Bioinformatics Advance Access published October 23, 2012

A Method for Integrative Structure Determination of Protein-Protein Complexes

Dina Schneidman-Duhovny[1*], Andrea Rossi[2], Agustin Avila-Sakar[4], Seung Joong Kim[1], Javier Velázquez-Muriel[1], Pavel Strop[2], Hong Liang[2], Kristin A. Krukenberg[3], Maofu Liao[4], Ho Min Kim[4], Solmaz Sobhanifar[5], Volker Dötsch[5], Arvind Rajpal[2], Jaume Pons[2], David A. Agard[6,7], Yifan Cheng[4], Andrej Sali[1*]

[1]Department of Bioengineering and Therapeutic Sciences, Department of Pharmaceutical Chemistry, and California Institute for Quantitative Biosciences (QB3), University of California, San Francisco, [2]Rinat-Pfizer Inc, [3]Graduate Program in Chemistry and Chemical Biology, University of California, San Francisco, [4]The W.M. Keck Advanced Microscopy Laboratory, Department of Biochemistry and Biophysics, University of California, San Francisco, [5]Institute of Biophysical Chemistry and Center for Biomolecular Magnetic Resonance and Cluster of Excellence Macromolecular Complexes, University of Frankfurt, [6]Department of Biochemistry and Biophysics, University of California, San Francisco , [7]The Howard Hughes Medical Institute, University of California, San Francisco

Associate Editor: Dr. Dina Schneidman-Duhovny

ABSTRACT

Motivation: Structural characterization of protein interactions is necessary for understanding and modulating biological processes. On one hand, X-ray crystallography or NMR spectroscopy provide atomic resolution structures but the data collection process is typically long and the success rate is low. On the other hand, computational methods for modeling assembly structures from individual components frequently suffer from high false positive rate, rarely resulting in an unique solution.

Results: Here, we present a combined approach that computationally integrates data from a variety of fast and accessible experimental techniques for rapid and accurate structure determination of protein-protein complexes. The integrative method uses atomistic models of two interacting proteins and one or more datasets from five accessible experimental techniques: a SAXS profile, 2D class average images from negative stain EM, a 3D density map from single particle negative stain EM, residue type content of the protein-protein interface from NMR spectroscopy, and chemical cross-linking detected by mass spectrometry. The method is tested on a docking benchmark consisting of 176 known complex structures and simulated experimental data. The near-native model is the top scoring one for up to 61% of benchmark cases depending on the included experimental datasets; in comparison to 10% for standard computational docking. We also collected SAXS, 2D class average images, and 3D density map from negative stain EM to model the PCSK9 antigen – J16 Fab antibody complex, followed by validation of the model by a subsequently available X-ray crystallographic structure.

Availability: http://salilab.org/idock

Supplementary information: Supplementary data are available at *Bioinformatics* online

Contact: dina@salilab.org; sali@salilab.org

1 INTRODUCTION

Biologists are identifying components of macromolecular assemblies and networks (Krogan, et al., 2006). To understand how these assemblies and networks underpin essential biological processes and to modulate them for therapeutic purposes, we need to describe the structures of both natural and engineered protein interactions (Robinson, et al., 2007). Due to the difficulty of determining the atomic structures of protein complexes by X-ray crystallography and NMR spectroscopy as well as inaccuracy of alternative methods, such as protein-protein docking, new techniques are necessary (Alber, et al., 2008).

One major computational approach to predicting structures of protein complexes relies on molecular docking of unbound single component structures. Even for complexes with two proteins, docking problem remains challenging despite recent advances (Lensink and Wodak, 2010). The major bottlenecks include dealing with protein flexibility and the absence of an accurate scoring function (Ritchie, 2008). Pairwise protein docking methods can be divided into three classes based on their configurational sampling algorithm (Vajda and Kozakov, 2009): (i) global methods using a fast Fourier transform (FFT) (Eisenstein and Katchalski-Katzir, 2004) or geometric matching (Schneidman-Duhovny, et al., 2005), (ii) medium-range methods such as Monte Carlo sampling (Fernandez-Recio, et al., 2003; Gray, et al., 2003), and (iii) methods guided by data, such as complex refinement based on NMR restraints, cross-linking, interface prediction, or site-directed mutagenesis (Dominguez, et al., 2003; Sivasubramanian, et al., 2006). It is common to begin docking two proteins with an unbiased global search followed by refinement of the best scoring models (Mashiach, et al., 2010).

Characterizing the structures of multi-subunit complexes benefits from using varied experimental datasets (Alber, et al., 2007; Alber,

et al., 2007; Russel, et al., 2012). In this hybrid or integrative approach, the datasets are encoded into a scoring function used to evaluate candidate models generated by a sampling method. Integrative structure determination typically iterates through the following stages: (i) gathering information, (ii) designing model representation and evaluation, (iii) sampling good models, and (iv) analyzing models and information.

