Published in final edited form as:

Phys Biol. 2011 June ; 8(3): 035007. doi:10.1088/1478-3975/8/3/035007.

Caught in self-interaction: evolutionary and functional mechanisms of protein homooligomerization

Kosuke Hashimoto⁺, Hafumi Nishi⁺, Stephen Bryant, and Anna R. Panchenko^{*}

National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA

Abstract

Many soluble and membrane proteins form homooligomeric complexes in a cell which are responsible for the diversity and specificity of many pathways, may mediate and regulate gene expression, activity of enzymes, ion channels, receptors, and cell adhesion processes. The evolutionary and physical mechanisms of oligomerization are very diverse and its general principles have not yet been formulated. Homooligomeric states may be conserved within certain protein subfamilies and might be important in providing specificity to certain substrates while minimizing interactions with other unwanted partners. Moreover, recent studies have led to a greater awareness that transitions between different oligomeric states may regulate protein activity and provide the switch between different pathways. In this review we summarize the biological importance of homooligomeric assemblies, physico-chemical properties of their interfaces, experimental and computational methods for their identification and prediction. We particularly focus on homooligomer evolution and describe the mechanisms to develop new specificities through the formation of different homooligomeric complexes. Finally, we discuss the possible role of oligomeric transitions in the regulation of protein activity and compile a set of experimental examples with such regulatory mechanisms.

Introduction

Recent findings have led to a greater awareness that only a small fraction of proteins function in isolation while the majority form complexes with identical or very similar chains (called “homooligomers” hereafter) or with different non-homologous chains (called “heterooligomers”). Many soluble and membrane-bound proteins form homooligomeric complexes in a cell [1–6] (Figure 1). For example, a majority of the enzymes in the BRENDA Enzyme Database [7] represent homooligomers and analysis of high-throughput protein–protein interaction networks has shown that there are significantly more self-interacting proteins than expected by chance [8]. Despite the importance and abundance of homooligomers in a cell, the mechanisms of oligomerization are not very well understood and general principles have not been formulated. One explanation for this situation comes from the ambiguity of homooligomer experimental characterization and difficulty regarding their computational prediction. Indeed, numerous papers are dedicated to the analyses of protein–protein interaction networks but the computational methods employed in these studies are not able to properly handle the self-interactions and usually neglect them. In this review we attempt to summarize the biological importance of homooligomeric assemblies, their evolution, physico-chemical properties, and role in the regulation of cellular processes.

^{*}Corresponding author Address: 8600 Rockville Pike, Building 38A 8S814, Bethesda, MD, 20894, USA, Tel: 301-435-5891, Fax: 301-480-9241, panch@ncbi.nlm.nih.gov.

⁺authors contributed equally to this work

Functional roles of homooligomers in a cell

It is difficult to overestimate the functional importance of protein homooligomerization, which provides the diversity and specificity of many pathways and may mediate and regulate gene expression, activity of enzymes, ion channels, receptors, and cell-cell adhesion processes [9–15]. It has been suggested that large assemblies consisting of many identical subunits have advantageous regulatory properties as they can undergo sensitive phase transitions [2]. Formation of homooligomers can also provide sites for allosteric regulation, generate new binding sites at interfaces to increase specificity, and increase diversity in the formation of regulatory complexes [16]. In addition, oligomerization allows proteins to form large structures without increasing genome size and provides stability, while the reduced surface area of the monomer in a complex can offer protection against denaturation [2,6,17].

Experimental characterization and computational prediction

The main experimental techniques used to study the oligomeric states of proteins include X-ray and neutron scattering, mass spectrometry, gel-filtration, dynamic light scattering, analytical ultracentrifugation, and fluorescence resonance energy transfer (FRET) [18], (see also references for Table 1). For example, analytical centrifugation and gel filtration chromatography provide data on molecular mass distribution, the subunit stoichiometry of the complexes and equilibrium constants. FRET characterizes the kinetics and dynamics of complex formation, monitoring the extent of energy transfer between donor and acceptor, while X-ray and neutron scattering offer the atomic details of interaction interfaces. Nowadays proteins are being crystallized using high-throughput techniques and very often without the extensive biochemical or biophysical characterization of their oligomeric states. Different computational methods have been proposed to identify the biological oligomeric complexes but only a few of them may decipher biological assemblies from crystalline states with high enough accuracy [19–22].

Such methods reconstruct both biological and crystal-packing interfaces by applying crystallographic symmetry operations, then differentiate the biological from the crystal-packing interfaces by computational criteria. For example, the PISA algorithm applies graph theory to find the set of stable assemblies, which fill all the crystal space in a systematic manner, where nodes and edges correspond to protein monomers and interfaces between them [20]. To distinguish “biologically relevant” from crystal packing interactions one can use ad-hoc scoring schemes which are based on interface area, amino acid composition, number of contacts, topological complementarity, and other characteristics [23–28]. Another group of methods tries to verify and predict oligomeric states and binding modes of a protein based on evolutionary conservation of homooligomeric binding modes of its homologs or on the presence of specific sequence and structural features [29–34]. For example, the NCBI IBIS server infers the interacting partners and the locations of binding sites for a given unknown protein by inspecting protein complexes formed by homologous proteins [32]. To ensure the biological relevance of binding sites, similar sites of homologous proteins are clustered together based on their sequence and structure conservation.

