Fig. 2k,l); modeling with RosettaCM proved difficult, as this segment lacked regular secondary structure. RosettaES determined a well-converged ensemble of models consistent with the 5.6-Å resolution density (Fig. 2m). As additional support for our assignment, a recent structure of the homologous pepino mosaic virus shows similar architecture for the terminus.²⁸

RosettaES should expand the range of atomic models that can be determined to include long unidentified segments, extended protein loops, and β sheets. RosettaES outperforms other approaches; it can reliably generate atomic models on unassigned segments up to 50 residues and can occasionally generate accurate models over 100 residues in length.

The improved efficiency of RosettaES results from the use of experimental data to incrementally guide sampling. Its high performance stems from the use of short three-residue fragments that can completely cover tripeptide conformational space (Supplementary Fig. 3; see Online Methods). Because a subset of fragments exists to generate an atomic model similar to the one deposited for any target in our benchmark, we can say with certainty that any failure to find an accurate atomic model occurs because the cap on the number of partial solutions is set too low. In some cases, a modest increase in cap size improves accuracy dramatically (Supplementary Fig. 4). Computational time scales linearly with this cap; a typical run for a 30-residue segment is 1.5 h on a 16-core machine (Supplementary Fig. 5).

There are several ways in which these failures due to insufficient exploration may be detected. When a segment is solved in the context of the rest of the protein, the convergence of the final solutions provides a good indication of the accuracy of the final model (Supplementary Fig. 2). Additionally, when multiple segments are assigned, the lack of convergence to a nonclashing conformation suggests that greater sampling is needed. RosettaES should be a useful tool for automated model building in near-atomic-resolution cryo-EM maps; its potential is here exemplified by its successful use in the determination of three novel protein structures.

METHODS

Methods, including statements of data availability and any associated accession codes and references, are available in the [online version of the paper](#).

Note: Any Supplementary Information and Source Data files are available in the [online version of the paper](#).

ACKNOWLEDGMENTS

This work was facilitated through the use of advanced computational, storage, and networking infrastructure provided by the Hyak supercomputer system and funded by the STF at the University of Washington. This work was also supported by the National Institutes of Health (NIH) under award number GM120553 (D.V.), GM035269 (E.H.E.), and T32GM008268 (A.C.W.).

AUTHOR CONTRIBUTIONS

B.F. performed the research and drafted the manuscript. F.D. supervised the research and edited the draft. D.V. and A.C.W. provided the reconstruction of MHV and HCoV-NL63 and performed manual tracing and analysis. E.H.E. provided the reconstruction of the bamboo mosaic virus.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

- McMullan, G., Chen, S., Henderson, R. & Faruqi, A.R. *Ultramicroscopy* **109**, 1126–1143 (2009).
- Campbell, M.G. *et al. Structure* **20**, 1823–1828 (2012).
- Merk, A. *et al. Cell* **165**, 1698–1707 (2016).
- Baker, M.L. *et al. J. Struct. Biol.* **174**, 360–373 (2011).
- Emsley, P. & Cowtan, K. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).
- Jones, T.A., Zou, J.Y., Cowan, S.W. & Kjeldgaard, M. *Acta Crystallogr. A* **47**, 110–119 (1991).
- Terwilliger, T.C. *Acta Crystallogr. D Biol. Crystallogr.* **59**, 38–44 (2003).
- DePristo, M.A., de Bakker, P.I.W., Johnson, R.J.K. & Blundell, T.L. *Structure* **13**, 1311–1319 (2005).
- Terwilliger, T.C. *et al. Acta Crystallogr. D Biol. Crystallogr.* **64**, 61–69 (2008).
- Trabuco, L.G., Villa, E., Mitra, K., Frank, J. & Schulten, K. *Structure* **16**, 673–683 (2008).
- Orzechowski, M. & Tama, F. *Biophys. J.* **95**, 5692–5705 (2008).
- Schröder, G.F., Brunger, A.T. & Levitt, M. *Structure* **15**, 1630–1641 (2007).
- Šali, A. & Blundell, T.L. *J. Mol. Biol.* **234**, 779–815 (1993).
- Chen, M., Baldwin, P.R., Ludtke, S.J. & Baker, M.L. *J. Struct. Biol.* **196**, 289–298 (2016).
- Lindert, S. *et al. Structure* **20**, 464–478 (2012).
- Wang, R.Y.-R. *et al. Nat. Methods* **12**, 335–338 (2015).
- Pirim, H., Bayraktar, E. & Eksioğlu, B. in *Tabu Search* (Ed. Jaziri, W.) Chapter 1 (InTech, 2008).
- Song, Y. *et al. Structure* **21**, 1735–1742 (2013).
- Kopp, J., Bordoli, L., Battey, J.N.D., Kiefer, F. & Schwede, T. *Proteins* **69**, 38–56 (2007).
- Cowtan, K. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 1002–1011 (2006).
- Walls, A.C. *et al. Nature* **531**, 114–117 (2016).
- Coleman, C.M. & Frieman, M.B. *J. Virol.* **88**, 5209–5212 (2014).
- Forgie, S. & Marrie, T.J. *Semin. Respir. Crit. Care Med.* **30**, 67–85 (2009).
- DiMaio, F. *et al. Nat. Methods* **12**, 361–365 (2015).
- Walls, A.C. *et al. Nat. Struct. Mol. Biol.* **23**, 899–905 (2016).
- Kirchdoerfer, R.N. *et al. Nature* **531**, 118–121 (2016).
- DiMaio, F. *et al. Nat. Struct. Mol. Biol.* **22**, 642–644 (2015).
- Agirrezabala, X. *et al. eLife* **4**, e11795 (2015).

