PNAS RESEARCH ARTICLE | EVOLUTION OPEN ACCESS

Creative destruction: New protein folds from old

Claudia Alvarez-Carreño[a,b,1], Rohan J. Gupta[b], Anton S. Petrov[a,b,1], and Loren Dean Williams[a,b,1]

Edited by Eugene Koonin, National Institutes of Health, Bethesda, MD; received May 7, 2022; accepted November 7, 2022

Mechanisms of emergence and divergence of protein folds pose central questions in biological sciences. Incremental mutation and stepwise adaptation explain relationships between topologically similar protein folds. However, the universe of folds is diverse and riotous, suggesting more potent and creative forces are at play. Sequence and structure similarity are observed between distinct folds, indicating that proteins with distinct folds may share common ancestry. We found evidence of common ancestry between three distinct β-barrel folds: Scr kinase family homology (SH3), oligonucleotide/oligosaccharide-binding (OB), and cradle loop barrel (CLB). The data suggest a mechanism of fold evolution that interconverts SH3, OB, and CLB. This mechanism, which we call creative destruction, can be generalized to explain many examples of fold evolution including circular permutation. In creative destruction, an open reading frame duplicates or otherwise merges with another to produce a fused polypeptide. A merger forces two ancestral domains into a new sequence and spatial context. The fused polypeptide can explore folding landscapes that are inaccessible to either of the independent ancestral domains. However, the folding landscapes of the fused polypeptide are not fully independent of those of the ancestral domains. Creative destruction is thus partially conservative; a daughter fold inherits some motifs from ancestral folds. After merger and refolding, adaptive processes such as mutation and loss of extraneous segments optimize the new daughter fold. This model has application in disease states characterized by genetic instability. Fused proteins observed in cancer cells are likely to experience remodeled folding landscapes and realize altered folds, conferring new or altered functions.

domain | DNA replication | tandem repeat | translation | OB-fold

Significance

Mechanisms of emergence and early diversification of structured proteins present deep and difficult problems in evolutionary biology. Here we excavate the deepest evolutionary history, found within the translation machinery, which is an ancient molecular fossil and the birthplace of all proteins. We provide evidence supporting common origins of some of the simplest, oldest, and most common protein folds. The data suggest a mechanism, that we call creative destruction, that explains at molecular level how old folds spawn new folds. In this mechanism, new folds emerge from old folds via gene duplication, protein expression, exploration of new folding landscapes, and adaptation. Creative destruction explains the facile emergence of complex from simple architectures in a funneled exploration.

The simplest and most ancient protein folds are built from a small set of supersecondary structures (1). The number of protein folds expanded over time to form the vast universe of protein function in contemporary biology (2–4). Protein folds diversified in a funneled exploration; there is insufficient time and resources in the universe to find novel folds by random searching of sequence space (5).

A fold is a specific arrangement of protein secondary structural elements and backbone topology (6) that incorporates information from various hierarchical levels of protein structure. At the base of the protein structure hierarchy, the polypeptide backbone forms intramolecular hydrogen bonds within α-helices, β-sheets, and loops (7, 8). At the next level of the hierarchy, these secondary structural elements combine to form supersecondary structural elements such as β-α-β or helix-turn-helix (9–12). At even higher levels of the hierarchy, secondary and supersecondary structural elements form globular self-assemblies (2, 13, 14).

The origins of protein folds and the evolutionary mechanisms of fold diversification pose central questions in biological sciences. How did ancient folds arise (1)? What is the role of the ribosomal exit tunnel and chaperones in the early evolution of protein folding (15)? What evolutionary mechanisms led to the diverse set of protein folds in contemporary biological systems? Why did nearly 4 billion years of fold evolution produce less than 2,000 distinct folds? Fold evolution must overcome one or more barriers (15, 16) and is seldom driven by point mutations (17). Numerous small stepwise changes rarely account for conversion of one protein fold to a fundamentally different fold (18–20). Incremental mutation can convert one type of secondary element to another (21) or can cause insertions that decorate a core structure (22).

Here, we describe a general mechanism of creation of daughter folds from ancestral folds. In our model, daughter folds can be different from ancestral folds and at the same time can inherit some elements. Fold innovation in this model starts with changes in gene structure that are known to be frequent. For example, an open reading frame can truncate (23), duplicate (24), or merge with another open reading frame (25). The product of the genetic transformation can be a polypeptide (Fig. 1 *A* and *B*) with a sequence that does not accommodate the ancestral fold. The ancestral fold can be destabilized in the new sequence by the absence of some secondary elements or by physical impingement between

Preprint servers: This manuscript was deposited as a preprint to bioRxiv under a CC-BY-NC-ND 4.0 International license.

Author contributions: C.A.-C., A.S.P., and L.D.W. designed research; C.A.-C. performed research; C.A.-C., R.J.G., and A.S.P. analyzed data; and C.A.-C., A.S.P., and L.D.W. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: ccarreno6@gatech.edu, anton.petrov@biology.gatech.edu, or loren.williams@chemistry.gatech.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2207897119/-/DCSupplemental.

Published December 19, 2022.

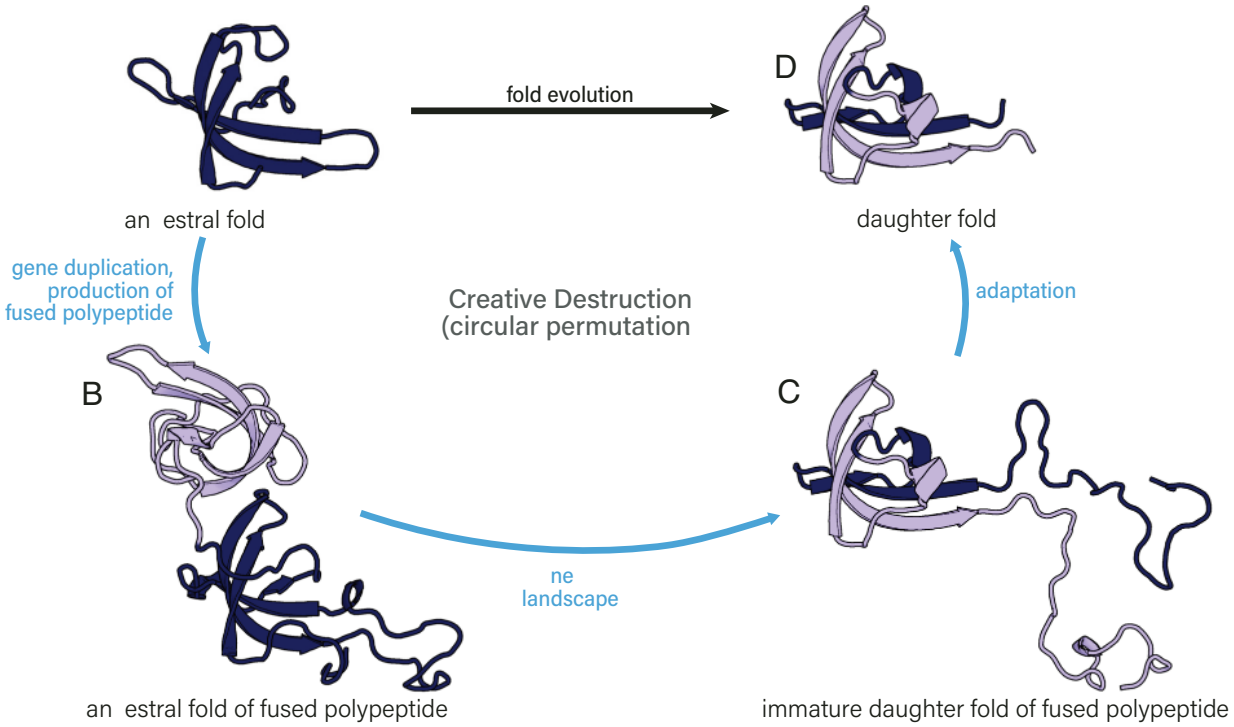

Fig. 1. Creative destruction. *Top*: Creative destruction as a mechanism of circular permutation; genes fuse, an ancestral fold is destroyed, and a daughter fold is created. This figure shows, in three-dimensions, (*A*) an ancestral fold (PDB: 5YYA), (*B*) the notional ancestral folds of the fused polypeptide (PDB: 5YYA), (*C*) the immature daughter fold of the fused polypeptide in which parts of the ancestral folds and some secondary elements have been destroyed and an immature daughter fold has been created (PDB: 7D4A, edited), and (*D*) the mature daughter fold (PDB: 7D4A), which has inherited some but not all supersecondary elements of the ancestors.

ancestral elements (26). The modified polypeptide can explore folding landscapes that are inaccessible to the ancestral sequence(s). Specific stabilizing interactions in the daughter fold might be less probable and would only arise for some sequences.

Our model of protein fold innovation has analogy with Schumpeter's model of economic innovation, called creative destruction (27). In Schumpeter's model, creation of daughter products involves destruction of ancestral products. Daughter products can inherit features of ancestors but can in essence be different from them. The evolution of smart phones is an example of creative destruction (28). Elements of ancestral wired phones, computers, cameras, global positioning, and other technologies merged to create a daughter—the smart phone. The daughter smart phone inherited many features of the ancestors. These features interact in specific ways in the daughter that are not possible in the ancestors. Smart phones created new functional niches that were not accessible to the ancestors. Schumpeter's creative destruction has strong analogy to the processes of fold evolution, illustrating a general and accessible pathway to fold innovation. Creative destruction of protein folds may account for much of observed diversity and affords experimental and computational approaches to exploration of new fold space.

