

Emergence of symmetric protein architecture from a simple peptide motif: evolutionary models

Michael Blaber, Jihun Lee & Liam Longo

Cellular and Molecular Life Sciences

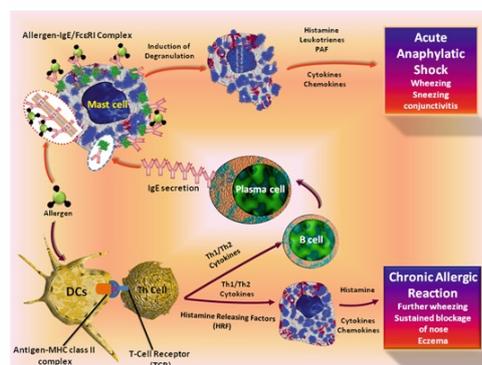
ISSN 1420-682X

Cell. Mol. Life Sci.

DOI 10.1007/s00018-012-1077-3

Cellular and
Molecular
Life Sciences

ONLINE
FIRST



 Springer

 Springer

Your article is protected by copyright and all rights are held exclusively by Springer Basel AG. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your work, please use the accepted author's version for posting to your own website or your institution's repository. You may further deposit the accepted author's version on a funder's repository at a funder's request, provided it is not made publicly available until 12 months after publication.

Emergence of symmetric protein architecture from a simple peptide motif: evolutionary models

Michael Blaber · Jihun Lee · Liam Longo

Received: 17 May 2012/Revised: 22 June 2012/Accepted: 26 June 2012
© Springer Basel AG 2012

Abstract Structural symmetry is observed in the majority of fundamental protein folds and gene duplication and fusion evolutionary processes are postulated to be responsible. However, convergent evolution leading to structural symmetry has also been proposed; additionally, there is debate regarding the extent to which exact primary structure symmetry is compatible with efficient protein folding. Issues of symmetry in protein evolution directly impact strategies for de novo protein design as symmetry can substantially simplify the design process. Additionally, when considering gene duplication and fusion in protein evolution, there are two competing models: “emergent architecture” and “conserved architecture”. Recent experimental work has shed light on both the evolutionary process leading to symmetric protein folds as well as the ability of symmetric primary structure to efficiently fold. Such studies largely support a “conserved architecture” evolutionary model, suggesting that complex protein architecture was an early evolutionary achievement involving oligomerization of smaller polypeptides.

Keywords Protein evolution · Protein symmetry · Protein design · Proteogenesis · Peptide motif · Tandem duplication

M. Blaber (✉) · L. Longo
Department of Biomedical Sciences, College of Medicine,
Florida State University, 1115 West Call St., Tallahassee,
FL 32306-4300, USA
e-mail: michael.blaber@med.fsu.edu

Present Address:

J. Lee
Celltrion Inc., 13-1 Songdo-dong, Yeonsu-gu,
Incheon 406-840, Korea

Symmetry in protein structure

Proteins, the essential workhorse molecules of living organisms, comprise typical proteome sizes of $1\text{--}4 \times 10^4$ unique proteins [1, 2]. However, examination of protein tertiary structure reveals the diverse application of a small number of common structural folds. Notably, the majority of such fundamental protein folds exhibit internal (i.e., rotational, C_n) structural symmetry. Among ten fundamental protein superfolds proposed by Thornton [3], six exhibit structural symmetry. The evolutionary origin of structural symmetry (in carp muscle calcium binding protein) was proposed by McLachlan [4] as having emerged from an ancient homo-oligomeric assembly into a single polypeptide chain via gene duplication and fusion events. Subsequent analyses of other proteins with evidence of primary or tertiary structure symmetry led to an increasing number of similar proposals (most notably by McLachlan) [5–13]. The discovery, for example, of four-helix up-down bundles in both homo-tetramer and homo-dimer forms provided compelling evidence for its evolution by a gene duplication and fusion event [13]. The rotational symmetry, C_n , observed in symmetric protein folds typically varies from $n = 2\text{--}8$ with examples of both even and odd values. A single gene duplication and fusion event can yield structural twofold (C_2) symmetry, the simplest level of symmetry (e.g., type I dockerin [14]). Higher symmetry (e.g., $n = 3\text{--}8$), requires subsequent duplication and fusion events. Odd-number symmetry (e.g., $n = 3, 5, 7$) requires either incomplete (i.e., sub-domain) duplication and fusion, or duplication and fusion followed by a truncation event (which can also lead to structural circular permutation) [15, 16]. Domain repeats are thought to arise via tandem duplication within a gene, where a segment is duplicated and the copy is inserted next to its origin. Although the

exact mechanism of tandem duplication is not fully understood, nonhomologous recombination appears most likely, and tandem duplication has been identified as a major evolutionary force in the emergence of extant protein architecture [17, 18]. Analyses show that proteins with domain repeats are abundant, with $\sim 5\%$ of the proteome of *Archaea* and bacteria containing proteins with domain repeats (typically 2–9 repeats) and $\sim 7\text{--}17\%$ of the proteome of eukaryotes containing proteins with domain repeats (typically 2–20+ repeats) [17]. Duplicated but non-fused genes are even more abundant, with 15–30% of *Archaea* and bacteria genes, and 30–65% of eukaryotic genes, showing evidence of duplication [19].

