

# Integrative Structure Modeling of Macromolecular Assemblies from Proteomics Data\*

Keren Lasker†§¶||, Jeremy L. Phillips†¶\*\*††, Daniel Russell†, Javier Velázquez-Muriel†, Dina Schneidman-Duhovny†§§, Elina Tjioe†, Ben Webb†, Avner Schlessinger†, and Andrej Sali†¶¶

**Proteomics techniques have been used to generate comprehensive lists of protein interactions in a number of species. However, relatively little is known about how these interactions result in functional multiprotein complexes. This gap can be bridged by combining data from proteomics experiments with data from established structure determination techniques. Correspondingly, integrative computational methods are being developed to provide descriptions of protein complexes at varying levels of accuracy and resolution, ranging from complex compositions to detailed atomic structures. *Molecular & Cellular Proteomics* 9:1689–1702, 2010.**

**A 3-D enhanced version of this article is available. The text is identical to this version but includes interactive figures.**

**Viewing the enhanced version of this article requires the use of a browser plug-in. Please install the plug-in when prompted. <http://www.thesgc.org/iSee/MCP/9/8/e2.html>**

## MOTIVATION: STRUCTURES FOR MECHANISTIC UNDERSTANDING OF PROCESSES

The cell contains hundreds of functional macromolecular assemblies responsible for performing critical cellular processes (1, 2). These include, among others, the ribosome (translation) (3, 4), chaperonins (protein folding) (5, 6), RNA polymerase (RNA synthesis) (7), and the proteasome (protein degradation) (8–10). A macromolecular machine is often built around a stable core of proteins that defines the basic function of the complex. This core assembly can be modulated through interactions with peripheral protein components, resulting in a multitude of functionally relevant states (11). A structural description of an assembly in all of its states often facilitates a mechanistic understanding of the corresponding process (3, 12, 13). Thus, a critical challenge in structural

biology is to identify biologically relevant states of macromolecular assemblies and to determine the structures of these states at the highest possible resolution.

## ASSEMBLY STRUCTURES OFTEN CANNOT BE RESOLVED BY A SINGLE TECHNIQUE

The structures of macromolecular assemblies in their biologically significant states generally cannot be resolved to atomic resolution by a single technique (14). Although x-ray crystallography remains the most powerful approach for visualizing a static snapshot of a complex at atomic resolution, it is limited to samples that can be purified in large quantities and crystallized (15). Similarly, NMR spectroscopy results in an ensemble of structures of a system in solution (16–18), but the technique is limited by the size of the complex and sample availability. Electron microscopy (EM)<sup>1</sup> techniques provide an alternative approach for visualizing multiple conformations of complexes *in vitro* and even within cells (19–22). However, in most cases, the resolution of an electron density map is too low to provide a full mechanistic description of a protein complex. Additional techniques, such as high throughput proteomics methods (23), small angle x-ray scattering (SAXS) (24, 25), and fluorescence resonance energy transfer (FRET) spectroscopy (26), are generally limited by low resolution (14) and at times by low accuracy (27–29) of the corresponding structural information.

## INTEGRATIVE STRUCTURE DETERMINATION

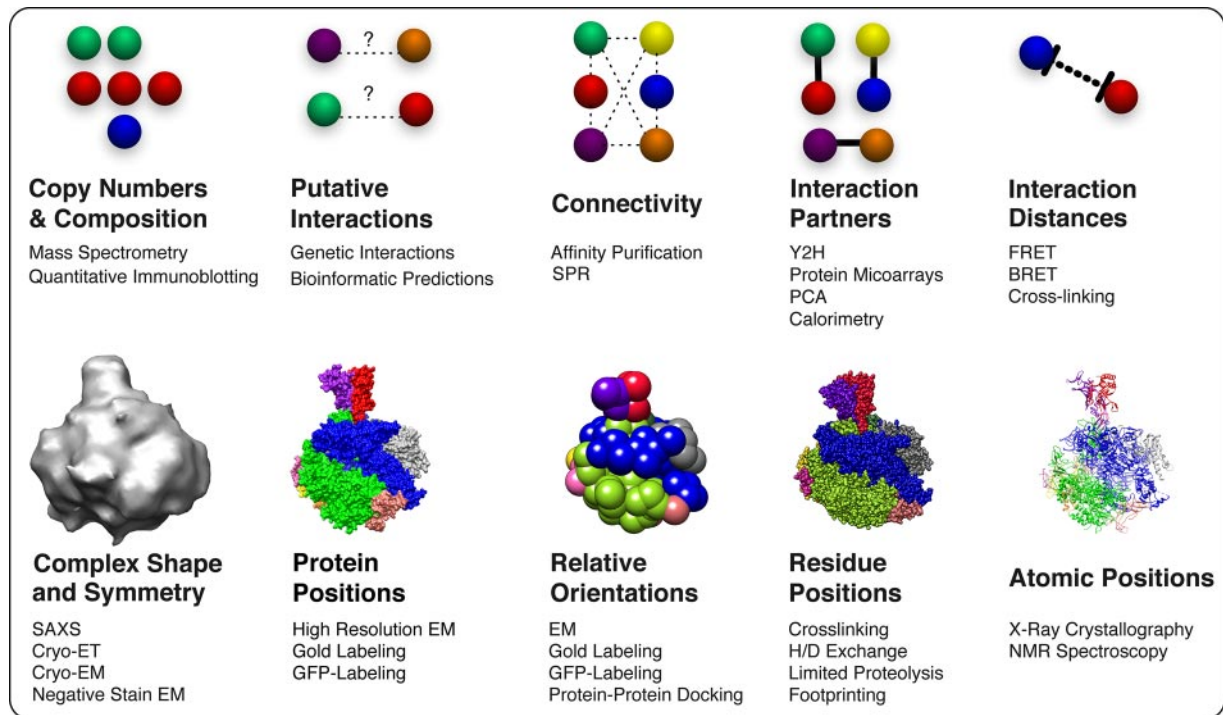
The limitations in the resolution, accuracy, and coverage of individual experimental methods can be bridged by simultaneous consideration of multiple types of information. Examples of techniques that specialize in integrating a few types of experimental data include (i) combining electron density maps of complexes with atomic structures of protein components to build high resolution structures of protein complexes (30–34); (ii) using atomic models to estimate the phases required for converting diffraction data into electron density maps (35); (iii) inferring the binary interaction map of a complex from

From the †Department of Bioengineering and Therapeutic Sciences, Department of Pharmaceutical Chemistry, and California Institute for Quantitative Biosciences (QB3) and the \*\*Graduate Group in Biological and Medical Informatics, University of California, San Francisco, California 94158 and §Blavatnik School of Computer Science, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, 69978 Tel Aviv, Israel

Received, May 5, 2010

Published, MCP Papers in Press, May 27, 2010, DOI 10.1074/mcp.R110.000067

<sup>1</sup> The abbreviations used are: EM, electron microscopy; FRET, fluorescence resonance energy transfer; SAXS, small angle x-ray scattering; NPC, nuclear pore complex; AAA-ATPase, adenosine triphosphatase associated with diverse cellular activities; RNAPII, RNA polymerase II; Rpb, RNA polymerase II subunit; H-RNAPII, human RNA polymerase II; Y2H, yeast two-hybrid; RMSD, root mean square deviation.



**FIG. 1. Structural information about a protein assembly.** Standard proteomics, biophysical, and computational methods can collectively determine the copy numbers (stoichiometry) and types (composition) of assembly components and predict or experimentally determine protein-protein connectivities (interactivity among a group of proteins) and protein-protein interactions (direct physical interactions). Many of these techniques are capable of a high degree of throughput, allowing for collection of a high volume of data about components of an assembly in a short period of time. Additional biophysical methods can determine distances between components in an assembly, positions of the components, and their relative orientations. Integration of data from varied methods, including low resolution proteomics data, generally increases the accuracy, precision, coverage, and efficiency of structure determination. Methods listed include the following: mass spectrometry (124–126), quantitative immunoblotting (127), genetic interactions (128, 129), bioinformatics predictions of protein-protein interactions (130), affinity purification (13, 39, 71, 72), surface plasmon resonance (*SPR*) (131), Y2H (111–116), protein microarrays (132–134), protein-fragment complementation assay (*PCA*) (135, 136), calorimetry (137, 138), FRET (139), bioluminescence resonance energy transfer (*BRET*) (140), SAXS (24, 25), electron tomography (*ET*) (21), EM (19, 20, 22), gold labeling (39, 141, 142), green fluorescent protein (*GFP*) labeling (143), protein-protein docking (144), cross-linking (36, 43, 145, 146), hydrogen/deuterium (*H/D*) (147), limited proteolysis (148), footprinting (149), x-ray crystallography (15), and NMR spectroscopy (16–18).

affinity purification, mass spectrometry, and comparative modeling data (36); and (iv) incorporating NMR-derived data into protein structure prediction (37, 38).

Recently, a number of macromolecular structures have been resolved by such integrative methods. For instance, the constituent proteins in the nuclear pore complex (NPC) were localized based on the shape and symmetry of the NPC from cryo-EM, positions of the proteins from immuno-EM, relative proximities of proteins from affinity purification, and the shapes of proteins from ultracentrifugation (13, 39). An atomic model of the AAA-ATPase ring of the 26 S proteasome was determined primarily by fitting comparative models of subunits into a single-particle cryo-EM map subject to protein interactions identified by proteomics (40). A structural model for a complete clathrin lattice (41) and a mechanistic model of the clathrin lattice assembly-disassembly cycle driven by chaperone Hsc70 (42) were suggested by combining data obtained by x-ray crystallography and single-particle cryo-EM. The architecture of RNA polymerase II in complex with its

initiation factors was determined by combining known crystal structures with data from chemical cross-linking coupled to mass spectrometry (43). An NMR solution structure for the interface between two subunits in the human immunodeficiency virus type 1 capsid was fitted to an electron density map of the whole complex, revealing a relative orientation of subunits different from that in the corresponding crystal structure (44).

