Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Structural Biology

Frustration, function and folding

Diego U Ferreiro[1], Elizabeth A Komives[2] and Peter G Wolynes[3,4,5,6]

Natural protein molecules are exceptional polymers. Encoded in apparently random strings of amino-acids, these objects perform clear physical tasks that are rare to find by simple chance. Accurate folding, specific binding, powerful catalysis, are examples of basic chemical activities that the great majority of polypeptides do not display, and are thought to be the outcome of the natural history of proteins. Function, a concept genuine to Biology, is at the core of evolution and often conflicts with the physical constraints. Locating the frustration between discrepant goals in a recurrent system leads to fundamental insights about the chances and necessities that shape the encoding of biological information.

Addresses
[1] Protein Physiology Lab, FCEyN-Universidad de Buenos Aires, IQUIBICEN/CONICET, Intendente Güiraldes 2160, Ciudad Universitaria, C1428EGA Buenos Aires, Argentina
[2] Department of Chemistry and Biochemistry, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92092-0378, USA
[3] Center for Theoretical Biological Physics, Rice University, Houston, TX, USA
[4] Department of Chemistry, Rice University, Houston, TX, USA
[5] Department of Physics, Rice University, Houston, TX, USA
[6] Department of Biosciences, Rice University, Houston, TX, USA

Current Opinion in Structural Biology 2018, **48**:68–73

This review comes from a themed issue on **Folding and binding**

Edited by **Amedeo Caflisch** and **Stefano Gianni**

https://doi.org/10.1016/j.sbi.2017.09.006

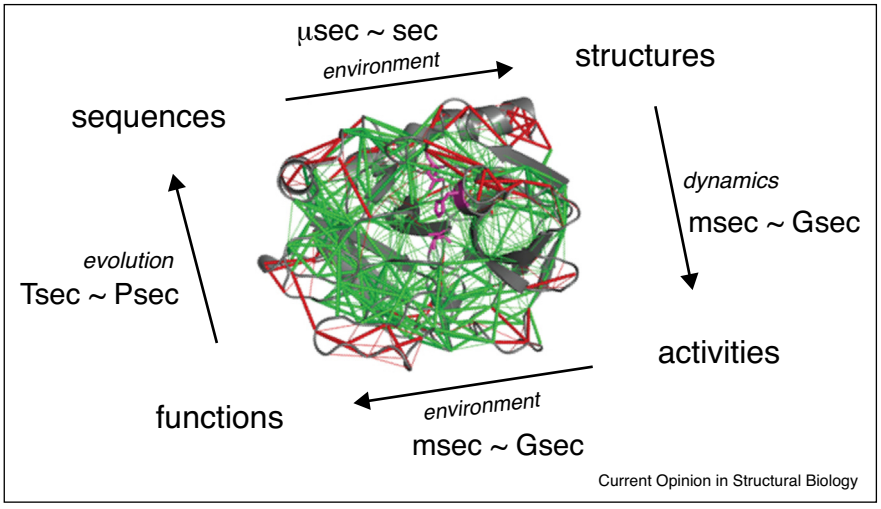
Introduction

Inorganic crystals are beautiful like wallpaper — every design element fits repetitively in place. Biomolecules are beautiful too but, like a painting by a great master, are made up of diverse parts, each still falling in place, where small details of how they are put together suggest action or life. Repetition is satisfying but action requires some conflict or frustration. Frustration occurs when a physical system is unable simultaneously to achieve minimum energy individually for each and every part of it [1]. Frustration can happen for geometric reasons (as in a triangular spin lattice or a complicated protein topology) and/or due to competition between the interactions of the basic elements. The application of this concept to protein molecules [2] paved the way to the development of the Energy Landscape Theory of protein folding, which provides powerful tools for understanding natural protein molecules [3,4]. The basic notion is the recognition that natural proteins are evolved polymers distinguishing them from random polypeptides thrown together entirely by chance. It is at the protein level that conflicting biological goals meet in the specification of the sequences. In order to fold robustly, proteins must satisfy a large number of local interactions simultaneously, a task that is feasible when frustration between interactions of the elements is low [2,5]. Beyond folding, however proteins perform chemical activities that impose further restrictions on the sequences that encode a given fold, possibly conflicting with the necessity of self-organization [6]. Looking for the deviations of the expectations of structural stability hints at other teleonomical goals that are needed for action (Figure 1).

