Repeat proteins challenge the concept of structural domains

Rocío Espada*[1], R. Gonzalo Parra*[1], Manfred J. Sippl†, Thierry Mora‡, Aleksandra M. Walczak§ and Diego U. Ferreiro*[2]

*Protein Physiology Lab, Dep de Química Biológica, Facultad de Ciencias Exactas y Naturales, UBA-CONICET-IQUIBICEN, Buenos Aires, C1430EGA, Argentina
†Center of Applied Molecular Engineering, Division of Bioinformatics, Department of Molecular Biology, University of Salzburg, 5020 Salzburg, Austria
‡Laboratoire de physique statistique, CNRS, UPMC and Ecole normale supérieure, 24 rue Lhomond, 75005 Paris, France
§Laboratoire de physique théorique, CNRS, UPMC and Ecole normale supérieure, 24 rue Lhomond, 75005 Paris, France

Abstract

Structural domains are believed to be modules within proteins that can fold and function independently. Some proteins show tandem repetitions of apparent modular structure that do not fold independently, but rather co-operate in stabilizing structural forms that comprise several repeat-units. For many natural repeat-proteins, it has been shown that weak energetic links between repeats lead to the breakdown of co-operativity and the appearance of folding sub-domains within an apparently regular repeat array. The quasi-1D architecture of repeat-proteins is crucial in detailing how the local energetic balances can modulate the folding dynamics of these proteins, which can be related to the physiological behaviour of these ubiquitous biological systems.

Introduction

It was early on noted that many natural proteins typically collapse stretches of amino acid chains into compact units, defining structural domains [1]. These domains typically correlate with biological activities and many modern proteins can be described as composed by novel 'domain arrangements' [2]. For globular proteins, this fact facilitates the description, evolution and construction of single amino acid chains that comprise a set of integrated biological functions, akin to tinkering [3] with modular components. Many natural proteins contain tandem repeats of similar amino acid stretches. These have been usefully classified into groups: short repeats up to five residues usually fold into fibrillar structures, whereas repeats longer than ∼60 residues fold as independent tandem domains [4]. There is a class of repeat proteins in which each repeat does not fold when isolated but only folds in the presence of neighbouring repeat-units. For these proteins, the separation into 'domains' is not obvious to identify. Solenoid repeat-proteins are made up of tandem arrays of <20–40 similar amino acid stretches that usually fold up into elongated architectures of stacked repeating structural motifs (Figure 1). A coarse representation of them as quasi-1D objects yields surprisingly rich insights into their folding dynamics [5]. This class of repeat proteins are thought to be stabilized only by interactions within each repeat and between neighbouring repeats, with no obvious contacts between residues much more distant in sequence. Thus they can be pictured as elongated objects that could be broken down to repeat-units, yet the folding of the repeat-array comprises subtle balances of the energetics within and between repeats, challenging the concept of structural domain.

Key words: ankyrin-repeat, local frustration, repeat-protein.

Abbreviation: DI; direct information.

[1] These authors contributed equally to this work.

[2] To whom correspondence should be addressed (email ferreiro@qb.fcen.uba.ar).

