Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Structural Biology

A building blocks perspective on protein emergence and evolution

Yishi Ezerzer[1], Moran Frenkel-Pinter[1,2], Rachel Kolodny[3] and Nir Ben-Tal[4]

Recent findings increasingly suggest the emergence of proteins by mix and match of short peptides, or 'building blocks'. What are these building blocks, and how did they evolve into contemporary proteins? We review two complementary approaches to tackling these questions. First, a bottom-up approach that involves identifying putative components of primordial peptides, and the synthetic routes through which these peptides may have emerged. Second, searches in protein space to reveal building blocks that make up the contemporary protein repertoire; proteins that are not closely related to one another may nevertheless have certain parts in common, suggesting common ancestry. Identifying such shared building blocks, and characterizing their functions, can shed light on the ancient molecules from which proteins emerged, and hint at the mechanisms that govern their evolution. A key challenge lies in merging these two approaches to create a cohesive narrative of how proteins emerged and continue to evolve.

Addresses
[1] Institute of Chemistry, The Hebrew University of Jerusalem, 9190401, Israel
[2] The Center for Nanoscience and Nanotechnology, The Hebrew University of Jerusalem, 9190401, Israel
[3] Department of Computer Science, University of Haifa, Haifa, Israel
[4] School of Neurobiology, Biochemistry and Biophysics, George S. Wise Faculty of Life Sciences, Tel Aviv University, Tel Aviv, Israel

Corresponding authors: Ben-Tal, Nir (bental@tauex.tau.ac.il); Kolodny, Rachel (trachel@cs.haifa.ac.il); Frenkel-Pinter, Moran (moran.fp@mail.huji.ac.il)

Current Opinion in Structural Biology 2025, **91**:102996

This review comes from a themed issue on **Folding and Binding (2025)**

Edited by **Yana Bromberg** and **Ben Schuler**

For complete overview of the section, please refer the article collection - Folding and Binding (2025)

Available online 6 February 2025

https://doi.org/10.1016/j.sbi.2025.102996

0959-440X/© 2025 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Keywords
Ancestral peptides, Primordial peptides, Protein space, Protein domains, Prebiotic chemistry, Origin of life.

Introduction

In our current era, the age-old question of the origin of life can be cast in molecular terms: How did the complex macromolecules that make up contemporary life emerge some 3.7 billion years ago? This question seems to be rooted in a paradox: Genetic information for protein synthesis is stored in DNA/RNA, which, in turn, is synthesized by proteins. Nevertheless, and in spite of arguments in support of an RNA-first hypothesis [1], recent findings increasingly suggest that a very plausible scenario for the emergence of proteins, perhaps entangled to the emergence of nucleotides, is by mix and match of short peptides, or 'building blocks' [2–4]. Indeed, Kocher and Dill recently suggested an abstract kinetic model to this effect, based on formation of peptides with cooperative autocatalytic activity [5].

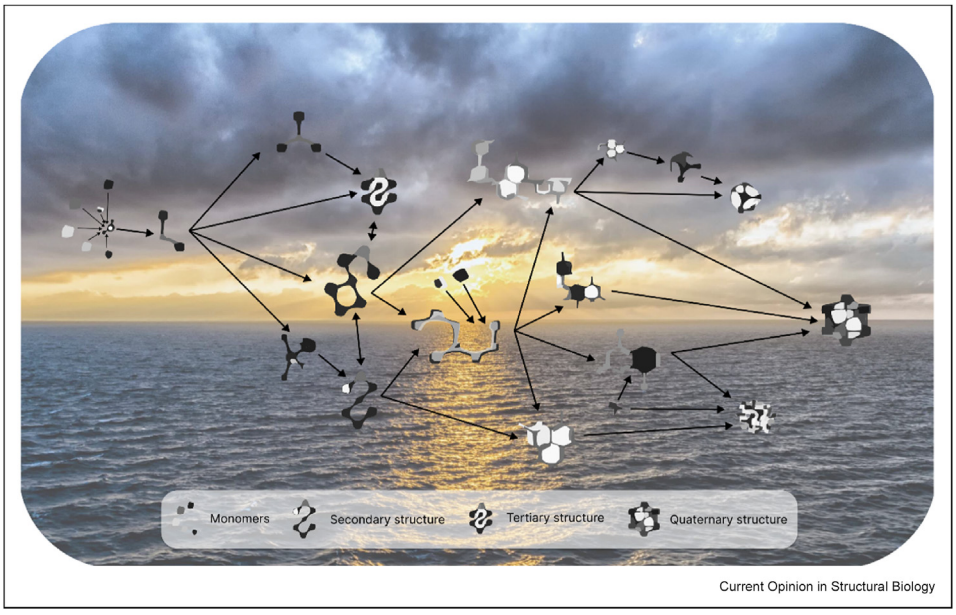
What, then, are these building blocks, and how did they lead to contemporary proteins? We review two complementary approaches to tackling these questions. First, drawing from works in prebiotic and organic chemistry, we describe a bottom-up approach that involves identifying putative components of primordial peptides (building blocks) and the synthetic routes through which these peptides may have emerged. The second approach, promoted by computational structural biologists, searches in protein space to reveal building blocks that make up the contemporary protein repertoire [6,7]. We hope that, ultimately, these two approaches will merge into a cohesive narrative that reveals how proteins emerged and continue to evolve (Figure 1).

The organic chemist's search for primordial peptides

Identifying abiotically plausible building-block components for early proteins

Researchers have explored various pathways for abiotic peptide formation [8] and have studied several plausible candidates for primordial peptide backbones (building blocks) (Figure 2a). Monomers that have been considered as possible components of primordial peptides include hydroxy acids, mercapto acids, β- and γ-derivatives of amino acids and of hydroxy acids, and non-

