# Lawrence Berkeley National Laboratory

**Title**
Assembly of macromolecular complexes by satisfaction of spatial restraints from electron microscopy images.

**Permalink**
https://escholarship.org/uc/item/3802m6kh

**Journal**
Proceedings of the National Academy of Sciences of USA, 109(46)

**Authors**
Velázquez-Muriel, Javier
Lasker, Keren
Russel, Daniel
et al.

**Publication Date**
2012-11-13

**DOI**
10.1073/pnas.1216549109

Peer reviewed

# Assembly of macromolecular complexes by satisfaction of spatial restraints from electron microscopy images

Javier Velázquez-Muriel[a,b,c], Keren Lasker[a,b,c,d], Daniel Russel[a,b,c], Jeremy Phillips[a,b,c], Benjamin M. Webb[a,b,c], Dina Schneidman-Duhovny[a,b,c], and Andrej Sali[a,b,c,1]

Departments of [a]Bioengineering and Therapeutic Sciences and [b]Pharmaceutical Chemistry, and [c]California Institute for Quantitative Biosciences, University of California, San Francisco, CA 94158-2552; and [d]Department of Developmental Biology, Stanford University, Stanford, CA 94305

To obtain a structural model of a macromolecular assembly by single-particle EM, a large number of particle images need to be collected, aligned, clustered, averaged, and finally assembled via reconstruction into a 3D density map. This process is limited by the number and quality of the particle images, the accuracy of the initial model, and the compositional and conformational heterogeneity. Here, we describe a structure determination method that avoids the reconstruction procedure. The atomic structures of the individual complex components are assembled by optimizing a match against 2D EM class-average images, an excluded volume criterion, geometric complementarity, and optional restraints from proteomics and chemical cross-linking experiments. The optimization relies on a simulated annealing Monte Carlo search and a divide-and-conquer message-passing algorithm. Using simulated and experimentally determined EM class averages for 12 and 4 protein assemblies, respectively, we show that a few class averages can indeed result in accurate models for complexes of as many as five subunits. Thus, integrative structural biology can now benefit from the relative ease with which the EM class averages are determined.

integrative modeling | structural determination | computational biology

**E**M is an increasingly useful approach for structural characterization of macromolecular assemblies (1–3). Different flavors of EM include electron crystallography, single-particle EM, and electron tomography (4), although tomography is generally limited to resolutions worse than 20 Å. Single-particle EM can be used with negative-stained or cryogenically frozen (cryo-EM) samples. Cryo-EM particularly presents some attractive advantages: It preserves a near-native conformation of the molecules, it can be applied to the study of large assemblies, and it can theoretically achieve atomic resolution (5). Currently, the standard way to analyze single-particle EM data is to align collected 2D single-particle images, cluster the images, calculate a 2D class-average image for each cluster, and finally perform 3D reconstruction to obtain a 3D density map. Pseudoatomic models for many assemblies have been generated by fitting X-ray crystallographic structures and/or comparative models of the individual components into a density map of the whole assembly (6).

Obtaining a high-resolution density map requires a large number of single-particle images and depends critically on determining an initial low-resolution density map as a template for reconstruction, as well as a high signal-to-noise ratio (SNR) in the particle images (7–10). Several methods exist to obtain the template map: random-conical reconstruction (11), common lines determination (12–14), and maximization of the posterior probability of observing the set of class averages (15). A challenge for template construction arises when the available class averages do not provide significant coverage of all the orientations of the complex; in such a case, an accurate template map cannot be computed. This situation is common for assemblies that have only a few preferred orientations on the sample grid. Therefore, there is a need for a general procedure to construct accurate assembly models based on only a few class averages and potentially other data.

Here, we address the problem of assembling multiple subunits of known stoichiometry into a macromolecular complex, based, in part, on the class averages. The inputs are the experimentally determined or modeled atomic structures of the assembly components; one or more class averages; and potentially other restraints, such as contacts and distances between neighboring proteins, derived from, for example, proteomics and chemical cross-linking experiments. Our approach relies on encoding the class averages as spatial restraints and potentially combining them with other restraints into a scoring function. Each image is used to score a model by the degree of the match between the optimal model projection and the image. In principle, any number of class averages can be used, without the need for single-particle reconstruction. Models are then obtained by sampling the scoring function. The sampling protocol relies on a simulated annealing Monte Carlo (SA-MC) optimization in continuous space, followed by the combination of the SA-MC solutions with the divide-and-conquer message-passing algorithm DOMINO (16). A unique aspect of our approach is that a complex structure is assembled from scratch without relying on an initial structure, although class averages have already been used to refine it (17).

We first present our benchmark results on 12 complexes. The results demonstrate that accurate models for various types of macromolecular assemblies can be obtained by using only a limited number of class averages in combination with other restraints not containing direct 3D information. Next, as an example of the utility of the approach, we use it to compute models for the human transferrin receptor–transferrin (TfR–Tf) complex (18), the human C3bB protoconvertase complex (19), the bovine mitochondrial supercomplex $I_1III_2IV_1$ (20), and the type I restriction modification complex EcoR124I from *Escherichia coli* (21). Finally, the applicability of the method and possible improvements are discussed.