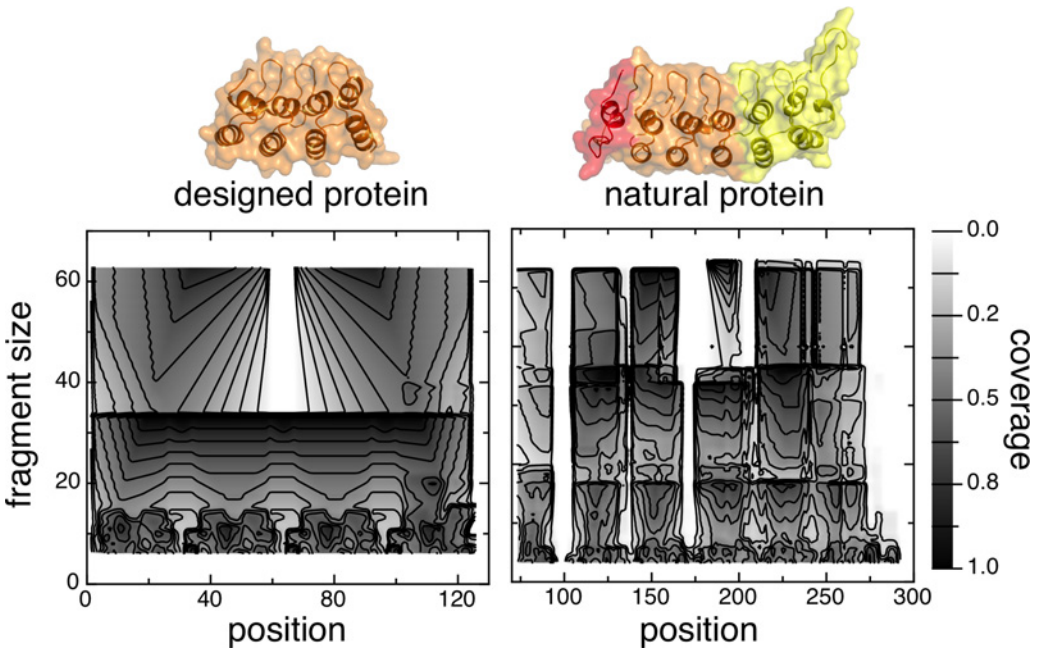
Definition of the repeat-units

Over the last years, several algorithms have been used to characterize repetitions in protein sequences. Most methods are based on the self-alignment of the primary structure and more sophisticated implementations use spectral analysis of pseudo-chemical characteristics of the amino acids [6]. It is not surprising that sequence-based methods fail to infer true structural repetitions since the same structural motif can be encoded by sequences that appear completely unrelated, which is the case in several repeat-protein families. There are only a few methods available to detect repetitions based on structural information. Some methods search for repeats by aligning the structure to itself [7,8]. Machine learning provided a fast method to recognize repeat regions in solenoid structures [9], which is being used to create a manually curated database [10]. There are many families of proteins with identified repeating motifs [4,11]. Nevertheless, the methods described above result in conflicting characterizations of the repeating units [6,12], even for basic parameters such as the size, the number and location of the repeating elements and the grouping of these into higher order patterns. In order to reconcile these views, we recently developed basics concepts and methods for the detection of repeats in protein structures that lead to an automated and consistent annotation of the size and location of the repeat-units [13].

The basic analogy underlying these developments is the treatment of a repeat-protein as a mosaic composed of repetitions of similar tiles. The algorithm exhaustively analyses the repetition of every possible continuous fragment of a protein structure and defines the portions that best

Figure 1 | **Folding sub-domains in two ankyrin repeat protein**

The maps show the coverage of how often each residue (*x*-axis) is covered by a repetition of a fragment of the protein of a given length (*y*-axis). At left, a designed protein four ankyrin (4ANK) (pdb: 1n0r, A) that appears as a periodic object at a fragment size of 33 residues. To the right, the natural protein IκBα (pdb:1ikn, D) in which structural sub-domains are apparent for every fragment size. On top, the ribbon representations of the structures coloured according to the number of sub-domains identified.