Here we will focus on gene fusions as initiating events in creative destruction of protein folds (Fig. 1*C*). Gene fusion, polypeptide expression, and exploration of new folding landscapes are followed by adaptive processes such as mutation and loss of extraneous segments to optimize the daughter fold. In this model, the folding landscape of a fused daughter polypeptide is not fully independent of those of the ancestral domains. Some secondary and supersecondary structural elements may be retained in the daughter fold. Creative destruction of folds is thus partially conservative in that a daughter fold inherits some motifs from ancestral folds and would also contain new elements (Fig. 1*D*).

Circular permutation is a common and explanatory example of fold evolution by creative destruction (Fig. 1 *A*–*D*). Two proteins related by circular permutation differ by connections between secondary elements, but otherwise appear conserved. Differences in circularly permuted ancestral (Fig. 1*A*) and daughter protein folds (Fig. 1*D*) might be interpreted to suggest that change is accomplished simply by rearrangements of linkages between secondary structural elements. That mechanism, at the polypeptide level, has been observed only in concanavalin A (29). The majority of circularly permuted proteins in nature were generated by evolutionary processes that involve gene duplication (21, 30) and expression of fused polypeptides with remodeled folding landscapes. A fused polypeptide can partially conserve secondary and supersecondary structural elements during folding (Fig. 1*C*). Ancestral folds are partially destroyed during circular permutation.

Here we document creative destruction of the ancestral fold of the zinc-binding ribosomal protein uL33, to give a circularly permuted variant (31). The mechanism entails internal duplication of the uL33 gene (Fig. 2*B*), fold destruction (Fig. 2*C*), fold creation (in a remodeled landscape), and adaptation (Fig. 2*D*). The secondary elements of the two ancestors are semi-conserved in the daughter fold (half of them are conserved and the other half are lost, Fig. 2*D*).

We provide support, on the level of sequence and three-dimensional (3D) structure, for creative destruction of protein folds. Our focus is on some of the oldest, simplest, and most ubiquitous folds in biology. Vestiges of creative destruction are observed by comparisons of three ancient β-barrel folds (Fig. 3): Scr kinase family homology 3 (SH3); oligonucleotide/oligosaccharide-binding (OB) (32); and cradle loop barrel (CLB) (33). We use CLB to refer to the Alanine Racemase C topology of the CLB fold (34). Proteins with SH3, OB, and CLB folds are found in central

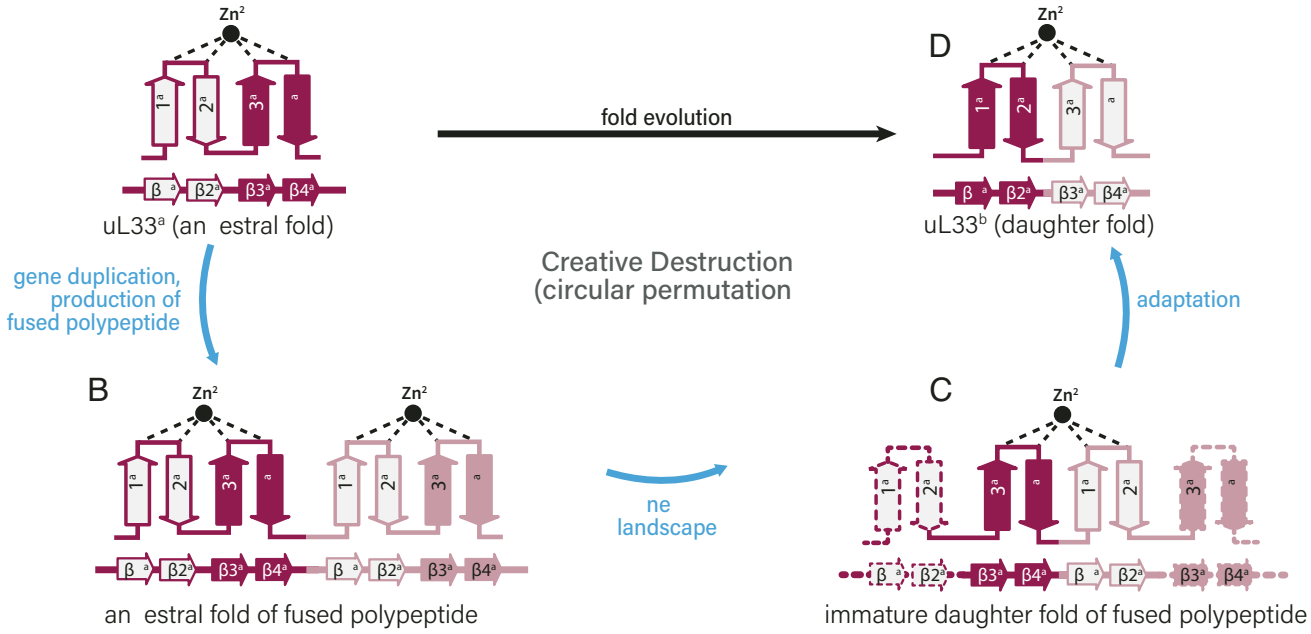

Fig. 2. Topological representation of circular permutation of ribosomal protein uL33 by creative destruction. (*A*) An ancestral uL33 fold, (*B*) the notional ancestral folds of two fused uL33 polypeptides, (*C*) the immature refolded daughterof uL33, and (*D*) mature circularly permuted daughter fold of uL33. A duplication of $\beta1^a\beta2^a\beta3^a\beta4^a$ gives the fused polypeptide $\beta1^a\beta2^a$ **3**^{a} **4**^{a} -- **1**^{a} **2**$^a\beta3^a\beta4^a$, (where -- is a linker). The circles represent zinc ions. Strands are selectively shaded to facilitate tracking through the creative destruction process. The fused polypeptide folds in a new landscape and resolves by adaption. The dashed secondary elements are lost in the mature daughter fold. The ancestral folds of the fused polypeptide are included in the schematic to illustrate destruction of the ancestral folds and inherence of some ancestral secondary motifs.

metabolic processes and throughout the translation system, including in ribosomal proteins, translation factors, and aminoacyl transfer RNA (tRNA) synthetases. Our results suggest that creative destruction is a mechanism of circular permutation and also explains common ancestry of SH3, OB, and CLB folds.

Results

3D structures of proteins are generally more conserved over evolution than their sequences (35, 36). Exceptions to this pattern are sequences that are conserved between proteins with different 3D structures and are called cross-fold sequence similarities. Cross-fold sequence similarities are considered evidence of fold evolution (10, 21, 37, 38). The probability of two unrelated proteins having significantly similar sequences just by chance is extremely low; thus, we use sequence similarity to identify shared ancestry between proteins both globally and locally (39–41).

To maximize the sensitivity of our sequence comparisons, we used protein sequence profiles (37) instead of single sequences (see *Methods* and *SI Appendix*, Fig. S1.) Protein sequence profiles capture the distribution of conserved and non-conserved positions across a multiple sequence alignment (MSA) with high sensitivity, allowing inference of homology with weak signals. The HHalign score is an estimated probability of homology between a query and a template profile (42). Here, we infer homology if a pair of profiles shares HHalign scores higher than 60% and E-values better than 5×10^{-5}. These cut-off values have been used previously for identification of ancestral relationships between folds (43).

We have detected significant sequence similarity within and between SH3, OB, and CLB proteins (Fig. 3). In several instances, global homology between two sequence profiles is inferred by way

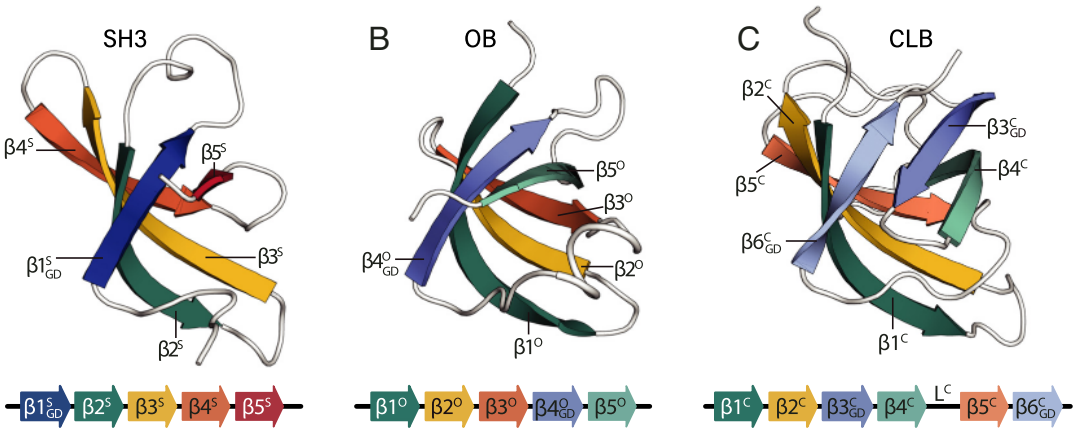

Fig. 3. Structures of SH3, OB, and CLB folds. (*A*) Structure of an SH3 fold (PDB: 1NZ9, chain A). (*B*) Structure of an OB fold (PDB: 2OQK, chain A). (*C*) Structure of a CLB fold (PDB: 4B43, chain A). SH3 and OB are five-stranded β-barrels, and CLB is a six-stranded β-barrel. GD: GD-box motif (described in the *Results*). The color scheme suggests common ancestry.

of homology to a common third sequence (38). This relationship is called transitivity (39). Globally, similar sequence profiles also display global structure similarity, as revealed by template modeling (TM)-scores (44) above 0.5 (*SI Appendix*, Tables S1-S3).