Physico-chemical properties of homooligomeric interfaces

The amino acid composition of homooligomeric interfaces differs from those of crystal packing interfaces, heterooligomeric interfaces and solvent-exposed surfaces and largely depends on the type of homooligomeric complexes [23–27,35–38]. For example, obligate homooligomers (complexes where monomers are unstable and usually non-functional upon isolation) and weakly associated dimers are characterized by the large fraction of hydrophobic and to a lesser extent aromatic residues on their interfaces while non-obligate (transient) complexes include more polar and charged residues [6,24,35,39]. Moreover,

interfaces of obligate homooligomers are usually larger [24] but contain fewer hydrogen bonds per residue compared to heterooligomers [36]. The analysis of kinetic and equilibrium data on dimeric proteins shows that per-residue interface and surface areas of “three state dimers” (monomers are stable on their own) is significantly smaller than that of “two state dimers” (monomers are not stable when separated from the complex) [40].

As can be seen in Figure 1 homooligomers mostly form dimers (38%), trimers (4%) and tetramers (10%) (Fig. 1), which have predominantly cyclic or dihedral symmetries [2,3,5,41]. Several explanations were proposed to account for these observations of self-attraction and symmetry, including stability, foldability and evolutionary optimization arguments [42–44]. The physical effect of a statistically enhanced self-attraction was modeled to show that interactions between identical random surfaces are stronger than attractive interactions between different random surfaces of the same size [43]. Furthermore, it was demonstrated that the efficiency of co-aggregation between different monomers and protein domains decreases with decreasing sequence identity [45]. Binding arrangements involving isologous homooligomeric interfaces with a two-fold symmetry axis seem to be more frequently conserved in evolution compared to non-isologous interfaces [46]. In addition, symmetrical dimers were shown to contain more residues in disordered regions than heterodimers, which might modulate the high specificity of interactions between the dimer complexes and their interacting partners and play an important role in allosteric regulation [47,48].

Evolutionary and physical mechanisms to form homooligomers

Different scenarios of protein oligomerization have been discussed in the literature. Some evolutionary pathways may follow kinetic pathways of two-state or three-state folding [49,50]. At the same time assembly pathways of oligomers may mimic evolutionary pathways and homooligomers with dihedral symmetry may evolve and assemble through their cyclic intermediates [41,51]. One of the major mechanisms of oligomerization is gene duplication with subsequent diversification when one copy of the gene retains its original function whereas another gene copy is under relaxed evolutionary constraints and so may develop new functional properties. This mechanism may lead to the formation of oligomeric paralogs and may create protein complexes with novel specificities [46,52–55]. It has been shown that in *S. cerevisiae* up to 20% of complexes evolved by step-wise partial duplications [52] whereas this mechanism was found to be less prevalent in *E. coli* [53].

Similarity in protein sequences, folds and functions between two orthologous proteins does not necessarily imply that they will have the same interacting partners [56]. Although homooligomeric states and binding modes have an overall tendency to be conserved within the clades on phylogenetic trees, they can only be reliably transferred from very close homologs (sharing higher than 30% sequence identity for oligomeric states and sharing higher than 50–70% identity for binding modes inference) [41,46]. Indeed, in some cases the oligomeric state can be evolutionarily conserved while the binding arrangement can be quite different. This points to the possibility that interactions and binding arrangements between paralogs are not necessarily inherited from the ancestral homooligomer but rather can develop anew in evolution. For example, proteins from the glycosyltransferase family may function as monomers, dimers, or tetramers in different organisms. While dimeric and tetrameric proteins appear to be as ancient as monomers, the binding modes can differ even between very similar proteins, for instance engaging completely opposite sites on the molecule [51]. Despite this diversity, there are certain features of evolutionarily conserved homooligomeric binding modes, for example, it has been found that binding modes with larger interfaces are more frequently conserved in evolution whereas smaller interfaces are often acquired more recently in evolution [41,46].

Next we will discuss several evolutionary mechanisms that play key roles in homooligomerization: domain swapping, formation of Leucine zippers, amino acid substitutions, and insertions/deletions on oligomeric interfaces.

Domain swapping

Domain swapping includes opening up of the monomeric conformation and exchanging identical regions between two monomers [57]. Overall, more than 60 examples of domain swapping are reported in protein structures from the Protein Data Bank [58]. Although swapping of terminal regions seems to be more common, the swapped regions can be located anywhere in the structure and the number of swapped regions can be greater than one [58,59]. For example, RNase A has two swapping regions at the N- and C- termini and can form a dimer, trimer, or higher order oligomers (Figure 2A) [60]. A number of studies analyzed the properties of inter-domain linker regions since these regions are flexible and might be responsible for domain swapping. It has been shown that insertions/deletions and substitution of certain residues (especially Pro) in the inter-domain linker region might affect monomer-dimer equilibrium [61–63].