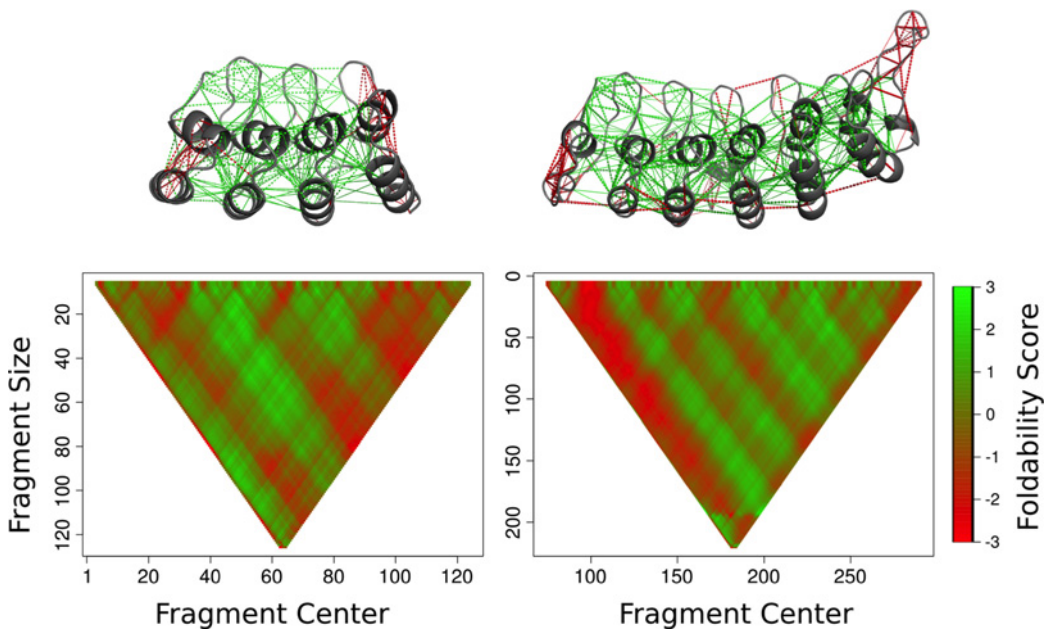
describe the overall structure. To do so, each fragment is repeated, translated, rotated and aligned with respect to the complete molecule. Using a fast and robust structural alignment protocol together with a well-defined metric [14,15], the subset of alignments that best describe the whole structure in terms of basic tiles is obtained. The tessellation lends itself to an intuitive visualization of the repeating units and their association into higher order patterns [13]. Notably, it was found that some architectures can be described as nearly periodic, whereas in some others, clear separations between repetitions exist. Figure 1 shows the results of the coverage of the structural space from the decomposition and tiling of every possible continuous fragment of two example repeat proteins. In the case of a natural protein, there appears to be boundaries between repeat-regions, corresponding to putative sub-domains. In contrast, a designed protein of the same class does not show boundaries but appears as a periodic object (Figure 1). This sequence-independent method was applied to several families of repeat-proteins and a continuous spectra of symmetrical arrangements of repeat-units have been described [13].

Exploring repeat-proteins energy landscapes

Over the last decade, the folding dynamics of several ankyrin-repeat proteins have been investigated. The first proteins to be studied were natural proteins consisting of short repeat-arrays, of ∼3–4 typical repeat-units. In most of these, the equilibrium folding is consistent with a two-state model: either all (or none) of the repeats are found folded [16–18]. This is consistent with the notion that short repeat-arrays constitute autonomous folding domains, in which all repeats fold co-operatively. The breakdown of co-operativity was first discovered in the ankyrin array of the Notch protein [19]. Equivalent mutations along the repeats showed that despite its apparent modular structure, the notch ankyrin domain unfolds as a co-operative unit consisting of six repeats. Only destabilizing substitutions in the last repeat leads to a multi-state unfolding transitions, showing that the coupling that gives rise to long-range co-operativity in the wild-type protein had a weak link in the C-terminal region. A similar break down in folding co-operativity was found in the ankyrin repeat region of inhibitor of kappa B alpha (IκBα). This protein displays two folding transitions: a non-co-operative conversion under weak perturbation that was localized to the C-terminal repeats and a major co-operative folding phase involving the N-terminal repeats [20]. Notably, stabilization of the weak regions leads to anomalous physiological responses [21] showing that the separation of the ankyrin-array in folding sub-domains is crucial for the biological activity of IκBα [22]. The largest repeat protein for which the folding mechanism has been characterized is D^{34}, a fragment of AnkyrinR. This protein populates an equilibrium folding intermediate and it has been shown that the folding kinetics can be described as that of two six-repeat sub-domains [23].

Figure 2 | Local frustration and folding routes

The maps represent the hierarchical foldability of every possible continuous fragment of a given size (*y*-axis) and centre (*x*-axis). The relative foldability is defined as $\Theta = \Delta E/(\delta E/\mathrm{sqrt}N)$, where ΔE is the difference between the energy of a given fragment with respect to the mean energy of all fragments of the same length and δE is the S.D. of the decoy set distribution of size *N*. On top, the ribbon representation of the structures of the proteins, a designed 4ANK at left and IκBα at right. The local frustration patterns are shown, with green lines being minimally frustrated interactions and red lines highly frustrated interactions [35].