Here, we present an integrative approach to pairwise protein docking. First, data from one or more of five different experiment types are translated into the corresponding scoring function terms. These data include (i) the pair-distance distribution function of the complex from a small angle X-ray scattering (SAXS) profile, (ii) 2D class average images of the complex from negative stain electron microscopy micrographs (EM2D), (iii) a 3D density map of the complex from single particle negative-stain electron microscopy (EM3D), (iv) residue type content at the protein interface from NMR spectroscopy (NMR-RTC) (Reese and Dötsch, 2003), and (v) chemical cross-linking detected by mass spectrometry (CXMS). These five experimental methods were selected due to their feasibility and efficiency of data collection: a SAXS profile of the complex in solution can be collected in several minutes (Hura, et al., 2009); a 3D EM density map can be reconstructed from a smaller sample amount than that for SAXS, but data collection process is significantly longer (Stahlberg and Walz, 2008); 2D class averages can be computed from micrographs more easily and rapidly than performing a full 3D reconstruction; the composition of interface residues from NMR (Reese and Dötsch, 2003) provides information about the interaction interface, unlike the SAXS and EM data; and cross-linking data (Rappsilber, 2011) provides information at intermediate resolution imposing an upper distance bound on inter-molecular pairs of residues. Second, complex models are sampled, relying on efficient global search methods developed for pairwise protein docking, followed by filtering based on fit to the experimental data, conformational refinement, and composite scoring. Third, good-scoring representatives of clusters of models are picked as final models.

To validate this approach, we apply the integrative method in two contexts. First, we test the method on a large benchmark for protein docking (Hwang, et al., 2010) with simulated experimental data and known complex structures. This test allows a robust assessment of the value of the individual types of experimental data for specific types of proteins. Second, we also collected SAXS, EM2D, and EM3D data to model the PCSK9 antigen – J16 Fab antibody complex, followed by validation of the model by a subsequently available X-ray crystallographic structure. This second test highlights the advantages of the integrative method that allows computing an accurate model in a timely manner.

2 METHODS

2.1 Integrative Docking Method Summary

Given the atomic structures of two proteins and one or more datasets from SAXS, EM2D, EM3D, NMR-RTC, and CXMS we compute the 3D structure of their complex. The approach involves four major stages (Fig. 1, Supplementary Data):

Stage 1: Global Search. A global search in the space of all possible docking models is performed using geometry-based molecular docking (Duhovny, et al., 2002). The configurational sampling precision is increased significantly compared to the default settings (Table S1: from $4.5{\cdot}10^3$ to $212{\cdot}10^3$ models) to ensure the interface and global shape of the complex are sampled with precision commensurate with that of the data.

Stage 2: Data-guided Filtering. Each available experimental dataset is used independently for scoring and filtering of models inconsistent with the data. To account for noise in the data, we convert the data into soft restraints. For SAXS profile, a model is filtered out if its radius of gyration is in significant disagreement with the experimentally derived one (Schneidman-Duhovny, et al., 2011). For EM2D class averages, there is no filter. For EM3D density map, a model is filtered only if it significantly protrudes out of the density map. For NMR-RTC data, a model is filtered out if it does not satisfy at least half of the specified residue type frequencies. For the cross-linking data, a model is filtered if it does not satisfy any of the cross-links. For each data type, the scores of the remaining models are normalized, using the average and standard deviation of their scores (S_{Data}). This normalization facilitates combining and comparing of scores for different data types with different noise levels. The models are clustered, and the cluster representative with the best fit to the data is selected. Top scoring 5000 cluster representatives are processed further. This number of models usually guarantees that near-native models are not excluded even in the case of noisy data (Table S1).

Stage 3: Conformational Refinement. The goal of this stage is to compute an interface energy score. Since rigid docking models may contain steric clashes, the side chain conformations as well as relative positions and orientations of the model components are refined, and an interface energy score (S_{Energy}) is computed (Andrusier, et al., 2007; Mashiach, et al., 2010).

Stage 4: Composite Scoring. The final models are scored and ranked by a composite score consisting of a normalized interface energy term and the fit to the data: $S_{Composite} = S_{Energy} + \sum_i S_{Data_i}$

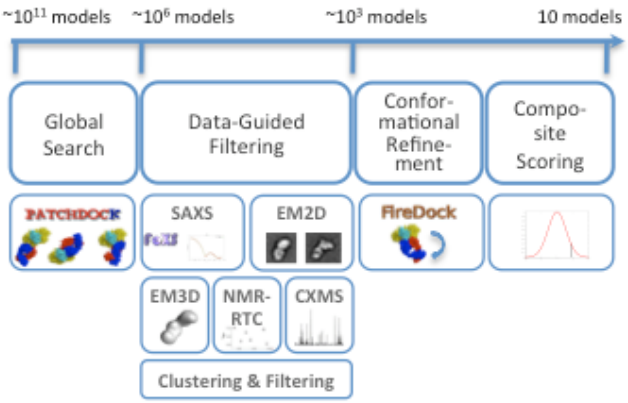

Fig. 1. Schematic representation of the integrative docking method. The number of possible configurations for two docked proteins is on the order of ~1011 (3 rotational degrees of freedom sampled in 5 degrees interval and 3 translational degrees of freedom sampled at 1Å interval). As the method proceeds, the number of considered configurations decreases.

