We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

4,800
Open access books available

122,000
International authors and editors

135M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS BOOK CITATION INDEX INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Chapter 11

The Assembly of Protein Oligomers — Old Stories and New Perspectives with Graph Theory

Claire Lesieur

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/58576

1. Introduction

Proteins are biological entities made of a chain of amino acids bound to one another in a specific order, called the primary structure or the amino acid sequence of the protein. Based on the sequence and the environment, the protein acquires a tridimensional shape called tertiary structure (3D-structure), conformation or fold, suitable for its biological function. The functional shape is the native structure of the protein. The set of reactions leading to the native structure is the folding of the protein. The vast majority of proteins are oligomers which function only after the association of several copies of their chains. Homo-oligomers have chains with identical sequences and hetero-oligomers have chains with different sequences. The number of associated chains defines the quaternary structure of the oligomer, or its stoichiometry [1]. According to the Protein Database (PDB) where all known 3D structures of proteins are stored, the most observed quaternary structure in all taxa is the dimer (Fig. 1A). A taxon is a set of living organisms grouped because of some shared characteristics. Besides dimers, there exists a large variety of assemblies in terms of quaternary structures and point group symmetries (Fig. 1). In forming a protein oligomer, subunit association has to be considered in addition to folding.

Folding involves the formation of interactions/bonds between atoms of the amino acids of a single chain. These are intramolecular (within a single molecule) amino acid interactions (Fig. 2A). Chain association involves the formation of interactions/bonds between atoms of the amino acids provided by at least two individual chains. These are intermolecular (between two molecules) amino acid interactions (Fig. 2B). Here the protein chain is considered as the molecule.

The twenty natural amino acids share four atoms called the backbone atoms and are distinguished by a set of atoms called the side chain atoms. These atoms can make different types

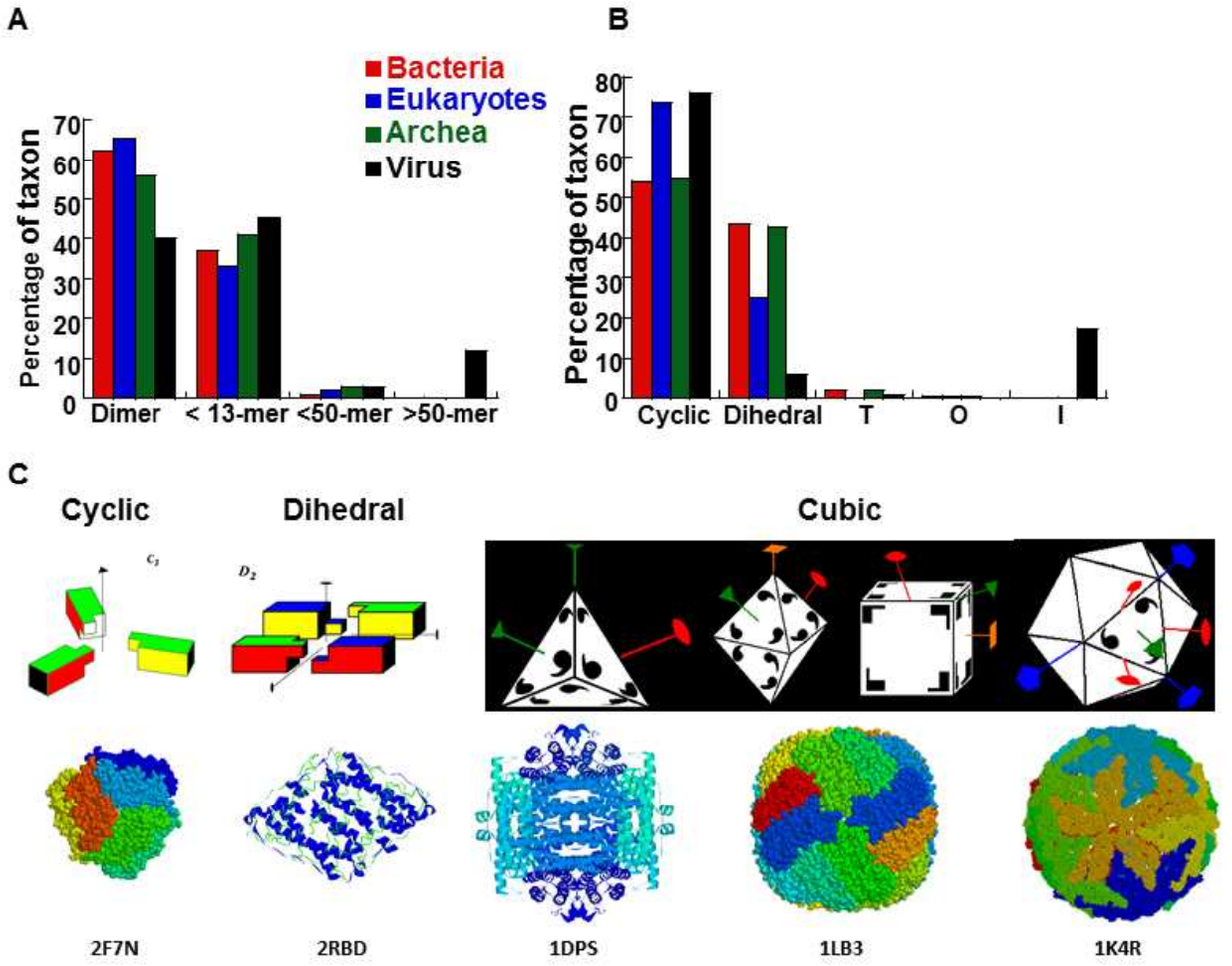

Figure 1. Protein oligomer features. The data in A and B are obtained by screening the PDB. **A.** Distribution of taxa according to quaternary structures. **B.** Distribution of taxa according to point group symmetries. **C.** Protein oligomers belong to three different point group symmetries. Cn, cyclic with n rotational axes. One ferritin is given as an example for a C3 symmetry (PDB 2F7N). Dn, dihedral with n rotational axes plus 2-fold perpendicular axes. One ferritin is given as an example for a D2 symmetry (PDB 2RBD). Cubic. T, tetrahedral with four 3-fold axes and six 2-fold axes. Ferritin is given as an example (PDB 1DPS) [2]. O, octoahedral, octahedron or hexahedron with six 4-fold axes, eight 3-fold axes and twelve 2-fold axes. One ferritin is given as an example (PDB 1LB3). I, icosahedral with twelve 5-fold axes, twenty 3-fold axes and thirty 2-fold axes. Ferritin is given as an example (PDB 1K4R). Icosahedral is not a cubic point group symmetry but has been conflated to the cubic point group symmetry in chemistry [3].

of chemical bonds. First, the amino acids are linked to one another by a covalent bond involving two backbone atoms and called the peptide bond. The covalent bonds are thus used to make a chain of amino acids arranged in a specific order, the primary sequence of the protein. Cysteine and methionine amino acids are the only amino acids that can make a supplementary covalent bond, called a disulfide bond, using the sulfur atom of their side chains. There can be intramolecular disulfide bonds (between two cysteines of one chain) or intermolecular disulfide bonds (between two cysteines, each one produced by a distinct chain), the latter making a covalent oligomer. Some collagen trimers are stabilized by inter-chain disulfide bonds [4]. The collagen oligomers have been reviewed recently [5, 6]. Some other examples of covalent oligomers can be found in the chapter on protein oligomerization by Giovanni Gotte

http://dx.doi.org/10.5772/58576

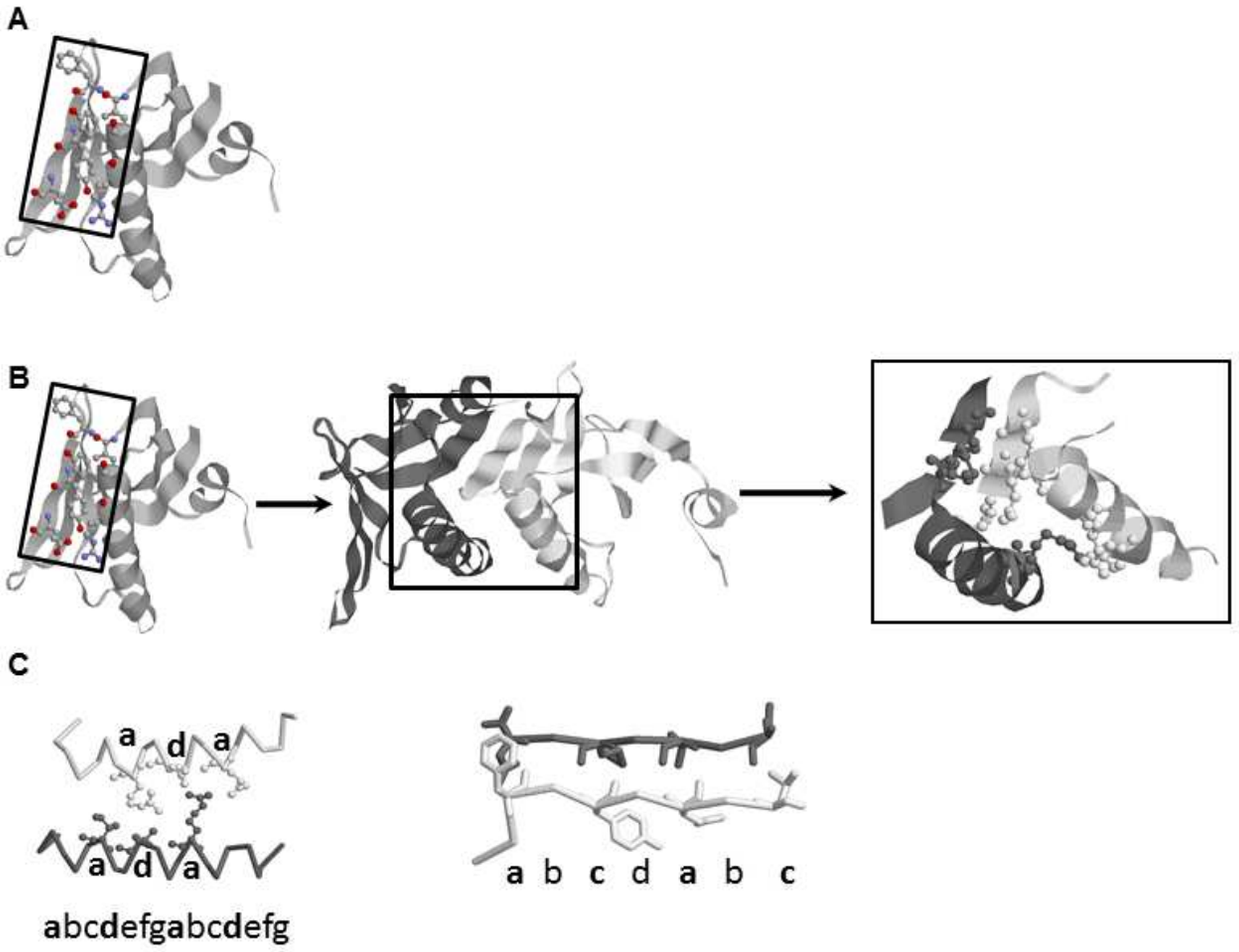

Figure 2. Interactions in proteins.A.Protein monomer. Monomeric proteins perform their biological function with a single chain. The formation and the stability of their native 3D structure involve only intramolecular atomic interactions (interaction within a chain). **B. Protein oligomer.** Oligomeric proteins need to assemble several copies of their chain to perform their biological function. The formation and the stability of their native structure involve intramolecular atomic interactions (interaction within one chain) to acquire their fold and intermolecular atomic interactions (interaction between chains) to acquire their quaternary structure. The pictures are generated with Rasmol. The protein chains are shown in ribbons of different colors. For a few amino acids, all atoms are indicated in balls and sticks to highlight intra and inter atomic interactions. **C. Recognition modes.** A protein interface is made of two set of atoms, one per chain, spatially organized to yield a chemical and geometrical complementarity. Two simple cases are presented. On the left are two interacting α-helices and on the right are two interacting β-strands. Because of the geometry, the interacting amino acids produced particular sequence motifs/pattern **a**bc**d**efg with *a* and *d* residues interacting in the α-helical interface and **a**b**c**da**b**d**c**d, with a and c residues interacting in the β-strands.

and Massimo Libonati. Disulfide bonds can significantly increase the stability of a chain or of an oligomer, but it is not necessarily true and that needs to be measured case by case [7]. Covalent bonds are strong interactions as it takes a large amount of energy to break them (110-50 kcal/mol). In living organisms, an enzyme (protease) is necessary to cut a covalent bond.

Second, the tertiary and quaternary structures of protein as well as the folding and the chain association involve mostly non-covalent bonds between the atoms of the amino acids, called weak bonds because it takes a small amount of energy to break them (1-7 kcal/mol). These are

hydrogen bonds, hydrophobic bonds, electrostatic bonds (between charges), polar bonds (between dipoles) and van der Waals interactions. Under physiological conditions, the weak bonds continuously form and break. The secondary structures of proteins, α-helices and β-sheets (intramolecular β-strand interactions) are stabilized by hydrogen bonds between atoms of the backbone of the amino acids. Likewise for intermolecular β-sheets but the hydrogen bonds are between atoms of the backbone of amino acids located on different chains. At last, but not least, worth noted amino acids in terms of folding and association is the proline. Its side chain geometry is particular and can adopt two positions named *cis* and *trans*, affecting the relative position of its neighboring amino acids accordingly. The consequence is the existence of two different local tridimensional states. The transition between the *cis and trans* conformation is called a *cis-trans* isomerization and is known to slow down the folding of a protein, and to also affect the association of chains indirectly [8-11].

The zone of contact between two associated chains is called the protein interface. The protein interface is made of intermolecular amino acid interactions. Every chain provides a domain that recognizes another domain, or the same domain, on another chain and associates with it. The association is based on the chemical and geometrical complementarities of the two domains. These complementarities are constructed on the spatial layout of the intermolecular amino acid interactions (Fig. 2C). These layouts are referred to as recognition modes and have been extensively studied [12-20]. Yet the rules that would enable us to predict recognition modes from sequences still remain elusive.

Understanding and predicting the modes of recognition of protein interfaces is essential for several reasons. First, because oligomers are involved in many cellular activities and when default interactions occur there are numerous consequences among which certain diseases. Second, because it is important to distinguish biologically significant interfaces from non-specific interfaces observed in protein crystals in order to properly assess biological assemblies in x-ray structures [21-24]. Along the same line, it is still not trivial to determine experimentally whether a protein is an oligomer and if so its quaternary structure, so any predictive quaternary structure tool is helpful. Third, because the knowledge on protein interfaces is used in synthetic biology to engineer artificial oligomers for several purposes from drug delivery devices to the development of new material [25-29]. As an example, one can read the chapter by Keqin Zhang on silkworm and spider protein fibers and their potential use in the fabric industry.

2. Overview of protein assembly

2.1. Intermolecular amino acid interactions

Protein interfaces have been extensively investigated [14, 30-35]. But because protein interfaces are large and rather flat in nature, they lack the spatial constraints achieved by a limited number of sequences. Thus the sequences of protein interfaces rarely share trivial profiles or patterns, in contrast to protein-small molecule interfaces or residues involved in enzymatic active sites [36, 37].