To see the conflicts encoded in protein sequence and structure one needs a reliable way for measuring the degree of satisfaction of a general energy function, a daunting task for large molecular objects such as proteins, where thousands of atoms interact by a myriad of weak forces. It is apparent, however, that we do not need to get to the fundamental quantum mechanics, as many of the forces can be accounted for in coarse-grained descriptions. Useful approximations to the energy functions are now at hand, and are being developed at different levels. Full atomic force fields are accurate enough to analyze the folding of small proteins (albeit at large computational cost), and multiple heuristics have led to ways to design sequences that fold to simple topologies [7–9]. These approximations rely on the fact that many of the interactions can be modeled with effective forces averaged over the solvent environment, such that the polymer can be described as being made of pseudoatoms that encode distinct properties [10–12].

Having a reliable way of measuring the overall free energy of a protein structure, one can explore how the free energy varies when the sequence or the structure of the protein changes. Ten years ago a simple heuristic method to explore these relations was presented [13,14••]. To analyze the existence of energetic conflicts in a folded protein, the energy of structural or sequence decoys is measured with respect to the native state. A local frustration index is defined as the Z-score of the free energy of parts of the native structure with respect to the distribution of the energy of rearranged decoys. If a native pair of interacting residues has an energy that lies in the most favorable end of

Figure 1



The recurrent flow of biological information conflicts in natural proteins. In appropriate environmental conditions, amino acid sequences encode the formation of specific structures. Structures interconvert due to frustration and give rise to chemical activities, contributing to the specification of multiple biological functions. The functional structures restrict the exploration and fixation of the genomic sequences. These processes occur in timescales that span many orders of magnitude. Conflicts between folding and function can be located in extant proteins, reflecting functional constraints. At center a representation of local frustration on Thrombin, a modern protease [17]. The backbone is shown as a continuous gray trace. The protein is networked by a connected set of minimally frustrated contacts (green) while there are patches of highly frustrated contacts (red). Thin lines represent water-mediated interactions and the catalytic residues are shown in magenta. Local frustration patterns were calculated with the frustratometer.tk server [14••].

the distribution, the interaction is labelled as minimally frustrated, as most changes in sequence in that location will destabilize the overall structure. In general, about 40% of the native contacts found in natural globular domains fall into this class, in line with the theoretical expectations and experimental results [13]. About half of the interactions can be labelled as neutral as they do not contribute distinctively to the total energy, and around 10% of the interactions are highly frustrated. These are regions in which most local sequence or structural changes would lower the free energy of the system. These frustrated regions are typically found as patches on the surface of globular domains. They must be held there over evolutionary time as well as physiological time at the expense of other interactions, that is, they conflict with the robust folding of a domain. The adaptive value for a molecule to tolerate spatially localized frustration arises from the way such frustration sculpts protein dynamics for specific functions. In a monomeric protein the alternate configurations caused by locally frustrating an otherwise largely unfrustrated structure provide specific control of the thermal motions guiding them in useful directions [15–19]. Alternatively a site that is frustrated in a monomeric protein may become less frustrated in the final larger assembly of this protein with partners, thus guiding specific association [13,20,21]. We will review here recent findings insights that come from analyzing the specifics of local frustration in several systems. For a detailed discussion of the basics of frustration biophysics the reader is referred to [22].

Topology can be frustrating

Frustration can be reflected in the topology of a proteins native states. Assuming that no energetic conflicts are present in a folded molecule, the chain connectivity by itself strongly restricts the sampling of the conformational space. Forming structure in one region may hinder the consolidation of structure in a distant part. Such topological frustration can be quantified using folding simulations with structure-based models [23]. Gosavi *et al.* have shown how subtle differences in the native topology can give rise to large biological effects was shown in the Interleukin (IL) family. Despite having similar three-dimensional structures and stabilities, IL-1β promotes downstream signaling, whereas IL-1Ra inhibits it. The folding traps caused by topology that distinguish these proteins make IL-1β fold more slowly than IL-1Ra. The differences in the landscape can be ascribed mainly in two loops, which when mutated can switch the functional forms [24]. Regions that participate in function are inferred to cause the different folding traps.

The flavodoxin-like fold is an old protein architecture. Proteins with this fold often misfold in search of their functionally active forms. This susceptibility to misfold is caused by the timing of the consolidation of structure, forming partially folded intermediates [25]. The differential stabilization of the intermediates can be related to the resistance to aggregation and degradation which change function in a system sense. Sequence