2.2 Benchmark Using Simulated Experimental Data

Pairwise protein docking benchmark 4.0 (Hwang, et al., 2010) is used to validate integrative docking method. This benchmark contains 176 complexes and their corresponding unbound structures, classified into 121 low-difficulty or rigid-body cases, 30 medium-difficulty cases, and 25 high-difficulty cases, based on the degree of conformational change at the interface upon complex formation. For testing EM2D and EM3D, only a subset of 27 complexes with more than 675 residues is used (EM benchmark). These complexes are divided into 16 rigid-body, 4 medium-difficulty, and 7 difficult cases. Each of the benchmark complexes also had SAXS,

EM2D, EM3D, NMR-RTC, and CXMS data simulated based on its native complex structure (Supplementary Data). We have also tested the method on three experimental SAXS dataset from pyDockSAXS benchmark (Niemann, et al., 2008; Pinotsis, et al., 2008; Pons, et al., 2010; Schubert, et al., 2002). Integrative docking is performed on each of these cases starting from the unbound structures and the predicted complex models are compared to the native complex and assessed for accuracy. Each model is assessed for accuracy by two measurements: orientation and interface accuracy, similar to CAPRI (Lensink, et al., 2007; Lensink and Wodak, 2010). Orientation accuracy (high, medium, acceptable or incorrect) is based on RMSD criteria (Supplementary Data), while interface accuracy is based on the fraction of correctly predicted interface residues (Supplementary Data). In line with previous docking papers, we define a near-native model as a model of high, medium, or acceptable accuracy. The success rate is the percentage of benchmark cases with at least one near-native model in the top N predictions (N is typically 10, referred to as top10).

3 RESULTS

3.1 Docking Benchmark Results

3.1.1 Docking Accuracy Increases Significantly for Individual Datasets. Integrative docking method shows 2-fold increase in the top10 success rate compared to standard docking (PatchDock-FireDock protocol) for SAXS and NMR-RTC, almost 3-fold increase for CXMS, and 4-fold increase for EM2D and EM3D (Table 1, Fig. 2a). The standard docking protocol succeeds to rank a near-native model in the top10 scoring models in 24% of benchmark cases. When SAXS data is used, this number goes up to 51%. If we consider only ~65 rigid body cases with less than 3% missing residues (unbound structures compared to complex), the success rate increases to 77% (Schneidman-Duhovny, et al., 2011). For EM2D and EM3D, the success rate is 82% and 79%, respectively. This success rate quadruples when compared to standard docking with the 19% success rate for the 27 complexes in the EM benchmark. For NMR-RTC, the success rate is 47%. With up to three cross-links, the success rate is 65%. If we consider the top scoring model, there is a 2-fold increase in the success rate for SAXS and NMR-RTC (22% and 18% vs. 10%), almost 4-fold increase for CXMS (36% vs. 10%), and almost 5-fold increase for EM2D and EM3D (33% vs. 7%).

Table 1. Success rate of integrative docking using individual experimental filters.

TopN	Standard docking	Standard docking EM cases	SAXS	EM2D	EM3D	NMR-RTC	CXMS
1	10%	7%	22%	33%	33%	18%	36%
10	**24%**	**19%**	**51%**	**82%**	**79%**	**47%**	**65%**
100	49%	26%	77%	89%	89%	76%	87%
Case #	176	27	176	27	27	176	138

While using any type of data significantly improves the results relative to standard docking, we are still far from the upper limit on the success rate, given the initial sampling by finer docking (97% of the benchmark cases have a near-native model sampled by a global search). When we allow for a near-native model in the top100 instead of top10 models, the success rate increases to 71-89%, depending on the data types. For the failing benchmark cases, the near-native model is usually among the top1000 models.

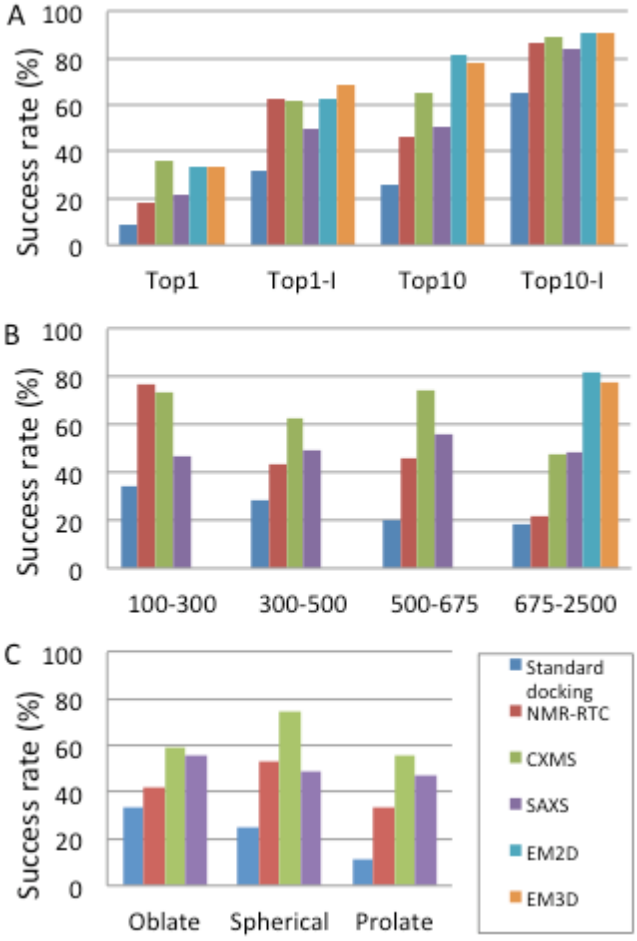

Fig. 2. Success rate of integrative docking for Benchmark 4.0. A) Success rate in prediction of orientation (top, top10) and binding site (top-I, top10-I) for standard docking and docking restrained by NMR-RTC, CXMS, SAXS, EM2D, and EM3D. B) Success rate for predicting a near-native model within the top10 models as a function of complex size for standard docking as well as docking restrained by NMR-RTC, CXMS, SAXS, EM2D, and EM3D. C) Success rate for predicting a near-native model within the top10 models as a function of complex shape for standard docking as well as docking restrained by NMR-RTC, CXMS, and SAXS.

