Review in Advance first posted online on June 26, 2013. (Changes may still occur before final publication online and in print.)

Supramolecular Evolution of Protein Organization

Ariel Fernández[1,2]

Instituto Argentino de Matemática "Alberto P. Calderón," CONICET (National Research Council), Buenos Aires 1083, Argentina; email: ariel@prowdsciences.com

[2]Collegium Basilea, Institute of Advanced Study, CH 4053, Basel, Switzerland

Annu. Rev. Genet. 2013. 47:1–17

The *Annual Review of Genetics* is online at genet.annualreviews.org

This article's doi: 10.1146/annurev-genet-111212-133310

Keywords

random drift, protein complex, molecular biophysics, molecular evolution, dehydron, protein wrapping

Abstract

Protein associations, whether transient or long-lasting, determine cellular processes and enable the cooperative and regulated functionalities characteristic of complex organisms. From a broad physical perspective, soluble natural proteins represent a unique kind of solute prone to associate but not to precipitate. Thus, discrete reproducible associations define the protein supramolecular organization. The evolutionary forces that enable and promote this complexity are the subject matter of this review. The central problem addressed involves the paradoxical constructive role of random genetic drift, typically mildly deleterious, in fostering interactome complexity. By introducing biophysical insights in molecular evolution, we are able to identify the adaptive and nonadaptive elements that define the protein association propensity. We emphasize the mechanistic importance of population size and selection inefficiency in creating an evolutionary niche to promote interactome complexity. Finally, we describe the fitness catastrophes that result from the prevailing evolutionary strategy.

INTRODUCTION AND PERSPECTIVES

As has been recognized for more than six decades, proteins are the molecular agents that perform biological functions. These functions achieve exquisite levels of efficacy and regulation through ephemeral or long-lasting protein associations (1, 2, 5, 7, 10, 41). Thus, protein activity may be agonized, switched on or off, or modulated through interactions with other protein molecules. These associations promote cooperative effects (2, 10, 14, 57) that may entail chemical change in the binding partners, may induce conformational change, or may do both (3, 19, 37, 39). The resulting complexes may actually constitute the functionally competent unit (51) or reflect a regulatory modality (55).

As an illustration of this phenomenological plexus, let us first revisit an example of ephemeral complexation that materializes a signal transduction event. Protein kinases are the typical signal transducers in the cell, and their activation translates into an enablement of their capacity to phosphorylate, which often requires their own phosphorylation by another upstream kinase (4, 11, 38, 65, 68). This process entails the formation of a transient complex. Within the complex, the upstream kinase performs a transphosphoesterification reaction, transferring a phosphate group to the substrate at a residue located within a floppy region known as the activation loop. This chemical change induces an activating conformational change in the substrate (downstream) kinase that enables the formation of a binding pocket for the capture of ATP (adenosine triphosphate). The complex dissociates when the activation of the downstream substrate kinase takes place, and the ligand-kinase association that follows activation enables the downstream kinase to phosphorylate its substrate. This takes place by transference of the gamma phosphate from the binding ATP through a transphosphoesterification reaction. In this way, the signal transduction entails the serial formation of transient protein complexes within which chemical and conformational changes (18, 38).

At the other end of the spectrum in terms of complex lifespan, there are proteins like HIV-1 protease that are only functional in the dimeric state (18, 27) and hence form obligatory complexes with far higher stability than the signaling complexes described. Unlike the case of protein associations realizing the signal transduction, the association of HIV-1-protease subunits entails considerably less induced folding and structural adaptation of the monomeric units. The rigidity of the attachment contributes to the stability of the complex because the conformational entropy cost of the association is significantly lower.

Protein activity is often exquisitely controlled through the formation of complexes that enable synergy and allostery, the two interrelated molecular attributes of proteomic complexity (6, 9, 31, 33, 36, 41, 53, 59, 66). Synergy refers to mutual enhancement of performance through coordinated action, whereas allostery implies modulation or enhancement of activity through an association-induced perturbation in a region of the molecule distant from the active site. To materialize this level of complexity, proteins must associate, forming highly organized quaternary structures, which are frequently obligatory rather than transient (42, 57). This supramolecular organization is familiar to most practitioners in the field of biochemistry and is often illustrated by the allosteric harvesting and release of oxygen carried out by tetrameric hemoglobin (14, 57, 62). The tetrameric version enables a dual affinity of the protein for oxygen: high when harvesting in lung alveoli, low when releasing in tissue. This dual affinity is impossible in a monomeric version of hemoglobin unless a conformational change modifying the binding site can be exogenously induced.

The previous examples illustrate the fact that, from a teleological standpoint, protein complexes are relatively easy to rationalize. However, if these complexes evolved from independent autonomous units (21, 30, 32, 36, 44, 56, 63), as seems to be the case, their evolutionary origin becomes problematic (26, 35). Most evolutionary change is deleterious

Signal transduction: chemical transference of a covalently attached phosphate group or other ligand from one biomolecule to another

Kinase: protein enzyme that catalyzes the phosphorylation of a biochemical substrate

Phosphorylation: chemical modification of a substrate molecule by covalent attachment of a phosphate group

Transphosphoesterification: transference of a phosphate group from one chemical substrate to another, entailing cleavage and formation of ester bonds

Adenosine triphosphate (ATP): the free energy–rich molecule that serves as a source of transferable phosphate in the cell

Protease: protein enzyme that catalyzes the cleavage of a protein chain at a specific site

Allostery: cooperative activity of a protein involving participation of additional molecules other than protein and a substrate

Quaternary structure: the spatial arrangement of a multiunit protein complex

rather than beneficial up to the point when an evolutionary niche opens up for natural selection (26). Thus, we arrive at the core question addressed by this review: What sort of evolutionary framework can reconcile the deleterious nature of prevailing evolutionary change with the constructive change needed to foster interactome complexity?

As this review hopes to show, the seemingly opposite attributes of evolutionary change can be reconciled by taking into account that structural defects in soluble proteins arising from mildly deleterious mutations are sticky in the sense that they invite biomolecular associations (29). This molecular property enables detrimental evolutionary change to introduce an opportunity for the natural selection of a higher level of protein organization.

TRACING THE ORIGIN OF SUPRAMOLECULAR ORGANIZATION

Researchers in the field of molecular evolution have examined possible scenarios that could promote proteins to higher levels of supramolecular organization. For example, Teichmann and her group examined the symmetry of homo-oligomers and noted that a higher-order oligomerization state endowed with higher-order point symmetry may evolve from complexes with lower symmetry, including monomers (34, 43–45, 52, 56, 60). Thus, complexes with cyclic symmetry (rotation invariant) may evolve into complexes with dihedral symmetry with twice as many elements in their symmetry group (rotations + reflexion). However, Kuriyan & Eisenberg (41) described scenarios that would facilitate coevolution of units leading to complex formation. Thus, colocalization resulting from spatially concurrent translation of peptide chains is likely to enhance significantly the rate of effective molecular encounters between the two chains, thereby increasing the evolutionary opportunity for complexation. This colocalization may arise in a variety of cellular situations, including (*a*) two translational products from a common gene, (*b*) translational products destined to a common cell microcompartment, (*c*) proximity in the attachment of the transmembrane portion of the individual subunits, and (*d*) gene fusion that generates a common open reading frame in which the two peptide chains are translated as two tethered domains (15, 40, 41, 46).