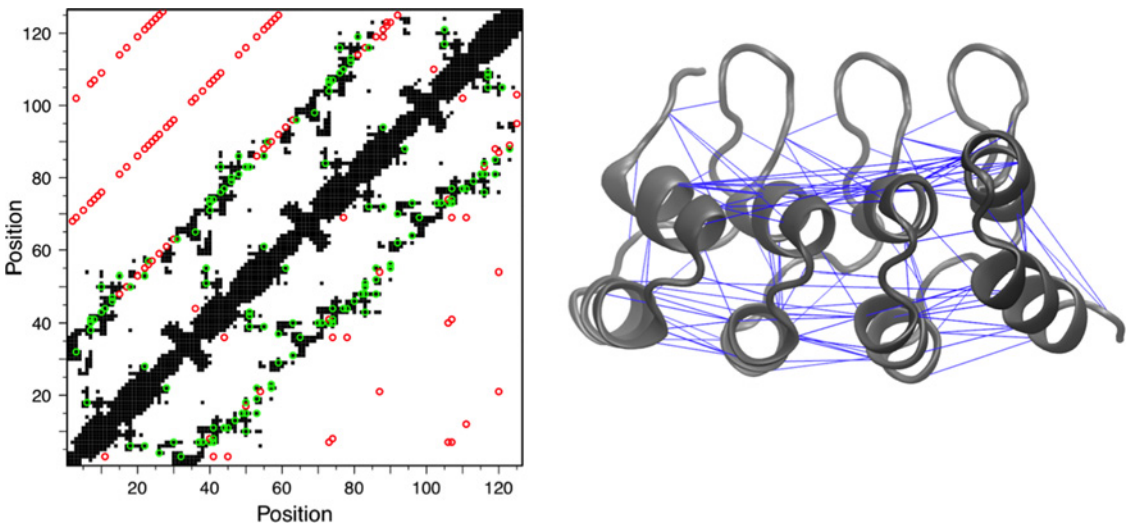
The different stabilities of these sub-domains hint at unevenness in the energy landscape that results in a broad ensemble of species of similar free energies [24]. Thus, it is apparent that repeat-arrays of relatively low repeat-number behave as single domains, yet the appearance of folding sub-domains in longer repeat-arrays is not an exception, but may be part of the natural physiological necessities in these systems.

In contrast with natural repeat proteins, the folding of todays' designed proteins does not show the appearance of folding sub-domains. An elegant series of experiments on the folding of designed repeat-arrays show that the folding of these typically behave in an all-or-none fashion: the longer the repeat-array, the more stable and more co-operative the folding transition is, namely the repeat-arrays behave as a single domains [25,26]. Theoretical and computational studies predict that these highly symmetric proteins should fold through parallel pathways on funnelled energy landscapes [27,28] and these have been recently characterized experimentally for the ankyrin [25] and the tetratricopeptide repeat (TPR) [29] protein families. The folding landscapes of repeat proteins can be manipulated, for example by the addition of terminal stabilizing repeats. It has been shown that these modifications can shift the transition state ensemble towards the stabilized regions, rerouting the folding of the whole repeat array [30]. Local energetics are thus crucial in determining the routes and traps that are actually followed during folding [31].

The inherent symmetries of repeat-proteins suggest that the overall folding properties of a 'domain' or its separation into 'sub-domains' (the stability and co-operativity of the array) may be derived from a microscopic description of the energy balance within each folding element and its interactions with it nearest neighbours [32]. Because of the delicate energetic balance in each subunit, subtle variations in the interactions within and between the repeats can give the impression of major changes in the folding landscape [33]. Such variations may 'decouple' the folding of the elements and partially folded species become populated, defining folding sub-domains. Sufficient information about the population of these states can yield quantitative models of the energetic distribution along the protein [5,34]. The traps that give rise to populated intermediates may arise from the mere topology of the protein, i.e. discontinuous packings, insertions, deletions, long loops. Alternatively, the population of intermediates can arise as an effect of local energetic frustration, regions of the protein that are found in energetic conflict upon folding [35]. Figure 2 shows the local frustration patterns of two example ankyrin repeat proteins, a designed and a natural protein. For both, the ends of the repeat arrays are enriched in frustrated interactions. The designed protein is strongly cross-linked by a web of minimally frustrated interactions, unlike the natural counterpart that shows other patches of highly frustrated interactions, corresponding to the natural binding site [36]. Since these proteins are quasi-1D, a