## Results

The *em2D* score based on an image is 1 minus the cross-correlation coefficient (*ccc*) between the image and the best-matching model projection (*Materials and Methods*). To find this optimal projection, a search over two translational and three rotational registration parameters is performed (Fig. 1*A*). When more than one image is available, the average *em2D* score (*<em2D>*) is computed from the individual *em2D* scores. Usually, the images are class averages obtained by classification and averaging of single-particle images from negative stains or cryo-EM; single-particle images from a tilt series and negative-stained samples can also be used in principle. In our benchmark, we assume that the sample used to produce the class averages is compositionally and conformationally homogeneous. Apart from the *<em2D>* score, our complete scoring function includes proteomics restraints describing proximity between components of the assembly, cross-linking restraints imposing a maximum distance

**Fig. 1.** Flow chart of the scoring and sampling algorithms. (*A*) To calculate the *em2D* score, we project the model in evenly spaced directions on the hemisphere with positive *y*-axis values. The resulting projections and the input images are preprocessed and subsequently aligned in two dimensions to obtain an initial coarse alignment. The best alignments are refined using the Simplex algorithm to minimize the squared difference between the pixels of the image and the projection, providing the *em2D* score of the image. The total score *<em2D>* for a model is the average of the individual image scores. (*B*) Inputs to the sampling protocol are the atomic structures of the assembly components and all the restraints. If chemical cross-linking data are available, we build a graph with nodes corresponding to assembly components and edges between cross-linked components; the edge weight is the number of cross-links. A pairwise rigid-body docking is performed between the elements connected by an edge in the maximum spanning tree of the graph. We use the docking sol-



utions to constrain the possible moves of the components during the SA-MC sampling. The models coming from multiple SA-MC runs are improved using the DOMINO algorithm to obtain a set of models for the macromolecular assembly.
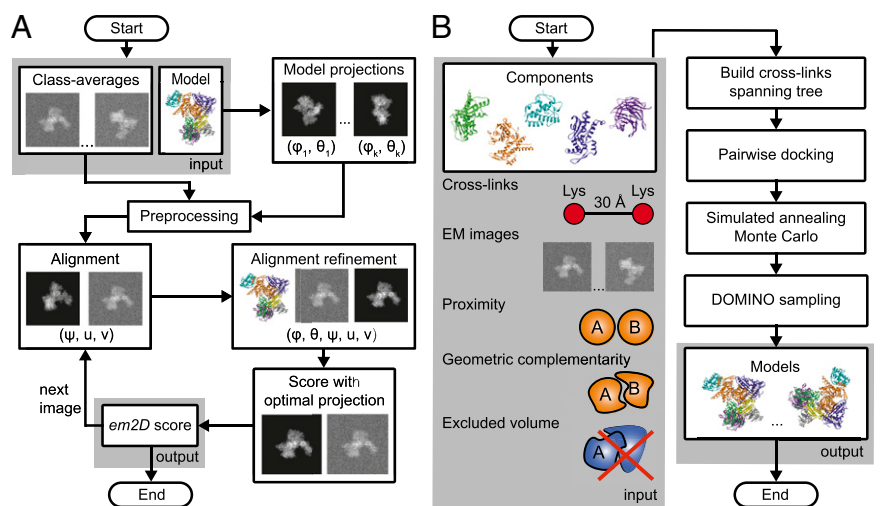
between specific residues, geometric complementarity restraints favoring large contact surfaces between components, and excluded volume restraints to avoid steric clashes between the components. Models for a macromolecular assembly are generated using a two-step sampling protocol consisting of an SA-MC optimization followed by discrete sampling with the DOMINO algorithm (Fig. 1*B*).

**Benchmark Preparation.** The method was tested on 12 assemblies from the Protein Data Bank (PDB), composed of two to five subunits (Table 1). The atomic structure (native configuration) of each complex was down-sampled to a resolution of 10 Å and projected in three random orientations to obtain a set of simulated

class averages (Fig. S1). Misalignment of an assembly in the image was simulated by adding random translations in the image plane within the range of (−3,3) pixels. To simulate experimental noise, zero-mean Gaussian white noise was added to obtain an SNR of 1. We performed 12 simulated experiments for each structure in the benchmark set. The first 6 simulated experiments (image-restrained experiments) included the class averages as restraints and differed in the number of proteomics and cross-linking restraints. The remaining 6 simulated experiments (control experiments) were identical to the first 6, except that they did not use the images.

**Benchmark.** The *<em2D>* score of a model correlates strongly with the all-atom distance-root-mean-square between the model

**Table 1. Results for the benchmark set of complexes modeled using simulated EM class averages as restraints and examples using experimental class averages (entries below the horizontal line)**

