Structure, Dynamics, Assembly, and Evolution of Protein Complexes

Joseph A. Marsh[1] and Sarah A. Teichmann[2,3]

[1]Medical Research Council Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Western General Hospital, Edinburgh EH4 2XU, United Kingdom; email: joseph.marsh@igmm.ed.ac.uk

[2]European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, United Kingdom

[3]Wellcome Trust Sanger Institute, Hinxton, Cambridge CB10 1SA, United Kingdom; email: sarah.teichmann@ebi.ac.uk

Annu. Rev. Biochem. 2015. 84:551–75

First published online as a Review in Advance on December 8, 2014

The *Annual Review of Biochemistry* is online at biochem.annualreviews.org

This article's doi: 10.1146/annurev-biochem-060614-034142

Keywords

quaternary structure, homomer, heteromer, oligomer, multimer, self-assembly, protein interactions

Abstract

The assembly of individual proteins into functional complexes is fundamental to nearly all biological processes. In recent decades, many thousands of homomeric and heteromeric protein complex structures have been determined, greatly improving our understanding of the fundamental principles that control symmetric and asymmetric quaternary structure organization. Furthermore, our conception of protein complexes has moved beyond static representations to include dynamic aspects of quaternary structure, including conformational changes upon binding, multistep ordered assembly pathways, and structural fluctuations occurring within fully assembled complexes. Finally, major advances have been made in our understanding of protein complex evolution, both in reconstructing evolutionary histories of specific complexes and in elucidating general mechanisms that explain how quaternary structure tends to evolve. The evolution of quaternary structure occurs via changes in self-assembly state or through the gain or loss of protein subunits, and these processes can be driven by both adaptive and nonadaptive influences.

Contents

INTRODUCTION 552
PRINCIPLES OF QUATERNARY STRUCTURE ORGANIZATION 553
Structural Characterization of Protein Complexes 553
Protein Self-Assembly into Homomeric Complexes 555
Quaternary Structure Diversity of Heteromeric Complexes 557
PROTEIN COMPLEX ASSEMBLY AND DYNAMICS 559
Flexibility, Disorder, and Conformational Changes Upon Protein–Protein Interaction 559
Ordered Assembly of Protein Complexes 561
Dynamic Protein Complexes 562
EVOLUTION OF QUATERNARY STRUCTURE 563
Evolutionary Changes in Self-Assembly State 563
Evolutionary Gain and Loss of Protein Complex Subunits 564
Adaptive and Nonadaptive Drivers of Quaternary Structure Evolution 566
OUTLOOK 567

INTRODUCTION

Within the crowded intracellular environment, individual proteins are constantly coming into physical contact with other proteins and biological macromolecules (1, 2). There is huge diversity in the frequency, specificity, and duration of these interactions. On one hand, a large fraction of proteins form long-lived homomeric or heteromeric assemblies that are amenable to experimental characterization and have well-defined molecular functions. On the other hand, cells are full of many short-lived, promiscuous interactions that have little biological relevance, have not been evolutionarily selected for, and are due mostly to intracellular crowding (3, 4). However, many transient interactions are of crucial functional importance, particularly in cell signaling (5, 6).

If the cell is such a dynamic place, with so many interactions occurring, then how should we define a protein complex? Do all interactions represent complexes of a sort, however transient or nonspecific? And if we want to distinguish ultratransient interactions from stable protein complexes, how do we define a threshold? Given that we are unlikely to obtain a complete picture of all interactions occurring within a cell, our concept of protein complexes depends largely on the experimental methods used to characterize them. For example, one can define a protein complex either as a collection of proteins that copurify together in a high-throughput proteomics experiment (7–9) or through the analysis of patterns within pairwise interaction data (10).

Quaternary structure: the structural arrangement of the different subunits of a protein complex, with respect to one another

Subunit: one of the individual protein chains of a complex

In this review, we focus on complexes that can be purified and characterized structurally, taking advantage of the vast amount of three-dimensional coordinate data now available. We first focus on how the quaternary structure of protein complexes is organized. We then discuss dynamic aspects of protein complexes, including their conformational changes upon binding, ordered assembly into multisubunit complexes, and dynamics that occur within the context of a fully assembled complex. Finally, we address quaternary structure evolution, which includes changes in self-assembly state, the gain and loss of subunits, and the adaptive and nonadaptive factors that influence these processes.

PRINCIPLES OF QUATERNARY STRUCTURE ORGANIZATION

Structural Characterization of Protein Complexes

Heteromer: a protein complex formed from multiple distinct subunit types

NMR: nuclear magnetic resonance

EM: electron microscopy

Monomer: an individual protein chain that is not part of a complex

Homomer: a protein complex formed from self-assembly of a single type of subunit

PDB: Protein Data Bank

Supplemental Material

Protein quaternary structure can be defined as the way the different individual protein chains (i.e., subunits) of a complex are organized with respect to one another. At its simplest, quaternary structure can be homomeric, formed from the self-assembly of repeated copies of a single subunit. In contrast, heteromeric complexes are composed of multiple distinct protein subunits, usually encoded by different genes, although heteromers can also be formed from proteolytic cleavage of single chains.

Three main experimental methods have been used to determine the structures of protein complexes: X-ray crystallography, nuclear magnetic resonance (NMR), and electron microscopy (EM). **Figure 1** depicts the rates at which new structures have been determined with each method for monomers (i.e., single protein chains that do not self-assemble), homomers, and heteromers. This figure is based upon a current snapshot of the Protein Data Bank (PDB) (11), filtered for redundancy and provided as **Supplemental Table 1** (follow the **Supplemental Material link** from the Annual Reviews home page at **http://www.annualreviews.org**).

Most protein structures have been determined using X-ray crystallography. The nonredundant snapshot of protein crystal structures contains 7,972 monomers, 9,206 homomers, and 2,677 heteromers. Thus, 87% of crystal structures involve only a single type of polypeptide chain, and a slight majority (54%) of these self-assemble into homomers. Interestingly, the number of new monomers and homomers peaked near the end of the last decade (**Figure 1*a***). This can probably be attributed to the proliferation of structural genomics projects during that decade, which focused primarily on the structures of novel protein folds (12). More recently, as structural genomics projects have been successful in determining a large fraction of the easily crystallizable protein folds, the number of new monomers and homomers per year has decreased. In contrast, it is clear that coverage of heteromeric quaternary structure space is still in its relative infancy.

NMR is the second most common method for protein structure determination. The nonredundant snapshot of NMR structures contains 3,743 monomers, 218 homomers, and 197 heteromers. Most NMR structures are composed of only a single type of polypeptide chain, although the vast majority are of monomers rather than homomers. There are more monomers because smaller proteins are much easier to characterize structurally with NMR than are larger proteins, so proteins that self-assemble into homomers are generally less amenable to structure determination. It is also conceivable that less consideration is given to correctly assigning quaternary structure during NMR structure determination, so some of the monomeric NMR structures might actually exist as homomers in solution. Moreover, it is interesting to note the dramatic increase in new monomeric NMR structures, peaking in 2005 and then dropping sharply (**Figure 1*b***). This increase is probably due to a combination of structural genomics proliferation, coupled with an increase, and then decrease, in interest in NMR as a structure-determination method.

Finally, various EM-based approaches have been used to determine protein structures. In total, there are 37 monomers, 102 homomers, and 148 heteromers in the nonredundant snapshot of EM structures. Note the large enrichment of heteromers and underrepresentation of monomers compared with other methods. The reason for this disparity is that EM is ideally suited for very large structures, which tend to be enriched in complexes, particularly heteromers with multiple distinct subunits. The number of new EM structures per year is currently increasing rapidly, facilitated by advances in cryogenic methods combined with single-particle analysis and tomography.

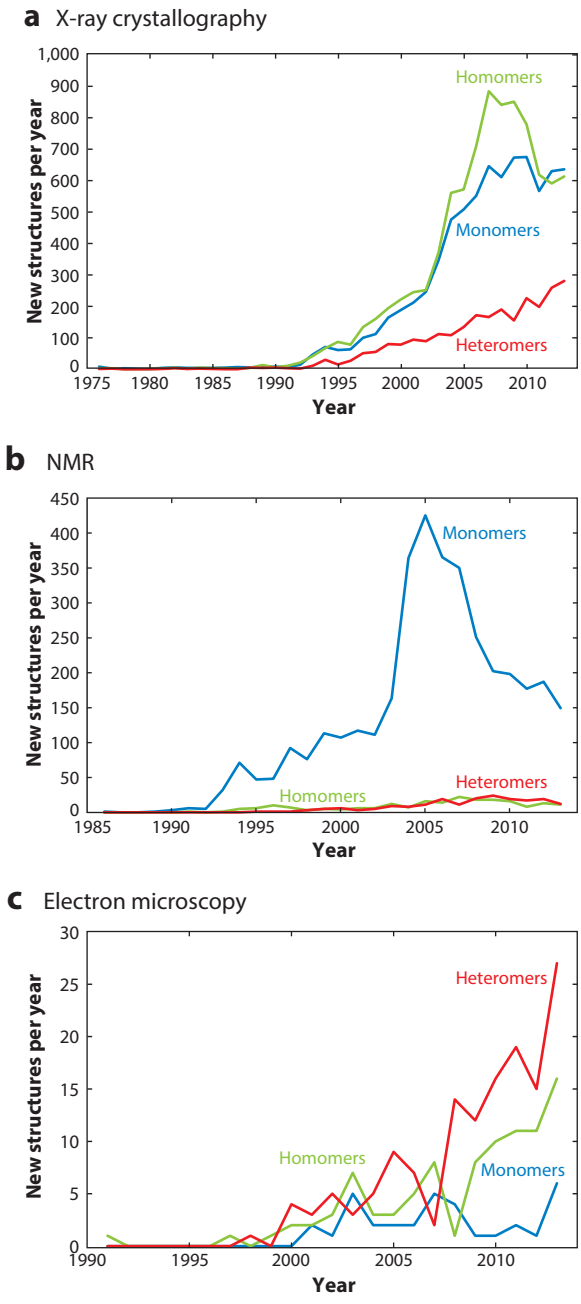

Figure 1

New nonredundant monomer (*blue*), homomer (*green*), and heteromer (*red*) structures released per year in the Protein Data Bank (PDB), determined using (*a*) X-ray crystallography, (*b*) nuclear magnetic resonance (NMR), and (*c*) electron microscopy. These plots are based upon a recent snapshot of the PDB (from June 14, 2014) and consider only new structures with sequences that differ substantially from those of existing structures (filtered at the level of 50% sequence identity). The nonredundant sets of structures used to generate these plots are provided in **Supplemental Table 1**.

Supplemental Material

Protein Self-Assembly into Homomeric Complexes

The reason there is so much more structural information on homomers than heteromers is not that people have specifically chosen to study homomers, but instead because the intrinsic propensity for individual proteins to self-assemble is high. Thus, because most structural studies have focused on individual protein chains, most protein structures are either monomeric or homomeric.