In some cases, distinct folds reveal multiple regions of cross-fold sequence similarity that are permuted relative to each other. The patterns of permuted cross-fold sequence similarities between distinct representatives of SH3, OB, and CLB folds are consistent with, and provide strong support for, creative destruction as a mechanism of fold evolution.

In addition to cross-fold sequence similarity, we analyze structural similarity. Shared structural motifs between SH3, OB and CLB folds include β-sheets, β-hairpins, and GD-box (45, 46). The GD-box motif contains a β-strand connected to a loop by a β-turn and portions of a second non-contiguous β-strand (45). A GD-box is stabilized by a network of hydrogen bonds and is characterized in part by the amino acid sequence ΨxΨxxGρxΨxΨ, where G is glycine, Ψ is aliphatic, ρ is polar, and x is anything.

To understand and explain similarities, we employ a reduced representation of 3D structure. In this representation, 1) α refers to an α-helix, β refers to a β-strand, and L refers to a loop; 2) the number following an α or β indicates the relative position in the sequence, N to C; 3) fold is indicated for each α-helix, β-strand, or loop by O, S, or C as α^O or β^O or L^O (indicates OB fold), α^S or β^S or L^S (indicates SH3), or α^C or β^C or L^C (indicates CLB); 4) L is only specified for loops that display cross-fold sequence similarity to an α-helix or a β-strand; 5) β_{GD} indicates a GD-box; 6) "--" refers to a linker between two domains; and 6) "~" indicates cross-fold sequence similarity between secondary elements. For example, $\beta1^O$~$\beta3^S$ indicates that the first β-strand of an OB fold protein has sequence similarity with the third β-strand of an SH3 fold protein.

SH3 and OB Folds. SH3 and OB folds are both five-stranded β-barrels with two antiparallel β-sheets (32). Although the topologis of SH3 and OB folds differ, the β-strands and some of the connective elements overlay in three dimensions (46–48). SH3 and OB folds share two regions of cross-fold similarity. One region of cross-fold sequence similarity corresponds to $\beta2^S\beta3^S\beta4^S$~$\beta1^O\beta2^O\beta3^O$ (Fig. 4*E*), and the second region of cross-fold sequence similarity corresponds to $\beta1_{GD}{}^S\ \beta2^S$~$\beta4_{GD}{}^O\ \beta5^O$ (Fig. 4*F*).

Cross-fold similarities between ribosomal proteins aS4 and uL2. Cross-fold sequence similarity is seen between ribosomal protein aS4 (an SH3 fold) and ribosomal protein uL2 (an OB fold). These two ribosomal proteins share one region of sequence (HHalign score 64%, E-value 6.5×10^{-5}) and structure similarity in the antiparallel β-sheet $\beta2^S\beta3^S\beta4^S$~$\beta1^O\beta2^O\beta3^O$ (Fig. 4*E* and *SI Appendix*, Table S4).

Cross-fold similarities between proteins RUVBL2 and MscS. Cross-fold sequence similarity is seen in the MscS (SH3) and RUVBL2 (OB) (Fig. 4*F* and *SI Appendix*, Table S4). The region of similarity is $\beta1_{GD}{}^S\beta2^S$~$\beta4_{GD}{}^O\beta5^O$ (HHalign score 85%, E-value 5.8×10^{-6}).

Creative destruction of SH3 generates OB. Homology relationships between SH3 folds and OB folds can be generalized. uL2 and aS4 ($\beta2^S\beta3^S\beta4^S$~$\beta1^O\beta2^O\beta3^O$) as well as RUVBL2 and MscS ($\beta1_{GD}{}^S\beta2^S$~$\beta4_{GD}{}^O\beta5^O$) share cross-fold sequence similarities. Additionally, we identified transitive homology relationships: 1) between the SH3 fold of aS4 and the SH3 fold of MscS (*SI Appendix*, Table S1) and 2) between the OB fold of uL2 and the OB fold of RUVBL2 (*SI Appendix*, Table S2). Thus, accounting for both cross-fold sequence similarity and transitive homology, the general relationship between SH3 and OB folds is as follows: 1) $\beta1^O\beta2^O\beta3^O$ is homologous to $\beta2^S\beta3^S\beta4^S$; 2) $\beta4_{GD}{}^O\beta5^O$ is homologous to $\beta1_{GD}{}^S\beta2^S$; and 3) the GD-box motifs in $\beta4_{GD}{}^O$ and in $\beta1_{GD}{}^S$ are homologous.

The patterns of cross-fold sequence similarity, topology, motif conservation, and 3D structure similarity between SH3 and OB support the creative destruction mechanism of fold evolution (Fig. 4 *A–D*, *SI Appendix*, Fig. S2). A duplication of $\beta1_{GD}{}^1\beta2^S\beta3^S\beta4^S\beta5^S$ yields the fused polypeptide $\beta1_{GD}{}^S\ \mathbf{2}^S\ \mathbf{3}^S\ \mathbf{4}^S\beta5^S$ -- $\mathbf{1}_{GD}{}^S\ \mathbf{2}^S\beta3^S\beta4^S\beta5^S$. In our formalism, the genes fuse, the fused polypeptide collapses, initially to the ancestral folds (Fig. 4*B*), then acquires an immature daughter fold (Fig. 4*C*) and adapts to establish the mature daughter fold (Fig. 4*D*). The collapse of the fused polypeptides to the ancestral folds is a formalism and is shown for illustrative purposes only; the fused polypeptide may collapse directly to the immature daughter fold. Creative destruction is supported by cross-fold sequence similarities $\beta1^O$~$\beta2^S$, $\beta2^O$~$\beta3^S$, $\beta3^O$~$\beta4^S$, $\beta4_{GD}{}^O$~$\beta1_{GD}{}^S$, $\beta5^O$~$\beta2^S$ (Fig. 4 *E* and *F*). The model is supported by structural prediction by AlphaFold (49). The fused polypeptide sequence containing $\beta2^S\beta3^S\beta4^S$--$\beta1_{GD}{}^S\beta2^S$ is folded by AlphaFold to a five-stranded β-barrel fold with OB topology: $\beta1\beta2\beta3\beta4_{GD}\beta5$ (Fig. 5*E* and *SI Appendix*, Fig. S3).

The simpler topology of SH3 compared to OB and the pattern of cross-fold sequence similarity suggests that SH3 is the ancestor of OB. However, the opposite polarity, where internal duplication of an OB results in the emergence of the SH3 fold, cannot be ruled out.

During creative destruction, SH3 is converted to OB via a remodeled folding landscape. $\beta1_{GD}{}^S\beta2^S$ in the SH3 domain and $\beta4_{GD}{}^O\beta5^O$ in the OB domain have common ancestry but are contained within different antiparallel β-sheets. $\beta5^O$ is on the edge of $\beta^O\beta^O\beta^O$, while $\beta2^S$ (similar in sequence to $\beta5^O$) is on the interior of $\beta^S\beta^S\beta^S$. The relative position of $\beta5^O$ is the same as $\beta5^S$ within the linear sequence, but not within the 3D structure. We previously observed that SH3 and OB folds share two conserved supersecondary structural motifs: $\beta1^O\beta2^O\beta3^O$~$\beta2^S\beta3^S\beta4^S$ and $\beta4_{GD}{}^O$~$\beta1_{GD}{}^S$ (46). These conserved structural motifs are circularly permuted in SH3 with respect to OB. However, $\beta5^O$ and $\beta5^S$ are structurally different; $\beta5^O$ has no structural equivalent in SH3, and $\beta5^S$ has no structural equivalent in OB. Here, we identify cross-fold sequence similarity between $\beta5^O$ and $\beta2^S$, which are not structurally conserved. Structural variability of $\beta5^O$ as well as differences in structure between $\beta5^O$ and $\beta2^S$ can be attributed to conformational adaptation, as part of the creative destruction process (see *Discussion*).

OB and CLB Folds. The CLB fold, with six β-strands, is larger and more complex than the OB fold; the number of β-strands and linkages between them differs between these folds. However, commonalities between OB and CLB folds suggest ancestry via creative destruction. Several distinct regions of some OB and CLB fold proteins show cross-fold sequence similarities and partially superimpose in three dimensions (Fig. 5 *E* and *F*). These folds share one region of cross-fold sequence similarity corresponding to $\beta1^O\beta2^O\beta4_{GD}{}^O$ and $\beta1^C\beta2^C\beta3_{GD}{}^C$ and a second region of cross-fold sequence similarity corresponding to $\beta1^O\beta2^O\beta3^O\beta4_{GD}{}^OL^O$ and $\beta4^CL^C\beta5^C\beta6_{GD}{}^C\alpha7^C$.

Cross-fold similarities between uL2 and bIF2. Cross-fold similarities are observed between the OB fold of uL2 and the CLB fold of bacterial translation initiation factor 2 (bIF2). A region of similarity (HHalign score 84%, E-value 3.9×10^{-6}) corresponds to $\beta1^O\beta2^O\beta4_{GD}{}^O$~$\beta1^C\beta2^C\beta3_{GD}{}^C$. $\beta4_{GD}{}^O$~$\beta3_{GD}{}^C$ have a GD-box motif. In three-dimensions, a β-hairpin formed by $\beta1^O\beta2^O$ is similar to the β-hairpin formed by $\beta1^S\beta2^S$. Strand $\beta3^S$ has no equivalent in CLB.

