Origin of Proteins

International Edition: DOI: 10.1002/anie.201609977
German Edition: DOI: 10.1002/ange.201609977

Functional Proteins from Short Peptides: Dayhoff's Hypothesis Turns 50

*M. Luisa Romero Romero, Avigayel Rabin, and Dan S. Tawfik**

Margaret O. Dayhoff · peptides · proteins · protein evolution · origin of life

1. Introduction

Proteins are unlikely to have emerged as the large, complex globular forms we know today. First, the primordial protein synthesis machinery would inevitably have been inefficient, and the earliest step in protein evolution may even have been the abiotic emergence of short peptides. Second, the de novo emergence of functional proteins through random condensation of amino acids demands an improbably extensive exploration of sequence space. We may assume a minimal set of primordial, abiotic amino acids,[1] even a "binary" system of polar/charged or hydrophobic[2] plus glycine for connecting loops. Nonetheless, the probability of the emergence of a typical protein domain that can carry out enzymatic functions (ca. 100 amino acids) seems dauntingly low ($1/3^{100} \approx 10^{-47}$). Hence, the first functional proteins presumably originated from short polypeptides. However, since short polypeptide segments cannot function in the same way that globular proteins do,[3] supramolecular self-assemblies may have provided peptides with the operative volume and network of interactions that are necessary, particularly for enzymatic functions. The first self-reproducing macromolecules are likely to have been ribonucleic acids (RNAs). Peptides appeared later, probably as cofactors that stabilized RNAs and augmented their catalytic capabilities.[4] Short functional peptides are therefore highly likely to have served as crucial intermediates between a primordial RNA world and the extant protein world.

In fact, the birth of new proteins is not rare. Certain classes of proteins, typically disordered proteins, or ordered all-beta proteins (e.g. β-propellers)[5] are constantly emerging. These new proteins may originate de novo from non-coding sequences or from sequences that encode proteins belonging to a different fold.[6] But how would relatively short peptides give rise to folded, globular, and functional proteins? Fifty years ago, Margaret O. Dayhoff pioneered the key hypothesis regarding the evolution of the first protein forms from short, simple peptides.

2. Margaret Oakley Dayhoff

Born in Philadelphia on 11 March 1925, Dayhoff grew up in New York City (Figure 1). Already in her early academic life, she stood out as an exceptional student, and was awarded a scholarship to New York University. In 1945, she graduated magna cum laude in mathematics and three years later she obtained a Ph.D. in quantum chemistry at Columbia University. Her thesis described pioneering steps in computational chemistry—using punch card machines, she calculated the resonance energies of organic molecules.[7] She further explored problems in theoretical chemistry, first at the Rockefeller Institute for Medical Research and later at the University of Maryland. In the 1960s, she became associate director of the National Biochemical Research Foundation, and in the early 1970s, she became a professor of physiology and biophysics at Georgetown University Medical Center.[8] She was the first woman to hold office in the Biophysical Society, and the first person to become both Secretary and

Figure 1. Margaret Oakley Dayhoff (1925–1983). A picture taken in 1980 (owned by her daughter, Ruth E. Dayhoff, and made available by the National Library of Medicine).

[*] Dr. M. L. Romero Romero, Prof. D. S. Tawfik
Department of Biomolecular Sciences
The Weizmann Institute of Science, Rehovot 76100 (Israel)
E-mail: dan.tawfik@weizmann.ac.il
A. Rabin
Current address: Department of Biological Chemistry
the Alexander Silberman Inst. of Life Sciences
the Hebrew University of Jerusalem
Edmond J. Safra Campus, Jerusalem 91904 (Israel)

Supporting information for this article can be found under: http://dx.doi.org/10.1002/anie.201609977.

President. Dayhoff was also a member of the American Association for the Advancement of Science, and in 1980, she became a councillor for the International Society for the Study of the Origins of Life. She served on the editorial boards of the journals *DNA*, *Journal of Molecular Evolution*, and *Computers in Biology and Medicine*.