The vast majority of homomeric protein complexes of known structure are symmetric (13–16). Thus, all homomers can be classified into a limited number of groups (**Figure 2*a***):

1. Twofold dimeric complexes, which can be represented by the C_2 symmetry group. In these complexes, a pair of subunits is related by a twofold axis of rotational symmetry. An important property of such a twofold axis is that the interaction between the two subunits is symmetric and necessarily involves identical parts of their surfaces. These are called isologous or head-to-head interfaces.
2. Cyclic complexes, which belong to the $C_{n(n>2)}$ symmetry groups. These are characterized by higher-order rotational symmetry (e.g., threefold in C_3 and fourfold in C_4 complexes). The interfaces of these cyclic complexes involve asymmetric interactions between different surfaces on each subunit, and are referred to as heterologous or head-to-tail interfaces. Although the twofold dimeric complexes technically belong to a cyclic symmetry group (i.e., C_2), they are generally considered separately because they do not form closed rings, as do the higher-order cyclic complexes.
3. Dihedral complexes, from the $D_{n(n>1)}$ symmetry groups. These complexes are characterized by two orthogonal symmetry axes. One is a twofold axis of rotational symmetry, whereas the other can be twofold or higher order. Thus, a D_2 complex can essentially be considered a "dimer of dimers," whereas the D_3 symmetry group is compatible with both a "dimer of trimers" and a "trimer of dimers."
4. Cubic complexes, which can belong to the tetrahedral (*T*), octahedral (*O*), or icosahedral (*I*) symmetry group. Tetrahedral homomers have 12 subunits with twofold and threefold symmetry axes. Octahedral homomers have 24 subunits with twofold, threefold, and fourfold symmetry axes. Icosahedral complexes have 60 subunits with twofold, threefold, and fivefold symmetry axes.
5. Helical complexes, with helical (*H*) symmetry. Helical symmetry can be considered rotation combined with translational motion (i.e., like a screw). Whereas all of the above complexes belong to closed symmetry groups, helical symmetry is open; therefore, in principle, they could self-assemble indefinitely. In practice, the number of subunits in complexes with helical symmetry is controlled by other factors, such as steric restrictions, subunit concentrations, and the kinetics of association and dissociation.
6. Asymmetric homomers, which can be represented by the trivial symmetry group C_1. These can be formed in a variety of ways, but by definition in any such asymmetric complex, different subunits must exist in nonequivalent positions.

It is interesting to observe the widely varying distributions of the different symmetry groups in the sets of homomeric structures determined with different experimental measurements (**Figure 2*b***). On one hand, whereas symmetric dimers make up the majority of both X-ray crystallography and NMR structures, they are quite rare (10%) in EM structures. The EM structures, on the other hand, are strongly enriched in homomers with cyclic (34%) and cubic (32%) symmetries. This observation can be explained by the fact that EM methods are generally applied to very large structures, which are likely to have more repeated subunits. Overall, these results emphasize the necessity of utilizing a variety of techniques to obtain the best coverage of quaternary structure space.

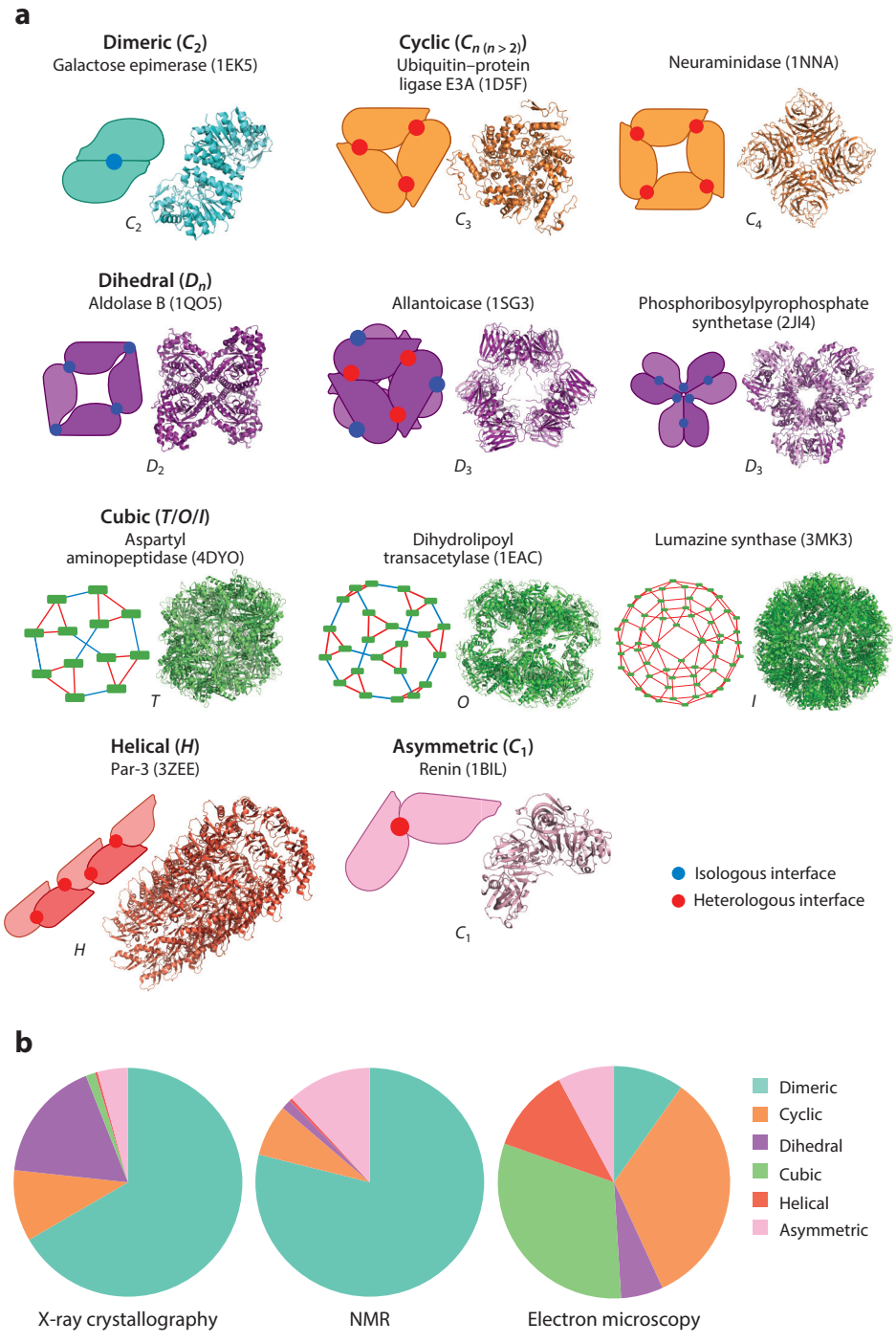

Figure 2

Diversity of quaternary structure in homomers. (*a*) Examples from the different classes of homomers. For each example, the quaternary structure topology, in which the interfaces between subunits are represented as red or blue circles or lines, is shown on the left, and a real protein structure is shown on the right. Protein Data Bank (PDB) identifiers are in parentheses. (*b*) Distribution of quaternary structure groups for nonredundant structures determined using X-ray crystallography, nuclear magnetic resonance (NMR), and electron microscopy, using the same data set as in **Figure 1**.

Finally, even among those complexes classified as symmetric on a global level, there are often deviations from perfect symmetry. For example, different copies of the same subunit within a single complex may exhibit small-to-large structural variations, although analysis of these variations can be complicated by the use of noncrystallographic symmetry constraints during structure determination (17). However, when such complexes are studied using NMR, there is generally no evidence for any asymmetry occurring in solution (18). These observations suggest that such asymmetric structural fluctuations can occur and are sometimes observed as the energetically preferred states in crystals, but they are averaged out in solution over an NMR timescale. Mixing of local symmetry elements can sometimes be observed as well. For example, in the tetrameric AMPA subtype glutamate receptor (PDB ID: 3KG2), the intramembrane channel regions adopt C_4 symmetry, whereas the extracellular domains form a pair of C_2 dimers (19).

Quaternary Structure Diversity of Heteromeric Complexes

Heteromer quaternary structure space is vast and largely unexplored. As highlighted in **Figure 1**, there are far fewer published heteromer structures compared with homomers. This situation contrasts with the scenario within cells, where most protein complexes are likely to be heteromeric. For example, there is evidence that crystal structures of protein complexes purified from native tissues tend to contain more distinct heteromeric subunits than those formed using recombinantly produced proteins (20, 21). This notion is further supported by numerous proteomic experiments, in which a very large fraction of proteins interact with other proteins in vivo (22) or copurify as part of multiprotein complexes (7–9). Even in *Mycoplasma pneumoniae*, which has one of the smallest genomes of any free-living organism, a proteome-wide screen revealed that 35% of the soluble complexes identified were homomeric, although investigators estimate that this fraction might be as high as 47% within the cell (23). Given the relatively low number of distinct proteins in its proteome and, thus, the lower number of possible heteromeric interactions that could form, we might speculate that this percentage is close to the upper limit for the fraction of homomeric complexes and, therefore, that in most other organisms heteromers are probably more common.

Despite this apparent bias against heteromers in previous structural studies, a huge diversity of heteromeric quaternary structure arrangements have been observed. **Figure 3** illustrates several examples of heteromeric complexes (15).

Perhaps the simplest heteromeric complexes are those formed from paralogous subunits (**Figure 3*a***). Such "paralogous heteromers" are fairly common, with complexes in which all subunits are paralogs comprising approximately 15% eukaryotic, 9% bacterial, and 19% archaeal heteromeric crystal structures (21). Often, these complexes adopt simple symmetric topologies that resemble symmetric homomers (**Figure 3*a***). For example, the human Rad9–Hus1–Rad1 heterotrimer has three different paralogous subunits that are arranged like a cyclic trimer. Thus, although this complex does not have perfect global symmetry and is technically categorized as C_1 (i.e., no symmetry), it does have a pseudo-C_3 symmetry. In fact, the individual polypeptide chains of this complex have internal pseudo-C_2 symmetry due to repeated domains; thus, the complex could be described as pseudo-D_3 if considered at the domain level. Another example is the archaeal group II chaperonin, the thermosome, which is formed from two paralogous subunits with overall D_4 symmetry that becomes pseudo-D_8 if we consider the structural similarity between the α- and β-subunits. However, not all paralogous heteromers adopt simple homomer-like arrangements: The Skp1 complex has 10 copies of 3 paralogous subunits arranged with overall D_5 symmetry. Because the different paralogous subunits occupy quite different positions within the complex, there is no higher-order pseudosymmetry.