Cross-fold similarities between initiation factors aeIF-1A and aeIF-5B. Cross-fold sequence similarity is detected between the entire OB fold of archaeal translation initiation factor aeIF-1A and the CLB fold in archaeal translation initiation factor aeIF-5B. Cross-

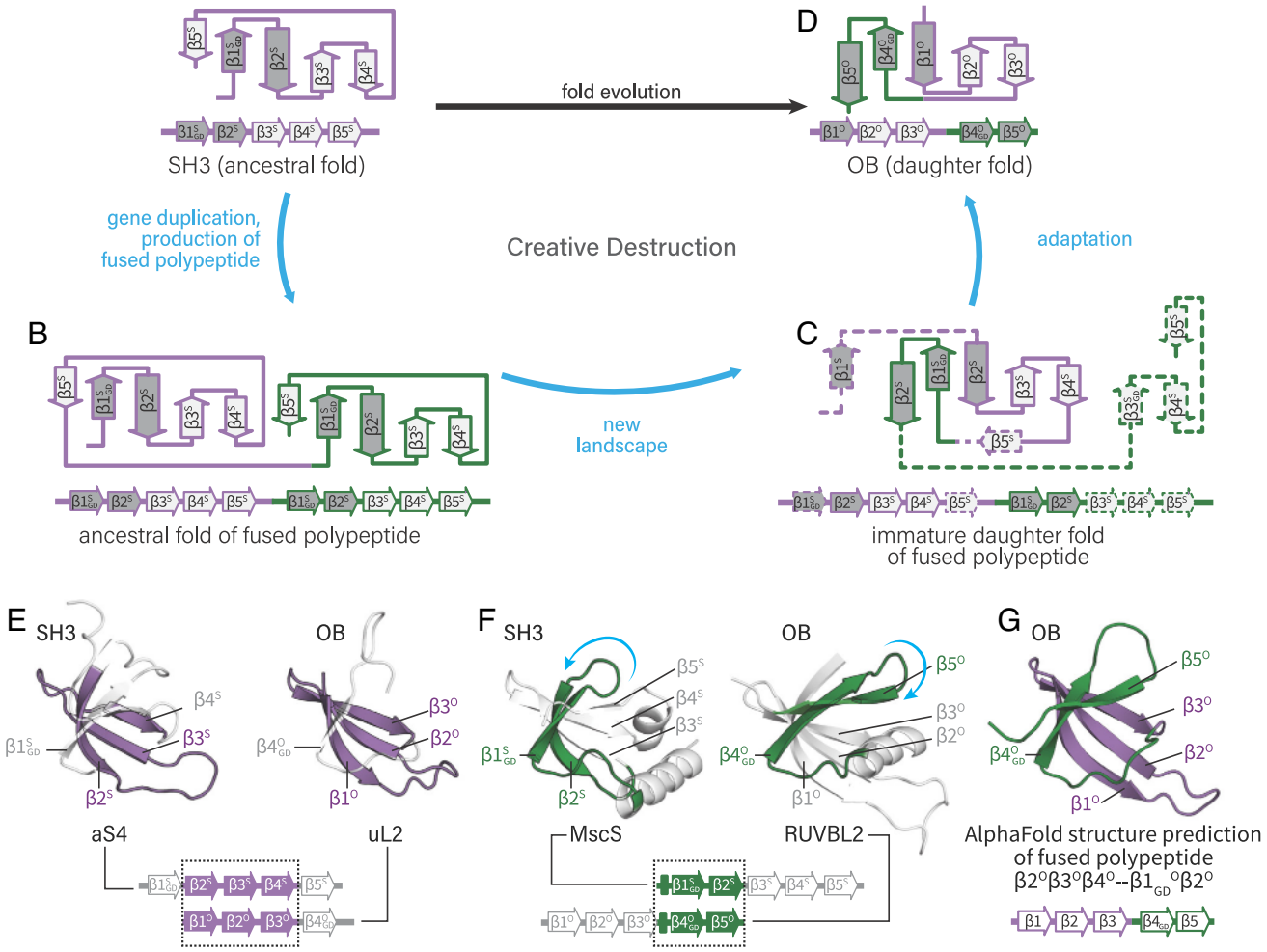

Fig. 4. Conversion from SH3 fold to OB fold by creative destruction. This figure shows topology diagrams of (*A*) the ancestral SH3 fold, (*B*) the notional ancestral folds of the fused polypeptide, (*C*) the immature re-folded fused polypeptide, and (*D*) mature daughter fold. A duplication of SH3 gives the fused polypeptide SH3-SH3. The fused polypeptide folds in a remodeled landscape and resolves by adaption, yielding an OB fold. GD indicates the GD-box motif. $\beta1_{GD}{}^{S}\beta2^{S}$ are shaded in dark gray to allow tracking the β-strand positions during the transition from ancestral fold to daughter fold. (*E*) Cross-fold sequence similarity (HHalign score 64%; E-value 6.5×10^{-5}; 32 aligned columns) between aS4 (PDB: 4V6U, chain AE) and uL2 (PDB: 1VY4, chain BA) corresponds to antiparallel β-strands $\beta2^{S}\beta3^{S}\beta4^{S}$ and $\beta1^{O}\beta2^{O}\beta3^{O}$. (*F*) Cross-fold sequence similarity (HHalign score 85%; E-value 5.8×10^{-6}; 25 aligned columns) between mechanosensitive channel of small conductance (MscS) (PDB: 2OAU) and RuvB-like AAA ATPase 2 (RUVBL2) (PDB: 2CQA) corresponds to $\beta1_{GD}{}^{S}\beta2^{S}$ and $\beta4_{GD}{}^{O}\beta5^{O}$. Regions without sequence similarity are white. Secondary structural elements that share sequence similarity are indicated by the same color in both members of the pair. Curved arrows follow secondary structural elements that display differences in conformation between folds. (*G*) Predicted fold of the fused polypeptide $\beta2^{S}\beta3^{S}\beta4^{S}$--$\beta1_{GD}{}^{S}\beta2^{S}$. The predicted structure has a characteristic five-stranded β-barrel structure with OB topology. pLDDT: 72.2 (predicted local difference distance test is a measure of local model quality, see Methods). pTM-score: 0.607 (pTM-score assesses the predicted full-length model, see *Methods*).

fold sequence similarities (HHalign score 77%, E-value 2.0×10^{-5}) between OB and CLB are mapped to the following secondary structural elements: $\beta1^{O}$~$\beta4^{C}$, $\beta2^{O}$~L^{C}; $\beta3^{O}$~$\beta5^{C}$; $\beta4_{GD}{}^{O}$~$\beta6_{GD}{}^{C}$; $\beta3^{O}$~$\beta5^{C}$; $\beta4_{GD}{}^{O}$~$\beta6_{GD}{}^{C}$; and L^{O}~$\alpha7^{C}$. A GD-box sequence motif is located in the loops of $\beta4_{GD}{}^{O}$~$\beta6_{GD}{}^{C}$.

Creative destruction of OB generates CLB. The CLB fold comprises a large group of folds that include six-stranded β-barrels with a variety of topologies (50, 51). Here we analyzed CLB folds of the Alanine Racemase C topology, which display no detectable sequence similarity to other CLB folds (51). The cross-fold sequence similarity of the OB fold can be mapped to two distinct regions of this CLB fold (Fig. 5 *E* and *F*). Compelling cross-fold sequence similarity within OB folds (HHscore more than 60%, E-value less than 5×10^{-5}, *SI Appendix*, Table S1) and within CLB folds (HHscore of 84%, E-value 6.3×10^{-6} *SI Appendix*, Table S3) suggests that homology relationships between OB and CLB folds can be generalized as follows: 1) $\beta1^{C}\beta2^{C}\beta3_{GD}{}^{C}$~$\beta1^{O}\beta2^{O}\beta4_{GD}{}^{O}$; 2) $\beta4^{C}L^{C}\beta5^{C}\beta6_{GD}{}^{C}\alpha7^{C}$~$\beta1^{O}\beta2^{O}\beta3^{O}\beta4_{GD}{}^{O}L^{O}$; and 3) a duplicated four-stranded OB $\mathbf{1}^{O}\beta2^{O}$ $\mathbf{3}^{O}$ $\mathbf{4}_{GD}{}^{O}\beta5^{O}$-- $\mathbf{1}^{O}$ $\mathbf{2}^{O}$ $\mathbf{3}^{O}$ $\mathbf{4}_{GD}{}^{O}$ $\mathbf{L}^{O}$ $\beta5^{O'}$ (where the prime indicates copy 2) retains the bold secondary elements, which are renumbered $\beta1^{C}\beta2^{C}\beta3_{GD}{}^{C}\beta4^{C}L^{C}\beta5^{C}\beta6_{GD}{}^{O}\alpha7^{C}$ in the final product.

