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

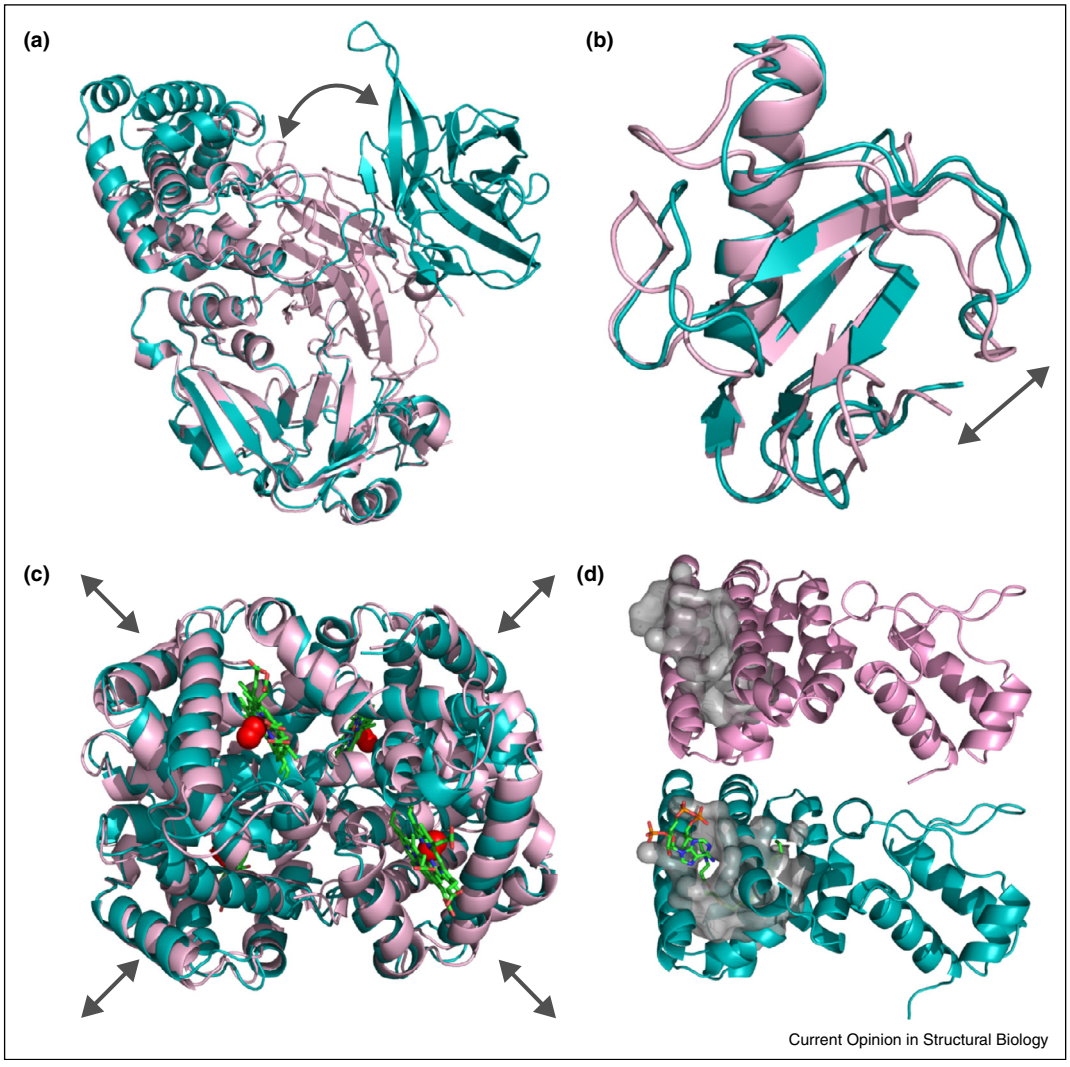
Recently, an overwhelming number of experiments and studies have unequivocally shown the dynamic nature of protein function. However, many experiments and a great deal of evidence were needed to challenge the early view of Linus Pauling's protein native state definition as a 'uniquely defined configuration' [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