Figure 1



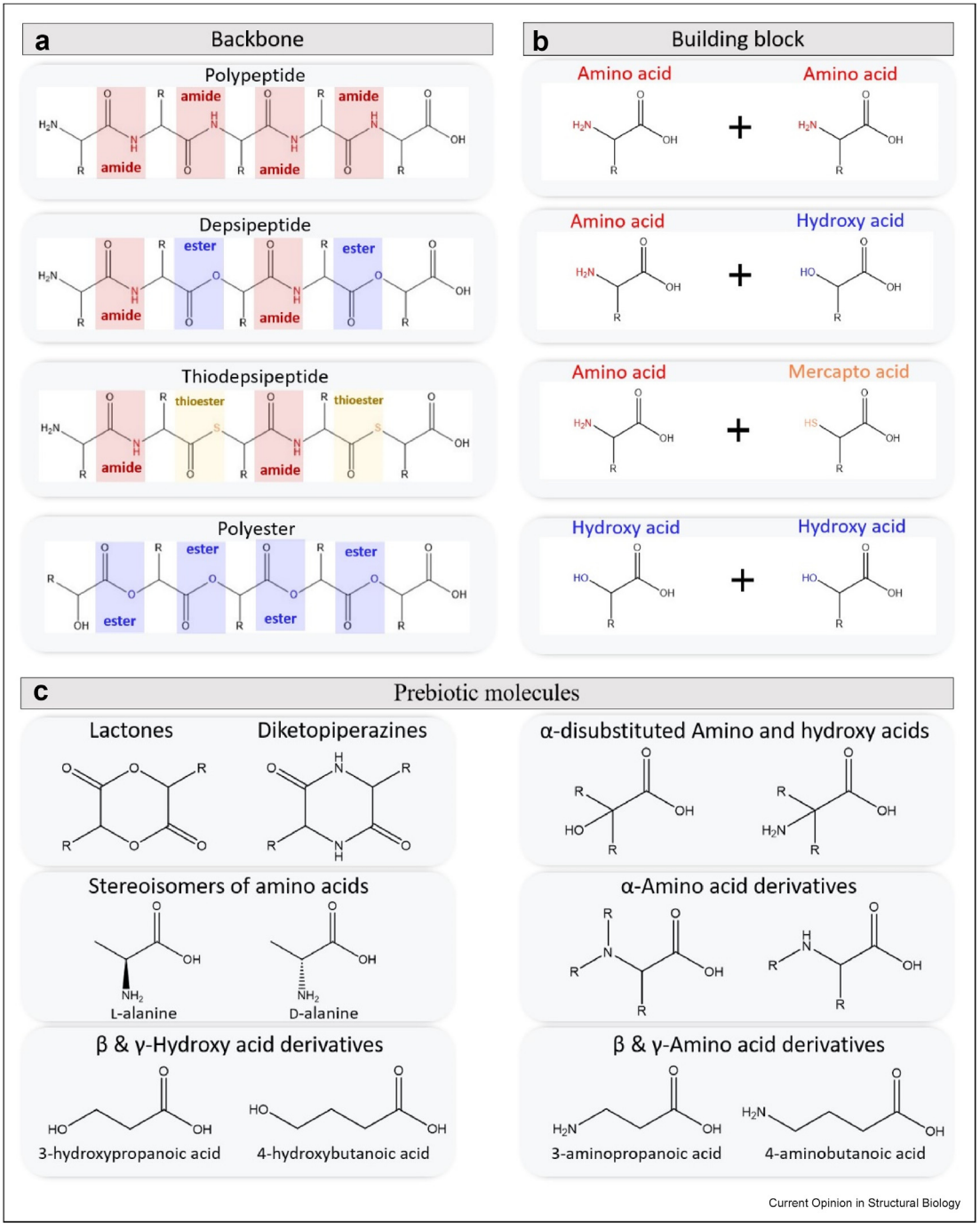
Protein emergence and evolution from building blocks. An abstract representation of how proteins might have emerged and continue to evolve through the mixing and matching of smaller building blocks. The nature of these building blocks may have changed over time, starting from biotic—and possibly abiotic—compounds to the complex proteins we observe today. The earliest proteins may have been small, with limited stability and minimal function, just sufficient to be selected for further processes. Over time some of these gradually improved in stability, foldability and function. Integration of perspectives from prebiotic chemistry and computational biology could ultimately allow replacement of this abstract view with a more realistic scenario. The annotations at the bottom of the figure indicate a possible interpretation as to the nature of the entities.

proteinogenic amino acids such as α-disubstituted amino acids (Figure 2b) [9–14]. Other, lesser-studied candidate molecules, sometimes referred to as xenobiological molecules, are closely related to amino acids, but feature slight modifications such as elongated carbon chains, cyclic variations, and derivatives lacking the carboxylic group (Figure 2c) [15–20]. In what follows we briefly discuss some of these molecules and the features that may have enabled them to give rise to the beginnings of peptides.

Hydroxy acids can readily polymerize into polyesters under prebiotic conditions [21–24]. In 1971, Rich proposed that these polyesters may have preceded polypeptides [25]. Hydroxy-acid-derived polyesters have various functions, but two features make them particularly interesting for our context. First, they can form microdroplets in aqueous solutions. Microdroplets are small assemblies that resemble vesicles or micelles, hinting at the emergence of compartmentalization through encapsulation [26]. These microdroplets can even encapsulate small molecules, offering the first clues for the development of self-replicating encapsulated systems [27,28]. Second, hydroxy-acid-derived polyesters are susceptible to ester-amide exchange reactions. These exchange reactions allow peptide bond formation under abiotic conditions, where peptide bond formation via direct condensation reactions between amino acid monomers is kinetically unfavorable. Specifically, co-oligomerization of amino acids and hydroxy acids results in the formation of depsipeptides (peptide-like molecules that contain mixed ester and peptide bonds in their backbone; Figure 2a, central panel), and the labile ester linkages are replaced by kinetically trapped amide bonds over time [8,13,21,29–32]. This reaction, which mimics peptide bond formation in the ribosome via ester intermediates (with tRNA), suggests an early mechanism for the evolution of modern peptides. Depsipeptides are fascinating to study as prebiotic protopeptides because they are abundant in today's biology and exhibit a variety of functions [33–36]. For example, they act as antimicrobial agents [34], anticancer compounds [33], and even as structural molecules [35,36]. These properties closely mirror some of the many functions of contemporary proteins in biological systems.

Mercapto acids, similarly to hydroxy acids, can readily polymerize into thioesters in drying reactions [12]. Although thioesters are not as hydrolytically stable as peptides, they can participate in various reactions that make them intriguing candidates for proto-peptide formation and proto-metabolites in the context of chemical evolution. Thioesters can undergo thioester-amide

Figure 2



Diverse prebiotic chemical space. (**a**) Chemical structures of peptides and postulated proto-peptides (depsipeptides, thiodepsipeptides, and polyesters), as well as their respective building blocks (**b**), which are mentioned in this paper. (**c**) Structures of several prebiotic molecules mentioned in this paper.

exchange reactions, similarly to esters, to form thiodepsipeptides enriched in both thioester bonds and peptide bonds [12]. Thioester bonds as intermediates to peptide bonds in prebiotic chemistry are of particular interest due to their significant role in modern biology. For example, in contemporary metabolism, acetyl-CoA acts as a cofactor and initiator of the Krebs cycle, the primary energy source for most life forms. Acetyl-CoA uses a thioester bond to deliver an acetyl group to the Krebs cycle, where it is oxidized to produce energy [37]. Moreover, thioester bonds are utilized in non-ribosomal peptide synthesis pathways in modern biology, making

them compelling candidates for research into the chemical evolution of proto-peptides [15,38,39].