Characteristic structural motifs

There are several known structural motifs used by proteins for oligomerization. The most common motif is the alpha-helical coiled-coil. The coiled-coil motif is observed as a series of continuous heptad repeats $(abcdefg)_n$ in protein sequences [64]. When the motif forms alpha-helical structure, hydrophobic residues at the “a” and “d” positions interact with each other to form a helix bundle. The “Leu zipper” motif is a type of the coiled-coil structure where leucine is frequently observed at the “d” position (Figure 2B) [65]. The coiled-coil motifs may form different oligomeric states depending on amino acids in the heptad repeat. For example, the yeast transcription factor GCN4 forms a dimeric coiled coil, however the replacements of Ile by Leu at the “a” and Leu by Ile at the “d” positions enables GCN4 to form a tetramer [66]. In addition it was shown that other combinations of Leu, Ile, and Val at the “a” and “d” positions may cause trimerization [66,67].

Amino acid substitutions

Amino acid substitutions introduced on the protein surface/interface may cause association or dissociation of homooligomers. A simple mechanism of amino acid substitutions that mediate the oligomeric states was proposed recently: the replacement of solvent exposed residues by more hydrophobic and larger protruding residues may shift the equilibrium toward the formation of oligomers [68] (Figure 2C). The artificial design of new oligomers by amino acid substitution was performed on examples of four assemblies [69]. Some of these proteins needed only a single amino acid replacement to associate to a higher-order complex and it was shown that introducing large non-polar side chains, such as phenylalanines or tryptophans, facilitated the complex formation [69]. Amino acid substitutions on homooligomer interfaces may change the functional activity of a protein and have been implicated in several diseases. For example, mutations leading to the disease fructose intolerance were shown to destabilize the tetramer of the enzyme fructose-1,6-biphosphate aldolase A and decrease its activity [70]. Certain substitutions in glutamate receptors may stabilize their dimer interface and reduce desensitization while amino acid substitutions in nuclear hormone receptors may disrupt the salt bridge stabilizing the functional dimer [71,72].

Insertions and deletions

Insertions and deletions at oligomer interfaces provide another important mechanism to modulate different protein oligomeric states in evolution (Figure 2D). The inspection of

insertions and deletions of homologous proteins in different oligomeric states revealed that about one quarter of them are located on interaction interfaces and are responsible for enabling or disabling the formation of oligomers [33,34,54]. According to these studies, insertions and deletions which differentiate monomers and dimers are located preferentially in loop regions and to a lesser extent on alpha-helices and beta-strands. It has been shown that insertions and deletions modulating oligomerization usually have a lower aggregation propensity and contain a larger fraction of polar, charged residues compared to conventional interfaces and protein surfaces [34]. Moreover, the removal of enabling regions from protein structures may result in the complete or partial loss of stability [34]. Interestingly, different oligomerization mechanisms can be employed by different proteins from the same family. For example, in the dihydrofolate reductase family, the loop of bacteriophage T4 leads to the formation of a homodimer [33] while homodimerization of the same enzyme from *Thermotoga maritima* is achieved by amino acid substitutions [68].

Evolution of new binding specificities through homooligomerization

As we described in the previous sections, homooligomeric states and binding modes can be very conserved within specific clades and protein families. At the same time protein regions modulating oligomerization may also be preserved in evolution, for example, the presence and locations of enabling insertions and deletions might be typical only for a given protein subfamily which in turn can be characterized by the well-defined oligomeric state. It implies that important homooligomerization features might function as specificity determinants providing high binding affinity to certain interacting partners while minimizing interactions with other unwanted partners. Such a mechanism would be essential for the separation of functional pathways of close paralogs, preventing the possible usage of similar surface regions interacting with the same or very similar partners; or facilitating through specific features the interactions with the novel partners. As shown in Table 1, proteins in different oligomeric states might have quite different binding affinities and therefore functional activities. Figure 3 and the two following examples illustrate the development of new protein specificities through homooligomerization in different organisms or in paralogs from the same organism.

For example, proteins from the human p53 C-terminal domain family function as homotetramers. Two of these family members (p63 and p73) can also form mixed heterotetramers, however p53 protein cannot associate with either p63 or p73 [73]. Interestingly, a recent study showed that p63 and p73 are different from p53 in that the former have an additional alpha-helix which stabilizes the tetramer [74]. The absence of this helix in p53 explains its less promiscuous binding which results in the separation of the p53 pathway from paralogous p63/p73 pathways. Another example involves the LIM-domain-binding protein Ldb1, a nuclear adaptor protein which interacts with diverse proteins containing LIM domains and plays essential roles in development and cellular differentiation [75,76]. Humans have two paralogs Ldb1 and Ldb2. Since the loss of Ldb1 causes severe developmental defects in embryos that are not compensated for by Ldb2 [77], Ldb1 and Ldb2 are likely to participate in different pathways. Interestingly, it has been shown that Ldb1 forms a trimer while Ldb2 exists in a monomer-tetramer-octamer equilibrium, suggesting that the oligomeric differences might enable Ldb1 and Ldb2 to interact with different partners and in different pathways [78].