Thus, it has been clear very early on that looking at the 3D structures of protein interfaces was a necessary alternative to sequence analysis [34, 38-43]. The increasing number of 3D-structures available for oligomeric proteins has also favored such investigations and the development of computational methods to identify and study the amino acids at protein interfaces on large scale datasets. In the chapter, we essentially review these computational advances but discuss little experimental progress. One can read the chapter on protein oligomerization by Giovanni Gotte and Massimo Libonati for information on experimental approaches or read information provided in the following publication [44].

The first benefit of computational approaches is the facility in discriminating intermolecular amino acid interactions from intramolecular amino acid interactions systematically and efficiently by relatively simple algorithms.

2.1.1. Identification of the intermolecular contacts at protein interfaces

Many algorithms are available to identify the amino acids involved in intermolecular interactions from the x-ray coordinates of the 3D-structure of a protein oligomer (reviewed in [13, 21, 45]). The coordinates are accessible at the protein database (PDB, http://www.rcsb.org/pdb/home/home.do)[46]. There are databases were interfaces have been classified according to their 3D organization, residue conservation and residue types [47]. Some databases are used to complement cellular networks (interactomes) with structural information on the binding modes between cellular partners [12, 18-20, 48].

The classical algorithms are based on three different measures: (i) accessibility surface area, (ii) voronoi cells and (iii) arithmetic distances. More novel algorithms use graph theory measures such as centrality and coefficient clustering.

Accessible surface area (ASA). The first method calculates the solvent accessible surface area by rolling a probe of a given radius around the Van der Waal's surface of the protein atoms whose centre is the accessible surface [49]. Typically, the probe has the same radius as water (1.4 Å) and hence the surface described is referred to as the solvent accessible surface. The ASA are calculated for the monomer and for the oligomer and the interface residues are obtained by the difference in their ASA. ASA is currently used to discriminate biological contacts (large ASA, 1600 ± 400 $Å^2$)) from crystal ones (small ASA <400 $Å^2$) [50, 51]. The PDB entries are now processed accordingly and provide both biological assembly and asymmetric unit coordinates. The biological assembly entry includes a remark to explain whether the oligomeric state is "author provided" (experimentally shown to be an oligomer) or "software determined" or both. Alternatively, the biological assembly can be downloaded directly from the Structure Server PQS (Protein Quaternary structure) at EBI (http://pqs.ebi.ac.uk) [50]. ASA can be calculated from different servers and programs such as PISA (Protein, Interfaces, Structures and Assemblies, http://www.ebi.ac.uk/msd-srv/prot_int/) or Naccess (http://www.bioinf.manchester.ac.uk/naccess/), both essentially implemented from the Lee & Richards method [52].

Voronoi cells. The second method selects interfacial residues based on the Voronoi diagram or its closely related power diagram [53-55]. The Voronoi diagram associates to each atom its Voronoi cell, namely the convex polyhedron that contains all points of space closer to that atom

than to any other atom. Instead of the Euclidean distance |ax| between a point *x* and an atom centered in *a*, this diagram use the power distance p(x) of *x* with respect to the ball of radius *r* that represents the atom,$p(x)=|ax|^2-r^2$. The Voronoi cell of an atom then comprises all points of space that have a power distance to that atom less than to any other atom. Its facets belong to the radical plane, which contains the intersection of the spheres if they do intersect. The Voronoi (or power) diagram offers a natural definition of contacts: two atoms are in contact if and only if their Voronoi cells share a facet. The use of Voronoi diagrams has been extended for assessing the reconstruction of protein assembly with the impressive example of the Nuclear Pore Complex [56].

Arithmetic distances. The third method also requires the 3D-structure (available in the PDB) and calculates Euclidian distances between atoms of the amino acids of different chains to detect only intermolecular atomic interactions [23–25]. The selection is the pairs of atoms which are within a cut-off distance from each other classically around 5.0 Å such that any type of chemical bonds between the atoms are considered (H-bonds, electrostatic interactions, van der Waals forces, salt bridges and hydrophobic attractions). The pairs of atoms selected as part of the interface, depend on the choice of the cut off distance. This is a serious issue because a distance cannot fully describe a spatial arrangement and there is a chance that the geometry of the interface is not faithfully represented by the set of selected pairs [57]. The need to use a cut off distance for the selection prevents from having a natural read of the geometry of the interface. Better alternatives select pairs of atoms in interactions as the nearest neighbor atoms instead of using a cut-off [40, 58-60]. This measure is more capable of reading the whole geometry of the interface and therefore supplies a more accurate set of pairs of the intermolecular contacts. Differences in the set of atoms selected according to distances are illustrated in figure 3.

In addition, residue conservation or spatial chemical conservation can be implemented to yield a set of intermolecular amino acid interactions based on structural and sequence information [47]. The method requires the PDBs and a multiple sequence alignment as input data. Individual residues are represented in terms of regional alignments that reflect both their structural environment and their evolutionary variation, as defined by the alignment of homologous sequences. Multiple alignments use either the Shannon or the Von Neumann entropy [61]. Conservation scores are also efficient in discriminating genuine biological assemblies from crystal contacts [22, 24]. There exist several algorithms, the most efficient are mapping conservation score to the 3D-structures such as Evolutionary Trace [62-65].

Hotspots-In the mid-nineties, the specific energetic contribution of the side chain atoms to protein interfaces was investigated using Alanine Scanning Mutagenesis because the mutation by alanine reduced the interactions to backbone atoms [66]. It was found that only a small subset of interfacial residues were sensitive to an alanine mutation indicating that the energetic contribution of the interfacial residues was not distributed uniformly. Some key 'hot spot' residues contributed dominantly to the binding free energy. Thorn and Bogan deposited hot spots from alanine scanning mutagenesis experiments in the ASEdb database (http://nic.ucsf.edu/asedb/) [67]. BID (The Binding Interface Database) is another database of experimental hot spots, which collects all available experimental data related to hot spots in protein

http://dx.doi.org/10.5772/58576

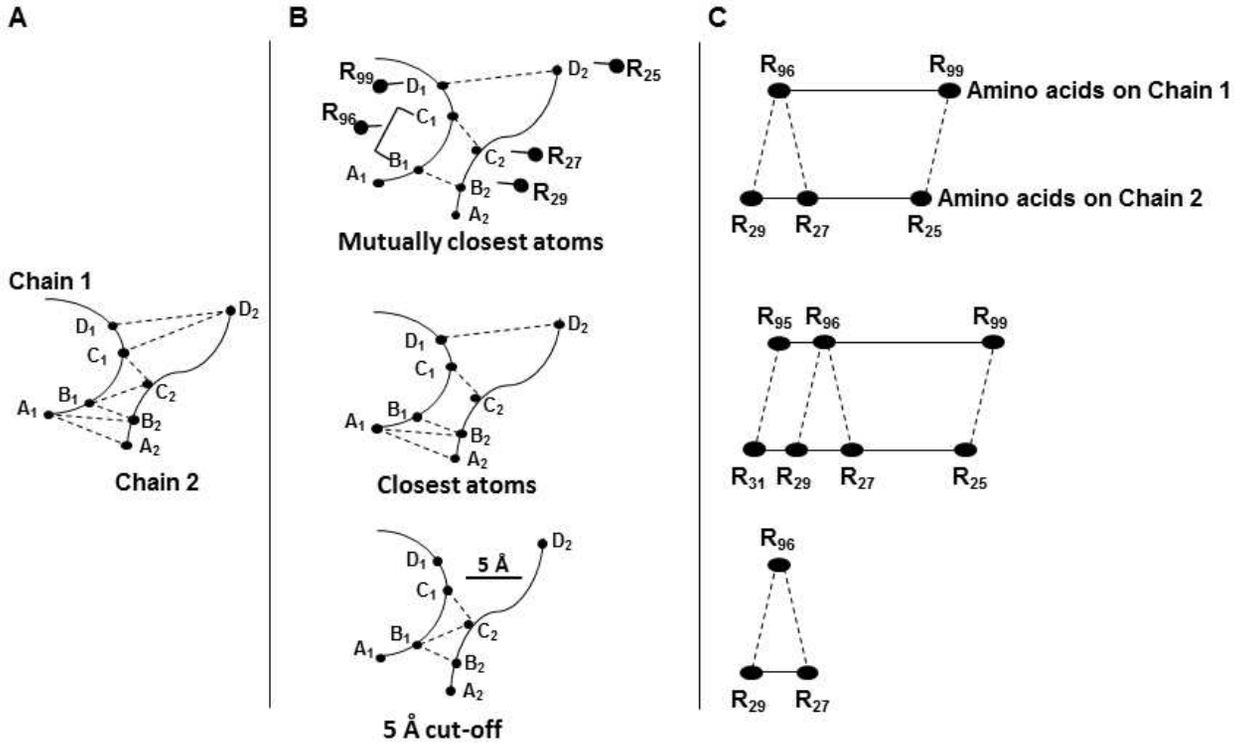

Figure 3. Selection of the amino acids in interactions at interfaces. A. Schematic of an interface between chain 1 and chain 2. Each chain is symbolized by a line and the chain respective atoms are indicated by black dot and letters with indices corresponding to the chain. Distances and so potential interactions between atoms are indicated by dotted lines. Only few interactions are indicated for the sake of clarity. **B. Selection of atoms in interactions at the interface.** The same schematic is reproduced after selection of the atoms in interaction at the interface. The top schematic is a selection based on mutually closest atoms, the middle one is a selection of all closest atoms and the bottom one is a selection of atoms at distances shorter than a cut-off of 5 Å. **C. Coarse-grained graphs of the interface.** Based on the selected atoms, a graph of the interactions between amino acid is drawn. The top, middle and bottom graphs correspond to the top, middle and bottom selections illustrated in B, respectively.

interfaces(http://tsailab.chem.pacific.edu/wikiBID/index.php/Main_Page)[68]. In parallel, there have been several experimental evidences not based on ala-scanning mutagenesis but on kinetics of assembly that showed the role of only some amino acids of the interfaces in regulating the chain association [69, 70]. Now hotspot (or hot spot) is a colloquial term that distinguishes a residue relevant for interface formation from others.

There exist several predictors of hotspots combined or not with evolution conservation based on ASA, voronoi and distances, some reviewed in [21, 47].

Algorithms based on graph theory-More recently, methods based on graph theory have also been proposed to identify hotspots. A graph or a network is a mathematical representation of pairwise relation between objects. A graph is made of vertices (or nodes) and lines, called edges (or links) that connect them. Proteins have been described as networks, with the amino acids of a protein chain considered as the nodes and the interactions between amino acids as the links. These networks, referred to as protein structure networks or amino acid networks, describe the entire protein and are used to infer global characteristics of the protein. Likewise, protein interfaces have been described as networks of hot spots in interactions with the hot

spots as nodes and the interactions between hot spots as links. Now, it is important to realize that protein interfaces are not networks but sub-networks (sub graph) as they describe local properties of the protein, namely the interfaces. A good overview of network measures can be found in [71].

The sub-graphs are built of pairs of amino acids in intermolecular interactions (Fig. 3C). As for the previous methods, the 3D-structures of the protein oligomer are used to build the graph and different measures are used to infer the amino acid in interactions. Basically, the atoms are considered as points in space, each chain of a protein oligomer constituting a distinct set of points. All distances between atoms of the different set are calculated and any two atoms within a given cut off distance or closest atoms are considered linked. A coarse-grain graph is built by replacing the interacting atoms by their respective interacting amino acids.

The measure of path length and centrality is used (Central nodes). In graphs, the notion of links between amino acids goes beyond chemical/physical bonds which are based on arithmetic distances. In a graph, any two amino acids are connected by a geodesic distance which is the shortest distance between them, measured as the minimum number of links that need to be crossed to connect them by the shortest path. This distance is called the path length and is symbolized by the letter *l*. The mean path length, $\langle l \rangle$, represents the average over the shortest paths between all pairs of nodes and offers a measure of the network's overall navigability. This introduces the notion of contacts through communication routes in addition to the more classical notion of geometrical/chemical contacts. This novel notion will have vast applications in the field of protein structure and protein dynamics. It has been already used for investigating protein allosteric mechanisms as discussed later [72-77].

Several measures of the centrality of a graph (closeness, betweenness) are associated with geodesic distances. Basically the numbers of shortcut paths going through every node are calculated and the most central nodes are those with the highest numbers of shortcut paths going through them. In other words, the centrality is finding nodes at the crossroad of communication routes. Centrality measures have been used to identify residues important for the function and for the fold of proteins [78-80]. It has also been used to identify hot spot residues in protein interfaces [81, 82]. One has to be careful to keep in mind that a central node needs not to be at the center of the protein or at the center of the interface, but it is necessarily at a crossroad of many paths.

The measure of degree and clustering coefficient is used. The number of links a node has, in the context, the number of interactions an amino acid has with other amino acids (number of contact amino acids), is called the degree and is symbolized by the letter *k*. The mean degree $\langle k \rangle$ represents the average degree over all the nodes of the network. The degree distribution of networks is informative on the characteristics of the network [71]. Networks with a power law degree distribution are called scale-free, a name that is rooted in statistical physics literature. It indicates the absence of a typical degree for the nodes in the network (one that could be used to characterize the rest of the nodes). This is in strong contrast to random networks, which have Poisson degree distributions, and for which the degree of all nodes is in the vicinity of the average degree $\langle k \rangle$, which can be considered typical. There are also exponential degree distributions which are single-scale networks. Scale free networks are

made of many nodes with few links and few nodes with many links, called hubs. Hubs are absent in random and single scale networks. Hubs are communication devices that allow most nodes of the networks to be connected with others. This has noticeable consequences in terms of the vulnerability of the networks to changes on the hubs or elsewhere [83]. This is discussed later in the chapter. Proteins are essentially random networks or single-scale networks [84].

The clustering coefficient, called *C*, is based on the degree of the nodes and it measures the probability that a node A which is connected to B, itself connected to C, has to be connected to C as well. Calculated over all nodes of the networks, it identifies clusters of nodes highly connected to one another and hence it discriminates different clusters as distinct communities. The calculation of clustering coefficient is detailed in [71]. In protein oligomers, the protein interface is made of many bonds between two adjacent chains and few bonds between non-adjacent chains. Hence the interface makes a cluster in terms of graph and hotspots have been successfully identified by clustering coefficient [85].

Protein interfaces are either analyzed based on all interfacial residues or on hotspots only.

2.2. Features of the intermolecular contacts at protein interfaces

To infer the features of protein interfaces, the method is simple: a dataset of protein oligomers/protein interfaces is built, an algorithm is applied to each one of the interfaces to identify intermolecular contacts and the features of the intermolecular contacts are analyzed using statistics. Classically the parameters to describe a protein interface are: (i) interface size, expressed either in number of amino acids or in ASA, (ii) the number of regions of interfaces over the full-length chain, (iii) the chemical properties of the amino acids (amino acid frequency, the interface propensity, namely the frequency of a residue a_i in the protein interface divided by its frequency in a reference set, generally the full-length chain). The size of protein interfaces is an important parameter because it may vary depending on the strength of the association [86].