Unfortunately, neither crowding effects nor colocalization or increased point symmetry per se provides mechanistic underpinnings of the evolutionary origin of interactome complexity. Any random mutation with a chance to get fixed in the species population is far more likely to be deleterious or neutral to the function of the free subunit, and mutations conducive to protein associations have not been physically characterized (12, 26). Thus, it is hard to imagine how evolutionary change, typically detrimental to the structure, may contribute to the generation of a suitable interface that enables protein association, but it undoubtedly must do so.

To address this problem, this review introduces and advocates a different approach rooted in a physical understanding of protein interfaces (1, 5, 28, 29, 60). From a physical perspective, their nature remains elusive because the forces driving protein associations—the same ones that govern intramolecular folding—remain poorly understood (23). In broad terms, one singular physical feature stands out as aqueous cellular environments are examined: There are no phase separations arising from spontaneous precipitation of proteins, no matter how high their crowding or effective concentration. In fact, the only instance when spontaneous unregulated phase separations are known to arise under physiological conditions signals an aberrant process known as amyloidosis, which leads to pathological conditions such as Alzheimer's disease. Phase separations do not occur spontaneously in biology, and when they do, they typically signal dysfunctional states (13, 20, 24, 25).

Thus, natural soluble proteins are special solutes, prone only to associate within a highly organized manner to form discrete complexes with specific point symmetries. As proteins form complexes, they reduce the interfacial tension with water. However, a phase separation

Supramolecular organization: spatial arrangements of protein subunits with well-defined protein-protein interfaces

Interactome: universe of specific protein associations within an organism

would also achieve the same goal and on a massive (mesoscopic) scale. Such a phenomenon does not take place spontaneously, at least not under physiological conditions. This implies that natural proteins are physically special, enabling associations but precluding spontaneous aggregation. From the physicist's perspective, it becomes imperative to understand what attributes of the biological interfacial tension enable proteins to interact in this special way. The physical underpinnings of protein-water (P-W) interfaces have remained somewhat elusive up to this point (26) because thermodynamic concepts useful in describing homogeneous phase separation (63) are inadequate in describing solutes that only associate reproducibly in organized discrete assemblies.

Free energy: thermodynamic potential that dictates the spontaneity or feasibility of a process or transformation

Dehydron: structural defect in a soluble protein, resulting from exposure of a backbone hydrogen bond to disruptive hydration

This situation has changed drastically as nanoscale models for water at the interface with proteins have now shown qualitative conceptual improvement (26). These advances enable the underpinning of the physical nature of hot spots in protein associations (8). Harnessing our physical understanding of protein interfaces, we are now reasonably well positioned to address core problems in molecular evolution: How did interactome complexity arise? (26) What is the nature of the evolutionary forces that ultimately led to the functional complexity of the natural proteins, enabling cooperative, synergistic, and allosteric modalities? These problems could have hardly been addressed a few years ago because a satisfactory physical characterization of protein-protein (P-P) interfaces was missing and the insights provided by structural biophysics were severely underrepresented in evolutionary studies. By focusing on the evolution of protein supramolecular organization, this review advocates the incorporation of molecular biophysical concepts that may be brought to fruition in the field of molecular evolution.

ASSOCIATION PROPENSITIES OF SOLUBLE PROTEINS

A soluble protein is likely to engage in associations if in its free state it spans a poor (unstable) interface with water, or in other words, if it has a high P-W interfacial tension (PWIT) (26). This thermodynamic parameter is typically defined as the free-energy increment ΔG_{if}, which is associated with spanning the P-W interface divided by the total protein surface area. Thus, a positive PWIT contributes to destabilizing the protein in the uncomplexed form, promoting protein associations. These associations may materialize spontaneously if the net free-energy change (ΔG_a) that results from the association is negative.

A detailed analysis of protein interfaces revealed some obvious and some less obvious contributions to the interfacial tension. Thus, a folded natural protein is known to expose polar groups to the solvent, thereby lowering the hydration free energy. Such P-W interactions solubilize the protein molecule as a whole and hence contribute to the decrease of PWIT, whereas the opposite holds true in regard to the exposure of nonpolar groups. However, as subsequently shown, the most significant contribution to the tension is perhaps the least studied: It is provided by mild structural deficiencies in the form of solvent-accessible backbone hydrogen bonds (**Figure 1**), the so-called dehydrons (16–22, 24–31, 58). These singularities represent structural vulnerabilities because the exposure of the protein backbone promotes the hydration of amides and carbonyls, which in turn is known to be a dominant folding-disruptive force. Thus, backbone hydration and the concurrent increase in conformational entropy resulting from chain unfolding compete with hydrophobic collapse to such an extent that the folded state becomes always marginally stable. However, dehydrons are inherently sticky (29): The exogenous protection, or wrapping, of a dehydron that takes place upon protein association contributes to the (*a*) enhancement of its stability by destabilizing the unbound state (consisting of buried unpaired amide and carbonyl) and the (*b*) enhancement of the underlying coulombic interaction by deshielding the net partial charges on the proton-donor (amide) and proton-acceptor (carbonyl) groups.

Computing PWIT is difficult because the classical thermodynamic concept of interfacial

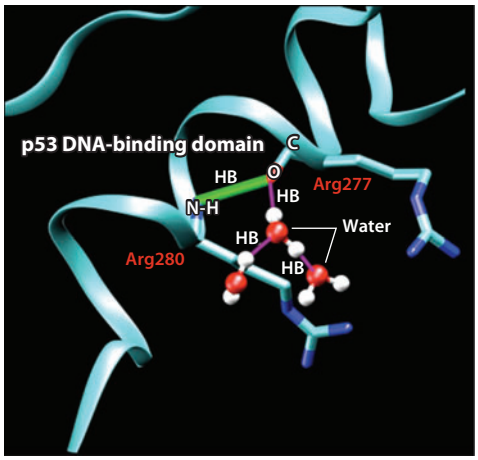

Figure 1

Dehydron as a structural defect in a soluble protein. Ribbon representation of a region of the DNA-binding domain for the tumor suppressor gene p53 (58). The backbone amide-carbonyl (>N-H . . . O = C<) hydrogen bond (HB) pairing residues Arg 277-Arg280 (*side chains shown*) is actually a dehydron, a poorly wrapped bond incompletely shielded from water attack, enabling partial hydration of the polar groups. This local backbone hydration is disruptive of the folded structure, introducing a dynamic instability. The dehydron also generates interfacial tension because the nearby water molecule has a partially hindered coordination to vicinal water molecules when compared with bulk water. Adapted with permission from Reference 58.

tension that is valid for homogeneous phase separations is not valid in this context: Biological processes involve neither homogeneous phases nor phase separations. To compute the PWIT, we need a rigorous nanoscale model of interfacial water that blends seamlessly with the well-known bulk properties as we depart from the nanoscale confinements of the biological interface and approach the macroscopic limit. To describe the structure of water at nanoscales, we introduce the scalar field $[g(\mathbf{r})]$, which associates a real positive number, the expected water coordination g, to each spatial position $\mathbf{r}$ (26). This field represents the equilibration time-averaged number of hydrogen bonds engaging a water molecule with its barycenter at position $\mathbf{r}$. In this way, the field coarsely represents the tension-generating features of water structure. In bulk, $g = 4$, whereas water molecules at the P-W interface may have reduced hydrogen-bonding opportunities ($g < 4$) generating tension (**Figure 2**). Thus, ΔG_{if} is determined by integrating unfavorable local decreases in g and favorable polarization contributions that result from P-W coulombic interactions to yield the free-energy cost of spanning the P-W interface. Within this physical picture, it is easy to visualize that dehydrons are the most significant contributors to the PWIT, as they generate a significant decrease in g value at the interface ($g < 3$) (**Figures 1** and **2**). A concave nonpolar region on the protein surface would have a similar effect, except that it becomes dewetted unless the cavity has a large enough curvature radius, but in that case, the confined water approaches bulk coordination levels. A high PWIT is indicative of a high propensity for P-P associations, which have a net consequence of reducing the protein-water interfacial (PWI) area.