ONLINE METHODS

RosettaES uses a greedy conformational sampling strategy to assemble protein backbones consistent with local sequence and experimental density data. It uses a beam search, in which a fixed-size ensemble of partial solutions is maintained throughout sampling. Sampling attempts to model each residue one at a time; starting at the terminal residue adjacent to a missing segment, we perform a ‘growing step’ in which one residue is added, and the conformation of up to the previous nine residues is sampled. Each generated solution is evaluated against the experimental data and added to the ‘beam’—that is, the pool of partial models. Following each sampling step, the model pool is culled to contain at most M solutions (for most experiments described in this manuscript, $M = 64$ or 128). This process is repeated until all missing residues have been assigned.

At each culling step conformations are selected to ensure only those that are physically realistic and in agreement with the data are carried forward. Additional filters are used to ensure that overly similar atomic models are removed. A final solution is selected based on the Rosetta energy augmented with a ‘fit to density’ energy term. Finally, in cases where multiple missing segments are present, partial solutions for each segment are sampled, then Monte Carlo sampling is used to find a consistent set of segments (that is, a set of solutions that do not clash with each other and that agree with the experimental data).

Conformational sampling. Sampling is guided by protein fragments, generated using Rosetta’s fragment picker²⁹, which finds high-resolution structures with similar local sequences. For benchmarking, all homologous proteins (psiblast E-value < 0.05) were excluded from the search. We have found that the best performance results from the use of 100 three-residue fragments and 20 nine-residue fragments at each position; the three residue fragments accurately recapitulate the diversity of these regions, while the nine residue fragments help in modeling the n to $n + 4$ hydrogen bond patterns in alpha helices. In each iteration a new residue is added to the expanding model, and an N -residue fragment manipulates the new residue and the $(N - 1)$ residues preceding it.

After each fragment insertion, all residues added by RosettaES (and up to 15 residues on the ‘stem’) are first minimized using a low-resolution ‘centroid’ representation that models side chains with a single interaction center. The ensemble is then culled to at most twice the size limit of the ensemble ($2 \times M$); each of the structures in the ensemble is then converted to an all-atom representation and subjected to two rounds of side-chain minimization and repacking. Since the structures we evaluate are incomplete, we use a modified centroid energy that only includes terms for the Ramachandran and omega angles, van der Waals energies, short- and long-range hydrogen bonds, and fit to density. Following all-atom refinement of the best $2 \times M$ centroid structures, only the fit-to-density score from the all-atom model is used, in combination with the energetic terms from the centroid model; the best M structures are then selected from this pool.

For internal polypeptide segments, when fewer than ten residues remain to be placed, distance constraints are used to penalize conformations that cannot be closed. These constraints make sure that the backbone N and C termini are within 32 Å when they are ten residues apart; they decrease by 3 Å as the gap closes by one.

They are also used in minimization, where a harmonic penalty is applied beyond this maximum allowed distance. The final minimization (with no gap) uses a very tight 0.5-Å penalty.

Filtering the pool of structures. A key component of the approach is the filtering step, where all sampled structures are culled to at most M structures, with M being the user-defined cap on ensemble size at each step. The first step in filtering is removing conformations that are inconsistent with the data; this is done by removing all conformations that do not score at least 85% as well (using Rosetta energy plus fit-to-density energy) as the best solution of the beam. If this results in an ensemble pool that is still larger than the cap, then these solutions are then filtered in two passes; first, ‘nearly redundant’ conformations (pairs where every residue has <1.2-Å r.m.s. deviation) are merged, keeping the lowest energy structure; next, all remaining structures are clustered with a 3-Å radius, with structures chosen from each cluster round robin. In this last step, models are chosen by energy, but two structures can be chosen from a single cluster until every cluster has one representative taken. Clustering in this way ensures that reasonable diversity is maintained throughout sampling.