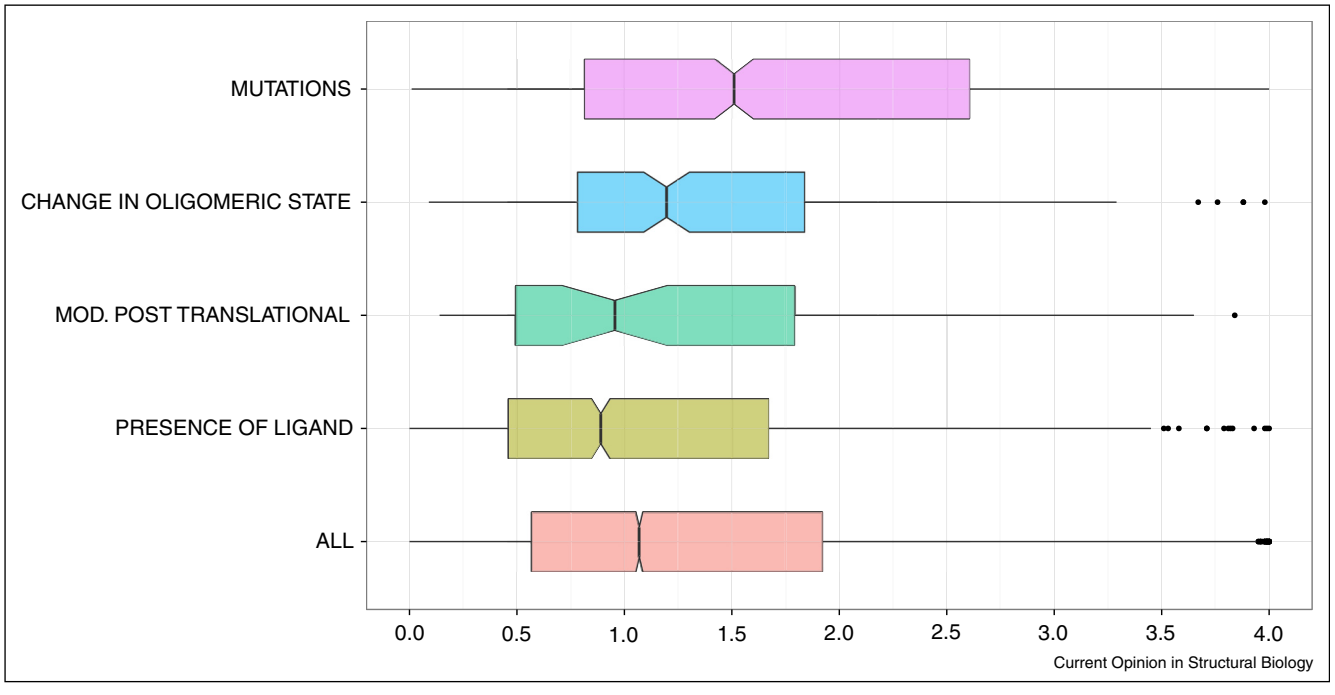
Structure motions. **(a)** Diphtheria toxin in complex with NAD (PDB code: 1tox_A — light blue) and crystallized under acidic conditions (PDB code: 4ow6_A — pink). The C-alpha RMSD between conformers in the superimposed region is 1.78 Å. This protein presents a hinge motion up to 65 Å that allows the 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).

co-crystallized with a given protein could differentially shift the conformational equilibrium towards different conformers. Large-scale comparisons of the structural differences between conformers taken from the Protein Data Bank (PDB) have shown that the root-mean-square deviation (RMSD) distribution between conformers has a peak at 0.3 Å but is largely skewed towards higher RMSD values up to a maximum of 24 Å [29]. An analysis of apo and substrate-bound forms of the same protein showed that 75% of the studied proteins have an RMSD between conformers below 1 Å and 95% below 2 Å [30]. More recently, an analysis of nearly 17 000 proteins with more than 11 million pairs of conformers from the CoDNaS database [31] showed a similar trend (Figure 2). In addition to the presence of ligands, the CoDNaS database allows for comparison of conformers differing in their post-translational modifications, oligomeric states and the presence of mutations assumed to be the cause of the conformers' population shift. The presence of mutations has been shown to produce the largest conformational changes. As shown in Figure 2, the majority of proteins have moderate to low numbers of displacements in their backbones; however, minor side-chain rotations could modify the size of inner cavities or favour the closing and opening of tunnels and pockets [32–33] (Figure 1d). Increasing evidence shows the role of tunnels and cavities is to allow the transit of different ligands from the surface to the active site or binding regions [34–35], generally without causing significant changes to the backbone structure. In addition, large movements in proteins are not necessary to define high-affinity and low-affinity conformers or even the existence of allosteric or cooperative effects [24].