The free-energy increment ΔG_{if} can be regarded as functional of $g(\mathbf{r})$ and is mathematically expressed as (26)

$$\Delta G_{if} = \frac{1}{2}\int \{\alpha|\nabla g|^2 - |\mathbf{P}[g(\mathbf{r})]|^2\}d\mathbf{r}. \quad (1)$$

The first elastic term $\frac{1}{2}\int \alpha|\nabla g|^2 d\mathbf{r}$ (α = scaling factor) accounts for the tension-generating decreases in water coordination ($|\nabla g| > 0$) and integrates all the resulting entropic penalizations, whereas the polarization term $-\frac{1}{2}\int |\mathbf{P}[g(\mathbf{r})]|^2 d\mathbf{r}$ ($\mathbf{P}$ = polarization) integrates all the counterbalances that result from favorable dipole-electrostatic field interactions with the water-exposed polar protein moieties. Notice that only spatial changes in coordination ($|\nabla g| > 0$) contribute to the elastic integral, and hence the entropic contribution to PWIT is solely determined by interfacial water structure. The scaling factor α in the elastic contribution is essential to extrapolate consistently from nanoscale water confinement to bulk water. Thus, bulk-water properties are incorporated because α may be obtained

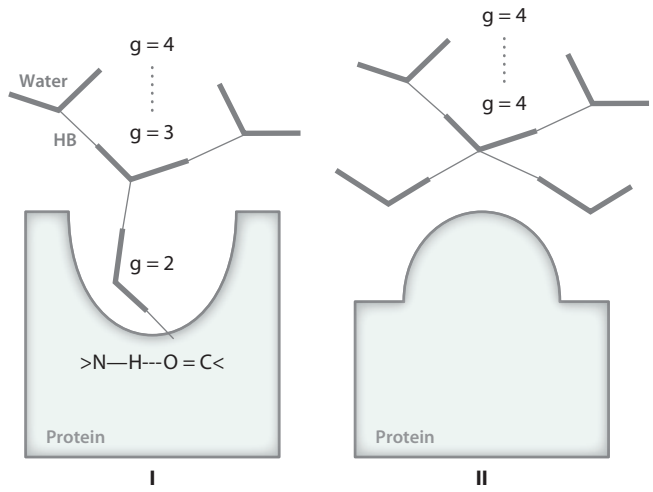

Figure 2

Structural deficiencies in soluble proteins induce association propensities. Hydration of backbone polar groups paired by a dehydron generates interfacial tension by causing water molecules near the defect to reduce their coordination ($g < 4$) relative to the level in a bulk solvent ($g = 4$). Hydrogen bonds (HB) are denoted by thin lines, and water molecules are represented by thick angular segments. In the left panel (*a*), protein structure causes local exposure and unfavorable hydration of the polar backbone, whereas (*b*) the absence of water-coordination hindrance in the well-wrapped protein removes the interfacial tension. In panel b, interfacial water is bulk-like, retaining the maximum coordination of $g = 4$. The protein in panel a has a propensity to associate to reduce interfacial tension, whereas the protein depicted in panel b does not.

from the interfacial tension of a large nonpolar sphere with radius θ in the macroscopic limit θ/1 nm→∞. At a physiological temperature T = 298K, we get $\alpha = 9.0\text{mJ/m} = \lim \gamma(4\pi\theta^2)/\int \frac{1}{2}|\nabla g|^2 d\mathbf{r}$, where $\gamma = 72\text{mJ/m}^2$ is the macroscopic surface tension of water at 298K, and $\int |\nabla g|^2 d\mathbf{r} = \frac{1}{2}4^2\ \text{m}^{-1}(4\pi\theta^2)$ because $\nabla g \neq 0$ holds only close to the interface, and the associated *g*-jump is 4 units in magnitude. Thus, all unfavorable P-W hydrophobic interactions are accounted for through the positive elastic contribution $\frac{1}{2}\int \alpha|\nabla g|^2 d\mathbf{r}$, which, as shown, extrapolates satisfactorily to macroscopic dimensions.

For a given protein structure or template-based structural model, the field $g = g(\mathbf{r})$ used in the numerical integration of Equation 1 is determined by equilibrating the water-embedded structure within an NPT ensemble with fixed parameters of N = number of particles, P = pressure, and T = temperature (47). It has been established that the watertightness of the structure of the soluble protein is the main determinant of the interfacial tension (26). This is expected because dehydrons typically create significant decreases in water coordination, as amides and carbonyls get incompletely hydrated (**Figures 1** and **2**). Thus, structurally deficient proteins with a large proportion of dehydrons have a higher propensity to engage in P-P interactions. To rigorously quantify these assertions, we defined the structural deficiency ν as the quotient (number of dehydrons)/(number of backbone hydrogen bonds). There exists a significant correlation between PWIT and ν, attesting to the importance of structural deficiency as an indicator of association propensity (26). Most importantly, it enables us to adopt the dehydron content as a surrogate for a complicated structural attribute in the form of a thermodynamic parameter to provide a simpler physical rationale for supramolecular organization.

The way structural deficiency promotes association can be best illustrated by examining protein complexes. **Figure 3*a–c*** displays a single insulin unit, itself a heterodimer composed of two covalently independent peptide chains tethered through two intermolecular disulfide bridges. This colocalization favors the coevolution of the chains into the complexed state, as emphasized by Kuriyan & Eisenberg (41). Given their significant levels of structural deficiency, none of the individual chains are able to sustain its structure autonomously in a free state. The longer chain has ν = 11/18, whereas the shorter chain has ν = 6/13 (**Figure 3*c***). This very high structural deficiency in the autonomous three-dimensional fold is the molecular diagnostic of a propensity to aberrantly aggregate in amyloid supramolecular structures, a phenomenon that has in fact been confirmed in the case of insulin. Therefore, these chains can only sustain the structure found in the crystal through cooperative interactions (**Figure 3*b***) or three-body correlations, whereby the two chains reciprocally correct their internal structural deficiencies.