Figure 3 | Direct Information analysis for the ANK family

At the left panel the contact map of 4ANK is shown, in black the pair of residues in structural contact. Upper triangle the DI hits are shown with circles, coloured in green when they match a contact and in red when they do not. Spurious correlations appear at a distance corresponding with one repeat-unit. The lower triangle shows the DI id hits with the same colour code. When correcting for translation symmetry, these spurious correlations are diminished and real co-evolutionary signals can be identified. At the right panel, the main co-evolutionary pairs of residues are shown in blue as contacts superimposed to a representative crystal structure.



pattern of the population of folding routes can be obtained by visualizing the 'relative foldability' of fragments (Figure 2) [37]. For this, the tertiary energy of the AWSEM (associated memory, water mediated, structure and energy model) energy function [38] of every possible continuous fragment is computed and compared with the energy of decoys, i.e. every other fragment of the same length. This hierarchical procedure based on the calculations of the relative foldability is analogous to a previously described method [39] and captures qualitative aspects of the overall folding landscape. It is apparent that both proteins can have complex folding patterns with multiple routes, for example the designed one is expected to populate parallel routes nucleated at the central repeats (Figure 2, left). In contrast, the foldability of the natural IκBα protein appears more polarized from N- to C-termini, with the very N-terminal repeat region folding last. Maybe it is no co-incidence that this protein was shown to fold in a polarized manner, only consolidating folding upon binding [40]. Thus, the traps that appear in folding IκBα must have a contribution from local topological effects, as recent simulations suggest, where disorder in some repeat regions initiates a domino-like effect partially destabilizing neighbouring regions, in effect showing symmetry-breaking at the level of primary structure [41].

Towards the search of an evolutionary energy function

We have reviewed here that the appearance of folding sub-domains in natural repeat-proteins is mainly dependent on local energetics. Since there is still no efficient way of accurately deconvoluting the energetic origins from first principles, an alternative is to infer energetic constraints from the variations observed in natural sequences. Amino acids that are in spatial proximity in the mean conformational ensemble are expected to co-vary on evolutionary timescales, as their energy contributions to fold stabilization are often localized to groups of residues. The maximum entropy principle proposes a scheme for approaching the problem of extracting essential pair couplings from multiple sequence alignments of families of homologous proteins [42–44]. Since the evolutionary record is inevitably incomplete, the sequences we find today constitute a biased sample of the possible outcomes; therefore, any search for the underlying constraints must take into account contingent factors that may confound the observed correlations. Indirect interactions may generate strong correlations; therefore, disentangling direct from indirect contributions is a fundamental step towards inferring the energetics underlying the observed couplings [43]. The direct interactions can be inferred using the heuristic of 'direct information (DI)' [43]. The mean structure of several globular protein domains can be reasonably well-predicted from the statistical analysis of variations in large sets of sequences [45,46]. Specific interactions between domains can be characterized and good approximations to the interaction energetics can be obtained [44,47–49].

The application of direct coupling analysis to repeat-proteins suffers from translational symmetry of their

sequences and thus spurious correlations appear in the calculation of the DI, precisely at the length of the repeat-unit (Figure 3, upper diagonal). In order to unbias this effect, we recently developed a method that equalized the weights of the sequences according to the identity of repeats in the same protein [50]. This procedure allows for the observation of true co-evolutionary signals from which native-contacts can be identified (Figure 3, lower diagonal). The Potts model underlying these calculations yields energetic parameters for the interactions at the single-residue level. The convolution of these energetic parameters with physically based force fields can yield profound insights into the evolution of natural proteins, such as the characteristic effective selection temperature at which foldable sequences can be selected in sequence space [51].

Concluding remarks

Repeat proteins are believed to be ancient folds. Their biological activity is usually attributed to mediating specific protein–protein interactions, crucial steps in modulating the biochemistry of any cell. When studied in detail, the coupling between folding and binding of natural proteins turns out to be intimately related to their biological function. There is a fine balance between the local folding signals that can be at play at the secondary structure and/or the tertiary contact levels, defining the appearance of folding sub-domains. A local effect is felt globally because the near neighbour interactions are extraordinarily relevant in stabilizing the repeats and, unlike most globular domains, weak biases can tip the balance to complete folding. Single substitutions that affect local biases, such as helix propensity, can exert profound effects on the overall folding of these proteins. The larger repeating arrays are more likely to tolerate 'cracks' and the folding at the repeats ends may become anti-correlated [5]. Such sensitivity allows for specific encoding of folding intermediates by means of making few sequence or environmental modifications. Owing to the symmetry of the repeating array, long arrays can be fine-tuned to populate partially folded states and these ensembles can be co-opted in functional mechanisms. Changing the stability of a single repeating element (by post-translational modifications or binding of other macromolecules) may affect the behaviour at a distant site, providing a coupling mechanism that can transmit allosteric signals to long distances within a single repeating geometry.