Conformational diversity and sequence evolution

To study the relationship between protein structure conservation and sequence divergence, several models of protein evolution have been proposed [36–39]. These models use an explicit representation of the protein structure to derive the structurally constrained substitution pattern using computational simulation; therefore, they are called *constrained* models. For example, in the Structurally Constrained Protein Evolutionary (SCPE) model [39], substitutions during simulations occur if the structural perturbation, evaluated using energetic calculations, is below a given threshold. Using the SCPE model and a set of 900 proteins with different folding patterns and conformations, we studied how the observed substitution pattern is explained by the different contributions of each available conformer of each protein in the dataset [40•]. When the individual contribution of a conformer is mapped in the observed substitution pattern of a protein family, we found that, on average, 37% of the sites evolve under structural constrains. However, when

Figure 2



Boxplot of the RMSD distribution of different conformers. The RMSD distribution of the maximal difference between conformers. 'All proteins' includes 17 276 proteins present in the CoDNaS database (two outliers are not displayed). 'Ligands' includes the subset of CoDNaS considering bound–unbound conformers without any other variation (1981 proteins). In the same way, 'post-translational modifications' includes the subset of CoDNaS proteins with post-translational modifications (70 proteins). 'Oligomeric state' includes the subset of CoDNaS proteins that undergo changes in the oligomeric state (246 proteins). 'Mutations' includes the subset of CoDNaS proteins that have mutations (987 proteins).

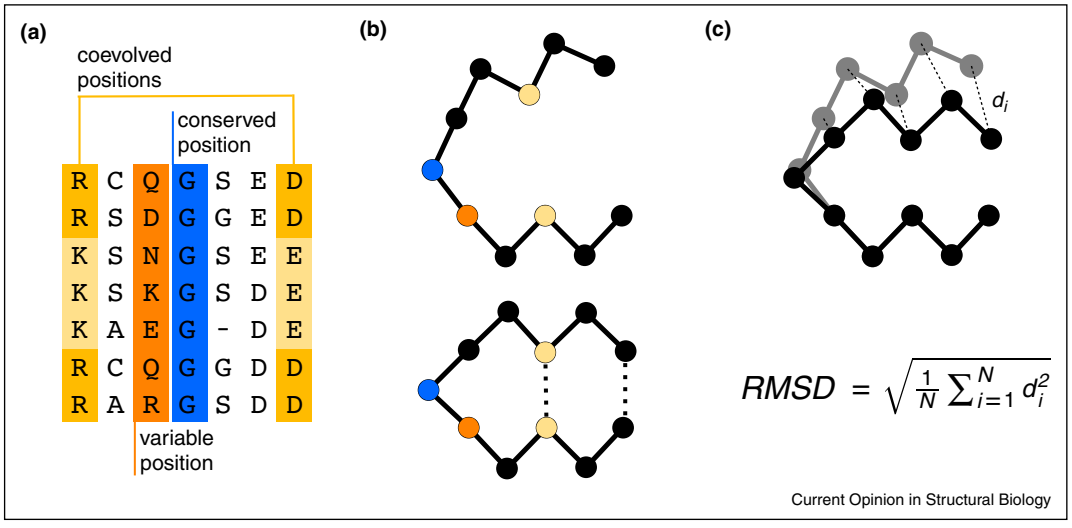
all the conformers were used to do the mapping, we found that this percentage increased to 48% of the sites. Although most of the conformers share the same constraints (and are mostly structurally similar), our results also indicate that individual conformers have their own constraints on the substitution pattern, modulating up to 30% of the protein positions. Interestingly, if each conformer contributes differentially to the substitution pattern, then the effect of mutations on protein stability should be studied over the whole conformational space. Considering disease and polymorphic mutations, we found that 35% of the mutations could either be classified as stabilizing, neutral or destabilizing depending on the selected conformer for the corresponding energetic evaluation, while in the rest of the mutations, the classification do not change even as the conformers change [41]. Our results indicate that the observed pattern of sequence divergence during evolution is then a complex outcome of the different contributions of the conformers found in the native ensemble. As a direct consequence of Lesk and Chothia results mentioned above, we hypothesized that proteins showing greater conformational diversity (measured as the maximum RMSD between two conformers) should have more structural constraints and then would probably evolve slower than proteins with fewer constraints between their conformers. We recently tested this hypothesis on a set of 16 species, each with 700 orthologous proteins and found a negative correlation between conformational diversity and evolutionary rate [19••]. Because protein dynamics is closely related to protein function, this result could indicate the indirect influence of protein function on the evolutionary rate.