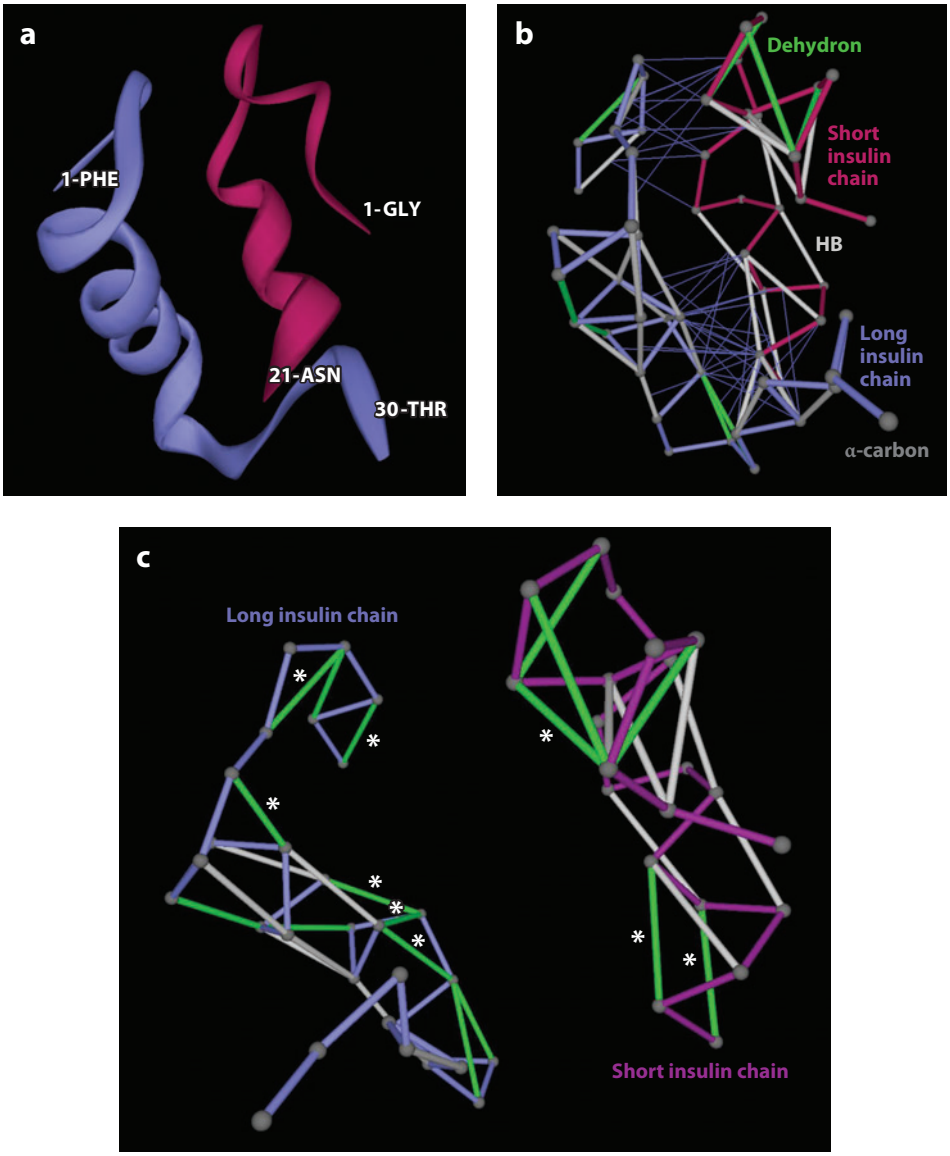

Figure 3

Cooperative effects arising from intermolecular wrapping sustain protein structure. (*a*) Ribbon representation of the two tethered structural chains of insulin, forming a complex. (*b*) Intermolecular wrapping of the two chains, enabling each chain to maintain its structural integrity. The backbone of each chain (*blue* and *magenta*) is represented by virtual bonds joining α-carbons. Well-wrapped backbone hydrogen bonds (HBs) are shown as light gray segments joining the α-carbons of the paired residues, and dehydrons are represented in green. A wrapping three-body interaction is depicted as a thin line from the wrapping residue to the center of mass of the hydrogen bond. Only intermolecular wrapping is shown. (*c*) In the absence of cooperativity provided by intermolecular wrapping, the high levels of structural deficiency of the individual structures would make them unsustainable in water. The long insulin chain (*blue*) contains six dehydrons (*asterisks*) that become well-wrapped hydrogen bonds upon association with the short insulin chain (*magenta*). Reciprocally, the small chain contains three dehydrons that become well-wrapped upon association. Adapted with permission from Reference 17.

Each three-body correlation represents the intermolecular wrapping of a hydrogen bond in one chain by the nonpolar groups of residues in its binding partner and only materializes upon association. According to the basic tenets of dehydron theory, a residue is a wrapper of a hydrogen bond if it contributes to the exclusion of water from the bond microenvironment by bringing to its proximity at least one nonpolar carbonaceous group. Thus, the two insulin chains are able to sustain their individual structures through a complex web of wrapping interactions (**Figure 3*b***) that promote water exclusion from the chain backbone, hence hampering the folding-disruptive effect of backbone hydration. Of crucial relevance to this analysis is the fact that a vast proportion of dehydrons in the free subunits become well-wrapped (buried) hydrogen bonds in the complex. More precisely, 6 out of the 11 dehydrons in the long chain and 3 out of 6 in the short chain become well-wrapped upon complexation; hence, the association introduces net relative decreases in PWIT on the order of 6/11 and 1/2, respectively.

The relevance of structural deficiency as a marker for protein association propensity has been further validated by contrasting PWIT computations with alanine-scanning dissection of P-P interfaces. This analysis effectively involves systematic site-specific truncation of individual side chains at the β-carbon and was implemented in order to identify hot spots, i.e., residues that contribute most significantly to the affinity. It should be emphasized that the experimental approach was originally motivated by the absence of a consistent physical picture to enable hot-spot prediction (8), a situation that has radically changed with the advent of theoretical methods to estimate PWIT. Thus, association free-energy differences ($\Delta\Delta G_a$) between mutant (m) and wild type (0) for the alanine substitution of each residue at the P-P interface have been contrasted with the interfacial free-energy difference between mutant and wild type for the free protein subunits (26). The association free energy is computed as $\Delta G_a = -RTlnK_a$, where K_a is the association equilibrium constant. The hot-spot residues (biggest contributors to affinity) obtained by alanine scanning were contrasted with the biggest generators of interfacial tension identified from an in silico shaving procedure that gives the change $-\Delta\Delta G_{if}$ introduced by a side-chain truncation at the β-carbon. A statistically significant correlation ($R^2 = 0.86$, $P < 10^{-7}$) exists between $-\Delta\Delta G_{if}$ and the association free-energy difference $\Delta\Delta G_a$ generated by site-specific substitution of interfacial residues. By focusing on this scanned P-P interface, this correlation upholds the relevance of PWIT as a determinant of protein associations.

Finally, the utility of PWIT (or its surrogate ν) as a measure of interactivity was established by examining a catalog of contact topologies for nonhomologous protein complexes with one to six subunits reported in the protein data bank (PDB) (26). For each complex, we computed the total P-P interface area after identifying the residues engaged in intermolecular contacts. For each protein subunit, the P-P interface is contained within the PWI region that generates tension in the free subunit, and there is a tight correlation between the surface areas for both regions, implying that regions on the protein surface that generate PWIT (i.e., those with $g < 4$ for nearby water) actually promote associations.

AN EVOLUTIONARY FRAMEWORK FOR INTERACTOME COMPLEXITY

A rather common mistake in delineating the origin of a phenotypic trait is to assume that the only evolutionary force at work is natural selection. The effect of random genetic drift is particularly important in species with low population size, where mildly deleterious mutations acquire a significant probability of getting fixed in the population (48–50). Transitions from prokaryotes to unicellular eukaryotes to multicellular eukaryotes involve orders-of-magnitude reductions in effective population size, with a resultant decline in the efficiency of selection and concurrent amplification of the

effects of random drift (26). Thus, a dehydron resulting from a mildly deleterious mutation that does not effectively disrupt the fold has a better chance of being fixed in a low-population species than in a unicellular organism. Being structural defects, dehydrons are not a priori expected to foster complexity, yet their stickiness suggests that they can act as promoters of constructive evolutionary change. This leads us to conjecture that as we compare homologous proteins across species (orthologs), we are likely to find higher levels of structural deficiency and concomitantly a higher association propensity in a multicellular eukaryote than in a unicellular organism. This is indeed the case (26), and hence population size becomes a crucial factor in determining the pathways underlying long-term phenotypic evolution. There is actually a phylogenetically broad inverse relationship between the power of random drift and the structural deficiency of protein subunits. This relation upholds the picture that low-population species accumulate dehydrons that promote secondary selection for P-P interactions. In turn, these associations make gene functions sustainable. This trend is supported by structural analysis and upholds the rather bold view that supramolecular protein organizations, the hallmarks of phenotypic diversity, may have arisen from nonadaptive mechanisms.