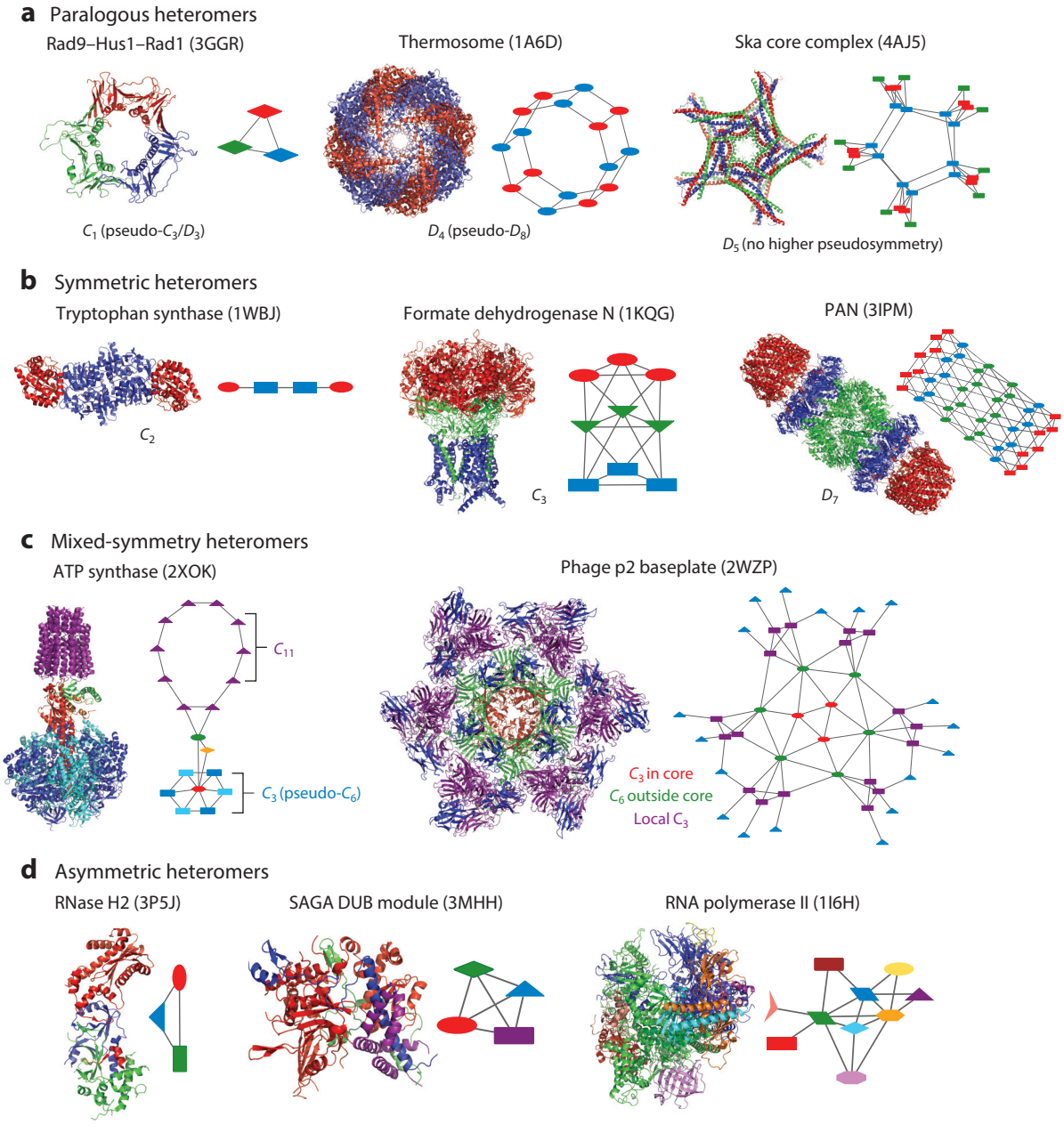

Figure 3

Diversity of heteromeric quaternary structure. (*a*) Paralogous heteromers, in which all subunits are paralogs, arising from gene duplication events. (*b*) Symmetric heteromers, in which the different subunit types are not all paralogous. (*c*) Mixed-symmetry heteromers, in which different symmetry groups are combined within a single complex. (*d*) Asymmetric heteromers, with no global or local symmetry. In the graph representations of quaternary structure, different-colored nodes indicate different types of subunits, whereas subunits from the same complex with different colors but the same shape are paralogous. Protein Data Bank (PDB) identifiers are in parentheses. Abbreviation: PAN, proteasome-activating nucleotidase.

Other heteromers adopt overall symmetric structures using nonparalogous subunits (**Figure 3*b***). For example, tryptophan synthase has two different subunits, each repeated twice with C_2 symmetry, whereas formate dehydrogenase N has three different subunits, each repeated three times with C_3 symmetry. Some complexes contain a mix of paralogous and nonparalogous subunits, as in the complex formed between the proteasome and the proteasome-activating nucleotidase (PAN) assembly, wherein the α- and β-subunits of the proteasome are similar to one another, but not to PAN. Each of the three subunit types is repeated 14 times, yielding an overall D_7 symmetry.

Some heteromers mix different elements of symmetry (**Figure 3*c***). Often, this mixing involves uneven subunit stoichiometry, in which there are different numbers of each type of subunit. For example, ATP synthase has a membrane-embedded region with C_{11} symmetry, which is connected via an asymmetric stalk to a hexamer formed from two paralogous subunits with C_3 (pseudo-C_6) symmetry. Similarly, the lactococcal phage p2 baseplate structure has a trimer with C_3 symmetry at its core, situated below a hexamer with C_6 symmetry. Moreover, there are six elements of local C_3 symmetry.

Finally, some heteromers have no symmetry at all (**Figure 3*d***). For example, RNase H2 has three distinct, asymmetrically arranged subunits, and the SAGA DUB module has four. RNA polymerase II is much more complex, with 10 different subunits, although two paralogous subunits in the core provide a degree of pseudosymmetry.

PROTEIN COMPLEX ASSEMBLY AND DYNAMICS

Flexibility, Disorder, and Conformational Changes Upon Protein–Protein Interaction

At the simplest level, protein complex assembly involves binary interactions between pairs of polypeptide chains. An important issue when considering these interactions is the extent of conformational changes that occur between the free and bound states. Interestingly, the strongest determinant of conformational change appears to be the intrinsic flexibility of the free state (24–28). Although there are exceptions, in general the more flexible a protein is as a monomer, the larger the conformational differences that will tend to be observed upon binding (29).

Unfortunately for those interested in protein flexibility and dynamics, the available structural information on monomeric proteins is heavily biased toward rigid proteins, because increasing flexibility tends to correspond with increasing difficulty of crystallization and lower-resolution crystal structures (30, 31). In fact, careful examination suggests that many of the crystal structures of apparently flexible proteins are the result of quaternary structure assignment errors, which are likely to be homomeric in solution (31, 32).

For these reasons, when considering the interactions of proteins that have had crystal structures determined in their monomeric states, the conformational changes that occur upon complex formation are usually quite minor (33). For some binary interactions involving two highly rigid proteins, there are hardly any conformational changes. For example, the interaction between cationic trypsin and buckwheat trypsin inhibitor (**Figure 4*a***) (34) reveals essentially no conformational differences between the free and bound states.

Despite the bias toward rigid proteins, many flexible proteins have still been crystallized in their free states. For example, cyclin-dependent kinase inhibitor 2 is moderately flexible (35) and undergoes substantial conformational rearrangements upon binding to cyclin A (**Figure 4*b***) (36).

An important class of protein interactions involves intrinsically disordered proteins. These proteins, which are quite common in eukaryotes, tend to be partially or completely disordered in

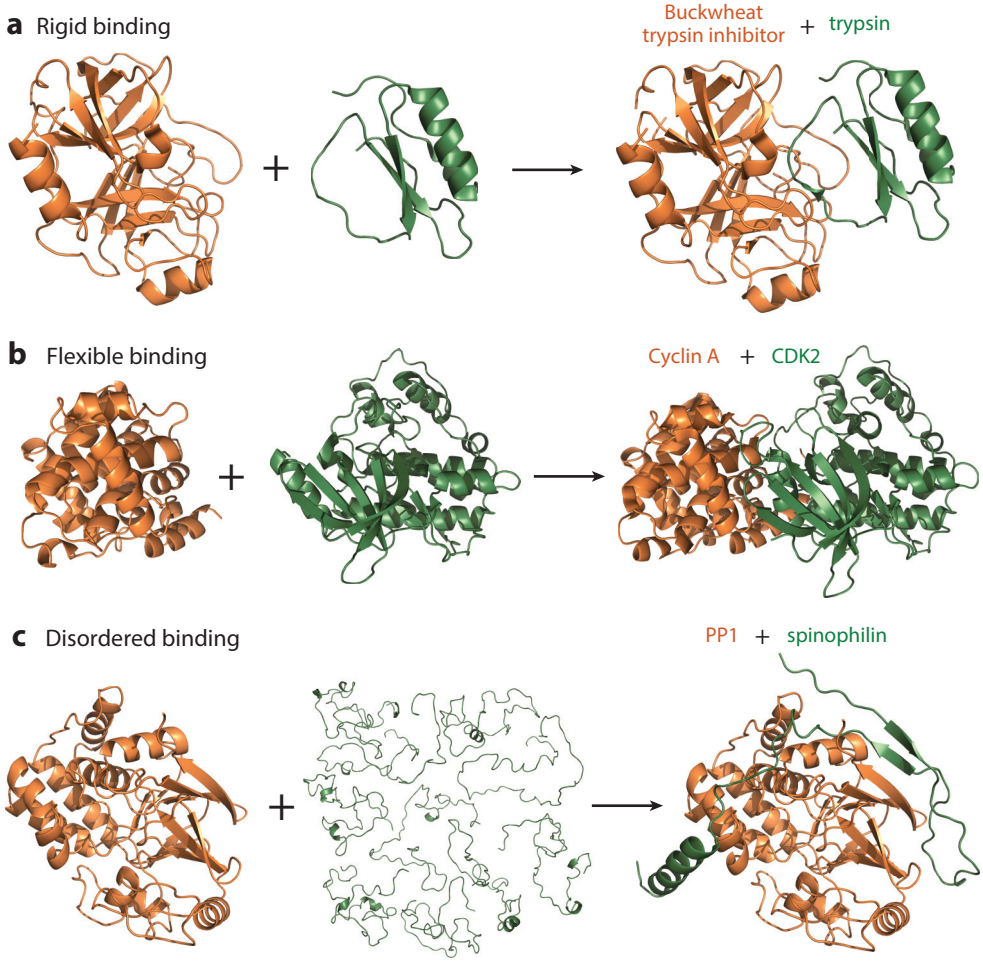

Figure 4

Protein flexibility and conformational changes upon binding. (*a*) Rigid binding between cationic trypsin and buckwheat trypsin inhibitor, where essentially no conformational changes occur in either subunit. (*b*) Flexible binding between cyclin A and CDK2, where CDK2 is moderately flexible in its unbound state and undergoes moderate conformational changes upon binding. (*c*) Disordered binding between protein phosphatase 1 (PP1) and the PP1-binding domain of spinophilin, where spinophilin is intrinsically disordered in its unbound state but undergoes a major folding transition upon binding. The Protein Data Bank (PDB) identifiers of the crystal structures depicted are 2A7H, 3RDY, 3RDZ, 2R3I, 1VIN, 2CCH, and 3EGG. No structure of unbound PP1 is available, so the bound-state structure is used. The ensemble model of spinophilin was generated from nuclear magnetic resonance and dynamic light-scattering measurements (40).

isolation when studied in vitro (37). Interestingly, however, many intrinsically disordered proteins form highly ordered structures within the context of a protein complex (38, 39). For example, a large region of spinophilin is intrinsically disordered. Although this intrinsic disorder prevents it from being crystallized, an ensemble model illustrating its extreme conformational heterogeneity (**Figure 4*c***) could be built by combining various NMR and dynamic light-scattering measurements (40, 41). However, upon interaction with protein phosphatase 1, this region of spinophilin

undergoes a huge disorder-to-order transition and folds completely, thereby allowing a crystal structure of the complex to be determined (42).