Funding

This work was supported by the Consejo Nacional de Investigaciones Científicas y Técnicas de Argentina (CONICET); the Agencia Nacional de Promociónn Científica y Tecnológica (ANPCyT) [grant number 2012-01647]; and the European Research Council Starting Grant (ERCStG) [grant number 306312].

References

1 Wetlaufer, D.B. (1973) Nucleation, rapid folding, and globular intrachain regions in proteins. Proc. Natl. Acad. Sci. U.S.A. **70**, 697–701 CrossRef PubMed
2 Nasir, A., Kim, K.M. and Caetano-Anollés, G. (2014) Global patterns of protein domain gain and loss in superkingdoms. PLoS Comput. Biol. **10**, e1003452 CrossRef PubMed
3 Jacob, F. (1977) Evolution and tinkering. Science **196**, 1161–1166 CrossRef PubMed
4 Marcotte, E.M., Pellegrini, M., Yeates, T.O. and Eisenberg, D. (1999) A census of protein repeats. J. Mol. Biol. **293**, 151–160 CrossRef PubMed
5 Ferreiro, D.U., Walczak, A.M., Komives, E.A. and Wolynes, P.G. (2008) The energy landscapes of repeat-containing proteins: topology, cooperativity, and the folding funnels of one-dimensional architectures. PLoS Comput. Biol. **4**, e1000070 CrossRef PubMed
6 Luo, H. and Nijveen, H. (2014) Understanding and identifying amino acid repeats. Brief. Bioinform. **15**, 582–591 CrossRef PubMed
7 Shih, E.S. and Hwang, M.J. (2004) Alternative alignments from comparison of protein structures. Proteins **56**, 519–527 CrossRef PubMed
8 Abraham, A.L., Rocha, E.P. and Pothier, J. (2008) Swelfe: a detector of internal repeats in sequences and structures. Bioinformatics **24**, 1536–1537 CrossRef PubMed
9 Walsh, I., Sirocco, F.G., Minervini, G., Di Domenico, T., Ferrari, C. and Tosatto, S.C. (2012) RAPHAEL: recognition, periodicity and insertion assignment of solenoid protein structures. Bioinformatics **28**, 3257–3264 CrossRef PubMed
10 Di Domenico, T., Potenza, E., Walsh, I., Parra, R.G., Giollo, M., Minervini, G. and Tosatto, S.C. (2014) RepeatsDB: a database of tandem repeat protein structures. Nucleic Acids Res. **42**, D352–D357 CrossRef PubMed
11 Kajava, A.V. (2012) Tandem repeats in proteins: from sequence to structure. J. Struct. Biol. **179**, 279–288 CrossRef PubMed
12 Schaper, E., Kajava, A.V., Hauser, A. and Anisimova, M. (2012) Repeat or not repeat?–statistical validation of tandem repeat prediction in genomic sequences. Nucleic Acids Res. **40**, 10005–10017 CrossRef PubMed
13 Parra, R.G., Espada, R., Sánchez, I.E., Sippl, M.J. and Ferreiro, D.U. (2013) Detecting repetitions and periodicities in proteins by tiling the structural space. J. Phys. Chem. B **117**, 12887–12897 CrossRef PubMed
14 Sippl, M.J. and Wiederstein, M. (2008) A note on difficult structure alignment problems. Bioinformatics **24**, 426–427 CrossRef PubMed
15 Sippl, M.J. (2008) On distance and similarity in fold space. Bioinformatics **24**, 872–873 CrossRef PubMed
16 Tang, K.S., Fersht, A.R. and Itzhaki, L.S. (2003) Sequential unfolding of ankyrin repeats in tumor suppressor p16. Structure **11**, 67–73 CrossRef PubMed
17 Lowe, A.R. and Itzhaki, L.S. (2007) Biophysical characterisation of the small ankyrin repeat protein myotrophin. J. Mol. Biol. **365**, 1245–1255 CrossRef PubMed
18 Zeeb, M., Rösner, H., Zeslawski, W., Canet, D., Holak, T.A. and Balbach, J. (2002) Protein folding and stability of human CDK inhibitor p19 INK4d. J. Mol. Biol. **315**, 447–457 CrossRef PubMed
19 Bradley, C.M. and Barrick, D. (2002) Limits of cooperativity in a structurally modular protein: response of the Notch ankyrin domain to analogous alanine substitutions in each repeat. J. Mol. Biol. **324**, 373–386 CrossRef PubMed
20 Ferreiro, D.U., Cervantes, C.F., Truhlar, S.M., Cho, S.S., Wolynes, P.G. and Komives, E.A. (2007) Stabilizing IκBα by "consensus" design. J. Mol. Biol. **365**, 1201–1216 CrossRef PubMed
21 Ferreiro, D.U. and Komives, E.A. (2010) Molecular mechanisms of system control of NF-κB signaling by IκBα. Biochemistry **49**, 1560–1567 CrossRef PubMed
22 Truhlar, S.M., Mathes, E., Cervantes, C.F., Ghosh, G. and Komives, E.A. (2008) Pre-folding IκBα alters control of NF-κB signaling. J. Mol. Biol. **380**, 67–82 CrossRef PubMed
23 Werbeck, N.D., Rowling, P.J., Chellamuthu, V.R. and Itzhaki, L.S. (2008) Shifting transition states in the unfolding of a large ankyrin repeat protein. Proc. Natl. Acad. Sci. U.S.A. **105**, 9982–9987 CrossRef PubMed
24 Itzhaki, L.S. and Lowe, A.R. (2012) From artificial antibodies to nanosprings: the biophysical properties of repeat proteins. Adv. Exp. Med. Biol. **747**, 153–166 CrossRef PubMed
25 Aksel, T. and Barrick, D. (2014) Direct observation of parallel folding pathways revealed using a symmetric repeat protein system. Biophys. J. **107**, 220–232 CrossRef PubMed