Throughout her career, Dayhoff met challenges that working women face to this day, especially in fields largely dominated by men. For example, she had to seek a new position after pregnancy and raising her two daughters.[8a] Nonetheless, her scientific achievements are remarkable and long-lasting. In her honour, the Biophysical Society established the Margaret Oakley Dayhoff Award in 1984, which is given to outstanding women researchers it the early stages of careers in biophysical research.

Dayhoff was fascinated by the origin of life and decided to address it by using computational analysis.[9] She developed thermodynamic models to study the prebiotic planetary atmosphere. Foremost, she established methods to infer evolutionary relationships through the comparison of protein sequences,[10] a ground-breaking contribution that marked the start of bioinformatics as a scientific discipline. In 1965, she initiated the first protein database, collecting the 70 known protein sequences at that time in the Atlas of Protein Sequence and Structure.[11] This was not an easy task because of the possessive attitudes held by many scientists regarding sequence data. Crucially, Dayhoff derived the first probability model of protein evolution that enabled the reconstruction of protein phylogenetic trees, the so-called PAM model (a rearranged acronym for accepted pointed mutation). She thus described the first reconstruction of a phylogenetic tree, and applied such trees to provide the first support of symbiosis as the origins of eukaryotes.[12] She introduced the concept of protein superfamilies and the hierarchical organization of related sequences,[13] thus initiating all modern protein-categorizing databases such as CATH or SCOP. Anyone analysing protein sequence, structure, or evolution is following in Dayhoff's giant footsteps (not least, when using the single-letter amino acids code she began to introduce in 1965).

In 1966, following the publication of the very first sequence of ferredoxin, Dayhoff, together with Richard Eck, published a paper in *Science* noting the internal sequence symmetry in ferredoxin and the profound implications of this symmetry.[14] Ferredoxin is a 55-residue protein containing key inorganic components in the form of iron–sulphur clusters that mediate its electron-transfer functions. Eck and Dayhoff considered ferredoxin to be a living protein fossil, not only because of its rudimentary function, but also because of its internal sequence symmetry. The symmetry they observed in the ferredoxin sequence known at the time (from *Clostridium pasteurianum*; Figure 2a) led them to propose that ferredoxin, and possibly other proteins, evolved through tandem duplications of a shorter protein, which itself may have emerged through the duplication of an even shorter and simpler ancestral peptide. By simpler, they meant a composition of amino acids that could largely form spontaneously, that is, through abiotic chemical and physical processes.

Remarkably, there were only a few ferredoxin sequences available, which were obtained through amino acid sequencing—DNA sequencing would emerge more than 10 years later. Furthermore, the 3D structure of ferredoxin was unknown, although the structure, once solved, revealed a strikingly high degree of symmetry in the tertiary structure with respect to the two halves (Figure 2b).[15] However, numerous scientific hypotheses are proven invalid with time, often, when the second or third model cases are examined. Still, the hundreds of sequences of ferredoxin that followed the few available to Dayhoff and Eck showed the same pattern of internal symmetry.[16] In fact, using the tools pioneered by Dayhoff, consistently higher internal symmetry can be observed in the inferred ancestral ferredoxin sequences compared to the extant ones (Figure 2c). Dayhoff and Eck's hypothesis of a simple (abiotic) composition for the ancestral peptide is also supported by this analysis (inset, Figure 2c). Finally, functional ferredoxin-like proteins have been experimentally obtained through the self-assembly of peptides as short as 16 amino acids.[17] The small size of these so-called maquettes supports the quarter-ferredoxin ancestor suggested by Dayhoff and Eck (although the sequence of a quarter ancestor is not readily inferred; see the Supporting Information).

The article "Evolution of the structure of ferredoxin based on living relics of primitive amino acid sequences"[14] not only

Maria Luisa Romero Romero studied chemistry at the University of Granada where, in 2012, she obtained her Ph.D. in the group of Prof. Sanchez-Ruiz. She continued as a Postdoctoral Researcher in Dan Tawfik's group at the Weizmann Institute of Science. Her current work is focused on the earliest stage of protein evolution.

Avigayel (Hefter) Rabin, obtained her B.Sc. in computer science at Jerusalem College of Technology (Machon Tal), during which period, she also carried out bioinformatics research in Dan Tawfik's group. She is currently pursuing a Ph.D. at the Hebrew university of Jerusalem, in Dr. Sebastian Kadener's lab, and her main interest is the evolution of circRNA biogenesis.