The success rate depends on the difficulty of the benchmark cases, but there is a significant increase when compared to standard docking, independent of the difficulty (Table S2, S3). The success rate also increases when only high or medium accuracy models are considered as near-natives (Table S4).

*3.1.2 Interface Prediction Accuracy***.** We find that the top-scoring model has a correctly predicted interface (Fig. 2a, Supplementary Data) in 50-68% of cases compared to 32% for standard docking. While NMR-RTC performs worse than other data types in orientation prediction, the success rate in interface prediction is comparable to that for other data types. Based on these benchmark results,

the probability of a correctly predicted interface in the top-scoring model is 50-70% (depending on data type used) vs. 32% for standard docking and it increases to 84-91% if top10 models are considered.

3.1.3 Dependence of Success Rate on Complex Size. For each of the five data types, we test the dependence of the success rate on the complex size. The 176 benchmark complexes were divided into four groups according to complex size, with the fourth group corresponding to the 27 complexes in the EM benchmark (Fig. 2b). Varying data types are most informative and applicable for different complex sizes. In particular, the success rate of standard docking decreases with the increase in complex size from 34% for small complexes to 19% for the EM benchmark. The reason is that the number of configurations and flexibility increase with protein size. The success rate for NMR-RTC drops sharply for complexes with more than 300 residues (for complexes with more than 675 residues, there is no significant difference between standard docking and docking with NMR-RTC). The reason is that the number of potential interfaces (*ie*, the size of the search space) increases with protein size. In contrast, the success rate of SAXS is not sensitive to complex size. Unsurprisingly, the success rate for CXMS decreases slightly for the larger and more challenging complexes.

3.1.4 Dependence of Success Rate on Protein Shapes. Protein complexes were classified into oblate, spherical, and prolate based on the eigenvalues of the gyration tensor (Pons, et al., 2010) (Fig. 2c). The success rate of standard docking is highest for oblate proteins and lowest for prolate proteins. The reason is that oblate proteins have larger interfaces and better shape complementarity. The success rate of NMR-RTC, CXMS, and SAXS is not sensitive to protein shapes due to a combination of data and energy scores. The most significant increase in the success rate compared to standard docking is for prolate proteins: 3-fold increase for NMR-RTC, 5-fold increase for CXMS, and 4-fold increase for SAXS. No analysis was performed for EM, due to a small size of the EM benchmark.

3.1.5 Combining Different Experimental Datasets Increases the Success Rate. We tested pairwise combinations of the five experimental data types. The top10 success rate increases from 42-82% for individual data types to 63-82% for pairwise combinations (Table 2). More important is the increase in the top1 success rate from 17-36% to 26-52%. CXMS data complements all other data types, with most significant improvement for the top-scoring model, where the success rate increases from 36% for CXMS alone to 47-52% for all four pairwise data type combinations. The top10 success rate for CXMS combined with SAXS or NMR-RTC is 80% and 81% respectively, and is comparable to the success rate of EM data types. Another successful pairwise combination is SAXS - NMR-RTC, improving the success rate for the whole benchmark from 47-51% for SAXS and NMR-RTC separately to 68% when both data types are used. No significant improvement in the top10 success rate is obtained by combining EM (2D or 3D) with other data types, since their independent success rate is already high (79-82%). For the EM - NMR-RTC combinations, there is even a slight decrease in the success rate, because the NMR-RTC data is not informative for large protein complexes in the EM benchmark. When all five data types are combined, the top10 success rate is similar to that for EM (83%), but more important is the increase in the top1 success rate to 61% from 33% for EM alone.

Table 2. Success rate of integrative docking using combined experimental filters.

TopN	SAXS, EM2D	SAXS, EM3D	SAXS, NMR-RTC	SAXS, CXMS	EM2D, EM3D	
1	26%	41%	27%	**51%**	**44%**	
10	74%	74%	**68%**	**80%**	82%	
100	82%	82%	85%	91%	89%	
Case #	27	27	176	138	27	
TopN	EM2D, NMR-RTC	EM2D, CXMS	EM3D, NMR-RTC	EM3D, CXMS	NMR-RTC, CXMS	All
1	26%	**52%**	30%	**48%**	**47%**	**61%**
10	63%	83%	67%	74%	**81%**	83%
100	85%	87%	85%	91%	94%	83%
Case #	27	23	27	23	138	23

Increase in the success rate by more than 10% as compared to individual datasets is marked in bold.

3.2 Application to an Antibody-Antigen Complex

To test the applicability of the integrative method for determining pairwise protein complexes in a biopharmaceutical setting, we applied it to an antibody-antigen complex with experimentally generated datasets. In a typical biopharmaceutical discovery project, antibodies for a specific target can be generated by mice immunization or by phage-display libraries. The next step, is selecting an optimal antibody out of several candidates for further development into a drug. Knowledge of the epitope is an important factor in antibody selection process. Therefore, a method that can model antibody-antigen complexes rapidly and accurately would be extremely useful.

