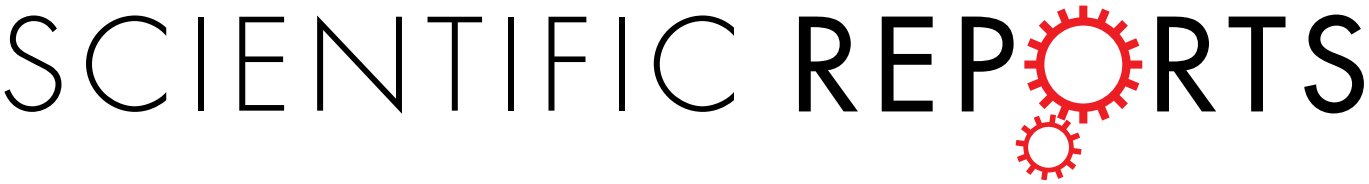

OPEN

EMBuilder: A Template Matching-based Automatic Model-building Program for High-resolution Cryo-Electron Microscopy Maps

Received: 4 January 2017
Accepted: 18 April 2017
Published online: 01 June 2017

Niyun Zhou[1,2], Hongwei Wang[1,2] & Jiawei Wang[3]

The resolution of electron-potential maps in single-particle cryo-electron microscopy (cryoEM) is approaching atomic or near- atomic resolution. However, no program currently exists for *de novo* cryoEM model building at resolutions exceeding beyond 3.5 Å. Here, we present a program, EMBuilder, based on template matching, to generate cryoEM models at high resolution. The program identifies features in both secondary-structure and Cα stages. In the secondary structure stage, helices and strands are identified with pre-computed templates, and the voxel size of the entire map is then refined to account for microscopic magnification errors. The identified secondary structures are then extended from both ends in the Cα stage via a log-likelihood (LLK) target function, and if possible, the side chains are also assigned. This program can build models of large proteins (~1 MDa) in a reasonable amount of time (~1 day) and thus has the potential to greatly decrease the manual workload required for model building of high-resolution cryoEM maps.

In recent years, substantial progress has been made in single-particle analysis (SPA) and has led to a resolution revolution in cryo-electron microscopy (cryoEM)[1]. The resolution of cryoEM structures is improving because of advancements in both hardware and software. An increasing number of structures of small proteins with low symmetry have achieved resolutions of ~3.5 Å[2] and even ~2 Å[3] or better. Interpreting these cryoEM maps as atomic models manually would be a difficult and tedious task. Therefore, fully automatic model-building programs for cryoEM maps are urgently needed.

Previous cryoEM model-building programs have mainly focused on medium-resolution (3.5–6 Å)[4] maps. The information content in a medium-resolution cryoEM map is insufficient to precisely interpret the main chain of a protein. Accordingly, most of the programs have added various types of external information (*i.e.*, structure-prediction-related information) to complement model building. A *de novo* model-building method[5] based on ROSETTA[6] combines the sequence-derived prediction of backbone conformations and side chain information with a cryoEM potential map to build a backbone and assign sequences. Gorgon[7] is an interactive modeling toolkit that performs *de novo* model building for cryoEM maps at resolutions of 3.5 to 10 Å, which combines sequence-based secondary structure prediction with feature detection and geometric modeling techniques to trace the backbones in cryoEM maps. EM-fold[8] is a method that combines secondary structure prediction with a Monte Carlo assembly algorithm and Rosetta refinement procedures to build topology models for medium-resolution cryoEM maps. Pathwalking[9] is based on the Traveling Salesman Problem (TSP) and traces the backbones in cryoEM maps. This method places the pseudo-atoms and traces the backbone by solving TSP on the basis of the placed pseudo-atoms. The secondary structure elements (SSEs) in the backbone are thus identified and corrected.

However, because of the resolution revolution of cryoEM maps (*i.e.*, beyond 3.5 Å), the information content in the map itself is now adequate enough to trace the main chain and part of the side chain. Therefore, the map can be directly interpreted by matching it with pre-built models or templates. In X-ray crystallography, numerous automatic model-building programs have been developed by using the template-matching method[10].

[1]MOE Key Laboratory of Protein Science, Tsinghua University, Beijing, 100084, China. [2]School of Life Sciences, Tsinghua University, Beijing, 100084, China. [3]State Key Laboratory of Membrane Biology, Tsinghua University, Beijing, 100084, China. Correspondence and requests for materials should be addressed to H.W. (email: hongweiwang@tsinghua.edu.cn) or J.W. (email: jwwang@tsinghua.edu.cn)

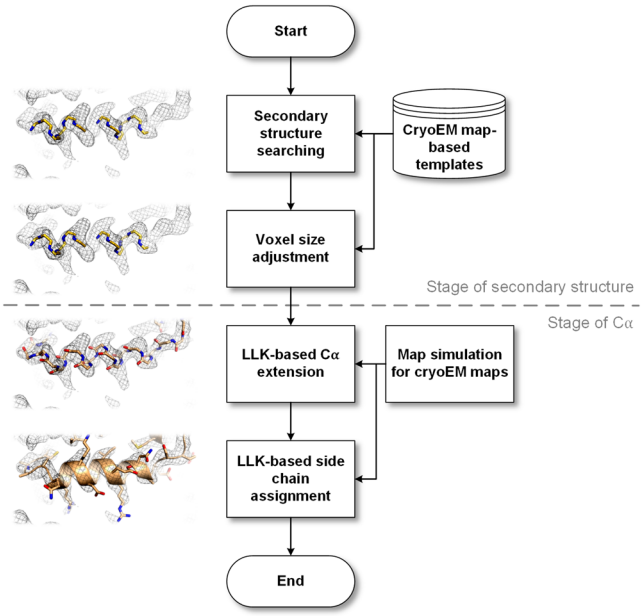

Figure 1. Workflow of EMBuilder. First, the program works on the secondary structure stage. Secondary structure templates are searched and placed in the map. Then, voxel size refinement is performed on these templates. Subsequently, the program works on the Cα stage. The reference map is adjusted to the same scale as the working map. The Cα target and side chain target functions are created on the basis of the scaled reference map, and then, Cα extension and side chain assignment are performed.