Dan S. Tawfik's research is focused on understanding how proteins, and enzymes in particular, evolve. The group studies the structural, biophysical, genetic, and biochemical aspects of protein evolution, examining a variety of case studies ranging from the emergence of pesticide-degrading enzymes in the last several decades to the evolution of the first protein enzyme more than 3.7 billion years ago.

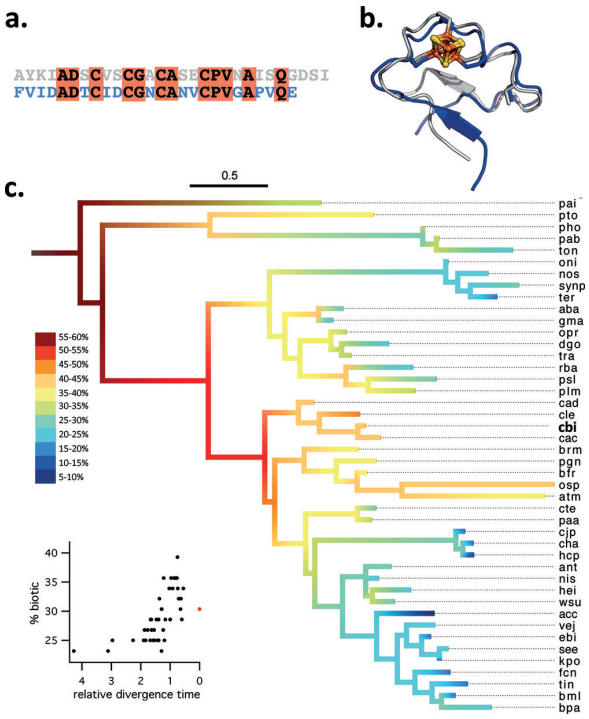

Figure 2. a) As noted by Dayhoff, ferredoxin shows high internal sequence symmetry, thus suggesting its emergence through duplication and fusion. Aligned are the two halves (blue and gray) of the ferredoxin from *Clostridium pasteurianum* that was originally analyzed.[14] b) The structure, published 30 years later, shows near-perfect 2-fold symmetry (PDB ID: 1CLF) between the two halves.[15] The iron–sulfur cluster is shown in orange/yellow. c) A phylogenetic tree and ancestral sequence inference of ferredoxin (also see the Supporting Information). The tree is based on 46 non-redundant ferredoxin sequences from Bacteria and Archaea (*C. pasteurianum* ferredoxin is noted in bold). Lineages are colored according to the internal symmetry in the structure of the inferred ancestral nodes. This phylogenetic analysis indicates that, although derived from essentially one sequence, both of Dayhoff's suppositions were correct: First, sequence identity systematically increases when going back in time, with the most ancient common ancestor we could reliably infer showing 59% internal identity compared to an average of 27% in contemporary sequences. Second, the fraction of abiotically synthesized amino acids is also systematically higher in the inferred ancestor (inset; the contemporary ferredoxin from *Clostridium pasteurianum* (relative divergence time = 0) is indicated as a red dot).

represents Dayhoff and Eck's intellectual vigor, vision, and boldness, but also an era prior to the current one in which data are acquired with alarming ease (and often with diminishing returns with respect to scientific insight; see Ref. [18]). Furthermore, as discussed below, their hypothesis has proven relevant far beyond ferredoxin, laying the groundwork for the following 50 years of studying the evolutionary origins of proteins.