In our case, the antigen, PCSK9, plays a major regulatory role in cholesterol homeostasis and it is an important drug target (Horton, et al., 2007). PCSK9 binds to the EGF-A domain of the low-density lipoprotein receptor (LDLR) and induces LDLR degradation. Reduced LDLR levels result in decreased metabolism of low-density lipoproteins, which may lead to hypercholesterolemia. The antibody, J16, inhibits the action of PCSK9 by preventing LDLR binding (Liang, et al., 2011). Recently, a crystal structure of PCSK9 in complex with J16 Fab showed that J16 is a competitive inhibitor of LDLR binding (Liang, et al., 2011).

*3.2.1 Complex Structure Modeling***.** The atomic structure of the unbound PCSK9 has been available since the beginning of this study (Protein Data Bank code 2P4E) (Cunningham, et al., 2007). For the J16 Fab, 20 comparative models corresponding to two different elbow angles (136° and 168°) and 10 different CDR loop conformations were selected based on the fit to the J16 Fab SAXS profile (Supplementary Data). In addition, the missing loops, N-termini, C-termini, and His tags were added for PCSK9 and the J16 Fab with MODELLER-9v8, to better model the SAXS data. The integrative docking protocol was applied to PCSK9 and 20 J16 Fab models; the final clustering considered all complex models simultaneously.

The structure of the complex was determined by X-ray crystallography during the course of this project, but was made available to this study only after the model of the complex was computed. Therefore, this application corresponds to a real life antibody discovery scenario, where the unbound structures of the drug target and the antibody are known, but the structure of the complex is not available.

*3.2.2 Assessment Against X-ray Structure***.** Since the accuracy of a docking prediction highly depends on the accuracy of the input structures, we first assess the accuracy of our input structures. The C_α-RMSD between the bound and unbound PCSK9 structures is 1.4Å and between the bound and modeled J16 Fabs is 1.0Å and 3.0Å for elbow angles of 136° and 168°, respectively. Thus, there are no major PCSK9 conformational changes upon binding. The elbow angle of the J16 Fab in the complex X-ray structure is 137.6°. Therefore, the prediction of the elbow angle based on the Fab SAXS profile was correct (Fig. S1). We have also tested the fit of the X-ray structure of the complex against each data type (Fig. S2) and observed high quality fits for SAXS (χ of 2.24), EM2D (cross-correlation coefficient of 0.87), and EM3D (cross-correlation coefficient of 0.78).

Next, we analyze the accuracy of the best-scored models in terms of orientation and interface accuracy for different datasets. The best-scored models with acceptable accuracy were ranked 14, 2, 2, and 2 for SAXS, EM2D, EM3D, and all three datasets combined, respectively (Table S5). The best-scored models with a correct epitope were ranked 3, 2, 1, and 1, respectively (Table S5). Docking results are slightly better for models with the elbow angle of 136° than for 168°, with acceptable accuracy models ranked 5, 2, 1, and 2 for SAXS, EM2D, EM3D, and all three data types combined, respectively (Table S5).

3.2.3 Data-guided Filtering and Funnel Analysis. Ideally, the normalized fitting scores would correlate strongly with the accuracy of the model over a broad range of accuracy (ie, I-RMSD of 0-5Å or L-RMSD of 0-15Å). We now examine whether or not such a "funnel" exists for each type of data (London and Schueler-Furman, 2008) and how these funnels relate to specific complex structures (Fig. 3a). The three experimental datasets indeed result in pronounced funnels, revealing similar complex structures (Fig. 3b). Typically, there are three or four funnels associated with complex structures related by the pseudo-symmetry of the antibody (*ie*, light chain *vs* heavy chain) and the triangular shape symmetry of PCSK9 (Fig. 3c).

The SAXS dataset produces four funnels. One of them includes the near-native models, although this funnel is the least pronounced among the four funnels. The EM2D dataset produces three funnels with comparable scores. One of the funnels is centered close to the native structure, demonstrating the predictive power of EM2D. The EM3D dataset produces four funnels similar to those from the SAXS dataset. In contrast to SAXS, the funnel with near-native models has the best EM3D scores, although this funnel is not centered on the native complex structure (its center is ~11Å RMSD away from the native structure). While the EM3D dataset is best in selecting the correct funnel, the EM2D score is better in picking the highest accuracy model once the correct funnel has been selected. The shift in the near-native EM3D funnel relative to the near-native EM2D funnel can be explained by a distortion of the 3D density map that results from inaccuracies in the initial density map used for the 3D reconstruction that was obtained from the 2D class averages by the random conical tilt method.