Symmetry in protein structure is of central interest not only to understanding protein evolution but also in the application of symmetric design principles to de novo protein design. The application of symmetry in de novo protein design substantially reduces the complexity of the design process; however, there remains uncertainty as to the limits of primary structure symmetry that will yield a foldable polypeptide [20–22]. Uncertainty regarding the foldability of symmetric primary structure calls into question the viability of tandem gene duplication in evolutionary processes, since immediately subsequent to gene duplication and fusion exact primary structure symmetry would exist. However, recent experimental work from several groups provides compelling support for gene duplication and fusion in the evolution of symmetric protein architecture, with support for an evolutionary model that is capable of yielding complex protein architecture with simple peptide motifs. Recent experimental work also shows that exact primary structure symmetry is not de facto incompatible with efficient protein folding, and is a critical result for both de novo protein design and models of protein evolution. Such experimental studies have been postulated to probe aspects of protein architecture and evolution prior to the last universal common ancestor (LUCA) of living systems.

Emergent versus conserved architecture evolutionary models

Two distinctly different types of models for the evolution of symmetric protein architecture from a primordial peptide motif, via gene duplication and fusion processes, have been proposed [23–29] and describe either an “emergent architecture” or a “conserved architecture” evolutionary pathway. The most detailed descriptions for the emergent and conserved architecture evolutionary models were proposed simultaneously by Mukhopadhyay [23] and Ponting and Russell [24], respectively, and each utilized the β -trefoil as the example symmetric protein fold. In the emergent architecture model for the β -trefoil fold a simple

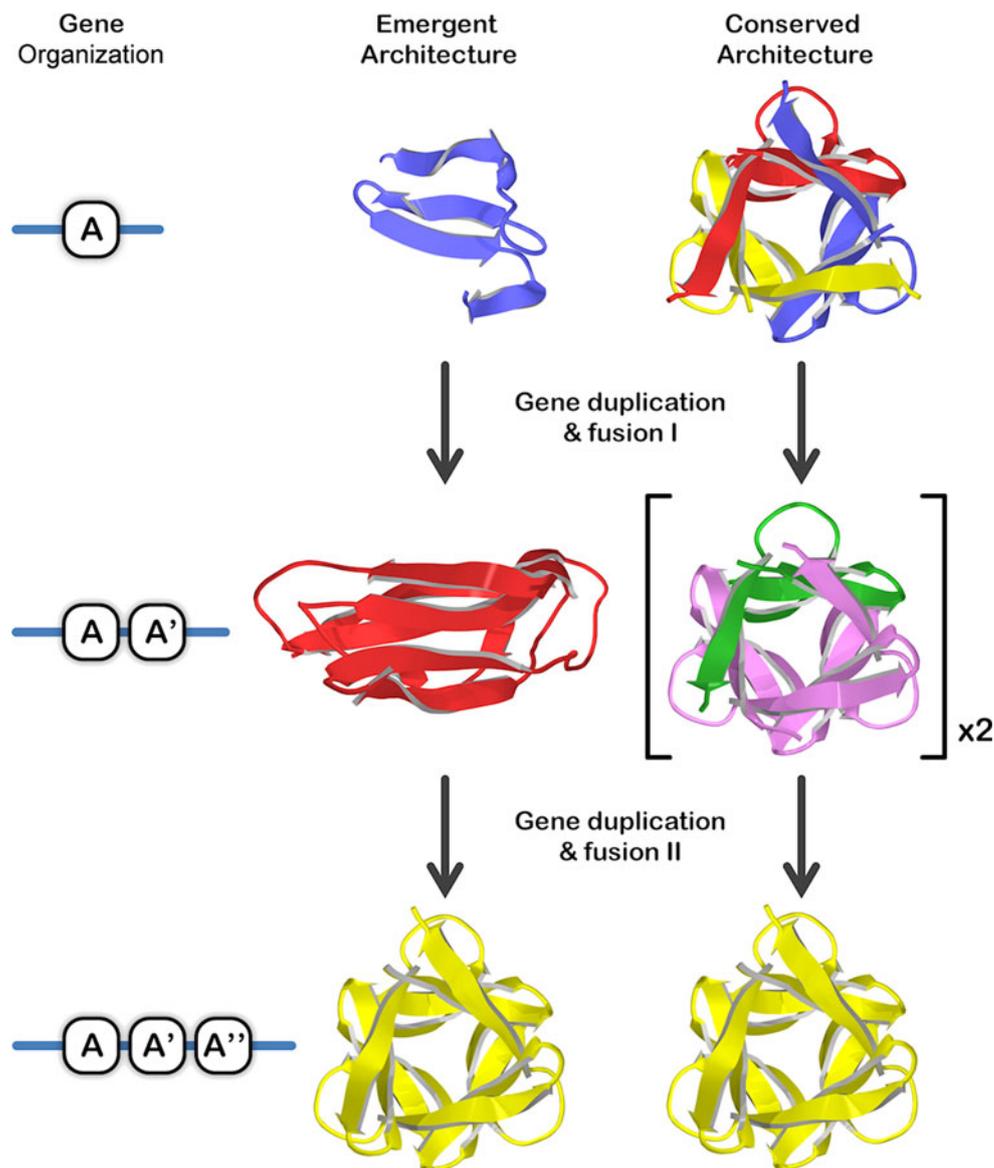
β_4 peptide motif (i.e., comprising two $\beta\beta$ -hairpins [13]) autonomously folded to yield a correspondingly simple architecture, similar to the epidermal growth factor (EGF) fold (Fig. 1). A gene duplication and fusion event yielded a twofold symmetric $(\beta_4)_2$ architecture as an independently folding polypeptide, similar to the ecotin fold. A subsequent gene duplication and fusion event involving a β_4 motif yielded the extant $(\beta_4)_3$ threefold symmetric β -trefoil fold. In contrast, in the conserved architecture model, the β_4 peptide motif was hypothesized not to fold as an independent simple architecture, rather, it oligomerized as a $(\beta_4)_3$ homo-trimer to produce a complete β -trefoil fold (Fig. 1). A gene duplication and fusion event yielded a tandem repeat of this motif, which also folded as a homo-trimer to generate two intact, yet inter-connected (via some form of domain-swapping) β -trefoil folds. A subsequent gene duplication and fusion event involving a β_4 motif (either by partial duplication or duplication followed by truncation) yielded the extant $(\beta_4)_3$ threefold symmetric β -trefoil fold as a single polypeptide. Corresponding structural details would hold for these alternative evolutionary models applied to other symmetric protein architectures such as the β -propeller and $(\beta\alpha)_8$ -barrel symmetric protein folds. Whereas the emergent architecture model implies “simple gene, simple polypeptide motif, simple architecture”, the conserved architecture model implies “simple gene, simple polypeptide motif, complex architecture” (achieved via oligomerization); thus, one primary distinction of the conserved architecture model is that the extant symmetric architecture is a comparatively early evolutionary achievement enabling simple genes to nonetheless code for structurally complex protein folds. Experimental testing of such evolutionary models of symmetric protein architecture is based upon characterization of the folding and structural properties of isolated peptide motifs derived from extant symmetric proteins.