| Complex ID | Sub units | Cross-links | ccc of best-scored model | | Placement score of best-scored model | | ccc of best-sampled model | | Placement score of best-sampled model | | Rank of best-sampled model | | Rank of native configuration | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| PDB 2uzx | 2 | 1 | 0.86 | 0.64 | 6.2, 18 | 11.9, 58 | 0.84 | 0.67 | 5.3, 10 | 10.0, 51 | 3 | 226 | 1 | 39 |
| PDB 3r5d | 3 | 7 | 0.84 | 0.63 | 5.7, 20 | 16.4, 43 | 0.86 | 0.75 | 1.80, 19 | 2.4, 46 | 50 | 90 | 1 | 29 |
| PDB 2wvy | 3 | 2 | 0.81 | 0.63 | 6.3, 29 | 11.2, 41 | 0.86 | 0.69 | 3.8, 13 | 9.0, 25 | 16 | 76 | 1 | 6 |
| PDB 3v6d | 3 | 5 | 0.87 | 0.61 | 2.9, 14 | 17.5, 65 | 0.85 | 0.71 | 2.6, 16 | 7.0, 56 | 3 | 67 | 1 | 97 |
| PDB 1z5s | 4 | 3 | 0.81 | 0.57 | 4.9, 49 | 13.5, 76 | 0.82 | 0.61 | 4.8, 46 | 10.8, 54 | 89 | 925 | 1 | 31 |
| PDB 3sfd | 4 | 7 | 0.93 | 0.75 | 1.1, 7 | 10.7, 51 | 0.93 | 0.76 | 1.1, 6 | 8.2, 45 | 3 | 420 | 1 | 12 |
| PDB 3pdu | 4 | 6 | 0.74 | 0.50 | 7.5, 62 | 10.3, 91 | 0.76 | 0.54 | 5.8, 53 | 8.4, 59 | 489 | 76 | 1 | 1 |
| PDB 2y7h | 4 | 6 | 0.73 | 0.47 | 9.2, 34 | 11.0, 63 | 0.79 | 0.57 | 4.1, 46 | 13.7, 58 | 102 | 355 | 1 | 12 |
| PDB 3nvq | 4 | 5 | 0.80 | 0.49 | 8.3, 26 | 17.9, 91 | 0.80 | 0.55 | 7.9, 16 | 15.4, 75 | 4 | 103 | 1 | 43 |
| PDB 3puv | 5 | 11 | 0.65 | 0.50 | 11.3, 50 | 26.9, 84 | 0.66 | 0.57 | 10.3, 65 | 19.5, 81 | 140 | 1,800 | 1 | 1 |
| PDB 3lu0 | 5 | 12 | 0.75 | 0.62 | 15.3, 35 | 18.3, 50 | 0.75 | 0.64 | 11.7, 49 | 17.9, 50 | 57 | 181 | 1 | 1 |
| PDB 1suv | 4 | 4 | 0.83 | 0.61 | 5.9, 11 | 26.8, 80 | 0.85 | 0.69 | 2.7, 11 | 15.7, 57 | 15 | 179 | 1 | 5 |
| PDB 1suv | 4 | 4 | 0.85 | 0.61 | 4.0, 52 | 26.8, 80 | 0.88 | 0.69 | 2.4, 4 | 15.7, 57 | 110 | 179 | 1 | 5 |
| EMDB 1853 | 2 | 2 | 0.86 | 0.73 | 3.5, 10 | 6.7, 40 | 0.89 | 0.73 | 0.9, 24 | 6.7, 40 | 105 | 1 | 33 | 1 |
| EMDB 1876 | 3 | 7 | 0.65 | 0.57 | 14.8, 14 | 17.6, 24 | 0.76 | 0.68 | 6.0, 13 | 7.5,19 | 102 | 1,193 | 1 | 1,200 |
| EMDB 1890 | 4 | 10 | 0.77 | 0.72 | 4.4,13 | 10.2, 43 | 0.78 | 0.74 | 3.4,16 | 5.0, 29 | 87 | 929 | 1 | 1 |

Column 1 shows the PDB or EMDB ID code of the complex, and column 2 shows its number of subunits. The models calculated using the class averages (white background) are compared with those calculated without class averages (gray background); the rest of the restraints are the same (number of cross-linking restraints, column 3). Two models are shown per complex: the model with the minimum value of the scoring function (best scored) and the model most similar to the native configuration (best sampled). The accuracy of a model is measured by the *ccc* between the density map of the native configuration and the density map of the model (columns 4 and 6), and the average placement score of the assembly subunits (columns 5 and 7). The placement score for a subunit is defined as the (placement distance in Å, placement angle in degrees) pair, resulting from comparing the model with the native configuration, in the least-squared superposition of all assembly atoms. The placement distance is given by the subunit centroid distance, and the placement angle is the screw rotation that superposes the model subunit on the native subunit. The rank of a model is its position in the sorted list of values of the scoring function (columns 8 and 9).

and the native configuration as well as with the *ccc* between the density map of the native configuration and the density map of the model (Fig. S2 *A* and *B*). In contrast, the scores for the remaining restraints are weakly correlated with the model accuracy (Fig. S2 *A* and *B*).

We applied principal component analysis (*Materials and Methods*) to determine weights for the restraint terms in the scoring function, thus maximizing the variance of the total score obtained for the models. The weights were set to the first principal component $w = (0.60, 0.33, 0.13, 0.18, 0.08)$. For the final ranking in the image-restrained experiments, however, only the <em2D> score was used because its weight was much larger than that for any other restraint.

We assessed the accuracy of an assembly model using the placement score (16) (Table 1) and the *ccc* between the density map of the native configuration and the density map of the model. For all the structures in the benchmark, the top model in the image-restrained experiments had a lower placement score and a greater *ccc* than the top model in the control experiments. On average, we found models with *ccc* = 0.80 using the class averages and *ccc* = 0.58 without them. We declared as acceptable a model with a placement distance ≤10 Å and a placement angle ≤36° (1/5 of the maximum rotation of 180°), based on an observation that such models tend to have a *ccc* >0.70 and are "visually" similar to the native configuration (Fig. S1). Using the <em2D> score allowed us to find an acceptable model in 8 of the 12 image-restrained experiments. None of the best-scoring models in the control experiments were acceptable (Table 1). Therefore, class averages provide useful information for assembly modeling.

Using the class averages also improved the sampling as well as the scoring. In all cases, the best-sampled model in an image-restrained experiment has a lower placement score and a greater *ccc* than its control counterpart (Table 1). The class averages did not improve sampling for structures containing globular (PDB ID code 1z5s) or cylindrical (PDB ID code 3pdu) components. The models selected in these two cases had overall shape similarity to the native configuration, with correct positions but incorrect orientations (Fig. S1). For PDB ID code 3puv, the method failed to find the native interface between the central subunits (Fig. S1). After clustering the 100 best-scoring models for each assembly, the best-scoring model was generally in the largest cluster, with a few exceptions (Fig. 2A and Fig. S2C). For PDB ID codes 2uzx and 2wvy, the best-scoring solution was not in the largest cluster but was more accurate than the solutions in the largest cluster; for PDB ID codes 3pdu and 3puv, both the largest cluster and the best-scoring solution had incorrect subunit orientations. Finally, the largest cluster from the image-restrained experiments was more accurate and its shape more closely resembled the native shape configuration than that from the control experiments (Fig. 2 *B* and *C* and Fig. S1). The lower bound on the positional and orientational accuracy of an individual subunit is its positional and orientational precision within the best-scoring cluster, respectively, as measured by the SDs of the placement distance and angle (Fig. S3*A*).