Buccaneer[11], the successor of FFFear[12], uses an oriented electron-density likelihood target function to identify likely Cα positions and then expands these Cαs to the extended main chain fragments. Side chain assignment is performed by applying the electron-density likelihood target function after the main chain extension. RESOLVE[13] identifies secondary structures and extends them with pre-built segment libraries constructed from the Protein Data Bank (PDB) database. C-Alpha Pattern Recognition Algorithm (CAPRA)[14] uses pattern-recognition techniques and a neural network to predict the candidate atoms closest to true Cα atoms. ARP/wARP[15] is a software suite for model building, refinement and validation. It uses density recognition-driven procedures to place and remove atoms, thereby limiting the resolution of the data to 2.5 Å or higher. Automatic Crystallographic Map Interpreter (ACMI)[16] uses probabilistic inference to predict the backbone layout and a statistical sampling method to produce an accurate and physically feasible set of structures and side chain templates to sample side chains. All of these programs were developed to address the problem of model building specific in crystallography. For example, the templates used in RESOLVE are derived from X-ray crystal structures and hence are suboptimal templates for cryoEM maps. Meanwhile the refinement procedure improves the phases of crystallographic maps iteratively, but no phase problem exists in cryoEM maps. Therefore, these distinct characteristics of cryoEM maps should stimulate the development of novel model-building programs for high-resolution cryoEM maps.

The most important problem that must be overcome in model building of high-resolution cryoEM maps is that the voxel size may be imprecise because of microscopic magnification errors that occur during data collection[17]. Hence, the voxel size should be corrected before model building to prevent errors in the final model. Furthermore, the noise level and scattering factors of cryoEM maps differ from that of crystallographic maps. Thus, the templates derived from crystallographic maps are not suitable for cryoEM maps, and it is crucial to re-design the algorithms and re-tune the parameters for cryoEM map applications. To this end, we present a dedicated program—EMBuilder—based on the template-matching method for model building specifically designed for high-resolution cryoEM maps.

Methods

Program Workflow. EMBuilder is based on the template-matching method and processes cryoEM maps through two stages: secondary-structure stage and Cα stage (Fig. 1). In the first stage, the positions of secondary structures are searched in the working map (input map) on the basis of the templates, including α-helix and β-stand, generated from previously known cryoEM structures. These potential secondary structure positions and voxel size are then refined using the correlation coefficient (CC) as the refinement target function. Once the voxel size is corrected, the program proceeds to the Cα stage. Another cryoEM map which is used as a reference is adjusted to the same scale as the working map by using a map simulation procedure. The scaled reference map, along with a known atomic model, is used to build Cα and side chain log-likelihood (LLK) target function. Then, candidate Cαs are extended from both ends of the secondary structures to complete the main chain model based on the Cα LLK target function learned from the scaled reference map. As an option, a side chain is assigned if a sequence of the working map is available. These LLK related procedures are similar to that in Buccaneer[11].

Clipper[18] and CCP4 coordinate libraries[19] are used to handle ccp4/mrc-format maps and PDB files. The individual functionalities are discussed below in details.

Templates from CryoEM Map. We generated a helix template that was 6 residues long and a strand template that was 4 residues long. Each template consisted of a PDB coordinate set, a mean map and a correlation map. These two templates were used for the subsequent secondary structure identification and voxel size refinement.

The calculation of the mean map and the correlation map of the template was similar to that used in RESOLVE[20]. A helix (133–138) in myoglobin (PDB entry 1a6m)[21,22] and a strand (105–108) in carboxypeptidase A (PDB entry 1bav)[21,23] were chosen as standard helix and strand respectively. The cryoEM map of glutamate dehydrogenase (GDH) (PDB entry 5a1a, EMDB entry 2984)[21,24,25] was low-pass filtered to 2.5 Å, 3.0 Å and 3.5 Å. All helices/strands of GDH were rotated and translated onto the above standard helix/strand, and only helices/strands with root-mean-square deviations (RMSDs) less than 0.5 were selected for further analysis. The cryoEM grid points of each selected helix/strand within 15 Å of the center of mass (COM) of the standard helix/strand were used for third-order interpolation on a standard grid (expressed as S_{nk}, where n represents the selected helices/strands, and k represents the interpolated grid points in each helix/strand). The mean map S^{mean} was calculated by separately averaging grid points from all of the selected helices/strands of GDH:

$$S_k^{mean} = \frac{1}{n}\sum_n S_{nk} \tag{1}$$

The correlation map was calculated by averaging the CCs between the above-mentioned mean map and individual cryoEM maps of helices/strands. For each S_{nk}, the distance between S_{nk} and the nearest atom in the standard helix/strand models was calculated (expressed as L_{nk}). All $L_{nk} < 15$ Å were divided into 30 groups with 0.5 Å intervals, expressed as G_i (where i represents the group number, $i \in \{1, 2, \ldots, 29, 30\}$). For each group G_i, the CCs between all S_{nk} within the same G_i and S_k^{mean} were calculated and averaged (expressed as CC_i):

$$CC_i = \frac{1}{n}\sum_n \rho_n(S_{nk}, S_k^{mean}),\ k \in G_i,\ i \in \{1, 2, \ldots, 29, 30\} \tag{2}$$

The value of each grid point of correlation map S_k^{corr} was assigned as the CC_i where the group i was the grid point belongs to.