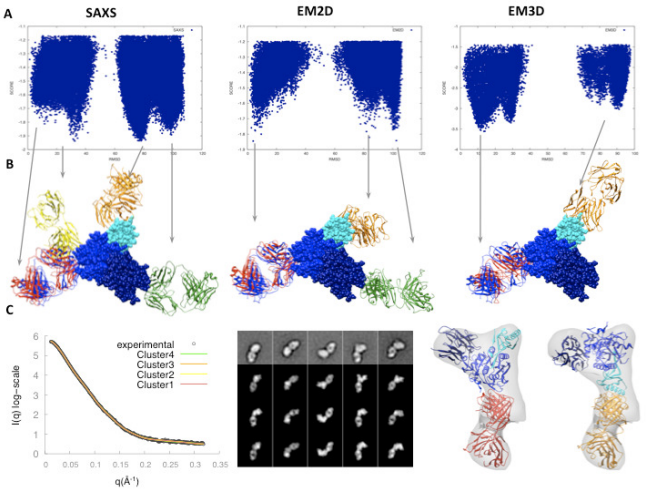

Fig. 3. Modeling of the PCSK9 – J16 Fab complex. A) Scoring funnels as a function of L-RMSD for different experimental filters. B) Top scoring cluster representatives (red, green, gold and yellow) for integrative docking with SAXS, EM2D, and EM3D filters, superimposed on X-ray crystallographic structure (blue). The models are superimposed on PCSK9 (prodomain in cyan, catalytic domain in blue, and C-terminal domain in dark blue). C) Fit of the top scoring cluster representatives to the SAXS profile, EM2D class averages, and EM3D density map.

4 DISCUSSION

We developed an integrative method for docking two protein structures by combining protein docking techniques with data from five experimental methods including SAXS, EM2D, EM3D, NMR-RTC, and CXMS. To assess the accuracy of the integrative method, we used a benchmark of 176 complex structures with simulated experimental data. We also applied the method to an antibody-antigen complex, relying on experimental datasets collected specifically for this study.

Additional information, such as sequence conservation and impact of site-directed mutagenesis on complex formation, has been used previously to increase the accuracy of pairwise protein docking (Lensink and Wodak, 2010; Mashiach, et al., 2010). Here, we analyze the improved docking success rates afforded by data from five accessible experimental methods. Our integrative framework can be modularly extended to support additional types of experimental data, such as those from footprinting, site-directed mutagenesis, FRET spectroscopy, and atomic force microscopy (Trinh, et al., 2012). In addition to the data types tested here, binding site residues and distance constraints, if available, can be added directly to the PatchDock input. In principle, experimental datasets can be used either to filter docking models or directly to drive the sampling. We select the first approach, because it allows seamless integration of any combination of datasets and we can rely on efficient global search methods already developed for pairwise protein docking. Moreover, driving the docking with global shape data, such as SAXS and EM2D, is algorithmically challenging.

4.1 Experimental Datasets and Their Impact on Docking

The experimental methods were chosen for their utility in a biotherapeutics discovery context where multiple artificial binding proteins (such as antibodies) are engineered to bind a specific drug target and rapid tools for epitope prediction are required. According to our large benchmark analysis, EM (2D and 3D) are the most informative of all datasets (Table 1). However, collecting experimental information to generate a 3D map is generally possible only for complexes larger than approximately 100 kDa and requires a relatively large amount of work. Here, we show that 2D class averages, which can be obtained significantly faster and for a wider range of samples, can provide the same information for pairwise docking as 3D density maps. In contrast to EM, SAXS has the advantage of being able to collect and analyze multiple samples in a few hours. As a result, purification, SAXS data collection, and docking of multiple antibodies binding the same target could be performed in a matter of days. Automation of collecting EM data is more challenging than that for SAXS, although recent advances in data acquisition and an increase in computing power have allowed more streamlined processes in single particle EM (Lyumkis, et al.; Wu, et al., 2012). While cross-linking with mass spectrometry is informative on its own, it also complements all other data types. With recent advances in data collection (Rappsilber, 2011), it is becoming a method of choice for combination with shape informative methods, such as SAXS and EM.

While validation of integrative docking by a large benchmark using simulated data has allowed a robust statistical analysis (Fig. 2), data collection and application to a specific target with real data has highlighted advantages and challenges of the integrative docking approach. Unlike NMR-RTC, which depends on protein expression in a cell-free expression system, both SAXS and EM gave useful data for the PCSK9-J16 Fab complex. In general, larger size and higher symmetry of a complex simplify EM data acquisition and interpretation. The larger mass of an IgG (150 kDa) compared to a Fab fragment (50 kDa) would simplify the data acquisition and image processing. However, the flexibility of an IgG may result in a conformationally heterogenous complex sample, favoring the use of the more rigid Fab fragment. While the EM3D data was most informative for identifying the near-native cluster of models and predicting the epitope, more accurate structural models could be selected by the EM2D score. Despite the relatively low information content of the SAXS profile, the SAXS score predicted the same clusters as the EM-based scores. Additionally, the J16 Fab SAXS profile was useful in predicting the Fab structure and its elbow angle.

4.2 Improvement Compared to Standard Docking

While integrative protocol succeeds in including a near-native model among the top10 models in 42-82% of the cases (Table 1), state-of-the-art docking methods succeed only in 30-40% of the cases, depending on the benchmark and accuracy criterion. ZDOCK-ZRANK ranks a model with I-RMSD < 4.0Å among the top10 models in 35-40% of the rigid-body cases of Benchmark 2.0 (Pierce and Weng, 2008). Recently developed residue potential, SIPPER (Pons, et al., 2011), succeeds to rank a model with L-RMSD < 10Å in 28% of the 81 Benchmark 3.0 complexes, where at least one model with L-RMSD < 10Å was generated by FTDock. In a recent CAPRI evaluation, an acceptable accuracy model was submitted by at least one participating group for 11 out of 13 complexes (Lensink and Wodak, 2010). However, top 8 predictors could only predict correctly 6 out of 13 complex structures. While predictors can use additional information manually, a fully automated method, ClusPro (Comeau, et al., 2004), succeeded to predict correctly 5 targets.