Evolutionary conservation, protein folds, secondary structures, quaternary structures or crystallographic B-factors can also be considered depending on the question and the criteria used to build the dataset. The idea is to find enough specific features to distinguish the residues of protein interfaces from the rest of the residues.

2.2.1. Dataset based on features of the full-length protein

Many dataset are built on proteins sharing properties at the level of their full-length chains (function, organism, superfamily, folds, and quaternary structures) but without necessarily sharing features at the level of their interfaces [39, 87-89]. In particular, the geometries of the interfaces are not necessarily looked at and therefore interfaces with different geometries are often compared. But it is generally assumed that proteins related in terms of folds or functions associate in similar ways. However, a screen over a large dataset of dimers, performed by Keskin *et al.* has shown that a non-negligible amount of protein oligomers have interfaces sharing features although they have different folds and functions [90]. This set is referred in the paper as "type II", following the term "type I" used for protein interfaces sharing features

and derived from protein oligomers having similar fold and/or functions. To establish the determinants of the construction of an interface it is simpler to look at type II interfaces because the pressure of evolution over the fold and the function of the protein chain is alleviated compared to type I interfaces.

Globally, the results of studies on protein interface dataset (mainly type I) revealed the importance of hydrophobic interactions in the formation of protein interfaces, greater residue conservation and chemical property similar to surface residues but packing like core residues [34, 38]. The two latter properties are coherent with the fate of a protein interface. The topology of a soluble protein is defined by surface residues which are accessible to the solvent and core residues which are, on the contrary, buried and inaccessible to the solvent. The amino acids of a protein interface have the solubility requirement of surface residues because the domains of the interface are initially accessible to the solvent to allow binding. To have stable binding, the domains need to minimize void and maximize packing as for the core residues.

If the role of hydrophobic residues is consistent over any dataset, the importance of polar and charged residues in interfaces varies very much between datasets. Altogether this indicates that hydrophobic residues are involved in promiscuous interactions while polar and charged residues yield alternative recognition modes and hence provide each type of interfaces its specificity.

Up to date, there is no single property sufficiently unambiguous to identify the protein interface from the rest of the protein, and considerable disagreement exists on which properties are actually useful. Conservation is an excellent example of a property both widely used and widely debated. De Vries and Bonvin as well as Neuvirth raise the matter of having so many algorithms and the absence of consensus on the parameters truly relevant to the formation of a protein interface [45, 91]. This may well explain the contradictory results on protein interface properties.

Most studies are performed on the features of individual hot spots. Yet protein interfaces result from intermolecular pairwise interactions and are likely encoded at the pair's level. Supporting this view, the few studies investigating the features of pairs of hotspots show sufficient specificity of the residue pair preferences for accurate prediction [40, 58, 92, 93].

2.2.2. Dataset based on features of the interfaces

To investigate properties responsible for interface formation, an alternative is to build a dataset based on the features of the interfaces and not on the features of the full-length chain. For instance, one can build a dataset of proteins sharing the same geometry of interface.

2.2.2.1. β-strand geometry

Interfaces made of two interacting β-strands (intermolecular β-strands) have been largely studied because it is present in many conformational diseases such as Alzheimer's disease, Parkinson's disease or serpinopathies [94-98]. Supporting the view of comparing interfaces with identical geometries, the proteins involve in conformational diseases share no functions,

http://dx.doi.org/10.5772/58576

fold, and quaternary structures yet they have a common local fold involved in the intermolecular contacts that lead to fiber formation. Their pathological form, whether a fiber or an oligomer, involves interactions between two β-strands, each provided by a different chain (intermolecular β-strands). These intermolecular β-strands share several structural properties. They are recognized by the same antibody A11 [99]. Their formation depends on interactions between atoms of the backbone, result which has led to the proposal that aggregation is a generic property of the polypeptide chain [100, 101]. They adopt a cross β structure which can be predicted from sequences by the PIRA (Parallel 'In Register' Arrangement) model, a network made of single pairs of residues [102-107]. Different predictors of the aggregation-prone sequences involved in the fiber formation are now available [96, 98, 108-111].

We have studied a dataset of 1056 interfaces present in 755 protein oligomers not known to be involved in conformational diseases [59]. As others, we found no specificity at the level of individual hot spots. The chemical properties of the individual hot spots and their distribution on the sequence characterize only the secondary structure and the solubility of the β-strands. In contrast the interaction pairs provide the interface some specificity. Interestingly, the interfaces are best described by two sets of interaction pairs, pairs involving backbone atoms made essentially of hydrophobic and/or small residues and pairs involving at least one atom of the side chain, preferentially made of charged, polar, long and medium residues. The backbone pairs have properties common to intramolecular β-strand interactions and intermolecular β-strands involved in fiber formation in terms of amino acid preferences. Thus hydrophobic amino acids whether in pairs or as individual are not giving any specificity to interfaces. That explains that they always appear in any dataset. On the other hand the side chain pairs have particular geometrical characteristic in terms of number of atoms, branching and length. They also show preferred chemical pairing different from those measured for β-fibers [98]. However this result is only based on comparison with the literature and it could be due to differences in the datasets and/or the algorithms.

The geometry of the side chains has been so far neglected when it appears in our study as a key parameter. Using Steiner Minimal Tree approach (SMT), MacGregor *Smith et al* proposed an elegant geometrical representation of the amino acids that was successfully applied to the problem of protein folding [112]. It will be interesting to extend this approach onto protein interfaces to see if the specificity of protein interfaces may be provided by the geometry of the amino acids rather than their sole chemistry. Similar double layer of interactions, has been observed at the interfaces between colicins and their cognate immunity proteins [113]. One set of the intermolecular residues was common to all colicin-immunity members and produced a low binding affinity between the colicin and its cognate immunity protein while the other set was made of variable residues providing high affinity and specificity to the colicin for a particular cognate. Double layer of interactions has also been reported in monomeric proteins (intramolecular networks) [84].

As mentioned earlier, proteins and protein interfaces are now described as networks of amino acids in interaction or as sub-networks of hot spots in interactions, respectively. This relatively new concept offers the possibility of looking at the layout of interactions in addition to the amino acid properties. It is clear now that the network of interactions is as much important as

the components of the network in providing the protein its properties in terms of folding, function, evolution and interface formation [59, 78, 84, 114-117].

We have observed in our 1056 β-strand interface dataset that the side chain pairs also have specific network features. The side chain hot spot sub-networks have nodes with more contacts than the backbone or the PIRA (Parallel In Register Alignment) networks, used to predict fiber sequences [98]. Yet the β-strand interfaces have no hubs and maintain a low interconnectedness (little communication between residues of the interface), probably a mechanism to resist the effect of mutation by secluding the nodes of the networks (Fig. 7). Simultaneously, the side chain residues make less than three contacts avoiding stringency on the choice of amino acids capable of making an interface and providing the β-strand interface high sequence plasticity. Robustness and plasticity of networks are well explored by graph theory and there are several very inspiring papers on that topic [83, 118-120]. This point is discussed in more details later in the chapter.

2.2.2.2. α-coiled interfaces

To date the only interfaces accurately predicted from sequences are α-coiled interfaces [121-124]. Intermolecular residues follow a so-called knobs-into-holes regular packing producing the α-coiled coil helix-helix assembly [125]. In the simplest case (dimer), the α-coiled coil sequence displays a repeat pattern of seven amino acids so-called heptad repeat, labeled *abcdefg*, with hydrophobic residues at the *a* and *d* positions (Fig. 2C). These hydrophobic intermolecular contacts constitute the seam of the core of the knobs-into-holes interface. The repeats can be shorter than 20 residues or span many hundreds of amino acids.

There are obvious reasons to why it has been possible to understand α-coiled coil interfaces when other geometries still elude us. First, α-helices are geometrically more constraint than β-strands and second backbone interactions do not participate in α-helix interfaces because the hydrogen bond networks are made intra-molecularly. Hence, there is no "backbone noise" information that interferes with the side chain information.

2.2.3. Interfaces and quaternary structures

The quaternary structures of the protein oligomers and the features of their interfaces are related and different methods are currently developed aimed at understanding such relations [126, 127].

In some cases such relation is more or less understood. For example, in higher-order α-coiled coil oligomers (above dimer) additional (peripheral) knobs-into-holes take place and broaden the helical contacts [128]. Such multiple repeats lead to multi-faceted helices, which combine repeats of different amino acid compositions to accommodate quaternary structures accordingly [129, 130]. Thus, it is possible by analyzing amino acid sequences to predict the quaternary structures of α-coiled coil oligomers [129-132].

The relation between the interfaces features and the quaternary structure is less understood in β-strand interfaces with few exception as the legume lectin family (81-82). Combining

clustering algorithms with sequence alignments, motifs of sequentially and structurally conserved residues are detected at the β-strand interfaces of lectins. The different motifs are built on a subset of residues at the interface that provide a specific 3D-orientation of the β-strands. Consensus patterns at the interfaces have been found for the different quaternary structures of the lectins. Briefly, there are nine different kinds of quaternary structures in legume lectins including Canonical, ECorL-type, GS4-type, DBL-type, ConAtype, PNA-type, GS1-type, DB58-type, and Arcelin-5-type (monomeric). Seven different consensus is observed including types II (canonical), X1 (DB58-type), X2 (noncanonical interface of ConA), X3 (ECorL-type, handshake), X4 GS4-type, back to back), and the unusual interfaces of PNA and GS1.

For a long time, there are experimental evidences of sequences that are responsible for the quaternary structure of protein. These sequences, called registration sequence, are located upstream the interface region and promote oligomerization of monomeric protein when genetically added [133]. Proline and histidine residues located upstream of interfaces have also been shown to regulate the association between chains [8, 11, 134-136]. Collagen α-fibers and silkworm/spider β-fibers contain several repeats composed of proline residues which also participate in the quaternary structures. Whether these residues belong to the interfaces or are systematically located outside the interface regions has not yet been established.

In summary, residues located within the interfaces and outside are participating in the quaternary structures and the chain assembly. This implies that protein assembly is regulated at two levels, at the level of intramolecular interactions (residues outside interfaces) and at the level of intermolecular interactions (residues in interfaces). Thus it is necessary to also investigate the residues involved in intramolecular interactions to discriminate those participating in folding reactions from those participating in both folding and interface formation. The latter residues are probably coordinating the whole assembly process by regulating communication between folding and association steps.

3. Intermolecular and intramolecular amino acid interactions: The mechanism of protein assembly

As mentioned at the beginning, protein assembly or protein oligomerization entails folding and association reactions. Thus, to have a full picture of the mechanism of assembly, besides studying intermolecular amino acid interactions, it is necessary to investigate intramolecular amino acid interactions and to apprehend how both types of interactions are coordinated. Different models of assembly have been recently reviewed [137, 138].

First, let's consider the simple case of the formation of a dimer. There are two routes to a dimer. One is through the three states model where unfolded monomers *U* (state 1) fold into monomers *M* (state 2) which subsequently associate into dimers *D (state 3)*. The alternative route is through the two states model in which unfolded monomers *U* (state 1) associate into folded dimers *D* (state 2). Intramolecular and intermolecular interactions occur sequentially in the three states model but concomitantly in the two states model. One can anticipate that folding

and association are going to be related but independent in the three states model but concerted in the two states model. In terms of networks, one can speculate that the three states model suggests a protein organized in two sub-graphs remotely connected, one governing the intramolecular interactions and the other the intermolecular reactions. On the contrary the two states model suggests two connected sub-graphs.

Discriminating the route of assembly is crucial in term of drug design strategy. In the three states model, it is likely that the interface in the folded monomer and in the folded dimer is similarly organized. Thus, it is consistent to use the x-ray structure of the native protein oligomer as a template to design assembly inhibitors. In contrast, in the two states model the interface in the unfolded monomer is different from the interface in the folded dimer so assembly inhibitors designed on the native structure of the protein oligomer are unlikely to recognize the unfolded monomer and block the assembly at the monomeric stage (or at early stage). This is one illustration on why it is important to anticipate the mechanism of assembly.

3.1. Can the mechanism of assembly be predicted by investigating evolutionary relationship?

D'Alesio offers good historical reviews on this question [139, 140]. The interfaces of dimers assembling by a two states model are found to share patterns with intramolecular interactions observed in monomers. It is proposed that such dimers have evolved from mutations in an existing monomer that led to its unfolding, followed by further mutations that yielded a viable dimer with intermolecular interactions similar to the intramolecular interactions present in the initial monomer. This mechanism is reminiscent of the domain swapping mechanism which is well presented in the chapter by Giovanni Gotte and Massimo Libonati or reviewed in [141]. In such a situation the evolution to the dimer depends initially on the evolution of a viable monomer towards unfolding induced by random mutation. There wouldn't be any folded monomer in the assembly route because it wouldn't bear the mutations in a folded state. The evolutionary route between the dimer and the monomer suggests the presence of epistatic mutations (mutations that have different effects in combination and individually). Here the fold and association steps are related and dependent on one another, in term of evolution and mechanism of assembly. In contrast, dimers assembling through a three states model were not found to share motifs with monomers. The folding of the monomer might then be a natural route towards association, and folding and association would appear evolutionary and mechanistically related but independent.

Next, let's look beyond the simple case of dimers. Possible relations between evolution and assembly mechanism have been further exploited by looking at protein oligomers sharing the same functions (superfamily) but adopting different quaternary structures. In one study the authors exploit the symmetry of oligomers to establish a relationship between evolution and assembly mechanisms [142]. From an initial screen of 5375 PDB structures, they found tetramers with D_2 symmetry having homologous dimers with C_2 point group symmetry and hexamers with D_3 point group symmetry homologous dimers with C_2 symmetry or homologue trimers with C_3 point group symmetry. In total 49 protein oligomers with a symmetry relation from C_n to D_n are reported. They found evolutionary links between the C_n and D_n counterparts

http://dx.doi.org/10.5772/58576

and isolated experimentally *the in vitro* disassembly C_n intermediates for ten of the D_n oligomers. In addition, for five cases, the C_n intermediates were also shown to be formed during *in vitro* reassemblies. They concluded that the evolutionary and assembly pathways were related and that assembly intermediates could be predicted solely from the atomic structure.

But this conclusion might be over optimistic because based on very little cases (49/5375) and because protein structure evolution and protein assembly are plastic in terms of mechanisms such that to date it remains difficult to establish either route.

For example a single mutation has been found responsible for a transition from Cn to Dn point group symmetry, it is not obvious how such global change could have been anticipated by simply considering the full-length 3D-structure of the wild-type protein [143]. On the other hand it tells that the protein assembly is regulated by local properties since a single mutation is enough to alter the global assembly. This strongly suggests that the solution lies in understanding the local properties and how they propagate information to regulate the global shape.

Hemoglobin is another complex example of a protein sharing a function but distinct quaternary structures and for which evolutionary and assembly routes are not easily drawn even if the structures are available (Fig. 4).