Secondary Structure Searching and Voxel Size Refinement. The refinement of the voxel size is typically performed by comparing the cryoEM map with a corresponding atomic model and adjusting the voxel size until the CC between the map and model is maximized[26]. However, no atomic model currently exists for such analysis before model building. In our method, we performed a 7-dimensional refinement (3D rotation, 3D transition and voxel size) on helix/strand templates to correct the voxel size of the maps.

The initial position of the helix/strand was roughly determined through a fast searching method, as used in Coot[27], to save the computation time compared with the CC-based searching method. It only uses several values of grid points of the map to determine the potential positions of helix/strand, *e.g.* the grid values at the Cα positions of the template. Therefore, it is very fast but quite imprecise. Once the rough positions are determined, the helix/strand template maps were then placed on these positions. The CC between the working map and the mean map of the placed helix/strand template was calculated and weighted by the corresponding correlation map. The top 20 helices/strands with CCs greater than 0.3 were selected (expressed as S_n, where n represents the number of helices/strands, $n \in \{1, 2, \ldots, 19, 20\}$). The simplex method with the CC target was used to perform six-dimensional positional refinement on these selected helices/strands. Then, the refined helices/strands were used for voxel size determination. For each S_n, the original voxel size of the working map (expressed as v_{ori}) was multiplied by a factor (expressed as m) from 0.9 to 1.1 with 0.001 intervals ($m \in \{0.9, 0.901, \ldots, 1.099, 1.1\}$). Notably, when the voxel size was altered, the entire map expanded or shrank along the x, y and z axes. The coordinates of every grid point (except the origin) on the map then mismatched relative to the already placed templates S_n. Therefore, the distance of this additional offset was complemented to adjust the position of S_n after multiplication. Subsequently, simplex positional refinement was performed on the S_n and then followed by CC calculation. For each particularly varying voxel size value $v_{ori} * m$, the CC between the refined template S_n and the working cryoEM map was calculated and expressed as CC_{mn}. All the CC_{mn} values were accumulated across the top 20 helices/strands as: $\sum_n CC_{mn}$. The maximum value of accumulated CC_{mn} values determined the corresponding voxel size adjustment scale, as:

$$v_{new} = v_{ori} * \arg\max_m \left\{\sum_n CC_{mn}\right\} \tag{3}$$

For each cryoEM map, helix and strand adjustment ratios were balanced to calculate an overall novel voxel size by weighted-averaging the voxel sizes from these two sources. The weight was dependent on the number of "good" helices and strands ($CC > 0.3$) found in the map.

Map Simulation. At the Cα stage of model building, the scaling factor between the reference and working maps was adjusted according to their power spectra. The Guinier plots of the working and reference structures were calculated separately. The natural logarithm of the structure factor (expressed as lnF) of the reference structure was interpolated to the lnF of the working structure at the resolution beyond 10 Å. At the low-resolution

region (<10 Å) the reference map was interpolated to the nearest 10 Å resolution to prevent overweighting. Once the scale was applied to the reference map, it was also low-pass filtered to the same resolution as that of the working map to generate the simulated reference map. The resulting simulated map was then used to accumulate the log-likelihood target function for the subsequent Cα identification and possible sequence assignment if the sequence was available.

Results

Model Building. The entire model-building procedure in EMBuilder was performed on benchmark test data sets containing 6 cryoEM maps from the EMDataBank[25] (EMDB) (Fig. 2), *i.e.* EMD8117[28] (M.W. 300 kDa, 2.95 Å), EMD6630[29] (M.W. 336 kDa, 3.26 Å), EMD3297[30] (M.W. 540 kDa, 3.3 Å), EMD3061[2] (M.W. 170 kDa, 3.4 Å), EMD3388[31] (M.W. 1.2 MDa, 3.4 Å) and EMD6534[32] (M.W. 700 kDa, 3.7 Å). The parameters used in the model building were as the followings: reference map: EMDB8194; resolution: reported in EMDB; number of residues to be built: estimated on the basis of molecular mass; and helix/strand templates: generated from EMD2984 after low-pass filtering to 2.5 Å.

We calculated the completeness, Cα RMSD (Fig. 3) and accuracy of side chain assignment (Supplementary Table S1) to measure the quality of the model-building results. Cα RMSD values were calculated by measuring the distance between each Cα in the model and the nearest Cα in the final structure deposited in the PDB. The RMSDs of the auto-built model ranged from 0.9 Å to 2.8 Å for our test data sets. On average, ~51% of the Cαs were built within 1 Å of the structure deposited in the PDB, and ~73% of the Cαs were built within 2 Å. Although the main chain, especially Cα positions could be identified to the reasonable quality in the map, the difficulty of assigning side chains for a cryoEM map is varying across the whole map because of the uneven resolution distribution of the map. If the side-chain density is good enough for visualization in a local area of a cryoEM map, the assignment will be successful by using our simulation and side-chain assignment method (Supplementary Table S1). The running times required for model building for EMD8117, EMD6630, EMD3297, EMD3061, EMD3388 and EMD6534 were 1.5 h, 3.3 h, 5.5 h, 1 h, 23.5 h and 5.5 h, respectively, on a 2.6-GHz central processing unit (CPU).