#### UNIFIED APPROACH FOR INTEGRATIVE MODELING

As outlined above, different studies on different systems will have a variety of different types of available data (Fig. 1 and Table I). Therefore, a unified approach for integrative modeling that can incorporate any type of information about a macromolecular assembly into the determination of its structure is needed. This information may include physical theories, statistical preferences extracted from biological databases, and heterogeneous experimental data at different resolutions, ranging from atomic structures to sets of inter-

TABLE I  
Common restraints that can be used for integrative structure determination

Restraint	Description	Source of information
Excluded volume restraint <sup>a</sup>	Prevents steric clashes between system particles	Physical first principles
Geometric complementarity restraint <sup>a</sup>	Restrains a protein interface to the tightest possible packing	Physical first principles
Statistical potential restraint	Restrains a structure to have contact frequencies similar to those in structurally defined complexes	Physical first principles, all previously determined protein structures
Distance restraint <sup>a</sup>	Restrains the distance between two particles	FRET, BRET, cross-linking, homology to a known structure
Protein localization restraint	Restrains a protein to a specific position	Immuno-EM, gold labeling, GFP labeling
Protein connectivity restraint <sup>a</sup>	Restrains all proteins in a set to interact directly or indirectly	Affinity purification
Angle restraint	Restrains the angle between three particles	EM, SAXS, homology to a known structure
Complex diameter restraint	Restrains the distance between the two most distant particles in a protein or complex	EM, SAXS
Symmetry restraint	Maintains the same configuration of equivalent particles across multiple symmetry units	EM, SAXS, homology to a known structure
EM quality-of-fit restraint <sup>a</sup>	Restrains the model to overlap with a density map	EM, SAXS
Radial distribution function restraint	Restrains the correlation between experimentally measured and computed radial distribution functions	SAXS

<sup>a</sup> Restraints used to determine the structure of H-RNAPII.

acting proteins. We have proposed a single unified approach that can leverage all information to describe a macromolecular structure (14, 39, 45). This approach consists of an iterative series of four steps, including 1) generation of data informative about the structure being determined, 2) design of system representation and translation of the data into spatial restraints, 3) calculation of an ensemble of structures that satisfy the spatial restraints, and 4) an analysis of the ensemble. In this procedure, spatial restraints derived from data about the structure are summed into a scoring function that assesses how well a structural model of an assembly agrees with the data. The scoring function is used to optimize the structural models and to generate a final ensemble of solutions that agrees with the data as much as possible. This four-step approach, by design, benefits from synergy among the input data sets, minimizing the drawback of incomplete, inaccurate, and/or imprecise data sets; although each individual restraint may contain little structural information, the concurrent satisfaction of all restraints derived from independent experiments may drastically reduce the degeneracy of the final structural models.

#### PROTEOMICS AS A KEY DATA SOURCE FOR INTEGRATIVE MODELING

Proteomics techniques have emerged as a powerful tool for mapping protein interactions in the cell. However, data produced by these techniques are rarely formally incorporated into macromolecular structure determination efforts. Here, we focus on the potential of proteomics techniques to contribute to the integrative modeling of macromolecular assemblies. Specifically, we describe how protein binding and association data can be interpreted as spatial restraints on a protein complex and thus reduce ambiguity in its structural description. These ideas have already been applied to determine the molecular architecture of the NPC

(13, 39) and a pseudoatomic model of the 20 S/AAA-ATPase ring of the 26 S proteasome (10, 40, 46). Below, we illustrate our integrative modeling approach by using real experimental data to determine the known architecture of the human RNA polymerase II.

#### INTEGRATIVE STRUCTURE CHARACTERIZATION OF HUMAN RNA POLYMERASE II (RNAPII)

The eukaryotic RNAPII is a central multiprotein machine that synthesizes messenger RNAs and small nuclear RNAs. It is composed of 12 protein subunits with a total molecular mass of 514 kDa (Fig. 2). Ten subunits (Rpb1, Rpb2, Rpb3, Rpb5, Rpb6, Rpb8, Rpb10, Rpb11, and Rpb12) form a structurally conserved core, whereas the Rpb4-Rpb7 heterodimer is located on the periphery (47, 48). Although the atomic structure of the *Saccharomyces cerevisiae* RNAPII has been solved by x-ray crystallography (49), the human RNAPII (H-RNAPII) has not been determined at atomic resolution mostly because of difficulties in obtaining sufficient quantities of pure sample (50). However, the molecular architecture of the H-RNAPII can be informed by that of its yeast homolog based on the homology between their constituent proteins (50).

Below, we demonstrate that our integrative structure determination procedure can be used to accurately model the known architecture of H-RNAPII using only proteomics-derived protein interactions, an electron density map at 20-Å resolution, comparative models of the protein subunits based on yeast and human crystallographic structures, and geometric complementarity between the interacting subunits. We describe the input data used for the modeling, the translation of these data into spatial restraints, an optimization procedure for determining the models that satisfy the restraints, and an analysis of the resulting set of solutions. We use a previously determined crystallographic structure of the full complex in yeast (51) to evaluate the results.

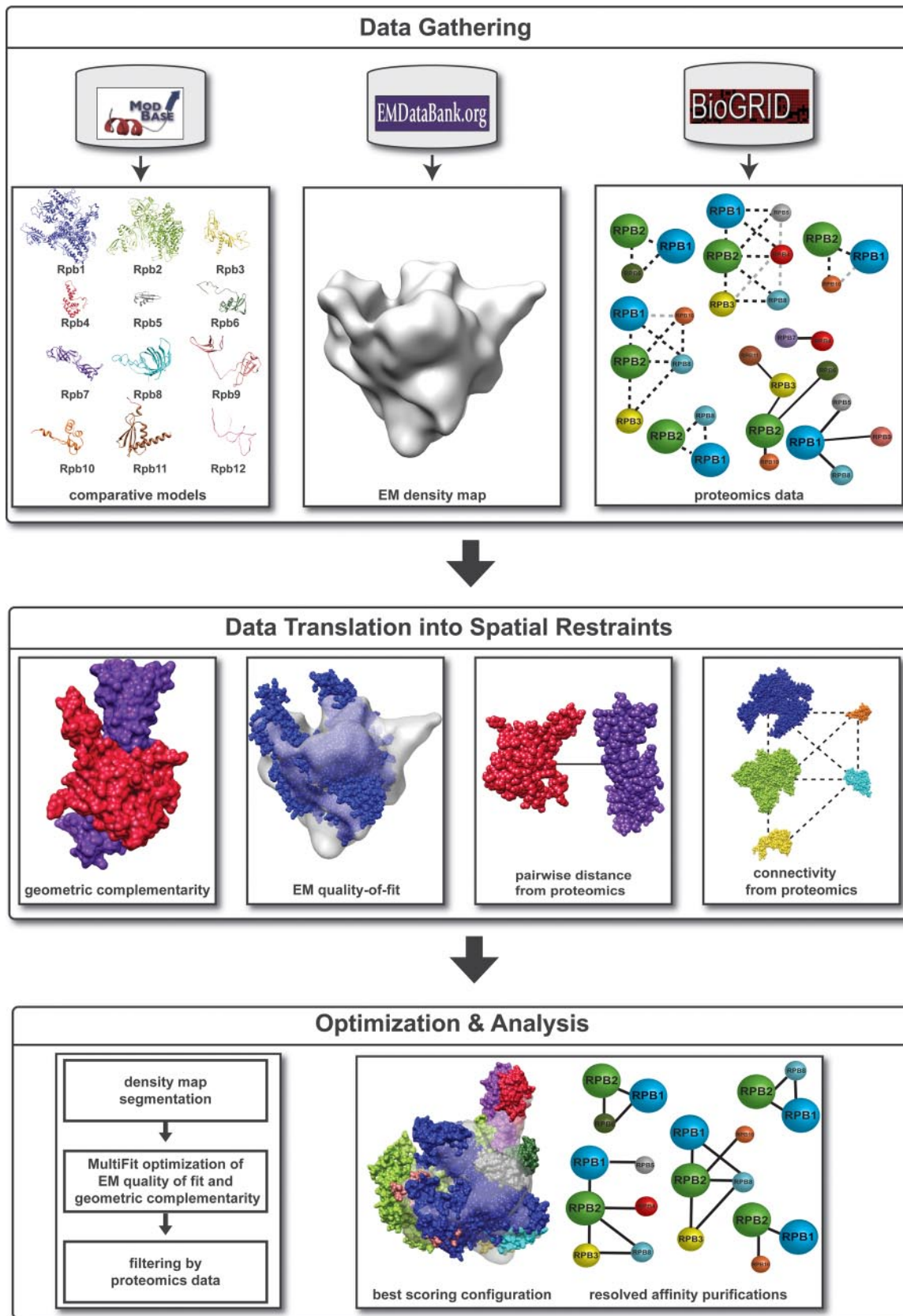


TABLE II  
Representation of H-RNAPII

Subunit (name, UniProt accession no.)	Sequence identity <sup>a</sup> (%)	Number of residues, residue range	Template (Protein Data Bank code and chain, residue range)
Rpb1, P24928	55	1970, 11–1475	1i6h A, 7–1445
Rpb2, P30876	63	1174, 15–1171	2vum B, 20–1216
Rpb3, P19387	47	275, 7–264	1twf C, 6–263
Rpb4, O15514	100	142, 14–142	2c35 A, 14–142
Rpb5, P1938	52	210, 146–209	1hmj A, 10–73
Rpb6, P61218	100	127, 1–127	1qkl A, 801–927
Rpb7, P62487	100	172, 1–171	2c35 B, 1–171
RPB8, P52434	100	150, 1–150	2f3i A, 1–150
Rpb9, P36954	47	125, 15–124	1twf I, 5–111
Rpb10, P62875	73	67, 1–64	1twf J, 1–65
Rpb11, P52435	52	1–105, 117	1twf K, 1–105
Rpb12, P53803	38	1–70, 70	2e2h L, 70

<sup>a</sup> The percentage of sequence identity between the target subunit and the template as calculated from their alignment used for comparative modeling.

### Data Generation by Experiments

Different techniques produce data that differ in types of measured features as well as in the accuracy, resolution, and coverage of the measurements (Fig. 1). An interpretation of the data in terms of a spatial restraint involves identifying the restrained structural components and the allowed values of the restrained feature implied by the data. For example, a result of a cross-linking experiment might be used to restrain the distance between two proteins (40, 52) or within one protein (53); the restraint parameters are a function of the length and flexibility of the cross-linker.

To determine the molecular architecture of the H-RNAPII, we use structural homologs of individual human protein subunits found in the ModBase database (54) (Table II), proteomics data for yeast RNAPII subunits extracted from the BioGRID database (55) (Table III), and an assembly electron density map of H-RNAPII determined at 20-Å resolution by single-particle cryo-EM (50) deposited in the EM data bank (56).