Most contemporary research on proteins is limited to only the 20 standard α-L-amino acids. However, β - and γ-derivatives of amino acids and of hydroxy acids present intriguing avenues for research, as they were abundant on prebiotic Earth [14,40,41] and exhibit diverse and noteworthy biological functions [42], for example in the context of non-ribosomal peptides. β- and γ-derivatives are capable of polymerization into peptides, polyesters, and depsipeptide derivatives [10,13]. Although they exhibit different properties compared to α-derivatives, including lower rigidity due to reduced hydrogen bonding density along the polymer backbone, their flexibility could yield unique and valuable structural properties [43]. Furthermore, β- and γ-amino acids can serve as modifications in biological peptides and cofactors, such as the β-alanine of coenzyme A, involved in acetyl-CoA metabolism [42]. While pure β- and γ-oligomers may not be as biologically active or stable, incorporating them into existing proto-oligomers could lead to fascinating insights and bring us closer to understanding the evolution of modern polymers and the subsequent selection leading to an α-peptide backbones.

Path forward in the bottom-up search for protein building blocks

Most research on chemical evolution focuses on studying contemporary biomolecules and attempting to recreate their chemical space without relying on biological processes [44]. However, it is possible that life started with polymers other than those found in extant life, and that biochemistry might have lost many molecules that were of high importance during prebiotic chemistry. Hence, it can be illuminating to investigate non-biological molecules as well [18], encompassing a wide range of molecules, e.g., cyclic variants of amino and hydroxy compounds (such as diketopiperazines and lactones), or amino acid derivatives featuring secondary amines in place of primary amines. Notably, proline itself is a unique amino acid in that it is the only proteinogenic amino acid that has a secondary amine in its backbone rather than a primary amine.

Moreover, the origin of life may have involved highly cooperative interactions between diverse classes of molecules. Given the role of proteins and amino acids in extant biology as key molecules in the networks connecting different classes of biomolecules, it is possible that early molecular evolution was influenced by the co-evolution of amino acids and primordial proteins with other classes of organic and inorganic molecules. Such co-evolution could have involved noncovalent interactions between peptides and other classes of molecules, as well as direct covalent linkages between amino acids/peptides and their interaction partners. Indeed, amino acids can form a variety of both covalent and non-covalent interactions with other molecules.

Understanding the composition and properties of peptides and proteins derived from biological systems provides significant insights into their synthesis pathways, selectivity, and biological importance. For example, more than 70 % of all secreted or membrane proteins are covalently linked to glycans [45]. Protein glycosylation, a result of co-translational or posttranslational modification, affects protein solubility and folding. Lipidation of peptides and proteins with long-chain amphiphiles has been shown to induce membrane association. Such modifications could have occurred on primordial peptides and further research is necessary to understand their role and importance in the context of prebiotic chemistry. It may well be that in the absence of evolved protein enzymes, prebiotic chemistry had also produced covalent adducts between amino acids/peptides and other organic molecules, similar to the common biological covalent conjugates of amino acids/peptides. For example, covalent bonds between amino acids and nucleic acids have been reported to form in the presence of condensing molecules [46,47] and dehydration reactions [48]. Lipid-amino acid conjugates (lipoamino acids), which are naturally occurring molecules [48], can also form abiotically [49—51].

Finally, given that the findings reviewed above establish that peptides are 'easy' to make abiotically, we believe that the next challenge lies in understanding how the vast array of possible peptides converged to the comparably limited repertoire of biological building blocks that we know today. The underlying assumption in the field is that the process was guided by evolutionary selection of sort, which can be revealed. One should keep in mind, though, that the process is stochastic, possibly involving 'frozen accidents'. The 'frozen accidents' explanation is, however, not helpful as a guide for research.

The computational search in protein space for protein building blocks

Proteins often share sequence and/or structure similarity [52]. For the most part this redundancy reflects emergence from a common evolutionary origin, i.e., the well understood homology; for a recent review see Ref. [53]. However, even proteins that are not entirely similar to each other may nevertheless share similar parts, e.g., the same stretch of amino acids or linear motif that mediates a specific function. Such short linear motifs are often recognition sites, e.g., for phosphorylation and other post-translational modifications, or phosphate binding [54]. Likewise, proteins with different overall sequence and structure may share a structural motif, like the relatively large DNA-binding

zinc finger, made of a beta-hairpin packed against an alpha-helix, or as few as 3–4 amino acids with the right stereochemistry for coordinating ion binding or enzymatic catalysis.

Top-down computational approaches, which cover all forms of similarities, can provide insights as to how proteins emerged and continue to evolve. In what follows, we focus on two aims: (1) identifying ancient entities (building blocks) of the protein world [7], and (2) revealing mechanistic aspects [53]. As more data accumulate in protein databases, the usefulness of such data-driven approaches may increase, but suitable methodologies for handling it are required. Because comprehensive searches in these large databases are computationally challenging [55], many searches thus far have focused on defined protein sets, especially ones that are likely ancient.

Identifying ancient entities

Protein space-wide studies

The best-known form of similarity in protein space is the domain. Domain databases, like SCOP, CATH, CDD, and ECOD, segment protein chains into domains based on sequence and structure similarity [52]. Yet the details of these segmentations vary, presumably reflecting different views about the nature of these shared building blocks as derived from the data available in the 1990s, when the databases were initiated. Two important decisions are implemented by all these databases: (a) an amino acid can belong only to one domain within a particular protein, and (b) a domain includes approximately 100 amino acids. The effects of these decisions are far-reaching because many protein studies are interpreted in light of domain knowledge.