These assertions were validated by structurally comparing established gene orthologies, with the goal of delineating the role of drift in protein structural evolution across broad phylogenetic patterns. Evolutionary change is unlikely to significantly disrupt the native fold of an essential protein because the loss of function would lead to lethal fitness consequences. However, the drift hypothesis, articulated by Lynch in his analysis of nonadaptive constructiveness (48–50), predicts a negative relationship between population size (N) and the accumulation of mildly deleterious (in our case, dehydron-yielding) amino acid substitutions. The examination of the dehydron patterns embossed in the structures of orthologous proteins from vastly different lineages reveals the enhanced power of drift in eukaryotes (multicellular species in particular), resulting in a significant destabilization of P-W interfaces via the partial exposure of paired backbone polar groups that are otherwise protected in prokaryotes. In effect, the reduced efficiency of selection in small-N species promotes dehydron accretion, leading to protein structures that are more vulnerable to fold-disruptive hydration and generate a higher PWIT (**Figure 2**).

Because of the inherent stickiness of dehydrons (29), destabilization of P-W interfaces promotes the secondary recruitment of novel P-P associations. In turn, these associations restore the structural stability of soluble proteins by reducing the unstable P-W interface. Thus, we may state that interactome complexity arises as an evolutionary remedy to the structural deterioration promoted by drift. Thus, complex organisms may frequently develop P-P interactions as compensatory mechanisms for retaining key gene functions. Given the inherent stickiness of dehydrons, a physical contact between two mildly degraded protein units provides a selective niche for the emergence of novel P-P interactions. This suggestion that the hallmark of eukaryotic evolution, the origin of interactome complexity, may have arisen passively as a consequence of the power of drift upholds the null hypothesis in natural selection. It also removes the need for an unwarranted a priori advocacy for selective advantages of phenotypic complexity.

To gain insight into the evolution of interactome complexity, we adopted PWIT and its surrogate ν as indicators of potential molecular interactivity or association propensity. Comparison of orthologous proteins with different levels of oligomerization in different species (44) supports the view that PWIT measures the propensity for P-P association. The ratio of P-P interfaces (lower to higher degrees of complexation) exhibits a strong positive correlation with the ratio of PWITs for the respective free subunits. This implies that the degree of oligomerization correlates with the PWIT of the basic subunit (26).

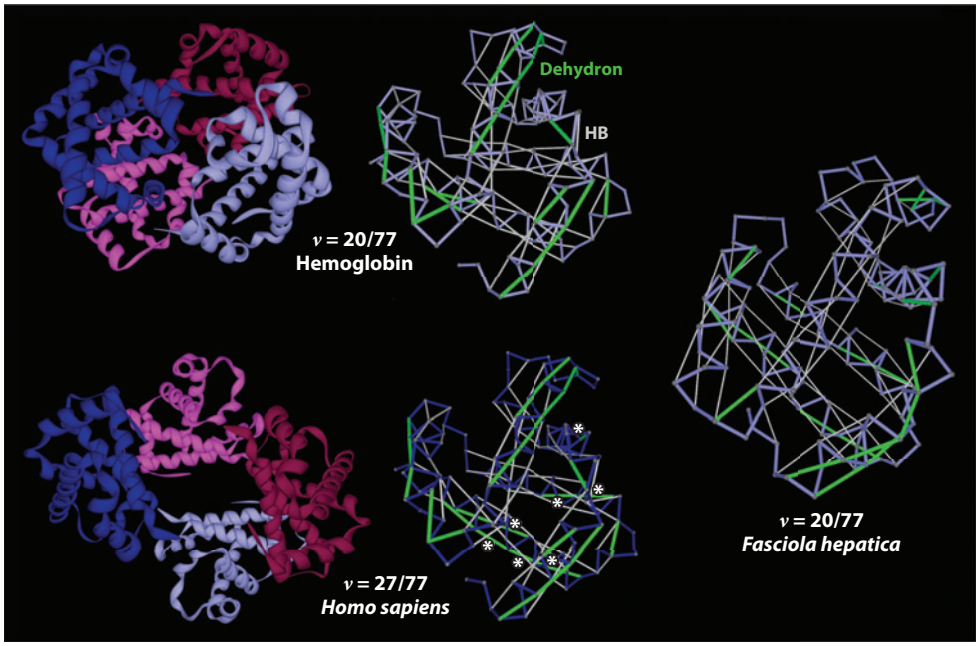

Figure 4

Comparison of dehydron patterns from hemoglobin orthologs from human and trematode. The orthologs present different oligomerization levels: tetrameric in human (*left* and *center*), monomeric in trematode (*right*). Backbone and hydrogen bond representation are consistent with **Figure 3**. The ribbon representations of the human complex and dissociated subunit are included as visual aids. The free subunit isolated from the tetramer in *Homo sapiens* has seven excess dehydrons (*asterisks*) when compared with the subunit within the tetrameric complex, where they become buried through intermolecular wrapping. As a consequence of this better wrapping, the overall extent of structural deficiency for the subunit within the human complex is identical to that of the natively monomeric hemoglobin, a protein that does not require associations to maintain its structural integrity, from the trematode *Fasciola hepatica*. This suggests that the accumulation of structural deficiencies in the human ortholog subunit promotes oligomerization as a means for reducing excess interfacial tension. However, the better-wrapped trematode hemoglobin does not need or promote oligomerization. Reproduced with permission from Reference 26.

As illustrated by the example in **Figure 3**, protein associations enable structurally deficient subunits to retain their structural integrity and are promoted by the stickiness of dehydrons. This evolutionary repair to structure degradation also opens up the possibility for innovative functional improvements, such as allosteric regulation. Thus, an isolated α-subunit of the human hemoglobin tetramer has seven dehydrons that become well protected within the tetrameric complex, implying that the oligomerization introduces structural stabilization. This stabilization within the complex reaches the same level as that attained by a free natively monomeric unit in another species, where hemoglobin is not allosterically regulated. Thus, ν for the subunit within the human complex is the same as that of an autonomously stable ortholog: the natively monomeric unit of hemoglobin from the trematode *Fasciola hepatica* (**Figure 4**). This implies that a structurally deficient unit not only necessitates and promotes association but, in so doing, opens up an opportunity to improve the regulation and efficiency of the underlying function. In general, as subsequently shown, the evolutionary fix to a degraded function is in effect a selective niche for functional improvement and innovation.

SELECTION INEFFICIENCY AS AN ENABLER OF SUPRAMOLECULAR ORGANIZATION

To evaluate whether association propensities in proteins are promoted by random genetic drift, we examined 106 groups of water-soluble proteins with orthologs in 36 species of vastly different population sizes (26). The analysis was limited by the need to have a minimal level of structural representation consisting of PDB-reported structures for at least two species. This dual representation enabled cross-validation of template-based, homology-threading models of three-dimensional structures for orthologs lacking PDB-reported structures (64, 67, 69). The structures were evaluated, ranked, and selected according to the energetic proximity between template and model (67). The accuracy of the homology-based prediction of PWIT was determined by adopting a test set of proteins with PDB-reported structures from two species, subjecting one member of each orthologous pair to homology threading through the other. Comparison of the indirect (homology-based) and direct (from PDB structure) estimates of PWIT demonstrates that when sequence identities are >35%, the predicted PWIT diverges <10% from the more direct estimate for the same protein. Given this threshold, only orthologous groups with at least 35% sequence identity were considered.