3. Symmetry Dominates the Protein World

Ferredoxin is a relatively small and simple protein. The symmetry primarily relates to its eight cysteine residues that assemble into two identical iron–sulphur clusters. The latter also dictate the fold through covalent sulphur–iron bonds, in contrast to the intricate weak hydrophobic interactions that pack other folds. However, as more protein structures were solved, it appeared that the internal symmetry of ferredoxin is not an exception—symmetry is a recurrent feature observed in many protein folds, including the most ancient ones.[4b,5,19] Examples of proteins with detectable internal sequence symmetry, however, are rare, with only detectable structural symmetry remaining (Figure 3). Symmetry is found not only at the level of tertiary organization, but also in higher assemblies, that is, at the level of quaternary structure.[20] The facile emergence of symmetric proteins also relates to function, for example, the binding of palindromic DNA/RNA sequences[21] or the allosteric regulation of proteins.[22] Today, through systematic mapping, we know that internal symmetry is widespread among protein folds from all of the different categories (all-alpha, all-beta, beta-alpha, etc.). Specialized tools have been developed that detect internal symmetry and its relationship to function at the level of both sequence (e.g., RADAR, HHrepID, TPRred, XTREAM, TRUST) and structure (e.g., COSEC2, HHpred). Experimental work has also provided a better understanding of the evolutionary pathways that underlie the emergence of symmetry and facilitated the de novo design of symmetrical proteins (sometimes called repeat proteins, although the latter also include stand-alone domains that are fused in tandem with no shared packing; see Ref. [23]).

If internal symmetry is considered a fingerprint of the peptide precursors of today's proteins, and these peptides were intimately interacting with RNA, the latter is also expected to exhibit symmetry. In fact, the ribosome core, which is remarkably conserved along all life forms, exhibits pseudo two-fold symmetry.[20b,24] The predominance of internal symmetry within domains of a variety of folds, and the construction of different domains through the duplication and fusion of shorter fragments (e.g., Ref. [23a,b,25]), including fragments composed of simple, abiotic amino acids,[1] suggest that domains as we see them today—compact, independently folding, and stable entities, typically of greater than or equal to 100 amino acids length—do not constitute the minimal evolutionary unit of proteins.

These observations support Dayhoff's hypothesis that the first folded domains arose through the duplication, fusion, and diversification of peptides. Peptide ancestors might also be detected in nonsymmetrical modern proteins, since globular proteins are universally built from fused loop-n-lock units of approximately 30 residues.[26] As further discussed below, the generally presumed peptide ancestors comprise two or three secondary structural elements (β-strands and/or α-helices) connected by either tight turns or longer loops that include functional residues for cofactor binding or catalysis.[27] The emergence of peptides comprising as many as 30 amino acids still imposes a challenge, in terms of both synthesis and sequence-space sampling. However, peptides of approximate-