Building blocks at the sub-domain level, including ones shared by domains classified in different folds, have long been recognized [56–58]. Nepomnyachiy et al. developed algorithmic tools and a computational pipeline to systematically search for shared protein segments of various lengths in a representative set of over 20,000 domains [59]. This search revealed shared ('reused') segments of 35–200 amino acids, denoted 'themes'. Indeed, reuse is prevalent in protein space and increases with shorter themes, which could be remnants of more ancient history. Interestingly, shared themes of 100 amino acids do not stand out, perhaps because domains, sometimes considered 'the evolutionary building blocks', are only one instance of similarity in protein space. Even though some domains can be traced all the way back to the last universal common ancestor (LUCA), traces in protein space of even earlier events can be found [59,60].

In an insightful paper, Lupas, Ponting, and Russel highlighted that shared fragments from domains considered of different lineages may be vestiges of ancient peptides from which these domains evolved [7]. Following this, Alva et al. compared domains of different folds and identified 40 ancient peptides that could be at the origins of folded domains [60]. Kolodny et al. searched for instances of the themes identified in Ref. [59] in domains of different folds; they denoted these 'bridging themes' [61–63]. Some 'bridging themes' bind ligands, a function that attests to their functional importance, and likely explains how their sequences were conserved so that they can be detected. Indeed, the ancient peptides and bridging themes may be building blocks left from the emergence of the earliest protein families.

Insights gained from studying a pre-defined set of proteins

Several studies identified shared patterns among proteins that still perform ancient interactions, and reasoned that these patterns may correspond to ancestral forms. Narunsky et al. explored protein binding to the ancient ligand adenine [64]. They found that adenine binding is often mediated by specific amino acid segments (themes) that recur across different proteins. Different themes bind different adenine-containing ligands, suggesting that adenine binding has emerged multiple times throughout evolution. Studying protein metal binding, Raanan et al. [65] analyzed protein structures with transition metals or electron transfer-related cofactors to identify common patterns that may be at the origin of modern oxidoreductases. Bromberg et al. [66] tailored a computational pipeline for comparison of metal binding sites and showed that even seemingly unrelated proteins may nevertheless share metal binding sites that are similar in sequence and structure. Zheng et al. [54] studied phosphate binding and showed that even radically different proteins that undergo the same post-translational modifications often share similar short linear motifs.

Others focused on deciphering the ancestral building blocks in a specific class or fold. Gruic-Sovulj et al. studied HUP domains, a class of ancient enzymes that includes Class I aminoacyl tRNA synthetases (AARSs), and enzymes that mediate NAD, FAD, and CoA biosynthesis [67]. Their analysis suggested that these domains emerged from a seed βαβ fragment and highlighted the prominent role of this building block in the emergence of enzymes. A systematic analysis of phosphate binding by Longo et al. also supported the hypothesis that in ancient enzymes, phosphate binding is often mediated by the N-terminal of the helix in the βαβ building block [68]. Studying outer membrane β-barrels (OMBBs)—a major class of outer membrane proteins from Gram-negative bacteria, mitochondria, and plastids—Remmert et al. showed that this class originated from a single, ancestral ββ hairpin [69].

Rather than analyzing a single fold, one could search for a shared building block in several folds to deduce their shared ancestral core. Longo et al. searched for structure and sequence elements shared by P-loop NTPases and Rossmann folds [63]. The two folds include ubiquitous enzyme families and are considered particularly ancient. They found homologous segments that span the first βαβ motif of both lineages, including the phosphate binding loop and a conserved aspartate at the tip of β2. It is likely that both lineages emerged from this shared βαβ building block, supporting the key role of this structural element as a scaffold for chemical functionality (e.g., the key aspartate) and, in that way, the eventual emergence of these broad enzymatic classes. Alvarez-Carreño et al. studied the similarities between SH3 and OB – the most ancient β barrel domains within the translation system [70]. SH3 and OB domains superimpose structurally well, but are topologically different [71], requiring sequence permutations to have evolved one from another. The authors predicted the structures of these suggested ancestors, and showed that an SH3-based permutated variant folded into an OB and *vice versa*. The structural similarity between these two domains was examined previously by Mura et al., within the overall context of structural similarity between domains of different topologies [72].

The Mura et al. study includes other cases of shared building blocks, where the N-to-C directionality of the elements were reversed, cases with alpha-to-beta interconversion between elements, and significant changes in the tilt angles between elements (e.g., a pair of helices) [72]. Their observations of common building blocks led them to suggest adding the Urfold level to the hierarchical classifications (where Ur is 'primitive,' or 'ancestral'), specifically between the Architecture and Topology levels in CATH. Based on the observation that there are instances where there is different order of secondary structure that is not necessarily due to a common evolutionary origin, they proposed that Urfold be more general and indicative of physicochemical qualities. Draizen et al. implemented these ideas in DeepUrfold, an AI model to explore distant protein structure relationships, that finds many fold similarities, including at the sub-domain level, which the authors proposed echo remote homologies [73].

Understanding mechanisms of protein evolution

Protein-space-wide studies

Studies tracing domain architectures, the usage of domain building blocks, have been conducted since domain databases were first established and have revealed much about mechanisms of protein evolution, including mutations, homologous and non-homologous recombination, accretion, fusions, and augmentation [74]. Recently, Smug et al. analyzed usage patterns of domains, and other fragments, in bacteriophage protein space to characterize the evolutionary mechanisms used by viruses [75]. They detected ample recent fragment shuffling events, e.g., in receptor-binding proteins, often among viruses of seemingly unrelated taxonomies, and across ecological boundaries. They suggested that the ongoing diversification via fragment shuffling is driven by viral need to overcome emerging bacterial resistance mechanisms.

Nepomnyachiy et al. found recursive patterns akin to 'Russian nested dolls', where long themes shared by a few proteins encompass shorter themes that appear in more proteins. In other words, relaxing the (arbitrary) constraints imposed by the domain databases and allowing an amino acid to belong to a short theme, shared by many proteins, and to longer themes shared by fewer proteins reveals a wealth of evolutionary traces [59]. These traces may be fossils of the evolutionary process that created the proteins in which they are found, with the shorter ones perhaps left from more ancient events that diverged and lengthened over time. Studying bridging themes shared by different folds, Kolodny et al. found that their structures are not necessarily conserved within different protein contexts [61,62]. This structural plasticity may underlie the capacity of bridging themes to fit into different environments. That they are shared between Rossmann, P-loop, TIM-barrels, and other ancient folds is a testimony to their evolutionary importance.