Ordered Assembly of Protein Complexes

In any complex composed of more than two subunits, assembly is necessarily more complex than a simple binary interaction. A useful analogy can be drawn from Levinthal's commonly stated folding paradox, whereby we can posit that protein complex assembly is likely to occur via an energetically favorable, ordered pathway, as the odds of all subunits simultaneously coming together in the correct orientation to form the fully assembled complex would be infinitesimally small. Therefore, the question that remains is essentially whether protein complexes assemble via a single pathway, or whether multiple parallel pathways are possible, similar to what has been proposed for protein folding (43).

Our ability to characterize protein assembly pathways has been greatly aided by the development of electrospray mass spectrometry techniques, which can be used to probe the identities of the different subcomplexes formed during solution-phase assembly and disassembly (44). These techniques can be complemented with ion mobility mass spectrometry, which allows structurally distinct intermediates with equivalent masses to be distinguished (45), and collision-induced dissociation, in which gas-phase disassembly pathways can be observed (46).

When electrospray mass spectrometry experiments were applied to a series of homomers, nearly all of the dihedral or tetrahedral complexes were observed to disassemble via a single subcomplex intermediate: either dimeric or cyclic (47). In contrast, no intermediates of cyclic complexes were observed. This result provided great insight into the assembly of protein complexes, suggesting that most homomers assemble via a single dominant pathway and that only very limited routes of assembly are possible, proceeding via intermediates with closed dimeric or cyclic symmetries (47).

For example, a monomer can assemble into a dimeric (**Figure 5*a***) or cyclic (e.g., C_3) (**Figure 5*b***) complex. Of course, as noted above, it is unlikely that a cyclic ring could instantaneously be formed from free monomeric subunits in a single step. Therefore, there are almost certainly transiently populated asymmetric assembly intermediates. However, the symmetric intermediates appear to be much more energetically favorable, and can therefore be detected experimentally.

A C_2 dimer can further dimerize to form a tetrameric D_2 complex (**Figure 5*c***) or trimerize to form a D_3 hexamer (**Figure 5*d***). A C_3 complex can dimerize to form a D_3 hexamer (**Figure 5*e***) or tetramerize to form a tetrahedral complex (**Figure 5*f***).

When heteromers have been characterized by electrospray mass spectrometry, a similarly strong tendency to detect intermediate subcomplexes has been observed, with a single (dis)assembly pathway in most cases (48, 49). However, there are interesting differences from the homomers. First, asymmetric subcomplexes are fairly common, suggesting that some of them are of comparable energetic favorability to the symmetric intermediates, at least in comparison to the hypothetical asymmetric homomeric intermediates. Second, evidence of parallel (dis)assembly has been observed, particularly for two large urease complexes with fairly complex topologies (49). Although the number of complexes studied thus far is limited, evidence to date suggests that simpler complexes may generally assemble via single pathways, whereas complexes with more complicated multisubunit topologies may be more likely to have multiple pathways. This hypothesis is consistent with the parallel pathways that have been observed for ribosome assembly (50).

An interesting observation concerning both homomeric and heteromeric assembly pathways is that one can accurately predict them if the three-dimensional structure of the complex is known (47, 49). Essentially, by assuming a correspondence between intersubunit interface size and dissociation energy, one can obtain remarkably good agreement between experiment and prediction. Although

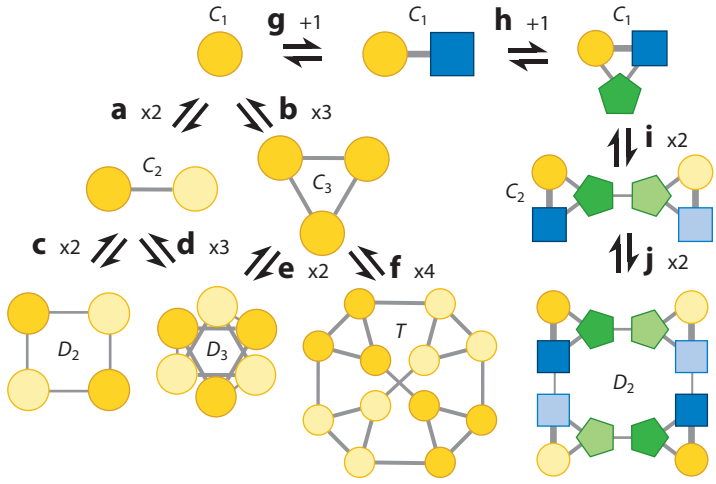

Figure 5

Multiple pathways of quaternary structure assembly and evolution. Protein complexes can assemble and evolve through changes in self-assembly state (*a–f*, *i*, *j*) or through the gain of new subunits (*g*,*h*).

the success of this approach might appear to contradict the generally low correlation between interface size and strength (51), it appears that within the context of disassembly of a single complex, in which a limited set of discrete subcomplexes might be formed at each step, the correspondence is strong enough to achieve accurate predictions (52).

Dynamic Protein Complexes

By analogy to the spectrum of flexibility observed for monomeric proteins, which ranges from highly rigid to intrinsically disordered, protein complexes can also experience a variety of dynamics within their complexed forms (29, 53–55). These dynamics are particularly obvious for the various energy-dependent molecular machines, in which large coordinated motions of different subunits are essential for carrying out a function (56). For example, ATP synthase (**Figure 3*c***) functions essentially as a rotary motor, wherein the movement of protons through the membrane-bound region drives the rotation of the stalk, which induces conformational changes in the β-subunits that drive ATP synthesis (57). In the thermosome (**Figure 3*a***), ATP hydrolysis drives conformational changes that facilitate substrate protein binding, folding, and release (58). Finally, RNA polymerase II (**Figure 3*d***) undergoes large structural rearrangements during transcription initiation and elongation (59, 60).

In addition to energy-dependent conformational changes, many proteins remain quite dynamic within their complexed forms. Previously, it was difficult to study these dynamics directly, as protein NMR becomes far more difficult as the mass of the protein or protein complex of interest increases. However, the development of methyl-TROSY (transverse relaxation optimized spectroscopy) NMR techniques has recently made the direct characterization of dynamic complexes far more accessible (61). For example, the archaeal proteasome (shown in complex with the activator PAN in **Figure 3*b***) undergoes substantial equilibrium fluctuations within its α-subunits between distinct open and closed conformations on a timescale of seconds, which seems important for controlling access to the proteolytic chamber (62–64). Similarly, slow motions were observed in the core of the archaeal exosome that appear to correlate with binding to the exosomal cap and

with RNA (65). Note that the archaeal homologs of eukaryotic complexes are commonly used for these studies because they often contain more repeated subunits (as opposed to the paralogous subunits often observed in eukaryotes), and this symmetry provides a significant advantage in NMR experiments due to increased signal strength and spectral simplicity.

At the extreme end of the dynamics spectrum are those proteins that remain highly dynamic or even partially disordered within the context of a complex. These can be considered similar to the intrinsically disordered proteins (53), and have sometimes been referred to as fuzzy complexes (54). Although an early report of disorder within a homomeric complex was later found to arise from experimental artefacts (66), several highly dynamic heteromers have since been reported. The development of ensemble modeling strategies combining various types of experimental measurements such as NMR and small-angle X-ray scattering has allowed detailed structural models of some of these complexes to be generated (40, 67, 68). For example, **Figure 6** shows an ensemble model of a complex formed between the regulatory (R) region of cystic fibrosis transmembrane conductance regulator (CFTR) and 14-3-3β (69). Rather than binding solely by use of a single part of the R region, which is intrinsically disordered in isolation (70), 14-3-3β dynamically interacts with eight different sites that are individually weak. Because seven of these eight sites are phosphorylated, this dynamic binding mechanism permits a graded response to phosphorylation (69).

In addition to the dynamics that can occur within a protein complex, quaternary structure topologies themselves can be intrinsically dynamic. For instance, many complexes exist in equilibrium between two or more different self-assembly states (e.g., monomer–dimer or dimer–tetramer) (71–73). A particularly striking example is the small heat-shock protein αB-crystallin, which self-assembles into a variety of interconverting quaternary structures, ranging from roughly 10 to 40 subunits (74–76).

EVOLUTION OF QUATERNARY STRUCTURE

Evolutionary Changes in Self-Assembly State

As discussed above, proteins can assemble into complexes with a wide variety of homomeric and heteromeric quaternary structure arrangements. There is considerable interest in understanding how these different types of quaternary structures can evolve. When considering homomers, a major question is whether a given complex evolved via a precursor, either monomeric or with different homomeric quaternary structure.

One way to trace the evolutionary history of a homomeric complex is to search for homologous proteins with differing quaternary structures (47, 77). For proteins with sequence identities greater than 90%, quaternary structure is nearly always conserved, whereas there is increasing divergence with lower sequence identity. For example, in proteins with 30–40% sequence identity, the likelihood of quaternary structure being conserved is approximately 70% (47).

Searching for homologous proteins of known structure on a large scale allowed the putative evolutionary transitions involving hundreds of homomers to be identified (47). Interestingly, a strong correspondence between evolutionary pathway and homomer symmetry has been observed. It appears that evolutionary steps that correspond to the same symmetric transitions observed for self-assembly (e.g., as shown in **Figure 5**) are strongly preferred. In other words, a monomer can evolve into a dimeric C_2 or cyclic $C_{n(n>2)}$ complex (**Figure 5*a*,*b***), and a dimeric C_2 complex can evolve into a dihedral D_n complex of varying size; but a cyclic complex can only dimerize to form a dihedral complex (excluding the rare transitions to cubic symmetries). Furthermore, nearly all of the putative evolutionary transitions corresponded perfectly with the assembly pathways that were either experimentally determined or predicted from the sizes of intersubunit interfaces (47).