Regulation of protein activity through transitions between different oligomeric states

Cellular processes are extremely complex requiring many factors and regulators to provide desired outcomes and prevent inefficient waste of energy. Control mechanisms change the

rate of particular cellular processes which in turn can be modulated at the level of gene expression or protein-protein interactions, by changing protein activity or post-translational modifications, by production of secondary messengers or other mechanisms. Oligomeric complexes between homologous proteins can integrate different pathways and provide the cross talk between them. Moreover, proteins might exist in dynamic equilibrium between different oligomeric states which can be controlled by physiological conditions (pH, certain ionic strength, temperature), ligands and post-translational modifications [79]. Therefore transitions between different oligomeric states may regulate protein activity and provide the switch between different pathways. It was shown that certain mutations may induce changes in oligomeric state and activity but do not compromise the stability [80]. It has also been shown that properties of weak transient homooligomers which exist in dynamic equilibrium are different from those of permanent dimers; the former contain smaller, more planar and polar interfaces [79]. In addition, reversible transitions between discrete conformations and oligomeric states might account for protein cooperative binding properties and allosteric mechanisms in signal transduction [16].

Here we summarize and present different scenarios for how shifting the equilibrium between different oligomeric states might serve as a regulatory mechanism. These scenarios include: a) homooligomerization might be important for protein self-activation; b) conformational changes accompanying transition may lead to exposure/suppression of active or protein binding sites; c) the formation of the oligomer may inhibit binding of a monomer to its substrate; d) finally, the post-translational modifications or binding of small molecules at or near the homooligomer interface may shift the equilibrium between different oligomeric states. In addition to experimental examples from the previous study [79] we manually compiled a list of experimentally verified examples of these mechanisms in Table 1.

We will describe a specific example of how the transition between the dimeric and tetrameric forms of pyruvate kinase can be implicated in tumor formation. Pyruvate kinase is a key glycolytic enzyme and its activity is consistently altered during tumorigenesis. It has been shown that during tumor formation the M2 isoform of pyruvate kinase is overexpressed [10]. The active tetrameric form of this enzyme from normal cells has high affinity to its substrate phosphoenolpyruvate (PEP), associates with the glycolytic enzyme complex, and produces high levels of ATP. The dimeric form has low affinity to PEP, does not associate with the glycolytic enzyme complex and is accompanied by low levels of ATP. Having the inactive dimeric form is advantageous for tumor cells as phosphometabolites above pyruvate kinases in the glycolytic cycle accumulate in tumor cells and can then be available as precursors for synthesis required by tumorigenesis. The tetramer-dimer ratio is regulated by different factors, for example by fructose 1,6-P2 and serine concentrations. Moreover Rous sarcoma virus may also phosphorylate M2-PK and lead to its dimerization and disassociation from the glycolytic enzyme complex [81].

Conclusions

The experimental characterization of homooligomeric structures, the dynamic equilibrium between different oligomeric states, and computational inference from the crystalline state has always been a challenge. Moreover, the diversity of homooligomeric binding modes among homologous proteins complicates the study of their evolutionary pathways and their computational prediction. Due to the advances of new experimental techniques, invaluable experimental data on homooligomers and their dynamics have been filling this gap, but most of them are focused on individual systems and do not provide the broader picture on the specific role of homooligomers in cellular function, especially in the regulation of protein functional activity. There remains an urgent need for new and improved experimental and computational methods for identification, modeling, and predicting homooligomers.

Acknowledgments

We thank Tom Madej for careful reading of the manuscript. This work was supported by National Institutes of Health/DHHS (Intramural Research program of the National Library of Medicine). K.H. was supported by a JSPS Research Fellowship from the Japan Society for the Promotion of Science.

References

1. Cornish-Bowden AJ, Koshland DE Jr. The quaternary structure of proteins composed of identical subunits. *J Biol Chem.* 1971; 246:3092–3102. [PubMed: 5574388]
2. Goodsell DS, Olson AJ. Structural symmetry and protein function. *Annu Rev Biophys Biomol Struct.* 2000; 29:105–153. [PubMed: 10940245]
3. Ali MH, Imperiali B. Protein oligomerization: how and why. *Bioorg Med Chem.* 2005; 13:5013–5020. [PubMed: 15993087]
4. Engstrom, A.; Strandberg, B., editors. *Symmetry and Function of Biological Systems at the Macromolecular Level.* New York, NY: 1969.
5. Ponstingl H, Kabir T, Gorse D, Thornton JM. Morphological aspects of oligomeric protein structures. *Prog Biophys Mol Biol.* 2005; 89:9–35. [PubMed: 15895504]
6. Jones S, Thornton JM. Protein-protein interactions: a review of protein dimer structures. *Prog Biophys Mol Biol.* 1995; 63:31–65. [PubMed: 7746868]
7. Schomburg I, Chang A, Ebeling C, Gremse M, Heldt C, et al. BRENDA, the enzyme database: updates and major new developments. *Nucleic Acids Res.* 2004; 32:D431–433. [PubMed: 14681450]
8. Ispolatov I, Yuryev A, Mazo I, Maslov S. Binding properties and evolution of homodimers in protein-protein interaction networks. *Nucleic Acids Res.* 2005; 33:3629–3635. [PubMed: 15983135]
9. Woodcock JM, Murphy J, Stomski FC, Berndt MC, Lopez AF. The dimeric versus monomeric status of 14-3-3zeta is controlled by phosphorylation of Ser58 at the dimer interface. *J Biol Chem.* 2003; 278:36323–36327. [PubMed: 12865427]
10. Mazurek S, Boschek CB, Hugo F, Eigenbrodt E. Pyruvate kinase type M2 and its role in tumor growth and spreading. *Semin Cancer Biol.* 2005; 15:300–308. [PubMed: 15908230]
11. Koike R, Kidera A, Ota M. Alteration of oligomeric state and domain architecture is essential for functional transformation between transferase and hydrolase with the same scaffold. *Protein Sci.* 2009; 18:2060–2066. [PubMed: 19670211]
12. Baisamy L, Jurisch N, Diviani D. Leucine zipper-mediated homo-oligomerization regulates the Rho-GEF activity of AKAP-Lbc. *J Biol Chem.* 2005; 280:15405–15412. [PubMed: 15691829]
13. Chen CP, Posy S, Ben-Shaul A, Shapiro L, Honig BH. Specificity of cell-cell adhesion by classical cadherins: Critical role for low-affinity dimerization through beta-strand swapping. *Proc Natl Acad Sci USA.* 2005; 102:8531–8536. [PubMed: 15937105]
14. Hattori T, Ohoka N, Inoue Y, Hayashi H, Onozaki K. C/EBP family transcription factors are degraded by the proteasome but stabilized by forming dimer. *Oncogene.* 2003; 22:1273–1280. [PubMed: 12618752]
15. Katsamba P, Carroll K, Ahlsen G, Bahna F, Vendome J, et al. Linking molecular affinity and cellular specificity in cadherin-mediated adhesion. *Proc Natl Acad Sci USA.* 2009; 106:11594–11599. [PubMed: 19553217]
16. Changeux JP, Edelstein SJ. Allosteric mechanisms of signal transduction. *Science.* 2005; 308:1424–1428. [PubMed: 15933191]
17. Miller S, Lesk AM, Janin J, Chothia C. The accessible surface area and stability of oligomeric proteins. *Nature.* 1987; 328:834–836. [PubMed: 3627230]
18. Dafforn TR. So how do you know you have a macromolecular complex? *Acta Crystallogr D Biol Crystallogr.* 2007; 63:17–25. [PubMed: 17164522]
19. Henrick K, Thornton JM. PQS: a protein quaternary structure file server. *Trends Biochem Sci.* 1998; 23:358–361. [PubMed: 9787643]