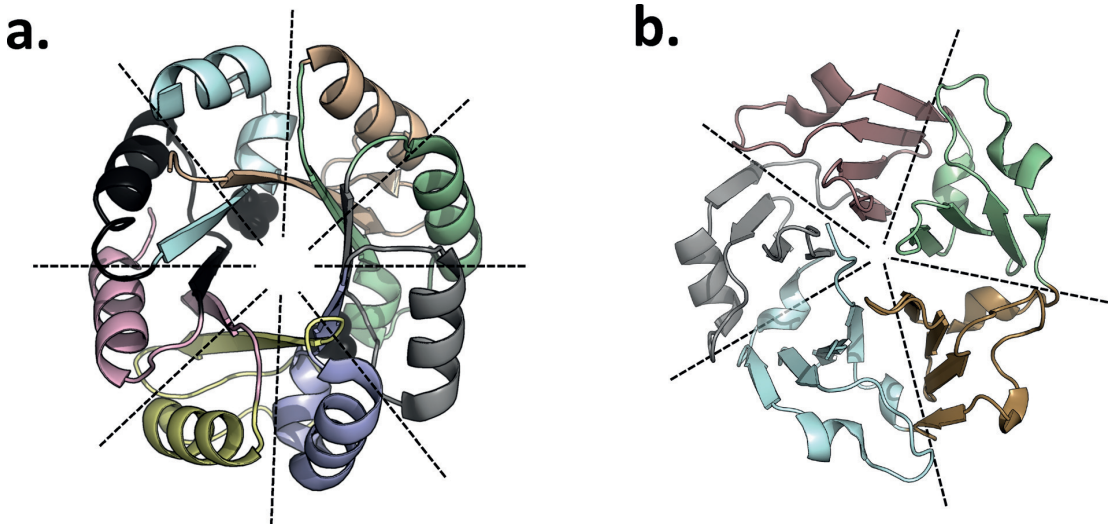

Figure 3. Two examples of protein folds in which internal symmetry is observed at different levels. a) The TIM barrel is an ancient and widely distributed fold in metabolic enzymes.[40] As in many other ancient folds, the repeating structural unit is a (β-strand)-loop-(α-helix), and this fold comprises eight such β-α repeats (as exemplified by PDB ID:1VZW). In contemporary TIM barrels, internal symmetry can be observed only in structure, not in sequence. However, barrels that are derived from the dimerization of a $(\beta\text{-}\alpha)_4$ half-barrel subunit have been engineered.[41] b) β-Propellers also show radial symmetry, with the repeating unit being a 4-stranded β-sheet (the blade). This fold, however, emerged much later in evolution, and β-propellers are still emerging de novo.[5,25c] Consequently, internal sequence identity can be still observed in a small fraction of natural sequences, and ancestral propellers with 100% sequence symmetry have been reconstructed. Shown is a reconstructed ancestor of a contemporary five-bladed lectin propeller (PDB ID: 5C2M).

ly 30 residue are probably the most ancient ancestral elements that can still be reliably detected in modern proteins. Sequences of shorter segments may be derived by analyzing extant proteins, but the shorter the sequence, the lower the reliability of the prediction. For example, despite hundreds of available sequences, the prediction of a quarter-ferredoxin ancestor is ambiguous (see the Supporting Information). Short peptides are also unlikely to be structured, although exceptions, for example, Trp-zippers[28] or β-hairpin peptides,[29] are known. However, cyclization or other configurations may also allow short peptides to adopt defined structures and carry out biological functions.[30] First and foremost, self-assembly presents a means by which relatively short peptides may display function, including catalysis.

4. Self-Assembly is a Facilitating Bridging Step

The likelihood of emergence would be greatly enhanced if the presumed ancestral polypeptide segments, be they approximately 30 amino acids long or even shorter, were functional on their own. Otherwise, emergence would depend not only on such peptides emerging from random sequences, but also on their tandem fusion in a manner that yields a functional protein. However, since short peptides are unlikely to be functional as they are, their self-assembly to yield larger structures comprises a crucial bridging step.

Spontaneous assembly is certainly feasible in the case of ferredoxin, the packing of which is driven by covalent cysteine–iron bonds.[17] However, in nearly all protein domains, and enzymes in particular, globular packing is driven by multiple weak, noncovalent interactions. Is self-assembly applicable here? In such cases, supramolecular structures such as oligomers, swapped domains, or even organized aggregates (amyloid-like) may endow peptides with the necessary minimal size and complexity, the properties needed to be functional. Function-wise, self-assembly also results in multiple functional residues per molecule, thus allowing avidity to enhance weak interactions. Fibrils provide a relatively simple solution in both respects;[31] for example, 7-mer peptides with zinc-ligating residues form amyloid fibrils and may acts as catalysts for aryl ester hydrolysis,[32] or catalyze phenol oxidation by ligating copper.[33] However, the transition from a short fibril-forming peptide into a globular protein demands major structural rearrangements. Other modes of self-assembly, whereby the self-assembled ancestor already adopts the architecture of the mature monomeric protein, are also feasible. For example, a β-sheet motif of 47 amino acids has been shown to oligomerize to form a five-bladed β-propeller. Assembly is a prerequisite for function, since oligomerization yields five binding sites that enable avid binding of glycoproteins despite a very low affinity per site.[25c] Self-assembly is obviously not an absolute prerequisite, since very short peptides are less likely to give soluble oligomeric assemblies, let alone display function. In the modern emergence of proteins, a pentatricopeptide repeat protein that binds RNA was constructed through the tandem fusion of a 35 amino acid helix-turn-helix motif,[25b] although the 35-mer motif is most likely unable to assemble and confer function on its own. The very first step may thus be the tandem amplification of a short motif (e.g., Dayhoff suggested ADSG as the amplified motif of ferredoxin). Tandem repeats of short

sequences are frequent in modern genomes, and are even more likely to have occurred with primordial replication systems. Taken together, it appears that self-assembling peptides would facilitate the emergence of functional protein domains. However, trajectories in which the first functional, selectable protein species is a tandemly fused motif are also feasible.