As the number of cross-linking restraints used for modeling an assembly decreased, the best-scoring models became less accurate. The positional accuracy was less sensitive to removing cross-linking restraints than the orientational accuracy (Table 1 and Tables S1–S5). Below a structure-specific threshold on the number of cross-linking restraints, the class averages were not sufficient to produce accurate models (Table S4). In the absence of cross-linking data, the orientation of the subunits could not be recovered (Table S5). Decreasing the number of cross-linking restraints further deteriorated the accuracy of the models in the control experiments (Tables S1–S5).

**Application to the Human TfR–Tf Complex.** Tf delivers iron as $Fe^{3+}$ to cells by endocytosis of the complex formed after its binding to
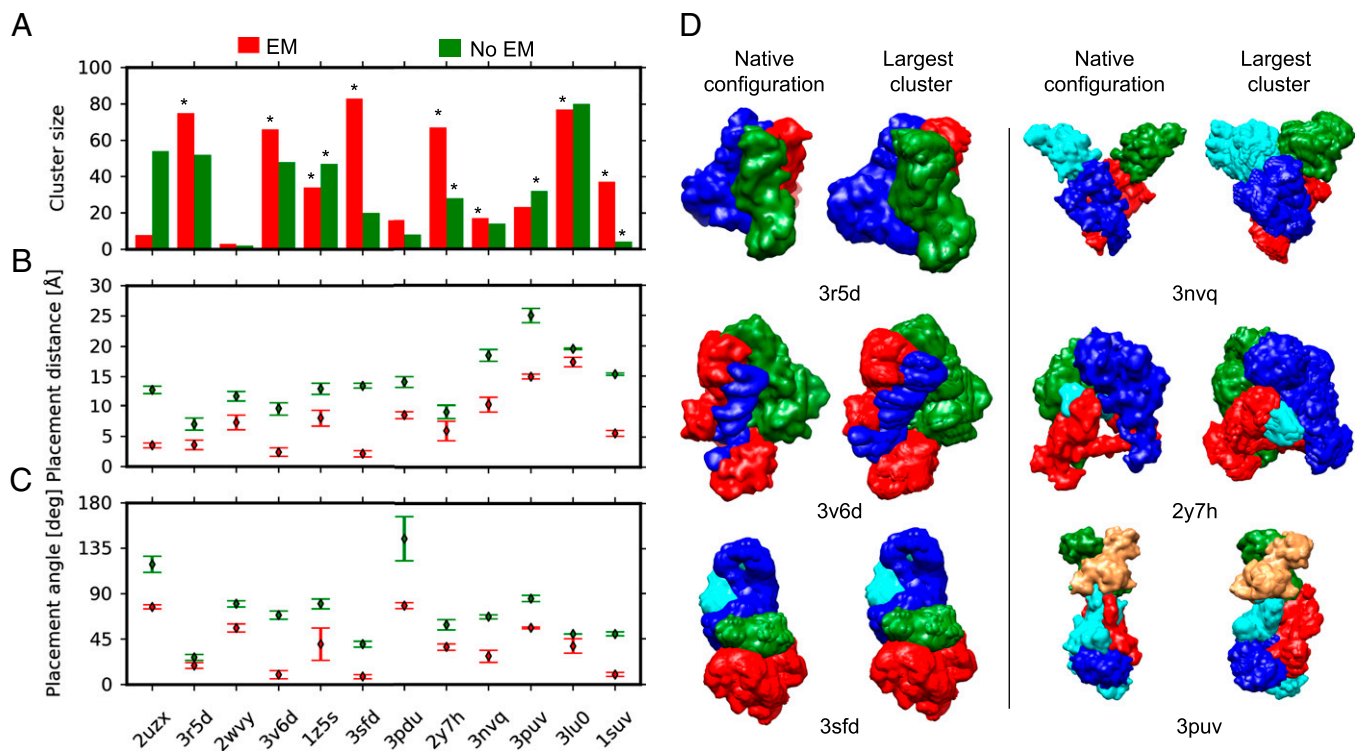


**Fig. 2.** Clusters of the 100 best-scoring models for each structure in the benchmark. (*A*) Size of the largest cluster (asterisk on top of a bar indicates that the best-scoring model is part of the cluster). (*B*) Placement distance of the models in the cluster. (*C*) Placement angle of the models in the cluster. The error bars in *B* and *C* correspond to 1 SD. (*D*) Simulated density map for the native configuration of six structures in the benchmark set compared with the simulated density map generated by the 10 best-scoring solutions in the largest cluster. Each component of the assembly appears in a different color.

the TfR. The atomic structure of the TfR–Tf complex, consisting of two Tfs bound to a single TfR, has been determined by fitting the crystallographic structures of Tf and TfR into a density map from cryo-EM (18). We built models for TfR-Tf using 10 cryo-EM class averages as restraints, as well as two proximity and four cross-linking restraints (Fig. 3 *A* and *B*). The class averages indicate a preferred orientation of the cryofrozen complex, and therefore represent a perfect example of the type of problems that our method addresses. The best-scoring model had a *ccc* with the experimental density map of 0.85, but one molecule of Tf had the N- and C-terminal domains swapped (with a placement score of 4.0 Å, 52°) (Fig. 3 *B* and *D*). After clustering, the largest cluster (35 elements: placement score of $8.6 \pm 0.9$ Å, $13 \pm 1°$) contained the correct positions and orientations. The size and accuracy were similar to those of the largest cluster for the simulated data of PDB ID code 1suv (37 elements: placement score of $5.6 \pm 0.5$ Å, $10 \pm 2°$). The clustering also produced two smaller clusters, corresponding to configurations with one (26 elements: placement score of $17.5 \pm 0.5$ Å, $60 \pm 1°$) and two (24 elements: placement score of $25.5 \pm 0.5$ Å, $158 \pm 2°$) swaps of the Tf N- and C-terminal domains. Removing the image restraints produced models of significantly lower accuracy (Table 1 and Fig. S1).