### System Representation

The first step in integrative structure determination is deciding on an appropriate representation for the system to be modeled as dictated by the resolution of the available data. At

TABLE III  
Proteomics data used for modeling the architecture of RNAPII

Interacting subunits	Source method
Rpb1, Rpb2, Rpb10	Affinity capture-MS (72)
Rpb1, Rpb2, Rpb3, Rpb4, Rpb5, Rpb8	Affinity capture-MS (72)
Rpb1, Rpb2, Rpb8	Affinity capture-MS (72)
Rpb1, Rpb2, Rpb3, Rpb8, Rpb10	Affinity capture-MS (72)
Rpb1, Rpb2, Rpb6	Affinity capture-MS (72)
Rpb1, Rpb5	Y2H (58, 59)
Rpb1, Rpb8	Y2H (60)
Rpb1, Rpb9	Y2H (58)
Rpb2, Rpb3	PCA (67)
Rpb2, Rpb6	PCA (67)
Rpb2, Rpb10	Y2H (58)
Rpb3, Rpb11	Y2H (58), reconstituted complex (65), PCA (67)
Rpb4, Rpb7	Y2H (61–66), reconstituted complex (69)

the finest representation granularity, an assembly structure can be represented by particles corresponding to its atoms, each associated with attributes such as position, radius, charge, and mass. Alternatively, a single-particle may be a sphere corresponding to a group of atoms, a whole amino acid residue, a secondary structure segment, a domain, a protein, a “subcomplex” consisting of a subset of proteins in a complete assembly, or even an entire assembly. Given the availability of high accuracy comparative models for the H-RNAPII subunits, we represent the structures of its subunits at atomic resolution. We use atomic models found in the ModBase database of comparative models for domains in ~2.4 million protein sequences that are detectably related to known structures (Table II) (57).

### Translation of Data into Spatial Restraints

A restraint is a function that reaches its minimum if the restrained feature (e.g. distance) is consistent with the data on which the restraint is based. Beyond that, a restraint can, in principle, have any functional form. For example, a restraint is frequently a harmonic function (of the form  $k \cdot x^2$  where  $x$  is the distance from the mean and  $k$  is proportional to the force constant) of the restrained feature. A restrained feature may be any structural attribute of a protein or assembly, including

FIG. 2. **Determining the molecular architecture of human RNAPII.** *Top*, data gathering. Comparative models of the H-RNAPII subunits were obtained from the ModBase database (54). A density map of H-RNAPII at 20-Å resolution (50) was obtained from the EM data bank (56). Proteomics data for *S. cerevisiae* RNAPII subunits were obtained from BioGRID (Table III) (55). All pairwise direct interactions are visualized in a single graph with *solid edges*, and each pull-down experiment is presented as a separate graph with *dashed edges* to indicate the missing underlying binary interaction network. Pull-downs Rpb1-Rpb2-Rpb3-Rpb4-Rpb5-Rpb8 and Rpb1-Rpb2-Rpb3-Rpb8-Rpb10 are missing some edges for clarity. *Gray edges* indicate interactions present in BioGRID but not in the yeast RNAPII crystallographic structure. *Middle*, scoring. The scoring function is the sum of the distance (illustrated between Rpb4 and Rpb7), connectivity (illustrated between Rpb1, Rpb2, Rpb3, Rpb8 and Rpb10), EM quality-of-fit (illustrated between the H-RNAPII density map and Rpb1), and geometric complementarity (illustrated between Rpb4 and Rpb7) restraints. *Bottom*, optimization. The configuration of the subunits in H-RNAPII was optimized using an extension of the divide-and-conquer MultiFit protocol to incorporate proteomics-based restraints. The optimization procedure resulted in a single model that satisfied all of the input restraints.

contact, proximity, charge, distance, angle, chirality, surface area, volume, excluded volume, shape, symmetry, and localization of particles or sets of particles (Table I). Below, we highlight some restraints in the context of the H-RNAPII structure determination process.

### *Dealing with Ambiguity*

Structural interpretation of data can be ambiguous, especially for proteomics data sets. For instance, if multiple copies of a protein exist in an assembly, a protein-protein interaction derived from a proteomics experiment may not be uniquely assigned to a specific pair of copies. Such ambiguous information must be translated into a restraint that considers all possible structural interpretations of the data; for example, an interaction between two protein types in an assembly with two symmetry units can occur either between the protein copies within each unit or between proteins across the two units (or both). We refer to such restraints as *conditional restraints* (45).

### *Distance Restraints from Proteomics*

We used direct physical interactions between eight pairs of eukaryotic RNAPII subunits as determined by the yeast two-hybrid (Y2H) system (58–66), protein complementation assays (67), co-localization (47), and complex reconstitution experiments (68, 69) (Table III). These interacting pairs were retrieved from the BioGRID database. Because we aim here to illustrate only what proteomics could do for structure determination, we selected true positive pairwise interactions and ignored the false positives; a discussion of techniques for addressing false positive interactions follows under “Dealing with Incorrect Data, Incomplete Data, and Multiple States”. There are also “indirect” interaction data in BioGRID. However, because BioGRID does not annotate which interactions are physical as opposed to indirect, we encoded as contact distance restraints only those experimentally measured interactions that have been detected by “pairwise” methods listed above.

In general, distance restraints may operate on multiple scales, ranging from the distance between two atoms or residues to the distance between two protein centers in an assembly. For example, if a direct interaction between two proteins has been identified, we may apply a restraint that penalizes deviations from a specified distance between the two protein centers. This distance restraint scores equally all relative orientations between the two proteins with the same intercenter distance. When the shape of the interacting proteins is known, we can achieve a more accurate score at the cost of additional computational time by restraining the distance between the closest pair of particles across the protein-protein interface. Because we do not know *a priori* which two atoms, residues, or domains are closest to each other, this ambiguity must be handled by a conditional restraint.

### *Connectivity Restraints from Proteomics*

In addition to the pairwise interactions described in the previous section, we also chose to use five sets of physically interacting RNAPII subunits as revealed by affinity purification and mass spectrometry (Table III). We searched three major large scale proteomics data sets (70, 71, 72) for all sets of interacting components that consist of RNAPII subunits only. We then disregarded sets of more than six subunits because such large affinity purification sets are relatively uninformative about the RNAPII structure (their inclusion does not significantly alter the results of our calculations; data not shown). In addition, because the majority of the sets (71 of the 103) were found in the Krogan *et al.* (72) data set, we used only the Krogan *et al.* (72) data set for our calculations. For affinity purification data, we know that at least one copy of each protein in a set directly interacts with at least one copy of another protein in the set; however, affinity purification data do not provide information on the stoichiometry of the proteins in the set, the number of complexes with distinct stoichiometry and configuration, or exactly which binary interactions occur, thus resulting in a great deal of ambiguity in the structural interpretation of the results. Because of this ambiguity, each affinity-purified set is encoded as a connectivity restraint that optimizes the assignment of binary interactions to proteins in the set along with the configuration of proteins (39). A putative binary interaction network for the proteins that best satisfies all available data for the system is assigned during each evaluation of the connectivity restraint during the optimization procedure.

### *Quality-of-fit Restraint from an Electron Density Map*

The fit of a model into an assembly density map is usually assessed by a cross-correlation measure between the assembly density and the model smoothed to the resolution of the map (22). Here, the configurations of the H-RNAPII subunits were restrained to fit an electron density map of the H-RNAPII complex (50).

### *Excluded Volume Restraint*

Molecules take up space that cannot be occupied by other molecules. This space filling property provides a key restraint on the conformations of the assembly. If the atomic structure is known, as is the case for H-RNAPII, the van der Waals radius for each atom is typically used to define the excluded volume (73). When the structure of a molecule is not known, it can be represented by a sphere; the volume of the sphere can be estimated from its composition (e.g. the number of residues in a protein (74)).

### *Geometric Complementarity Restraint from First Principles*

Protein-protein interfaces are typically geometrically complementary, characterized by tight packing with little space

between them. This geometric complementarity is commonly used as a restraint in protein-protein docking (75, 76). Because atomic models are used for H-RNAPII subunit structures, this consideration was enforced with an explicit restraint. The geometric complementarity restraint may be less informative if used on coarsely represented subunits.

#### *Additional Restraints*

Although not applied in our integrative structure determination of H-RNAPII, many additional restraint types can also be used.

#### *Radial Distribution Restraint*

An approximate radial distribution function of an assembly can be measured by an SAXS experiment (24, 25). Correspondingly, the SAXS restraint on a model can penalize the difference between the experimental and computed radial distribution functions (77). This restraint was used, for example, to select among several putative configurations of domains for the chaperone Hsp90 (78).

#### *Symmetry Restraint*

Symmetry is a recurrent theme in macromolecular assembly structures (79–81). For example, cyclic, helical, dihedral, and icosahedral symmetries are found in many important molecular machines such as viruses, the NPC, and chaperonins. The similarity between corresponding particle configurations in each symmetry unit can be enforced by imposing a restraint that maintains the same particle-particle distances within each unit (39, 82).

#### *Physical Energy and Statistical Potential Restraints*

Positions and orientations of interacting proteins can also be restrained by potentials based on the laws of physics (83–86) as well as statistical potentials extracted from databases of known protein structures (87–92). For example, a statistical potential can be derived from the observed distance distributions or contact frequencies of different atom type pairs in structurally defined proteins or complexes (93–96).

#### *Combining Restraints into a Scoring Function*

Once the data sets are encoded as restraints, they are combined into a scoring function, usually the sum of all the restraints. In this sum, the degree of uncertainty encoded by each restraint is effectively its weight. Ideally, the restraint on a spatial feature should be a probability density function on the feature given the corresponding measurement (39); for example, the lower and upper bounds on a distance should reflect the uncertainty of the corresponding distance measurement and its interpretation.

#### *Calculation of an Ensemble of Structures by Satisfaction of Spatial Restraints*

Next, all structural models that minimize the scoring function and therefore fit the original data must be found. An optimization procedure performs a search through the space of all possible macromolecular complex configurations by minimizing the violations of all restraints simultaneously. It is helpful to have many optimization methods available and to choose one that works best with a given representation and set of restraints. We have implemented several different optimizers as part of the Integrative Modeling Platform package. These optimizers can be classified as whole-system and divide-and-conquer optimizers.

#### *Whole-system Optimizers*

In this class of optimizers, an algorithm usually starts with a random initial configuration. The space of conformations is then explored iteratively by computing the next assembly configuration based on the values of all restraints for the configuration in the current optimization step with the intent of moving closer to the minimum value of the scoring function. Optimizers in this class include traditional conjugate gradients (97), quasi-Newton (98) and molecular dynamics schemes (99), Monte Carlo procedures as well as more sophisticated methods such as self-guided Langevin dynamics (100), and the replica exchange protocol (101). Because of the stochastic nature of these optimizations and the need to find all good scoring solutions, many independent runs are generally performed, each starting with a different random initial configuration.

#### *Divide-and-conquer Optimizers*

Divide-and-conquer optimizers can separate the particles and restraints in a system into smaller “suboptimizations,” ultimately resulting in more rapid sampling of structures. We have recently suggested a general divide-and-conquer approach to more efficiently sample protein assembly configurations (32). In this approach, the set of variables is decomposed into relatively uncoupled but potentially overlapping subsets that can be sampled independently of each other (*i.e.* are not required to be sampled together in a single calculation and can be sampled in parallel) and then efficiently gathered to compute the global minimum. The strength of this approach is derived from the decomposition procedure, which helps to reduce the size of the search space from exponential in the number of components in the whole system to exponential in the number of components in the largest subset. Similar ideas have been used for various modeling tasks such as side chain packing (102–104), sequence-structure threading (103), *ab initio* RNA folding (105), and prediction of quaternary structures of multiprotein complexes (106).