In addition to the scoring terms described above, we have modified the score function in three ways, which helps to improve accuracy by penalizing incorrect conformations, effectively increasing the size of the ensemble pool. These terms are described below.

Continuous density penalty. The density energy calculated in Rosetta is based on the correlation between the density expected from the model and the experimental density. However, we found that, with the conformational sampling of RosettaES, solutions which jumped between distinct backbone paths in the density were not properly penalized. Therefore, we came up with a penalty scheme that attempted to more strongly penalize paths that traveled through density discontinuities.

The ‘fast density’ scoring function in Rosetta³⁰ computes the density score as the sum of per-atom scores—which are quickly computed by convoluting one atom’s density with the experimental map—and interpolating the resulting map to calculate scores. The normal density score $E_d = \sum_i e_i$ is computed as the sum scores for all atoms. We can also compute a ‘local discontinuity’ score that considers the worst scoring atom k in a stretch from residues $N - 8$ to $N + 8$; a modified score E_d^* computes the score for each atom as $\sum_i e_k + 0.3(e_i - e_k)$; that is, we only take 30% of the score above the worst scoring atom in the local region. Finally, we use the ratio of E_d^* to E_d to scale the (correlation-based) density score used by Rosetta.

β -sheet sampling. In order to form a proper β sheet correct hydrogen bond patterning must exist between complementary strands. When building a strand in a partial model in which the adjacent strand(s) is (are) missing, properly orienting hydrogen bond donors and acceptors is challenging. To handle this, we consider explicitly modeling adjacent strands during sampling. When in a β region of Ramachandran space, we consider an idealized four-residue complementary β strand on both sides of the four most recently placed residues. Strands are aligned in an antiparallel orientation with perfect hydrogen bond geometry. Grown strands that clash with any other atom in the model are removed, and any remaining strands are refined with the growing segment. Following refinement, if strands poorly match density (if any atom scores lower than $0.7e_k$, with e_k defined in the previous

subsection), they are removed. Remaining strands are added to the energy of the respective partial model (the full Rosetta energy), and 50% of the density energy is added.

Side-chain density. Refinement is primarily carried out in Rosetta's low-resolution 'centroid' mode, largely for reasons of speed, as all-atom refinement is slower, and the landscape is much more rugged, requiring significantly more optimization. However, a key disadvantage of this approach is that side-chain density—which is limited but present at these modest resolutions—is ignored. Thus, we have developed a strategy to use low-resolution modeling augmented with all-atom fit to density.

Following low-resolution refinement, the ensemble is filtered to contain twice the number of structures allowed by the cap. These structures are then converted to all atom, and side-chain rotamers are discretely optimized and then minimized. The density score of the all-atom model is then used to replace the density score of the low-resolution model. To penalize side chain density discontinuities, we use a strategy similar to that used for penalizing backbone density discontinuities: all atoms receive a score equal to the lesser of (i) the per-atom score or (ii) the score of the worst atom of C β or C γ plus 30% of the difference between this and the per-atom score. This avoids giving a bonus to, for example, arginines placing their side chains in regions of unassigned backbone.

Multiloop assembly. When multiple missing segments are present in the atomic model (Fig. 1a), we subdivide the problem into two steps. First, for each segment (independently), we find the ensemble of plausible paths. Then, we find all sets of per-segment solutions that are consistent with each other. To do so, we have implemented a Monte Carlo search algorithm, referred to as Monte Carlo Assembly (MCA), which takes the results from each segment and attempts to find a nonclashing set. The energy of an assignment is the sum of individual density scores, scaled by segment size, plus all pairwise van der Waals (vdw) scores:

$$E(\mathbf{F}) = \sum_i \text{dens}(f_i) + w_{vdw} \sum_{i,j} vdw(f_i, f_j) \quad (1)$$

For all experiments, we use $w_{vdw} = 1.0$. All pairwise scores are precomputed, and the Monte Carlo simulation is carried out for 1,500 steps; with kT initially at 200 and halving every 250 steps. Each 'move' in the simulation replaces a single segment with another in the ensemble. 100 trajectories of MCA are performed, with a composite model being created for each, and the best scoring is selected. If the sum of vdw energies of this solution is greater than 100, then additional rounds of sampling are carried out.