For each protein, the differences in association propensity across orthologs were then obtained by determining the differences in the interfacial free energy (ΔG_{if}) among species. To quantify organizational differences for a given protein compared across species, we introduced the relative complexation propensity ($M_{j,n}$) of a protein in ortholog group j (1, . . . , 106) from species n (1, . . . , 36), adopting *E. coli* as the reference species ($n = 1$) (26):

$$M_{j,n} = [(\Delta G_{if})_{j,n} - (\Delta G_{if})_{j,1}]/(\Delta G_{if})_{j,1}. \quad (2)$$

Thus, $M_{j,1}$ equals zero for all proteins in *E. coli*, whereas taxa with more deficient protein structures have positive values.

To establish a relative measure of the supramolecular organizational complexity of proteomes across species, we introduced a mean value (M_n of $M_{j,n}$) over all proteins evaluated for species n. This parameter is negatively correlated with the approximate effective population sizes of the species (**Figure 5**), given that the average ranking of the latter is prokaryotes > unicellular eukaryotes > invertebrates > vertebrates and land plants. An illustration of the trend toward increasing organizational complexity with reduced population sizes is provided in **Figure 6**, where the dehydron patterns and ν values for orthologs of the common metabolic enzyme dehydrofolate reductase (DHFR) are compared across three species with vastly different population size.

These results uphold the hypothesis that large organisms with small population sizes are under a stronger influence of random genetic drift, which results in an exacerbated accretion of mild structural deficiencies that promote higher levels of supramolecular organization. This phenomenon is not only an evolutionary remedy needed to retain structural integrity but also a promoter of higher regulatory modalities, leading to improved functional efficiency. By contrast, mutations yielding dehydrons are more frequently excluded by more efficient selection in species with larger population sizes (e.g., prokaryotes). Given that dehydrons are the main determinants of PWIT, the proteins of large organisms have a greater inherent tendency to increase their interactive complexity. Interactome complexity is then shaped in multicellular species by nonadaptive forces that result in the structural degradation of individual proteins.

The importance of nonadaptive forces in shaping interactome complexity is nowhere more apparent than in the regulatory machinery for gene expression (61), precisely at the core of phenotype realization. Thus, transcription factors (TFs) in unicellular species, such as phages and bacteria, are less unstable than in higher eukaryotes and often operate as single entities with high affinity (K_D~nM) for

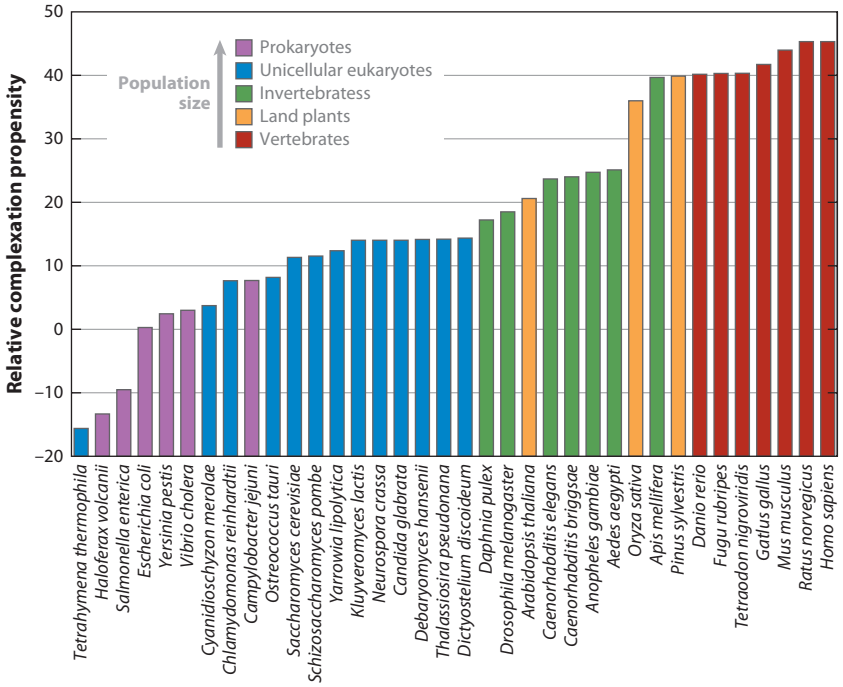

Figure 5

Structural degradation induces complexation propensity in species with low population size. Propensity for interactome complexity of 36 species with broadly diverse population sizes, relative to *Escherichia coli*. To highlight the relative power of random genetic drift, bars are color coded to indicate species grouping in broad population-size classes. Adapted with permission from Reference 26.

a relatively long (~17 base pairs), and therefore unique, DNA recognition sequence. By contrast, TFs for higher eukaryotes have low stability (higher PWIT) (58), are often disordered in the free state, and have much fuzzier low-affinity recognition of rather common DNA regions (~6 base pairs). However, they achieve higher levels of regulation efficiency through associations that enable a cooperative control of gene expression (61). Again, the nonadaptive forces contributing to the individual structural degradation of TFs, making them individually fuzzy, also open up the possibility of highly regulated gene expression patterns through cooperative binding of TF assemblies to DNA. This example highlights the importance of population size on the level of complexity of gene expression regulation, a hallmark of phenotypic complexity.

The generic nonadaptive scenario has been independently validated by comparing proteins from related species, such as endosymbiotic/intracellular bacteria (54) and their free-living relatives, that have experienced relatively recent divergences in effective population sizes but no major modifications in intra- or intercellular complexity. Thus, endosymbionts are thought to have experienced substantial reductions in effective population sizes relative to the free-living species, and there have been suggestions that endosymbionts experience elevated levels of random genetic drift based on ratios of substitution rates at silent and replacement sites. A comparative structural analysis of established orthologies in α- and γ-proteobacteria consistently shows that free-living species with larger effective population sizes have smaller ν

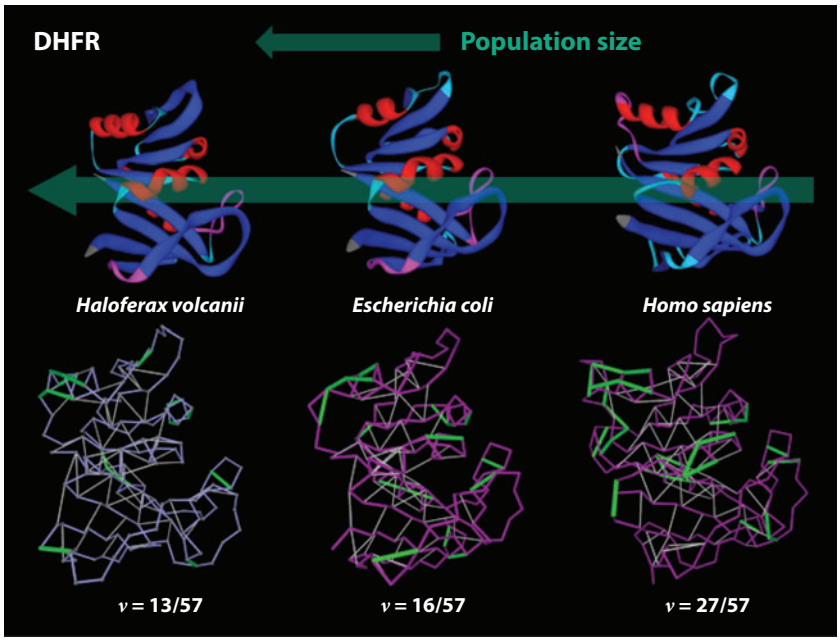

Figure 6

Structural deficiency of orthologs of the common metabolic enzyme dehydrofolate reductase (DHFR) revealing a progressive accumulation of dehydrons (*green segments*) in the orthologs of the archaea *Haloferax volcanii*, the bacterium *Escherichia coli*, and the mammal *Homo sapiens*. The extent of structural degradation increases as species population decreases, in accord with the general trend shown in **Figure 5**. The ribbon representation is adopted to illustrate the well-established structural conservation across proteins of common ancestry.

values for their orthologous proteins than their endosymbiont counterparts (26).