20. Krissinel E, Henrick K. Inference of macromolecular assemblies from crystalline state. *J Mol Biol.* 2007; 372:774–797. [PubMed: 17681537]
21. Zhu H, Domingues FS, Sommer I, Lengauer T. NOXclass: prediction of protein-protein interaction types. *BMC Bioinformatics.* 2006; 7:27. [PubMed: 16423290]
22. Levy ED. PiQSi: protein quaternary structure investigation. *Structure.* 2007; 15:1364–1367. [PubMed: 17997962]
23. Ponstingl H, Henrick K, Thornton JM. Discriminating between homodimeric and monomeric proteins in the crystalline state. *Proteins.* 2000; 41:47–57. [PubMed: 10944393]
24. Bahadur RP, Chakrabarti P, Rodier F, Janin J. Dissecting subunit interfaces in homodimeric proteins. *Proteins.* 2003; 53:708–719. [PubMed: 14579361]
25. Janin J, Rodier F. Protein-protein interaction at crystal contacts. *Proteins.* 1995; 23:580–587. [PubMed: 8749854]
26. Dasgupta S, Iyer GH, Bryant SH, Lawrence CE, Bell JA. Extent and nature of contacts between protein molecules in crystal lattices and between subunits of protein oligomers. *Proteins.* 1997; 28:494–514. [PubMed: 9261866]
27. Saha RP, Bahadur RP, Chakrabarti P. Interresidue contacts in proteins and protein-protein interfaces and their use in characterizing the homodimeric interface. *J Proteome Res.* 2005; 4:1600–1609. [PubMed: 16212412]
28. Shoemaker BA, Panchenko AR, Bryant SH. Finding biologically relevant protein domain interactions: conserved binding mode analysis. *Protein Sci.* 2006; 15:352–361. [PubMed: 16385001]
29. Xu Q, Dunbrack RL Jr. The protein common interface database (ProtCID)-a comprehensive database of interactions of homologous proteins in multiple crystal forms. *Nucleic Acids Res.* 2011; 39:D761–770. [PubMed: 21036862]
30. Aytuna AS, Gursoy A, Keskin O. Prediction of protein-protein interactions by combining structure and sequence conservation in protein interfaces. *Bioinformatics.* 2005; 21:2850–2855. [PubMed: 15855251]
31. Ogmen U, Keskin O, Aytuna AS, Nussinov R, Gursoy A. PRISM: protein interactions by structural matching. *Nucleic Acids Res.* 2005; 33:W331–336. [PubMed: 15991339]
32. Shoemaker BA, Zhang D, Thangudu RR, Tyagi M, Fong JH, et al. Inferred Biomolecular Interaction Server--a web server to analyze and predict protein interacting partners and binding sites. *Nucleic Acids Res.* 2009; 38:D518–D524. [PubMed: 19843613]
33. Akiva E, Itzhaki Z, Margalit H. Built-in loops allow versatility in domain-domain interactions: lessons from self-interacting domains. *Proc Natl Acad Sci USA.* 2008; 105:13292–13297. [PubMed: 18757736]
34. Hashimoto K, Panchenko AR. Mechanisms of protein oligomerization, the critical role of insertions and deletions in maintaining different oligomeric states. *Proc Natl Acad Sci USA.* 2010; 107
35. Ofra Y, Rost B. Analysing six types of protein-protein interfaces. *J Mol Biol.* 2003; 325:377–387. [PubMed: 12488102]
36. Zhanhua C, Gan JG, Lei L, Sakharkar MK, Kanguane P. Protein subunit interfaces: heterodimers versus homodimers. *Bioinformation.* 2005; 1:28–39. [PubMed: 17597849]
37. Anashkina A, Kuznetsov E, Esipova N, Tumanyan V. Comprehensive statistical analysis of residues interaction specificity at protein-protein interfaces. *Proteins.* 2007; 67:1060–1077. [PubMed: 17357164]
38. Brock K, Talley K, Coley K, Kundrotas P, Alexov E. Optimization of electrostatic interactions in protein-protein complexes. *Biophys J.* 2007; 93:3340–3352. [PubMed: 17693468]
39. Dey S, Pal A, Chakrabarti P, Janin J. The subunit interfaces of weakly associated homodimeric proteins. *J Mol Biol.* 2010; 398:146–160. [PubMed: 20156457]
40. Gunasekaran K, Tsai CJ, Nussinov R. Analysis of ordered and disordered protein complexes reveals structural features discriminating between stable and unstable monomers. *J Mol Biol.* 2004; 341:1327–1341. [PubMed: 15321724]
41. Levy ED, Boeri Erba E, Robinson CV, Teichmann SA. Assembly reflects evolution of protein complexes. *Nature.* 2008; 453:1262–1265. [PubMed: 18563089]