The *Synechocystis cyanoglobin* produces a monomeric hemoglobin (PDB 1S69) with a C_1 point group symmetry, the human hemoglobin is tetrameric (PDB 2HHB) with a C_2 point group symmetry, the *Oligobrachia mashikoi* produces a 12-mer hemoglobin (PDB 2ZS1) with a D_3 point group symmetry while the giant earthworm hemoglobin contains 144 chains (PDB 2GTL) with a D_6 point group symmetry. In such case, the different hemoglobin point group symmetries and quaternary structures may result from coding constraints of their respective organisms rather than from a relation in terms of evolution or assembly intermediates. As described by Crick in the early 60s, symmetric assemblies require fewer distinct kinds of specific interaction interfaces compared to asymmetric assemblies [144]. Likewise, higher symmetries require fewer distinct interfaces compared to lower symmetries and thus, the smaller a genome the more often its protein structural complexity may rely on high symmetry. This is consistent with the large occurrences of proteins with icosahedral symmetry in viruses while most eukaryotic molecular machines have C1 symmetry (Fig. 1B). Now, some high point group symmetry oligomers have been also discovered in eukaryotes, bacteria and archea with the vault proteins assembling 78 copies (2zuo, 2zv4, 2zv5), the encapsulin (3dkt) and the vault from *Pyrococcus furiosus* (2E0Z). Clearly one has to be cautious in interpreting data and statistics derived from the PDB.

In addition, hemoglobin is also an interesting example of how a unique function is provided by a combinatory of assemblies using the same protein fold. There are many other examples (e.g. ferritin, rubisco) but probably α-coiled coil oligomers offer the largest combinatory of assemblies. It was recently shown that they formed before LUCA (last universal common ancestor), by independent routes and most likely as the result of all possible geometric solutions to packing helices in a stable way [145]. Again, that illustrates how diverse evolution routes are.

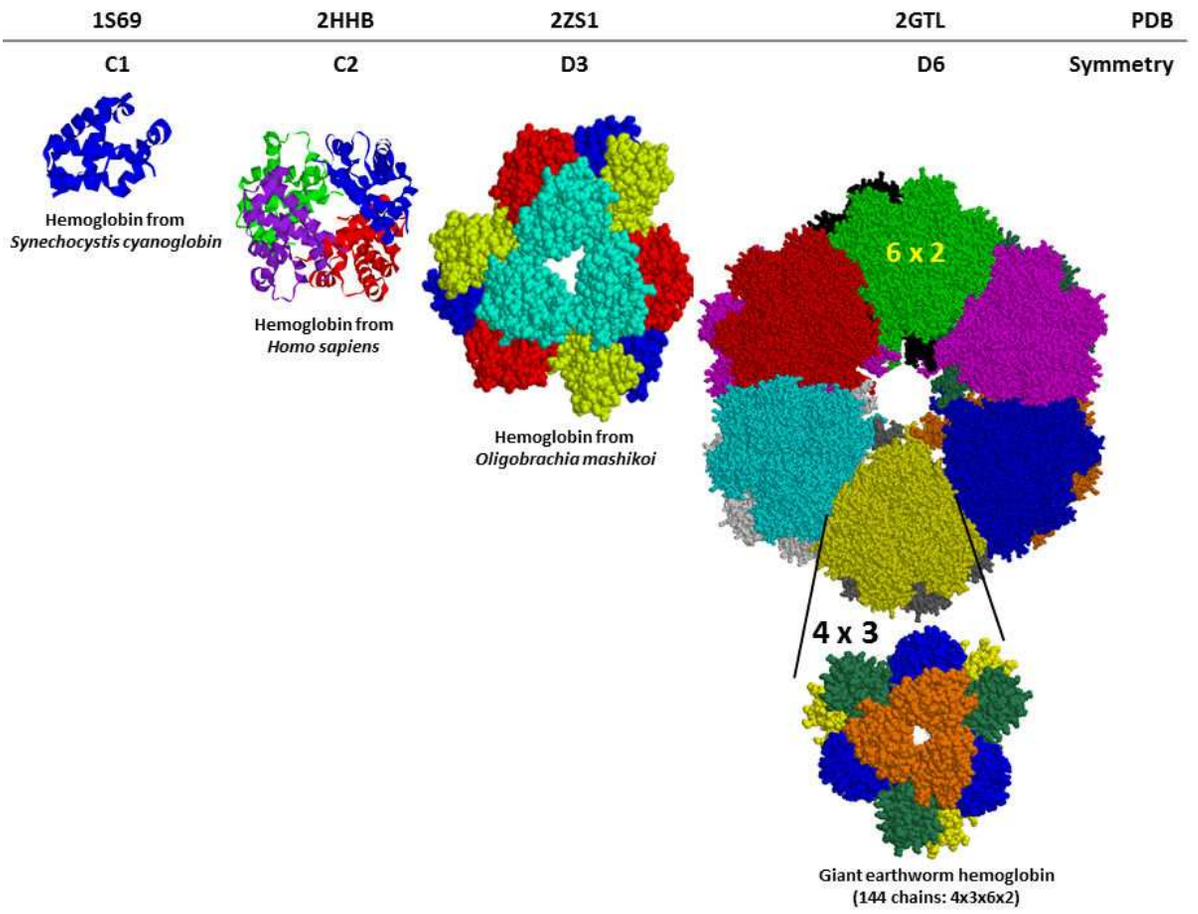

Figure 4. Plasticity of quaternary structures fulfilling a single biological function: the hemoglobin example. The hemoglobin chain exists as a single fold which is copied and assembled with different stoichiometries (number of chains) and different symmetries across species to maintain the same biological function. Few cases are represented from a hemoglobin monomer to a 144-mer assembly. The structures are shown in ribbons except when spacefill is better to illustrate the symmetry of the assembly. The pictures are generated with Rasmol. The PDB codes and the symmetries of the hemoglobins are indicated above their respective structures.

The relation between evolution and assembly routes assumes that an oligomer evolves/assembles from a monomeric entity. But reverse situation exists as for the native tachylectin-2 monomer which has been proposed to have evolved from a pentameric ancestor through short, functional gene segments that, at later stages, duplicated, fused, and rearranged [146]. The authors propose that new folds evolved through the structural plasticity of assembly intermediates.

This last example illustrates quite ironically that protein folds and quaternary structures still hold surprises and a direct relation between the evolution of protein oligomers and the mechanism of their assembly is not readily systematic. Both evolution and assembly certainly involve multiple parameters making their prediction rather challenging.

3.2. Can the mechanism of assembly be predicted by experimental approaches?

The two and the three state models are depicted in the figure 5A.

http://dx.doi.org/10.5772/58576

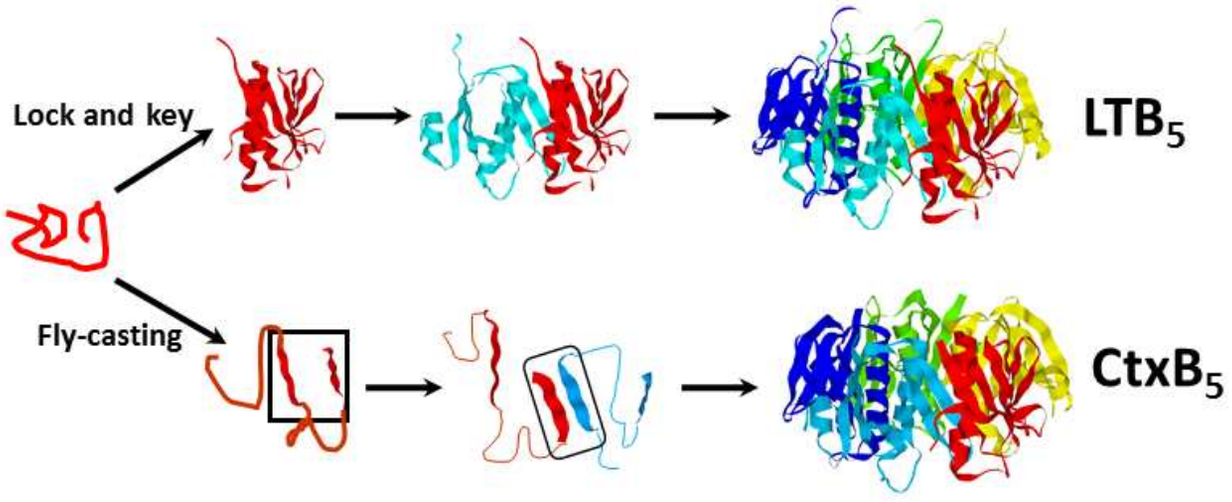

Figure 5. Mechanisms of assembly. Two mechanisms of assembly have been described and experimentally observed. The protein chain folds before association in the lock and key mechanism, also called the three states model because the protein can be observed in three states, unfolded monomer, folded monomer and native oligomer (top route). The protein chains associate in a more or less partially folded state, and only subsequently acquire native folded conformation, in the fly-casting mechanism also referred to as the two states model because because the protein exist only in two states in a dimer, unfolded monomer or native dimer. These two models are illustrated with the assemblies of the two related AB_5 toxins, the heat labile enterotoxin B pentamer (LTB_5) and the cholera toxin B pentamer ($CtxB_5$). The two toxins share 94 % sequence identity and almost superimposable atomic structures but nevertheless assemble through two different mechanisms.

The three states model is the oldest and most classical mechanism observed, it is generally referred to as the lock and key mechanism. There are plenty of experimental evidences of both the two and three states mechanisms. Non-native oligomers, namely oligomers with native quaternary structures but not native folds have been isolated experimentally for a long time and are common intermediates of assemblies [69, 70, 133, 134, 147-153]. Such intermediates are typical border line cases as they might be produced by a lock and key mechanism or by a fly-casting mechanism. There are clear examples of protein associating by a fly-casting assembly with unfolded monomers able to associate [8, 134, 151, 154, 155]. The RING domain protein family of scaffolding oligomers presents an interesting case of the formation of a stable partially folded assembly tetramer along the oligomerization route to a native 24-mer [156-158]. The C_4 symmetry tetramer populates because of its fast formation from monomers and its slow disappearance into a D_4 24-mer (6 x 4). The transition to the D4 symmetry 24-mer is rate-limiting, because of the slow folding Proline cis/trans isomerization that regulates the association of two monomers via the ligation of Zn sites. Likewise dimer, trimer and tetramer assembly intermediates are isolated along the route to the native cholera toxin B pentamer ($CtxB_5$) because the formation of one of the toxin interface is regulated by a *cis-trans* Proline isomerization [8]. The CtxB assembly intermediates acquire some of their native secondary structure along with association because their main interface involves the formation of an intermolecular β-sheet, this folding/association step is regulated by histidine residues [134].

Proline and histidine residues are rare at interfaces but are often found upstream the region of interfaces and are indirectly acting on their formation, as mentioned at the beginning of the

chapter (see introduction). Registration sequences that control the quaternary structure of protein oligomers are also located outside interfaces. In fact, several cases of residues located outside interfaces have been shown to be involved indirectly in the chain association and a variety of small amino acid modules have been proposed to act upon assembly by different processes. Basically they introduce the flexibility required to modulate the 3D position of interface domains so to increase the chance of successful encounters [138, 159].

3.3. Can the mechanism of assembly be predicted by computational approaches?

The two states model was revisited by Wolynes' laboratory showing that an unfolded protein has a greater capture radius for a specific binding site than the folded state with its restricted conformational freedom [160]. In this scenario of binding, the unfolded state binds weakly at a relatively large distance followed by folding as the protein approaches the binding site: the ''fly-casting mechanism'' (Fig. 5A). In 2004, Wolynes introduces the notion that certain characteristics of the atomic structure like the interface size and hydrophobicity, the ratio of the number of interfacial contacts to the number of intramonomeric contacts enabled to determine whether a homodimer assembled into a fly casting or lock and key mechanism [161]. A large ratio of interfacial to monomeric contacts is typical of a two-state model and of the fly-casting mechanism.

Computational approaches also provide evidences supporting the lock and key mechanism, the fly-casting mechanism and a series of in-between mechanisms attesting of back and forth between folding and association reactions and whose idea lies on an "induced-fit" principle during which intermolecular contacts "catalyze" folding (allostery, conformational gating, induced fit) [162-165].

Recently, molecular dynamic (MD) simulations have been combined to network analysis to provide detail understanding of the route of assembly. For example, coarse-grained transition networks (CGTNs) can be derived from MD simulation to show the transition between oligomers of different sizes [166, 167]. In a recent report, the role of the sequences in the aggregation kinetics and assembly mechanisms was described in great details [168]. Briefly, MD is performed and the state of each conformation/state observed in the MD is defined by a set of digit. Based on the MD, a transition matrix $N \times N$ is built with N states and with the matrix elements defined by the occurrences of any transition between two states. The matrix transition is converted into a graph called KTN (Kinetic Transition Network) with the nodes corresponding to the states and the edges to the transitions. Such graphs provide measures of the population of different states and the probability of transition between them. Energy barriers are associated to the transitions and disconnectivity graphs are constructed to evaluate the energy barrier to go from one conformation to another with min-cut algorithms. The dynamics of aggregation was also evaluated using FPTD (First Passage Time Distribution) which informs on the most populated states and kinetics. Although such approach has not yet been applied to a protein assembly on a full-length protein, there is no doubt that such combination of molecular dynamics with graph theory would provide new directions in predicting protein assembly mechanisms.

http://dx.doi.org/10.5772/58576

In fact, graph theory is the ideal tool to investigate the residues involved in intramolecular interactions, the residues involved in intermolecular interactions and their cross-talk communications. Basically sub-graphs or clusters are produced and allosteric communication between the different clusters is investigated. This has been used in enzyme/ligand intermolecular interactions and in interfaces [74]. It appears that the intramolecular networks maintained the robustness of the structure while the interface residues are more plastic to accommodate the flexible motion required for association.

Obviously folding and association reactions intertwine to orchestrate the protein assembly. This means that the key factors for protein assembly is the balance between intra and inter molecular interactions.

4. Local key contacts regulate global conformations

There are cases of proteins sharing functions, high sequence identities, folds and quaternary structures but following distinct assembly mechanisms. For example the two related AB_5 toxins, heat labile enterotoxin (LTB_5) and cholera toxin ($CtxB_5$) have 94 % sequence identity, almost superimposable 3D structures, and identical quaternary structures but nevertheless assemble through different mechanisms under identical experimental conditions (Fig. 5A). LTB_5 follows a lock and key mechanism whereas $CtxB_5$ assembles through a fly-casting mechanism [8, 134, 135]. Out of 103 amino acids, 11 are different among which only two in the interface. The cpn10 heptamers are another of such example [70].

The role of only few residues in controlling an assembly or a disassembly mechanism is also evidenced in diseases called conformational diseases where a single amino acid mutation is enough to redirect the protein native conformation to an aberrant conformation such as a fiber, through unfolding/refolding steps [169-175]. Consequently the protein loses its function leading to the disease.