In the central core regions of these six maps, the features of each residue are unambiguous, such that the LLK target function can identify the spatial orientations of the main and side chains. Therefore EMBuilder produced low-RMSD models (*e.g.* the transmembrane domains of TRPV1 and γ-secretase). Nonetheless in the peripheral region, which usually only has medium resolution because of radiation damage or particle misalignment or domain flexibility, the recognition of residues by the target function is difficult. Some error may occur even in main chain tracing, especially in loop building. The RMSD of the model in this region is ~0.5 Å greater than that at core region. Therefore, EMBuilder is suitable for generating high-resolution cryoEM models. Moreover, EMBuilder can build models for large cryoEM maps. The PDB model of EMD3388 comprises ~7300 residues. In our model, 3282 Cαs were built within 1 Å of the Cαs in the PDB model, and 6009 Cαs were built within 2 Å. These results indicate that substantial model-building time can be saved using EMBuilder compared with manual building.

Voxel Size Refinement. The voxel sizes of cryoEM maps may be inaccurate, and the error can be as great as 5%[17]. Therefore, the voxel size must be assessed or corrected before model building. The voxel size refinement subroutine in EMBuilder was designed to perform this task. We manually introduced voxel size error to evaluate the accuracy of our voxel size refinement subroutine. We assumed that all the maps deposited in the EMDB had the correct voxel sizes. The error was introduced before refinement by multiplying the voxel size by a factor ranging from 0.95 to 1.05 in the intervals of 0.01, which corresponded to errors of ±5%. We used 4 templates to evaluate the accuracy: one from RESOLVE[20], and 3 computed from a cryoEM map after low-pass filtering to 2.5 Å, 3.0 Å and 3.5 Å (described in Methods). The refined voxel size was compared with the correct voxel size to validate the accuracy (Fig. 4). The detailed data of voxel size refinement can be found as Supplementary Figure S1.

Our voxel size refinement subroutine reduced the voxel size error to 0.59% on average. The best refinement result reduced the voxel size error to 0.37%, a value acceptable for model building. Our template with low-pass filtering to 2.5 Å yielded the best accuracy for voxel size refinement among the templates tested (data not shown). Therefore, it was used as the default template in EMBuilder. Moreover, the template from RESOLVE clearly had relatively high error (~4%) in voxel size refinement compared with those of our templates (Fig. 4), thus suggesting that the features of cryoEM maps differ from those of the maps used in crystallography. Therefore, the parameters and algorithms used for crystallography may not generate optimal results when they are directly applied to cryoEM maps.

Map Simulation. During the Cα stage of model building, the generation of the LLK target function requires a reference map and a working map on the same scale. The purpose of the map simulation subroutine is to adjust the scale of the reference map to that of the working map. The map simulation and Cα finding were performed with EMBuilder and Buccaneer[11] on the test data sets to specifically evaluate the effectiveness of our map simulation subroutine. Buccaneer, which was designed specifically for model building in crystallography, may also be used for cryoEM map model building, but its power is suboptimal. The working map was EMD3061, and the reference map was EMD8194. We calculated the Guinier plot (lnF *vs.* d^{-2}) of the reference and the working structures (Fig. 5A,D) and Cα distance between the structures deposited in the PDB and the Cα finding (Fig. 5B,E).

The map simulation method used by EMBuilder was effective for cryoEM maps and offered better results than Buccaneer. The percentage of Cαs placed by EMBuilder within 2 Å of those in the structures deposited in the PDB was 70.8%. The Cα finding in Buccaneer was influenced by the density of the detergent in the map because of inappropriate map simulation (Fig. 5C). In contrast, most of the Cαs identified by EMBuilder were located on the