**Other Examples.** We further illustrate our method by using experimental negative-stain and cryo-EM class averages as restraints to build accurate models for three additional protein assemblies in the Electron Microscopy Data Bank (EMDB): the human proto-convertase C3bB (EMDB ID code 1583) (19), the bovine mitochondrial supercomplex $I_1III_2IV_1$ (EMDB ID code 1876) (20), and the type I restriction modification complex EcoR124I from *E. coli* (EMDB ID code 1890) (21) (Table 1 and Fig. S4).

**Comparison with the MultiFit Algorithm.** We compared models for four sample assemblies with experimentally determined class averages with models obtained by simultaneous rigid fitting of the subunits into the assembly 3D EM density map, using the MultiFit algorithm (16). As expected, the use of 3D information was advantageous. The MultiFit algorithm produced more accurate models, with a slight improvement in the subunits' positions and a larger improvement in their orientations (Table S6).

## Discussion

An integrative approach to structure determination combines spatial restraints derived from multiple experimental techniques together with theoretical and statistical knowledge (22–25). Here, we propose including spatial restraints derived from EM class averages into such an integrative approach. As a result, structural determination can benefit from incomplete EM data (e.g., EM images without sufficient angular coverage to perform a full single-particle reconstruction). We define the *<em2D>* score to rank candidate models according to their similarity to the class averages and describe a sampling algorithm suitable for finding assembly models that fit these images.

Our benchmark demonstrates that including image restraints improves the accuracy of the scoring function as well as the thoroughness of configurational sampling. Sampled models had more accurate component positions and orientations than the models obtained without the images, as was apparent from the improved *ccc* between the density maps of the models and the density maps of the native configurations (Table 1). Furthermore, using the *<em2D>* score alone produces more accurate models than using the rest of the restraints combined (Table 1). Due to the weak correlation between the accuracy and other types of restraints used here (proteomics, cross-linking, geometric complementarity, and excluded volume), the information contained in the class averages is critical for computing accurate models. In particular, a few cross-linking restraints that otherwise would not produce acceptable models may be all that is required in addition
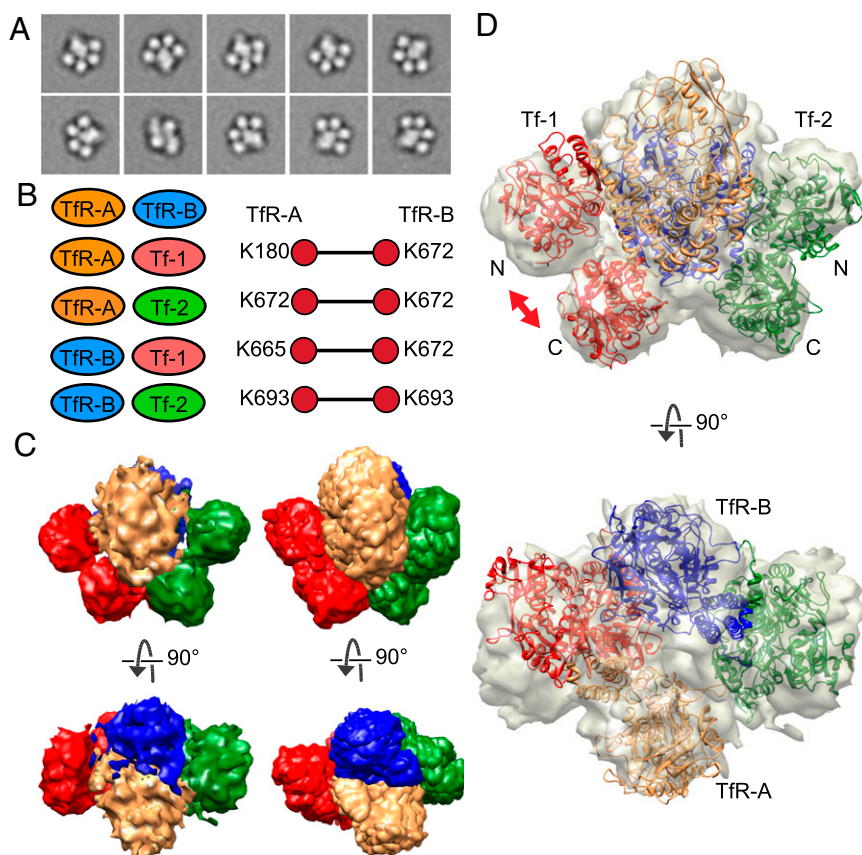


**Fig. 3.** Model for the TfR–Tf complex. (*A*) Class averages used for modeling, corresponding to the 10 most populated class averages from the cryo-EM study. (*B*) Proximity restraints (colored ellipses) and cross-linking restraints (connected circles) used for modeling. Tf-1, first molecule of Tf (red); Tf-2, second molecule of Tf (green); TfR-A, first monomer of the receptor (gold); TfR-B, second monomer of the receptor (blue). (*C*) Comparison of the experimental cryo-EM density map (*Left*) with the simulated density map (8 Å) of the 10 best-scoring solutions in the largest cluster (*Right*). (*D*) Fitting of the best-scoring model into the experimental cryo-EM density map. The N- and C-terminal domains of Tf-1 are swapped with respect to the correct configuration (red arrow).

to the class averages to determine the relative orientations of the subunits accurately (Table 1). Integrating additional restraints into our scoring and sampling protocol is straightforward. EM class averages are expected to be generally helpful when other data do not include complete 3D shape information (e.g., fluorescence resonant energy transfer spectroscopy, small angle X-ray scattering, ion mobility spectrometry).

For the TfR–Tf complex example, we were able to generate and identify accurate models using two proximity restraints, four cross-linking restraints, and restraints from 10 experimental cryo-EM class averages (Fig. 3). The results observed using simulated class averages for the TfR–Tf complex were similar to those obtained with experimental images, increasing our confidence in the simulated benchmark. Finding the correct positions and orientations of the N- and C-terminal domains of Tf presented a challenge in the original work (18). In our modeling, we computed three clusters of solutions with the correct positions and orientations, one domain swap, and two domain swaps, respectively. Because the sizes and scores of the clusters were comparable, we could not pick the correct solution, given the used information. In the original work, the correct solution was identified based on additional negative-stain class averages of complexes of isolated Tf N- and C-domains with TfR.