Residue coevolution networks and conformational diversity: implications in protein function and structure

Another approach to study sequence/structure evolution is through the analysis of the interdependencies between residues in a protein. Multiple sequence alignments (MSAs) of homologous proteins carry at least two levels of information. One is given by the amino acid conservation at each position in the protein sequence, and the second is provided by the inter-relationship between two or more positions. The extent of the relationship between two positions in a protein family can be estimated measuring their mutual information (MI). Positions that suffered concerted compensatory changes are said that have coevolved. Figure 3 shows a schematic representation of the different types of positions (Figure 3a), their possible location in the 3D structure and their difference in RMSD (Figure 3b,c). Coevolutionary studies have been applied to many biologically relevant questions, such as functionally important residue prediction [42,43,44•], allosteric pathway studies [45–46], protein conformation prediction [47–49], protein–protein interaction studies [50] and the study of conformational diversity [51•,52,53].

Thus, some protein functions occur through the activity of conserved residues while others are maintained by concerted changes in a group of residues forced to coevolve. Such is the case of enzymes, where the catalytic

Figure 3

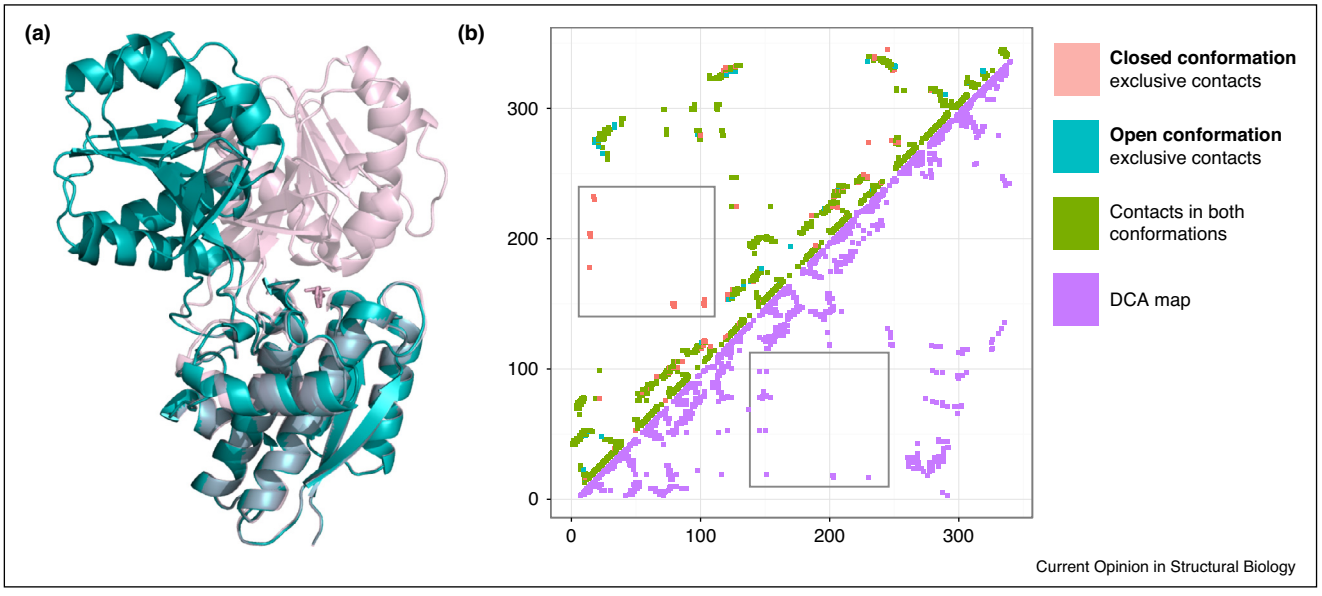


(a) Representation of a multiple sequence alignment of homologous proteins. The conserved position is highlighted in blue, and the variable position in orange. The positions that coevolved are highlighted in dark yellow. The residues within these positions where changes occurred are shown in light yellow. **(b)** Simplified alpha carbon representation of a protein conformational change from an open to closed conformation. Residue contacts in the closed conformation are marked with a dotted line. Residues are coloured as in the multiple sequence alignment, where residues in yellow show a coevolution signal. **(c)** The degree of conformational change between two conformers can be calculated after a rigid structure superposition and RMSD calculation of the alpha carbons.