These results support the hypothesis that the range of population sizes encountered in different species promotes significantly different patterns of evolution at the level of supramolecular organization of the species proteomes. The nonadaptive structural degradation in species with small populations provides an opportunity for the recruitment of stabilizing P-P interactions, yielding a plausible mechanism for the emergence of supramolecular complexities that may ultimately result in phenotypic divergence. The evolutionary scenario described in this review does not neglect the significant role of natural selection in marshaling and exploiting the association propensities subsequent to their establishment in a nonadaptive setting. However, it also introduces a note of skepticism in regard to the need to attribute a selective advantage to organismal complexity.

FUTURE DIRECTIONS

This review revisits and upholds an evolutionary scenario in which nonadaptive factors play a significant role as enablers and promoters of supramolecular organization. According to the results described, interactome complexity constitutes an evolutionary fix to the structural degradation fostered by the selection inefficiency prevalent in multicellular eukaryotes. This view is not counterintuitive unless one is biased to think that complex traits must necessarily be an outcome of natural selection. However, the proposed scenario prompts the question: What are the long-term fitness consequences of such an evolutionary strategy?

Fitness catastrophe: evolutionary accident resulting in a highly deleterious trait that compromises survival

Humans have already experienced some of the fitness catastrophes that are likely to arise as a consequence of inefficient selection. The proteins with the largest accumulation of structural defects are the prions (25, 58), soluble proteins so poorly wrapped that they relinquish their soluble fold and form aberrant aggregates that lead to degenerative neuropathies and other debilitating conditions. These extreme cases illustrate the high level of genetic risk to which we are exposed as a result of our small population. The prion is a fitness catastrophe that provides clues as to where the evolutionary strategy to foster complexity may lead. This argument prompts us to delineate two medical imperatives for the 21st century: (*a*) the assessment of the long-term fitness cost of our progressive structural degradation and (*b*) the development of therapeutic agents to mitigate the impact of nonadaptive forces on the integrity of basic biological functions.

SUMMARY POINTS

1. The supramolecular organization of proteins is determined by discrete reproducible associations in the form of protein complexes.
2. An evolutionary force that enables and promotes supramolecular organizational complexity involves the paradoxically constructive role of random genetic drift.
3. For species with relatively low population size, selection inefficiency becomes a crucial factor in shaping an evolutionary niche to promote interactome complexity.
4. The evolutionary strategy to foster complexity involves partial structural degradation resulting from mildly deleterious mutation, which has a better chance to get fixed in species with low population size.
5. This evolutionary strategy may ultimately lead to fitness catastrophes arising from deregulated protein aggregation.

FUTURE ISSUES

1. What are the long-term fitness consequences of the evolutionary strategy to promote interactome complexity by exploiting the mild structural degradation of proteins?
2. In the case of humans, what are the therapeutic challenges that arise from the need to mitigate the impact of nonadaptive forces on the structural integrity of proteins?

DISCLOSURE STATEMENT

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

ACKNOWLEDGMENTS

This work was made possible through support from CONICET, the National Research Council of Argentina. The author is indebted to Professor Wen-Hsiung Li for our most helpful discussions.

LITERATURE CITED

1. Bahadur RP, Chakrabarti P, Rodie F, Janin J. 2003. Dissecting subunit interfaces in homodimeric proteins. *Proteins* 53:708–19

2. Bennett MJ, Choe S, Eisenberg D. 1994. Domain swapping: entangling alliances between proteins. *Proc. Natl. Acad. Sci. USA* 91:3127–31
3. Blundell TL, Srinivasan N. 1996. Symmetry, stability and dynamics of multidomain and multicomponent protein systems. *Proc. Natl. Acad. Sci. USA* 93:14243–48
4. Brasher BB, Van Etten RA. 2000. c-Abl has high intrinsic tyrosine kinase activity that is stimulated by mutation of the Src homology 3 domain and by autophosphorylation at two distinct regulatory sites. *J. Biol. Chem.* 275:35631–37
5. Brinda KV, Vishveshwara S. 2005. Oligomeric protein structure networks: insights into protein-protein interactions. *BMC Bioinforma.* 6:296
6. Changeux JP, Edelstein SJ. 2005. Allosteric mechanisms of signal transduction. *Science* 308:1424–28
7. Chen J, Liang H, Fernández A. 2008. Protein structure protection commits gene expression patterns. *Genome Biol.* 9:R107
8. Clackson T, Ultsch MH, Wells JA, de Vos AM. 1998. Structural and functional analysis of the 1:1 growth hormone:receptor complex reveals the molecular basis for receptor affinity. *J. Mol. Biol.* 277:1111–28
9. Cooper A, Dryden DT. 1984. Allostery without conformational change. A plausible model. *Eur. Biophys. J.* 11:103–9
10. Cui Q, Karplus M. 2008. Allostery and cooperativity revisited. *Protein Sci.* 17:1295–307
11. Deindl S. 2007. Structural basis for the inhibition of tyrosine kinase activity of ZAP-70. *Cell* 129:735–46
12. DePristo MA, Weinreich DM, Hartl DL. 2005. Missense meanderings in sequence space: a biophysical view of protein evolution. *Nat. Rev. Genet.* 6:678–87
13. Dobson CM. 2003. Protein folding and misfolding. *Nature* 426:884–90
14. Eaton WA, Henry ER, Hofrichter J, Mozzarelli A. 1999. Is cooperative oxygen binding by hemoglobin really understood? *Nat. Struct. Biol.* 6:351–58
15. Enright AJ, Iliopoulos I, Kyrpides NC, Ouzounis CA. 1999. Protein interaction maps for complete genomes based on gene fusion events. *Nature* 402:86–90
16. Fernández A. 2004. Keeping dry and crossing membranes. *Nat. Biotech.* 22:1081–84
17. Fernández A. 2012. Epistructural tension promotes protein associations. *Phys. Rev. Lett.* 108:188102
18. Fernández A. 2010. *Transformative Concepts for Drug Design: Target Wrapping*. Heidelberg, Ger.: Springer
19. Fernández A, Bazán S, Chen J. 2009. Taming the induced folding of drug-targeted kinases. *Trends Pharm. Sci.* 30:66–71
20. Fernández A, Berry RS. 2003. Proteins with hydrogen-bond packing defects are highly interactive with lipid bilayers: implications for amyloidogenesis. *Proc. Natl. Acad. Sci. USA* 100:2391–96
21. Fernández A, Berry RS. 2004. Molecular dimension explored in evolution to promote proteomic complexity. *Proc. Natl. Acad. Sci. USA* 101:13460–65
22. Fernández A, Berry RS. 2010. Golden rule for buttressing vulnerable soluble proteins. *J. Proteome Res.* 9:2643–48
23. Fernández A, Colubri A, Berry RS. 2000. Topology to geometry in protein folding: β-lactoglobulin. *Proc. Natl. Acad. Sci. USA* 97:14062–66
24. Fernández A, Crespo A. 2008. Protein wrapping: a marker for association, aggregation and drug design. *Chem. Soc. Rev. R. Soc. Chem.* 37:2373–82
25. Fernández A, Kardos J, Scott R, Goto Y, Berry RS. 2003. Structural defects and the diagnosis of amyloidogenic propensity. *Proc. Natl. Acad. Sci. USA* 100:6446–51
26. Fernández A, Lynch M. 2011. Nonadaptive origins of interactome complexity. *Nature* 474:502–5
27. Fernández A, Rogale K, Scott R, Scheraga HA. 2004. Inhibitor design by wrapping packing defects in HIV-1 proteins. *Proc. Natl. Acad. Sci. USA* 101:11640–45
28. Fernández A, Scheraga HA. 2003. Insufficiently dehydrated hydrogen bonds as determinants for protein interactions. *Proc. Natl. Acad. Sci. USA* 100:113–18
29. Fernández A, Scott R. 2003. Adherence of packing defects in soluble proteins. *Phys. Rev. Lett.* 91:018102
30. Fernández A, Scott R, Berry RS. 2004. The nonconserved wrapping of conserved protein folds reveals a trend towards increasing connectivity in proteomic networks. *Proc. Natl. Acad. Sci. USA* 101:2823–27
31. Fernández A, Zhang X, Chen J. 2008. Folding and wrapping soluble proteins: exploring the molecular basis of cooperativity and aggregation. *Prog. Nucleic Acids Res. Transl. Sci.* 83:57–87