42. Wolynes PG. Symmetry and the energy landscapes of biomolecules. *Proc Natl Acad Sci U S A*. 1996; 93:14249–14255. [PubMed: 8962034]
43. Lukatsky DB, Shakhnovich BE, Mintseris J, Shakhnovich EI. Structural similarity enhances interaction propensity of proteins. *J Mol Biol*. 2007; 365:1596–1606. [PubMed: 17141268]
44. Andre I, Strauss CE, Kaplan DB, Bradley P, Baker D. Emergence of symmetry in homooligomeric biological assemblies. *Proc Natl Acad Sci USA*. 2008; 105:16148–16152. [PubMed: 18849473]
45. Wright CF, Teichmann SA, Clarke J, Dobson CM. The importance of sequence diversity in the aggregation and evolution of proteins. *Nature*. 2005; 438:878–881. [PubMed: 16341018]
46. Dayhoff JE, Shoemaker BA, Bryant SH, Panchenko AR. Evolution of protein binding modes in homooligomers. *J Mol Biol*. 2010; 395:860–870. [PubMed: 19879880]
47. Simon SM, Sousa FJ, Mohana-Borges R, Walker GC. Regulation of *Escherichia coli* SOS mutagenesis by dimeric intrinsically disordered umuD gene products. *Proc Natl Acad Sci USA*. 2008; 105:1152–1157. [PubMed: 18216271]
48. Fong JH, Shoemaker BA, Garbuzynskiy SO, Lobanov MY, Galzitskaya OV, et al. Intrinsic disorder in protein interactions: insights from a comprehensive structural analysis. *PLoS Comput Biol*. 2009; 5:e1000316. [PubMed: 19282967]
49. Xu D, Tsai CJ, Nussinov R. Mechanism and evolution of protein dimerization. *Protein Sci*. 1998; 7:533–544. [PubMed: 9541384]
50. Tiana G, Broglia RA. Design and folding of dimeric proteins. *Proteins*. 2002; 49:82–94. [PubMed: 12211018]
51. Villar G, Wilber AW, Williamson AJ, Thiara P, Doye JP, et al. Self-assembly and evolution of homomeric protein complexes. *Phys Rev Lett*. 2009; 102:118106. [PubMed: 19392244]
52. Pereira-Leal JB, Teichmann SA. Novel specificities emerge by stepwise duplication of functional modules. *Genome Res*. 2005; 15:552–559. [PubMed: 15805495]
53. Reid AJ, Ranea JA, Orengo CA. Comparative evolutionary analysis of protein complexes in *E. coli* and yeast. *BMC Genomics*. 2010; 11:79. [PubMed: 20122144]
54. Hashimoto K, Madej T, Bryant SH, Panchenko AR. Functional states of homooligomers: insights from the evolution of glycosyltransferases. *J Mol Biol*. 2010; 399:196–206. [PubMed: 20381499]
55. Pereira-Leal JB, Levy ED, Kamp C, Teichmann SA. Evolution of protein complexes by duplication of homomeric interactions. *Genome Biol*. 2007; 8:R51. [PubMed: 17411433]
56. Yu H, Luscombe NM, Lu HX, Zhu X, Xia Y, et al. Annotation transfer between genomes: protein-protein interologs and protein-DNA regulogs. *Genome Res*. 2004; 14:1107–1118. [PubMed: 15173116]
57. Bennett MJ, Choe S, Eisenberg D. Domain swapping: entangling alliances between proteins. *Proc Natl Acad Sci USA*. 1994; 91:3127–3131. [PubMed: 8159715]
58. Gronenborn AM. Protein acrobatics in pairs--dimerization via domain swapping. *Curr Opin Struct Biol*. 2009; 19:39–49. [PubMed: 19162470]
59. Liu Y, Eisenberg D. 3D domain swapping: as domains continue to swap. *Protein Sci*. 2002; 11:1285–1299. [PubMed: 12021428]
60. Liu Y, Gotte G, Libonati M, Eisenberg D. A domain-swapped RNase A dimer with implications for amyloid formation. *Nat Struct Biol*. 2001; 8:211–214. [PubMed: 11224563]
61. Rousseau F, Schymkowitz JW, Wilkinson HR, Itzhaki LS. Three-dimensional domain swapping in p13suc1 occurs in the unfolded state and is controlled by conserved proline residues. *Proc Natl Acad Sci USA*. 2001; 98:5596–5601. [PubMed: 11344301]
62. Barrientos LG, Louis JM, Botos I, Mori T, Han Z, et al. The domain-swapped dimer of cyanovirin-N is in a metastable folded state: reconciliation of X-ray and NMR structures. *Structure*. 2002; 10:673–686. [PubMed: 12015150]
63. Dickason RR, Huston DP. Creation of a biologically active interleukin-5 monomer. *Nature*. 1996; 379:652–655. [PubMed: 8628400]
64. Mason JM, Arndt KM. Coiled coil domains: stability, specificity, and biological implications. *ChemBiochem*. 2004; 5:170–176. [PubMed: 14760737]
65. O'Shea EK, Rutkowski R, Kim PS. Evidence that the leucine zipper is a coiled coil. *Science*. 1989; 243:538–542. [PubMed: 2911757]