residues (CR) carry a particular signature defined by networks of close proximity residues with high MI [44•]. Such a signature is due to the fact that the requirement to maintain a particular function places limitations on the diversification of the structural environment along the course of evolution. Residues undergo sequence variations as they evolve and form spatial clusters in the protein structure that may be part of binding sites, catalytic sites or allosteric pathways. In accordance with the observation of Halabi *et al.* [54], it was demonstrated that groups of coevolving residues tend to be close, forming a *sector* (which we called the *MI3D cluster*) when mapped onto the 3D structure [55•]. Furthermore, it was found that, amongst the many MI3D clusters usually present in a protein domain, those containing catalytic residues have distinguishable network properties [55•]. Finally, as residues in an MI3D cluster coevolve with residues within their group and to a lesser extent with residues of other MI3D clusters, they can be considered as units of quasi-independent evolution. Thus, a protein could be thought of as a network of MI3D clusters, which are physically connected in the tertiary structure, and each one has a different functional role [54,55•]. The distinctive topological characteristics of the nodes in such a network allows us to identify the clusters devoted to the catalytic function in the protein [55•]. Some current methods to infer a coevolutionary measure between residues include the following: MISTIC, a web server for the graphical representation of the information contained within an MSA and a complete analysis tool for MI networks in protein families [53]; EVCouplings [48]; and PSICOV [47]. The last two methods focus on residue contact prediction.

The alignment and analysis of evolutionarily related protein sequences can also serve as sources of information to infer conformational diversity [52,56•]. Jeon *et al.* [51•] investigated the relationship between sequence evolution and protein conformational changes and discussed that structural transitions are encoded in the amino acid sequence as coevolving residue pairs. They found that highly coevolving residues are clustered in the flexible regions of proteins and facilitate structural transitions by forming and disrupting their interactions cooperatively. Along the same lines, Liu and Bahar [52] studied the relationship between the relative mobility in the collective dynamics (using a Gaussian network model), the amino acid conservation and the correlated mutation propensities at each sequence position, for a set of 34 families of enzymes. The most important conclusions from this work are that conserved residues have minimal fluctuations in global modes and that an increase in sequence variability is accompanied with an increase in conformational mobility. Residues acting as global hinges during collective dynamics are often conserved. It also explains that a large number of sequence variations at highly flexible regions are neutral. However, some are accompanied by compensatory mutations, and those coevolving pairs (or clusters) at high mobility regions in global modes are presumably involved in substrate

Figure 4



(a) Structural comparison between the apo (light blue, PDB code: 1usg) and the holo states (pink, PDB code: 1usi) of the L-leucine binding protein, showing domain closure [56•]. **(b)** Upper triangle: Contact map of these structures. The two conformations share the majority of contacts (green dots); however, there are specific contacts of the open (pink) and closed forms (light blue). Lower triangle: DCA map (violet points) calculated using FreeContact [58]. DCA finds some of the contacts (grey box).

recognition, suggesting that their behaviour is driven by functional requirements. This study shows that, at least for enzymes, mobility/restrictions and sequence variability/conservations exhibit a weak but statistically significant correlation.

It has also been shown that by integrating direct coupling analysis (DCA) for protein contact prediction [57] into coarse-grained physical models of proteins, such as structure-based models (SBMs), it might be possible to predict both, the open and closed conformations of a protein as well as its intermediate states [56•] (see Figure 4). These predictions may be possible because DCA captures evolutionarily significant residue–residue correlations, regardless of whether the interaction stabilizes the final or intermediate states.

Conclusions

The determination of the complete functional conformational landscape of proteins remains a challenge. However, we have described interesting advances in the field, mostly related to the capacity to detect specific sequential signatures related to conformational diversity. As conformational diversity is related to protein function, these results indicate that we may be near to the prediction of biological function using specific conformational diversity signatures at the sequence level. However, these methods have been applied on a small set of proteins, and, therefore, further testing is needed to assess the relationship between conformational diversity and sequence divergence. In addition, structurally constrained models of protein evolution should include an explicit representation of protein dynamics and be able to simulate differential contributions of specific conformers accordingly to the equilibrium representation in the ensemble. Although the results reviewed are promising, some limitations need to be addressed, such as the high amount of information required for both DCA and coevolutionary estimation and the highly demanding computational models of evolution in order to apply these methodologies on a genomic, proteomic or metabolomic scale.

Acknowledgements

GP and CMB are researchers of the Argentinean National Research Council (CONICET). DJZ and AM are fellows of the Argentinean National Research Council (CONICET). CMB is also funded by CONICET and the Leloir Institute Foundation. GP is also funded by the National University of Quilmes.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Mirsky AE, Pauling L: **On the structure of native denatured, and coagulated proteins**. *Proc Natl Acad Sci U S A* 1936, **22**:439-447.