To form a stable protein, protein building blocks must complement each other (to a certain degree) in their geometrical, stereochemical and other physicochemical qualities. But proteins are dynamic entities that have to undergo conformational changes for proper function, e.g., between active and inactive states. With that in mind, Kutlu et al. [76] studied the correlation between themes and equilibrium dynamics, approximated using the Gaussian network model (GNM), a close relative of normal modes analysis, that allowed the decomposition of domain motion into dynamic elements. Analysis of a diverse set of 150 domains showed that most often the decomposition into dynamic elements correlates with the theme decomposition of the domains. Unless it is due to a yet unknown confounding element, the correlation provides further support for the role of themes in protein evolution and highlights the importance of dynamic match between themes (or other building blocks) in protein 'evolutionary design'.

Insights gained from studying a pre-defined set of proteins

Interestingly, cases of similarities among different folds can also provide insights into the mechanisms of evolution. Chakravarty et al. computationally analyzed a

family of ∼600,000 DNA-binding bacterial response regulator proteins [77]. Their analysis revealed an evolutionary pathway between two DNA-binding motifs: helix-turn-helix vs. winged-helix. Longo et al. [78] detected a non-trivial bridging theme linking β-trefoil and other all-β domains, especially in IgG-like domains. Although bridging themes often appear in different environment-dependent conformations, this purely-sequence based search yielded structures that superimposed very well, including a water molecule, considered important for the folding and stability of β-trefoils and that was also found in the IgG-like domains. Thus, the authors suggested an evolutionary mechanism for the emergence of the relatively recent β-trefoils from the very ancient IgG-like domains. According to their model, a bridging theme budded from an IgG-like domain, and evolved to form a homotrimer, with subsequent monomeric fusions that formed the 3-repeats β-trefoil.

Analyzing the shared building blocks of small, ancient, β-barrel folds that have common ancestors also revealed evolutionary mechanisms. Alvarez-Carreño et al. proposed a 'creative destruction' mechanism by comparing SH3, OB and CLB folds [79]. Creative destruction is based on gene duplication followed by deletions. It offers a model where SH3 domains evolved to OB domains, which then further evolved to CLB domains. Yagi & Tagami used clever structure-based engineering to artificially evolve proteins between four contemporary small β-barrel folds: double-psi β-barrels (DPBBs), RIFT, OB, and SH3 [80]. A key to the moves that they implemented is an additional metamorphic fold DZBB. A major step towards the leap from the original DPBB fold to the three other folds are alterations that shift the population of DZBB towards a less populated conformation, and then further engineering led to the other β-barrel folds. Conceptually, the engineering was based on comparison of the building blocks composing the four folds, with changes mostly in the linkers between them or removal of a complete building block (even though the term 'building blocks' was not used). Even though the proposed evolutionary pathway between all four folds is via DZBB, DZBB itself did not survive, and the authors discuss possible reasons. As DPBBs are found in RNA polymerase and the three other beta-barrel folds are found in ribosomal proteins, they suggested evolutionary pathway links between these two macromolecular machines.

Path forward in the top-down analysis of protein space

The protein sequence databases are expanding at an ever-fast pace, and enabled by the dramatic AI success in structure prediction, structure databases closely follow. Within these datasets may lie non-trivial evolutionary pathways, that can indicate how proteins have emerged. However, the opportunity of mining this data for hints of the emergence of proteins comes with the challenge of analyzing such huge datasets. We believe AI will be a significant contributor in these efforts, enrolling generative [73] and discriminative models. Specifically, analyzing embeddings of protein space by Protein Language Models (PLMs) may reveal non-trivial relationships and submerged building blocks.

Concluding remarks

Excluding creationist explanations for the origin of life, the very seemingly possible scenario for the emergence of proteins is by mix and match of short peptides, or protein building blocks (Figure 1). We have reviewed two approaches for identifying these building blocks and revealing their evolutionary path into contemporary life: a bottom-up approach, in which prebiotic chemists explore various means of abiotic peptide formation, and a top-down approach, in which computational structural biologists investigate common sequences shared by seemingly distant proteins. Linking the two approaches, perhaps using AI, could be key to providing a clear explanation of the initiation of life.

Uncovering the exact origins of life as they occurred is an impractical goal, given the countless bits of information lost since it has emerged some 3.7 billion years ago. Instead, the purpose of studying chemical evolution is to better understand how life may emerge from chemistry by identifying plausible pathways for its formation. This is not just about recreating the conditions of Earth's early history, which we may never fully verify, but about uncovering universal principles of life's emergence in any form. This broader perspective is what makes this fundamental research question so interesting.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Nir Ben-Tal reports financial support was provided by Israel Science Foundation. Rachel Kolodny reports financial support was provided by Israel Science Foundation. Moran Frenkel Pinter reports financial support was provided by Israel Science Foundation. Moran Frenkel Pinter reports was provided by Azrieli Foundation. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Nir Ben-Tal and Moran Frenkel Pinter are supported by the Minerva Foundation. Nir Ben-Tal's research is supported by the Abraham E. Kazan Chair in Structural Biology, Tel Aviv University. Moran Frenkel Pinter is supported by the Azrieli Foundation Early Career Faculty Grant, the Israel Science Foundation (ISF) grant (1611/22), and the FEBS Foundation Excellence Award. Rachel Kolodny and Nir Ben-Tal are supported by ISF grant 1764/21.

Data availability

No data was used for the research described in the article.

References

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

* of special interest
** of outstanding interest

1. Gilbert W: **Origin of life: the RNA world**. *Nature* 1986, **319**:618.-618.

2. Fried SD, Fujishima K, Makarov M, Cherepashuk I, Hlouchova K: **Peptides before and during the nucleotide world: an origins story emphasizing cooperation between proteins and nucleic acids**. *J R Soc Interface* 2022, **19**, 20210641.

3. ** Kocher CD, Dill KA: **Origins of life: the protein folding problem all over again?** *Proc Natl Acad Sci U S A* 2024, **121**, e2315000121.
The authors highlight parallels between the problem of how proteins emerged and the problem of how to find the minimal energy conformation of a protein, the so-called protein folding problem.

4. Kessel A, Ben-Tal N: *From molecules to cells: the origin of life on Earth*. 2022.

5. ** Kocher CD, Dill KA: **The prebiotic emergence of biological evolution**. *R Soc Open Sci* 2024, **11**, 240431.
An abstract kinetic model based on formation of peptides with cooperative autocatalytic activity. The model is framed as a set of coupled differential equations, and for some model parameters persistent propagation, required for evolution, is not rare.

6. Eck RV, Dayhoff MO: **Evolution of the structure of ferredoxin based on living relics of primitive amino acid sequences**. *Science* 1966, **152**:363–366.