4.3 Comparison to Other Hybrid Docking Methods

Docking has been previously combined with additional data. HADDOCK (Dominguez, et al., 2003) benefits from a consensus interface predictor CPORT (de Vries and Bonvin, 2011), succeeding to rank an acceptable accuracy model among the top10 models for ~19% of the Benchmark 2.0 complexes. pyDockSAXS (Pons, et al., 2010), which combines FTDock sampling with the pyDock scoring function and a SAXS profile, succeeds to rank a model with L-RMSD < 10Å in 43% of the Benchmark 2.0 complexes (*ie*, for 70 of the 84 complexes with similar molecular mass for bound and unbound structures). In comparison, our approach applied to the same benchmark with a SAXS profile only, results in the significantly increased 63% success rate. The increase in the success rate is due to the increased precision of configurational sampling and higher accuracy of interface energy score in FireDock (Tables S6, S7). The integrative approach had a similar performance for three experimental SAXS dataset from pyDockSAXS benchmark (Niemann, et al., 2008; Pinotsis, et al., 2008; Pons, et al., 2010; Schubert, et al., 2002) (Table S7).

4.4 Sampling *versus* Scoring

The current work highlights the challenges in protein docking. Sufficient sampling is required in the global search stage to maximize model accuracy, hit rate, and the quality-of-fit to the experimental data (Table S1, S6). Our integrative approach is designed to benefit from the increasingly focused molecular docking search space afforded by consideration of experimental data. While an acceptable accuracy model is contained among the ~200,000 models generated by a global search for 97% of benchmark complexes, our integrative protocol succeeds to rank such models among the top10 scoring models only in 42% to 82% of the test cases, depending on the data used (Tables 1, 2); correct binding sites are identified among the top10 scoring models in 84% to 91% of the cases (Fig. 2a). We suggest that a combination of finer sampling methods (including flexible docking) and improved scoring functions with physico-chemical and/or statistical terms can be helpful for further improving the success rate of pairwise protein docking. Integrative docking, such as that described here, may provide the best compromise between the relative expediency and inaccuracy of standard docking on one hand and relative complexity and accuracy of experimental structure determination by X-ray crystallography or NMR spectroscopy on the other hand.

Software. The package is downloadable from http://salilab.org/idock. SAXS, EM2D, EM3D, NMR-RTC, and CXMS scoring functions are implemented in our open source Integrative Modeling Platform (IMP; http://salilab.org/imp). PatchDock and FireDock are available at http://bioinfo3d.cs.tau.ac.il. Docking with a SAXS profile can also be done *via* a webserver at http://salilab.org/foxsdock.

ACKNOWLEDGEMENTS

We wish to thank Javier Chaparro-Riggers for engineering J16 Fab and Michael Chin for the PCSK9 and J16 Fab samples. We thank Carles Pons, Juan Fernandez-Recio, and Dmitri Svergun for kindly providing access to the experimental SAXS datasets.

Funding: DSD has been funded by the Weizmann Institute Advancing Women in Science Postdoctoral Fellowship. We also acknowledge support from NIH R01 GM083960, NIH U54 RR022220, NIH PN2 EY016525, and Rinat (Pfizer) Inc. The SIBYLS beamline at Lawrence Berkeley National Laboratory is supported by the DOE program Integrated Diffraction Analysis Technologies (IDAT). We are also grateful for computer hardware gifts from Ron Conway, Mike Homer, Intel, Hewlett-Packard, IBM, and NetApp.

REFERENCES

Alber, F., *et al.* (2007) Determining the architectures of macromolecular assemblies, *Nature*, **450**, 683-694.

Alber, F., *et al.* (2007) The molecular architecture of the nuclear pore complex, *Nature*, **450**, 695-701.

Alber, F., *et al.* (2008) Integrating diverse data for structure determination of macromolecular assemblies, *Annu Rev Biochem*, **77**, 443-477.

Andrusier, N., Nussinov, R. and Wolfson, H.J. (2007) FireDock: fast interaction refinement in molecular docking, *Proteins*, **69**, 139-159.

Comeau, S.R., *et al.* (2004) ClusPro: an automated docking and discrimination method for the prediction of protein complexes, *Bioinformatics*, **20**, 45-50.

Cunningham, D., *et al.* (2007) Structural and biophysical studies of PCSK9 and its mutants linked to familial hypercholesterolemia, *Nat Struct Mol Biol*, **14**, 413-419.

de Vries, S.J. and Bonvin, A.M. (2011) CPORT: a consensus interface predictor and its performance in prediction-driven docking with HADDOCK, *PLoS ONE*, **6**, e17695.

Dominguez, C., Boelens, R. and Bonvin, A.M. (2003) HADDOCK: a protein-protein docking approach based on biochemical or biophysical information, *J Am Chem Soc*, **125**, 1731-1737.

Duhovny, D., Nussinov, R. and Wolfson, H.J. (2002) Efficient Unbound Docking of Rigid Molecules. In Guigó, R. and Gusfield, D. (eds), *Second International Workshop, WABI 2002*. Springer Berlin / Heidelberg, Rome, Italy, pp. 185-200.

Eisenstein, M. and Katchalski-Katzir, E. (2004) On proteins, grids, correlations, and docking, *C R Biol*, **327**, 409-420.

Fernandez-Recio, J., Totrov, M. and Abagyan, R. (2003) ICM-DISCO docking by global energy optimization with fully flexible side-chains, *Proteins*, **52**, 113-117.