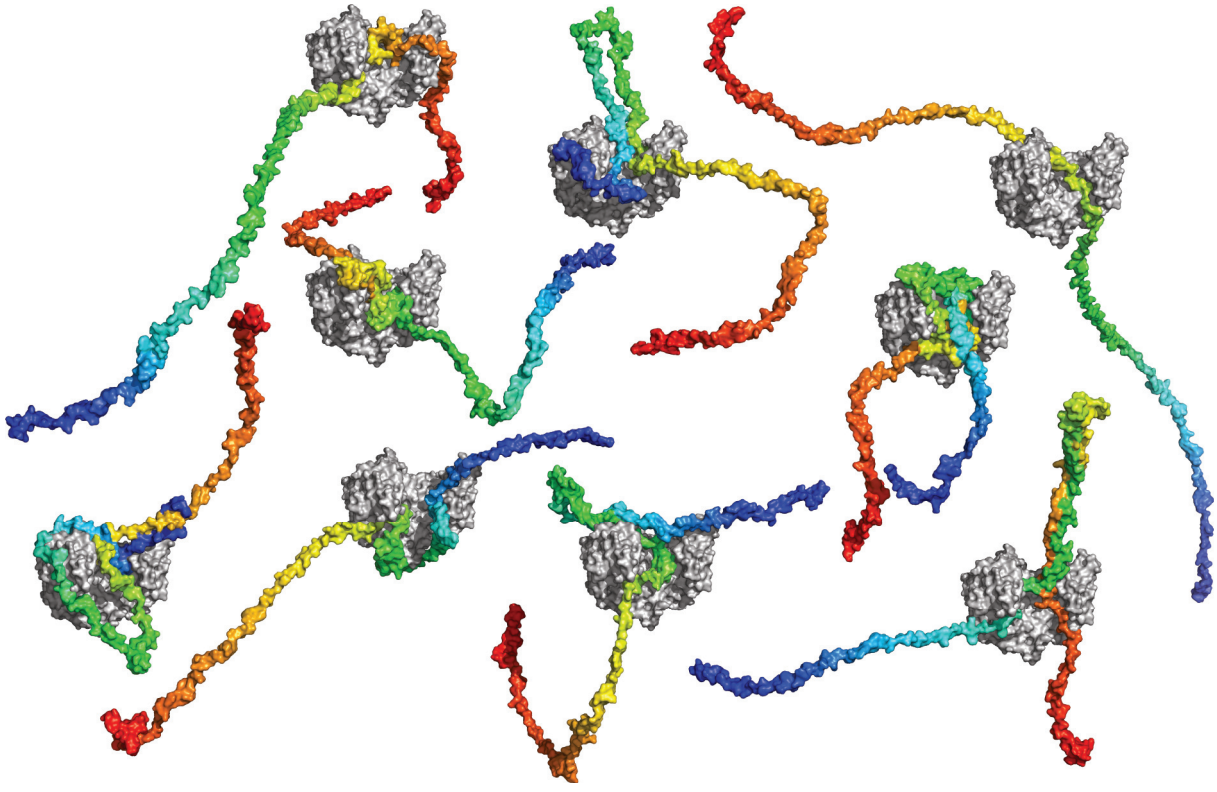

Figure 6

Ensemble model of the dynamic complex between the cystic fibrosis transmembrane conductance regulator (CFTR) regulatory region (R region), from the N terminus (*blue*) to the C terminus (*red*), and 14-3-3β homodimer (*gray*), as previously calculated (69). The 14-3-3β structure undergoes only very small conformational changes between the different ensemble members. All ensemble members are shown in the same orientation. The R region remains highly disordered in complex, undergoing only very local ordering around its binding sites.

In other words, the more ancient interfaces of homomeric complexes tend to be larger than more recently formed interfaces (78, 79). Thus, the energetically favorable subcomplexes formed in solution during assembly have a very strong tendency to correlate with evolutionary precursors (52).

A few different mechanisms have been identified by which proteins can vary their self-assembly state. Quaternary structure can change due to simple point mutations, which either can be located directly in an intersubunit interface or exert an indirect, allosteric effect by being located away from any interface (77, 80, 81). Sometimes these mutations cause a monomer to exchange an element of its internal tertiary structure with another copy of itself, thus forming a "domain-swapped" dimer (82). Overall, this type of domain swapping has been observed in 5% of protein families (83), whereas 24% of homomers have been identified as having intertwined structures, which include domain swapping as well as exchange of shorter segments (84). Finally, quaternary structure can also be modulated by insertions and deletions, particularly at the interface region (85–88).

Evolutionary Gain and Loss of Protein Complex Subunits

In addition to changes in self-assembly state, protein complex evolution can occur through the gain and loss of subunits. A very common mechanism for acquiring new subunits involves gene

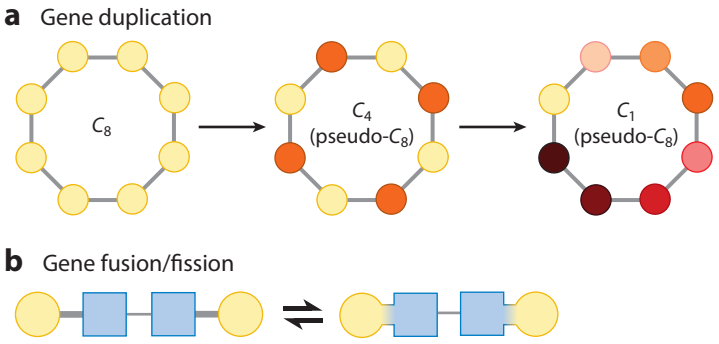

Figure 7

Two mechanisms for heteromer evolution. (*a*) Duplication of a gene encoding a protein that assembles into a homomeric complex can lead to a paralogous heteromer. Multiple duplication events can occur so that all subunits become distinct paralogs. (*b*) Gene fusion can result in permanent, covalent association between two previously separate subunits of the same complex. The reverse process of gene fission can also occur, although this is less common.

duplication events (89–93). If a gene encoding a protein that self-assembles into a homomeric complex is duplicated, initially the two identical genes will still form the same complex, consisting of mixtures of identical molecules from both genes. However, as the two genes diverge in sequence and become paralogs, the two slightly different proteins will then be able to assemble into paralogous heteromers containing the products of both genes. Should mutations occur that increase the energetic favorability of the heteromer relative to the homomers, then eventually the complex could become exclusively heteromeric. This phenomenon is particularly evident in eukaryotic complexes, in which multiple gene duplication events often occur, resulting in several paralogs of a single protein. This process can cause each subunit of a cyclic ring (or half a dihedral complex) to be encoded by a separate gene. For example, the eukaryotic forms of the archaeal thermosome and proteasome shown in **Figure 3** are encoded by 8 and 14 paralogous genes, respectively. **Figure 7*a*** illustrates the process of gene duplication leading to the evolution of a paralogous heteromer.

Although paralogous heteromers are fairly common, most heteromers in bacteria, archaea, and eukaryotes are not composed entirely of paralogous subunits (21). For these complexes, it is likely that new subunits were generally acquired sequentially, over the course of evolution, as the likelihood of multiple new subunits simultaneously being able to assemble into a heteromeric complex seems low. Inferring evolutionary gain of subunits by comparing different protein structures, as was done for homomer evolution (47), is challenging given the nature of protein structural data. The presence of a specific subunit in a given complex, but not in a homologous complex in a different species, does not necessarily indicate a genuine evolutionary difference. Instead, it often reflects the experimenter's choice of proteins that are amenable to being structurally characterized together.

Sequence data are much more abundant than structure data, so by comparing the genes present in different organisms of varying evolutionary relatedness, one can reconstruct some aspects of the stepwise evolution (94, 95). For example, a detailed study of NADH–ubiquinone oxidoreductase (Complex I) attempted to trace its evolution from bacteria, in which it has 14 distinct subunits, to the mitochondria of mammals, in which it has 46 (94). The authors of this study were able to date the likely relative evolutionary ages of different subunits of the mammalian complex and to track six subunits back to the bacterial ancestor of the mitochondria that were subsequently lost in bacteria. Interestingly, all of the new subunits appear to have been acquired in a stepwise manner,

in contrast with the early evolution of the complex, in which separate heteromers appear to have been combined to form a chimera (96, 97).

A similar comparative genomic approach was used to identify the evolutionarily older and newer subunits of human complexes (21). Strikingly, there was a very strong tendency for the evolutionarily more recent subunits of a given complex to be more flexible than the evolutionarily more ancient subunits. The reason appears to be that greater intrinsic flexibility of subunits greatly facilitates the assembly of heteromeric complexes with more distinct subunits. In other words, the more distinct subunit types there are within a complex, the more flexible those subunits will tend to be (21). Thus, as new subunits are sequentially added to a complex in evolution, they will need to be more flexible than existing subunits. This research suggests that the relative flexibilities of the different subunits of a heteromer, as can be assessed from the complex structure (28, 29, 31), may be useful in future evolutionary reconstructions.

Closely related to the issue of heteromer evolution is how strongly protein interactions are conserved across different species. The results of several studies based upon high-throughput experiments have suggested that protein interactomes have surprisingly weak overall conservation (98–103). Some of these results are undoubtedly due to the abundance of false positives and nonfunctional interactions detected in proteomic experiments (4). Interestingly, when membership in a protein complex is accounted for, the tendency for interactions to be conserved appears to be much higher. In particular, when considering pairs of human proteins that interact within the same complex, 90% appear to copurify together in yeast, when orthologs of both human proteins are present (104).

Finally, the quaternary structure of a complex can change through the evolutionary process of gene fusion, in which two previously separate genes fuse into one open reading frame. Gene fusion has been studied extensively as a mechanism for the evolution of multidomain proteins (105–108). Interestingly, gene fusion commonly involves pairs of genes encoding subunits of the same protein complex (109, 110). Thus, fusion provides a mechanism by which the number of distinct subunits within a complex can be reduced without disrupting the overall structure or assembly of the complex (20, 49, 78). **Figure 7*b*** shows the process of gene fusion within a protein complex. The reverse process of gene fission, in which a single gene splits into two, can also occur as a mechanism for subunit gain, although it is much less common than fusion (111, 112).

Adaptive and Nonadaptive Drivers of Quaternary Structure Evolution

Why have proteins evolved to assemble into complexes? For heteromers, the potential functional benefits are obvious; they include the bringing together of distinct functions on different polypeptide chains and the tremendous regulatory potential afforded by doing so. It is also generally easier to assemble a heteromeric complex from multiple smaller subunits than it is to fold a very large protein containing all of the subunits covalently fused together (14).

Numerous functional benefits associated with homomerization have also been postulated, which could have led to strong positive selection for proteins that self-assemble. These functional advantages have been reviewed elsewhere (14, 20), but we briefly list some here:

1. The clearest link between function and self-assembly is when the interfaces formed between identical subunits participate directly in protein function. For example, active sites (113, 114) and ligand-binding sites (115) can be formed at homomeric interfaces. Moreover, ligand binding can sometimes induce self-assembly (116).
2. Allostery is commonly associated with homomeric complexes and provides a simple mechanism for achieving cooperative function. Although allostery is possible in monomeric

proteins (117), the earliest models of allostery involved symmetric homomers (118, 119), and numerous examples have been identified. A strong association between self-assembly and allosteric regulation has been continuously observed, particularly in complexes with dihedral symmetry (14).