7. Lupas AN, Ponting CP, Russell RB: **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* 2001, **134**:191–203.

8. Frenkel-Pinter M, Frenkel-Pinter M, Samanta M, Ashkenasy G, Leman LJ, Leman LJ: **Prebiotic peptides: molecular hubs in the origin of life**. *Chem Rev* 2020, **120**:4707–4765.

9. Dickerson RE. *Chemical evolution and the origin of life*, **239**; 1978:70–87.

10. Brack A: **Selective emergence and survival of early polypeptides in water**. In *Origins of life*, **17**; 1987:367–379.

11. Ruiz-Mirazo K, Briones C, De La Escosura A: **Prebiotic systems chemistry: new perspectives for the origins of life**. In *Chemical reviews*, **114**; 2014:285–366.

12. ** Frenkel-Pinter M, Bouza M, Fernández FM, Leman LJ, Williams LD, Hud NV, Guzman-Martinez A: **Thioesters provide a plausible prebiotic path to proto-peptides**. *Nat Commun* 2022, **13**.
In this paper a robust prebiotically plausible pathway for peptide formation using mercaptoacids was found. Thiodepsipeptide formation was achieved under a wide range of pH and temperature by simply drying and heating mercaptoacids with amino acids. These results support the hypothesis that thiodepsipeptides and thiol-terminated peptides formed readily on prebiotic Earth.

13. Frenkel-Pinter M, Jacobson KC, Eskew-Martin J, Forsythe JG, Grover MA, Williams LD, Hud NV: **Differential oligomerization of alpha versus beta amino acids and hydroxy acids in abiotic proto-peptide synthesis reactions**. *Life* 2022, **12**.

14. Wolman Y, Haverland WJ, Miller SL: **Nonprotein amino acids from spark discharges and their comparison with the murchison meteorite amino acids**. *Proc Natl Acad Sci USA* 1972, **69**:809–811.

15. Roy RS, Gehring AM, Milne JC, Belshaw PJ, Walsh CT: **Thiazole and oxazole peptides: biosynthesis and molecular machinery**. *Nat Prod Rep* 1999, **16**:249–263.

16. Mittapalli GK, Reddy KR, Xiong H, Munoz O, Han B, De Riccardis F, Krishnamurthy R, Eschenmoser A: **Mapping the landscape of potentially primordial informational oligomers: oligodipeptides and oligodipeptoids tagged with triazines as recognition elements**. *Angew Chem Int Ed* 2007, **46**:2470–2477.

17. Eschenmoser A, Eschenmoser A: **Etiology of potentially primordial biomolecular structures: from vitamin B12 to the nucleic acids and an inquiry into the chemistry of life's origin: a retrospective**. *Angew Chem Int Ed* 2011, **50**:12412–12472.

18. Chandru K, Jia TZ, Mamajanov I, Bapat N, Cleaves HJ: **Prebiotic oligomerization and self-assembly of structurally diverse xenobiological monomers**. *Sci Rep* 2020, **10**:1–14. 10:1 2020.

19. Fialho DM, Roche TP, Hud NV: **Prebiotic syntheses of noncanonical nucleosides and nucleotides**. In *Chemical reviews*, **120**. American Chemical Society; 2020:4806–4830.

20. Fialho DM, Karunakaran SC, Greeson KW, Martínez I, Schuster GB, Krishnamurthy R, Hud NV: **Depsipeptide nucleic acids: prebiotic formation, oligomerization, and self-assembly of a new proto-nucleic acid candidate**. *J Am Chem Soc* 2021, **143**:13525–13537.

21. Forsythe JG, Yu S-S, Mamajanov I, Grover MA, Krishnamurthy R, Fernández FM, Hud NV: **Ester-mediated amide bond formation driven by wet-dry cycles: a possible path to polypeptides on the prebiotic Earth**. *Angew Chem* 2015, **127**:10009–10013.

22. Chen-Yu Chien S, Yu S-S, Clarence T, Wong CT, P Ng DK: **Ester-mediated peptide formation promoted by deep eutectic solvents: a facile pathway to proto-peptides**. *Chem Commun* 2020, **56**:11949–11952.

23. Pengthong P, Bopp PA, Jungsuttiwong S, Nanok T: **Mechanistic insights into the self-esterification of lactic acid under neutral and acidic conditions**. *J Mol Struct* 2023:1273.

24. Mamajanov I, Callahan MP, Dworkin JP, Cody GD: **Prebiotic alternatives to proteins: structure and function of hyperbranched polyesters**. *Orig Life Evol Biosph* 2015, **45**:123–137.

25. Rich A: **The possible participation of esters as well as amides in prebiotic polymers**. In *Chemical evolution and the origin of life; third international conference*; 1971.

26. Chandru K, Mamajanov I, Cleaves HJ, Jia TZ: **Polyesters as a model system for building primitive biologies from non-biological prebiotic chemistry**. *Life* 2020, **10**.

27. Matsuo M, Kurihara K: **Proliferating coacervate droplets as the missing link between chemistry and biology in the origins of life**. *Nat Commun* 2021, **12**.

28. Jia TZ, Chandru K, Hongo Y, Afrin R, Usui T, Myojo K, Cleaves HJ: **Membraneless polyester microdroplets as primordial compartments at the origins of life**. *Proc Natl Acad Sci USA* 2019, **116**:15830–15835.

29. Frenkel-Pinter M, Haynes JW, Martin C, Petrov AS, Burcar BT, Krishnamurthy R, Hud NV, Leman LJ, Williams LD: **Selective incorporation of proteinaceous over nonproteinaceous cationic amino acids in model prebiotic oligomerization reactions**. *Proc Natl Acad Sci USA* 2019, **116**:16338–16346.

30. Forsythe JG, Petrov AS, Millar WC, Yu S-S, Krishnamurthy R, Grover MA, Hud NV, Fernández FM: **Surveying the sequence**

diversity of model prebiotic peptides by mass spectrometry. *Proc Natl Acad Sci USA* 2017, **114**:E7652–E7659.

31. Yu S-S, Krishnamurthy R, Fernández FM, Hud NV, Schork FJ, Grover MA: **Kinetics of prebiotic depsipeptide formation from the ester–amide exchange reaction**. *Phys Chem Chem Phys* 2016, **18**:28441–28450.

32. Frenkel-Pinter M, Sargon AB, Glass JB, Hud NV, Williams LD: **Transition metals enhance prebiotic depsipeptide oligomerization reactions involving histidine**. *RSC Adv* 2021, **11**: 3534–3538.