Although our approach is less limited by incomplete and noisy EM data than single-particle reconstruction, our scoring function may still not be able to distinguish between a large number of different models, or the sampling algorithm may not find the best-scoring solution. In the former case, the variability of the ensemble of good-scoring solutions may be a reasonable lower bound on the accuracy (Fig. S3A). In the latter case, sufficient sampling is indicated when additional sampling does not generate structurally distinct good-scoring solutions (Fig. S3B). Insufficient sampling results in best-scoring models corresponding to local minima of the scoring function, producing many clusters of small size (Fig. S3C).

Many factors can affect the quality of the EM images, including the contrast transfer function of the microscope, noise, distortions of the assembly shape due to the stain, conformational or compositional heterogeneities, and lack of characteristic views of the assembly in the images. Noise and stain distortions are not severely limiting, because obtaining class averages with an improved SNR is often possible by collecting more single-particle images. However, images of an assembly in characteristic orientations, resulting in the unique projections of the assembly, are critical for the success of the method. Therefore, globular assemblies are more difficult to determine accurately based on class averages (e.g., PDB ID codes 1z5s and 3pdu in Table 1). Here, we have not addressed using class averages of an assembly in different configurations or of varied composition.

In principle, the $<em2D>$ score can be improved by removing the assumption that all pixels in all images are affected by the same noise. This goal can be achieved by assigning specific noise $\sigma_j$ to each pixel (Eq. 2), if known. When computing the $<em2D>$ score using class averages, $\sigma_j$ could be obtained from the variance images that are calculated together with the class averages. However, this generalization would be computationally costly because the images could then not be aligned using a Fourier method.

We expect our approach to be helpful for determining the structures of macromolecular assemblies that show preferred orientations during EM experiments as well as assemblies on the lower end of the EM size spectrum, where obtaining a density map using single-particle reconstruction generally takes considerably longer than obtaining a relatively small number of class-average images.

## Materials and Methods

**Representation.** We assume that experimentally determined or computationally predicted atomic models are available for all components of the assembly. Different model representations are used to calculate different types of restraints. First, 10 residues are represented with a single bead for computing the proximity and excluded volume restraints. Second, one residue is represented with a bead for computing cross-linking restraints. Third,

atoms are projected onto the closest grid point on a 1-Å cubic lattice for computing the geometric complementarity restraint. Fourth, the original atomic resolution is used for computing the em2D restraint.

**Scoring.** The em2D score quantifies the similarity between a model and a set of EM images (Fig. 1A). An EM image **d** is a projection of the density map of the assembly **s**, corrupted by additive noise **n** (i.e., $\mathbf{d} = \mathbf{s} + \mathbf{n}$). The Bayes theorem states that the probability of a model **X**, given the observed data **d** is (15, 26, 27):

$$P(\mathbf{X}|\mathbf{d}) = \frac{P(\mathbf{d}|\mathbf{X})P(\mathbf{X})}{P(\mathbf{d})} \propto P(\mathbf{d}|\mathbf{X}). \qquad [1]$$

The proportionality between the posterior probability $P(\mathbf{X}|\mathbf{d})$ and the likelihood $P(\mathbf{d}|\mathbf{X})$ applies because $P(\mathbf{X})$ and $P(\mathbf{d})$ do not change the ranking of different models **X**, given image **d**. Thus, maximizing the posterior probability is equivalent to maximizing the likelihood. When the noise is modeled by a Gaussian distribution $N(0, \sigma_j)$ affecting each pixel independently, the log-likelihood is:

$$-\log P(\mathbf{d}|\mathbf{X}) = -\log \prod_{j=1}^{M} \exp\left[\frac{-1}{2}\left(\frac{d_j - p_j}{\sigma}\right)^2\right] = \frac{1}{2\sigma^2}\sum_{j=1}^{M}\left(d_j - p_j\right)^2, \qquad [2]$$

where $p$ is the projection of the model **X**, $j$ indicates the overlapping pixels of the image and the model projection, and $M$ is the number of pixels. The specific noise $\sigma_j$ is presumed to be constant for all images. We define the em2D score as the minimal difference between the image and a model projection (28, 29):

$$em2D = \min_\alpha \sum_{j=1}^{M}\left(d_j - p_j(\alpha)\right)^2/2M, \qquad [3]$$

where $\alpha$ represents the five registration parameters, including the three Euler angles of rotation $(\Phi, \theta, \psi)$ and the in-plane translation distances $(u, v)$. It follows from Eqs. 2 and 3 that computing the em2D score is equivalent to minimizing the log-likelihood. If both the image and the projection are normalized to the mean of 0 and SD of 1, the em2D score is 1 minus the ccc between the image and the optimal model projection (1 − ccc). Thus, the em2D score is 0 if the model matches the image perfectly (e.g., for the native configuration and a noise-free image) and 1 if there is no correlation between the projection of the model and the image (e.g., for an image with only noise). In practice, model errors and/or image noise invariably results in em2D > 0, with the best possible value of $1 - (SNR/(1+SNR))^{1/2}$ for an image with the SNR and its perfectly matching noise-free projection (30). When more than one image of a macromolecule in a single configuration is available, minimizing the log-likelihood of observing all images is equivalent to minimizing the average $<em2D>$.