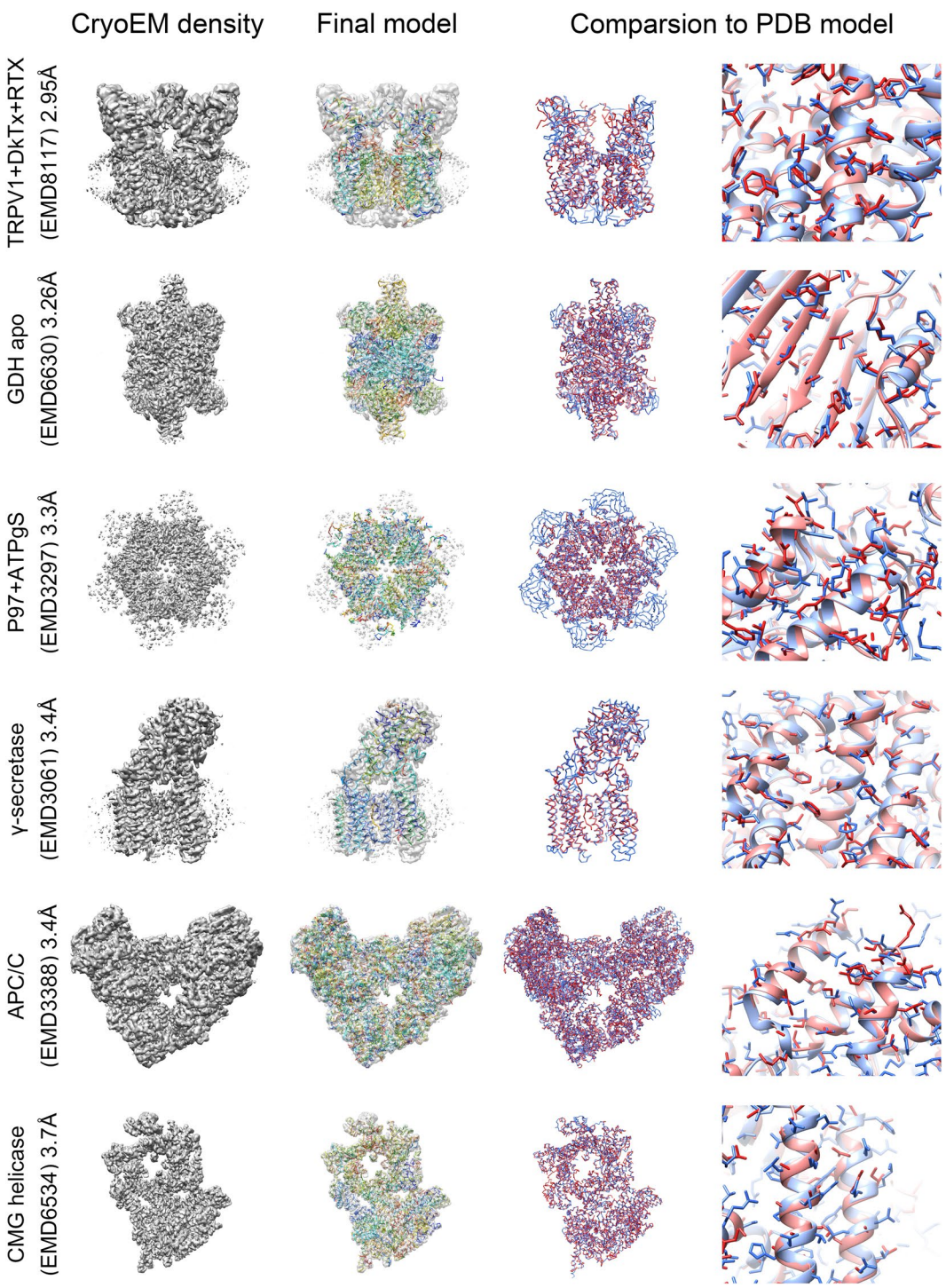

Figure 2. Model-building results from EMBuilder and comparison with PDB models. The cryoEM density map is presented in gray surface in the first column with a threshold from EMDB. Models built with EMBuilder are shown in the second column as cartoon representation. The overall comparison of the models built with EMBuilder (red ribbon) and the PDB models (blue ribbon) is presented in the third column. The side chain assignments of the models built with EMBuilder (red stick) and the PDB model (blue stick) are given in the fourth column.

backbone of the protein (Fig. 5F), thus suggesting that the algorithm used for crystallography may not perform optimally when applied to cryoEM maps.

Discussion

We developed EMBuilder, which uses a template-matching method for model building of cryoEM maps. This program is capable of correcting the voxel size error and building atomic models for high-resolution (<3.5 Å)

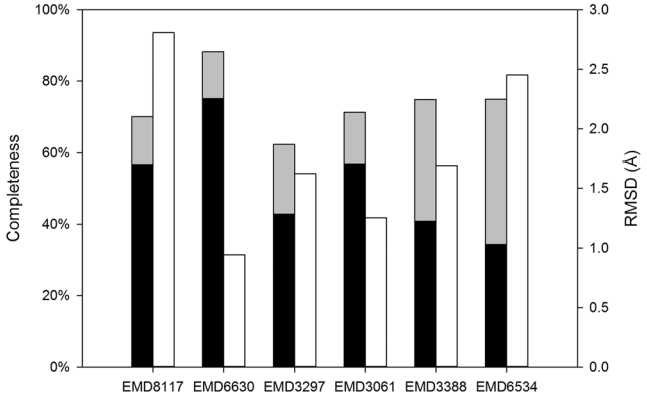

Figure 3. Completeness and RMSDs of the models built by EMBuilder from the test data sets. The shaded bars and white bars represent the completeness and RMSD, respectively. In the shaded bars, black and gray represent the percentages of Cα built between 1 Å and 2 Å and within 1 Å, respectively.