33. Zeng M, Tao J, Xu S, Bai X, Zhang H: **Marine organisms as a prolific source of bioactive depsipeptides**. *Mar Drugs* 2023, **21**:120. -120.

34. Casalme LO, Katayama K, Hayakawa Y, Nakamura K, Yamauchi A, Nogata Y, Yoshimura E, Matsuda F, Umezawa T: **Effective synthesis and antifouling activity of dolastatin 16 derivatives**. *Mar Drugs* 2022, **20**.

35. Alonzo DA, Schmeing TM: **Biosynthesis of depsipeptides, *or* Depsi: the peptides with varied generations**. *Protein Sci* 2020, **29**:2316–2347.

36. Wang X, Gong X, Li P, Lai D, Zhou L: **Structural diversity and biological activities of cyclic depsipeptides from fungi**. *Molecules* 2018, **23**:169.

37. Muchowska KB, Varma SJ, Moran J: **Nonenzymatic metabolic reactions and life's origins**. In *Chemical reviews*, **120**. American Chemical Society; 2020:7708–7744.

38. ** Zhang L, Ying J: **Amino acid analogues provide multiple plausible pathways to prebiotic peptides**. *J R Soc Interface* 2024, **21**. Edited by: Royal Society Publishing.
This recent review discusses the reported pathways of polypeptide formation and amino acid analogues that likely existed on early Earth and participated in the formation of primordial peptides. The authors also discuss potential co-evolution between amino acids and their analogues.

39. Marahiel MA: **Working outside the protein-synthesis rules: insights into non-ribosomal peptide synthesis**. *J Pept Sci* 2009, **15**:799–807. Edited by.

40. Cronin JR, Pizzarello S, Cooper GW: **Characteristics and formation of amino acids and hydroxy acids of the Murchison meteorite**. *Adv Space Res* 1994, **15**:91–97.

41. Bada JL: **New insights into prebiotic chemistry from Stanley Miller's spark discharge experiments**. *Chem Soc Rev* 2013, **42**:2186.

42. Seebach D, Beck AK, Bierbaum DJ: **The world of β- and γ-peptides comprised of homologated proteinogenic amino acids and other components**. *Chem Biodivers* 2004, **1**: 1111–1239.

43. Forsythe JG, English SL, Simoneaux RE, Weber AL: **Synthesis of β-peptide standards for use in model prebiotic reactions**. *Orig Life Evol Biosph* 2018, **48**:201–211.

44. ** Matange K, Rajaei V, Capera-Aragonès P, Costner JT, Robertson A, Kim JS, Petrov AS, Bowman JC, Williams LD, Pinter MF: *Evolution of complex chemical mixtures reveals combinatorial compression and population synchronicity*. 2023.
In this paper, an experimental platform to study the evolution of complex chemical systems has been established, as well as some general guidelines for conducting chemical evolution. It was found that when subjecting complex chemical mixtures to wet-dry cycling at low temperatures (45 °C), the chemical composition of the system continuously changes during each of the 15 investigated wet-dry cycles. Moreover, the system demonstrates chemical selection, during which combinatorial explosion is avoided, and maintains synchronicity of molecular sub-populations.

45. Ajit Varki RDC, Esko Jeffrey D, Stanley Pamela, Hart Gerald W, Aebi Markus, Mohnen Debra, Kinoshita Taroh, Packer Nicolle H, Prestegard James H, Schnaar Ronald L, Seeberger Peter H: *Essentials of glycobiology. 4th* ed. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 2022.

46. Müller F, Escobar L, Xu F, Węgrzyn E, Nainytė M, Amatov T, Chan CY, Pichler A, Carell T: **A prebiotically plausible scenario of an RNA–peptide world**. *Nature* 2022, **605**:279–284.

47. ** Welsch F, Kervio E, Tremmel P, Richert C: **Prolinyl nucleotides drive enzyme-free genetic copying of RNA**. *Angew Chem* 2023, **135**.
In the papar it was demonstated that prolinyl nucleotides with a phosphoramidate linkage are active building blocks in enzyme- and ribozyme-free copying of RNA in the presence of monosubstituted imidazoles as organocatalyst.

48. Suárez-Marina I, Abul-Haija YM, Turk-Macleod R, Gromski PS, Cooper GJT, Olivé AO, Colón-Santos S, Cronin L: **Integrated synthesis of nucleotide and nucleosides influenced by amino acids**. *Commun Chem* 2019, **2**.

49. Joshi MP, Sawant AA, Rajamani S: **Spontaneous emergence of membrane-forming protoamphiphiles from a lipid–amino acid mixture under wet–dry cycles**. *Chem Sci* 2021, **12**:2970–2978.

50. Lancet D, Segrè D, Kahana A: **Twenty years of "lipid world": a fertile partnership with David Deamer**. *Life* 2019, **9**:77.

51. ** Lago I, Black L, Wilfinger M, Maurer SE: **Synthesis and characterization of amino acid decyl esters as early membranes for the origins of life**. *Membranes* 2022, **12**:858.
In the papar the formation of amino acids decyl esters was demonstrated by simple dehydration reactions of decanol and various amino acids. Products were analyzed using LCMS and microscopy. All amino acids tested produced decyl esters, and some of the products formed membranes when rehydrated in water.

52. Kessel A, Ben-Tal N: *Introduction to proteins: structure, function, and motion*. edn 2. CRC Press; 2018.

53. Jayaraman V, Toledo-Patiño S, Noda-García L, Laurino P: **Mechanisms of protein evolution**. *Protein Sci* 2022, **31**:e4362.

54. Zheng Z, Goncearenco A, Berezovsky IN: **Back in time to the Gly-rich prototype of the phosphate binding elementary function**. *Curr Res Struct Biol* 2024, **7**, 100142.

55. Kolodny R: **Searching protein space for ancient sub-domain segments**. *Curr Opin Struct Biol* 2021, **68**:105–112.

56. Nepomnyachiy S, Ben-Tal N, Kolodny R: **Global view of the protein universe**. *Proc Natl Acad Sci USA* 2014, https://doi.org/10.1073/pnas.1403395111.

57. Alva V, Remmert M, Biegert A, Lupas AN, Söding J: **A galaxy of folds**. *Protein Sci* 2010, **19**:124–130.

58. Berger B: **Algorithms for protein structural motif recognition**. *J Comput Biol* 1995, **2**:125–138.

59. Nepomnyachiy S, Ben-Tal N, Kolodny R: **Complex evolutionary footprints revealed in an analysis of reused protein segments of diverse lengths**. *Proc Natl Acad Sci USA* 2017, **114**: 11703–11708.