This tends to show that the assembly of a protein is in fact regulated by only few amino acids, indicating that little differences are enough to go from a fly-casting to an induced-fit mechanism. This is in good agreement with allosteric mechanisms and the MWC (Monod, Wyman, Changeux) theory that unifies fly-casting and an induced-fit routes into a single mechanism [176]. Accordingly, protein assembly can be expressed as a single scheme with transitions between the fly-casting and the induced-fit mechanisms depending on thermodynamic equilibrium and kinetic rates (Fig. 6).

There exist several evidences of such transitions in biology, some of which are illustrated on figure 7. For example, in the course of evolution proteins may change their folding and/or assembly routes upon random mutations. Or proteins very similar in sequences and structures may favor different routes because of small amino acid differences in their sequence and/ or environmental factors. This illustrates the plasticity of proteins in terms of mechanism of formation and in terms of quaternary structures but also supports the fact that it is the local characteristics (few amino acids) that impact on the global structure of a protein.

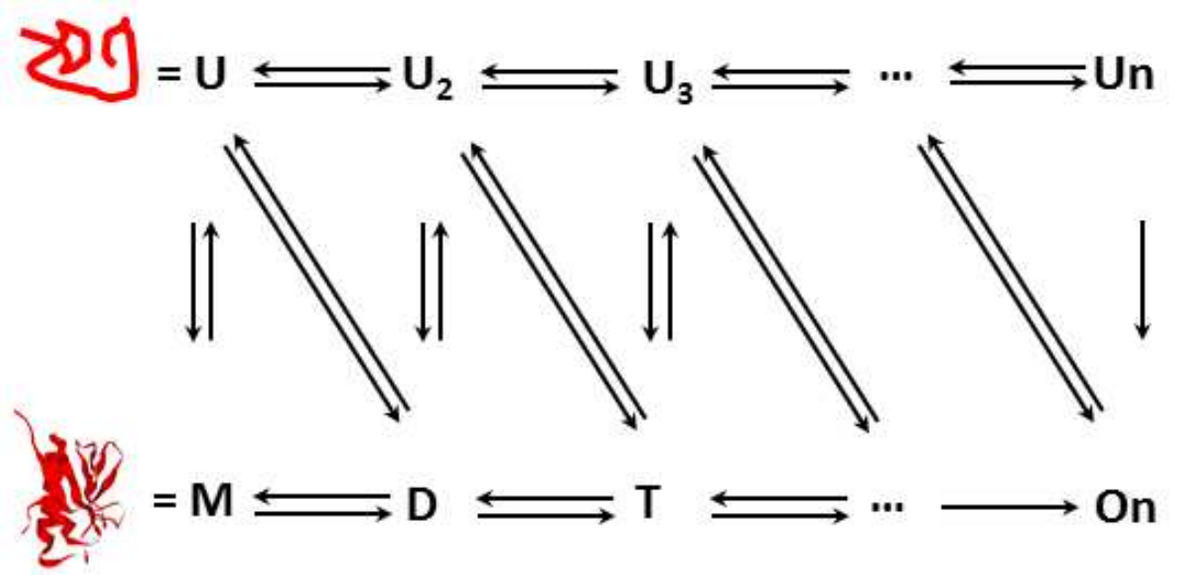

Figure 6. Kinetic scheme of protein assembly. A protein monomer may exist in an unfolded state ***U*** and folds into a folded state ***M***. A protein oligomer may assemble from (top reactions) unfolded monomers ***U*** which associate in "partially"folded states ***Ui*** with *i* going from 2 to *n*, *n* being the number chains; until they assemble into a non-native oligomeric state *Un* which finally folds into a native folded oligomer *On*. Alternatively, a protein oligomer may assemble from (bottom reactions) unfolded monomers ***U*** which fold and associate into dimers *D*, trimers *T,* etc until they reach the native oligomeric state *On.* Each of the conformational state exists in equilibrium and may go from one state to another according to k_{on} and k_{off} rates of the reaction. Only the formation of the native state are considered irreversible. The population of every species and the transition species depend on kinetic parameters.

How to identify few amino acids as key determinant for a protein fold or an assembly and what properties they must have to affect the mechanism of assembly and/or its final output? Graph theory is at present probably one of the most suitable approach to investigate such questions. For example, the effects of the mutations involved in conformational diseases have been considered in terms of network. Recently, a novel approach using graph-based signatures has shown that the impact of a mutation correlated with the atomic-distance patterns surrounding an amino acid residue [177]. They showed that the signatures can be used to predict stability changes of a wide range of mutations occurring in the tumor suppressor protein p53.

We have also investigated the effects of mutation on graph features and the possible consequences in terms of the disease mechanism [59]. As briefly mentioned earlier, we have seen that the networks of the 1056 intermolecular β-strands present in "healthy" protein oligomers, avoid hubs (highly connected residues) to be robust to mutation. The intermolecular β-strands are essentially disconnected graphs so any mutation would not spread damages far in the network. We compared these "healthy" networks with the β-strand interface of the p53 tetramer, which has known familial mutations related to dissociation of the tetramer, fiber formation, and associated with cancer [169]. The p53 network has a higher interconnectedness because its nodes have higher degrees (ie more contacts) than those in the "healthy" networks, with the consequences that a single node modification (ie a mutation) is enough to reorganize the interactions in the entire network. Thus the higher connectivity of the p53 network leads to a greater sensitivity to rewiring (rearrangement of links upon node modification) than the

http://dx.doi.org/10.5772/58576

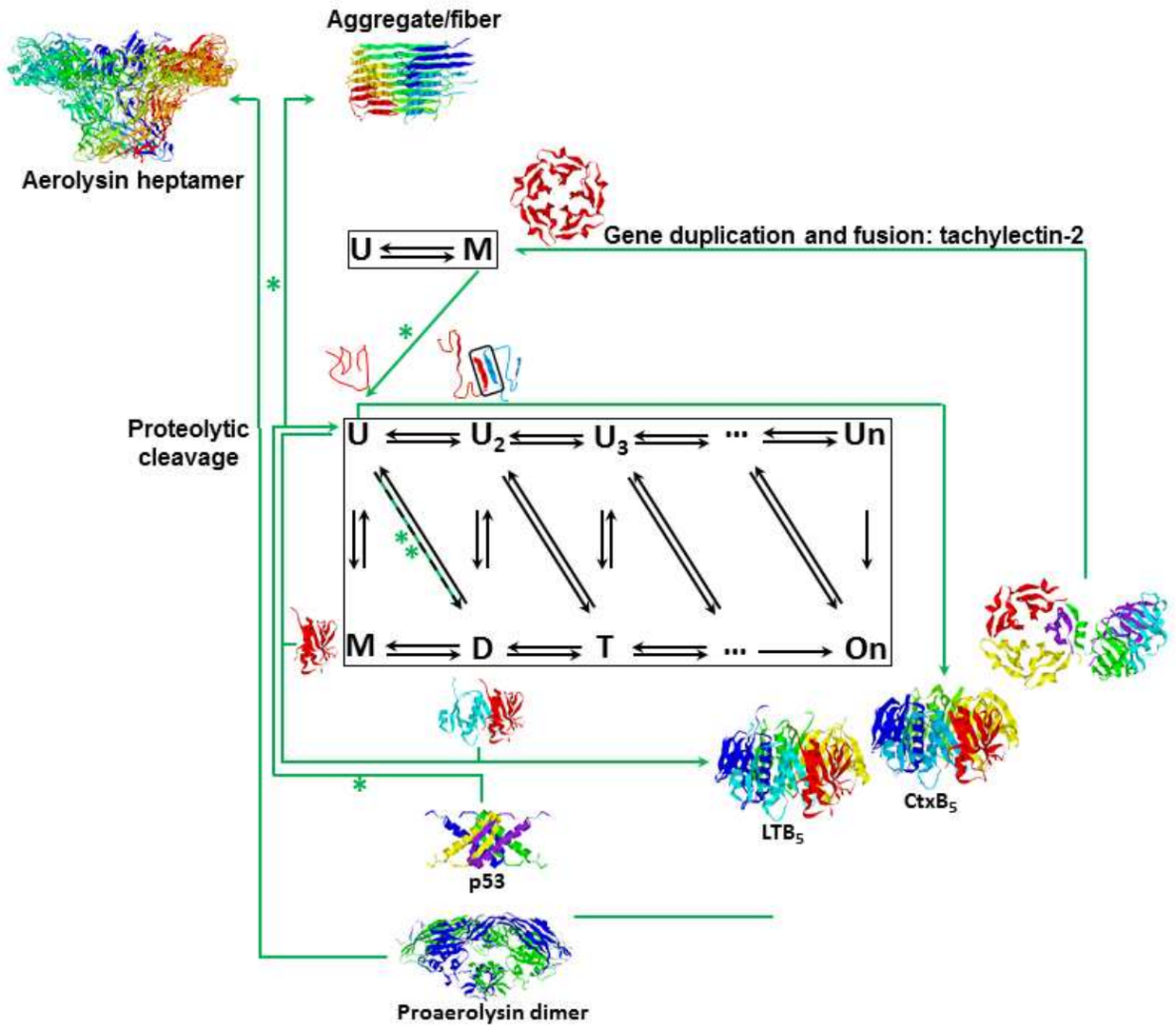

Figure 7. Evidences of transitions between different states and different reaction paths. The kinetic scheme described in figure 6 is reported here at the center of the figure. The transitions between different protein conformations or different reaction paths are indicated by green arrows. The x-ray structures of protein cases undergoing such transition are shown as illustration of these transitions. The transition can take place during evolution and because of mutation (e.g. tachylectin-2, two states model). It can take place because of mutation and lead to conformational diseases as for the p53 tumor suppressor p53. The main route may depend on a difference of few amino acids as for the two related toxins $CtxB_5$ and LTB_5. Or else, the protein as the pore-forming toxin aerolysin may adopt different quaternary structures and go from one to another because of environmental factors (e.g. pH, proteolyic cleavage, cell receptor etc…).

disconnected graph observed in healthy proteins. In some cases, such ample rewiring probably promotes chain dissociation, first step to fiber formation. We have now started to investigate why the p53 network has a higher interconnectedness than the "healthy" networks. The p53 tetramer has a D_2 point group symmetry and its interfaces adopt a local central symmetry because the two interacting domains have identical sequences and the residues are paired in an anti-parallel manner. In contrast 60 % of the protein interfaces of "healthy" proteins have domains made of different sequences and their β-interfaces have no local symmetry.

Let's consider three intermolecular networks, one with different amino acid sequences and no local symmetry, a second with identical sequence arranged in a parallel manner (horizontal axe symmetry) and a third with identical sequence arranged in an anti-parallel manner (rotational axe symmetry) (Fig. 8).

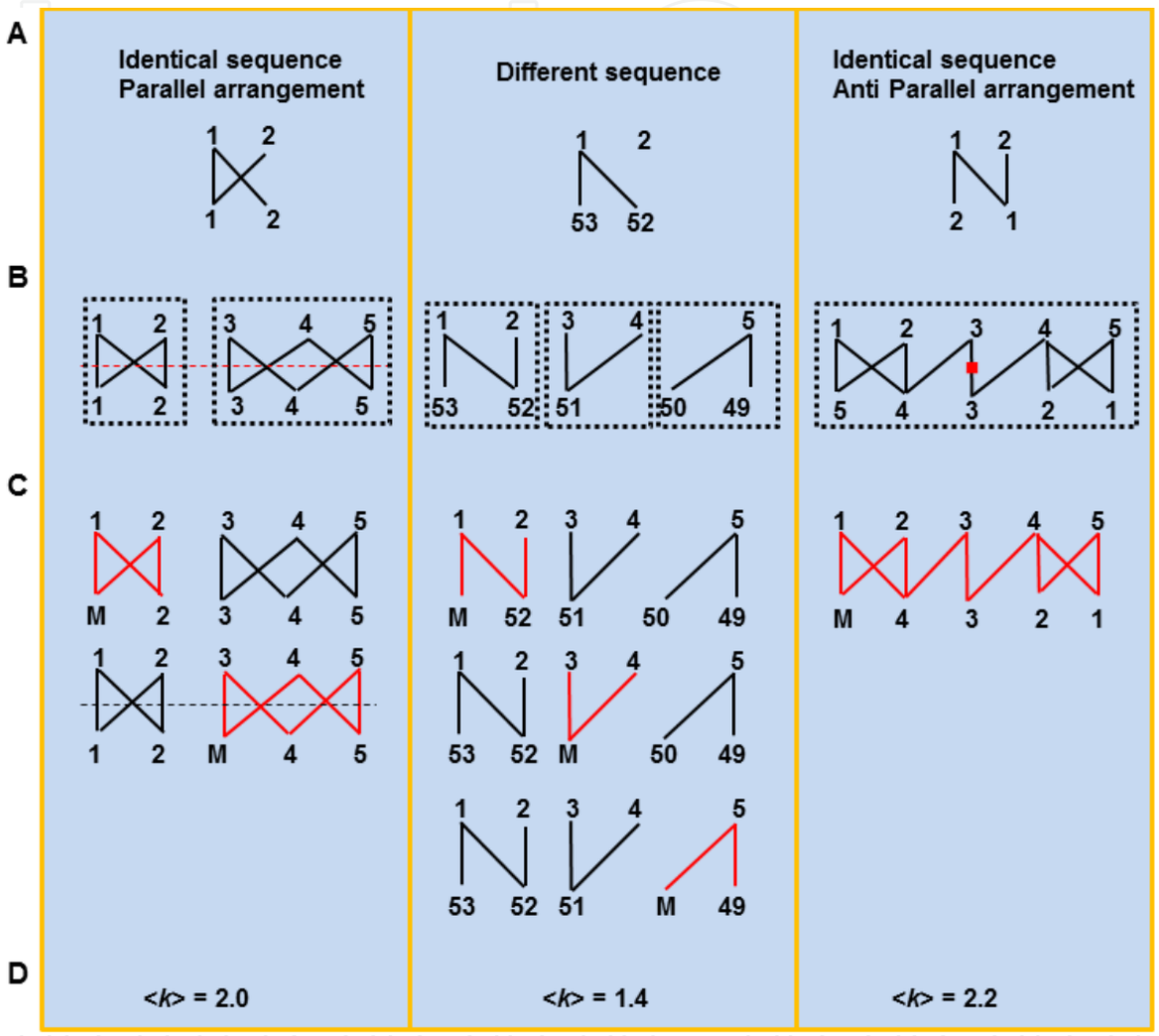

Figure 8. Effect of local symmetry on network features.A. Protein interfaces may be formed by association of domains with identical sequences (left and right pannels) or different sequences (middle pannel). In the former cases, the two domains maybe aligned in a parallel or antiparallel manner (in register or out register arrangement). The identical sequences will have an intrinsic symmetry in their amino acid pairing, namely if the residue 1 is in interaction with the residue 2 then the residue 2 is in interaction with the residue 1. Such symmetrical constraint will produce some motifs in the network which are not necessarily present in "asymmetrical" interfaces made of domains with two different sequences. This is illustrated on a simple network. **B. Connected components.** Considering a slightly more complex network one can see that the motifs results from the elements of symmetry, a horizontal axial or a rotational axe-symmetry for a parallel or antiparallel arrangement, respectively. The dotted boxes indicated the connected components, namely the residues which are connected to each other. One can see the effect of the symmetry on the total number of connected component. **C. Propagation of changes.** The effect of a single node modification on the network, indicated by a M for mutation, is considered. Assuming there is an effect as long as there is a link between two nodes, the symmetry enables the changes (red link) to propagate within the network. **D.** The average degree <*k*> of the nodes is given for each of the networks.

http://dx.doi.org/10.5772/58576

Let's now look at the consequences on the network features. The first consequence is a multiplicity of the number of interactions in interfaces with local symmetry and when the sequences are identical and therefore an intrinsic increase of the network interconnectedness (Fig. 8B). As observed for the p53 case, such increase would lead to network sensitivity to rewiring which in terms of protein may introduce a vulnerability to chain dissociation or chain reorganization. The second consequence is the decrease of the number of distinct connected components (Fig. 8B). This again would increase the propagation of changes within the network upon mutation because the amino acids are not secluded from one another. It is interesting that local symmetry is enough to improve the communication within the network without altering significantly the average degree <*k*>. This means that even without hubs the protein interfaces become highly connected by long paths. This preliminary analysis suggests that interfaces made of domains with different sequences might be more resistant to fold plasticity because of an absence of sequence symmetry. Protein oligomers which undergo a transition to pathological assemblies (fiber or oligomers) probably have global and local properties that make them amenable to fold plasticity. How the local properties alter the global properties remain to be explored.