Shared features between OB and CLB folds are consistent with creative destruction. This model supports cross-fold sequence similarities of $\beta1^{C}$~$\beta1^{O}$, $\beta2^{C}$~$\beta2^{O}$, $\beta3_{GD}{}^{C}$~$\beta4_{GD}{}^{O}$, $\beta4^{C}$~$\beta1^{O}$, L^{C}~$\beta2^{O}$, $\beta5^{C}$~$\beta3^{O}$, $\beta6_{GD}{}^{C}$~$\beta4_{GD}{}^{O}$, and $\alpha7^{C}$~L^{O}. In the daughter fold, $\beta5^{O}$ and $\beta5^{O'}$ are extruded, $\beta3^{O}$ is lost, and $\beta2^{O'}$ corresponds to a loop. $\beta4^{O'}$ forms the terminus of the β-sheet, next to $\beta1^{O}$. AlphaFold (49) predicts that a sequence construct containing $\beta1^{O}\beta2^{O}\beta4_{GD}{}^{O}$--$\beta1^{O}\beta2^{O}\beta3^{O}\beta4_{GD}{}^{O}$ could result in a six-stranded β-barrel fold with CLB topology: $\beta1\beta2\beta3_{GD}\beta4L\beta5\beta6_{GD}$ (Fig. 5 *G* and *SI Appendix*, Fig. S4).

Discussion

Protein structure is generally more conserved over evolution than sequence (35, 36). The converse, sequence conservation between divergent folds (cross-fold sequence similarity) suggests that evolution of folds (21) can in some instances outpace changes in

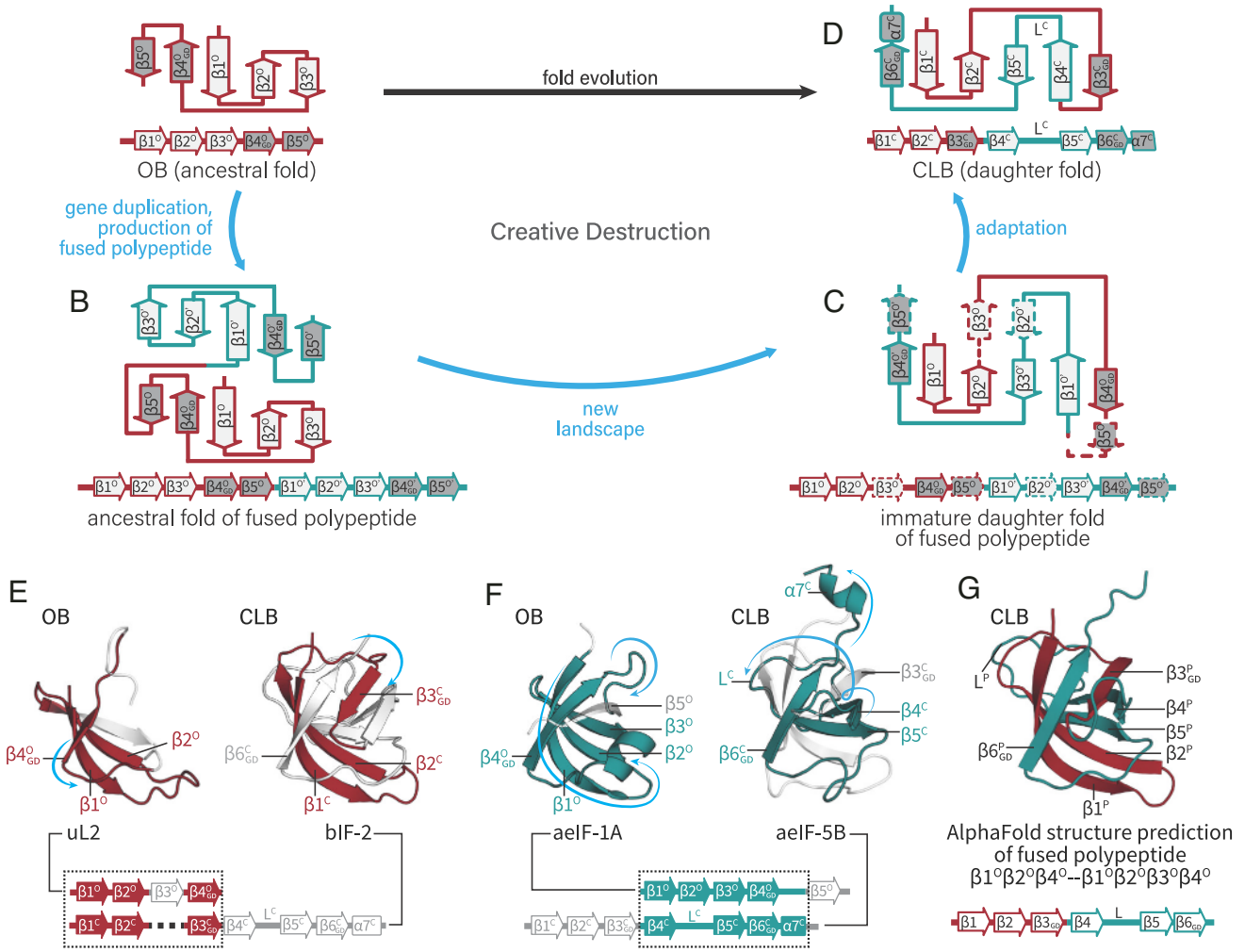

Fig. 5. Conversion of OB fold to CLB fold by creative destruction. This figure shows topology diagrams of (*A*) the ancestral OB fold, (*B*) the notional ancestral folds of the fused polypeptide, (*C*) the immature daughter fold of the fused polypeptide, and (*D*) the mature daughter fold. A duplication of SH3 gives the fused polypeptide OB-OB. The fused polypeptide folds in a remodeled landscape and resolves by adaption, yielding a CLB fold. $\beta4^{O}$~$\beta3^{C}$, and $\beta4^{O}$~$\beta6^{C}$ are shaded dark gray to allow tracking of their positions during conversion from ancestral fold to daughter fold. (*E*) Cross-fold sequence similarity (HHalign score 77%; E-value 3.9×10^{-6}; 36 aligned columns) between uL2 (PDB: 1VY4, chain BA) and bIF2 (PDB: 1ZO1, chain B) corresponds to $\beta1^{O}\beta2^{O}\beta4_{GD}{}^{O}$ and $\beta1^{C}\beta2^{C}\beta3_{GD}{}^{C}$. (*F*) Cross-fold sequence similarity (HHalign probability 77%; E-value 2.0×10^{-5}; 75 aligned columns) between eIF-1A (PDB: 2OQK) and aIF-5B (PDB: 4V8Z, chain CV) corresponds to $\beta1^{O}\beta2^{O}\beta3^{O}\beta4_{GD}{}^{O}L^{O}$ and $\beta4^{C}L^{C}\beta5^{C}\beta6_{GD}{}^{C}\alpha7^{C}$. Mapping of sequence similarity between OB and CLB: Regions that yield no sequence similarity are white. Secondary structural elements that share sequence similarity are indicated by the same color in both members of the pair. Curved arrows follow secondary structural elements that display differences in conformation between folds. (*G*) Predicted fold of the fused polypeptide $\beta1^{O}\beta2^{O}\beta4_{GD}{}^{O}$--$\beta1^{O}\beta2^{O}\beta3^{O}\beta4_{GD}{}^{O}L$. The predicted structure has a characteristic six-stranded β-barrel structure. pLDDT: 73.4 (predicted local difference distance test is a measure of local model quality, see *Methods*). pTM-score: 0.625 (predicted template modeling score assesses the predicted full-length model, see *Methods*).

sequence. Here, we identified cross-fold sequence similarities between three ancient β-barrel folds: SH3, OB, and CLB.

We describe a general mechanism of conversion between protein folds that appears to be widely applicable and ongoing. Patterns of cross-fold sequence similarities between SH3 and OB proteins are consistent with this mechanism. It appears that the OB fold arose via duplication of an ancestral SH3 gene, then expression of a fused polypeptide with a remodeled folding landscape, followed by adaptation (46). Extant OB fold proteins retain sequence and structure imprints from ancestral SH3 fold proteins. The polarity (SH3 -> OB versus OB -> SH3) is not fully resolved here and does not affect the conclusion. However, it seems most probable that SH3 is the ancestor of OB.

Patterns of cross-fold sequence similarity suggest an analogous origin of the CLB fold. The data suggest that the CLB fold arose from OB gene duplication, protein expression, folding of the fused polypeptide in a new landscape, and adaption. Final maturation of the CLB fold required loss of internal elements and tuning mutations. Extant CLB fold proteins retain both sequence and structural imprints of OB fold proteins.

Creative Destruction. The combined data support a general model of fold evolution that we call creative destruction. This model explains how ancestral folds beget daughter folds. Steps of creative destruction are gene fusion or truncation, production of an altered polypeptide with a remodeled folding landscape, and adaptation to produce a mature daughter fold. Each of these steps is known to be independently accessible and relatively frequent. An initial gene fusion step can cause insertion of one coding sequence in tandem with or within another (25, 52). The two ancestral genes, before fusion, may or may not express stable protein domains. A second step is expression of a fused polypeptide. The fusion of the polypeptide alters the sequence and spatial context of the ancestral polypeptides, destabilizes the ancestral folds, and opens alternative folding landscapes. The fused polypeptide collapses to a daughter fold that differs from either of the ancestral folds.

Creative destruction is applicable beyond the examples discussed here. It can explain relationships between complex folds. For example, two distinct cross-fold regions of sequence similarity have been identified between Flavodoxin-like folds and triose phosphate isomerase (TIM)-barrel folds (38, 53), indicating that

these folds share ancestry. Flavodoxin-like domain sequences partially align to sequences in the two halves of a TIM-barrel domain (53). As predicted by creative destruction, the TIM-barrel fold is not a replica of the Flavodoxin-like fold but is a distinct fold; some secondary and supersecondary structural elements of the ancestral fold are conserved, others have been destroyed, and some are created in the daughter.