5. How Did the First Protein(s) Evolve?

Polypeptides of variable lengths, essentially from 7 to 47 amino acids, as exemplified above, can become functional through self-assembly and/or tandem fusion, and this may eventually yield globular, functional proteins. However, none of these examples directly relates to the emergence of the first protein(s). In particular, none of the core enzymatic functions, be they anabolic (synthetic) or catabolic, is represented in these studies. The most rudimentary functions make use of metal ions and/or ribonucleoside cofactors or cosubstrates such as ATP or NAD/H that probably emerged within the RNA world.[34] Accordingly, the most ancient peptide motifs mediate the binding of such cofactors and are omnipresent in the contemporary protein world. Systematic analyses have provided catalogues of ancient motifs, and accordingly, of the most ancient enzyme folds (for recent studies, see Refs. [27,35]). The most probable cofactor-binding peptide motifs that repetitively emerge in all analyses are: 1) the P-loop motif that comprises a β-strand and α-helix connected by a Gly-rich phosphate-binding loop (Walker A motif) that typically binds the transferred phosphate moiety of ATP (see also Ref. [34,36]), and 2) The Rossmann motif, which comprises a β-α-β unit and contains a Gly-rich phosphate-binding loop (a mirror image of the P-loop) and an Asp or Glu residue at the tip of the second β-strand; the latter forms a bidentate interaction with the hydroxy groups of the ribose moiety of various cofactors (NAD, FAD, or SAM). These motifs occur in the most abundant and functionally diverse enzyme folds, which all comprise β-α repeat proteins, including the Rossmann fold.[37] However, there is currently no evidence for the emergence of a rudimentary protein, let alone a functional enzyme, through the duplication and fusion of β-α and/or β-α-β segments containing functional motifs, and in particular motifs such as the P-loop or Rossmann cofactor-binding motifs.

And thus, 50 years after Dayhoff's seminal paper, her hypothesis has been reinforced and validated, and now represents a general and proven model for the emergence of large, globular, and functional proteins from relatively short, simple peptides. However, the fundamental question underlining Dayhoff's hypothesis, namely, how did the first protein(s) emerge, still stands. Nonetheless, the tools and hypotheses pioneered by Dayhoff, and the fact that, by now, more than 65 million protein sequences and 12000 structures are known (as opposed to 70 and 3, respectively, in 1966)[38] enables those who follow in Dayhoff's footsteps to provide new and deeper insight. These tools and databases have already enabled the identification of the most ancient motifs: 25–35 amino acid segments that include functional elements such as cofactor binding and/or catalytic residues.[27,35,36,39] However, whether these motifs display biochemical function as they are, or only upon assembly, is yet to be examined. Reconstructing simple proteins that comprise tandem repeats of these motifs is also an unmet challenge. The uncertainties associated with inferring ancestral sequences after more than 3.7 billions years of diversification are vast. However, computational protein design is a powerful tool that may enable bridging of the inference gap and thus the generation of simple prototypes of primordial proteins along the lines envisaged by Dayhoff.

Acknowledgements

Financial support by the Israel Science Foundation grant 980/14 is gratefully acknowledged. MLRR was supported by the Ramón Areces Foundation. DST is the Nella and Leon Benoziyo Professor of Biochemistry. We thank Liam M. Longo for insightful discussions, his comments on this manuscript, and analysis of the fraction of biotic amino acids (Figure 2c, inset).

How to cite: *Angew. Chem. Int. Ed.* **2016**, *55*, 15966–15971
Angew. Chem. **2016**, *128*, 16198–16203