60. Alva V, Söding J, Lupas AN: **A vocabulary of ancient peptides at the origin of folded proteins**. *Elife* 2015, **4**, e09410.

61. Kolodny R, Nepomnyachiy S, Tawfik DS, Ben-Tal N: **Bridging themes: short protein segments found in different architectures**. *Mol Biol Evol* 2021, **38**:2191–2208.

62. Qiu K, Ben-Tal N, Kolodny R: **Similar protein segments shared between domains of different evolutionary lineages** 2022, **31**, e4407.

63. Longo LM, Jabłońska J, Vyas P, Kanade M, Kolodny R, Ben-Tal N, Tawfik DS: **On the emergence of P-Loop NTPase and Rossmann enzymes from a Beta-Alpha-Beta ancestral fragment**. *Elife* 2020, **9**, e64415.

64. Narunsky A, Kessel A, Solan R, Alva V, Kolodny R, Ben-Tal N: **On the evolution of protein–adenine binding**. *Proc Natl Acad Sci USA* 2020, **117**:4701–4709.

65. Raanan H, Poudel S, Pike DH, Nanda V. *Small protein folds at the root of an ancient metabolic network*, **117**; 2020:7193–7199.

66. Bromberg Y, Aptekmann AA, Mahlich Y, Cook L, Senn S, Miller M, Nanda V, Ferreiro DU, Falkowski PG: **Quantifying structural relationships of metal-binding sites suggests origins of biological electron transfer**. *Sci Adv* 2022, **8**, eabj3984.

67. Gruic-Sovulj I, Longo LM, Jabłońska J, Tawfik DS: **The evolutionary history of the HUP domain**. *Crit Rev Biochem Mol Biol* 2022, **57**:1–15.

68. Longo LM, Petrović D, Kamerlin SCL, Tawfik DS: **Short and simple sequences favored the emergence of N-helix phospho-ligand binding sites in the first enzymes**. *Proc Natl Acad Sci USA* 2020, **117**:5310–5318.

69. Remmert M, Biegert A, Linke D, Lupas AN, Söding J: **Evolution of outer membrane β-barrels from an ancestral ββ hairpin**. *Mol Biol Evol* 2010, **27**:1348–1358.

70. Alvarez-Carreño C, Penev PI, Petrov AS, Williams LD: **Fold evolution before LUCA: common ancestry of SH3 domains and OB domains**. *Mol Biol Evol* 2021, **38**:5134–5143.

71. Youkharibache P, Veretnik S, Li Q, Stanek KA, Mura C, Bourne PE: **The small β-barrel domain: a survey-based structural analysis**. *Structure* 2019, **27**:6–26.

72. Mura C, Veretnik S, Bourne PE: **The Urfold: structural similarity just above the superfold level?** *Protein Sci* 2019, **28**: 2119–2126.

73. ** Draizen EJ, Veretnik S, Mura C, Bourne PE: **Deep generative models of protein structure uncover distant relationships across a continuous fold space**. *Nat Commun* 2024, **15**:8094.
DeepUrFold is a deep generative modeling framework for studying the relationships between protein sets. It learns a high-dimensional embedding for proteins, and generating a mostly-continuous view of fold space. This allows studying the relationships between proteins by anlayzing the relationships among their embeddings.

74. Forslund SK, Kaduk M, Sonnhammer ELL: **Evolution of protein domain architectures**. In *Evolutionary genomics: statistical and computational methods.* Edited by Anisimova M, New York: Springer; 2019:469–504, https://doi.org/10.1007/978-1-4939-9074-0_15.

75. ** Smug BJ, Szczepaniak K, Rocha EP, Dunin-Horkawicz S, Mostowy RJ: **Ongoing shuffling of protein fragments diversifies core viral functions linked to interactions with bacterial hosts**. *Nat Commun* 2023, **14**:7460.
The authors study the use of protein building blocks in bacteriophages. Relying on sensitive tools for homology detection they study instances where fragments are shared between two otherwise unrelated proteins. It beutifully demonstrates how insights can be derived by data-mining the protein databases.

76. * Kutlu Y, Axel G, Kolodny R, Ben-Tal N, Haliloglu T: **Reused protein segments linked to functional dynamics**. *Mol Biol Evol* 2024, **41**, msae184.
Proteins are dynamic entities that must undergo conformational changes to function properly, e.g., between active and inactive states. This work studies the correlation between themes and equilibrium dynamics, approximated using a form of normal modes analysis called Gaussian network model (GNM). They show a correlation between the motion decomposition into dynamic elements and the themes decomposition which is based on reuse. This highlights the importance of dynamic match between building blocks in 'evolutionary design' of proteins.

77. Chakravarty D, Sreenivasan S, Swint-Kruse L, Porter LL: **Identification of a covert evolutionary pathway between two protein folds**. *Nat Commun* 2023, **14**:3177.

78. * Longo LM, Kolodny R, McGlynn SE: **Evidence for the emergence of β-trefoils by 'peptide Budding'from an IgG-like β-sandwich**. *PLoS Comput Biol* 2022, **18**, e1009833.
The authors address the mystery of β-trefoils emergence, a relatively young fold that was even rumored to have emerged *de novo*. Running a directed search in protein space found similarities to segments in IgG-like β-sandwich. Thus they propose that β-trefoils emerged by stiching together parts from the very old IgG-like β-sandwich, in a process they called 'budding'.

79. ** Alvarez-Carreño C, Gupta RJ, Petrov AS, Williams LD: **Creative destruction: new protein folds from old**. *Proc Natl Acad Sci USA* 2022, **119**, e2207897119.
The authors focus on three distinct β-barrel folds that are found in the translation machinery, and as such are considered to be fossils of their ancient ancestors. By analyzing the relationships amongst them, leads to a proposed mechanism for early diversification that explains how old folds spawn new ones.

80. ** Yagi S, Tagami S: **An ancestral fold reveals the evolutionary link between RNA polymerase and ribosomal proteins**. *Nat Commun* 2024, **15**:5938.
The authors use protein engineering to propose mechanisms for the diversification of the ancient and distinct β-barrel folds. They reconstruct a metamorphic fold that is no longer found within contemporary protein space, DZBB (double zeta β-barrel). They show a plausible diversification from DZBB into these folds.