3. Large structures are often required for morphological purposes within the cell. Assembling a large homomeric complex from multiple copies of a small protein subunit may sometimes be simpler than using a single, much larger protein. Because the smaller protein is encoded using less genetic material, this may represent a form of coding economy (120).
4. Stability has often been proposed as a benefit of self-assembly. In particular, a large homomer has a much smaller surface-to-volume ratio than that of free monomers, which may lead to enhanced stability (14, 121). Exposing a smaller fraction of surface residues also has the potential benefit of minimizing promiscuous interactions with other proteins (3, 122).

Although the advantages and sometimes necessity of protein self-assembly are often clear, Lynch (123) cautioned that speculation on the functional benefits has often run ahead of the evidence. He emphasized that in some cases, nonadaptive, stochastic processes may adequately explain evolutionary variation in homomerization, so great care should be taken before ascribing an adaptive purpose to protein self-assembly (123, 124). Furthermore, it seems that a strong propensity for self-assembly may be an intrinsic property of proteins, due to the inherent energetic favorability of symmetric, self-associating interfaces (125, 126). This idea is also consistent with the commonly observed tendencies of proteins to crystallize (127) and aggregate (128).

The evolution of heteromers may also often be nonadaptive, especially for the paralogous heteromers discussed above, which have evolved from homomers via gene duplications followed by divergence (**Figure 7*a***). Essentially, this phenomenon may represent a unidirectional evolutionary "ratchet" (89, 129, 130). That is, after gene duplication, random neutral mutations begin to accumulate with no effect on protein function. At first, a functional complex might be able to form from a random sampling of the two closely related proteins. Eventually, mutations might occur that disrupt the homomeric interactions, yet preserve a fully functional paralogous heteromer. At this point, both paralogous genes are now necessary, and random mutations are unlikely to cause reversion to a homomeric state. Such a paralogous heteromer is easy to form, yet the process is extremely unlikely to be reversed. In fact, a recent study using ancestral gene resurrection provided direct evidence for this model (131). Therefore, such nonadaptive processes might account for the high proportion of such paralogous heteromers observed in eukaryotes.

It is also possible that nonadaptive factors could drive the evolution of nonparalogous heteromers. For instance, mildly deleterious mutations that disrupt the stability of individual proteins could be tolerated in species with small population sizes in which negative selection is weak (132), which in turn might lead to secondary selection for new interactions that rescue the stability of the protein (133).

OUTLOOK

Our understanding of protein interactions and protein complexes has progressed dramatically, yet there is still much that remains to be learned. In terms of quaternary structure, although our coverage of homomeric complexes is very good, the number of heteromer structures is still quite low. However, the number of new heteromeric structures per year is still increasing at a fast rate. Thus, these existing experimental methods, as well as hybrid strategies integrating different experimental techniques (134, 135), are likely to play an important role in the elucidation of heteromer quaternary structure space in years to come. There is also a strong need to develop methods for modeling complexes with diverse quaternary structures. Considerable progress has

been made in terms of modeling binary protein–protein interactions (136–139), and some research has begun to model multiprotein assemblies (126, 140, 141). Further advances can be made by combining these approaches with the principles of quaternary structure organization and by integrating proteomic and lower-resolution structural data [e.g., chemical cross-linking provides some structural constraints (142)]. Such research will be tremendously beneficial for increasing our coverage of quaternary structure space without having to directly solve the structure of every possible heteromeric complex.

Another critical aspect lies in improving our understanding of protein complexes within cells. Advances in cryo–electron tomography are hugely exciting in this regard for the ability to characterize proteins in situ (143). There is also great potential in trying to understand how assembly occurs in vivo. The fact that in vitro assembly pathways have a strong tendency to be evolutionarily conserved strongly suggests that similar pathways are often occurring within the cell (49). However, many questions remain, such as the role of chaperones and the location within cells where assembly occurs. For example, recent research has suggested that a large fraction of protein complexes may assemble cotranslationally (144). The advancement of experimental techniques for characterizing proteins in vivo combined with new computational methods will continue to greatly improve our knowledge of protein complex biology in the near- to long-term future.

Finally, further study of protein complexes has potential medical and practical benefits in the short to long term. For example, it is very interesting that disease-associated mutations tend to be enriched in the interface regions between two subunits (145), whereas different mutations of the same protein can cause different diseases if they disrupt interfaces with different binding partners (146). Therefore, a better understanding of quaternary structure and assembly could improve predictions of the phenotypic effects of mutations and the prioritization of potential causative variants. Furthermore, understanding assembly could aid in the engineering of protein complexes (147) or the development of new drugs to inhibit complex assembly (148) or induce disassembly (149).

SUMMARY POINTS

1. Protein complexes can adopt a wide variety of homomeric and heteromeric quaternary structures that can be characterized in detail using X-ray crystallography, NMR, and EM.
2. Homomers are mostly symmetric, whereas heteromers can be symmetric, asymmetric or a mixture of the two.
3. A variety of dynamics are integral to protein complexes, including conformational changes and folding ranging from binding to ordered multistep assembly processes, conformational variability within fully assembled complexes, and intrinsic fluctuations between different quaternary structure states.
4. Quaternary structure can evolve in various ways, including through changes in self-assembly state or through gain and loss of heteromeric subunits.
5. Both adaptive and nonadaptive factors appear to be important for the evolution of protein complexes.

FUTURE ISSUES

1. How similar are protein complex quaternary structure and assembly pathways observed in vitro to what is actually occurring within the cell?
2. Can we better quantify how often proteins participate in weak interactions within the cell and, furthermore, how often are they important for function?
3. Where does assembly occur within the cell, and how often does assembly occur cotranslationally, that is, while at least one protein is still in the process of being translated?
4. Can we quantify the extent to which evolutionary changes in quaternary structure are driven by adaptive versus nonadaptive influences?

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

We thank Dinesh Soares and Andrew Wood for helpful comments on the manuscript and Julie Forman-Kay and Mickaël Krzeminski for providing the CFTR R-region ensemble mode.

LITERATURE CITED

1. Ellis RJ. 2001. Macromolecular crowding: an important but neglected aspect of the intracellular environment. *Curr. Opin. Struct. Biol.* 11:114–19
2. McGuffee SR, Elcock AH. 2010. Diffusion, crowding and protein stability in a dynamic molecular model of the bacterial cytoplasm. *PLOS Comput. Biol.* 6:e1000694
3. Levy ED, De S, Teichmann SA. 2012. Cellular crowding imposes global constraints on the chemistry and evolution of proteomes. *PNAS* 109:20461–66
4. Landry CR, Levy ED, Abd Rabbo D, Tarassov K, Michnick SW. 2013. Extracting insight from noisy cellular networks. *Cell* 155:983–89
5. Perkins JR, Diboun I, Dessailly BH, Lees JG, Orengo C. 2010. Transient protein–protein interactions: structural, functional, and network properties. *Structure* 18:1233–43
6. Tompa P, Davey NE, Gibson TJ, Babu MM. 2014. A million peptide motifs for the molecular biologist. *Mol. Cell* 55:161–69
7. Krogan NJ, Cagney G, Yu H, Zhong G, Guo X, et al. 2006. Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. *Nature* 440:637–43
8. Gavin A-C, Aloy P, Grandi P, Krause R, Boesche M, et al. 2006. Proteome survey reveals modularity of the yeast cell machinery. *Nature* 440:631–36
9. Havugimana PC, Hart GT, Nepusz T, Yang H, Turinsky AL, et al. 2012. A census of human soluble protein complexes. *Cell* 150:1068–81
10. Bader GD, Hogue CW. 2003. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinform.* 4:2
11. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, et al. 2000. The Protein Data Bank. *Nucleic Acids Res.* 28:235–42
12. Khafizov K, Madrid-Aliste C, Almo SC, Fiser A. 2014. Trends in structural coverage of the protein universe and the impact of the protein structure initiative. *PNAS* 111:3733–38

13. Klotz IM, Langebman N, Dahnall D. 1970. Quaternary structure of proteins. *Annu. Rev. Biochem.* 39:25–62
14. Goodsell DS, Olson AJ. 2000. Structural symmetry and protein function. *Annu. Rev. Biophys. Biomol. Struct.* 29:105–53
15. Levy ED, Pereira-Leal JB, Chothia C, Teichmann SA. 2006. 3D complex: a structural classification of protein complexes. *PLOS Comput. Biol.* 2:e155
16. Venkatakrishnan AJ, Levy ED, Teichmann SA. 2010. Homomeric protein complexes: evolution and assembly. *Biochem. Soc. Trans.* 38:879–82
17. Kleywegt GJ. 1996. Use of non-crystallographic symmetry in protein structure refinement. *Acta Crystallogr. D* 52:842–57
18. Godoy-Ruiz R, Krejcirikova A, Gallagher DT, Tugarinov V. 2011. Solution NMR evidence for symmetry in functionally or crystallographically asymmetric homodimers. *J. Am. Chem. Soc.* 133:19578–81
19. Sobolevsky AI, Rosconi MP, Gouaux E. 2009. X-ray structure, symmetry and mechanism of an AMPA-subtype glutamate receptor. *Nature* 462:745–56
20. Perica T, Marsh JA, Sousa FL, Natan E, Colwell LJ, et al. 2012. The emergence of protein complexes: quaternary structure, dynamics and allostery. *Biochem. Soc. Trans.* 40:475–91
21. Marsh JA, Teichmann SA. 2014. Protein flexibility facilitates quaternary structure assembly and evolution. *PLOS Biol.* 12:e1001870
22. Tarassov K, Messier V, Landry CR, Radinovic S, Serna Molina MM, et al. 2008. An in vivo map of the yeast protein interactome. *Science* 320:1465–70
23. Kühner S, Van Noort V, Betts MJ, Leo-Macias A, Batisse C, et al. 2009. Proteome organization in a genome-reduced bacterium. *Science* 326:1235–40
24. Tama F, Sanejouand YH. 2001. Conformational change of proteins arising from normal mode calculations. *Protein Eng.* 14:1–6
25. Tobi D, Bahar I. 2005. Structural changes involved in protein binding correlate with intrinsic motions of proteins in the unbound state. *PNAS* 102:18908–13
26. Dobbins SE, Lesk VI, Sternberg MJE. 2008. Insights into protein flexibility: the relationship between normal modes and conformational change upon protein–protein docking. *PNAS* 105:10390–95
27. Bakan A, Bahar I. 2009. The intrinsic dynamics of enzymes plays a dominant role in determining the structural changes induced upon inhibitor binding. *PNAS* 106:14349–54
28. Marsh JA, Teichmann SA. 2011. Relative solvent accessible surface area predicts protein conformational changes upon binding. *Structure* 19:859–67
29. Marsh JA, Teichmann SA, Forman-Kay JD. 2012. Probing the diverse landscape of protein flexibility and binding. *Curr. Opin. Struct. Biol.* 22:643–50
30. Yu L, Reutzel-Edens SM, Mitchell CA. 2000. Crystallization and polymorphism of conformationally flexible molecules: problems, patterns, and strategies. *Org. Process Res. Dev.* 4:396–402
31. Marsh JA. 2013. Buried and accessible surface area control intrinsic protein flexibility. *J. Mol. Biol.* 425:3250–63
32. Levy ED. 2007. PiQSi: protein quaternary structure investigation. *Structure* 15:1364–67
33. Stein A, Rueda M, Panjkovich A, Orozco M, Aloy P. 2011. A systematic study of the energetics involved in structural changes upon association and connectivity in protein interaction networks. *Structure* 19:881–89
34. Wang L, Zhao F, Li M, Zhang H, Gao Y, et al. 2011. Conformational changes of rBTI from buckwheat upon binding to trypsin: implications for the role of the P_8' residue in the potato inhibitor I family. *PLOS ONE* 6:e20950
35. Bartova I, Koca J, Otyepka M. 2008. Functional flexibility of human cyclin-dependent kinase 2 and its evolutionary conservation. *Protein Sci.* 17:22–33
36. Cheng K-Y, Noble MEM, Skamnaki V, Brown NR, Lowe ED, et al. 2006. The role of the phospho-CDK2/cyclin A recruitment site in substrate recognition. *J. Biol. Chem.* 281:23167–79
37. Wright PE, Dyson HJ. 1999. Intrinsically unstructured proteins: re-assessing the protein structure–function paradigm. *J. Mol. Biol.* 293:321–31
38. Dyson HJ, Wright PE. 2002. Coupling of folding and binding for unstructured proteins. *Curr. Opin. Struct. Biol.* 12:54–60