5. Conclusion

The novel results obtained by graph theory are that the layout of the interactions, called the network topology, is extremely important for understanding the formation of an interface and the plasticity of fold and quaternary changes. It is also important to understand that the keys are not in any hot spot features but are in the residues whose local properties spread enough global effects to regulate/affect the full-length chain structure. In other words, the formation of interfaces and the quaternary plasticity lay on the residues that control allosteric transitions, mechanisms now revisited using propagation measures in networks. This local to global transition is also investigated by mathematical concepts in the chapter by Laurent Vuillon and Claire Lesieur.

The take home message of the chapter is to exhibit the usefulness of computational approaches to efficiently complement experimental approaches and gain insight in protein assembly. Obviously, future challenges are on understanding how intramolecular and intermolecular interactions are coordinated and the determinants of allosteric transitions. Graph theory and networks approaches open new venues to explore such problems and are certainly going to provide important breakthrough. Briefly, graph theory can help in identifying intramolecular and intermolecular key interactions as well as in investigating their communication means by analyzing the topology of the networks, isolating appropriate clusters and determining propagation route (allostery).

Now, it may be yet too early to grasp what are the network measures most relevant to the problem of protein assembly and how they can be interpreted in terms of protein's needs. For example, proteins and protein interfaces have been described as random networks with Poisson degree distributions centered to a characteristic average <*k*> degree [178]. This means

all nodes have on average the same number of links (or contacts) and there are no hubs. Proteins have also been described as single-scale network with exponential degree distribution and no hubs again [59, 84]. In some case, the random network is attributed to the backbone interactions while the single-scale network is attributed to the side chain interactions. A network made of a minimum number of contacts seems rather coherent as proteins probably minimize the number of links (bonds) per amino acid to reduce the "building" cost in terms of bonds and the sequence stringency.

Simultaneously, proteins are described as small world because they have small average path length <*l*>. Small <*l*> generally indicates that most nodes, namely amino acids, of the network are within the reach of each other. Such node accessibility would suggest that a single modification anywhere in the network (ie any mutation in a protein) would easily spread changes in the whole network, a hazardous situation for a protein and in contradiction with the fact that protein folds and functions resist most mutations. Small world networks generally have hubs, highly connected nodes that govern the network communication routes. But proteins are random or single-scale networks and as such are not expected to have hubs, at least not hubs with many more links than other nodes. The absence of hubs is good as it reduces the protein vulnerability to mutation. We have measured <*l*> from 10 to 19 in protein interfaces for networks made of about 300 nodes (unpublished). For comparison the world wide web has similar <*l*>=19, but 800 million nodes. Maybe it just happens that some worlds are smaller than other.

It is therefore not so simple to deconvoluate the topology of a network with a small average <*l*> depending if it is a random, single scale or scale-free (power law degree distribution) network. Theoretical developments aiming at this understanding are proposed and allow considering distribution of connected components, distribution of clustering coefficient and approximation of <*l*>. Such work will help analyzing the network measures obtained for amino acid networks [179].

One problem of network is the number of interactions and nodes generated to describe a protein network and how to discriminate a hierarchy within these set of interactions to understand the determinant ones. To this goal, one elegant strategy is to experimentally measure kinetics and affinity to prioritize interactions in networks [180]. Such approaches would complement MD simulations and help discriminating the good from the bad.

Acknowledgements

A special thanks to Kave Salamatian and Laurent Vuillon for numerous fruitful discussions on network and graph theories. We thank the federation of research FR2914 MSIF (Modelization, Simulations, Interaction Fundamentals) for supporting our work on interdisciplinary research.

Author details

Claire Lesieur*

Address all correspondence to: claire.lesieur@agim.eu

AGIM-FRE3405 UJF-CNRS, Grenoble, France

References

[1] Goodsell DS, Olson AJ. Structural symmetry and protein function. Annu Rev Biophys Biomol Struct. 2000;29:105-53. PubMed PMID: 10940245.

[2] Grant RA, Filman DJ, Finkel SE, Kolter R, Hogle JM. The crystal structure of Dps, a ferritin homolog that binds and protects DNA. Nat Struct Biol. 1998 Apr;5(4):294-303. PubMed PMID: 9546221.

[3] Lloyd DR. Cubic Icosahedra? A Problem in Assigning Symmetry. Journal of Chemical Education. 2010;87(8):823-6.

[4] Cheung DT, DiCesare P, Benya PD, Libaw E, Nimni ME. The presence of intermolecular disulfide cross-links in type III collagen. J Biol Chem. 1983 Jun 25;258(12):7774-8. PubMed PMID: 6863264.

[5] Ricard-Blum S. The collagen family. Cold Spring Harbor perspectives in biology. 2011 Jan;3(1):a004978. PubMed PMID: 21421911. Pubmed Central PMCID: 3003457.

[6] Gordon MK, Hahn RA. Collagens. Cell and tissue research. 2010 Jan;339(1):247-57. PubMed PMID: 19693541. Pubmed Central PMCID: 2997103.

[7] Lesieur C, Frutiger S, Hughes G, Kellner R, Pattus F, van der Goot FG. Increased stability upon heptamerization of the pore-forming toxin aerolysin. J Biol Chem. 1999 Dec 17;274(51):36722-8. PubMed PMID: 10593978.

[8] Lesieur C, Cliff MJ, Carter R, James RF, Clarke AR, Hirst TR. A kinetic model of intermediate formation during assembly of cholera toxin B-subunit pentamers. J Biol Chem. 2002 May 10;277(19):16697-704. PubMed PMID: 11877421.

[9] McLaughlin SH, Bulleid NJ. Molecular recognition in procollagen chain assembly. Matrix Biol. 1998 Feb;16(7):369-77. PubMed PMID: 9524357.

[10] Reimer U, Scherer G, Drewello M, Kruber S, Schutkowski M, Fischer G. Side-chain effects on peptidyl-prolyl cis/trans isomerisation. J Mol Biol. 1998 Jun 5;279(2):449-60. PubMed PMID: 9642049.

[11] Tacnet P, Cheong EC, Goeltz P, Ghebrehiwet B, Arlaud GJ, Liu XY, et al. Trimeric reassembly of the globular domain of human C1q. Biochim Biophys Acta. 2008 Mar; 1784(3):518-29. PubMed PMID: 18179779.

[12] Tuncbag N, Gursoy A, Nussinov R, Keskin O. Predicting protein-protein interactions on a proteome scale by matching evolutionary and structural similarities at interfaces using PRISM. Nature Protocols. 2011;6(9):1341-54.

[13] Tuncbag N, Kar G, Keskin O, Gursoy A, Nussinov R. A survey of available tools and web servers for analysis of protein–protein interactions and interfaces. Briefings in Bioinformatics. 2009;10(3):217.

[14] Guney E, Tuncbag N, Keskin O, Gursoy A. HotSprint: database of computational hot spots in protein interfaces. Nucleic acids research. 2008;36(suppl 1):D662-D6.

[15] Shoemaker BA, Panchenko AR. Deciphering protein-protein interactions. Part II. Computational methods to predict protein and domain interaction partners. PLoS Comput Biol. 2007 Apr 27;3(4):e43. PubMed PMID: 17465672. Pubmed Central PMCID: 1857810.

[16] Wass MN, David A, Sternberg MJ. Challenges for the prediction of macromolecular interactions. Curr Opin Struct Biol. 2011 Jun;21(3):382-90. PubMed PMID: 21497504.

[17] Juan D, Pazos F, Valencia A. High-confidence prediction of global interactomes based on genome-wide coevolutionary networks. Proc Natl Acad Sci U S A. 2008 Jan 22;105(3):934-9. PubMed PMID: 18199838. Pubmed Central PMCID: 2242690.

[18] Mosca R, Ceol A, Stein A, Olivella R, Aloy P. 3did: a catalog of domain-based interactions of known three-dimensional structure. Nucleic Acids Res. 2014 Jan 1;42(1):D374-9. PubMed PMID: 24081580.

[19] Mosca R, Pons T, Ceol A, Valencia A, Aloy P. Towards a detailed atlas of protein-protein interactions. Curr Opin Struct Biol. 2013 Dec;23(6):929-40. PubMed PMID: 23896349.

[20] Mosca R, Ceol A, Aloy P. Interactome3D: adding structural details to protein networks. Nature methods. 2013 Jan;10(1):47-53. PubMed PMID: 23399932.

[21] Janin J, Bahadur RP, Chakrabarti P. Protein-protein interaction and quaternary structure. Q Rev Biophys. 2008 May;41(2):133-80. PubMed PMID: 18812015.

[22] Valdar WS, Thornton JM. Conservation helps to identify biologically relevant crystal contacts. J Mol Biol. 2001 Oct 19;313(2):399-416. PubMed PMID: 11800565.

[23] Ponstingl H, Henrick K, Thornton JM. Discriminating between homodimeric and monomeric proteins in the crystalline state. Proteins. 2000 Oct 1;41(1):47-57. PubMed PMID: 10944393.

[24] Elcock AH, McCammon JA. Identification of protein oligomerization states by analysis of interface conservation. Proc Natl Acad Sci U S A. 2001 Mar 13;98(6):2990-4. PubMed PMID: 11248019. Pubmed Central PMCID: 30594.

[25] Lai YT, King NP, Yeates TO. Principles for designing ordered protein assemblies. Trends Cell Biol. 2012 Dec;22(12):653-61. PubMed PMID: 22975357.

[26] Woolfson DN, Bartlett GJ, Bruning M, Thomson AR. New currency for old rope: from coiled-coil assemblies to alpha-helical barrels. Curr Opin Struct Biol. 2012 Aug; 22(4):432-41. PubMed PMID: 22445228.

[27] Channon K, Bromley EH, Woolfson DN. Synthetic biology through biomolecular design and engineering. Curr Opin Struct Biol. 2008 Aug;18(4):491-8. PubMed PMID: 18644449.

[28] Ringler P, Schulz GE. Self-assembly of proteins into designed networks. Science. 2003 Oct 3;302(5642):106-9. PubMed PMID: 14526081.

[29] King NP, Lai Y-T. Practical approaches to designing novel protein assemblies. Current opinion in structural biology. 2013.

[30] Gursoy A, Keskin O, Nussinov R. Topological properties of protein interaction networks from a structural perspective. Biochemical Society Transactions. 2008;36:1398-403.

[31] Tuncbag N, Gursoy A, Guney E, Nussinov R, Keskin O. Architectures and functional coverage of protein-protein interfaces. J Mol Biol. 2008 Sep 5;381(3):785-802. PubMed PMID: 18620705. Pubmed Central PMCID: 2605427.

[32] Shulman-Peleg A, Shatsky M, Nussinov R, Wolfson HJ. Spatial chemical conservation of hot spot interactions in protein-protein complexes. BMC Biol. 2007;5:43. PubMed PMID: 17925020.

[33] Ma B, Nussinov R. Trp/Met/Phe hot spots in protein-protein interactions: potential targets in drug design. Current topics in medicinal chemistry. 2007;7(10):999-1005.

[34] Jones S, Thornton JM. Principles of protein-protein interactions. Proc Natl Acad Sci U S A. 1996 Jan 9;93(1):13-20. PubMed PMID: 8552589.

[35] Laskowski RA, Thornton JM. Understanding the molecular machinery of genetics through 3D structures. Nature Reviews Genetics. 2008;9(2):141-51.

[36] Toogood PL. Inhibition of protein-protein association by small molecules: approaches and progress. Journal of medicinal chemistry. 2002 Apr 11;45(8):1543-58. PubMed PMID: 11931608.

[37] Gadek TR, Nicholas JB. Small molecule antagonists of proteins. Biochemical pharmacology. 2003 Jan 1;65(1):1-8. PubMed PMID: 12473372.

[38] Lo Conte L, Chothia C, Janin J. The atomic structure of protein-protein recognition sites. J Mol Biol. 1999 Feb 5;285(5):2177-98. PubMed PMID: 9925793.

[39] Chakrabarti P, Janin J. Dissecting protein-protein recognition sites. Proteins. 2002 May 15;47(3):334-43. PubMed PMID: 11948787.

[40] Ofran Y, Rost B. Analysing six types of proteinâ€“protein interfaces. Journal of molecular biology. 2003;325(2):377-87.

[41] Bashton M, Chothia C. The geometry of domain combination in proteins. J Mol Biol. 2002 Jan 25;315(4):927-39. PubMed PMID: 11812158.

[42] Xu D, Tsai CJ, Nussinov R. Hydrogen bonds and salt bridges across protein-protein interfaces. Protein Eng. 1997 Sep;10(9):999-1012. PubMed PMID: 9464564.

[43] Selbig J, Argos P. Relationships between protein sequence and structure patterns based on residue contacts. Proteins. 1998 May 1;31(2):172-85. PubMed PMID: 9593191.

[44] Shoemaker BA, Panchenko AR. Deciphering protein-protein interactions. Part I. Experimental techniques and databases. PLoS Comput Biol. 2007 Mar 30;3(3):e42. PubMed PMID: 17397251. Pubmed Central PMCID: 1847991.

[45] de Vries SJ, Bonvin AM. How proteins get in touch: interface prediction in the study of biomolecular complexes. Curr Protein Pept Sci. 2008 Aug;9(4):394-406. PubMed PMID: 18691126.

[46] Guex N, Peitsch MC. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis. 1997 Dec;18(15):2714-23. PubMed PMID: 9504803.

[47] Shulman-Peleg A, Shatsky M, Nussinov R, Wolfson HJ. MultiBind and MAPPIS: webservers for multiple alignment of protein 3D-binding sites and their interactions. Nucleic Acids Res. 2008 Jul 1;36(Web Server issue):W260-4. PubMed PMID: 18467424.