### *Use of Restraints to Restrain the Search Space for Optimization*

Efficiency can be increased by designing an optimization scheme to avoid considering configurations that clearly violate a subset of the data. Examples include segmenting an electron density map for the entire assembly into components that likely correspond to individual proteins prior to fitting the assembly proteins into the map (32), eliminating geometrically unlikely protein-protein docking solutions (75, 107), and restricting the search space to symmetric configurations (108, 109).

### *Human RNAPII Optimization*

For our H-RNAPII example, we used the sum of the distance, connectivity, EM quality-of-fit, and geometric complementarity restraints described above as a scoring function. The configuration of the subunits in H-RNAPII was optimized using an extension of the divide-and-conquer MultiFit protocol (Fig. 2) (32, 33).<sup>2</sup> We began by segmenting the electron density map into 12 regions, each one of which served to localize one of the 12 constituent H-RNAPII proteins. This procedure resulted in 479,001,600 (12!) possible H-RNAPII subunit configurations. Next, we eliminated all H-RNAPII subunit configurations that did not satisfy a majority of the proteomics restraints (Table III), keeping only 2,576 configurations for further refinement. We then refined each of these 2,576 configurations to optimize the EM quality-of-fit and geometric complementarity restraints using the standard MultiFit protocol (32); 63 of the 2,567 configurations resulted in refined models with “good” scores. These models had equivalent positions for Rpb1, Rpb2 and Rpb3; however, the models varied in the positions of the remaining subunits. Finally, we filtered the 63 models by all proteomics restraints, resulting in a single model that satisfied all proteomics restraints as well as the EM quality-of-fit and geometric complementarity restraints (Fig. 3).

### *Analysis of the Ensemble*

#### *Precision*

There are three possible outcomes of an optimization procedure. First, if only a single structural model satisfies all restraints and thus all input information, there is probably sufficient data for prediction of the unique native state. Second, if two or more different models are consistent with the restraints, the data are insufficient to define the single native state, or there are multiple significantly populated states. If the number of distinct models is small, the structural differences between the models may suggest additional experiments to narrow down the possible solutions. Third, if no models satisfy all input information, the data or their interpretation in

terms of the restraints are incorrect. For example, it might be that a complex exists in several functional states and that the available data cover more than one of them.

In the case of the H-RNAPII model, optimization resulted in a single model that satisfied all the data. Thus, sufficient information was available to predict the positions and orientations of the H-RNAPII subunits. The ensemble of possible models in the absence of proteomics data was much larger (2,576 coarse configurations) and defined the structure far less precisely. Therefore, proteomics data were crucial for providing an unambiguous determination of a precise molecular architecture of H-RNAPII.

#### *Accuracy*

Assessing the accuracy of a structure, defined as the difference between the model and the native structure, is difficult but important (45). It is impossible to know with certainty the accuracy of the proposed structure without knowing the real native structure. Nevertheless, our confidence can be modulated by five considerations: (a) self-consistency of independent experimental data; (b) structural similarity among all configurations in the ensemble that satisfy the input restraints; (c) simulations where a native structure is assumed, corresponding restraints are simulated from it, and the resulting calculated structure is compared with the assumed native structure; (d) confirmatory spatial data that were not used in the calculation of the structure (e.g. a criterion similar to the crystallographic free *R*-factor (110) can be used to assess both the model accuracy and the harmony among the input restraints); and (e) patterns emerging from a mapping of independent and unused data on the structure that are unlikely to occur by chance (13, 39).

In the case of H-RNAPII, we can estimate the accuracy directly because we know the crystallographic structure of the yeast RNAPII, which is likely to be highly similar to that of H-RNAPII (50) (c.f. the high degree of sequence similarity between yeast and human subunit orthologs (Table II) and the high correlation coefficient of 0.65 between the crystallographic yeast RNAPII structure and the electron density map of H-RNAPII). The H-RNAPII model clearly recapitulates the molecular architecture of yeast RNAPII (Fig. 3), preserving all of its protein interactions. More quantitatively, the subunits in the H-RNAPII model share a  $C\alpha$  root mean square deviation (RMSD) of only 11.4 Å with the human subunits individually superposed on their orthologs in the yeast RNAPII structure.

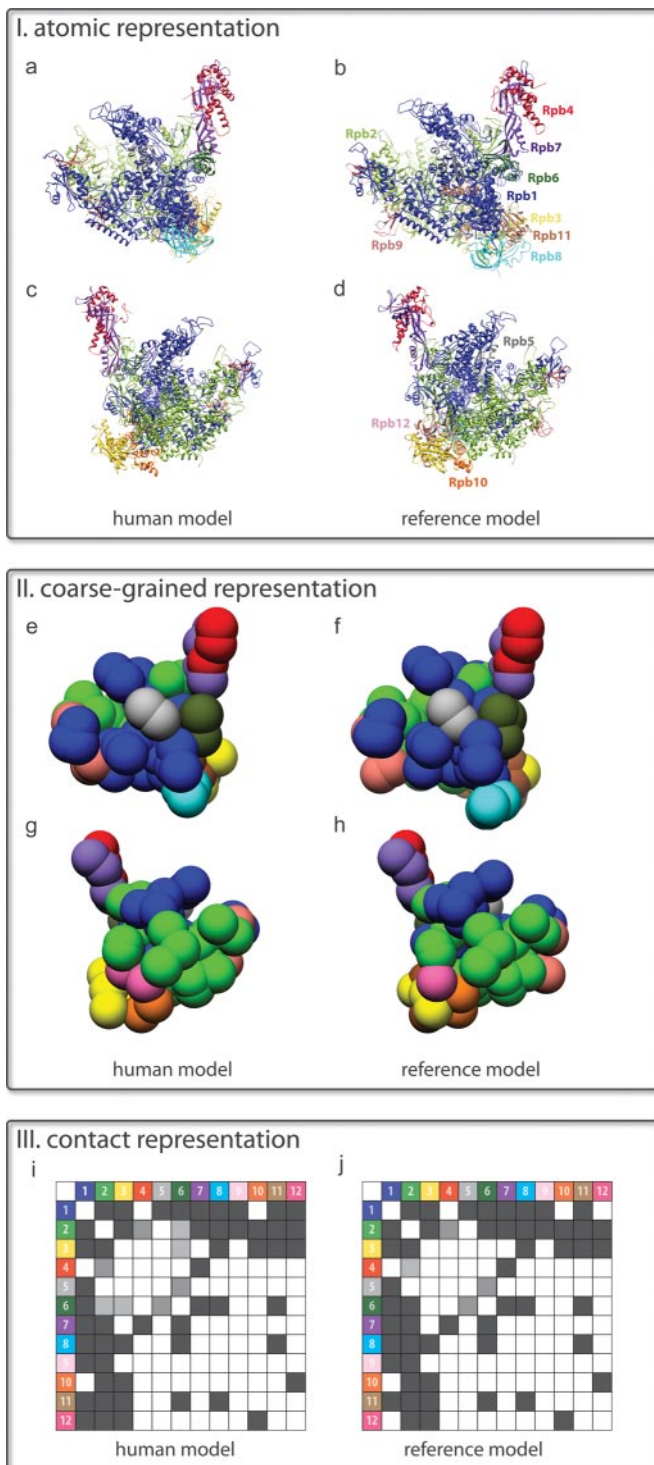
#### *Dealing with Incorrect Data, Incomplete Data, and Multiple States*

Proteome-wide protein-protein interaction maps have been produced by high throughput assays, such as affinity purification (11, 71) and yeast two-hybrid system (111–116). However, these data sets can be limited in three respects (117–119). First, the data can be incomplete in the sense that a

---

<sup>2</sup> K. Lasker, unpublished data.





**FIG. 3. Comparison of the crystallographic structure of yeast RNAPII and the integrative model of human RNAPII.** *I*, *a–d*, atomic representations of the integrative model of H-RNAPII and the reference structure in two views; the reference structure is composed of human subunits individually superposed on their orthologs in the yeast RNAPII structure. The configuration of the H-RNAPII subunits (*a* and *c*) is very similar to that in the reference structure (*b* and *d*); the C $\alpha$  RMSD is only 11.4 Å. *II*, *e–h*, coarse representations of the H-RNAPII model (*e* and *g*) and the reference structure (*f* and *h*) in the

number of interactions insufficient to describe the studied system were detected. Second, the data can be inaccurate in the sense that some detected interactions do not apply to the studied system. Third, the data can be “frustrated” in the sense that different subsets of the data apply to compositionally and/or conformationally different states of the studied system. For example, prior to filtering, a significant fraction of the affinity purification data for RNAPII subunits corresponds to false positive interactions (defined as a set of interacting subunits that do not have a connecting interaction path in the crystallographic structure of the complex (51)). In particular, 31, 35, and 0% of the 71, 26, and six affinity purification sets with two or more RNAPII subunits as reported by Krogan *et al.* (72), Gavin *et al.* (71), and Ho *et al.* (70) were false positives, respectively. In addition, 33% of the 12 reported binary interactions extracted from the BioGRID database were false positives.

A reasonable goal of structural modeling is to find the minimum number of system states that account for the observed data. If the data sets are correct and complete and describe a single state of the system, the optimization procedure should, in principle, result in a single solution that satisfies the data. If the data sets are inaccurate or incomplete, irrespective of the number of system states, the sampling should result in different states, some of which may or may not satisfy all the data. Next, we describe these possible outcomes in more detail.

*Correct, Complete Data, Single State*—The optimization procedure should result in a single solution that satisfies all restraints. If the data set is redundant, it is possible to cross-validate the solution by rerunning the modeling procedure using only random subsets of the data (120).

*Correct, Incomplete Data, Single State*—The optimization procedure should produce multiple solutions, all of which should satisfy all restraints. For example, this situation may occur when the proteomics data do not apply to all subunits of a system or only cover a small subset of interactions. It is possible to identify the least precisely localized components of the system within the set of solutions, directing future experiments for the largest possible gain in the next iteration of integrative modeling.

*Incorrect, Complete Data, Single State*—The optimization procedure should produce multiple solutions, each satisfying a fraction of the restraints. If there are redundant correct data, it may be possible to identify the conflicting incorrect data by cross-validation.

same two views as in *a–d* further illustrate the high similarity between the model and the reference. In the coarse representation, sets of 30 contiguous residues are shown as a single *bead*. *III*, *i* and *j*, protein contact maps for the H-RNAPII model and the reference structure (*white*, no contact; *gray*, weak contact; *black*, contact). The maps are essentially identical, differing only in the interactions of Rpb6 with Rpb2 and Rpb3, and Rpb1 with Rpb12.