For simplicity, the examples here focus on gene duplication. However, creative destruction might be initiated upon truncation, deletion, or incorporation of exogenous coding or non-coding sequences or by stop codon read-through (54). Any genetic process that alters a folding landscape has the potential to spawn evolution of new folds.

Circular Permutation. Circular permutation has been documented previously in hundreds of proteins (55) including concanavalin A (29) and other lectins (52), saposins (56), DNA methyltransferases (30, 57), and zinc ribbons (31, 58, 59). Circular permutation can arise by two distinct mechanisms. One mechanism, seen only for concanavalin A, occurs at the polypeptide level (29) and is a posttranslational re-wiring of the connections between secondary structural elements: 1) the N and C termini are joined; and 2) a loop is cleaved, generating new N and C termini. The second mechanism, as proposed here, involves creative destruction and occurs on the gene level (30, 52). Although the results appear indistinguishable, the two mechanisms are fundamentally different.

Circular permutation by creative destruction results from gene duplication, protein expression, exploration of new folding landscapes, and adaptation (Fig. 1). The fused polypeptide product collapses to a new fold and adapts by loss of terminal elements. In circular permutation, the new folding landscape produces secondary elements inherited from the ancestral domains albeit with different linkages. Even though the ancestral domains and the daughter domain of circularly permuted proteins have common secondary structural elements, secondary elements of each ancestral domain are destroyed during the process. The ancestral fold is not accessible to the daughter sequence. The daughter sequence might duplicate again, prompting another round of creative destruction resulting in formation of another fold. Although improbable, this process could revert to the original ancestral fold.

We previously noted that archaeal and bacterial versions of universal ribosomal protein uL33 are related by circular permutation (Fig. 2 *A*–*D*) (31, 59, 60). uL33 is a zinc ribbon, with two amphipathic β-hairpins (antiparallel ββ) linked by a zinc ion. The ordering of the ββ elements is switched in archaeal uL33 compared to bacterial uL33 ($\beta1^a\beta2^a$-$\beta3^b\beta4^b$ and $\beta3^a\beta4^a$-$\beta1^b\beta2^b$). Creative destruction provides a simple mechanism of conversion of archaeal to bacterial uL33 (and vice versa). Duplication of $\beta1^a\beta2^a\beta3^a\beta4^a$ gives fused polypeptide $\beta1^a\beta2^a$ **3^a 4^a** -- **1^a 2^a**$\beta3^a\beta4^a$. The fused protein collapses to a new fold, which omits ancestral elements $\beta1^a\beta2^a$ and$\beta3^a\beta4^a$, retaining the bold secondary elements to give daughter $\beta1^b\beta2^b\beta3^b\beta4^b$, where $\beta1^a$ $\beta3^b$, $\beta2^a$ $\beta4^b$, $\beta3^a$ $\beta1^b$ and $\beta4^a$ $\beta2^b$. The daughter fold adapts by loss of ancestral sequences for $\beta1^a\beta2^a$ and $\beta3^a\beta4^a$.

Motif Inheritance. Fold evolution by creative destruction is best described as a fold-space funnel, rather than a random walk. Daughter folds are derived by partially conservative and readily accessible processes from ancestral folds. Daughter folds are contingent on ancestral folds and inherit sequence and structural elements from them.

The creative destruction model of fold evolution provides a mechanism of conversion of one protein fold to another and offers a basis for establishing evolutionary relationships among diverse folds. The model can potentially explain the ubiquitous distribution of certain motifs, such as the GD-box (45) and the Rossmann-like motif (61). These motifs are seen in folds that appear to be otherwise unrelated. The GD-box demonstrates persistence between SH3, OB, CLB, and other folds. This persistent motif appears to provide stabilization of β-barrels by bringing together a β-turn and a non-contiguous β-strand (45). SH3 and OB have one GD-box motif, and CLB has two (Fig. 3). The creative destruction model explains the locations of these motifs in the sequence.

Fold Plasticity. In creative destruction, a gene that encodes a protein with secondary elements ABCD can fuse with a gene that encodes a protein with secondary elements EFGH. The result can be a fused polypeptide with ancestral secondary structure A**BC**D–E**FG**H. The fused polypeptide collapses in a new folding landscape and resolves by adaptation to daughter fold BCKG. The topology has been rewired, and F has changed conformation to K. Ancestral elements A, D, E, and H are lost. Some secondary elements of ancestors are retained, while others are modified or lost. The most frequent conformational changes are expected to be conversion of α-helices to β-strands and vice versa (21). Conformational plasticity enables the integration of folding-competent sequences into new supersecondary elements.

Creative destruction explains patterns of conserved and non-conserved sequence and structure between SH3, OB, and CLB folds. The data suggest conformational changes take place during fold evolution: β2 in an OB fold converts to a loop in CLB (consistent with cross-fold sequence similarity between uL2 and bIF2, Fig. 5*E*). A loop of OB converts to an α-helix in CLB (consistent with cross-fold sequence similarity between aeIF-1A and aeIF-5B, Fig. 5*F*). β5 in the ancestral OB fold is absent in the daughter CLB fold. Similarly, β5 in the ancestral SH3 fold is absent in the daughter OB fold. It is possible that the ancestral folds were four-stranded. Alternatively, the ancestors may have been five-stranded, but β5 was destroyed during creative destruction, either during the fusion event or later, during adaptation.

Mechanisms of Fold Evolution. A previously proposed mechanistic model of fold evolution (37, 43, 62) attributes cross-fold sequence similarities to mobility of ancestral peptides that are smaller than domains and are ancestral to them. Under this model, new folds emerge by repetition of (63) and decoration by small peptide elements (10, 64). Other models suggest that fold evolution occurs by preadaptation, combinatorial shuffling of supersecondary structures and transfer of isolated folding-incompetent motifs between proteins (62, 65, 66).

Creative destruction is an alternative model of fold evolution that, by contrast, acts at the level of domains. In creative destruction, well-characterized and frequent genetic processes, (25) such as full-length or partial gene duplication and gene fusion, provide access to partially conservative new folding landscapes. Creative destruction depends on fold plasticity, which is the facile exploration of new folding landscapes, as observed in protein switches (67), and by a tendency of many polypeptide sequences to self-associate, as observed in amyloid formation (68). In this model, emergence of a new daughter fold is prompted by the new combination of amino acids in the fused polypeptide and involves breaking of hydrogen bonds in the canonical ancestral folds and formation of new hydrogen bonds. Because these changes are sequence-dependent, only some gene fusion events may result in the emergence of a new fold.

The relationships described here include the OB domain of uL2 and the CLB domain of bIF2, which display robust folding (69), and SH3 domains, which bear conformational diversity (70–72). Creative destruction resolves cross-fold similarities by a

biologically plausible mechanism and is in agreement with the observation that the universe of protein folds is better described as a network than as a tree (53, 73).

Creative Destruction in Real Time. Creative destruction is not limited to the genesis of new folds at the dawn of life but appears to be ongoing today. In many cancers for example, chromosomal translocations commonly cause gene fusion (74–76). Breakpoint cluster region-Abelson murine leukemia (BCR-ABL) (77, 78) and Echinoderm microtubule associated protein like 4-Anaplastic lymphoma kinase (EML4-ALK) (79) are transforming genes that produce fused proteins. It seems likely that fused proteins such as BCR-ABL1 and EML4-ALK experience remodeled folding landscapes and altered folds compared to the non-fused gene products. The gain and altered functions of these fused polypeptides might arise in part from altered folding, as described by the creative destruction model proposed here.

Methods

Representative SH3, OB, and CLB Domains of Proteins from the Translation System. We studied the evolutionary relationship between the most prevalent and simplest β-barrel folds within the translation machinery: SH3, OB, and CLB. The information for specific representative proteins (uL2, aS4, and aS28, bIF2, aeIF-5B) that contain these folds is summarized in Table 1. MSAs of these representative proteins containing orthologous proteins from Sparse and Efficient Representation of Extant biology (36) were obtained from ProteoVision (80). The MSAs of ribosomal proteins were trimmed to the domain boundaries as defined by the Evolutionary Classification of Domains (ECOD) (34), and by Phase of Ribosomal evolution (1). The MSAs for ribosomal protein domains (deposited in FigShare, DOI 10.6084/m9.figshare.19412180, as Files 2–6 (81)) were transformed to profile hidden Markov models using the HHsuite version 3.3.0 (82).

Finding Transitive Relationships and Cross-Fold Sequence Similarity. To understand the relationships of SH3, OB, and CLB within and across folds, we searched the ECOD database (34). Profile files were retrieved from ECOD v283 (available at http://prodata.swmed.edu/ecod/complete/distribution). An initial sequence similarity search to identify sequence similarity was performed using HHsearch with our SH3, OB, and CLB profiles as queries. To distinguish transitive homology relationships from cross-fold sequence similarities according to the hierarchical classification of ECOD, the following criteria were applied: 1) HHalign scores greater than 60 and E-values better than 5×10^{-5} within the same X-, H-, and T-level groups in ECOD were considered transitive homologous relationships; 2) protein domains yielding HHalign scores greater than 60 and E-values better than 5×10^{-5} within different X-, H-, or T-level groups in ECOD were considered cross-fold sequence similarities.