Experimental support for evolutionary models of symmetric protein folds

The β -propeller fold

The β -propeller fold has rotational symmetry typically involving 4–8 repeats (i.e., $n = 4\text{--}8$) of a (β_4) structural motif of ~ 50 amino acids known as a “blade” (Fig. 2, bottom of panel a). In addition to the variability in number of repeating blade motifs, the sequence identity between β -propeller proteins is often quite low, suggesting an ancient protein fold. Fragmentation studies of a five-bladed (C5 rotational symmetry) β -propeller (Tachylectin 2) by Tawfik and coworkers yielded a tandem repeat of the fundamental (β_4) motif that oligomerized as a pentamer to

Fig. 1 Comparison of emergent and conserved architecture evolutionary models (using the threefold symmetric β -trefoil fold as example) [23, 24]. A fundamental difference between these models is the ability of simple peptide motifs to achieve complex (symmetric) architecture (via oligomerization) in the conserved architecture model. *Coloring* is used to indicate individual polypeptide chains. The relative order of the repeated domains A, A', and A'' is potentially variable and dependent upon the particular mechanism of duplication [18]



yield two integral copies of the $(\beta_4)_5$ β -propeller (Fig. 2, panel a) and with two distinct domain-swapped arrangements [26, 30]. Further studies utilizing consensus sequence analysis and random mutagenesis produced a near-symmetric pentameric repeat of a 47 amino acid blade subdomain that was soluble and exhibited lectin affinity (indicating correct folding into a β -propeller) [21].

The $(\beta\alpha)_8$ -barrel (TIM barrel) fold

The $(\beta\alpha)_8$ -barrel (“TIM barrel”) fold has a symmetric architecture with eightfold (C8) rotational symmetry of a repeating $(\beta\alpha)$ structural motif of ~ 35 amino acids comprising a central β -strand, reverse-turn, outer α -helix, and reverse turn [31] (Fig. 2, bottom of panel b). Because of a general lack of sequence homology, the ancestry of this

molecule is difficult to trace by sequence analyses alone and is presumed to be a very ancient fold. The evolutionary pathway of the $(\beta\alpha)_8$ -barrel has been the subject of debate, and arguments have been made in favor of both convergent [32] and divergent evolutionary processes [33]. Kirschner and colleagues demonstrated the ability of phosphoribosylanthranilate isomerase (PRAI; a $(\beta\alpha)_8$ -barrel protein) to efficiently fold with circularly permuted termini [34] including the transposition of the last two $(\beta\alpha)$ motifs (i.e., $(\beta\alpha)_{7-8}$) onto the N terminus. Subsequently, wild-type PRAI was fragmented to yield separate amino terminus $(\beta\alpha)_{1-6}$ and carboxyl-terminus $(\beta\alpha)_{7-8}$ domains, which were capable of stoichiometric assembly to yield an intact functional enzyme [35]. These results raised the prospect of subdomain oligomerization in the evolutionary pathway of the $(\beta\alpha)_8$ barrel.