[48] Orchard S, Ammari M, Aranda B, Breuza L, Briganti L, Broackes-Carter F, et al. The MIntAct project--IntAct as a common curation platform for 11 molecular interaction databases. Nucleic Acids Res. 2014 Jan 1;42(1):D358-63. PubMed PMID: 24234451.

[49] Chothia C, Janin J. Principles of protein-protein recognition. Nature. 1975 Aug 28;256(5520):705-8. PubMed PMID: 1153006.

[50] Henrick K, Thornton JM. PQS: a protein quaternary structure file server. Trends Biochem Sci. 1998 Sep;23(9):358-61. PubMed PMID: 9787643.

[51] Janin J, Rodier F. Protein-protein interaction at crystal contacts. Proteins. 1995 Dec; 23(4):580-7. PubMed PMID: 8749854.

[52] Lee B, Richards FM. The interpretation of protein structures: estimation of static accessibility. J Mol Biol. 1971 Feb 14;55(3):379-400. PubMed PMID: 5551392.

[53] Poupon A. Voronoi and Voronoi-related tessellations in studies of protein structure and interaction. Curr Opin Struct Biol. 2004 Apr;14(2):233-41. PubMed PMID: 15093839.

[54] Cazals F, Proust F, Bahadur RP, Janin J. Revisiting the Voronoi description of protein-protein interfaces. Protein Sci. 2006 Sep;15(9):2082-92. PubMed PMID: 16943442.

[55] Bouvier B, Grunberg R, Nilges M, Cazals F. Shelling the Voronoi interface of protein-protein complexes reveals patterns of residue conservation, dynamics, and composition. Proteins. 2009 Aug 15;76(3):677-92. PubMed PMID: 19280599.

[56] Dreyfus T, Doye V, Cazals F. Probing a continuum of macro-molecular assembly models with graph templates of complexes. Proteins. 2013 Nov;81(11):2034-44. PubMed PMID: 23609891.

[57] Faure G, Bornot A, de Brevern AG. Protein contacts, inter-residue interactions and side-chain modelling. Biochimie. 2008;90:626-39.

[58] Ofran Y, Rost B. Protein–protein interaction hotspots carved into sequences. PLoS computational biology. 2007;3(7):e119.

[59] Feverati G, Achoch M, Vuillon L, Lesieur C. Intermolecular β-Strand Networks Avoid Hub Residues and Favor Low Interconnectedness: A Potential Protection Mechanism against Chain Dissociation upon Mutation. PloS one. 2014;9(4):e94745.

[60] Feverati G, Lesieur C. Oligomeric interfaces under the lens: gemini. PloS one. 2010;5(3):e9897.

[61] Caffrey DR, Somaroo S, Hughes JD, Mintseris J, Huang ES. Are protein-protein interfaces more conserved in sequence than the rest of the protein surface? Protein Sci. 2004 Jan;13(1):190-202. PubMed PMID: 14691234.

[62] Lichtarge O, Bourne HR, Cohen FE. An evolutionary trace method defines binding surfaces common to protein families. J Mol Biol. 1996 Mar 29;257(2):342-58. PubMed PMID: 8609628.

[63] Armon A, Graur D, Ben-Tal N. ConSurf: an algorithmic tool for the identification of functional regions in proteins by surface mapping of phylogenetic information. J Mol Biol. 2001 Mar 16;307(1):447-63. PubMed PMID: 11243830.

[64] Pupko T, Bell RE, Mayrose I, Glaser F, Ben-Tal N. Rate4Site: an algorithmic tool for the identification of functional regions in proteins by surface mapping of evolutionary determinants within their homologues. Bioinformatics. 2002;18 Suppl 1:S71-7. PubMed PMID: 12169533.

[65] Landgraf R, Xenarios I, Eisenberg D. Three-dimensional cluster analysis identifies interfaces and functional residue clusters in proteins. J Mol Biol. 2001 Apr 13;307(5): 1487-502. PubMed PMID: 11292355.

[66] Clackson T, Wells JA. A hot spot of binding energy in a hormone-receptor interface. Science. 1995;267(5196):383-6.

[67] Bogan AA, Thorn KS. Anatomy of hot spots in protein interfaces. J Mol Biol. 1998 Jul 3;280(1):1-9. PubMed PMID: 9653027.

[68] Fischer TB, Arunachalam KV, Bailey D, Mangual V, Bakhru S, Russo R, et al. The binding interface database (BID): a compilation of amino acid hot spots in protein interfaces. Bioinformatics. 2003 Jul 22;19(11):1453-4. PubMed PMID: 12874065.

[69] Guidry JJ, Shewmaker F, Maskos K, Landry S, Wittung-Stafshede P. Probing the interface in a human co-chaperonin heptamer: residues disrupting oligomeric unfolded state identified. BMC Biochem. 2003 Oct 2;4:14. PubMed PMID: 14525625.

[70] Luke K, Perham M, Wittung-Stafshede P. Kinetic folding and assembly mechanisms differ for two homologous heptamers. J Mol Biol. 2006 Oct 27;363(3):729-42. PubMed PMID: 16979655.

[71] Barabasi AL, Oltvai ZN. Network biology: understanding the cell's functional organization. Nature reviews Genetics. 2004 Feb;5(2):101-13. PubMed PMID: 14735121.

[72] Bhattacharyya M, Vishveshwara S. Probing the allosteric mechanism in pyrrolysyl-tRNA synthetase using energy-weighted network formalism. Biochemistry. 2011 Jul 19;50(28):6225-36. PubMed PMID: 21650159.

[73] De Ruvo M, Giuliani A, Paci P, Santoni D, Di Paola L. Shedding light on protein-ligand binding by graph theory: the topological nature of allostery. Biophys Chem. 2012 May;165-166:21-9. PubMed PMID: 22464849.

[74] Daily MD, Gray JJ. Allosteric communication occurs via networks of tertiary and quaternary motions in proteins. PLoS Comput Biol. 2009 Feb;5(2):e1000293. PubMed PMID: 19229311. Pubmed Central PMCID: 2634971.

[75] Tsai CJ, Del Sol A, Nussinov R. Protein allostery, signal transmission and dynamics: a classification scheme of allosteric mechanisms. Molecular bioSystems. 2009 Mar; 5(3):207-16. PubMed PMID: 19225609. Pubmed Central PMCID: 2898650.

[76] Gunasekaran K, Ma B, Nussinov R. Is allostery an intrinsic property of all dynamic proteins? Proteins: Structure, Function, and Bioinformatics. 2004;57(3):433-43.

[77] del Sol A, Tsai CJ, Ma B, Nussinov R. The origin of allosteric functional modulation: multiple pre-existing pathways. Structure. 2009 Aug 12;17(8):1042-50. PubMed PMID: 19679084. Pubmed Central PMCID: 2749652.

[78] Amitai G, Shemesh A, Sitbon E, Shklar M, Netanely D, Venger I, et al. Network analysis of protein structures identifies functional residues. J Mol Biol. 2004 Dec 3;344(4): 1135-46. PubMed PMID: 15544817.

[79] Vendruscolo M, Dokholyan NV, Paci E, Karplus M. Small-world view of the amino acids that play a key role in protein folding. Phys Rev E Stat Nonlin Soft Matter Phys. 2002 Jun;65(6 Pt 1):061910. PubMed PMID: 12188762.

[80] Dokholyan NV, Li L, Ding F, Shakhnovich EI. Topological determinants of protein folding. Proc Natl Acad Sci U S A. 2002 Jun 25;99(13):8637-41. PubMed PMID: 12084924. Pubmed Central PMCID: 124342.

[81] del Sol A, O'Meara P. Small-world network approach to identify key residues in protein-protein interaction. Proteins. 2005 Feb 15;58(3):672-82. PubMed PMID: 15617065.

[82] Brinda KV, Kannan N, Vishveshwara S. Analysis of homodimeric protein interfaces by graph-spectral methods. Protein Eng. 2002 Apr;15(4):265-77. PubMed PMID: 11983927.

[83] Albert R, Jeong H, Barabasi AL. Error and attack tolerance of complex networks. Nature. 2000 Jul 27;406(6794):378-82. PubMed PMID: 10935628.

[84] Greene LH, Higman VA. Uncovering network systems within protein structures. J Mol Biol. 2003 Dec 5;334(4):781-91. PubMed PMID: 14636602.

[85] Brinda KV, Vishveshwara S. Oligomeric protein structure networks: insights into protein-protein interactions. BMC Bioinformatics. 2005;6:296. PubMed PMID: 16336694. Pubmed Central PMCID: 1326230.

[86] Dey S, Pal A, Chakrabarti P, Janin J. The subunit interfaces of weakly associated homodimeric proteins. J Mol Biol. 2010 Apr 23;398(1):146-60. PubMed PMID: 20156457.

[87] Talavera D, Robertson DL, Lovell SC. Characterization of protein-protein interaction interfaces from a single species. PloS one. 2011;6(6):e21053.

[88] Kim WK, Henschel A, Winter C, Schroeder M. The many faces of protein-protein interactions: A compendium of interface geometry. PLoS Comput Biol. 2006 Sep 29;2(9):e124. PubMed PMID: 17009862.

[89] Winter C, Henschel A, Kim WK, Schroeder M. SCOPPI: a structural classification of protein-protein interfaces. Nucleic Acids Res. 2006 Jan 1;34(Database issue):D310-4. PubMed PMID: 16381874.

[90] Keskin O, Tsai CJ, Wolfson H, Nussinov R. A new, structurally nonredundant, diverse data set of proteinâ€“protein interfaces and its implications. Protein Science. 2004;13(4):1043-55.

[91] Neuvirth H, Heinemann U, Birnbaum D, Tishby N, Schreiber G. ProMateus--an open research approach to protein-binding sites analysis. Nucleic Acids Res. 2007 Jul;

35(Web Server issue):W543-8. PubMed PMID: 17488838. Pubmed Central PMCID: 1933218.

[92] Yan C, Wu F, Jernigan RL, Dobbs D, Honavar V. Characterization of protein-protein interfaces. Protein J. 2008 Jan;27(1):59-70. PubMed PMID: 17851740.

[93] Yan C, Dobbs D, Honavar V. A two-stage classifier for identification of protein-protein interface residues. Bioinformatics. 2004 Aug 4;20 Suppl 1:i371-8. PubMed PMID: 15262822.

[94] Cheng P-N, Pham JD, Nowick JS. The Supramolecular Chemistry of β-Sheets. Journal of the American Chemical Society. 2013.

[95] Khakshoor O, Nowick JS. Artificial beta-sheets: chemical models of beta-sheets. Curr Opin Chem Biol. 2008 Dec;12(6):722-9. PubMed PMID: 18775794.

[96] López De La Paz M, Serrano L. Sequence determinants of amyloid fibril formation. Proceedings of the National Academy of Sciences of the United States of America. 2004;101(1):87.

[97] Lopez De La Paz M, Goldie K, Zurdo J, Lacroix E, Dobson CM, Hoenger A, et al. De novo designed peptide-based amyloid fibrils. Proc Natl Acad Sci U S A. 2002 Dec 10;99(25):16052-7. PubMed PMID: 12456886.

[98] Trovato A, Chiti F, Maritan A, Seno F. Insight into the structure of amyloid fibrils from the analysis of globular proteins. PLoS computational biology. 2006;2(12):e170.

[99] Kayed R, Head E, Thompson JL, McIntire TM, Milton SC, Cotman CW, et al. Common structure of soluble amyloid oligomers implies common mechanism of pathogenesis. Science. 2003;300(5618):486-9.

[100] Guijarro JI, Sunde M, Jones JA, Campbell ID, Dobson CM. Amyloid fibril formation by an SH3 domain. Proc Natl Acad Sci U S A. 1998 Apr 14;95(8):4224-8. PubMed PMID: 9539718. Pubmed Central PMCID: 22470.

[101] Dobson CM. Protein misfolding, evolution and disease. Trends Biochem Sci. 1999;24:329-32.

[102] Petkova AT, Ishii Y, Balbach JJ, Antzutkin ON, Leapman RD, Delaglio F, et al. A structural model for Alzheimer's beta-amyloid fibrils based on experimental constraints from solid state NMR. Proc Natl Acad Sci U S A. 2002 Dec 24;99(26):16742-7. PubMed PMID: 12481027. Pubmed Central PMCID: 139214.

[103] Der-Sarkissian A, Jao CC, Chen J, Langen R. Structural organization of alpha-synuclein fibrils studied by site-directed spin labeling. J Biol Chem. 2003 Sep 26;278(39): 37530-5. PubMed PMID: 12815044.

[104] Kajava AV, Aebi U, Steven AC. The parallel superpleated beta-structure as a model for amyloid fibrils of human amylin. J Mol Biol. 2005 Apr 29;348(2):247-52. PubMed PMID: 15811365.

[105] Krishnan R, Lindquist SL. Structural insights into a yeast prion illuminate nucleation and strain diversity. Nature. 2005 Jun 9;435(7043):765-72. PubMed PMID: 15944694. Pubmed Central PMCID: 1405905.

[106] Margittai M, Langen R. Template-assisted filament growth by parallel stacking of tau. Proc Natl Acad Sci U S A. 2004 Jul 13;101(28):10278-83. PubMed PMID: 15240881. Pubmed Central PMCID: 478563.

[107] Lv G, Kumar A, Giller K, Orcellet ML, Riedel D, Fernandez CO, et al. Structural comparison of mouse and human alpha-synuclein amyloid fibrils by solid-state NMR. J Mol Biol. 2012 Jun 29;420(1-2):99-111. PubMed PMID: 22516611.

[108] Fernandez-Escamilla A-M, Rousseau F, Schymkowitz J, Serrano L. Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. Nature Biotechnology. 2004;22(10):1302-6.

[109] Galzitskaya OV, Garbuzynskiy SO, Lobanov MY. Prediction of amyloidogenic and disordered regions in protein chains. PLoS computational biology. 2006;2(12):e177.

[110] Thompson MJ, Sievers SA, Karanicolas J, Ivanova MI, Baker D, Eisenberg D. The 3D profile method for identifying fibril-forming segments of proteins. Proceedings of the National Academy of Sciences of the United States of America. 2006;103(11):4074-8.

[111] Belli M, Ramazzotti M, Chiti F. Prediction of amyloid aggregation in vivo. EMBO Rep. 2011 Jul;12(7):657-63. PubMed PMID: 21681200. Pubmed Central PMCID: 3128957.

[112] Smith JM, Jang Y, Kim MK. Steiner minimal trees, twist angles, and the protein folding problem. Proteins: Structure, Function, and Bioinformatics. 2007;66(4):889-902.

[113] Levin KB, Dym O, Albeck S, Magdassi S, Keeble AH, Kleanthous C, et al. Following evolutionary paths to protein-protein interactions with high affinity and selectivity. Nature structural & molecular biology. 2009;16(10):1049-55.

[114] Caflisch A. Network and graph analyses of folding free energy surfaces. Curr Opin Struct Biol. 2006 Feb;16(1):71-8. PubMed PMID: 16413772.

[115] Bahar I, Chennubhotla C, Tobi D. Intrinsic dynamics of enzymes in the unbound state and relation to allosteric regulation. Curr Opin Struct Biol. 2007 Dec;17(6): 633-40. PubMed PMID: 18024008. Pubmed Central PMCID: 2197162.