2. Karush F: **Heterogeneity of the binding sites of bovine serum albumin1**. *J Am Chem Soc* 1950, **72**:2705-2713.

3. Monod J, Wyman J, Changeux JP: **On the nature of allosteric transitions: a plausible model**. *J Mol Biol* 1965, **12**:88-118.

4. Koshland DJ, Ray WJJ, Erwin MJ: **Protein structure and enzyme action**. *Fed Proc* 1958, **17**:1145-1150.

5. Changeux J-P, Edelstein S: **Conformational selection or induced fit? 50 years of debate resolved**. *F1000 Biology Reports* 2011, **3**:19.

6. James L, Roversi P, Tawfik D: **Antibody multispecificity mediated by conformational diversity**. *Science* 2003, **299**:1362-1367.

7. del Sol A, Tsai C, Ma B, Nussinov R: **The origin of allosteric functional modulation: multiple pre-existing pathways**. *Structure* 2009, **17**:1042-1050.

8. Ma B, Nussinov R: **Enzyme dynamics point to stepwise conformational selection in catalysis**. *Curr Opin Chem Biol* 2010, **14**:652-659.

9. •• Cui Q, Karplus M: **Allostery and cooperativity revisited**. *Protein Sci* 2008, **17**:1295-1307.
This paper is an excellent up-to-date review on the mechanism underling cooperativity and allostery. Using an atomistic level, the authors use several examples and computational simulations enhancing the description made by classical allostery models (Monod and Koshland models).

10. • Tsai C-J, Nussinov R: **A unified view of 'how allostery works'**. *PLoS Comput Biol* 2014, **10**:e1003394.
Although allostery is a well established concept; in this manuscript the authors describe this property using three different views: thermodynamics, population shift and structure. Using these mechanisms to describe the same process, the authors offer an unified view of allostery.

11. •• Motlagh HN, Wrabl JO, Li J, Hilser VJ: **The ensemble nature of allostery**. *Nature* 2014, **508**:331-339.
This work focuses on the conformational ensemble nature of allosteric proteins. It explains the full spectrum of strategies that nature uses to facilitate allostery, such as, rigid body movement, folded yet dynamic structure, and intrinsic disorder.

12. Hilser VJ: **An ensemble view of allostery**. *Science* 2010, **327**:653-654.

13. Boehr DD, McElheny D, Dyson HJ, Wright PE: **The dynamic energy landscape of dihydrofolate reductase catalysis**. *Science* 2006, **313**:1638-1642.

14. James LC, Roversi P, Tawfik DS: **Antibody multispecificity mediated by conformational diversity**. *Science* 2003, **299**:1362-1367.

15. Smock RG, Gierasch LM: **Sending signals dynamically**. *Science (New York, NY)* 2009, **324**:198-203.

16. Yogurtcu ON, Erdemli SB, Nussinov R, Turkay M, Keskin O: **Restricted mobility of conserved residues in protein–protein interfaces in molecular simulations**. *Biophys J* 2008, **94**:3475-3485.

17. Lynch TJ, Bell DW, Sordella R, Gurubhagavatula S, Okimoto RA, Brannigan BW, Harris PL, Haserlat SM, Supko JG, Haluska FG *et al.*: **Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib**. *N Engl J Med* 2004, **350**:2129-2139.

18. Tokuriki N, Stricher F, Serrano L, Tawfik DS: **How protein stability and new functions trade off**. *PLoS Comput Biol* 2008, **4**:e1000002.

19. •• Zea DJ, Alexander Miguel M, Fornasari MS, Marino-Buslje C, Parisi G: **Protein conformational diversity correlates with evolutionary rate**. *Mol Biol Evol* 2013, **30**:1500-1503.
This work demonstrates a negative correlation between conformational diversity and protein evolutionary rate. It also explains that it is independent of the protein expression level and comparable in magnitude and sign with the correlation between gene expression level and evolutionary rate. They suggest that the structural constraints underlying protein dynamics and function could modulate protein divergence.

20. Lesk A, Chothia C: **How different amino acid sequences determine similar protein structures: the structure and evolutionary dynamics of the globins**. *J Mol Biol* 1980, **136**:225-270.

21. Chothia C, Lesk A: **The relation between the divergence of sequence and structure in proteins**. *EMBO J* 1986, **5**:823-826.