32. Fong JH, Geer LY, Panchenko AR, Bryant SH. 2007. Modeling the evolution of protein domain architectures using maximum parsimony. *J. Mol. Biol.* 366:307–15
33. Formaneck MS, Ma L, Cui Q. 2006. Reconciling the "old" and "new" views of protein allostery: a molecular simulation study of chemotaxis Y protein (CheY). *Proteins* 63:846–67
34. Goodsell DS, Olson AJ. 2000. Structural symmetry and protein function. *Annu. Rev. Biophys. Biomol. Struct.* 29:105–53
35. Gray MW, Lukes J, Archibald JM, Keeling PJ, Doolittle WF. 2001. Cell biology. Irremediable complexity? *Science* 330:920–21
36. Ispolatov I, Yuryev A, Mazo I, Maslov S. 2005. Binding properties and evolution of homodimers in protein-protein interaction networks. *Nucleic Acids Res.* 33:3629–35
37. Janin J. 2008. Biochemistry. Dicey assemblies. *Science* 319:165–66
38. Johnson LN, Lewis RJ. 2001. Structural basis for control by phosphorylation. *Chem. Rev.* 101:2209–42
39. Kern D, Zuiderweg ER. 2003. The role of dynamics in allosteric regulation. *Curr. Opin. Struct. Biol.* 13:748–57
40. Kummerfeld SK, Teichmann SA. 2005. Relative rates of gene fusion and fission in multi-domain proteins. *Trends Genet.* 21:25–30
41. Kuriyan J, Eisenberg D. 2007. The origin of protein interactions and allostery in colocalization. *Nature* 450:983–90
42. Laskowski RA, Gerick F, Thornton JM. 2009. The structural basis of allosteric regulation in proteins. *FEBS Lett.* 583:1692–98
43. Levy ED. 2007. PiQSi: Protein quaternary structure investigation. *Structure* 15:1364–67
44. Levy ED, Boeri Erba E, Robinson CV, Teichmann SA. 2008. Assembly reflects evolution of protein complexes. *Nature* 453:1262–65
45. Levy ED, Pereira-Leal JB, Chothia C, Teichmann SA. 2006. 3D complex: a structural classification of protein complexes. *PLoS Comput. Biol.* 2:e155
46. Liang H, Sandberg WS, Terwilliger TC. 1993. Genetic fusion of subunits of a dimeric protein substantially enhances its stability and rate of folding. *Proc. Natl Acad. Sci. USA* 90:7010–14
47. Lindahl E, Hess B, Van der Spoel D. 2001. GROMACS 3.0: A package for molecular simulation and trajectory analysis. *J. Mol. Model.* 7:302–317
48. Lynch M. 2007. *The Origins of Genome Architecture*. Sunderland, Massachusetts: Sinauer
49. Lynch M. 2007. The frailty of adaptive hypotheses for the origins of organismal complexity. *Proc. Natl. Acad. Sci. USA* 104:8597–604
50. Lynch M, Conery JS. 2003. The origins of genome complexity. *Science* 302:1401–4
51. Marianayagam NJ, Sunde M, Matthews JM. 2004. The power of two: protein dimerization in biology. *Trends Biochem. Sci.* 29:618–25
52. Monod J. 1968. *Nobel Symposium 11: Symmetry and Function of Biological Systems at the Macromolecular Level*. Stockholm, Swed.: Almqvist & Wiksell
53. Monod J, Wyman J, Changeux JP. 1965. On the nature of allosteric transitions: a plausible model. *J. Mol. Biol.* 12:88–118
54. Moran NA. 1996. Accelerated evolution and Muller's ratchet in endosymbiotic bacteria. *Proc. Natl. Acad. Sci. USA* 93:2873–78
55. Pardee AB. 2006. Regulatory molecular biology. *Cell Cycle* 5:846–52
56. Pereira-Leal JB, Levy ED, Kamp C, Teichmann SA. 2007. Evolution of protein complexes by duplication of homomeric interactions. *Genome Biol.* 8:R51
57. Perutz MF. 1970. Stereochemistry of cooperative effects in haemoglobin. *Nature* 228:726–39
58. Pietrosemoli N, Crespo A, Fernández A. 2007. Dehydration propensity of order-disorder intermediate regions in soluble proteins. *J. Proteome Res.* 6:3519–26
59. Popovych N, Sun S, Ebright RH, Kalodimos CG. 2006. Dynamically driven protein allostery. *Nat. Struct. Mol. Biol.* 13:831–38
60. Powers ET, Powers DL. 2003. A perspective on mechanisms of protein tetramer formation. *Biophys. J.* 85:3587–99
61. Ptashne M, Gann A. 2002. *Genes & Signals*. New York: Cold Spring Harbor Lab. Press

62. Royer WE, Zhu H, Gorr TA, Flores JF, Knapp JE. 2005. Allosteric hemoglobin assembly: diversity and similarity. *J. Biol. Chem.* 280:27477–80
63. Rowlinson JS, Widom B. 1982. *Molecular Theory of Capillarity*. New York: Oxford Univ. Press
64. Sali A, Blundell TL. 1993. Comparative protein modeling by satisfaction of spatial restraints. *J. Mol. Biol.* 234:779–815
65. Schlessinger J. 2000. Cell signaling by receptor tyrosine kinases. *Cell* 103:211–25
66. Swain JF, Gierasch LM. 2006. The changing landscape of protein allostery. *Curr. Opin. Struct. Biol.* 16:102–8
67. Wiederstein M, Sippl MJ. 2007. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res.* 35:W407–410
68. Xu W, Harrison SC, Eck MJ. 1997. Three-dimensional structure of the tyrosine kinase c-Src. *Nature* 385:595–602
69. Zhou H, Skolnick J. 2010. Improving threading algorithms for remote homology modeling by combining fragment and template comparisons. *Proteins* 78:2041–48