Gray, J.J., *et al.* (2003) Protein-protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations, *J Mol Biol*, **331**, 281-299.

Horton, J.D., Cohen, J.C. and Hobbs, H.H. (2007) Molecular biology of PCSK9: its role in LDL metabolism, *Trends Biochem Sci*, **32**, 71-77.

Hura, G.L., *et al.* (2009) Robust, high-throughput solution structural analyses by small angle X-ray scattering (SAXS), *Nat Methods*, **6**, 606-612.

Hwang, H., *et al.* (2010) Protein-protein docking benchmark version 4.0, *Proteins*, **78**, 3111-3114.

Krogan, N.J., *et al.* (2006) Global landscape of protein complexes in the yeast Saccharomyces cerevisiae, *Nature*, **440**, 637-643.

Lensink, M.F., Mendez, R. and Wodak, S.J. (2007) Docking and scoring protein complexes: CAPRI 3rd Edition, *Proteins*, **69**, 704-718.

Lensink, M.F. and Wodak, S.J. (2010) Blind predictions of protein interfaces by docking calculations in CAPRI, *Proteins*, **78**, 3085-3095.

Lensink, M.F. and Wodak, S.J. (2010) Docking and scoring protein interactions: CAPRI 2009, *Proteins*, **78**, 3073-3084.

Liang, H., *et al.* (2011) PCSK9 Antagonism Reduces LDL-cholesterol in Statin-treated Hypercholesterolemic Non-human Primates, *J Pharmacol Exp Ther*.

London, N. and Schueler-Furman, O. (2008) Funnel hunting in a rough terrain: learning and discriminating native energy funnels, *Structure*, **16**, 269-279.

Lyumkis, D., *et al.* Automation in single-particle electron microscopy connecting the pieces, *Methods Enzymol*, **483**, 291-338.

Mashiach, E., Nussinov, R. and Wolfson, H.J. (2010) FiberDock: Flexible induced-fit backbone refinement in molecular docking, *Proteins*, **78**, 1503-1519.

Mashiach, E., *et al.* (2010) An integrated suite of fast docking algorithms, *Proteins*, **78**, 3197-3204.

Niemann, H.H., *et al.* (2008) X-ray and neutron small-angle scattering analysis of the complex formed by the Met receptor and the Listeria monocytogenes invasion protein InlB, *Journal of Molecular Biology*, **377**, 489-500.

Pierce, B. and Weng, Z. (2008) A combination of rescoring and refinement significantly improves protein docking performance, *Proteins*, **72**, 270-279.

Pinotsis, N., *et al.* (2008) Molecular basis of the C-terminal tail-to-tail assembly of the sarcomeric filament protein myomesin, *The EMBO journal*, **27**, 253-264.

Pons, C., *et al.* (2010) Structural characterization of protein-protein complexes by integrating computational docking with small-angle scattering data, *J Mol Biol*, **403**, 217-230.

Pons, C., *et al.* (2011) Scoring by intermolecular pairwise propensities of exposed residues (SIPPER): a new efficient potential for protein-protein docking, *J Chem Inf Model*, **51**, 370-377.

Rappsilber, J. (2011) The beginning of a beautiful friendship: cross-linking/mass spectrometry and modelling of proteins and multi-protein complexes, *Journal of structural biology*, **173**, 530-540.

Reese, M.L. and Dötsch, V. (2003) Fast mapping of protein-protein interfaces by NMR spectroscopy, *J Am Chem Soc*, **125**, 14250-14251.

Ritchie, D.W. (2008) Recent progress and future directions in protein-protein docking, *Curr Protein Pept Sci*, **9**, 1-15.

Robinson, C.V., Sali, A. and Baumeister, W. (2007) The molecular sociology of the cell, *Nature*, **450**, 973-982.

Russel, D., *et al.* (2012) Putting the pieces together: integrative modeling platform software for structure determination of macromolecular assemblies, *PLoS biology*, **10**, e1001244.

Schneidman-Duhovny, D., Hammel, M. and Sali, A. (2011) Macromolecular docking restrained by a small angle X-ray scattering profile, *J Struct Biol*, **173**, 461-471.

Schneidman-Duhovny, D., *et al.* (2005) Geometry-based flexible and symmetric protein docking, *Proteins*, **60**, 224-231.

Schubert, W.D., *et al.* (2002) Structure of internalin, a major invasion protein of Listeria monocytogenes, in complex with its human receptor E-cadherin, *Cell*, **111**, 825-836.

Sivasubramanian, A., *et al.* (2006) Structural model of the mAb 806-EGFR complex using computational docking followed by computational and experimental mutagenesis, *Structure*, **14**, 401-414.

Stahlberg, H. and Walz, T. (2008) Molecular electron microscopy: state of the art and current challenges, *ACS Chem Biol*, **3**, 268-281.

Trinh, M.H., *et al.* (2012) Computational reconstruction of multidomain proteins using atomic force microscopy data, *Structure*, **20**, 113-120.

Vajda, S. and Kozakov, D. (2009) Convergence and combination of methods in protein-protein docking, *Curr Opin Struct Biol*, **19**, 164-170.

Wu, S., *et al.* (2012) Fabs Enable Single Particle cryoEM Studies of Small Proteins, *Structure*, **20**, 582-592.