The em2D score of a model is calculated in three steps: preprocessing, a coarse search, and a refined search in the vicinity of the coarse solutions. The preprocessing starts by generating $k$ (typically $20 \le k \le 100$) projections of the model, with directions of projection uniformly distributed in the hemisphere with positive y-axis values. The direction of projection provides an estimation of the first two registration parameters of $\alpha$ $(\Phi, \theta)$. A projection **p** is calculated as the sum of projected 3D Gaussian functions centered on each atom, with weights depending on atom type and projection resolution (31); for efficiency, the Gaussian functions are preprojected and stored in memory. The autocorrelation function (ACF) in the polar coordinates and fast Fourier transform (FFT) of the ACF are calculated for each projection. In addition, the FFT of each image is computed. In the coarse search, each projection $\mathbf{p}_j$ is aligned with each image $\mathbf{d}_i$, resulting in the estimated values of the registration parameters $(\psi, u, v)$ of $\alpha_{ij}$, using a rotational alignment based on the ACF followed by translational alignment based on the FFT (4, 32). In the refinement step, the $k$ vectors $\alpha_{ij}$ for image $\mathbf{d}_i$ obtained from the coarse search are sorted by descending value of the corresponding ccc, and the first $l \le k$ of them are refined (typically, $l = 2$). Each $\alpha_{ij}$ is refined by the Simplex algorithm to maximize the ccc between image $\mathbf{d}_i$ and projection $\mathbf{p}_j$ (28). The vector $\alpha_{ij}$ that gives the maximum ccc after the refinement is used to compute the em2D score. Calculating the em2D score of a model scales linearly with the number of class averages used as restraints.

Our complete scoring function $F$ is $w_{em2d}em2d + w_c f_c + w_d f_d + w_g f_g + w_e f_e$, where $\mathbf{w} = (w_{em2d}, w_c, w_d, w_g, w_e)$ are the weights for the five different restraint types. The proximity term $f_c$ is a sum of all proximity restraints, each one of which is calculated as a harmonic upper-bound function $k(d - d_0)^2/2$ of the shortest distance $d$ between beads of the two involved components.

The harmonic function is set to 0 when $d < d_0 = 10$ Å and $k = 0.0656$ Å$^{-2}$. Similarly, the term $f_d$ is a sum of all distance restraints between a pair of residues, where each individual restraint is calculated as a harmonic upper-bound function with $d_0 = 30$ Å and $k = 0.148$ Å$^{-2}$. The term $f_g$ corresponds to the sum of all geometric complementarity restraints (33), and $f_e$ is a sum of all excluded volume restraints, each one of which is calculated as a harmonic lower-bound function that is set to 0 when $d > d_0 = 2$ and has $k = 2$ Å. To compute the weights $\mathbf{w}$, we selected all models for all benchmark structures with a placement score better than (10 Å, 36°) and performed a principal component analysis on the values of the restraints ($em2d, f_c, f_d, f_g, f_e$). The first eigenvector was selected as the weight $\mathbf{w}$.

**Sampling.** We search for the models of the macromolecular assembly that minimally frustrate the restraints using an SA-MC optimization in continuous space, followed by further optimization with the divide-and-conquer message-passing algorithm DOMINO (16) as follows.

First, if cross-linking data are available, we build a graph with nodes corresponding to the assembly components and edges between the cross-linked components; the edge weight is the number of cross-links. The component with the largest number of connections in the maximum spanning tree of the graph is used as an anchor, and its coordinates are fixed. Each edge in the maximum spanning tree triggers a pairwise docking between the two connected components, in the order of decreasing edge weight. A pairwise docking is performed with HEXDOCK (34). Only solutions satisfying the cross-linking restraints are stored, flagging as the receptor the component with the largest edge weight.

Second, we run a number of independent SA-MC optimizations (*N,* typically 500), each of which starts with random positions of the assembly components. In a single independent run, ~100,000 MC moves of two types are applied with approximately even probability, iterating through a simulated annealing schedule with increasing and decreasing temperatures (see the configuration files of the benchmark). In the first type of a move, one of the stored pairwise docking solutions is randomly chosen and used to propose a new relative position for the ligand component relative to its receptor. In the second type, all components are reoriented randomly for up to 180° and displaced randomly for up to a subunit size in an attempt to compensate for docking inaccuracy.

Third, we attempt to improve on the *N* final SA-MC solutions by a divide-and-conquer DOMINO algorithm (16). Briefly, the DOMINO algorithm identifies subsets of these solutions that can be assembled self-consistently into a better scoring output solution. In more detail, the DOMINO algorithm proceeds in four steps. First, the scoring function is represented as a graph, where the nodes are the variables to be sampled (here, the positions and orientations of the assembly components) and the edges are the pairwise terms acting on these variables. Second, the set of variables is decomposed into overlapping subsets that are "loosely" coupled (i.e., subsets contain common variables); these subsets are represented as nodes in a junction tree where edges are drawn between subsets containing the same variables. Third, the possible discrete states for each subset are generated by the *m* best SA-MC solutions according to the <*em2D*> score (typically, $50 < m < 100$). Finally, the DOMINO algorithm uses message passing to gather self-consistent combinations of these subset states, and thus to obtain final good-scoring solutions. We decrease the DOMINO algorithm running time and memory requirements by keeping only the 2,000 best-scoring configurations from each gathering step, which results in a 100- to 1,000-fold increase in speed without significantly affecting the accuracy of the final solutions.

Finally, we cluster the 100 best-scoring solutions from the DOMINO algorithm using the betweenness centrality clustering algorithm (35), with the all-atom rmsd between the models as the metric.

The method is implemented in the EMageFit module of our open source package IMP (http://integrativemodeling.org/) (25) (Version 15319 of the SVN repository of the package). The benchmark and scripts are available at ftp://salilab.org/javi/benchmark_em2d.tgz.

To compare using 2D class averages vs. 3D density maps, we used a version of the MultiFit algorithm (16) modified to include cross-linking restraints in the scoring function. MultiFit applies inferential optimization on a discrete sampling grid to determine the best-scoring positions and orientations of multiple subunits, given their structures, the assembly density map, and the cross-linking restraints. MultiFit is implemented in the MultiFit module of IMP.