66. Harbury PB, Zhang T, Kim PS, Alber T. A switch between two-, three-, and four-stranded coiled coils in GCN4 leucine zipper mutants. *Science*. 1993; 262:1401–1407. [PubMed: 8248779]
67. Harbury PB, Kim PS, Alber T. Crystal structure of an isoleucine-zipper trimer. *Nature*. 1994; 371:80–83. [PubMed: 8072533]
68. Nishi H, Ota M. Amino acid substitutions at protein-protein interfaces that modulate the oligomeric state. *Proteins*. 2010; 78:1563–1574. [PubMed: 20112251]
69. Grueninger D, Treiber N, Ziegler MO, Koetter JW, Schulze MS, et al. Designed protein-protein association. *Science*. 2008; 319:206–209. [PubMed: 18187656]
70. Malay AD, Allen KN, Tolan DR. Structure of the thermolabile mutant aldolase B, A149P: molecular basis of hereditary fructose intolerance. *J Mol Biol*. 2005; 347:135–144. [PubMed: 15733923]
71. Sun Y, Olson R, Horning M, Armstrong N, Mayer M, et al. Mechanism of glutamate receptor desensitization. *Nature*. 2002; 417:245–253. [PubMed: 12015593]
72. Estebanez-Perpina E, Arnold LA, Jouravel N, Togashi M, Blethrow J, et al. Structural insight into the mode of action of a direct inhibitor of coregulator binding to the thyroid hormone receptor. *Mol Endocrinol*. 2007; 21:2919–2928. [PubMed: 17823305]
73. Davison TS, Vagner C, Kaghad M, Ayed A, Caput D, et al. p73 and p63 are homotetramers capable of weak heterotypic interactions with each other but not with p53. *J Biol Chem*. 1999; 274:18709–18714. [PubMed: 10373484]
74. Joerger AC, Rajagopalan S, Natan E, Veprintsev DB, Robinson CV, et al. Structural evolution of p53, p63, and p73: implication for heterotetramer formation. *Proc Natl Acad Sci USA*. 2009; 106:17705–17710. [PubMed: 19815500]
75. Matthews JM, Visvader JE. LIM-domain-binding protein 1: a multifunctional cofactor that interacts with diverse proteins. *EMBO Rep*. 2003; 4:1132–1137. [PubMed: 14647207]
76. Matthews JM, Bhati M, Craig VJ, Deane JE, Jeffries C, et al. Competition between LIM-binding domains. *Biochem Soc Trans*. 2008; 36:1393–1397. [PubMed: 19021562]
77. Mukhopadhyay M, Teufel A, Yamashita T, Agulnick AD, Chen L, et al. Functional ablation of the mouse Ldb1 gene results in severe patterning defects during gastrulation. *Development*. 2003; 130:495–505. [PubMed: 12490556]
78. Cross AJ, Jeffries CM, Trewheella J, Matthews JM. LIM domain binding proteins 1 and 2 have different oligomeric states. *J Mol Biol*. 2010; 399:133–144. [PubMed: 20382157]
79. Nooren IM, Thornton JM. Structural characterisation and functional significance of transient protein-protein interactions. *J Mol Biol*. 2003; 325:991–1018. [PubMed: 12527304]
80. Devenish SR, Gerrard JA. The role of quaternary structure in (beta/alpha)(8)-barrel proteins: evolutionary happenstance or a higher level of structure-function relationships? *Org Biomol Chem*. 2009; 7:833–839. [PubMed: 19225662]
81. Presek P, Reinacher M, Eigenbrodt E. Pyruvate kinase type M2 is phosphorylated at tyrosine residues in cells transformed by Rous sarcoma virus. *FEBS Lett*. 1988; 242:194–198. [PubMed: 2462512]
82. Jia-Lin Ma N, Stern DF. Regulation of the Rad53 protein kinase in signal amplification by oligomer assembly and disassembly. *Cell Cycle*. 2008; 7:808–817. [PubMed: 18239457]
83. Lin X, Liu J, Maley F, Chu E. Role of cysteine amino acid residues on the RNA binding activity of human thymidylate synthase. *Nucleic Acids Res*. 2003; 31:4882–4887. [PubMed: 12907731]
84. Genovese F, Ferrari S, Guaitoli G, Caselli M, Costi MP, et al. Dimer-monomer equilibrium of human thymidylate synthase monitored by fluorescence resonance energy transfer. *Protein Sci*. 2010; 19:1023–1030. [PubMed: 20306493]
85. Arnold SF, Vorojeikina DP, Notides AC. Phosphorylation of tyrosine 537 on the human estrogen receptor is required for binding to an estrogen response element. *J Biol Chem*. 1995; 270:30205–30212. [PubMed: 8530431]
86. Randez-Gil F, Sanz P, Entian KD, Prieto JA. Carbon source-dependent phosphorylation of hexokinase PII and its role in the glucose-signaling response in yeast. *Mol Cell Biol*. 1998; 18:2940–2948. [PubMed: 9566913]