*Incorrect, Incomplete Data, Single State*—The optimization procedure should produce multiple solutions, each satisfying a fraction of the restraints. It is difficult to identify the incorrect data as well as to detect a solution corresponding to the correct state. This situation arose in a preliminary attempt to model the molecular architecture of the 19 S regulatory particle of the 26 S proteasome (46). In that case, we have concluded that additional data are required.

*Multiple States*—Even when all data are correct and complete, the optimization procedure may be inadequate and produce multiple solutions, each satisfying only a fraction of the restraints. The same outcome is obtained when using incorrect data. Thus, multiple states are difficult to deconvolve from incorrect data (such as false positive interactions from proteomics).

In conclusion, when no solution is found that satisfies all data, it is difficult to identify the correct state(s). Formally, a similar problem exists in protein structure determination based on NMR spectroscopy. There, structural features, such as interatomic distances and dihedral angles, are obtained experimentally and used in the form of spatial restraints for finding the set of structural models that satisfies these restraints. One approach to dealing with incorrect data for one or more states looks at the frequency with which each restraint is violated in an ensemble of calculated structures (121, 122); if a given restraint is violated often, the bounds on the distances allowed by the restraint can be loosened. Other approaches use cross-validation to assess the completeness of the experimental restraints (123). Another development, the inferential structure determination method, formulates structure determination as an inference problem, handling incorrect and incomplete data as well as multiple states in a Bayesian framework (43). Adaptations of these methods and development of new methods should improve future handling of incorrect and incomplete data in integrative structure determination of conformationally and compositionally heterogeneous assemblies.

### DISCUSSION

As illustrated above, proteomics techniques can now facilitate the characterization of the structure of macromolecular assemblies via integrative modeling. We have demonstrated that by using atomic subunit structures, an electron density map of their assembly, and proteomics data restraining relative subunit proximities we can extend the scope of macromolecular structure determination beyond what is possible with single methods. Specifically, using the RNAPII structure as an example, we have shown that proteomics data, although traditionally not considered a source of formal structural information, can play a key role in assembly structure determination.

One key challenge for integrating proteomics data into structure determination remains the treatment of assemblies that exist in multiple functional states, corresponding to dif-

ferent configurations and compositions of the assembly. Although integrative methods can already restrain the structure of the modeled assembly based on all available information, some of the proteomics data may in fact apply to only a subset of all functional states of the assembly. For example, proteomics techniques often detect peripheral interactions that are not part of the core assembly but could be vital for one of the biologically relevant states. Thus, future protocols need to be able to simultaneously determine structures for all biologically relevant states. These methods will need to associate specific interactions with specific functionally relevant states of an assembly as well as remove false positive interactions that are not relevant to a given state.

As the quantity and variety of experimental data about macromolecular assemblies grows, integrative structure determination will be vital for characterization of these machines and the corresponding cellular processes. Methods are needed that are more accurate in translation of heterogeneous data into spatial restraints as well as combination of these restraints into a scoring function. New sampling and optimization schemes should improve the accuracy and level of detail with which we can describe assemblies. In addition, as a generalization of treating systems with multiple configurations and compositions, we should address the challenge of characterizing the dynamics of macromolecular assemblies by satisfying both spatial and temporal restraints for a system of multiple components. As integrative structure determination techniques advance, we will be able to describe an increasing number of key cellular structures, progressing toward a comprehensive structural, temporal, and logical model of the cell.

*Acknowledgments*—We thank Frank Alber, Michael P. Rout, Brian Chait, Wolfgang Baumeister, and Friedrich Förster for discussions about integrative structure determination based on proteomics data; Haim Wolfson for collaborating on optimization methods; and Hannes Braberg and Javier Fernandez-Martinez for discussing interpretation of proteomics data.

\* This work was supported, in whole or in part, by National Institutes of Health Grants R01 GM54762, U54 RR022220, PN2 EY016525, and R01 GM083960 (to A. Sali).

¶ Both authors contributed equally to this work.

|| Supported by the Clore Foundation Ph.D. Scholars program and carried out research in partial fulfillment of the requirements for the Ph.D. degree at Tel Aviv University. To whom correspondence may be addressed. E-mail: kerenl@salilab.org.

‡‡ To whom correspondence may be addressed. E-mail: jeremy@salilab.org.

§§ Supported by a Weizmann Institute Advancing Women in Science postdoctoral fellowship.

¶¶ Supported by the Sandler Family Supporting Foundation, National Science Foundation Grant IIS-0705196, Ron Conway, Mike Homer, Hewlett-Packard, NetApp, IBM, and Intel. To whom correspondence may be addressed: University of California, MC 2552, Byers Hall at Mission Bay, Suite 503B, 1700 4th St., San Francisco, CA 94158. Tel.: 415-514-4227; Fax: 415-514-4231; E-mail: sali@salilab.org.

## REFERENCES

- Alberts, B. (1998) The cell as a collection of protein machines: preparing the next generation of molecular biologists. *Cell* **92**, 291–294
- Abbott, A. (2002) Proteomics: the society of proteins. *Nature* **417**, 894–896
- Schmeing, T. M., and Ramakrishnan, V. (2009) What recent ribosome structures have revealed about the mechanism of translation. *Nature* **461**, 1234–1242
- Allen, G. S., and Frank, J. (2007) Structural insights on the translation initiation complex: ghosts of a universal initiation complex. *Mol. Microbiol.* **63**, 941–950
- Horwich, A. L., and Fenton, W. A. (2009) Chaperonin-mediated protein folding: using a central cavity to kinetically assist polypeptide chain folding. *Q. Rev. Biophys.* **42**, 83–116
- Spiess, C., Meyer, A. S., Reissmann, S., and Frydman, J. (2004) Mechanism of the eukaryotic chaperonin: protein folding in the chamber of secrets. *Trends Cell Biol.* **14**, 598–604
- Cramer, P., Armache, K. J., Baumli, S., Benkert, S., Brueckner, F., Buchen, C., Damsma, G. E., Dengl, S., Geiger, S. R., Jasiak, A. J., Jawhari, A., Jennebach, S., Kamenski, T., Kettenberger, H., Kuhn, C. D., Lehmann, E., Leike, K., Sydow, J. F., and Vannini, A. (2008) Structure of eukaryotic RNA polymerases. *Annu. Rev. Biophys.* **37**, 337–352
- Cheng, Y. (2009) Toward an atomic model of the 26S proteasome. *Curr. Opin. Struct. Biol.* **19**, 203–208
- Murata, S., Yashiroda, H., and Tanaka, K. (2009) Molecular mechanisms of proteasome assembly. *Nat. Rev. Mol. Cell Biol.* **10**, 104–115
- Förster, F., Lasker, K., Nickell, S., Sali, A., and Baumeister, W. (2010) Towards an integrated structural model of the 26S proteasome. *Mol. Cell. Proteomics*
- Gavin, A. C., Bösch, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J. M., Michon, A. M., Cruciat, C. M., Remor, M., Höfert, C., Schelder, M., Brajnovic, M., Ruffner, H., Merino, A., Klein, K., Hudak, M., Dickson, D., Rudi, T., Gnau, V., Bauch, A., Bastuck, S., Huhse, B., Leutwein, C., Heurtier, M. A., Copley, R. R., Edelmann, A., Querfurth, E., Rybin, V., Drewes, G., Raida, M., Bouwmeester, T., Bork, P., Seraphin, B., Kuster, B., Neubauer, G., and Superti-Furga, G. (2002) Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* **415**, 141–147
- Mitra, K., and Frank, J. (2006) Ribosome dynamics: insights from atomic structure modeling into cryo-electron microscopy maps. *Annu. Rev. Biophys. Biomol. Struct.* **35**, 299–317
- Alber, F., Dokudovskaya, S., Veenhoff, L. M., Zhang, W., Kipper, J., Devos, D., Suprpto, A., Karni-Schmidt, O., Williams, R., Chait, B. T., Sali, A., and Rout, M. P. (2007) The molecular architecture of the nuclear pore complex. *Nature* **450**, 695–701
- Robinson, C. V., Sali, A., and Baumeister, W. (2007) The molecular sociology of the cell. *Nature* **450**, 973–982
- Blundell, T. L., and Johnson, L. (1976) *Protein Crystallography*, Academic Press, New York
- Bonvin, A. M., Boelens, R., and Kaptein, R. (2005) NMR analysis of protein interactions. *Curr. Opin. Chem. Biol.* **9**, 501–508
- Fiaux, J., Bertelsen, E. B., Horwich, A. L., and Wüthrich, K. (2002) NMR analysis of a 900K GroEL GroES complex. *Nature* **418**, 207–211
- Neudecker, P., Lundström, P., and Kay, L. E. (2009) Relaxation dispersion NMR spectroscopy as a tool for detailed studies of protein folding. *Biophys. J.* **96**, 2045–2054
- Stahlberg, H., and Walz, T. (2008) Molecular electron microscopy: state of the art and current challenges. *ACS Chem. Biol.* **3**, 268–281
- Chiu, W., Baker, M. L., Jiang, W., Dougherty, M., and Schmid, M. F. (2005) Electron cryomicroscopy of biological machines at subnanometer resolution. *Structure* **13**, 363–372
- Lucic, V., Leis, A., and Baumeister, W. (2008) Cryo-electron tomography of cells: connecting structure and function. *Histochem. Cell Biol.* **130**, 185–196
- Frank, J. (2006) *Three-Dimensional Electron Microscopy of Macromolecular Assemblies: Visualization of Biological Molecules in Their Native State*, Oxford University Press, New York
- Berggård, T., Linse, S., and James, P. (2007) Methods for the detection and analysis of protein-protein interactions. *Proteomics* **7**, 2833–2842
- Svergun, D. I., Petoukhov, M. V., and Koch, M. H. (2001) Determination of domain structure of proteins from X-ray solution scattering. *Biophys. J.* **80**, 2946–2953
- Hura, G. L., Menon, A. L., Hammel, M., Rambo, R. P., Poole, F. L., 2nd, Tsutakawa, S. E., Jenney, F. E., Jr., Classen, S., Frankel, K. A., Hopkins, R. C., Yang, S. J., Scott, J. W., Dillard, B. D., Adams, M. W., and Tainer, J. A. (2009) Robust, high-throughput solution structural analyses by small angle X-ray scattering (SAXS). *Nat. Methods* **6**, 606–612
- Joo, C., Balci, H., Ishitsuka, Y., Buranachai, C., and Ha, T. (2008) Advances in single-molecule fluorescence methods for molecular biology. *Annu. Rev. Biochem.* **77**, 51–76
- Hart, G. T., Ramani, A. K., and Marcotte, E. M. (2006) How complete are current yeast and human protein-interaction networks? *Genome Biol.* **7**, 120
- Collins, S. R., Kemmeren, P., Zhao, X. C., Greenblatt, J. F., Spencer, F., Holstege, F. C., Weissman, J. S., and Krogan, N. J. (2007) Toward a comprehensive atlas of the physical interactome of *Saccharomyces cerevisiae*. *Mol. Cell. Proteomics* **6**, 439–450
- Cusick, M. E., Yu, H., Smolyar, A., Venkatesan, K., Carvunis, A. R., Simonis, N., Rual, J. F., Borick, H., Braun, P., Dreze, M., Vandenhautte, J., Galli, M., Yazaki, J., Hill, D. E., Ecker, J. R., Roth, F. P., and Vidal, M. (2009) Literature-curated protein interaction datasets. *Nat. Methods* **6**, 39–46
- Topf, M., Lasker, K., Webb, B., Wolfson, H., Chiu, W., and Sali, A. (2008) Protein structure fitting and refinement guided by cryo-EM density. *Structure* **16**, 295–307
- Topf, M., Baker, M. L., Marti-Renom, M. A., Chiu, W., and Sali, A. (2006) Refinement of protein structures by iterative comparative modeling and CryoEM density fitting. *J. Mol. Biol.* **357**, 1655–1668
- Lasker, K., Topf, M., Sali, A., and Wolfson, H. J. (2009) Inferential optimization for simultaneous fitting of multiple components into a cryoEM map of their assembly. *J. Mol. Biol.* **388**, 180–194
- Lasker, K., Sali, A., and Wolfson, H. J. (in press) Determining macromolecular assembly structures by molecular docking and fitting into an electron density map. *Proteins*
- Lindert, S., Stewart, P. L., and Meiler, J. (2009) Hybrid approaches: applying computational methods in cryo-electron microscopy. *Curr. Opin. Struct. Biol.* **19**, 218–225
- Qian, B., Raman, S., Das, R., Bradley, P., McCoy, A. J., Read, R. J., and Baker, D. (2007) High-resolution structure prediction and the crystallographic phase problem. *Nature* **450**, 259–264
- Taverner, T., Hernández, H., Sharon, M., Ruotolo, B. T., Matak-Vinković, D., Devos, D., Russell, R. B., and Robinson, C. V. (2008) Subunit architecture of intact protein complexes from mass spectrometry and homology modeling. *Acc. Chem. Res.* **41**, 617–627
- Bowers, P. M., Strauss, C. E., and Baker, D. (2000) De novo protein structure determination using sparse NMR data. *J. Biomol. NMR* **18**, 311–318
- Raman, S., Lange, O. F., Rossi, P., Tyka, M., Wang, X., Aramini, J., Liu, G., Ramelot, T. A., Eletsky, A., Szyperski, T., Kennedy, M. A., Prestegard, J., Montelione, G. T., and Baker, D. (2010) NMR structure determination for larger proteins using backbone-only data. *Science* **327**, 1014–1018
- Alber, F., Dokudovskaya, S., Veenhoff, L. M., Zhang, W., Kipper, J., Devos, D., Suprpto, A., Karni-Schmidt, O., Williams, R., Chait, B. T., Rout, M. P., and Sali, A. (2007) Determining the architectures of macromolecular assemblies. *Nature* **450**, 683–694
- Förster, F., Lasker, K., Beck, F., Nickell, S., Sali, A., and Baumeister, W. (2009) An Atomic Model AAA-ATPase/20S core particle sub-complex of the 26S proteasome. *Biochem. Biophys. Res. Commun.* **388**, 228–233
- Fotin, A., Cheng, Y., Sliz, P., Grigorieff, N., Harrison, S. C., Kirchhausen, T., and Walz, T. (2004) Molecular model for a complete clathrin lattice from electron cryomicroscopy. *Nature* **432**, 573–579
- Xing, Y., Böcking, T., Wolf, M., Grigorieff, N., Kirchhausen, T., and Harrison, S. C. (2010) Structure of clathrin coat with bound Hsc70 and auxilin: mechanism of Hsc70-facilitated disassembly. *EMBO J.* **29**, 655–665
- Chen, Z. A., Jawhari, A., Fischer, L., Buchen, C., Tahir, S., Kamenski, T., Rasmussen, M., Lariviere, L., Bukowski-Wills, J. C., Nilges, M., Cramer, P., and Rappsilber, J. (2010) Architecture of the RNA polymerase II-TFIIF complex revealed by cross-linking and mass spectrometry. *EMBO J.* **29**, 717–726
- Byeon, I. J., Meng, X., Jung, J., Zhao, G., Yang, R., Ahn, J., Shi, J., Concel, J., Aiken, C., Zhang, P., and Gronenborn, A. M. (2009) Structural con-