26 Wetzel, S.K., Settanni, G., Kenig, M., Binz, H.K. and Plückthun, A. (2008) Folding and unfolding mechanism of highly stable full-consensus ankyrin repeat proteins. J. Mol. Biol. **376**, 241–257 CrossRef PubMed

27 Ferreiro, D.U., Cho, S.S., Komives, E.A. and Wolynes, P.G. (2005) The energy landscape of modular repeat proteins: topology determines folding mechanism in the ankyrin family. J. Mol. Biol. **354**, 679–692 CrossRef PubMed

28 Hagai, T., Azia, A., Trizac, E. and Levy, Y. (2012) Modulation of folding kinetics of repeat proteins: interplay between intra and interdomain interactions. Biophys. J. **103**, 1555–1565 CrossRef PubMed

29 Javadi, Y. and Main, E.R. (2009) Exploring the folding energy landscape of a series of designed consensus tetratricopeptide repeat proteins. Proc. Natl. Acad. Sci. U. S. A. **106**, 17383–17388 CrossRef PubMed

30 Tripp, K.W. and Barrick, D. (2008) Rerouting the folding pathway of the Notch ankyrin domain by reshaping the energy landscape. J. Am. Chem. Soc. **130**, 5681–5688 CrossRef PubMed

31 Street, T.O. and Barrick, D. (2009) Predicting repeat protein folding kinetics from an experimentally determined folding energy landscape. Protein Sci. **18**, 58–68 PubMed

32 Aksel, T. and Barrick, D. (2009) Analysis of repeat protein folding using nearest neighbor statistical mechanical models. Methods Enzymol. **455**, 95–125 CrossRef PubMed

33 Ferreiro, D.U. and Komives, E.A. (2007) The plastic landscape of repeat proteins. Proc. Natl. Acad. Sci. U.S.A. **104**, 7735–7736 CrossRef PubMed

34 Schafer, N.P., Hoffman, R.M., Burger, A., Craig, P.O., Komives, E.A. and Wolynes, P.G. (2012) Discrete kinetic models from funneled energy landscape simulations. PLoS One **7**, e50635 CrossRef PubMed