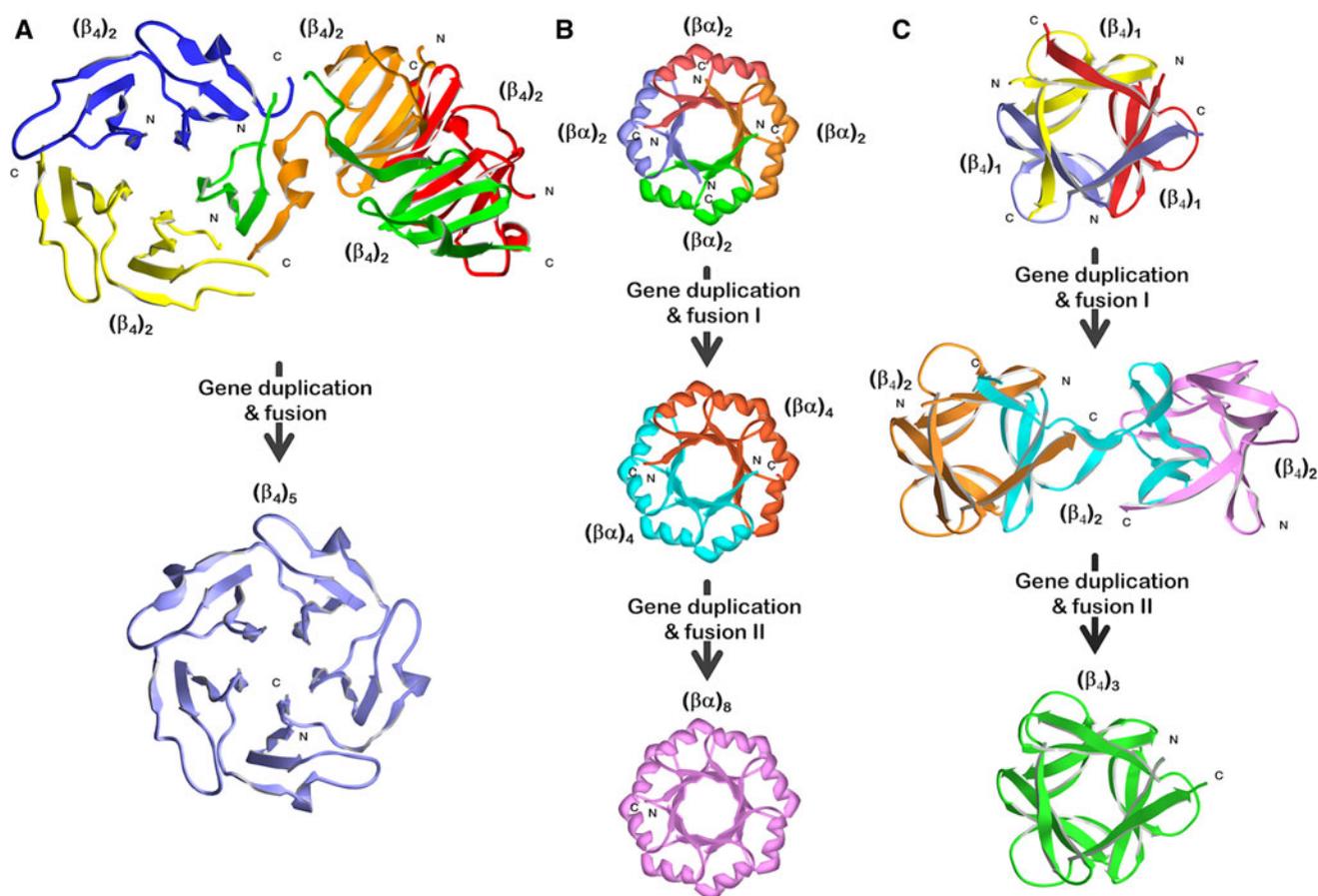


Fig. 2 Experimental studies of the folding and structural properties of simple peptide motifs derived from three common symmetric protein folds and the relationship to plausible gene duplication and fusion events in their evolution. **a** Tandem repeats of a ~ 50 -amino-acid (β_4) motif from a five-bladed β -propeller fold [21, 26, 30];

b tandem and quadruplicate repeats of a ~ 35 -amino-acid ($\beta\alpha$) motif from a $(\beta\alpha)_8$ -barrel fold [28, 35, 38, 59–61]; **c** single, tandem, and triple repeats of a ~ 40 -amino-acid (β_4) motif from a β -trefoil fold [29, 47]. Coloring is used to indicate individual polypeptide chains

Sterner and coworkers in studying the molecular structure of the HisA and HisF $(\beta\alpha)_8$ -barrel proteins of the histidine biosynthetic pathway of *Thermotoga maritima* noted a pronounced twofold structural relationship between reverse turn regions of the N and C terminus half-barrels [11]. This structural similarity led to the proposal of a specific evolutionary pathway for the emergence of the $(\beta\alpha)_8$ -barrel architecture via gene duplication and fusion of an ancient $(\beta\alpha)_4$ -half-barrel. In later work published by this group, it was hypothesized that this putative ancient $(\beta\alpha)_4$ -half-barrel was unlikely to have existed as an independently folding motif (due to exposure of hydrophobic groups); thus, the $(\beta\alpha)_4$ -half-barrel likely existed as a homo-dimer forming an intact $(\beta\alpha)_8$ -barrel [36]. A soluble, stable, and cooperatively folding $(\beta\alpha)_8$ -barrel protein with an almost exact primary structure symmetry was constructed from duplicated C-termini $(\beta\alpha)_4$ -half-barrels of *T. maritima* HisF [37]. Working with *Escherichia coli* *N*-(5'-phosphoribosyl)anthranilate isomerase (TrpF), a $(\beta\alpha)_8$ -barrel protein of the tryptophan biosynthetic pathway, and

evaluating circular permutations of $(\beta\alpha)_4$ -half-barrels definitions, Akanuma and Yamagishi showed that a $(\beta\alpha)_{3-6}$ -half-barrel fragment could both autonomously-fold as a $(\beta\alpha)_4$ -half-barrel monomer and also oligomerize as a homo-dimer [38]. Furthermore, a single polypeptide containing a tandem duplication of the TrpF $(\beta\alpha)_{3-6}$ sequence yielded a stable monomeric $(\beta\alpha)_8$ -barrel structure. These results provide additional clues to the potential evolutionary process, indicating that $(\beta\alpha)_4$ -half-barrels dimerized to produce an intact $(\beta\alpha)_8$ -barrel (although single $(\beta\alpha)_4$ -half-barrels could have existed as independent simpler structural motifs), and that circular permutation may have occurred as part of the duplication and fusion process in the emergence of the $(\beta\alpha)_8$ -barrel.

Sequence analysis of the *T. maritima* HisF protein suggests that its $(\beta\alpha)_2$ -quarter barrel subdomains may have a common evolutionary origin, and that the first $(\beta\alpha)_2$ -quarter barrel (HisF-N1) within this protein is the slowest evolving and most conserved subdomain [28]. Expression and refolding of the HisF-N2 subdomain yielded evidence

for formation of a stable disulfide-linked $(\beta\alpha)_8$ -barrel tetramer. Based upon these results, Sterner and coworkers have proposed an evolutionary pathway for the $(\beta\alpha)_8$ -barrel architecture that involves two sequential gene duplication and fusion events starting from a $(\beta\alpha)_2$ -quarter barrel subdomain, and describing a conserved architecture evolutionary model (Fig. 2, panel b).

