Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Structural Biology

Conformational diversity and the emergence of sequence signatures during evolution

Gustavo Parisi[1], Diego Javier Zea[1], Alexander Miguel Monzon[1] and Cristina Marino-Buslje[2]

Proteins' native structure is an ensemble of conformers in equilibrium, including all their respective functional states and intermediates. The induced-fit first and the pre-equilibrium theories later, described how structural changes are required to explain the allosteric and cooperative behaviours in proteins, which are key to protein function. The conformational ensemble concept has become a key tool in explaining an endless list of essential protein properties such as function, enzyme and antibody promiscuity, signal transduction, protein–protein recognition, origin of diseases, origin of new protein functions, evolutionary rate and order–disorder transitions, among others. Conformational diversity is encoded by the amino acid sequence and such a signature can be evidenced through evolutionary studies as evolutionary rate, conservation and coevolution.

Addresses
[1] Departamento de Ciencia y Tecnología, Universidad Nacional de Quilmes, Roque Saenz Pena 182, 1876 Bernal, Argentina
[2] Bioinformactics Unit, Fundación Instituto Leloir, Av. Patricias Argentinas 435, C1405BWE Buenos Aires, Argentina

Corresponding author: Marino-Buslje, Cristina (cmb@leloir.org.ar)

Current Opinion in Structural Biology 2015, **32**:58–65

This review comes from a themed issue on **Sequences and topology**

Edited by **M Madan Babu** and **Anna R Panchenko**

http://dx.doi.org/10.1016/j.sbi.2015.02.005

Introduction

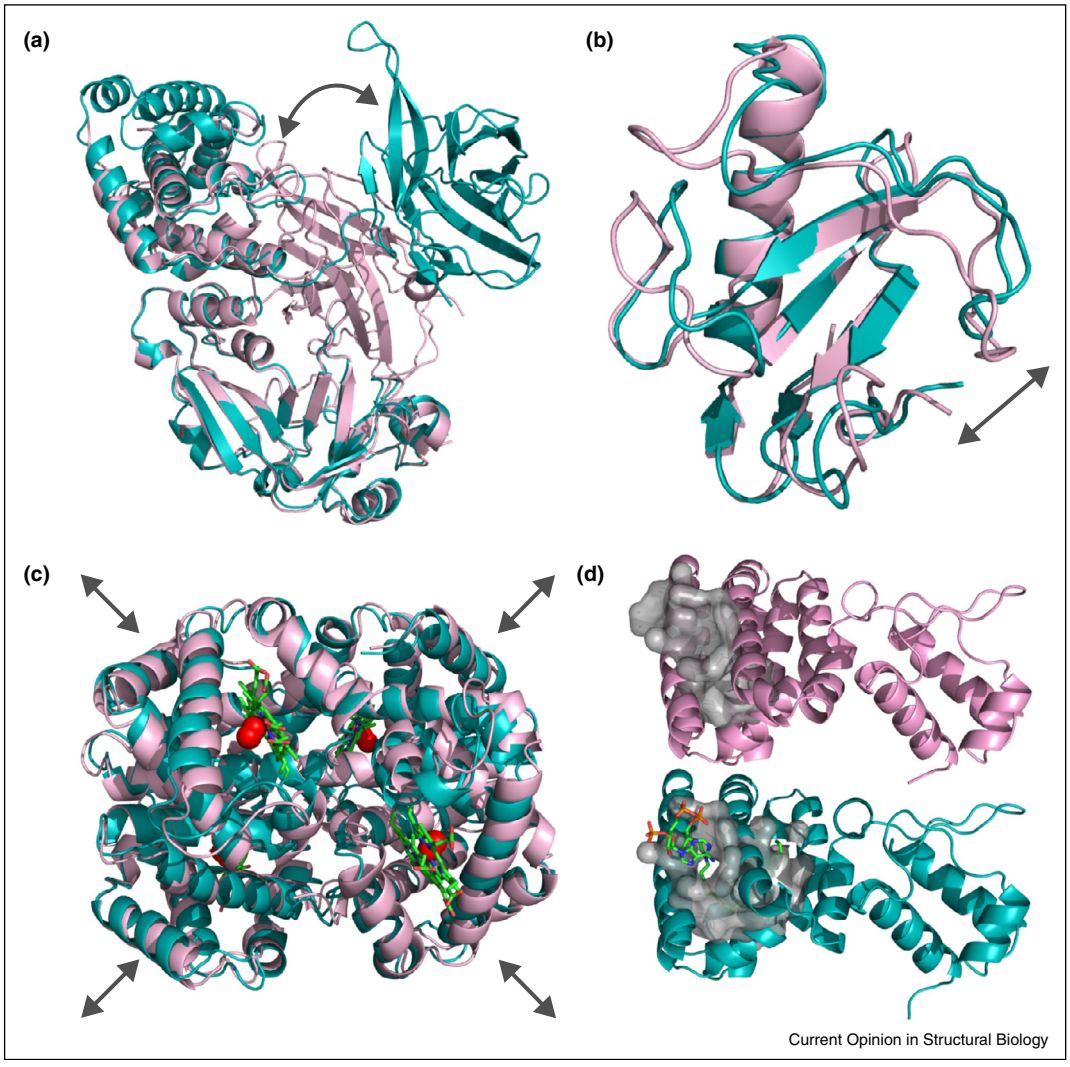
Recently, an overwhelming number of experiments and studies have unequivocally shown the dynamic nature of protein function. However, many experiments and a great deal of evidence were needed to challenge the early view of Linus Pauling's protein native state definition as a 'uniquely defined configuration' [1]. This change started in the 1950 with the experiments of F. Karush, who proposed the notion that the native state of proteins could contain different conformers with similar energies in dynamic equilibrium to explain the binding heterogeneity in seroalbumin [2]. The structural differences between the different conformers define differential binding capacities towards ligands, a property that Karush called *configurational adaptability*. Later, Monod, Wyman and Changeux [3] and Koshland [4] proposed models to explain the allosteric and cooperative behaviours in proteins, which are key to protein function (for a review see [5]). Both models emphasize protein conformational changes but in very different ways. Monod and co-workers proposed that the native state of proteins is described by a pre-existing conformational equilibrium, which is shifted by the differential stabilization of the conformers in the presence of ligands. The induced-fit theory proposed by Koshland describes the native state of a protein as a unique structure, which undergoes conformational changes induced by the ligand in order to obtain a better fit. Recently, several studies have supported the existence of an equilibrium between pre-existing conformations, giving support to Monod's model [6–8]. However, both models are phenomenological and do not provide a residue-level or atomistic-level explanation of the allosteric or cooperative behaviour [9••]. More recently, thermodynamic and structure-based models have also been suggested, which combine properties from both models [10•,11••]. Whatever the mechanisms underlying conformational change are, it is clear that in many cases, protein function requires switching between different structures in the native state. The conformational ensemble concept has become a key tool in explaining an endless list of essential protein properties, such as function [7,12–13], enzyme and antibody promiscuity [14], signal transduction [15], protein–protein recognition [16], origin of diseases [17], emergence of new protein functions [18], evolutionary rate [19••] and order–disorder transitions, among others. Based on the studies of Lesk and Chothia [20–21] it is clear that protein structure is conserved more than protein sequence during evolution. Therefore, the conservation of protein structure introduces additional constraints on sequence divergence to preserve biological function. Thus, the dynamic nature of the native state presents new challenges to explain the specific contributions of different conformers within the ensemble to the observed residue substitution pattern and how different mutations can modify the conformational equilibrium.