- vergence between Cryo-EM and NMR reveals intersubunit interactions critical for HIV-1 capsid function. *Cell* **139**, 780–790
45. Alber, F., Förster, F., Korkein, D., Topf, M., and Sali, A. (2008) Integrating diverse data for structure determination of macromolecular assemblies. *Annu. Rev. Biochem.* **77**, 443–477
  46. Nickell, S., Beck, F., Scheres, S. H., Korinek, A., Förster, F., Lasker, K., Mihalache, O., Sun, N., Nagy, I., Sali, A., Plitzko, J. M., Carazo, J. M., Mann, M., and Baumeister, W. (2009) Insights into the molecular architecture of the 26S proteasome. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 11943–11947
  47. Jasiak, A. J., Hartmann, H., Karakasili, E., Kalocsay, M., Flatley, A., Krenmer, E., Strässer, K., Martin, D. E., Söding, J., and Cramer, P. (2008) Genome-associated RNA polymerase II includes the dissociable Rpb4/7 subcomplex. *J. Biol. Chem.* **283**, 26423–26427
  48. Hahn, S. (2004) Structure and mechanism of the RNA polymerase II transcription machinery. *Nat. Struct. Mol. Biol.* **11**, 394–403
  49. Cramer, P., Bushnell, D. A., and Kornberg, R. D. (2001) Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science* **292**, 1863–1876
  50. Kostek, S. A., Grob, P., De Carlo, S., Lipscomb, J. S., Garczarek, F., and Nogales, E. (2006) Molecular architecture and conformational flexibility of human RNA polymerase II. *Structure* **14**, 1691–1700
  51. Kettenberger, H., Armache, K. J., and Cramer, P. (2004) Complete RNA polymerase II elongation complex structure and its interactions with NTP and TFIIIS. *Mol. Cell* **16**, 955–965
  52. Maiolica, A., Cittaro, D., Borsotti, D., Sennels, L., Ciferri, C., Tarricone, C., Musacchio, A., and Rappsilber, J. (2007) Structural analysis of multi-protein complexes by cross-linking, mass spectrometry, and database searching. *Mol. Cell. Proteomics* **6**, 2200–2211
  53. Young, M. M., Tang, N., Hempel, J. C., Oshiro, C. M., Taylor, E. W., Kuntz, I. D., Gibson, B. W., and Dollinger, G. (2000) High throughput protein fold identification by using experimental constraints derived from intramolecular cross-links and mass spectrometry. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 5802–5806
  54. Pieper, U., Chiang, R., Seffernick, J. J., Brown, S. D., Glasner, M. E., Kelly, L., Eswar, N., Sauder, J. M., Bonanno, J. B., Swaminathan, S., Burley, S. K., Zheng, X., Chance, M. R., Almo, S. C., Gerlt, J. A., Raushel, F. M., Jacobson, M. P., Babbitt, P. C., and Sali, A. (2009) Target selection and annotation for the structural genomics of the amidohydrolase and endonuclease superfamilies. *J. Struct. Funct. Genomics* **10**, 107–125
  55. Stark, C., Breitkreutz, B. J., Reguly, T., Boucher, L., Breitkreutz, A., and Tyers, M. (2006) BioGRID: a general repository for interaction datasets. *Nucleic Acids Res.* **34**, D535–D539
  56. Henrick, K., Newman, R., Tagari, M., and Chagoyen, M. (2003) EMDep: a web-based system for the deposition and validation of high-resolution electron microscopy macromolecular structural information. *J. Struct. Biol.* **144**, 228–237
  57. Pieper, U., Eswar, N., Webb, B. M., Eramian, D., Kelly, L., Barkan, D. T., Carter, H., Mankoo, P., Karchin, R., Marti-Renom, M. A., Davis, F. P., and Sali, A. (2009) MODBASE, a database of annotated comparative protein structure models and associated resources. *Nucleic Acids Res.* **37**, D347–D354
  58. Flores, A., Briand, J. F., Gadal, O., Andrau, J. C., Rubbi, L., Van Mullem, V., Boschiero, C., Gousso, M., Marck, C., Carles, C., Thuriaux, P., Sentenac, A., and Werner, M. (1999) A protein-protein interaction map of yeast RNA polymerase III. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 7815–7820
  59. Zeros, C., Briand, J. F., Boulard, Y., Labarre-Mariotte, S., Garcia-Lopez, M. C., Thuriaux, P., and Navarro, F. (2007) Functional organization of the Rpb5 subunit shared by the three yeast RNA polymerases. *Nucleic Acids Res.* **35**, 634–647
  60. Briand, J. F., Navarro, F., Rematier, P., Boschiero, C., Labarre, S., Werner, M., Shpakovski, G. V., and Thuriaux, P. (2001) Partners of Rpb8p, a small subunit shared by yeast RNA polymerases I, II and III. *Mol. Cell. Biol.* **21**, 6056–6065
  61. Tan, Q., Prysak, M. H., and Woychik, N. A. (2003) Loss of the Rpb4/Rpb7 subcomplex in a mutant form of the Rpb6 subunit shared by RNA polymerases I, II, and III. *Mol. Cell. Biol.* **23**, 3329–3338
  62. Qi, H., and Zakian, V. A. (2000) The *Saccharomyces* telomere-binding protein Cdc13p interacts with both the catalytic subunit of DNA polymerase alpha and the telomerase-associated est1 protein. *Genes Dev.* **14**, 1777–1788
  63. Sampath, V., Rekha, N., Srinivasan, N., and Sadhale, P. (2003) The conserved and non-conserved regions of Rpb4 are involved in multiple phenotypes in *Saccharomyces cerevisiae*. *J. Biol. Chem.* **278**, 51566–51576
  64. Khazak, V., Sadhale, P. P., Woychik, N. A., Brent, R., and Golemis, E. A. (1995) Human RNA polymerase II subunit hSRPB7 functions in yeast and influences stress survival and cell morphology. *Mol. Biol. Cell* **6**, 759–775
  65. Sareen, A., Choudhry, P., Mehta, S., and Sharma, N. (2005) Mapping the interaction site of Rpb4 and Rpb7 subunits of RNA polymerase II in *Saccharomyces cerevisiae*. *Biochem. Biophys. Res. Commun.* **332**, 763–770
  66. Selitrennik, M., Duek, L., Lotan, R., and Choder, M. (2006) Nucleocytoplasmic shuttling of the Rpb4p and Rpb7p subunits of *Saccharomyces cerevisiae* RNA polymerase II by two pathways. *Eukaryot. Cell* **5**, 2092–2103
  67. Tarassov, K., Messier, V., Landry, C. R., Radinovic, S., Serna Molina, M. M., Shames, I., Malitskaya, Y., Vogel, J., Bussey, H., and Michnick, S. W. (2008) An in vivo map of the yeast protein interactome. *Science* **320**, 1465–1470
  68. Benga, W. J., Grandemange, S., Shpakovski, G. V., Shematorova, E. K., Keding, C., and Vigneron, M. (2005) Distinct regions of RPB11 are required for heterodimerization with RPB3 in human and yeast RNA polymerase II. *Nucleic Acids Res.* **33**, 3582–3590
  69. Orlicky, S. M., Tran, P. T., Sayre, M. H., and Edwards, A. M. (2001) Dissociable Rpb4-Rpb7 subassembly of RNA polymerase II binds to single-strand nucleic acid and mediates a post-recruitment step in transcription initiation. *J. Biol. Chem.* **276**, 10097–10102
  70. Ho, Y., Gruhler, A., Heilbut, A., Bader, G. D., Moore, L., Adams, S. L., Millar, A., Taylor, P., Bennett, K., Boutillier, K., Yang, L., Wolting, C., Donaldson, I., Schandorff, S., Shewnarane, J., Vo, M., Taggart, J., Goudreau, M., Musk, B., Alfano, C., Dewar, D., Lin, Z., Michalickova, K., Willems, A. R., Sassi, H., Nielsen, P. A., Rasmussen, K. J., Andersen, J. R., Johansen, L. E., Hansen, L. H., Jespersen, H., Podtelejnikov, A., Nielsen, E., Crawford, J., Poulsen, V., Sørensen, B. D., Matthies, J., Hendrickson, R. C., Gleeson, F., Pawson, T., Moran, M. F., Durocher, D., Mann, M., Hogue, C. W., Figeys, D., and Tyers, M. (2002) Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature* **415**, 180–183
  71. Gavin, A. C., Aloy, P., Grandi, P., Krause, R., Boesche, M., Marzioch, M., Rau, C., Jensen, L. J., Bastuck, S., Dümpelfeld, B., Edelmann, A., Heurter, M. A., Hoffman, V., Hoefert, C., Klein, K., Hudak, M., Michon, A. M., Schelder, M., Schirle, M., Remor, M., Rudi, T., Hooper, S., Bauer, A., Bouwmeester, T., Casari, G., Drewes, G., Neubauer, G., Rick, J. M., Kuster, B., Bork, P., Russell, R. B., and Superti-Furga, G. (2006) Proteome survey reveals modularity of the yeast cell machinery. *Nature* **440**, 631–636
  72. Krogan, N. J., Cagney, G., Yu, H., Zhong, G., Guo, X., Ignatchenko, A., Li, J., Pu, S., Datta, N., Tikuisis, A. P., Punna, T., Peregrín-Alvarez, J. M., Shales, M., Zhang, X., Davey, M., Robinson, M. D., Paccanaro, A., Bray, J. E., Sheung, A., Beattie, B., Richards, D. P., Canadien, V., Lalev, A., Mena, F., Wong, P., Starostine, A., Canete, M. M., Vlasblom, J., Wu, S., Orsi, C., Collins, S. R., Chandran, S., Haw, R., Rilstone, J. J., Gandi, K., Thompson, N. J., Musso, G., St Onge, P., Ghanny, S., Lam, M. H., Butland, G., Altaf-Ul, A. M., Kanaya, S., Shilatifard, A., O’Shea, E., Weissman, J. S., Ingles, C. J., Hughes, T. R., Parkinson, J., Gerstein, M., Wodak, S. J., Emili, A., and Greenblatt, J. F. (2006) Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. *Nature* **440**, 637–643
  73. Connolly, M. L. (1983) Solvent-accessible surfaces of proteins and nucleic acids. *Science* **221**, 709–713
  74. Shen, M., Davis, F. P., and Sali, A. (2005) The optimal size of a globular protein domain: a simple sphere-packing model. *Chem. Phys. Lett.* **405**, 224–228
  75. Katchalski-Katzir, E., Shariv, I., Eisenstein, M., Friesem, A. A., Aflalo, C., and Vakser, I. A. (1992) Molecular surface recognition: determination of geometric fit between proteins and their ligands by correlation techniques. *Proc. Natl. Acad. Sci. U.S.A.* **89**, 2195–2199
  76. Duhovny, D., Nussinov, R., and Wolfson, H. J. (2002) Efficient unbound docking of rigid molecules, in *Second International Workshop on Algo-*