[116] Chennubhotla C, Bahar I. Signal propagation in proteins and relation to equilibrium fluctuations. PLoS Comput Biol. 2007 Sep;3(9):1716-26. PubMed PMID: 17892319. Pubmed Central PMCID: 1988854.

[117] Jin Y, Turaev D, Weinmaier T, Rattei T, Makse HA. The evolutionary dynamics of protein-protein interaction networks inferred from the reconstruction of ancient networks. PLoS One. 2013;8(3):e58134. PubMed PMID: 23526967. Pubmed Central PMCID: 3603955.

[118] Liu YY, Slotine JJ, Barabasi AL. Controllability of complex networks. Nature. 2011;473:167-73.

[119] Albert R, Barabasi AL. Topology of evolving networks: local events and universality. Physical review letters. 2000 Dec 11;85(24):5234-7. PubMed PMID: 11102229.

[120] Callaway DS, Newman ME, Strogatz SH, Watts DJ. Network robustness and fragility: percolation on random graphs. Physical review letters. 2000 Dec 18;85(25):5468-71. PubMed PMID: 11136023.

[121] Lupas A, Van Dyke M, Stock J. Predicting coiled coils from protein sequences. Science. 1991 May 24;252(5010):1162-4. PubMed PMID: 2031185.

[122] Gruber M, Soding J, Lupas AN. Comparative analysis of coiled-coil prediction methods. J Struct Biol. 2006 Aug;155(2):140-5. PubMed PMID: 16870472.

[123] Bartoli L, Fariselli P, Krogh A, Casadio R. CCHMM_PROF: a HMM-based coiled-coil predictor with evolutionary information. Bioinformatics. 2009 Nov 1;25(21):2757-63. PubMed PMID: 19744995.

[124] Wolf E, Kim PS, Berger B. MultiCoil: a program for predicting two-and three-stranded coiled coils. Protein Sci. 1997 Jun;6(6):1179-89. PubMed PMID: 9194178. Pubmed Central PMCID: 2143730.

[125] Crick FHC. The packing of alpha-helices: simple coiled-coils. Acta Crystallogr. 1953;6:689–97.

[126] Poupon A, Janin J. Analysis and prediction of protein quaternary structure. Methods Mol Biol. 2010;609:349-64. PubMed PMID: 20221929.

[127] Comeau SR, Camacho CJ. Predicting oligomeric assemblies: N-mers a primer. J Struct Biol. 2005 Jun;150(3):233-44. PubMed PMID: 15890272.

[128] Walshaw J, Woolfson DN. Extended knobs-into-holes packing in classical and complex coiled-coil assemblies. J Struct Biol. 2003 Dec;144(3):349-61. PubMed PMID: 14643203.

[129] Calladine CR, Luisi BF, Pratap JV. A "mechanistic" explanation of the multiple helical forms adopted by bacterial flagellar filaments. J Mol Biol. 2013 Mar 11;425(5):914-28. PubMed PMID: 23274110. Pubmed Central PMCID: 3605589.

[130] Calladine CR, Sharff A, Luisi B. How to untwist an alpha-helix: structural principles of an alpha-helical barrel. J Mol Biol. 2001 Jan 19;305(3):603-18. PubMed PMID: 11152616.

[131] Moutevelis E, Woolfson DN. A periodic table of coiled-coil protein structures. J Mol Biol. 2009 Jan 23;385(3):726-32. PubMed PMID: 19059267.

[132] Testa OD, Moutevelis E, Woolfson DN. CC+: a relational database of coiled-coil structures. Nucleic Acids Res. 2009 Jan;37(Database issue):D315-22. PubMed PMID: 18842638.

[133] Papanikolopoulou K, Forge V, Goeltz P, Mitraki A. Formation of highly stable chimeric trimers by fusion of an adenovirus fiber shaft fragment with the foldon domain of bacteriophage t4 fibritin. J Biol Chem. 2004b Mar 5;279(10):8991-8. PubMed PMID: 14699113.

[134] Zrimi J, Ng Ling A, Giri-Rachman Arifin E, Feverati G, Lesieur C. Cholera toxin B subunits assemble into pentamers-proposition of a fly-casting mechanism. PLoS One. 2010;5(12):e15347. PubMed PMID: 21203571.

[135] Ruddock LW, Coen JJ, Cheesman C, Freedman RB, Hirst TR. Assembly of the B subunit pentamer of Escherichia coli heat-labile enterotoxin. Kinetics and molecular basis of rate-limiting steps in vitro. J Biol Chem. 1996b Aug 9;271(32):19118-23. PubMed PMID: 8702586.

[136] Dang LT, Purvis AR, Huang RH, Westfield LA, Sadler JE. Phylogenetic and functional analysis of histidine residues essential for pH-dependent multimerization of von Willebrand factor. Journal of Biological Chemistry. 2011.

[137] Hashimoto K, Nishi H, Bryant S, Panchenko AR. Caught in self-interaction: evolutionary and functional mechanisms of protein homooligomerization. Phys Biol. 2011 Jun;8(3):035007. PubMed PMID: 21572178. Pubmed Central PMCID: 3148176.

[138] Csermely P, Palotai R, Nussinov R. Induced fit, conformational selection and independent dynamic segments: an extended view of binding events. Trends Biochem Sci. 2010 Oct;35(10):539-46. PubMed PMID: 20541943. Pubmed Central PMCID: 3018770.

[139] D'Alessio G. The evolutionary transition from monomeric to oligomeric proteins: tools, the environment, hypotheses. Prog Biophys Mol Biol. 1999;72(3):271-98. PubMed PMID: 10581971.

[140] D'Alessio G. Oligomer evolution in action? Nat Struct Biol. 1995 Jan;2(1):11-3. PubMed PMID: 7719846.

[141] Eisenberg D, Jucker M. The Amyloid State of Proteins in Human Diseases. Cell. 2012;148(6):1188-203.

[142] Levy ED, Erba EB, Robinson CV, Teichmann SA. Assembly reflects evolution of protein complexes. Nature. 2008;453(7199):1262-5.

[143] Luo M, Singh RK, Tanner JJ. Structural determinants of oligomerization of delta(1)-pyrroline-5-carboxylate dehydrogenase: identification of a hexamerization hot spot. J Mol Biol. 2013 Sep 9;425(17):3106-20. PubMed PMID: 23747974. Pubmed Central PMCID: 3743950.

[144] Crick FH, Watson JD. Structure of small viruses. Nature. 1956 Mar 10;177(4506): 473-5. PubMed PMID: 13309339.

[145] Rackham OJ, Madera M, Armstrong CT, Vincent TL, Woolfson DN, Gough J. The evolution and structure prediction of coiled coils across all genomes. J Mol Biol. 2010 Oct 29;403(3):480-93. PubMed PMID: 20813113.

[146] Yadid I, Kirshenbaum N, Sharon M, Dym O, Tawfik DS. Metamorphic proteins mediate evolutionary transitions of structure. Proc Natl Acad Sci U S A. 2010 Apr 20;107(16):7287-92. PubMed PMID: 20368465. Pubmed Central PMCID: 2867682.

[147] King J, Wood WB. Assembly of bacteriophage T4 tail fibers: the sequence of gene product interaction. J Mol Biol. 1969 Feb 14;39(3):583-601. PubMed PMID: 5390559.

[148] King J. Assembly of the tail of bacteriophage T4. J Mol Biol. 1968 Mar 14;32(2):231-62. PubMed PMID: 4868421.

[149] Rennell D, Bouvier SE, Hardy LW, Poteete AR. Systematic mutation of bacteriophage T4 lysozyme. J Mol Biol. 1991 Nov 5;222(1):67-88. PubMed PMID: 1942069.

[150] Goldenberg DP, Berget PB, King J. Maturation of the tail spike endorhamnosidase of Salmonella phage P22. J Biol Chem. 1982 Jul 10;257(13):7864-71. PubMed PMID: 7045114.

[151] Perham M, Chen M, Ma J, Wittung-Stafshede P. Unfolding of heptameric co-chaperonin protein follows "fly casting" mechanism: observation of transient nonnative heptamer. J Am Chem Soc. 2005 Nov 30;127(47):16402-3. PubMed PMID: 16305220.

[152] Bascos N, Guidry J, Wittung-Stafshede P. Monomer topology defines folding speed of heptamer. Protein Sci. 2004 May;13(5):1317-21. PubMed PMID: 15075408.

[153] Tacnet P, Thielens N, Arifin Giri Rachman E, Hirst TR, Lesieur C. Cholera toxin B assembly intermediates provide some explanation to the existence of a pentameric toxin state.

[154] Pell LG, Cumby N, Clark TE, Tuite A, Battaile KP, Edwards AM, et al. A conserved spiral structure for highly diverged phage tail assembly chaperones. J Mol Biol. 2013 Jul 24;425(14):2436-49. PubMed PMID: 23542344.

[155] Aghera N, Udgaonkar JB. Kinetic studies of the folding of heterodimeric monellin: evidence for switching between alternative parallel pathways. J Mol Biol. 2012 Jul 13;420(3):235-50. PubMed PMID: 22542529.

[156] Kentsis A, Gordon RE, Borden KL. Control of biochemical reactions through supramolecular RING domain self-assembly. Proc Natl Acad Sci U S A. 2002 Nov 26;99(24):15404-9. PubMed PMID: 12438698. Pubmed Central PMCID: 137729.

[157] Kentsis A, Gordon RE, Borden KL. Self-assembly properties of a model RING domain. Proc Natl Acad Sci U S A. 2002 Jan 22;99(2):667-72. PubMed PMID: 11792829. Pubmed Central PMCID: 117363.

[158] Kentsis A, Borden KL. Construction of macromolecular assemblages in eukaryotic processes and their role in human disease: linking RINGs together. Curr Protein Pept Sci. 2000 Jul;1(1):49-73. PubMed PMID: 12369920.

[159] Pereira-Leal JB, Levy ED, Teichmann SA. The origins and evolution of functional modules: lessons from protein complexes. Philos Trans R Soc Lond B Biol Sci. 2006 Mar 29;361(1467):507-17. PubMed PMID: 16524839. Pubmed Central PMCID: 1609335.

[160] Shoemaker BA, Portman JJ, Wolynes PG. Speeding molecular recognition by using the folding funnel: the fly-casting mechanism. Proc Natl Acad Sci U S A. 2000 Aug 1;97(16):8868-73. PubMed PMID: 10908673.

[161] Levy Y, Wolynes PG, Onuchic JN. Protein topology determines binding mechanism. Proc Natl Acad Sci U S A. 2004 Jan 13;101(2):511-6. PubMed PMID: 14694192.

[162] Spaar A, Dammer C, Gabdoulline RR, Wade RC, Helms V. Diffusional encounter of barnase and barstar. Biophys J. 2006 Mar 15;90(6):1913-24. PubMed PMID: 16361332.

[163] Ehrlich LP, Nilges M, Wade RC. The impact of protein flexibility on protein-protein docking. Proteins. 2005 Jan 1;58(1):126-33. PubMed PMID: 15515181.

[164] Gabdoulline RR, Wade RC. Protein-protein association: investigation of factors influencing association rates by brownian dynamics simulations. J Mol Biol. 2001 Mar 9;306(5):1139-55. PubMed PMID: 11237623.

[165] Gabdoulline RR, Wade RC. Simulation of the diffusional association of barnase and barstar. Biophys J. 1997 May;72(5):1917-29. PubMed PMID: 9129797.

[166] Noe F, Fischer S. Transition networks for modeling the kinetics of conformational change in macromolecules. Curr Opin Struct Biol. 2008 Apr;18(2):154-62. PubMed PMID: 18378442.

[167] Wales DJ. Energy landscapes: some new horizons. Curr Opin Struct Biol. 2010 Feb; 20(1):3-10. PubMed PMID: 20096562.

[168] Barz B, Wales DJ, Strodel B. A kinetic approach to the sequence-aggregation relationship in disease-related protein assembly. J Phys Chem B. 2014 Jan 30;118(4):1003-11. PubMed PMID: 24401100. Pubmed Central PMCID: 3908877.

[169] Higashimoto Y, Asanomi Y, Takakusagi S, Lewis MS, Uosaki K, Durell SR, et al. Unfolding, aggregation, and amyloid formation by the tetramerization domain from mutant p53 associated with lung cancer. Biochemistry. 2006;45(6):1608-19.

[170] Fujita T, Kiyama M, Tomizawa Y, Kohno T, Yokota J. Comprehensive analysis of p53 gene mutation characteristics in lung carcinoma with special reference to histological subtypes. International journal of oncology. 1999;15(5):927-34.

[171] Mateu MG, Fersht AR. Nine hydrophobic side chains are key determinants of the thermodynamic stability and oligomerization status of tumour suppressor p53 tetramerization domain. The EMBO Journal. 1998;17(10):2748-58.

[172] Reixach N, Foss TR, Santelli E, Pascual J, Kelly JW, Buxbaum JN. Human-murine transthyretin heterotetramers are kinetically stable and non-amyloidogenic. A lesson in the generation of transgenic models of diseases involving oligomeric proteins. J Biol Chem. 2008 Jan 25;283(4):2098-107. PubMed PMID: 18006495.

[173] Jiang X, Buxbaum JN, Kelly JW. The V122I cardiomyopathy variant of transthyretin increases the velocity of rate-limiting tetramer dissociation, resulting in accelerated amyloidosis. Proc Natl Acad Sci U S A. 2001 Dec 18;98(26):14943-8. PubMed PMID: 11752443. Pubmed Central PMCID: 64963.

[174] Lomas DA, Carrell RW. Serpinopathies and the conformational dementias. Nature Reviews Genetics. 2002;3(10):759-68.

[175] Chiti F, Stefani M, Taddei N, Ramponi G, Dobson CM. Rationalization of the effects of mutations on peptide and protein aggregation rates. NATURE-LONDON-. 2003:805-8.

[176] Changeux JP, Edelstein SJ. Allosteric mechanisms of signal transduction. Science. 2005 Jun 3;308(5727):1424-8. PubMed PMID: 15933191.

[177] Pires DE, Ascher DB, Blundell TL. mCSM: predicting the effects of mutations in proteins using graph-based signatures. Bioinformatics. 2014 Feb 1;30(3):335-42. PubMed PMID: 24281696. Pubmed Central PMCID: 3904523.

[178] Bode C, Kovacs IA, Szalay MS, Palotai R, Korcsmaros T, Csermely P. Network analysis of protein dynamics. FEBS Lett. 2007 Jun 19;581(15):2776-82. PubMed PMID: 17531981.

[179] Newman ME, Strogatz SH, Watts DJ. Random graphs with arbitrary degree distributions and their applications. Phys Rev E Stat Nonlin Soft Matter Phys. 2001 Aug;64(2 Pt 2):026118. PubMed PMID: 11497662.

[180] Peysselon F, Ricard-Blum S. Heparin-protein interactions: From affinity and kinetics to biological roles. Application to an interaction network regulating angiogenesis. Matrix Biol. 2013 Nov 16. PubMed PMID: 24246365.