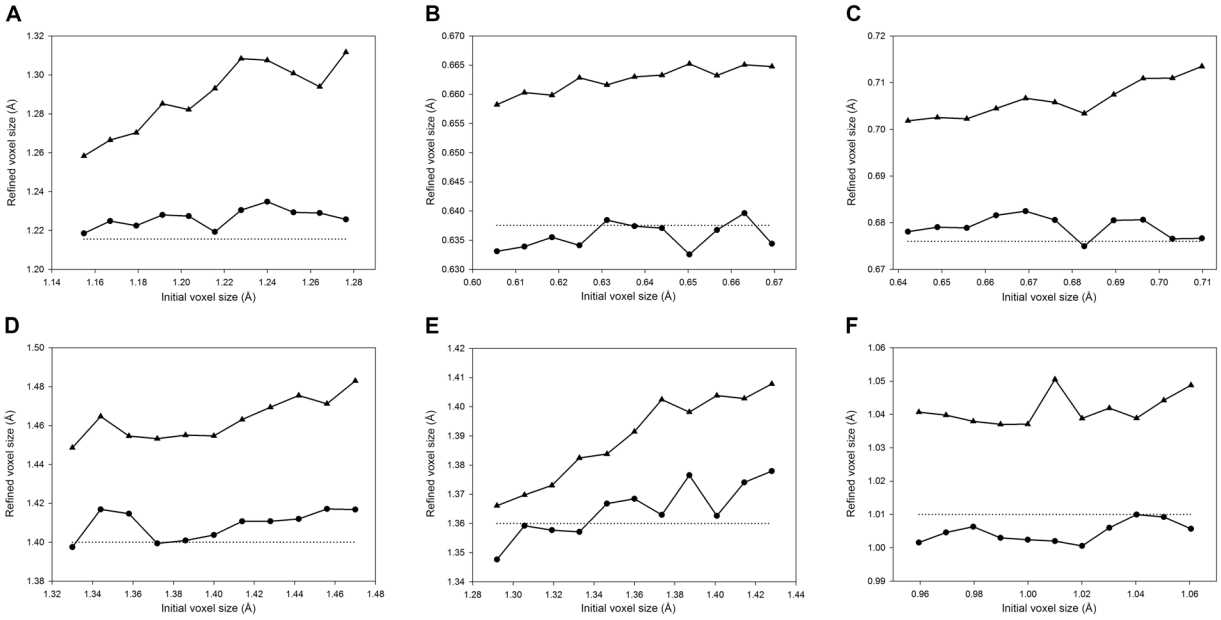

Figure 4. Results of voxel size refinement. The test data sets are: EMD8117 (**A**), EMD6630 (**B**), EMD3297 (**C**), EMD3061 (**D**), EMD3388 (**E**) and EMD6534 (**F**). RESOLVE template and our template (2.5 Å) were used in the refinement and are depicted as triangles and circles, respectively. The dotted line represents the correct voxel size of the map (from EMDB).

cryoEM maps. EMBuilder uses two stages to build an atomic model. The positions of helices and strands are identified and refined in seven dimensions in the secondary structure stage. In the subsequent Cα stage the reference cryoEM map is adjusted to the same scale as that of the working map to generate the LLK target functions of the main chain and side chain. Then, the Cαs are extended into main chain fragments according to the density map. The side chains are assigned on the basis of the assembled main chain fragments. The time required for model building is trivial enough compared with manual building. For a 300-kDa protein, 2 to 3 h was required for EMBuilder to build a model with ~70% completeness, a time considerably less than that required in manual building. Additionally, the time required to build a very large cryoEM density map (~1 MDa) was acceptable (~24 h) with a single thread.

A previous study has reported that the voxel size error in cryoEM maps can be as great as ~5%[17]. Under normal circumstances, the voxel size error of a cryoEM map may be ~2%. Additionally, the voxel size error can accumulate across several residues during residue extension, thus severely affecting the fitting between the model and map. Therefore, voxel size correction is an essential procedure before model building. However, it is difficult