In this manuscript, we review recent findings about the relevance of the conformational diversity in proteins and its possible influence in sequence divergence. We also review the use of the evolutionary rate, evolutionary models and residue coevolution models to examine the relationship between sequence divergence and protein dynamics during evolution.

Extension of conformational diversity in protein space

Structural differences between conformers define the conformational diversity of the protein. These differences can be as large as relative movements of subunits or entire domains as well as changes in smaller segments, such as loops or secondary structural elements [22]. These elements can move as rigid bodies [23] involving hinge or shear displacements, or they can undergo tertiary structural rearrangements [24] (Figure 1). Although conformational diversity can be studied using computational methods, such as molecular dynamics [25] and coarse-grain normal mode analysis [26], experimental-based evidence of conformational diversity comes from the analysis of protein crystals and nuclear magnetic resonance (NMR) of proteins. A collection of structures of a given protein obtained under different conditions can be viewed as snapshots of the conformational space of the native ensemble [27–28]. Assuming the pre-existing equilibrium hypothesis, different ligands (such as the natural substrates or inhibitors)

Figure 1



Structure motions. **(a)** Diphtheria toxin in complex with NAD (PDB code: 1tox_A — light blue) and crystallized under acidic conditions (PDB code: 4ow6_A — pink). The C-alpha RMSD between conformers in the superimposed region is 1.78 Å. This protein presents a hinge motion up to 65 Å that allows the domain rotation indicated with the grey arrow. **(b)** Guanine-specific ribonuclease F1 in a complex with pyroglutamic acid (NMR structure, PDB code: 1rck_A model 21 — light blue) and free form (PDB code: 1fus_A — pink). These structures show a shear motion with a C-alpha RMSD of 1.99 Å, as indicated with the grey arrow. **(c)** Structures of haemoglobin in the R-state (oxy) (PDB code: 1hho — light blue) and T-state (deoxy) (PDB code: 2hhb — pink). The two conformations present a C-alpha RMSD of 2.34 Å and show a rigid body motion, as indicated by grey arrows. **(d)** FadR transcription factor, the apo (PDB code: 1e2x_A — pink) and holo states in complex with myristoyl-CoA (PDB code: 1h9g_A — light blue) with a C-alpha RMSD of 1.28 Å. The main structural differences between these conformers are the change in the volume of the cavity (1586 $Å^3$ between conformers, shown as the grey surface) and the number of tunnels that allow ligand access to the protein (1 and 3, respectively).