The β -trefoil fold

The β -trefoil fold has threefold (C3) rotational symmetry of a repeating (β_4) structural motif of ~ 40 amino acids that forms an overall six-stranded β -barrel capped at one end by three β -hairpins [8, 39, 40] (Fig. 2, bottom of panel c). As with the β -propeller and $(\beta\alpha)_8$ -barrel architectures, the general lack of sequence homology between repeating domains, as well as between different members of this family, suggest it is an ancient fold [41, 42]. The β -trefoil fold is widely held to have evolved via gene duplication and fusion processes; however, within this framework both an emergent architecture [23] and a conserved architecture evolutionary pathway have been proposed [24] (Fig. 1).

Starting with fibroblast growth factor-1 (FGF-1), a β -trefoil protein, Blaber and coworkers introduced cumulative symmetric constraint mutations, ultimately producing a purely threefold symmetric primary structure that efficiently folded into a thermostable β -trefoil architecture [29, 43–47]. Notably, all the intermediary mutations (14 total) in this symmetric design also efficiently folded, demonstrating a plausible pathway through foldable sequence space from the extant (and highly asymmetric) FGF-1 protein to a purely symmetric β -trefoil protein (representative of a putative ancient form). Subsequent fragmentation of this threefold symmetric β -trefoil protein into single (42-amino-acid) and tandem (84-amino-acid) peptide motifs successfully yielded, in both cases, spontaneously folding and thermostable homotrimer oligomeric β -trefoil folds (Fig. 2, panel c). Meiering and coworkers [48] pursued a symmetric β -trefoil protein design starting with ricin and utilizing a consensus sequence computational approach. This effort yielded a soluble and thermostable β -trefoil protein with exact primary structure symmetry (although, subsequent fragmentation did not yield foldable polypeptides). The central hydrophobic core packing group in the β -trefoil fold is formed from a set of 12–18 residues [48–50] and is postulated to be a key determinant of the structural symmetry [42]. Notably, the symmetric solutions for the β -trefoil folds determined independently starting from FGF-1 or ricin have an essentially identical core-packing arrangement, suggesting the possibility of a common ancestor for these β -trefoil proteins.

Informatics studies suggest repeated emergence of β -propeller and β -trefoil protein folds

An informatics study of the evolution of the β -propeller fold [51] concluded that this fold has risen repeatedly throughout evolution by amplification of single blades (and not dimeric or trimeric repeats). Underlying symmetry (e.g., twofold or threefold within six-bladed propellers) is not observed, suggesting that the evolutionary pathway may be distinctly different from that of the $(\beta\alpha)_8$ -barrel (where evidence of an underlying twofold or fourfold symmetry typically exists). This analysis also suggested that the single ancestral blade may likely have had the capacity to fold as an oligomer; although extant examples of oligomeric assemblies yielding a β -propeller fold are currently limited to a single example of a trimer of double-bladed units [52]. This analysis therefore implies a likely evolutionary jump from a multi-bladed oligomeric structure to multiple (4–8) repeat arrangements within a single polypeptide (involving multiple gene duplication and fusion events within an evolutionarily brief period). Assuming the hypothesis that a single ancestral blade had the capacity to oligomerize is correct, the evolutionary pathway of the β -propeller is consistent with a conserved architecture model. Similarly, an analysis of the β -trefoil fold by Meiering and coworkers [48] concluded that the β -trefoil fold has arisen repeatedly throughout evolution (and that this is an ongoing evolutionary process). The mechanism of this emergence appears similar to the β -propeller, in that single (or in some cases double) trefoil-fold subdomains within existing β -trefoil proteins have undergone duplication events spawning new β -trefoil proteins (as concluded from identity analyses between trefoil-folds within a single β -trefoil protein, and between the trefoil-folds of different β -trefoil proteins). This proposed “modular” evolutionary process is also consistent with a conserved architecture model.

Conclusions

Although it cannot be absolutely ruled out that symmetric protein architectures emerged through a convergent evolutionary process (and not by gene duplication and fusion events), the preponderance of experimental data, involving several different symmetric protein folds, are consistent with gene duplication and fusion evolutionary processes. The folding of the tandem $(\beta_4)_2$ motif of the five-bladed β -propeller protein Tachylectin 2 to form two intact five-bladed β -propeller folds supports a conserved architecture model of evolution. Similarly, the folding properties of

$(\beta\alpha)_2$ -quarter barrel and $(\beta\alpha)_4$ -half barrel domains of HisF and TrpF also support a conserved architecture model for evolution (although some evidence exists for soluble, independently folded $(\beta\alpha)_4$ -half barrel folds). The folding properties of the single and tandem repeating (β_4) β -trefoil sub-domains are remarkably consistent with the hypothesized conserved architecture model of Ponting and Russell. Overall, therefore, the above experimental studies, involving three different symmetric protein folds, provide broad support for a conserved architecture evolutionary model. An important hypothesis emerging from these results is that complex protein architecture is possible in primitive organisms having a simple genome and capable of encoding only comparatively small (i.e., 35–50mer) peptides. However, no such examples of oligomeric assemblies of simple peptides yielding β -propeller, $(\beta\alpha)_8$ -barrel or β -trefoil folds (as shown in Fig. 2) have been identified in any extant proteome. Such architecture is therefore postulated to represent an extremely primitive type of protein structure, possibly preceding the LUCA of bacteria and *Archaea* [28, 29].