22. Gerstein M, Krebs W: **A database of macromolecular motions**. *Nucleic Acids Res* 1998, **26**:4280-4290.

23. Daily MD, Gray JJ: **Allosteric communication occurs via networks of tertiary and quaternary motions in proteins**. *PLoS Comput Biol* 2009, **5**:e1000293.

24. Tsai C-J, del Sol A, Nussinov R: **Protein allostery, signal transmission and dynamics: a classification scheme of allosteric mechanisms**. *Mol Biosyst* 2009, **5**:207-216.

25. Karplus M, Kuriyan J: **Molecular dynamics and protein function**. *Proc Natl Acad Sci U S A* 2005, **102**:6679-6685.

26. Bahar I, Rader AJ: **Coarse-grained normal mode analysis in structural biology**. *Curr Opin Struct Biol* 2005, **15**:586-592.

27. Zoete V, Michielin O, Karplus M: **Relation between sequence and structure of HIV-1 protease inhibitor complexes: a model system for the analysis of protein flexibility1**. *J Mol Biol* 2002, **315**:21-52.

28. Best RB, Lindorff-Larsen K, DePristo MA, Vendruscolo M: **Relation between native ensembles and experimental structures of proteins**. *Proc Natl Acad Sci U S A* 2006, **103**:10901-10906.

29. Burra PV, Zhang Y, Godzik A, Stec B: **Global distribution of conformational states derived from redundant models in the PDB points to non-uniqueness of the protein structure**. *Proc Natl Acad Sci U S A* 2009, **106**:10505-10510.

30. Gutteridge A, Thornton J: **Conformational changes observed in enzyme crystal structures upon substrate binding**. *J Mol Biol* 2005, **346**:21-28.

31. Monzon AM, Juritz E, Fornasari MS, Parisi G: **CoDNaS: a database of conformational diversity in the native state of proteins**. *Bioinformatics* 2013, **29**:2512-2514.

32. Gora A, Brezovsky J, Damborsky J: **Gates of enzymes**. *Chem Rev* 2013, **113**:5871-5923.

33. Brylinski M, Skolnick J: **What is the relationship between the global structures of apo and holo proteins?** *Proteins Struct Funct Bioinform* 2008, **70**:363-377.

34. Sonavane S, Chakrabarti P: **Cavities and atomic packing in protein structures and interfaces**. *PLoS Comput Biol* 2008, **4**:e1000188.

35. Binkowski TA, Freeman P, Liang J: **pvSOAR: detecting similar surface patterns of pocket and void surfaces of amino acid residues on proteins**. *Nucleic Acids Res* 2004, **32**:W555-W558.

36. Bastolla U, Roman HE, Vendruscolo M: **Neutral evolution of model proteins: diffusion in sequence space and overdispersion**. *J Theor Biol* 1999, **200**:49-64.

37. Robinson DM, Jones DT, Kishino H, Goldman N, Thorne JL: **Protein evolution with dependence among codons due to tertiary structure**. *Mol Biol Evol* 2003, **20**:1692-1704.

38. Kleinman CL, Rodrigue N, Lartillot N, Philippe H: **Statistical potentials for improved structurally constrained evolutionary models**. *Mol Biol Evol* 2010, **27**:1546-1560.

39. Parisi G, Echave J: **Structural constraints and emergence of sequence patterns in protein evolution**. *Mol Biol Evol* 2001, **18**:750-756.

40. Juritz E, Palopoli N, Fornasari MS, Fernandez-Alberti S, Parisi G:
• **Protein conformational diversity modulates sequence divergence**. *Mol Biol Evol* 2013, **30**:79-87.
It is well established that conservation of protein structure during evolution constrains sequence divergence. However, the native state of a protein is not unique and is better represented by an ensemble of conformers. In this manuscript the authors explore the contribution of different conformers to the substitution pattern using 900 different protein families.

41. Juritz E, Fornasari MS, Martelli PL, Fariselli P, Casadio R, Parisi G: **On the effect of protein conformation diversity in discriminating among neutral and disease related single amino acid substitutions**. *BMC Genomics* 2012, **13**:S5.

42. Kuipers RKP, Joosten H-J, Verwiel E, Paans S, Akerboom J, van der Oost J, Leferink NGH, van Berkel WJH, Vriend G, Schaap PJ: **Correlated mutation analyses on super-family alignments reveal functionally important residues**. *Proteins Struct Funct Bioinform* 2009, **76**:608-616.

43. Chakrabarti S, Panchenko AR: **Coevolution in defining the functional specificity**. *Proteins Struct Funct Bioinform* 2009, **75**:231-240.