Profiles of domains displaying either transitive homologous relations (*SI Appendix*, Tables S1–S3) or cross-fold similarities (*SI Appendix*, Tables S4 and S5) were retrieved and were compared in an all-versus-all fashion using HHalign with default parameters (82). HHalign scores of pairwise comparisons were deposited in a 2×2 matrix (*SI Appendix*, Fig. S2).

Table 1. Representative SH3, OB, and CLB folds

Protein name	Fold name	PDB code and chain	PDB fold range
uL2	OB	1VY4, chain BA	74–119
aS4	SH3	4V6U, chain AE	180–243
aS28	OB	4V6U, chain M	2–71
bIF2	CLB	1ZO1, chain B	190–267
aIF-5B	CLB	4V8Z, chain CV	477–559

Structural Analysis. Topology diagrams for these coordinate files were generated in ProteoVision using PDBsum (83). 3D representations of specific folds in the selected Protein Data Bank (PDB) files (Table 1) were rendered in PyMol (84). Pairs of folds displaying global sequence similarity were superimposed using TM-align (44). Pairs of folds displaying cross-fold sequence similarity were superimposed using Click (54) and manually adjusted using the pair fitting option in PyMol. Residues involved in cross-fold sequence similarities were highlighted by various colors in the topology diagrams and 3D structure representations.

Structure Predictions. Ancestral sequence reconstructions were calculated for aS4, MscS, uL2, and aeIF-1A (*SI Appendix*, Figs. S2 and S3). Sequence constructs were designed using HHalign alignments as reference (*SI Appendix*, Supplementary Methods). The 3D structures of these constructs (deposited in FigShare, DOI 10.6084/m9.figshare.19412180 as Files 7 and 8 (81)) were predicted with AlphaFold (49) in ColabFold (85) using the single-sequence option. An HHpred search (86) was performed to determine the best PDB model to be used as template. We report predicted template modeling score (pTM-score) and per-residue confidence score (pLDDT) for the best-scoring predictions. The pTM-score has a value between (0,1], where 1 indicates no predicted error of the full-length model (87). The pLDDT has a value between 0 and 100. Higher values indicate higher local structure confidence.

Data, Materials, and Software Availability. Sequence alignments and structure predictions associated with this manuscript have been deposited in the FigShare repository https://doi.org/10.6084/m9.figshare.19412180.

ACKNOWLEDGMENTS. The authors thank Jessica Bowman, Petar Penev, Eric Smith, Cameron Mura, Stella Veretnik, Philip E. Bourne, and Vijay Jayaraman for insightful discussions. This work was funded by the National Aeronautics and Space Administration grant 80NSSC18K1139. Claudia Alvarez-Carreño's research was supported by the NASA Postdoctoral Program, administered by Oak Ridge Associated Universities under contract with NASA.

Author affiliations: [a]NASA Center for the Origin of Life, Georgia Institute of Technology, Atlanta, GA 30332-0400; and [b]School of Chemistry and Biochemistry, Georgia Institute of Technology, Atlanta, GA 30332

1. N. A. Kovacs, A. S. Petrov, K. A. Lanier, L. D. Williams, Frozen in time: The history of proteins. *Mol. Biol. Evol.* **34**, 1252–1260 (2017).
2. C. Chothia, M. Gerstein, Protein evolution. How far can sequences diverge? *Nature* **385**, 579–581 (1997).
3. M. Levitt, Nature of the protein universe. *Proc. Natl. Acad. Sci. U.S.A.* **1 6**, 11079–11084 (2009).
4. A. V. Efimov, Structural trees for protein superfamilies. *Proteins* **28**, 241–260 (1997).
5. E. V. Koonin, M. Y. Galperin, *Sequence-Evolution-Function: Computational Approaches in Comparative Genomics* (Kluwer Academic, Boston, 2003), chap. 8.
6. C. Chothia, T. Hubbard, S. Brenner, H. Barns, A. Murzin, Protein folds in the all-beta and all-alpha classes. *Annu. Rev. Biophys. Biomol. Struct.* **26**, 597–627 (1997).
7. L. Pauling, R. B. Corey, H. R. Branson, The structure of proteins: Two hydrogen-bonded helical configurations of the polypeptide chain. *Proc. Natl. Acad. Sci. U.S.A.* **37**, 205–211 (1951).
8. W. L. Bragg, J. C. Kendrew, M. F. Perutz, Polypeptide chain configurations in crystalline proteins. *Proc. R. Soc. Lond. A Math. Phys. Sci.* **2 3**, 321–357 (1950).
9. S. T. Rao, M. G. Rossmann, Comparison of super-secondary structures in proteins. *J. Mol. Biol.* **76**, 241–256 (1973).
10. J. Söding, A. N. Lupas, More than the sum of their parts: On the evolution of proteins from peptides. *Bioessays* **25**, 837–846 (2003).
11. W. R. Taylor, J. M. Thornton, Recognition of super-secondary structure in proteins. *J. Mol. Biol.* **173**, 487–512 (1984).
12. M. Levitt, C. Chothia, Structural patterns in globular proteins. *Nature* **261**, 552–558 (1976).
13. C. Chothia, A. V. Finkelstein, The classification and origins of protein folding patterns. *Annu. Rev. Biochem.* **59**, 1007–1039 (1990).
14. G. D. Rose, P. J. Fleming, J. R. Banavar, A. Maritan, A backbone-based theory of protein folding. *Proc. Natl. Acad. Sci. U.S.A.* **1 3**, 16623–16633 (2006).
15. J. C. Bowman, A. S. Petrov, M. Frenkel-Pinter, P. I. Penev, L. D. Williams, Root of the tree: The significance, evolution, and origins of the ribosome. *Chem. Rev.* **12** , 4848–4878 (2020).
16. S. D. Fried, K. Fujishima, M. Makarov, I. Cherepashuk, K. Hlouchova, Peptides before and during the nucleotide world: An origins story emphasizing cooperation between proteins and nucleic acids. *J. R. Soc. Interface* **19**, 20210641 (2022).
17. E. Bornberg-Bauer, F. Beaussart, S. K. Kummerfeld, S. A. Teichmann, J. Weiner 3rd, The evolution of domain arrangements in proteins and interaction networks. *Cell Mol. Life Sci.* **62**, 435–445 (2005).
18. C. Vogel, V. Morea, Duplication, divergence and formation of novel protein topologies. *Bioessays* **28**, 973–978 (2006).
19. J. Weiner 3rd, E. Bornberg-Bauer, Evolution of circular permutations in multidomain proteins. *Mol. Biol. Evol.* **23**, 734–743 (2006).