Tabu sampling. When building multiple loops, if no nonclashing solutions are found, we carry out additional sampling. However, this sampling is aware of the previous rounds' sampling and tries to explore additional regions of conformational space. For these subsequent rounds, this sampling is used in the ensemble filtering stage; when we select round robin from the 3-Å clusters, we limit only 10% of the cluster cap to clusters that were generated in previous rounds—all other structures are sampled from clusters that did not have a representative in previous rounds. Following this additional round of sampling, these new solutions are added to the previous pool, and assembly is attempted again.

Ensemble diversity. The convergence of the final ensemble provides a good indication of whether an accurate solution has

been sampled, with lower diversity suggesting that a correct solution has been found (Supplementary Fig. 2). To determine ensemble diversity for each new residue, the center of mass across the ensemble's C α positions is calculated; and then the r.m.s. deviation between the actual C α positions and their respective centers of mass are calculated for all models in the ensemble.

Ensemble size cap and run time. One key parameter of model building is the cap on ensemble size, which allows a tradeoff between run time and completeness of conformational sampling. For all experiments in the manuscript, a cap of 32 was considered. In addition, in one case where a cap of 32 failed to find a solution (a 65-residue segment in FrhG), we showed the performance as a function of increasing the cap size incrementally from 32 up to 320. The results are shown in Supplementary Figure 4. In this case, which RosettaES failed to sample accurately with a cap of 32 (r.m.s. deviation ≈ 3 Å), an increased ensemble size cap helps significantly: when the cap was set to 288 or greater, a model with r.m.s. deviation of 0.9 Å was sampled.

The typical run time for a missing segment of ~ 30 residues with the ensemble size capped at 32, and parallelized to 16 cores, is roughly 1.5 h. Supplementary Figure 5 shows that run time scales linearly with increasing beam size. Run times are reported for a single round of model building.

Map sharpening. To test how our method performed on maps of varying sharpness, we took the unsharpened map for the hCoV-NL63 and ran RosettaES on increasing levels of B-factor sharpening using Rosetta's ScaleMapIntensities mover. The high-resolution cutoff was set to 3.4, and the low resolution was set arbitrarily high; fade_width was set to 0.2. Backbone and C β r.m.s. deviation was calculated in comparison to the deposited model. As shown in Supplementary Figure 6, while RosettaES performance was dampened at extreme values, performance was largely consistent over a wide range of sharpening values.

Data sets used. We used a previously generated benchmark set of nine proteins, for which 3–5-Å cryo-EM data and a deposited model were available¹⁴. The round 1 models of the Rosetta *de novo* density application were used as inputs for all benchmarking except for comparison to alternate methods where the target segment was extracted from the deposited model in order to create a fair comparison. For the bamboo mosaic virus, the initial model was a homology model built from the Papaya Mosaic Virus. For the Mouse Hepatitis Virus (MHV), the starting model was a fragment docked by the Rosetta *de novo* density application. For hCoV-NL63, the initial model was a homology model built from MHV.

Statistics. The high-accuracy version of the global distance test¹⁹ was used to compare models. This test assigned a score to each atom pair based on their distance with the following criteria:

- distance < 0.5 Å; 4
- distance < 1.0 Å; 3
- distance < 2.0 Å; 2
- distance < 4.0 Å; 1
- distance > 4.0 Å; 0

The final score is equal to the sum of the atom pair scores divided by $4N$, where N is the number of atom pairs.

Figures were generated using UCSF Chimera³¹.

Code availability. This protocol and source code are freely available for academic use in weekly releases, after week 17 2017, of the Rosetta software suite found at <https://www.rosettacommons.org/>. Instructions for using RosettaES can be found in the **Supplementary Protocol**.

Data availability statement. The accession codes used are as follows: mouse hepatitis virus at 4 Å (EMDB 6526; PDB 3jcl), human coronavirus NL63 at 3.4 Å (EMDB 8331; PDB 5szs), and bamboo mosaic virus at 5.6 Å (EMDB 3020; PDB 5a2t).

29. Gront, D., Kulp, D.W., Vernon, R.M., Strauss, C.E.M. & Baker, D. *PLoS One* **6**, e23294 (2011).
30. DiMaio, F., Tyka, M.D., Baker, M.L., Chiu, W. & Baker, D. *J. Mol. Biol.* **392**, 181–190 (2009).
31. Goddard, T.D., Huang, C.C. & Ferrin, T.E. *J. Struct. Biol.* **157**, 281–287 (2007).