[1] L. M. Longo, J. Lee, M. Blaber, *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 2135–2139.
[2] S. Kamtekar, J. M. Schiffer, H. Xiong, J. M. Babik, M. H. Hecht, *Science* **1993**, *262*, 1680–1685.
[3] D. R. Corey, M. A. Phillips, *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 4106–4109.
[4] a) W. Gilbert, *Nature* **1986**, *319*, 618; b) J. Söding, A. N. Lupas, *Bioessays* **2003**, *25*, 837–846.
[5] K. O. Kopec, A. N. Lupas, *PLoS One* **2013**, *8*, e77074.
[6] S. Chen, B. H. Krinsky, M. Long, *Nat. Rev. Genet.* **2013**, *14*, 645–660.
[7] M. B. Oakley, G. E. Kimball, *J. Chem. Phys.* **1949**, *17*, 706–717.
[8] a) L. Hunt, *Bull. Math. Biol.* **1984**, *46*, 467–472; b) B. J. Strasser in *eLS*, Wiley, Chichester, **2012**.
[9] H. M. Martinez, *Bull. Math. Biol.* **1984**, *46*, 461–465.
[10] a) M. O. Dayhoff, *J. Theor. Biol.* **1965**, *8*, 97; b) M. O. Dayhoff, *Precambrian Res.* **1983**, *20*, 299–318; c) M. O. Dayhoff, W. C. Barker, L. T. Hunt, *Methods Enzymol.* **1983**, *91*, 524–545.
[11] M. O. Dayhoff, N. B. R. Foundation, *Atlas of Protein Sequence and Structure*, National Biomedical Research Foundation, Washington, DC, **1979**.
[12] R. M. Schwartz, M. O. Dayhoff, *Science* **1978**, *199*, 395–403.
[13] M. O. Dayhoff, P. J. Mclaughlin, W. C. Barker, L. T. Hunt, *Naturwissenschaften* **1975**, *62*, 154–161.
[14] R. V. Eck, M. O. Dayhoff, *Science* **1966**, *152*, 363.
[15] I. Bertini, A. Donaire, B. A. Feinberg, C. Luchinat, M. Piccioli, H. P. Yuan, *Eur. J. Biochem.* **1995**, *232*, 192–205.
[16] a) E. Otaka, T. Ooi, *J. Mol. Evol.* **1989**, *29*, 246–254; b) B. K. Davis, *Prog. Biophys. Mol. Biol.* **2002**, *79*, 77–133; c) J. Meyer, *J. Biol. Inorg. Chem.* **2008**, *13*, 157–170.
[17] a) M. L. Antonkine, M. S. Koay, B. Epel, C. Breitenstein, O. Gopta, W. Gartner, E. Bill, W. Lubitz, *Biochim. Biophys. Acta Bioenerg.* **2009**, *1787*, 995–1008; b) B. R. Gibney, S. E. Mulholland, F. Rabanal, P. L. Dutton, *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 15041–15046.
[18] a) D. Graur, Y. Zheng, N. Price, R. B. Azevedo, R. A. Zufall, E. Elhaik, *Genome Biol. Evol.* **2013**, *5*, 578–590; b) G. A. Petsko, *EMBO Rep.* **2009**, *10*, 1282.