39. Hegyi H, Schad E, Tompa P. 2007. Structural disorder promotes assembly of protein complexes. *BMC Struct. Biol.* 7:65
40. Marsh JA, Dancheck B, Ragusa MJ, Allaire M, Forman-Kay JD, Peti W. 2010. Structural diversity in free and bound states of intrinsically disordered protein phosphatase 1 regulators. *Structure* 18:1094–103
41. Krzeminski M, Marsh JA, Neale C, Choy W-Y, Forman-Kay JD. 2013. Characterization of disordered proteins with ENSEMBLE. *Bioinformatics* 29:398–99
42. Ragusa MJ, Dancheck B, Critton DA, Nairn AC, Page R, Peti W. 2010. Spinophilin directs protein phosphatase 1 specificity by blocking substrate binding sites. *Nat. Struct. Mol. Biol.* 17:459–64
43. Dill KA, Chan HS. 1997. From Levinthal to pathways to funnels. *Nat. Struct. Biol.* 4:10–19
44. Hernández H, Robinson CV. 2007. Determining the stoichiometry and interactions of macromolecular assemblies from mass spectrometry. *Nat. Protoc.* 2:715–26
45. Ruotolo BT, Benesch JLP, Sandercock AM, Hyung S-J, Robinson CV. 2008. Ion mobility–mass spectrometry analysis of large protein complexes. *Nat. Protoc.* 3:1139–52
46. Hall Z, Hernández H, Marsh JA, Teichmann SA, Robinson CV. 2013. The role of salt bridges, charge density, and subunit flexibility in determining disassembly routes of protein complexes. *Structure* 21:1325–37
47. Levy ED, Erba EB, Robinson CV, Teichmann SA. 2008. Assembly reflects evolution of protein complexes. *Nature* 453:1262–65
48. Hall Z, Politis A, Robinson CV. 2012. Structural modeling of heteromeric protein complexes from disassembly pathways and ion mobility–mass spectrometry. *Structure* 20:1596–609
49. Marsh JA, Hernández H, Hall Z, Ahnert S, Perica T, et al. 2013. Protein complexes are under evolutionary selection to assemble via ordered pathways. *Cell* 153:461–70
50. Talkington MWT, Siuzdak G, Williamson JR. 2005. An assembly landscape for the 30S ribosomal subunit. *Nature* 438:628–32
51. Brooijmans N, Sharp KA, Kuntz ID. 2002. Stability of macromolecular complexes. *Proteins* 48:645–53
52. Marsh JA, Teichmann SA. 2014. Parallel dynamics and evolution: protein conformational fluctuations and assembly reflect evolutionary changes in sequence and structure. *BioEssays* 36:209–18
53. Mittag T, Kay LE, Forman-Kay JD. 2010. Protein dynamics and conformational disorder in molecular recognition. *J. Mol. Recognit.* 23:105–16
54. Tompa P, Fuxreiter M. 2008. Fuzzy complexes: polymorphism and structural disorder in protein–protein interactions. *Trends Biochem. Sci.* 33:2–8
55. Fuxreiter M, Tóth-Petróczy Á, Kraut DA, Matouschek AT, Lim RYH, et al. 2014. Disordered proteinaceous machines. *Chem. Rev.* 114:6806–43
56. Alberts B. 1998. The cell as a collection of protein machines: preparing the next generation of molecular biologists. *Cell* 92:291–94
57. Stock D, Leslie AG, Walker JE. 1999. Molecular architecture of the rotary motor in ATP synthase. *Science* 286:1700–5
58. Ditzel L, Löwe J, Stock D, Stetter KO, Huber H, et al. 1998. Crystal structure of the thermosome, the archaeal chaperonin and homolog of CCT. *Cell* 93:125–38
59. Cramer P, Bushnell DA, Kornberg RD. 2001. Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science* 292:1863–76
60. Hahn S. 2004. Structure and mechanism of the RNA polymerase II transcription machinery. *Nat. Struct. Mol. Biol.* 11:394–403
61. Sprangers R, Velyvis A, Kay LE. 2007. Solution NMR of supramolecular complexes: providing new insights into function. *Nat. Methods* 4:697–703
62. Sprangers R, Kay LE. 2007. Quantitative dynamics and binding studies of the 20S proteasome by NMR. *Nature* 445:618–22
63. Religa TL, Sprangers R, Kay LE. 2010. Dynamic regulation of archaeal proteasome gate opening as studied by TROSY NMR. *Science* 328:98–102
64. Latham MP, Sekhar A, Kay LE. 2014. Understanding the mechanism of proteasome 20S core particle gating. *PNAS* 111:5532–37

65. Audin MJC, Dorn G, Fromm SA, Reiss K, Schütz S, et al. 2013. The archaeal exosome: identification and quantification of site-specific motions that correlate with cap and RNA binding. *Angew. Chem. Int. Ed. Engl.* 52:8312–16
66. Nourse A, Mittag T. 2014. The cytoplasmic domain of the T-cell receptor ζ subunit does not form disordered dimers. *J. Mol. Biol.* 426:62–70
67. Wells M, Tidow H, Rutherford TJ, Markwick P, Jensen MR, et al. 2008. Structure of tumor suppressor p53 and its intrinsically disordered N-terminal transactivation domain. *PNAS* 105:5762–67
68. Mittag T, Marsh J, Grishaev A, Orlicky S, Lin H, et al. 2010. Structure/function implications in a dynamic complex of the intrinsically disordered SIC1 with the CDC4 subunit of an SCF ubiquitin ligase. *Structure* 18:494–506
69. Bozoky Z, Krzeminski M, Muhandiram R, Birtley JR, Al-Zahrani A, et al. 2013. Regulatory R region of the CFTR chloride channel is a dynamic integrator of phospho-dependent intra- and intermolecular interactions. *PNAS* 110:e4427–36
70. Baker JMR, Hudson RP, Kanelis V, Choy W-Y, Thibodeau PH, et al. 2007. CFTR regulatory region interacts with NBD1 predominantly via multiple transient helices. *Nat. Struct. Mol. Biol.* 14:738–45
71. Deprez E, Tauc P, Leh H, Mouscadet JF, Auclair C, Brochon JC. 2000. Oligomeric states of the HIV-1 integrase as measured by time-resolved fluorescence anisotropy. *Biochemistry* 39:9275–84
72. Jaffe EK. 2005. Morpheeins—a new structural paradigm for allosteric regulation. *Trends Biochem. Sci.* 30:490–97
73. Dey S, Pal A, Chakrabarti P, Janin J. 2010. The subunit interfaces of weakly associated homodimeric proteins. *J. Mol. Biol.* 398:146–60
74. Haley DA, Horwitz J, Stewart PL. 1998. The small heat-shock protein, αB-crystallin, has a variable quaternary structure. *J. Mol. Biol.* 277:27–35
75. Aquilina JA, Benesch JLP, Bateman OA, Slingsby C, Robinson CV. 2003. Polydispersity of a mammalian chaperone: Mass spectrometry reveals the population of oligomers in αB-crystallin. *PNAS* 100:10611–16
76. Stengel F, Baldwin AJ, Painter AJ, Jaya N, Basha E, et al. 2010. Quaternary dynamics and plasticity underlie small heat shock protein chaperone function. *PNAS* 107:2007–12
77. Perica T, Chothia C, Teichmann SA. 2012. Evolution of oligomeric state through geometric coupling of protein interfaces. *PNAS* 109:8127–32
78. Kim WK, Henschel A, Winter C, Schroeder M. 2006. The many faces of protein–protein interactions: a compendium of interface geometry. *PLOS Comput. Biol.* 2:e124
79. Dayhoff JE, Shoemaker BA, Bryant SH, Panchenko AR. 2010. Evolution of protein binding modes in homooligomers. *J. Mol. Biol.* 395:860–70
80. Malay AD, Allen KN, Tolan DR. 2005. Structure of the thermolabile mutant aldolase B, A149P: molecular basis of hereditary fructose intolerance. *J. Mol. Biol.* 347:135–44
81. Nishi H, Ota M. 2010. Amino acid substitutions at protein–protein interfaces that modulate the oligomeric state. *Proteins* 78:1563–74
82. Bennett MJ, Schlunegger MP, Eisenberg D. 1995. 3D domain swapping: a mechanism for oligomer assembly. *Protein Sci.* 4:2455–68
83. Huang Y, Cao H, Liu Z. 2012. Three-dimensional domain swapping in the protein structure space. *Proteins* 80:1610–19
84. Mackinnon SS, Malevanets A, Wodak SJ. 2013. Intertwined associations in structures of homooligomeric proteins. *Structure* 21:638–49
85. Akiva E, Itzhaki Z, Margalit H. 2008. Built-in loops allow versatility in domain–domain interactions: lessons from self-interacting domains. *PNAS* 105:13292–97
86. Hashimoto K, Panchenko AR. 2010. Mechanisms of protein oligomerization, the critical role of insertions and deletions in maintaining different oligomeric states. *PNAS* 107:20352–57
87. Hashimoto K, Madej T, Bryant SH, Panchenko AR. 2010. Functional states of homooligomers: insights from the evolution of glycosyltransferases. *J. Mol. Biol.* 399:196–206
88. Nishi H, Koike R, Ota M. 2011. Cover and spacer insertions: small nonhydrophobic accessories that assist protein oligomerization. *Proteins* 79:2372–79
89. Archibald JM, Logsdon JM Jr, Doolittle WF. 1999. Recurrent paralogy in the evolution of archaeal chaperonins. *Curr. Biol.* 9:1053–56