35 Ferreiro, D.U., Komives, E.A. and Wolynes, P.G. (2014) Frustration in biomolecules. Q. Rev. Biophys. **47**, 285–363 CrossRef PubMed

36 Ferreiro, D.U., Hegler, J.A., Komives, E.A. and Wolynes, P.G. (2007) Localizing frustration in native proteins and protein assemblies. Proc. Natl. Acad. Sci. U.S.A. **104**, 19819–19824 CrossRef PubMed

37 Panchenko, A.R., Luthey-Schulten, Z. and Wolynes, P.G. (1996) Foldons, protein structural modules, and exons. Proc. Natl. Acad. Sci. U.S.A. **93**, 2008–2013 CrossRef PubMed

38 Schafer, N.P., Kim, B.L., Zheng, W. and Wolynes, P.G. (2014) Learning to fold proteins using energy landscape theory. Isr. J. Chem. **54**, 1311–1337 CrossRef PubMed

39 Tsai, C.J., Maizel, J.V. and Nussinov, R. (2000) Anatomy of protein structures: visualizing how a one-dimensional protein chain folds into a three-dimensional shape. Proc. Natl. Acad. Sci. U.S.A. **97**, 12038–12043 CrossRef PubMed

40 Lamboy, J.A., Kim, H., Dembinski, H., Ha, T. and Komives, E.A. (2013) Single-molecule FRET reveals the native-state dynamics of the IκBα ankyrin repeat domain. J. Mol. Biol. **425**, 2578–2590 CrossRef PubMed

41 Sivanandan, S. and Naganathan, A.N. (2013) A disorder-induced domino-like destabilization mechanism governs the folding and functional dynamics of the repeat protein iκbα. PLoS Comput. Biol. **9**, e1003403 CrossRef PubMed

42 Neher, E. (1994) How frequent are correlated changes in families of protein sequences? Proc. Natl. Acad. Sci. U.S.A. **91**, 98–102 CrossRef PubMed

43 Weigt, M., White, R.A., Szurmant, H., Hoch, J.A. and Hwa, T. (2009) Identification of direct residue contacts in protein–protein interaction by message passing. Proc. Natl. Acad. Sci. U.S.A. **106**, 67–72 CrossRef PubMed

44 Mora, T., Walczak, A.M., Bialek, W. and Callan, C.G. (2010) Maximum entropy models for antibody diversity. Proc. Natl. Acad. Sci. U.S.A. **107**, 5405–5410 CrossRef PubMed

45 Morcos, F, Hwa, T, Onuchic, JN and Weigt, M (2014) Direct coupling analysis for protein contact prediction.. Methods Mol. Biol. **1137**, 55–70 CrossRef PubMed

46 Sułkowska, J.I., Morcos, F., Weigt, M., Hwa, T. and Onuchic, J.N. (2012) Genomics-aided structure prediction. Proc. Natl. Acad. Sci. U.S.A. **109**, 10340–10345 CrossRef PubMed

47 Marks, D.S., Colwell, L.J., Sheridan, R., Hopf, T.A., Pagnani, A., Zecchina, R. and Sander, C. (2011) Protein 3D structure computed from evolutionary sequence variation. PLoS One **6**, e28766 CrossRef PubMed

48 Cheng, R.R., Morcos, F., Levine, H. and Onuchic, J.N. (2014) Toward rationally redesigning bacterial two-component signaling systems using coevolutionary information. Proc. Natl. Acad. Sci. U.S.A. **111**, E563–E571 CrossRef PubMed

49 Lui, S. and Tiana, G. (2013) The network of stabilizing contacts in proteins studied by coevolutionary data. J. Chem. Phys. **139**, 155103 CrossRef PubMed

50 Espada, R., Parra, R.G., Mora, T., Walczak, A.M. and Ferreiro, D.U. (2015) Capturing coevolutionary signals in repeat proteins. BMC Bioinformatics. doi: 10.1186/s12859-015-0648-3.

51 Morcos, F., Schafer, N.P., Cheng, R.R., Onuchic, J.N. and Wolynes, P.G. (2014) Coevolutionary information, protein folding landscapes, and the thermodynamics of natural selection. Proc. Natl. Acad. Sci. U.S.A. **111**, 12408–12413 CrossRef PubMed

Received 21 April 2015
doi:10.1042/BST20150083