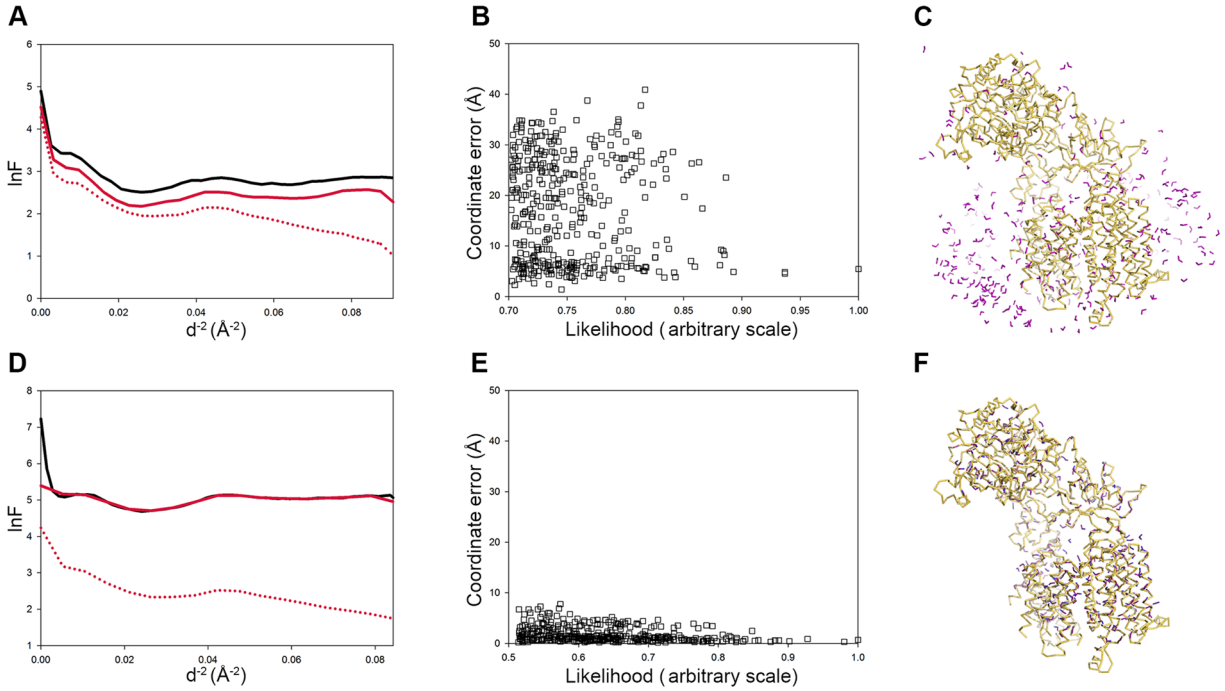

Figure 5. Accuracy evaluation of the map simulations of Buccaneer and EMBuilder. The *lnF vs* d^{-2} plots of the working map, original reference map and simulated reference map are shown as a solid black line, dotted red line and solid red line, respectively (panels A and D). The Cα distances between the structures deposited in PDB and the Cα finding are presented in panels B and E, respectively. The structures deposited in PDB and the Cα finding are shown as yellow ribbons and purple sticks, respectively (panels C and F).

to correct the voxel size of a map manually when the corresponding model is unavailable. Our solution involves 1) identifying secondary structure elements (SSEs) in the map, 2) correcting the voxel size of the SSEs, and 3) using the results of the corrected SSEs' voxel sizes to determine the overall voxel size. Thus, the method corrects the SSE voxel size by calculating the CC with the pre-computed template and adjusts it until the CC is maximized. After obtaining the voxel sizes of the SSEs, the 20 helices and strands with the highest CCs are used to determine the overall voxel size. By using our test data sets, we demonstrated that this process is robust. The entire refinement procedure required only ~2 minutes in most cases. Thus, our voxel size refinement subroutine provides an easy method of correcting the voxel size of a cryoEM map.

We also found that the choice of templates significantly influences the accuracy of voxel size refinement. The template derived from the crystallography density map yielded ~4% error in voxel size refinement. By using our templates, we found that the voxel size refinement subroutine reduced the voxel size error to 0.59% on average; this value is acceptable for model building. One possible reason for this difference is that the density distribution and pattern of noise of crystallography maps and cryoEM maps are not similar.

In a crystallographic model-building program, such as Buccaneer, the map simulation method is used to adjust the scaling factors of two maps according to the Wilson statistics[11]. However, we found that the Buccaneer simulation method produced minimal effect in cryoEM maps, thus resulting in the misidentification of Cαs, possibly because of the higher noise level in the cryoEM maps. We have tested several methods to equivalently scale the cryoEM maps and found that linear interpolation of *lnF vs* d^{-2} between the reference and working maps was rapid and produced acceptable accuracy.

Resolution typically varies widely across cryoEM maps. Therefore, a different model-building strategy should be used if a map contains both high-resolution and medium-resolution regions. For medium-resolution and ambiguous regions, more external information, such as secondary structure prediction, can be added to aid in determining the molecular topology. In addition, the map can be segmented by resolution. In these sub-maps, different algorithms and templates can be used to build local models. In regions of lower resolution, automatic docking between the pre-computed homology model and the cryoEM map can be performed. Thus, the resolution range of input map for EMBuilder can be extended to a lower level. Moreover, a mask of the asymmetric unit might be created for a homo-multimer structure. Then, the speed of model building can be accelerated by building the model into only one asymmetric unit.

In conclusion, we present a model-building program—EMBuilder—for high-resolution cryoEM maps. This program is based on a template-matching method that uses pre-computed templates to correct voxel sizes and build atomic models. EMBuilder is an effective program that can help researchers build cryoEM structure models rapidly and easily.