90. Archibald JM, Logsdon JM Jr, Doolittle WF. 2000. Origin and evolution of eukaryotic chaperonins: phylogenetic evidence for ancient duplications in CCT genes. *Mol. Biol. Evol.* 17:1456–66
91. Ispolatov I, Yuryev A, Mazo I, Maslov S. 2005. Binding properties and evolution of homodimers in protein–protein interaction networks. *Nucleic Acids Res.* 33:3629–35
92. Pereira-Leal JB, Levy ED, Kamp C, Teichmann SA. 2007. Evolution of protein complexes by duplication of homomeric interactions. *Genome Biol.* 8:R51
93. Reid AJ, Ranea JA, Orengo CA. 2010. Comparative evolutionary analysis of protein complexes in *E. coli* and yeast. *BMC Genomics* 11:79
94. Gabaldón T, Rainey D, Huynen MA. 2005. Tracing the evolution of a large protein complex in the eukaryotes, NADH:ubiquinone oxidoreductase (complex I). *J. Mol. Biol.* 348:857–70
95. Van der Sluis EO, Driessen AJM. 2006. Stepwise evolution of the sec machinery in Proteobacteria. *Trends Microbiol.* 14:105–8
96. Friedrich T. 2001. Complex I: a chimaera of a redox and conformation-driven proton pump? *J. Bioenerg. Biomembr.* 33:169–77
97. Mathiesen C, Hägerhäll C. 2003. The "antiporter module" of respiratory chain complex I includes the MRPC/NUOK subunit—a revision of the modular evolution scheme. *FEBS Lett.* 549:7–13
98. Matthews LR, Vaglio P, Reboul J, Ge H, Davis BP, et al. 2001. Identification of potential interaction networks using sequence-based searches for conserved protein–protein interactions or "interologs." *Genome Res.* 11:2120–26
99. Aloy P, Ceulemans H, Stark A, Russell RB. 2003. The relationship between sequence and interaction divergence in proteins. *J. Mol. Biol.* 332:989–98
100. Yu H, Luscombe NM, Lu HX, Zhu X, Xia Y, et al. 2004. Annotation transfer between genomes: protein–protein interologs and protein–DNA regulogs. *Genome Res.* 14:1107–18
101. Beltrao P, Serrano L. 2007. Specificity and evolvability in eukaryotic protein interaction networks. *PLOS Comput. Biol.* 3:e25
102. Lewis ACF, Jones NS, Porter MA, Deane CM. 2012. What evidence is there for the homology of protein–protein interactions? *PLOS Comput. Biol.* 8:e1002645
103. Andreani J, Guerois R. 2014. Evolution of protein interactions: from interactomes to interfaces. *Arch. Biochem. Biophys.* 554:65–75
104. Van Dam TJP, Snel B. 2008. Protein complex evolution does not involve extensive network rewiring. *PLOS Comput. Biol.* 4:e1000132
105. Björklund AK, Ekman D, Light S, Frey-Skött J, Elofsson A. 2005. Domain rearrangements in protein evolution. *J. Mol. Biol.* 353:911–23
106. Pasek S, Risler J-L, Brézellec P. 2006. Gene fusion/fission is a major contributor to evolution of multi-domain bacterial proteins. *Bioinformatics* 22:1418–23
107. Buljan M, Frankish A, Bateman A. 2010. Quantifying the mechanisms of domain gain in animal proteins. *Genome Biol.* 11:R74
108. Marsh JA, Teichmann SA. 2010. How do proteins gain new domains? *Genome Biol.* 11:126
109. Marcotte EM, Pellegrini M, Ng HL, Rice DW, Yeates TO, Eisenberg D. 1999. Detecting protein function and protein–protein interactions from genome sequences. *Science* 285:751–53
110. Enright AJ, Iliopoulos I, Kyrpides NC, Ouzounis CA. 1999. Protein interaction maps for complete genomes based on gene fusion events. *Nature* 402:86–90
111. Kummerfeld SK, Teichmann SA. 2005. Relative rates of gene fusion and fission in multi-domain proteins. *Trends Genet.* 21:25–30
112. Fong JH, Geer LY, Panchenko AR, Bryant SH. 2007. Modeling the evolution of protein domain architectures using maximum parsimony. *J. Mol. Biol.* 366:307–15
113. Dobson RCJ, Valegård K, Gerrard JA. 2004. The crystal structure of three site-directed mutants of *Escherichia coli* dihydrodipicolinate synthase: further evidence for a catalytic triad. *J. Mol. Biol.* 338:329–39
114. Navia MA, Fitzgerald PMD, McKeever BM, Leu C-T, Heimbach JC, et al. 1989. Three-dimensional structure of aspartyl protease from human immunodeficiency virus HIV-1. *Nature* 337:615–20
115. Heldin C-H. 1995. Dimerization of cell surface receptors in signal transduction. *Cell* 80:213–23

116. Cunningham BC, Ultsch M, Vos AD, Mulkerrin MG, Clauser KR, Wells JA. 1991. Dimerization of the extracellular domain of the human growth hormone receptor by a single hormone molecule. *Science* 254:821–25
117. Gunasekaran K, Ma B, Nussinov R. 2004. Is allostery an intrinsic property of all dynamic proteins? *Proteins* 57:433–43
118. Monod J, Wyman J, Changeux J-P. 1965. On the nature of allosteric transitions: a plausible model. *J. Mol. Biol.* 12:88–118
119. Koshland DE, Némethy G, Filmer D. 1966. Comparison of experimental binding data and theoretical models in proteins containing subunits. *Biochemistry* 5:365–85
120. Crick FHC, Watson JD. 1956. Structure of small viruses. *Nature* 177:473–75
121. Bershtein S, Mu W, Wu W, Shakhnovich EI. 2012. Soluble oligomerization provides a beneficial fitness effect on destabilizing mutations. *PNAS* 109:4857–62
122. Yang J-R, Liao B-Y, Zhuang S-M, Zhang J. 2012. Protein misinteraction avoidance causes highly expressed proteins to evolve slowly. *PNAS* 109:e831–40
123. Lynch M. 2013. Evolutionary diversification of the multimeric states of proteins. *PNAS* 110:e2821–28
124. Lynch M. 2012. The evolution of multimeric protein assemblages. *Mol. Biol. Evol.* 29:1353–66
125. Lukatsky DB, Shakhnovich BE, Mintseris J, Shakhnovich EI. 2007. Structural similarity enhances interaction propensity of proteins. *J. Mol. Biol.* 365:1596–606
126. André I, Strauss CEM, Kaplan DB, Bradley P, Baker D. 2008. Emergence of symmetry in homooligomeric biological assemblies. *PNAS* 105:16148–52
127. Durbin SD, Feher G. 1996. Protein crystallization. *Annu. Rev. Phys. Chem.* 47:171–204
128. Knowles TPJ, Vendruscolo M, Dobson CM. 2014. The amyloid state and its association with protein misfolding diseases. *Nat. Rev. Mol. Cell Biol.* 15:384–96
129. Ruano-Rubio V, Fares MA. 2007. Testing the neutral fixation of hetero-oligomerism in the archaeal chaperonin CCT. *Mol. Biol. Evol.* 24:1384–96
130. Doolittle WF. 2012. Evolutionary biology: a ratchet for protein complexity. *Nature* 481:270–71
131. Finnigan GC, Hanson-Smith V, Stevens TH, Thornton JW. 2012. Evolution of increased complexity in a molecular machine. *Nature* 481:360–64
132. Lynch M. 2007. The frailty of adaptive hypotheses for the origins of organismal complexity. *PNAS* 104(Suppl. 1):8597–604
133. Fernández A, Lynch M. 2011. Non-adaptive origins of interactome complexity. *Nature* 474:502–5
134. Alber F, Förster F, Korkin D, Topf M, Sali A. 2008. Integrating diverse data for structure determination of macromolecular assemblies. *Annu. Rev. Biochem.* 77:443–77
135. Robinson CV, Sali A, Baumeister W. 2007. The molecular sociology of the cell. *Nature* 450:973–82
136. Mukherjee S, Zhang Y. 2011. Protein–protein complex structure predictions by multimeric threading and template recombination. *Structure* 19:955–66
137. Stein A, Mosca R, Aloy P. 2011. Three-dimensional modeling of protein interactions and complexes is going 'omics. *Curr. Opin. Struct. Biol.* 21:200–8
138. Kundrotas PJ, Zhu Z, Janin J, Vakser IA. 2012. Templates are available to model nearly all complexes of structurally characterized proteins. *PNAS* 109:9438–41
139. Mosca R, Céol A, Aloy P. 2013. Interactome3D: adding structural details to protein networks. *Nat. Methods* 10:47–53
140. Inbar Y, Benyamini H, Nussinov R, Wolfson HJ. 2005. Prediction of multimolecular assemblies by multiple docking. *J. Mol. Biol.* 349:435–47
141. Esquivel-Rodríguez J, Yang YD, Kihara D. 2012. Multi-LZerD: multiple protein docking for asymmetric complexes. *Proteins* 80:1818–33
142. Rappsilber J. 2011. The beginning of a beautiful friendship: cross-linking/mass spectrometry and modelling of proteins and multi-protein complexes. *J. Struct. Biol.* 173:530–40
143. Lučić V, Rigort A, Baumeister W. 2013. Cryo–electron tomography: the challenge of doing structural biology in situ. *J. Cell Biol.* 202:407–19
144. Duncan CDS, Mata J. 2011. Widespread cotranslational formation of protein complexes. *PLOS Genet.* 7:e1002398

145. David A, Razali R, Wass MN, Sternberg MJE. 2012. Protein–protein interaction sites are hot spots for disease-associated nonsynonymous SNPs. *Hum. Mutat.* 33:359–63
146. Wang X, Wei X, Thijssen B, Das J, Lipkin SM, Yu H. 2012. Three-dimensional reconstruction of protein networks provides insight into human genetic disease. *Nat. Biotechnol.* 30:159–64
147. Lai Y-T, King NP, Yeates TO. 2012. Principles for designing ordered protein assemblies. *Trends Cell Biol.* 22:653–61
148. Jubb H, Higueruelo AP, Winter A, Blundell TL. 2012. Structural biology and drug discovery for protein–protein interactions. *Trends Pharmacol. Sci.* 33:241–48
149. Kim B, Eggel A, Tarchevskaya SS, Vogel M, Prinz H, Jardetzky TS. 2012. Accelerated disassembly of IgE–receptor complexes by a disruptive macromolecular inhibitor. *Nature* 491:613–17