If β -propeller, $(\beta\alpha)_8$ -barrel and β -trefoil symmetric protein architecture were present at an evolutionarily early time point, then in one sense subsequent tandem duplication played no role in the evolution of these protein folds per se—tandem duplication simply enabled the structure to be contained within a single polypeptide chain. Tandem duplication would, however, have enabled an increase in the information content (i.e., Shannon's entropy [53]) of the architecture; in this sense, the conserved architecture evolutionary model is not associated with conserved entropy. Tandem duplication would subsequently also enable markedly diverse functional evolution since it would permit the emergence of asymmetry (and functional adaptation) not possible in homo-oligomeric (i.e., purely symmetric) assemblies. Furthermore, an intriguing aspect of the conserved architecture evolutionary model for protein folds with odd number symmetry (i.e., C_n , $n = 1, 3, 5, 7$) is the potential for intermediate forms with multiple integral folds (e.g., the dimer architectures observed for the β -trefoil and five-bladed β -propeller folds in Fig. 2). These would not be strictly identical domains as point mutations would be unequally distributed throughout the twofolds (Fig. 3). In such cases, the resulting physical proximity of two related yet distinctly different proteins provides a structural foundation for the emergence of efficient biochemical pathways (i.e., there would be a high local concentration of both folds and a “metabolite” produced by one fold could be efficiently bound yet differentially acted upon by the second-fold) [11]. Thus, intrinsic to the conserved architecture model for odd symmetry protein architecture is an associated mechanism that potentially supports the emergence of efficient biochemical pathways.

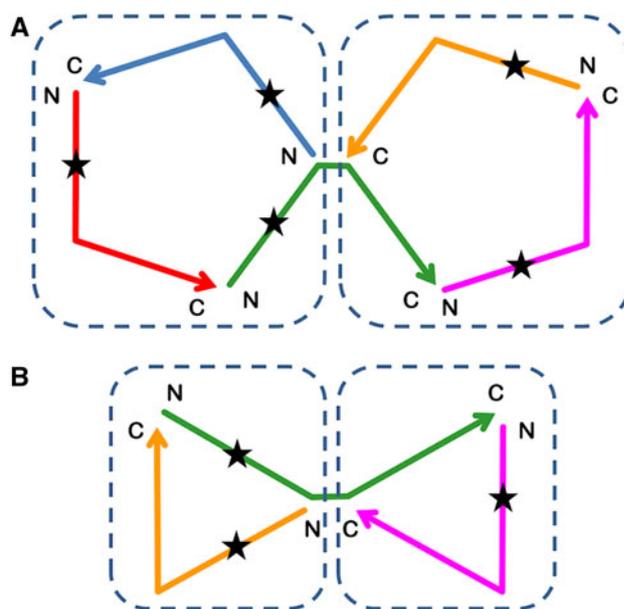


Fig. 3 Intermediates in the conserved architecture evolutionary model [24, 29] of the (five-bladed) β -propeller (top) and β -trefoil folds (bottom) (see Fig. 2a, c). The star indicates the asymmetric effect of a point mutation upon the two integral folds formed by homo-trimer assembly. Coloring is used to indicate individual polypeptide chains

The symmetric protein folds discussed above contain, as fundamental repeating motifs, peptides that comprise 35–50 amino acids and encode (β_4) or $(\beta\alpha)$ structural elements. Such structural motifs can be conceptually divided further (e.g., the (β_4) motif can be separated into two (β_2) β -hairpins). However, the oligomerization of such smaller domains to construct the extant complex symmetric folds is associated with a substantially increased entropic penalty. Thus, an emergent architecture evolutionary pathway may be more likely when generating the 35–50 amino acid motifs from yet simpler polypeptides and represents a more fundamental and major evolutionary step in the development of complex protein structure (Fig. 4).

Tracing the evolutionary pathway of symmetric protein architecture back in time can be considered akin to a “top-down” approach to protein design [54] with the goal of identifying a simple peptide motif having the capacity to spontaneously self-assemble to achieve the desired target architecture. Recent experimental work along these lines has identified a 47-amino-acid peptide motif that can oligomerize (with limited asymmetry) to yield an intact 5-bladed β -propeller [55]; a 74-amino-acid peptide that can oligomerize to yield an intact $(\alpha\beta)_8$ -barrel [28]; and a 42-amino acid peptide that can oligomerize to yield an intact β -trefoil [29, 47]. Typically, such solutions are shown to be thermophile proteins [28, 48, 56]; thus having the capacity to tolerate mutational change in order to

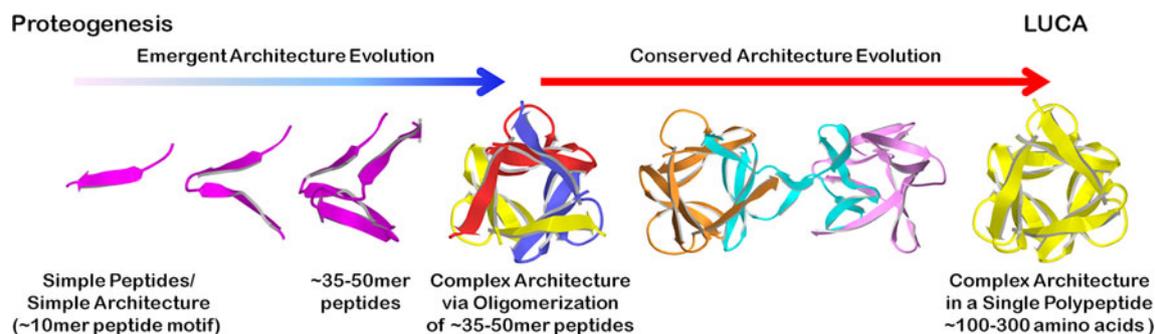


Fig. 4 One possible evolutionary pathway leading from proteogenesis and simple peptide motifs, via gene duplication and fusion, to the emergence of complex architecture present at the time of the LUCA. *Coloring* is used to indicate individual polypeptide chains

introduce novel function (at the expense of stability [57, 58]). Thus, elucidating protein evolutionary pathways can also present new opportunities in de novo protein design.