1. Orlova EV, Saibil HR (2011) Structural analysis of macromolecular assemblies by electron microscopy. *Chem Rev* 111(12):7710–7748.
2. Rossmann MG, Battisti AJ, Plevka P (2011) Future prospects. *Adv Protein Chem Struct Biol* 82:101–121.
3. Zhou ZH (2011) Atomic resolution cryo electron microscopy of macromolecular complexes. *Adv Protein Chem Struct Biol* 82:1–35.
4. Frank J (2006) *Three-Dimensional Electron Microscopy of Macromolecular Assemblies: Visualization of Biological Molecules in Their Native State* (Oxford Univ Press, New York), 2nd Ed.
5. Glaeser RM, Hall RJ (2011) Reaching the information limit in cryo-EM of biological macromolecules: Experimental aspects. *Biophys J* 100(10):2331–2337.
6. Volkmann N (2012) Putting structure into context: Fitting of atomic models into electron microscopic and electron tomographic reconstructions. *Curr Opin Cell Biol* 24(1):141–147.
7. Grigorieff N (2000) Resolution measurement in structures derived from single particles. *Acta Crystallogr D Biol Crystallogr* 56(Pt 10):1270–1277.
8. Stewart A, Grigorieff N (2004) Noise bias in the refinement of structures derived from single particles. *Ultramicroscopy* 102(1):67–84.
9. Yang S, Yu X, Galkin VE, Egelman EH (2003) Issues of resolution and polymorphism in single-particle reconstruction. *J Struct Biol* 144(1-2):162–171.
10. Zhou ZH (2008) Towards atomic resolution structural determination by single-particle cryo-electron microscopy. *Curr Opin Struct Biol* 18(2):218–228.
11. Radermacher M (1988) Three-dimensional reconstruction of single particles from random and nonrandom tilt series. *J Electron Microsc Tech* 9(4):359–394.
12. Van Heel M (1987) Angular reconstitution: A posteriori assignment of projection directions for 3D reconstruction. *Ultramicroscopy* 21(2):111–123.
13. Penczek PA, Zhu J, Frank J (1996) A common-lines based method for determining orientations for N > 3 particle projections simultaneously. *Ultramicroscopy* 63(3-4):205–218.
14. Singer A, Shkolnisky Y (2011) Three-Dimensional Structure Determination from Common Lines in Cryo-EM by Eigenvectors and Semidefinite Programming(). *SIAM J Imaging Sci* 4(2):543–572.
15. Jaitly N, Brubaker MA, Rubinstein JL, Lilien RH (2010) A Bayesian method for 3D macromolecular structure inference using class average images from single particle electron microscopy. *Bioinformatics* 26(19):2406–2415.
16. Lasker K, Topf M, Sali A, Wolfson HJ (2009) Inferential optimization for simultaneous fitting of multiple components into a CryoEM map of their assembly. *J Mol Biol* 388(1):180–194.
17. Zhang J, Minary P, Levitt M (2012) Multiscale natural moves refine macromolecules using single-particle electron microscopy projection images. *Proc Natl Acad Sci USA* 109(25):9845–9850.
18. Cheng Y, Zak O, Aisen P, Harrison SC, Walz T (2004) Structure of the human transferrin receptor-transferrin complex. *Cell* 116(4):565–576.
19. Torreira E, Tortajada A, Montes T, Rodríguez de Córdoba S, Llorca O (2009) 3D structure of the C3bB complex provides insights into the activation and regulation of the complement alternative pathway convertase. *Proc Natl Acad Sci USA* 106(3):882–887.
20. Althoff T, Mills DJ, Popot JL, Kühlbrandt W (2011) Arrangement of electron transport chain components in bovine mitochondrial supercomplex I1III2IV1. *EMBO J* 30(22):4652–4664.
21. Kennaway CK, et al. (2012) Structure and operation of the DNA-translocating type I DNA restriction enzymes. *Genes Dev* 26(1):92–104.
22. Alber F, et al. (2007) Determining the architectures of macromolecular assemblies. *Nature* 450(7170):683–694.
23. Robinson CV, Sali A, Baumeister W (2007) The molecular sociology of the cell. *Nature* 450(7172):973–982.
24. Steven AC, Baumeister W (2008) The future is hybrid. *J Struct Biol* 163(3):186–195.
25. Russel D, et al. (2012) Putting the pieces together: Integrative modeling platform software for structure determination of macromolecular assemblies. *PLoS Biol* 10(1):e1001244.
26. Rieping W, Habeck M, Nilges M (2005) Inferential structure determination. *Science* 309(5732):303–306.
27. Scheres SHW, Valle M, Grob P, Nogales E, Carazo J-M (2009) Maximum likelihood refinement of electron microscopy data with normalization errors. *J Struct Biol* 166(2):234–240.
28. Yang Z, Penczek PA (2008) Cryo-EM image alignment based on nonuniform fast Fourier transform. *Ultramicroscopy* 108(9):959–969.
29. Yang C, Ng EG, Penczek PA (2005) Unified 3-D structure and projection orientation refinement using quasi-Newton algorithm. *J Struct Biol* 149(1):53–64.
30. Frank J (1975) Practical resolution criterion in optics and electron microscopy. *Optik (Stuttg)* 43(1):25–34.
31. Alber F, Förster F, Korkin D, Topf M, Sali A (2008) Integrating diverse data for structure determination of macromolecular assemblies. *Annu Rev Biochem* 77:443–477.
32. Penczek P, Radermacher M, Frank J (1992) Three-dimensional reconstruction of single particles embedded in ice. *Ultramicroscopy* 40(1):33–53.
33. Lawrence MC, Colman PM (1993) Shape complementarity at protein/protein interfaces. *J Mol Biol* 234(4):946–950.
34. Ritchie DW, Venkatraman V (2010) Ultra-fast FFT protein docking on graphics processors. *Bioinformatics* 26(19):2398–2405.
35. Brandes U (2001) A faster algorithm for betweenness centrality. *J Math Sociol* 25(2): 163–177.