20. A. Prakash, A. Bateman, Domain atrophy creates rare cases of functional partial protein domains. *Genome Biol.* **16**, 88 (2015).
21. N. V. Grishin, Fold change in evolution of protein structures. *J. Struct. Biol.* **134**, 167–185 (2001).
22. B. H. Dessailly, O. C. Redfern, A. L. Cuff, C. A. Orengo, Detailed analysis of function divergence in a large and diverse domain superfamily: Toward a refined protocol of function classification. *Structure* **18**, 1522–1535 (2010).
23. K. W. Kinzler, B. Vogelstein, Lessons from hereditary colorectal cancer. *Cell* **87**, 159–170 (1996).
24. A. K. Björklund, D. Ekman, A. Elofsson, Expansion of protein domain repeats. *PLoS Comput. Biol.* **2**, e114 (2006).
25. S. Lauer, D. Gresham, An evolving view of copy number variants. *Curr. Genet.* **65**, 1287–1295 (2019).
26. A. Lafita, P. Tian, R. B. Best, A. Bateman, TADOSS: Computational estimation of tandem domain swap stability. *Bioinformatics* **35**, 2507–2508 (2019).
27. J. A. Schumpeter, "The process of creative destruction" in *Capitalism Socialism and Democracy* (Harper Torchbooks, New York, 1950), pp. 81–86.
28. J. L. Xing, N. Sharif, From creative destruction to creative appropriation: A comprehensive framework. *Res. Policy* **49**, 104060 (2020).
29. B. A. Cunningham, J. J. Hemperly, T. P. Hopp, G. M. Edelman, Favin versus concanavalin a: Circularly permuted amino acid sequences. *Proc. Natl. Acad. Sci. U.S.A.* **76**, 3218–3222 (1979).
30. A. Jeltsch, Circular permutations in the molecular evolution of DNA methyltransferases. *J. Mol. Evol.* **49**, 161–164 (1999).
31. N. A. Kovacs, P. I. Penev, A. Venapally, A. S. Petrov, L. D. Williams, Circular permutation obscures universality of a ribosomal protein. *J. Mol. Evol.* **86**, 581–592 (2018).
32. A. G. Murzin, OB(oligonucleotide/oligosaccharide binding)-fold: Common structural and functional solution for non-homologous sequences. *EMBO J.* **12**, 861–867 (1993).
33. M. Coles *et al.*, AbrB-like transcription factors assume a swapped hairpin fold that is evolutionarily related to double-psi beta barrels. *Structure* **13**, 919–928 (2005).
34. H. Cheng *et al.*, ECOD: An evolutionary classification of protein domains. *PLoS Comput. Biol.* **1** , e1003926 (2014).
35. K. Illergård, D. H. Ardell, A. Elofsson, Structure is three to ten times more conserved than sequence–a study of structural response in protein cores. *Proteins: Struct. Funct. Bioinform.* **77**, 499–508 (2009).
36. C. R. Bernier, A. S. Petrov, N. A. Kovacs, P. I. Penev, L. D. Williams, Translation: The universal structural core of life. *Mol. Biol. Evol.* **35**, 2065–2076 (2018).
37. S. Nepomnyachiy, N. Ben-Tal, R. Kolodny, Complex evolutionary footprints revealed in an analysis of reused protein segments of diverse lengths. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 11703–11708 (2017).
38. J. A. Farías-Rico, S. Schmidt, B. Höcker, Evolutionary relationship of two ancient protein superfolds. *Nat. Chem. Biol.* **1** , 710–715 (2014).
39. W. R. Pearson, An introduction to sequence similarity ("homology") searching. *Curr. Protoc. Bioinformatics.* **42**, 311-318 (2013).
40. W. R. Pearson, BLAST and FASTA similarity searching for multiple sequence alignment. *Methods Mol. Biol.* **1 79**, 75–101 (2014).
41. A. G. Murzin, How far divergent evolution goes in proteins. *Curr. Opin. Struct. Biol.* **8**, 380–387 (1998).
42. J. Söding, Protein homology detection by HMM-HMM comparison. *Bioinformatics* **21**, 951–960 (2005).
43. V. Alva, J. Söding, A. N. Lupas, A vocabulary of ancient peptides at the origin of folded proteins. *eLife* **4**, e09410 (2015).
44. Y. Zhang, J. Skolnick, TM-align: A protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res.* **33**, 2302–2309 (2005).
45. V. Alva, S. Dunin-Horkawicz, M. Habeck, M. Coles, A. N. Lupas, The GD box: A widespread noncontiguous supersecondary structural element. *Protein Sci.* **18**, 1961–1966 (2009).
46. C. Alvarez-Carreño, P. I. Penev, A. S. Petrov, L. D. Williams, Fold evolution before LUCA: Common ancestry of SH3 domains and OB domains. *Mol. Biol. Evol.* **38**, 5134–5143 (2021).
47. V. Agrawal, R. K. Kishan, Functional evolution of two subtly different (similar) folds. *BMC Struct. Biol.* **1**, 5 (2001).
48. P. Youkharibache *et al.*, The small β-barrel domain: A survey-based structural analysis. *Structure* **27**, 6–26 (2019).
49. J. Jumper *et al.*, Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
50. M. Ammelburg *et al.*, A CTP-dependent archaeal riboflavin kinase forms a bridge in the evolution of cradle-loop barrels. *Structure* **15**, 1577–1590 (2007).
51. V. Alva, K. K. Koretke, M. Coles, A. N. Lupas, Cradle-loop barrels and the concept of metafolds in protein classification by natural descent. *Curr. Opin. Struct. Biol.* **18**, 358–365 (2008).
52. J. J. Hemperly, B. A. Cunningham, Circular permutation of amino acid sequences among legume lectins. *Trends Biochem. Sci.* **8**, 100–102 (1983).
53. N. Ferruz *et al.*, Identification and analysis of natural building blocks for evolution-guided fragment-based protein design. *J. Mol. Biol.* **432**, 3898–3914 (2020).
54. M. N. Nguyen, M. S. Madhusudhan, Biological insights from topology independent comparison of protein 3D structures. *Nucleic. Acids Res.* **39**, e94 (2011).
55. W. C. Lo, C. C. Lee, C. Y. Lee, P. C. Lyu, CPDB: A database of circular permutation in proteins. *Nucleic Acids Res.* **37**, D328–D332 (2009).
56. C. P. Ponting, R. B. Russell, Swaposins: Circular permutations within genes encoding saposin homologues. *Trends Biochem. Sci.* **2** , 256–256 (1995).
57. S. G. Peisajovich, L. Rockah, D. S. Tawfik, Evolution of new protein topologies through multistep gene rearrangements. *Nat. Genet.* **38**, 168–174 (2006).
58. S. S. Krishna, I. Majumdar, N. V. Grishin, Structural classification of zinc fingers: Survey and summary. *Nucleic Acids Res.* **31**, 532–550 (2003).
59. D. J. Klein, P. B. Moore, T. A. Steitz, The roles of ribosomal proteins in the structure assembly, and evolution of the large ribosomal subunit. *J. Mol. Biol.* **34** , 141–177 (2004).
60. N. Ban, P. Nissen, J. Hansen, P. B. Moore, T. A. Steitz, The complete atomic structure of the large ribosomal subunit at 2.4 a resolution. *Science* **289**, 905–920 (2000).
61. K. E. Medvedev, L. N. Kinch, R. Dustin Schaeffer, J. Pei, N. V. Grishin, A fifth of the protein world: Rossmann-like proteins as an evolutionarily successful structural unit. *J. Mol. Biol.* **433**, 166788 (2021).
62. R. Kolodny, S. Nepomnyachiy, D. S. Tawfik, N. Ben-Tal, Bridging themes: Short protein segments found in different architectures. *Mol. Biol. Evol.* **38**, 2191–2208 (2021).
63. H. Zhu *et al.*, Origin of a folded repeat protein from an intrinsically disordered ancestor. *eLife* **5**, e16761 (2016).
64. A. N. Lupas, C. P. Ponting, R. B. Russell, On the evolution of protein folds: Are similar motifs in different protein folds the result of convergence, insertion, or relics of an ancient peptide world? *J. Struct. Biol.* **134**, 191–203 (2001).
65. L. M. Longo, R. Kolodny, S. E. McGlynn, Evidence for the emergence of β-trefoils by "peptide budding" from an IgG-like β-sandwich. *PLoS Comput. Biol.* **18**, e1009833 (2022).
66. K. Qiu, N. Ben-Tal, R. Kolodny, Similar protein segments shared between domains of different evolutionary lineages. *Protein Sci.* **31**, e4407 (2022).
67. L. L. Porter, L. L. Looger, Extant fold-switching proteins are widespread. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 5968–5973 (2018).
68. Y. O. Chernoff, Amyloidogenic domains, prions and structural inheritance: Rudiments of early life or recent acquisition? *Curr. Opin. Chem. Biol.* **8**, 665–671 (2004).
69. P. To, B. Whitehead, H. E. Tarbox, S. D. Fried, Nonrefoldability is pervasive across the E. coli proteome. *J. Am. Chem. Soc.* **143**, 11435–11448 (2021).
70. L. C. James, D. S. Tawfik, Conformational diversity and protein evolution–a 60-year-old hypothesis revisited. *Trends Biochem. Sci.* **28**, 361–368 (2003).
71. P. Galaz-Davison *et al.*, Differential local stability governs the metamorphic fold switch of bacterial virulence factor RfaH. *Biophys J.* **118**, 96–104 (2020).
72. Y. Huang, J. Fang, M. T. Bedford, Y. Zhang, R. M. Xu, Recognition of histone H3 lysine-4 methylation by the double tudor domain of JMJD2A. *Science* **312**, 748–751 (2006).
73. S. Nepomnyachiy, N. Ben-Tal, R. Kolodny, Global view of the protein universe. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 11691–11696 (2014).
74. P. A. Futreal *et al.*, A census of human cancer genes. *Nat. Rev. Cancer* **4**, 177–183 (2004).
75. H. Rahi *et al.*, Gene fusions in gastrointestinal tract cancers. *Genes Chromosomes Cancer* **61**, 285–297 (2022).
76. Y. Zhou, M. El-Bahrawy, Gene fusions in tumourigenesis with particular reference to ovarian cancer. *J. Med. Genet.* **58**, 789–795 (2021).
77. T. P. Braun, C. A. Eide, B. J. Druker, Response and resistance to BCR-ABL1-targeted therapies. *Cancer Cell* **37**, 530–542 (2020).
78. J. Groffen, N. Heisterkamp, The chimeric BCR-ABL gene. *Baillieres Clin. Haematol.* **1** , 187–201 (1997).
79. M. Soda *et al.*, Identification of the transforming EML4–ALK fusion gene in non-small-cell lung cancer. *Nature* **448**, 561–566 (2007).
80. P. I. Penev *et al.*, ProteoVision: Web server for advanced visualization of ribosomal proteins. *Nucleic Acids Res.* **49**, W578–W588 (2021).
81. C. Alvarez-Carreño, R. J. Gupta, A. S. Petrov, L. D. Williams, CreativeDestruction_Supplementary_Files_2-8. figshare. Dataset. https://doi.org/10.6084/m9.figshare.19412180.v3. Deposited 30 November 2022.
82. M. Steinegger *et al.*, HH-suite3 for fast remote homology detection and deep protein annotation. *BMC Bioinf.* **2** , 473 (2019).
83. R. A. Laskowski, J. Jabłońska, L. Pravda, R. S. Vařeková, J. M. Thornton, PDBsum: Structural summaries of PDB entries. *Protein Sci.* **27**, 129–134 (2018).
84. Schrödinger LLC, The PyMOL molecular graphics system, (2.4.0, Schrodinger, LLC, 2021).
85. M. Mirdita *et al.*, ColabFold: Making protein folding accessible to all. *Nat. Methods* **19**, 679–682 (2022).
86. F. Gabler *et al.*, Protein sequence analysis using the MPI bioinformatics toolkit. *Curr. Protoc. Bioinf.* **72**, e108 (2020).
87. Y. Zhang, J. Skolnick, Scoring function for automated assessment of protein structure template quality. *Proteins* **57**, 702–710 (2004).