References

- Harrison PM, Kumar A, Lang N, Snyder M, Gerstein M (2002) A question of size: the eukaryotic proteome and problems in defining it. *Nucleic Acids Res* 30:1083–1090
- Schad E, Tompa P, Hegyi H (2011) The relationship between proteome size, structural disorder and organism complexity. *Genome Biol* 12:R120
- Thornton JM, Orengo CA, Todd AE, Pearl FM (1999) Protein folds, functions and evolution. *J Mol Biol* 293(2):333–342
- McLachlan AD (1972) Gene duplication in carp muscle calcium binding protein. *Nat New Biol* 240(98):83–85
- McLachlan AD (1976) Evidence for gene duplication in collagen. *J Mol Biol* 107(2):159–174
- McLachlan AD, Walker JE (1977) Evolution of serum albumin. *J Mol Biol* 112(4):543–558
- Tang J, James MN, Hsu IN, Jenkins JA, Blundell TL (1978) Structural evidence for evolution of the beta/alpha barrel scaffold by gene duplication and fusion. *Science* 289(5484):1546–1550
- Hocker B, Schmidt S, Sterner R (2002) A common evolutionary origin of two elementary enzyme folds. *FEBS Lett* 510(3):133–135
- Soding J, Lupas AN (2003) More than the sum of their parts: on the evolution of proteins from peptides. *Bioessays* 25:837–846
- Carvalho AL, Dias FM, Prates JA, Nagy T, Gilbert HJ, Davies GJ, Ferreira LM, Romão MJ, Fontes CM (2003) Cellulose assembly revealed by the crystal structure of the cohesin-dockerin complex. *Proc Natl Acad Sci USA* 100:13809–13814
- Jeltsch A (1999) Circular permutations the molecular evolution of DNA methyltransferases. *J Mol Evol* 49:161–164
- Peisajovich SG, Rockah L, Tawfik DS (2006) Evolution of new protein topologies through multistep gene rearrangements. *Nat Genet* 38:168–174
- Bjorklund AK, Ekman D, Elofsson A (2006) Expansion of protein domain repeats. *PLoS One* 2:959–969
- Nacher JC, Hayashida M, Akutsu T (2010) The role of internal duplication in the evolution of multi-domain proteins. *Biosystems* 101:127–135
- Zhang J (2003) Evolution by gene duplication: an update. *Trends Ecol Evol* 18:292–298
- Wright CF, Teichmann SA, Clarke J, Dobson CM (2005) The importance of sequence diversity in the aggregation and evolution of proteins. *Nature* 438:878–881
- Yadid I, Tawfik DS (2011) Functional β -propeller lectins by tandem duplications of repetitive units. *Prot Eng Des Sel* 24(1–2):185–195
- Borgia MB, Borgia A, Best RB, Steward A, Nettels D, Wunderlich BS, Schuler B, Clarke J (2011) Single-molecule fluorescence reveals sequence-specific misfolding in multidomain proteins. *Nature* 474:662–665
- Mukhopadhyay D (2000) The molecular evolutionary history of a winged bean α -chymotrypsin inhibitor and modeling of its mutations through structural analysis. *J Mol Evol* 50:214–223
- Ponting CP, Russell RB (2000) Identification of distant homologues of fibroblast growth factors suggests a common ancestor for all beta-trefoil proteins. *J Mol Biol* 302:1041–1047
- Liu L, Iwata K, Yohda M, Miki K (2002) Structural insight into gene duplication, gene fusion and domain swapping in the evolution of PLP-independent amino acid racemases. *FEBS Lett* 528:114–118
- Yadid I, Tawfik DS (2007) Reconstruction of functional β -propeller lectins via homo-oligomeric assembly of shorter fragments. *J Mol Biol* 365:10–17
- Akanuma S, Matsuba T, Ueno E, Umeda N, Yamagishi A (2010) Mimicking the evolution of a thermally stable monomeric four-helix bundle by fusion of four identical single-helix peptides. *J Biochem* 147:371–379
- Richter M, Bosnali M, Carstensen L, Seitz T, Durchschlag H, Blanquart S, Merkl R, Sterner R (2010) Computational and experimental evidence for the evolution of a $(\beta\alpha)_8$ -barrel protein from an ancestral quarter-barrel stabilized by disulfide bonds. *J Mol Biol* 398:763–773
- Lee J, Blaber M (2011) Experimental support for the evolution of symmetric protein architecture from a simple peptide motif. *Proc Natl Acad Sci USA* 108:126–130
- Yadid I, Kirshenbaum N, Sharon M, Dym O, Tawfik DS (2010) Metamorphic proteins mediate evolutionary transitions of structure. *Proc Natl Acad Sci USA* 107:7287–7292