- rithms in Bioinformatics* (Guido, R., and Gusfield, D., eds) pp. 185–200, Springer-Verlag, London
77. Förster, F., Webb, B., Krukenberg, K. A., Tsuruta, H., Agard, D. A., and Sali, A. (2008) Integration of small-angle X-ray scattering data into structural modeling of proteins and their assemblies. *J. Mol. Biol.* **382**, 1089–1106
  78. Krukenberg, K. A., Förster, F., Rice, L. M., Sali, A., and Agard, D. A. (2008) Multiple conformations of *E. coli* Hsp90 in solution: insights into the conformational dynamics of Hsp90. *Structure* **16**, 755–765
  79. Goodsell, D. S., and Olson, A. J. (2000) Structural symmetry and protein function. *Annu. Rev. Biophys. Biomol. Struct.* **29**, 105–153
  80. Tama, F., and Brooks, C. L. (2006) Symmetry, form, and shape: guiding principles for robustness in macromolecular machines. *Annu. Rev. Biophys. Biomol. Struct.* **35**, 115–133
  81. Levy, E. D., Boeri Erba, E., Robinson, C. V., and Teichmann, S. A. (2008) Assembly reflects evolution of protein complexes. *Nature* **453**, 1262–1265
  82. Alber, F., Kim, M. F., and Sali, A. (2005) Structural characterization of assemblies from overall shape and subcomplex compositions. *Structure* **13**, 435–445
  83. Brooks, B. R., Brucoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., and Karplus, M. (1983) *CHARMM: a Program for Macromolecular Energy, Minimization, and Dynamics Calculations*, Wiley, New York
  84. Pearlman, D. (1995) AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Comput. Phys. Commun.* **91**, 1–41
  85. Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E., and Berendsen, H. J. (2005) GROMACS: fast, flexible, and free. *J. Comput. Chem.* **26**, 1701–1718
  86. Jorgensen, W. L., and Tirado-Rives, J. (1988) *The OPLS Potential Functions for Proteins. Energy Minimizations for Crystals of Cyclic Peptides and Crambin*. *J. Am. Chem. Soc.* **110**, 657–666
  87. Shen, M. Y., and Sali, A. (2006) Statistical potential for assessment and prediction of protein structures. *Protein Sci.* **15**, 2507–2524
  88. Zhang, C., Liu, S., and Zhou, Y. (2004) Accurate and efficient loop selections by the DFIRE-based all-atom statistical potential. *Protein Sci.* **13**, 391–399
  89. Melo, F., Sánchez, R., and Sali, A. (2002) Statistical potentials for fold assessment. *Protein Sci.* **11**, 430–448
  90. Misura, K. M., Chivian, D., Rohl, C. A., Kim, D. E., and Baker, D. (2006) Physically realistic homology models built with ROSETTA can be more accurate than their templates. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 5361–5366
  91. Simons, K. T., Kooperberg, C., Huang, E., and Baker, D. (1997) Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions. *J. Mol. Biol.* **268**, 209–225
  92. Simons, K. T., Ruczinski, I., Kooperberg, C., Fox, B. A., Bystroff, C., and Baker, D. (1999) Improved recognition of native-like protein structures using a combination of sequence-dependent and sequence-independent features of proteins. *Proteins* **34**, 82–95
  93. Sippl, M. J. (1990) Calculation of conformational ensembles from potentials of mean force. An approach to the knowledge-based prediction of local structures in globular proteins. *J. Mol. Biol.* **213**, 859–883
  94. Davis, F. P., and Sali, A. (2005) PIBASE: a comprehensive database of structurally defined protein interfaces. *Bioinformatics* **21**, 1901–1907
  95. Davis, F. P., Braberg, H., Shen, M. Y., Pieper, U., Sali, A., and Madhusudhan, M. S. (2006) Protein complex compositions predicted by structural similarity. *Nucleic Acids Res.* **34**, 2943–2952
  96. Eswar, N., Eramian, D., Webb, B., Shen, M. Y., and Sali, A. (2008) Protein structure modeling with MODELLER. *Methods Mol. Biol.* **426**, 145–159
  97. Shanno, D. F., and Phua, K. H. (1980) Minimization of unrestrained multivariate functions. *ACM Trans. Math. Soft.* **6**, 618–622
  98. Ponder, J. W., and Richards, F. M. (1987) An efficient newton-like method for molecular mechanics energy minimization of large molecules. *J. Comput. Chem.* **8**, 1016–1024
  99. Karplus, M., and McCammon, J. A. (2002) Molecular dynamics simulations of biomolecules. *Nat. Struct. Biol.* **9**, 646–652
  100. Wu, X., and Brooks, B. R. (2003) Self-guided Langevin dynamics simulation method. *Chem. Phys. Lett.* **381**, 512–518
  101. Sugita, Y., and Okamoto, Y. (1999) Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **314**, 141–151
  102. Canutescu, A. A., Shelenkov, A. A., and Dunbrack, R. L., Jr. (2003) A graph-theory algorithm for rapid protein side-chain prediction. *Protein Sci.* **12**, 2001–2014
  103. Xu, J., Jiao, F., and Berger, B. (2005) A tree-decomposition approach to protein structure prediction. *Proc. IEEE Comput. Syst. Bioinform. Conf.* 247–256
  104. Yanover, C., Schueler-Furman, O., and Weiss, Y. (2008) Minimizing and learning energy functions for side-chain prediction. *J. Comput. Biol.* **15**, 899–911
  105. Zhao, J., Malmberg, R. L., and Cai, L. (2008) Rapid ab initio prediction of RNA pseudoknots via graph tree decomposition. *J. Math. Biol.* **56**, 145–159
  106. Inbar, Y., Benyamini, H., Nussinov, R., and Wolfson, H. J. (2005) Prediction of multimolecular assemblies by multiple docking. *J. Mol. Biol.* **349**, 435–447
  107. Schneidman-Duhovny, D., Inbar, Y., Polak, V., Shatsky, M., Halperin, I., Benyamini, H., Barzilai, A., Dror, O., Haspel, N., Nussinov, R., and Wolfson, H. J. (2003) Taking geometry to its edge: fast unbound rigid (and hinge-bent) docking. *Proteins* **52**, 107–112
  108. Schneidman-Duhovny, D., Inbar, Y., Nussinov, R., and Wolfson, H. J. (2005) Geometry-based flexible and symmetric protein docking. *Proteins* **60**, 224–231
  109. André, I., Bradley, P., Wang, C., and Baker, D. (2007) Prediction of the structure of symmetrical protein assemblies. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 17656–17661
  110. Brünger, A. T. (1992) Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **355**, 472–475
  111. Stelzl, U., Worm, U., Lalowski, M., Haenig, C., Brembeck, F. H., Goehler, H., Stroedicke, M., Zenkner, M., Schoenherr, A., Koeppen, S., Timm, J., Mintzlauff, S., Abraham, C., Bock, N., Kietzmann, S., Goedde, A., Toksöz, E., Droege, A., Krobitsch, S., Korn, B., Birchmeier, W., Lehrach, H., and Wanker, E. E. (2005) A human protein-protein interaction network: a resource for annotating the proteome. *Cell* **122**, 957–968
  112. Rual, J. F., Venkatesan, K., Hao, T., Hirozane-Kishikawa, T., Dricot, A., Li, N., Berriz, G. F., Gibbons, F. D., Dreze, M., Ayivi-Guedehoussou, N., Klitgord, N., Simon, C., Boxem, M., Milstein, S., Rosenberg, J., Goldberg, D. S., Zhang, L. V., Wong, S. L., Franklin, G., Li, S., Albal, J. S., Lim, J., Fraughton, C., Llamosas, E., Cevik, S., Bex, C., Lamesch, P., Sikorski, R. S., Vandenhaute, J., Zoghbi, H. Y., Smolyar, A., Bosak, S., Sequerra, R., Doucette-Stamm, L., Cusick, M. E., Hill, D. E., Roth, F. P., and Vidal, M. (2005) Towards a proteome-scale map of the human protein-protein interaction network. *Nature* **437**, 1173–1178
  113. Giot, L., Bader, J. S., Brouwer, C., Chaudhuri, A., Kuang, B., Li, Y., Hao, Y. L., Ooi, C. E., Godwin, B., Vitols, E., Vijayadamar, G., Pochart, P., Machineni, H., Welsh, M., Kong, Y., Zerhusen, B., Malcolm, R., Varrone, Z., Collis, A., Minto, M., Burgess, S., McDaniel, L., Stimpson, E., Spriggs, F., Williams, J., Neurath, K., Ioigne, N., Agee, M., Voss, E., Furtak, K., Renzulli, R., Aanensen, N., Carroll, S., Bickelhaupt, E., Lazovatsky, Y., DaSilva, A., Zhong, J., Stanyon, C. A., Finley, R. L., Jr., White, K. P., Braverman, M., Jarvie, T., Gold, S., Leach, M., Knight, J., Shimkets, R. A., McKenna, M. P., Chant, J., and Rothberg, J. M. (2003) A protein interaction map of *Drosophila melanogaster*. *Science* **302**, 1727–1736
  114. Walhout, A. J., Sordella, R., Lu, X., Hartley, J. L., Temple, G. F., Brasch, M. A., Thierry-Mieg, N., and Vidal, M. (2000) Protein interaction mapping in *C. elegans* using proteins involved in vulval development. *Science* **287**, 116–122
  115. Uetz, P., Giot, L., Cagney, G., Mansfield, T. A., Judson, R. S., Knight, J. R., Lockshon, D., Narayan, V., Srinivasan, M., Pochart, P., Qureshi-Emili, A., Li, Y., Godwin, B., Conover, D., Kalbfleisch, T., Vijayadamar, G., Yang, M., Johnston, M., Fields, S., and Rothberg, J. M. (2000) A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature* **403**, 623–627
  116. Ito, T., Chiba, T., Ozawa, R., Yoshida, M., Hattori, M., and Sakaki, Y. (2001) A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 4569–4574
  117. Bader, G. D., and Hogue, C. W. (2002) Analyzing yeast protein-protein interaction data obtained from different sources. *Nat. Biotechnol.* **20**,