[19] a) S. Balaji, *Curr. Opin. Struct. Biol.* **2015**, *32*, 156 – 166; b) C. A. Orengo, D. T. Jones, J. M. Thornton, *Nature* **1994**, *372*, 631 – 634; c) W. R. Taylor, A. S. Aszódi, *Protein geometry, classification, topology and symmetry: a computational analysis of structure*, Institute of Physics, Bristol, **2005**.
[20] a) G. Villar, A. W. Wilber, A. J. Williamson, P. Thiara, J. P. Doye, A. A. Louis, M. N. Jochum, A. C. Lewis, E. D. Levy, *Phys. Rev. Lett.* **2009**, *102*, 118106; b) C. P. Jones, A. R. Ferre-D'Amare, *Trends Biochem. Sci.* **2015**, *40*, 211 – 220.
[21] Z. S. Juo, T. K. Chiu, P. M. Leiberman, I. Baikalov, A. J. Berk, R. E. Dickerson, *J. Mol. Biol.* **1996**, *261*, 239 – 254.
[22] J. Monod, J. Wyman, J. P. Changeux, *J. Mol. Biol.* **1965**, *12*, 88 – 118.
[23] a) A. Broom, A. C. Doxey, Y. D. Lobsanov, L. G. Berthin, D. R. Rose, P. L. Howell, B. J. McConkey, E. M. Meiering, *Structure* **2012**, *20*, 161 – 171; b) M. Blaber, J. Lee, *Curr. Opin. Struct. Biol.* **2012**, *22*, 442 – 450; c) T. J. Brunette, F. Parmeggiani, P. S. Huang, G. Bhabha, D. C. Ekiert, S. E. Tsutakawa, G. L. Hura, J. A. Tainer, D. Baker, *Nature* **2015**, *528*, 580; d) A. R. Voet, H. Noguchi, C. Addy, D. Simoncini, D. Terada, S. Unzai, S. Y. Park, K. Y. Zhang, J. R. Tame, *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 15102 – 15107.
[24] C. Davidovich, M. Belousoff, I. Wekselman, T. Shapira, M. Krupkin, E. Zimmerman, A. Bashan, A. Yonath, *Isr. J. Chem.* **2010**, *50*, 29 – 35.
[25] a) J. Lee, M. Blaber, *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 126 – 130; b) S. Coquille, A. Filipovska, T. Chia, L. Rajappa, J. P. Lingford, M. F. Razif, S. Thore, O. Rackham, *Nat. Commun.* **2014**, *5*, 5729; c) R. G. Smock, I. Yadid, O. Dym, J. Clarke, D. S. Tawfik, *Cell* **2016**, *164*, 476 – 486.
[26] I. N. Berezovsky, A. Y. Grosberg, E. N. Trifonov, *FEBS Lett.* **2000**, *466*, 283 – 286.
[27] A. Goncearenco, I. N. Berezovsky, *Phys. Biol.* **2015**, *12*, 045002.
[28] A. G. Cochran, N. J. Skelton, M. A. Starovasnik, *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 5578 – 5583.
[29] a) F. J. Blanco, G. Rivas, L. Serrano, *Nat. Struct. Biol.* **1994**, *1*, 584 – 590; b) V. Munoz, P. A. Thompson, J. Hofrichter, W. A. Eaton, *Nature* **1997**, *390*, 196 – 199.
[30] T. Machida, S. Dutt, N. Winssinger, *Angew. Chem. Int. Ed.* **2016**, *55*, 8595 – 8598.
[31] a) M. P. Friedmann, V. Torbeev, V. Zelenay, A. Sobol, J. Greenwald, R. Riek, *PLoS One* **2015**, *10*, e0143948; b) C. Rufo, Y. Moroz, O. Moroz, O. Makhlynets, P. Gosavi, J. Stohr, T. Smith, X. Z. Hu, W. DeGrado, I. Korendovych, *Protein Sci.* **2015**, *24*, 188.
[32] C. M. Rufo, Y. S. Moroz, O. V. Moroz, J. Stohr, T. A. Smith, X. Z. Hu, W. F. DeGrado, I. V. Korendovych, *Nat. Chem.* **2014**, *6*, 303 – 309.
[33] O. V. Makhlynets, P. M. Gosavi, I. V. Korendovych, *Angew. Chem. Int. Ed.* **2016**, *55*, 9017 – 9020.
[34] G. Caetano-Anollés, M. W. Wang, D. Caetano-Anollés, J. E. Mittenthal, *Biochem. J.* **2009**, *417*, 621 – 637.
[35] a) V. Alva, J. Soding, A. N. Lupas, *Elife* **2015**, *4*, e09410; b) P. Laurino, A. Toth-Petroczy, R. Meana-Paneda, W. Lin, D. G. Truhlar, D. S. Tawfik, *PLoS Biol.* **2016**, *14*, e1002396.
[36] A. N. Lupas, C. P. Ponting, R. B. Russell, *J. Struct. Biol.* **2001**, *134*, 191 – 203.
[37] A. Tóth-Petróczy, D. S. Tawfik, *Curr. Opin. Struct. Biol.* **2014**, *26*, 131 – 138.
[38] A. R. Fersht, *Nat. Rev. Mol. Cell Biol.* **2008**, *9*, 650 – 654.
[39] A. Goncearenco, I. N. Berezovsky, *Bioinformatics* **2010**, *26*, i497 – i503.
[40] N. Nagano, C. A. Orengo, J. M. Thornton, *J. Mol. Biol.* **2002**, *321*, 741 – 765.
[41] B. Hocker, J. Claren, R. Sterner, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 16448 – 16453.

Received: October 12, 2016
Published online: November 16, 2016

