Supplementary Materials

1 DATA INTEGRATION FOR MULTIPLE COMPONENT FITTING

As an example, we assume that protein binding sites are known that form a protein interaction in the assembly. In such a case it is possible to identify the feature points in each component that are closest to the binding interface (Figure S1). The distance between these feature points must be below a distance cutoff C so that the interface information is satisfied. The intercomponent matrix R can be modified to incorporate such binding interface information. For two interacting components the inter-component matrix, values are set to a positive value. A positive value in R ensures that the corresponding distance between the feature points of two different components will be considered in the objective function and therefore optimization process. We define the cutoff distance between the two interacting feature points as the sum of the Voronoi radii of the two feature points.

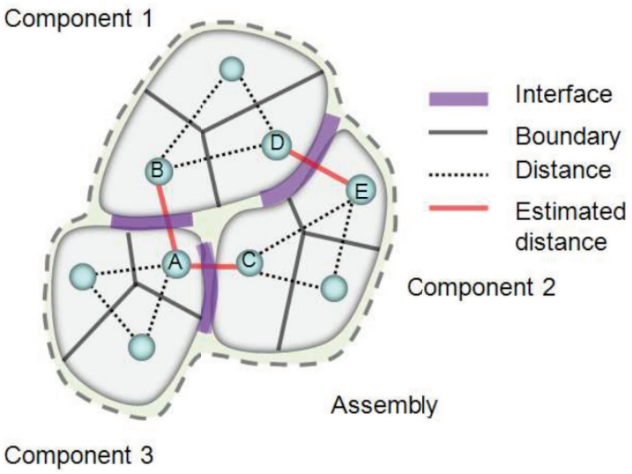

Fig. S1. Integration of protein interaction information in the IQP fitting process. A schematic view of interacting proteins and their binding interfaces . An assembly with three feature points per component is shown. Binding interfaces between proteins are indicated by thick lines. The volume of the assembly is divided into Voronoi cells, each containing a single feature point at its center. Feature points are divided into two classes: Feature points in Voronoi cells that share a binding interface with another component (dark spheres) and feature points of Voronoi without binding interfaces (light spheres). If a binding interface spans two Voronoi cell, the feature point representing the Voronoi cell with the largest fraction of the binding interface is selected as interaction feature point. Expected binding distance between interaction feature points can be estimated from the distance between feature points and binding interfaces.

2 AVERAGE NUMBER OF IQP SOLUTIONS

The IQP is a local optimization method. Fortunately, it only produce a limited number of solutions for current feature point matching problem. The averaged number of solutions of 10 IQP runs is about 2.2 (Figure S2).

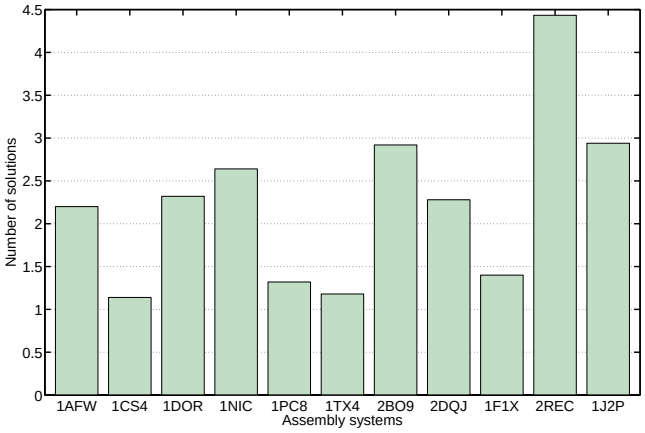

Fig. S2. Average number of IQP solutions for each assembly system. The average number of solutions for IQP is calculated based on 10 IQP runs on 50 groups of feature points computed by VQ.

3 WICP REFINEMENT

The weighted ICP framework can be stated as follows.
Flowchart of the weighted ICP algorithm:

- **Input:** The two sets of voxel point coordinates $X = [x_1, x_2, ..., x_N]$ and $Y = [y_1, y_2, ..., y_M]$ with N and M points as well their corresponding density values $\rho(x_i)$ and $\rho(y_j)$ are given. Initial transformation matrix T_0 determined by the IQP procedure.
- **Output:** The 'optimal' index map ϕ, transformation matrix T and wRMSD.
- **Step-0**: Initialize the procedure and set the iteration index $k = 0$.
- **Step-1:** Apply the transformation: $X^{k+1} = T_{k+1} \cdot X^k$. Set $k = k + 1$.
- **Step-2**: Compute the closest voxel points and determine the correspondence ϕ. Set

$$\phi(i) = \text{arg}min_{j\in\{1,\cdots,M\}} w_{ij} \|X_i^k - Y_{\phi(i)}\|$$

  So that $Y_{\phi(i)}$ is the closest point to the X_i.
- **Step-3:** Compute new transformation T_{k+1} by a weighted least-square fitting using the wRMSD metric. Set

$$T_{k+1} = \text{arg}min \sum_{i=1}^{n_c} w_{i\phi(j)} \|X_i^k - Y_{\phi(i)}\|$$

  T_{k+1} is computed by the singular value decomposition (SVD) algorithm.

- **Step-4:** Terminate iteration if the change of wRMSD or T falls below a preset threshold $\tau > 0$ or the number of iterations is larger than a preset maximum step. Otherwise go to **Step-1**.

4 CONVERGENCE RADIUS OF WICP

To assess the radius of convergence of the wICP algorithm, we generate a suit of starting structures of different orientations to the native structure of 1PC8a. Specifically, we generate new orientations based on the rotation matrix $R_{xyz}(\alpha, \beta, \gamma) = R_z(\gamma)R_y(\beta)R_x(\alpha)$ with the three rotation angles α,β,γ. The three angles indicate the degree of deviation from the original map. We calculate the C_α RMSD as the measure of convergence. An RMSD of close to zero, indicates that wICP converges to the optimal solution (Figure S2). The wICP algorithm converges to the best solution from relatively large deviations, indicating the robustness of the method and its large 'radius of convergence'.

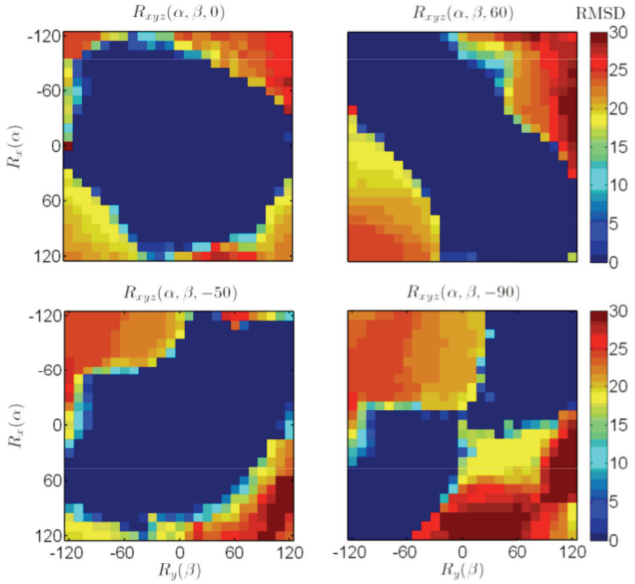

Fig. S3. Convergence Radius of wICP.