- 991–997
118. von Mering, C., Krause, R., Snel, B., Cornell, M., Oliver, S. G., Fields, S., and Bork, P. (2002) Comparative assessment of large-scale data sets of protein-protein interactions. *Nature* **417**, 399–403
119. Mann, M., and Kelleher, N. L. (2008) Precision proteomics: the case for high resolution and high mass accuracy. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 18132–18138
120. Duda, R. O., Hart, P. E., and Stork, D. G. (2001) *Pattern Classification*, Wiley, New York
121. Clore, G. M., Robien, M. A., and Gronenborn, A. M. (1993) Exploring the limits of precision and accuracy of protein structures determined by nuclear magnetic resonance spectroscopy. *J. Mol. Biol.* **231**, 82–102
122. Clore, G. M., and Gronenborn, A. M. (1998) New methods of structure refinement for macromolecular structure determination by NMR. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 5891–5898
123. Brünger, A. T., Clore, G. M., Gronenborn, A. M., Saffrich, R., and Nilges, M. (1993) Assessing the quality of solution nuclear magnetic resonance structures by complete cross-validation. *Science* **261**, 328–331
124. Gingras, A. C., Gstaiger, M., Raught, B., and Aebersold, R. (2007) Analysis of protein complexes using mass spectrometry. *Nat. Rev. Mol. Cell Biol.* **8**, 645–654
125. Zhou, M., and Robinson, C. V. (in press) When proteomics meets structural biology. *Trends Biochem. Sci.*
126. Bich, C., and Zenobi, R. (2009) Mass spectrometry of large complexes. *Curr. Opin. Struct. Biol.* **19**, 632–639
127. Towbin, H., Staehelin, T., and Gordon, J. (1979) Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. *Proc. Natl. Acad. Sci. U.S.A.* **76**, 4350–4354
128. Roguev, A., Bandyopadhyay, S., Zofall, M., Zhang, K., Fischer, T., Collins, S. R., Qu, H., Shales, M., Park, H. O., Hayles, J., Hoe, K. L., Kim, D. U., Ideker, T., Grewal, S. I., Weissman, J. S., and Krogan, N. J. (2008) Conservation and rewiring of functional modules revealed by an epistasis map in fission yeast. *Science* **322**, 405–410
129. Phillips, P. C. (2008) Epistasis—the essential role of gene interactions in the structure and evolution of genetic systems. *Nat. Rev. Genet.* **9**, 855–867
130. Skrabanek, L., Saini, H. K., Bader, G. D., and Enright, A. J. (2008) Computational prediction of protein-protein interactions. *Mol. Biotechnol.* **38**, 1–17
131. Visser, N. F., and Heck, A. J. (2008) Surface plasmon resonance mass spectrometry in proteomics. *Expert Rev. Proteomics* **5**, 425–433
132. Stoevesandt, O., Taussig, M. J., and He, M. (2009) Protein microarrays: high-throughput tools for proteomics. *Expert Rev. Proteomics* **6**, 145–157
133. Wolf-Yadlin, A., Sevecka, M., and MacBeath, G. (2009) Dissecting protein function and signaling using protein microarrays. *Curr. Opin. Chem. Biol.* **13**, 398–405
134. Korf, U., and Wiemann, S. (2005) Protein microarrays as a discovery tool for studying protein-protein interactions. *Expert Rev. Proteomics* **2**, 13–26
135. Kerppola, T. K. (2006) Visualization of molecular interactions by fluorescence complementation. *Nat. Rev. Mol. Cell Biol.* **7**, 449–456
136. Remy, I., and Michnick, S. W. (2007) Application of protein-fragment complementation assays in cell biology. *BioTechniques* **42**, 137–145
137. Freyer, M. W., and Lewis, E. A. (2008) Isothermal titration calorimetry: experimental design, data analysis, and probing macromolecule/ligand binding and kinetic interactions. *Methods Cell Biol.* **84**, 79–113
138. Velazquez-Campoy, A., Leavitt, S. A., and Freire, E. (2004) Characterization of protein-protein interactions by isothermal titration calorimetry. *Methods Mol. Biol.* **261**, 35–54
139. Piston, D. W., and Kremers, G. J. (2007) Fluorescent protein FRET: the good, the bad and the ugly. *Trends Biochem. Sci.* **32**, 407–414
140. Pflieger, K. D., and Eidne, K. A. (2006) Illuminating insights into protein-protein interactions using bioluminescence resonance energy transfer (BRET). *Nat. Methods* **3**, 165–174
141. Lucocq, J. (2008) Quantification of structures and gold labeling in transmission electron microscopy. *Methods Cell Biol.* **88**, 59–82
142. Hainfeld, J. F., and Powell, R. D. (2000) New frontiers in gold labeling. *J. Histochem. Cytochem.* **48**, 471–480
143. Drummond, S. P., and Allen, T. D. (2008) From live-cell imaging to scanning electron microscopy (SEM): the use of green fluorescent protein (GFP) as a common label. *Methods Cell Biol.* **88**, 97–108
144. Vajda, S., and Kozakov, D. (2009) Convergence and combination of methods in protein-protein docking. *Curr. Opin. Struct. Biol.* **19**, 164–170
145. Sinz, A. (2006) Chemical cross-linking and mass spectrometry to map three-dimensional protein structures and protein-protein interactions. *Mass Spectrom. Rev.* **25**, 663–682
146. Trester-Zedlitz, M., Kamada, K., Burley, S. K., Fenyö, D., Chait, B. T., and Muir, T. W. (2003) A modular cross-linking approach for exploring protein interactions. *J. Am. Chem. Soc.* **125**, 2416–2425
147. Tsutsui, Y., and Wintrode, P. L. (2007) Hydrogen/deuterium exchange-mass spectrometry: a powerful tool for probing protein structure, dynamics and interactions. *Curr. Med. Chem.* **14**, 2344–2358
148. Dokudovskaya, S., Williams, R., Devos, D., Sali, A., Chait, B. T., and Rout, M. P. (2006) Protease accessibility laddering: a proteomic tool for probing protein structure. *Structure* **14**, 653–660
149. Guan, J. Q., and Chance, M. R. (2005) Structural proteomics of macromolecular assemblies using oxidative footprinting and mass spectrometry. *Trends Biochem. Sci.* **30**, 583–592