31. Banner DW, Bloomer A, Petsko GA, Phillips DC, Wilson IA (1976) Atomic coordinates for triose phosphate isomerase from chicken. *Biochem Biophys Res Commun* 72(1):146–155
32. Lesk AM, Branden CL, Chothia C (1989) Structural principles of alpha/beta barrel proteins: the packing of the interior of the sheet. *Proteins* 5(2):139–148
33. Reardon D, Farber GK (1995) The structure and evolution of α/β barrel proteins. *FASEB J* 9:497–503
34. Luger K, Hommel U, Herold M, Hofsteenge J, Kirschner K (1989) Correct folding of circularly permuted variants of a $\beta\alpha$ barrel enzyme in vivo. *Science* 243:206–210
35. Eder J, Kirschner K (1992) Stable substructures of eightfold $\beta\alpha$ -barrel proteins: fragment complementation of phosphoribosylanthranilate isomerase. *Biochemistry* 31:3617–3625
36. Henn-Sax M, Hocker B, Wilmanns M, Sterner R (2001) Divergent evolution of $(\beta\alpha)_8$ -barrel enzymes. *Biol Chem* 382:1315–1320
37. Seitz T, Bocola M, Claren J, Sterner R (2007) Stabilization of a (beta-alpha)₈-barrel protein designed from identical half barrels. *J Mol Biol* 372:114–129
38. Akanuma S, Yamagishi A (2008) Experimental evidence for the existence of a stable half-barrel subdomain in the $(\beta\alpha)_8$ -barrel fold. *J Molec Biol* 382:458–466
39. Blow DM, Janin J, Sweet RM (1974) Mode of action of soybean trypsin inhibitor (Kunitz) as a model for specific protein–protein interactions. *Nature* 249:54–57
40. Sweet RM, Wright HT, Janin J, Chothia CH, Blow DM (1974) Crystal structure of the complex of porcine trypsin with soybean trypsin inhibitor (Kunitz) at 2.6 angstrom resolution. *Biochemistry* 13:4212–4228
41. Rutenber E, Ready M, Robertus JD (1987) Structure and evolution of ricin B chain. *Nature* 326:624–626
42. Feng J, Li M, Huang Y, Xiao Y (2010) Symmetric key structural residues in symmetric proteins with beta-trefoil fold. *PLoS One* 5:e14138. doi:10.1371/journal.pone.0014138
43. Brych SR, Blaber SI, Logan TM, Blaber M (2001) Structure and stability effects of mutations designed to increase the primary sequence symmetry within the core region of a β -trefoil. *Protein Sci* 10:2587–2599
44. Brych SR, Kim J, Logan TM, Blaber M (2003) Accommodation of a highly symmetric core within a symmetric protein superfold. *Protein Sci* 12:2704–2718
45. Brych SR, Dubey VK, Bienkiewicz E, Lee J, Logan TM, Blaber M (2004) Symmetric primary and tertiary structure mutations within a symmetric superfold: a solution, not a constraint, to achieve a foldable polypeptide. *J Mol Biol* 344(3):769–780
46. Dubey VK, Lee J, Blaber M (2005) Redesigning symmetry-related “mini-core” regions of FGF-1 to increase primary structure symmetry: thermodynamic and functional consequences of structural symmetry. *Protein Sci Publ Protein Soc* 14(9):2315–2323
47. Lee J, Blaber SI, Dubey VK, Blaber M (2011) A polypeptide “building block” for the β -trefoil fold identified by “top-down symmetric deconstruction”. *J Mol Biol* 407:744–763
48. Broom A, Doxey AC, Lobsanov YD, Berthin LG, Rose DR, Howell PL, McConkey BJ, Meiering EM (2012) Modular evolution and the origins of symmetry: reconstruction of a three-fold symmetric globular protein. *Structure* 20:1–11
49. Murzin AG, Lesk AM, Chothia C (1992) β -Trefoil fold. Patterns of structure and sequence in the Kunitz inhibitors interleukins-1 β and 1 β and fibroblast growth factors. *J Mol Biol* 223:531–543
50. Blaber M, DiSalvo J, Thomas KA (1996) X-ray crystal structure of human acidic fibroblast growth factor. *Biochemistry* 35:2086–2094
51. Chaudhuri I, Soding J, Lupas AN (2008) Evolution of the β -propeller fold. *Proteins* 71:795–803
52. Kostlánová N, Mitchell EP, Lortat-Jacob H, Oscarson S, Lahmann M, Gilboa-Garber N, Chambat G, Wimmerová M, Imberty A (2005) The fucose-binding lectin from *Ralstonia solanacearum*. A new type of beta-propeller architecture formed by oligomerization and interacting with fucoside, fucosyllactose, and plant xyloglucan. *J Biol Chem* 280:27839–27849
53. Shannon CE (1948) A mathematical theory of communication. *Bell Syst Tech J* 27(3):379–423
54. Blaber M, Lee J (2012) Designing proteins from simple motifs: opportunities in top-down symmetric deconstruction. *Curr Opin Structural Biol.* doi:10.1016/j.sbi.2012.05.008
55. Nikkhah M, Jawad-Alami Z, Demydchuk M, Ribbons D, Paoli M (2006) Engineering of β -propeller protein scaffolds by multiple gene duplication and fusion of an idealized WD repeat. *Biomol Eng* 23:185–194
56. Alsenaidy MA, Wang T, Kim JH, Joshi SB, Lee J, Blaber M, Volkin DB, Middaugh CR (2012) An empirical phase diagram approach to investigate conformational stability of “second-generation” functional mutants of acidic fibroblast growth factor (FGF-1). *Protein Sci Publ Protein Soc* 21(3):418–432
57. Beadle BM, Shoichet BK (2002) Structural basis of stability: function tradeoffs in enzymes. *J Mol Biol* 321:285–296
58. Tokuriki N, Stricher F, Serrano L, Tawfik DS (2008) How protein stability and new functions trade off. *PLoS Comput Biol* 4(2):e1000002. doi:10.1371/journal.pcbi.1000002
59. Hocker B, Beismann-Driemeyer S, Hettwer S, Lustig A, Sterner R (2001) Dissection of a $(\beta\alpha)_8$ -barrel enzyme into two folded halves. *Nat Struct Biol* 8:32–36
60. Hocker B, Claren J, Sterner R (2004) Mimicking enzyme evolution by generating new (beta-alpha)₈-barrels from (beta-alpha)₄-half-barrels. *Proc Nat Acad Sci USA* 101:16448–16453
61. Hocker B, Lochner A, Seitz T, Claren J, Sterner R (2009) High-resolution crystal structure of an artificial (beta-alpha)₈-barrel protein designed from identical half-barrels. *Biochemistry* 48(6):1145–1147. doi:10.1021/bi802125b