87. Lee KK, Yonehara S. Phosphorylation and dimerization regulate nucleocytoplasmic shuttling of mammalian STE20-like kinase (MST). *J Biol Chem.* 2002; 277:12351–12358. [PubMed: 11805089]
88. Hwang I, Thorgeirsson T, Lee J, Kustu S, Shin YK. Physical evidence for a phosphorylation-dependent conformational change in the enhancer-binding protein NtrC. *Proc Natl Acad Sci USA.* 1999; 96:4880–4885. [PubMed: 10220387]
89. Levinson NM, Visperas PR, Kuriyan J. The tyrosine kinase Csk dimerizes through Its SH3 domain. *PLoS One.* 2009; 4:e7683. [PubMed: 19888460]

Figure 1. Distribution of homooligomeric states in a non-redundant set of Protein Data Bank (PDB) structures

The non-redundant set was obtained using the criteria of BLAST p-value < 10e-07 on all PDB chains and oligomeric states were annotated by PISA [20].

Figure 2. Mechanisms of homooligomerization

A. Domain swapping of RNase A: monomer (1RTB) and domain-swapped dimer (1A2W) of RNase A. The swapped regions and domain linker regions are shown as red and yellow, respectively. B. Leu-zipper of GCN4 (1YSA). Leu at position “d” is shown as stick model and colored in red. C. amino acid substitutions for designed homooctamer of L-rhamnulose-1-phosphatase aldolase (2UYU). The original protein exists as a tetramer which is shifted to the octamer upon a single substitution (A88F, shown in red). D. Insertions and deletions of N-acetyl-L-glutamate kinase hexamer (2BUF). Inserted N-terminal helix shown in red enables the hexamer formation.

Figure 3.
Illustration for the development of novel protein specificities and regulation of protein activity through homooligomerization.

Table 1

Experimental examples of proteins regulated through transitions between different oligomeric states

Protein name	Regulation of transition	Function of one oligomeric state	Function of the other oligomeric state	Reference
Pyruvate kinase M2	Tetramer-dimer ratio is regulated by fructose 1,6-P2, serine and phosphorylation	Inactive dimer leads to accumulation of phosphometabolites in tumor cells	Active tetramer associates with glycolytic enzyme complex to perform its functions in normal cell	[10]
Rad53, DNA damage checkpoint kinase in yeast	Assembly/disassembly is regulated through phosphorylation		Rad53 forms oligomers upon DNA damage. Rad53/Chk2 homooligomerization is a mechanism to activate signal in DNA damage responses	[82]
Human thymidylate synthase		Conformational changes accompanying dimer/monomer transition lead to exposure of Cys residues of monomer and binding to TS mRNA and autoregulation of translation by its protein product	Dimeric form is an active thymidylate synthase	[83,84]
Estrogen receptor	Phosphorylation of Tyr547 by Src tyrosine kinases controls receptor dimerization	Monomer probably does not bind DNA	Dimerization induces DNA binding	[85]
Hexokinase PII	Reversible phosphorylation is carbon source dependent Glucose promotes dephosphorylation.	When monomer is phosphorylated cells can not provide glucose repression of invertase	Dimer is dephosphorylated	[86]
Mammalian STE20- like kinase (MST)	Okadaic acid induces phosphorylation in the activation loop and monomerization	Phosphorylated active monomeric enzyme is translocated into nucleus	Dimerization contributes to the cytoplasmic retention of MST	[87]
Bacterial enhancer- binding protein nitrogen regulatory protein C (NtrC)	Phosphorylation induces oligomerization and changes interdomain interactions		Oligomerization activates the ATPase activity	[88]
C-terminal Src Kinase (Csk)			SH3 domains form homodimers and bury recognition surface for SH3 interactors including tyrosine phosphatase	[89]
D-hydantoinase	Removal of C-terminal Arg induces monomer formation	Monomer shows 40% activity reduction	Dimer and tetramers are active	[80]
14-3-3 proteins	Phosphorylation of Ser58 disrupts dimers	Monomer-dimer transition modulates target protein activity	Monomer-dimer transition modulates target protein activity	[9]