44. Marino Buslje C, Teppa E, Di Doménico T, Delfino JM, Nielsen M:
• **Networks of high mutual information define the structural proximity of catalytic sites: implications for catalytic residue identification**. *PLoS Comput Biol* 2010, **6**:e1000978.
This work demonstrates that catalytic residues carry a particular signature defined by networks of close proximity residues with high mutual information (MI), and that this signature can be applied to distinguish functional from other non-functional conserved residues. It also concludes that the requirement to maintain a particular function places limitations on the diversification of the structural environment along the course of evolution.

45. Lockless SW, Ranganathan R: **Evolutionarily conserved pathways of energetic connectivity in protein families**. *Science* 1999, **286**:295-299.

46. Shi Z, Resing KA, Ahn NG: **Networks for the allosteric control of protein kinases**. *Curr Opin Struct Biol* 2006, **16**:686-692.

47. Jones DT, Buchan DWA, Cozzetto D, Pontil M: **PSICOV: precise structural contact prediction using sparse inverse covariance estimation on large multiple sequence alignments**. *Bioinformatics* 2012, **28**:184-190.

48. Marks DS, Hopf TA, Sander C: **Protein structure prediction from sequence variation**. *Nat Biotechnol* 2012, **30**:1072-1080.

49. Kosciolek T, Jones DT: ***De novo* structure prediction of globular proteins aided by sequence variation-derived contacts**. *PLOS ONE* 2014, **9**:e92197.

50. Ovchinnikov S, Kamisetty H, Baker D: **Robust and accurate prediction of residue–residue interactions across protein interfaces using evolutionary information**. *eLife* 2014.

51. Jeon J, Nam H-J, Choi YS, Yang J-S, Hwang J, Kim S: **Molecular**
• **evolution of protein conformational changes revealed by a network of evolutionarily coupled residues**. *Mol Biol Evol* 2011, **28**:2675-2685.
This work investigates the relationship between sequence evolution and protein conformational changes and discusses that structural transitions are encoded in the amino acid sequence as coevolving residue pairs. Highly coevolving residues are clustered in the flexible regions and facilitate structural transitions by forming and disrupting their interactions cooperatively.

52. Liu Y, Bahar I: **Sequence evolution correlates with structural dynamics**. *Mol Biol Evol* 2012, **29**:2253-2263.

53. Simonetti FL, Teppa E, Chernomoretz A, Nielsen M, Marino Buslje C: **MISTIC: mutual information server to infer coevolution**. *Nucleic Acids Res* 2013, **41**:W8-W14.

54. Halabi N, Rivoire O, Leibler S, Ranganathan R: **Protein sectors: evolutionary units of three-dimensional structure**. *Cell* 2009, **138**:774-786.

55. Aguilar D, Oliva B, Marino Buslje C: **Mapping the mutual**
• **information network of enzymatic families in the protein structure to unveil functional features**. *PLOS ONE* 2012, **7**:e41430.
This work combines mutual information and structural data to describe the amino acid networks within a protein and their interactions. They found that clusters of co-evolving residues usually evolve independently, which could be related to a fail-safe mechanism. Among those clusters, the ones related to the catalytic site of an enzyme have distinguishable topological properties in the network. Finally, a significant enrichment of functional residues (e.g. metal binding, susceptibility to detrimental mutations) in the clusters was discovered, which could be the foundation of new prediction tools.

56. • Morcos F, Jana B, Hwa T, Onuchic JN: **Coevolutionary signals across protein lineages help capture multiple protein conformations**. *Proc Natl Acad Sci U S A* 2013, **110**:20533-20538.
Conformational plasticity can be retrieved from a collection of evolutionary related protein sequences. This is due to that diverse states (intermediates/closed), which are of functional importance, are selected by evolution. DCA captures evolutionarily significant residue–residue correlations, regardless of whether the interaction stabilizes the final or intermediate state.

57. Morcos F, Pagnani A, Lunt B, Bertolino A, Marks DS, Sander C, Zecchina R, Onuchic JN, Hwa T, Weigt M: **Direct-coupling analysis of residue coevolution captures native contacts across many protein families**. *Proc Natl Acad Sci U S A* 2011, **108**:E1293-E1301.

58. Kajan L, Hopf T, Kalas M, Marks D, Rost B: **FreeContact: fast and free software for protein contact prediction from residue co-evolution**. *BMC Bioinformatics* 2014, **15**:85.