References

1. Kuhlbrandt, W. The Resolution Revolution. *Science* **343**, 1443–1444, doi:10.1126/science.1251652 (2014).
2. Bai, X. C. *et al*. An atomic structure of human gamma-secretase. *Nature* **525**, 212–217, doi:10.1038/nature14892 (2015).
3. Merk, A. *et al*. Breaking Cryo-EM Resolution Barriers to Facilitate Drug Discovery. *Cell* **165**, 1698–1707, doi:10.1016/j.cell.2016.05.040 (2016).
4. DiMaio, F. & Chiu, W. Tools for Model Building and Optimization into Near-Atomic Resolution Electron Cryo-Microscopy Density Maps. *Methods Enzymol* **579**, 255–276, doi:10.1016/bs.mie.2016.06.003 (2016).
5. Wang, R. Y. *et al*. De novo protein structure determination from near-atomic-resolution cryo-EM maps. *Nat Methods* **12**, 335–338, doi:10.1038/nmeth.3287 (2015).
6. Gront, D., Kulp, D. W., Vernon, R. M., Strauss, C. E. & Baker, D. Generalized fragment picking in Rosetta: design, protocols and applications. *PLoS One* **6**, e23294, doi:10.1371/journal.pone.0023294 (2011).
7. Baker, M. L. *et al*. Modeling protein structure at near atomic resolutions with Gorgon. *Journal of structural biology* **174**, 360–373, doi:10.1016/j.jsb.2011.01.015 (2011).
8. Lindert, S. *et al*. EM-fold: de novo atomic-detail protein structure determination from medium-resolution density maps. *Structure* **20**, 464–478, doi:10.1016/j.str.2012.01.023 (2012).
9. Chen, M., Baldwin, P. R., Ludtke, S. J. & Baker, M. L. De Novo modeling in cryo-EM density maps with Pathwalking. *Journal of structural biology* **196**, 289–298, doi:10.1016/j.jsb.2016.06.004 (2016).
10. Kleywegt, G. J. & Jones, T. A. Detecting folding motifs and similarities in protein structures. *Methods Enzymol* **277**, 525–545 (1997).
11. Cowtan, K. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta crystallographica. Section D, Biological crystallography* **62**, 1002–1011, doi:10.1107/S0907444906022116 (2006).
12. Cowtan, K. Fast Fourier feature recognition. *Acta crystallographica. Section D, Biological crystallography* **57**, 1435–1444 (2001).
13. Terwilliger, T. C. *et al*. Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta crystallographica. Section D, Biological crystallography* **64**, 61–69, doi:10.1107/S090744490705024X (2008).
14. Ioerger, T. R. & Sacchettini, J. C. Automatic modeling of protein backbones in electron-density maps via prediction of C(alpha) coordinates. *Acta Crystallogr D* **58**, 2043–2054, doi:10.1107/S0907444902016724 (2002).
15. Cohen, S. X. *et al*. Towards complete validated models in the next generation of ARP/wARP. *Acta crystallographica. Section D, Biological crystallography* **60**, 2222–2229, doi:10.1107/S0907444904027556 (2004).
16. DiMaio, F. *et al*. Creating protein models from electron-density maps using particle-filtering methods. *Bioinformatics* **23**, 2851–2858, doi:10.1093/bioinformatics/btm480 (2007).
17. Rossmann, M. G., Bernal, R. & Pletnev, S. V. Combining electron microscopic with x-ray crystallographic structures. *Journal of structural biology* **136**, 190–200, doi:10.1006/jsbi.2002.4435 (2001).
18. Cowtan, K. Fitting molecular fragments into electron density. *Acta crystallographica. Section D, Biological crystallography* **64**, 83–89, doi:10.1107/S0907444907033938 (2008).
19. Krissinel, E. B. *et al*. The new CCP4 Coordinate Library as a toolkit for the design of coordinate-related applications in protein crystallography. *Acta crystallographica. Section D, Biological crystallography* **60**, 2250–2255, doi:10.1107/S0907444904027167 (2004).
20. Terwilliger, T. C. Maximum-likelihood density modification using pattern recognition of structural motifs. *Acta crystallographica. Section D, Biological crystallography* **57**, 1755–1762 (2001).
21. Berman, H. M. *et al*. The Protein Data Bank. *Nucleic Acids Res* **28**, 235–242 (2000).
22. Vojtechovsky, J., Chu, K., Berendzen, J., Sweet, R. M. & Schlichting, I. Crystal structures of myoglobin-ligand complexes at near-atomic resolution. *Biophysical journal* **77**, 2153–2174, doi:10.1016/S0006-3495(99)77056-6 (1999).
23. Massova, I. *et al*. Crystallographic and computational insight on the mechanism of zinc-ion-dependent inactivation of carboxypeptidase a by 2-benzyl-3-iodopropanoate. *Journal of the American Chemical Society* **118**, 12479–12480, doi:10.1021/ja963234k (1996).
24. Bartesaghi, A. *et al*. 2.2 A resolution cryo-EM structure of beta-galactosidase in complex with a cell-permeant inhibitor. *Science* **348**, 1147–1151, doi:10.1126/science.aab1576 (2015).
25. Lawson, C. L. *et al*. EMDataBank unified data resource for 3DEM. *Nucleic Acids Res* **44**, D396–403, doi:10.1093/nar/gkv1126 (2016).
26. Rossmann, M. G. Fitting atomic models into electron-microscopy maps. *Acta crystallographica. Section D, Biological crystallography* **56**, 1341–1349 (2000).
27. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta crystallographica. Section D, Biological crystallography* **66**, 486–501, doi:10.1107/S0907444910007493 (2010).
28. Gao, Y., Cao, E., Julius, D. & Cheng, Y. TRPV1 structures in nanodiscs reveal mechanisms of ligand and lipid action. *Nature* **534**, 347–351, doi:10.1038/nature17964 (2016).
29. Borgnia, M. J. *et al*. Using Cryo-EM to Map Small Ligands on Dynamic Metabolic Enzymes: Studies with Glutamate Dehydrogenase. *Molecular pharmacology* **89**, 645–651, doi:10.1124/mol.116.103382 (2016).
30. Banerjee, S. *et al*. 2.3 A resolution cryo-EM structure of human p97 and mechanism of allosteric inhibition. *Science* **351**, 871–875, doi:10.1126/science.aad7974 (2016).
31. Zhang, S. *et al*. Molecular mechanism of APC/C activation by mitotic phosphorylation. *Nature* **533**, 260–264, doi:10.1038/nature17973 (2016).
32. Yuan, Z. *et al*. Structure of the eukaryotic replicative CMG helicase suggests a pumpjack motion for translocation. *Nature structural & molecular biology* **23**, 217–224, doi:10.1038/nsmb.3170 (2016).

Acknowledgements

We thank Prof. L. Cheng, Dr. X. Wang, Dr. C. Yan, Dr. Q. Zhou, and F. Yang for testing the program and providing helpful suggestions. We thank Prof. Z. Wang, Prof. J. Wu and X. Chen for their helpful discussions about the manuscript. This work was supported by the Chinese Ministry of Science and Technology (Nos 2015CB910104 and 2016YFA0501103) to J.W.

Author Contributions

N.Z. performed the research and drafted the manuscript. H.W. and J.W. supervised the research and edited the draft. All authors reviewed the manuscript.

Additional Information

Supplementary information accompanies this paper at doi:10.1038/s41598-017-02725-w

Competing Interests: The authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.