










Redefining the Limits of Functional Continuity in the Early Evolution of P-Loop NTPases

Andrey O. Demkiv ¹, Saacnicteh Toledo-Patiño ^{2,3}, Encarnación Medina-Carmona ⁴,
Andrej Berg ¹, Gaspar P. Pinto ¹, Antonietta Parracino ¹, Jose M. Sanchez-Ruiz ⁴,
Alvan C. Hengge ⁵, Paola Laurino ^{2,6}, Liam M. Longo ^{7,8,*},
Shina Caroline Lynn Kamerlin ^{9,10,*}

¹Department of Chemistry—BMC, Uppsala University, Uppsala S-751 23, Sweden

²Protein Engineering and Evolution Unit, Okinawa Institute of Science and Technology, Graduate University (OIST), Okinawa 904-0495, Japan

³Molecular Bioengineering Group, Okinawa Institute of Science and Technology, Graduate University (OIST), Okinawa 904-0495, Japan

⁴Departamento de Química Física, Facultad de Ciencias, Unidad de Excelencia de Química aplicada a Biomedicina y Medioambiente (UEQ), Universidad de Granada, Granada 18071, Spain

⁵Department of Chemistry and Biochemistry, Utah State University, Logan, UT 84322-0300, USA

⁶Institute for Protein Research, Osaka University, Suita, Japan

⁷Blue Marble Space Institute of Science, Seattle, WA 98104, USA

⁸Earth-Life Science Institute, Institute of Science Tokyo, Tokyo 152-8550, Japan

⁹School of Chemistry and Biochemistry, Georgia Institute of Technology, Atlanta, GA 30332, USA

¹⁰Department of Chemistry, Lund University, Box 124, Lund 22100, Sweden

*Corresponding author: E-mails: llongo@elsi.jp, skamerlin3@gatech.edu.

Associate editor: Banu Ozkan

Abstract

At the heart of many nucleoside triphosphatases is a conserved phosphate-binding sequence motif. A current model of early enzyme evolution proposes that this six to eight residue motif could have sparked the emergence of the very first nucleoside triphosphatases—a striking example of evolutionary continuity from simple beginnings, if true. To test this provocative model, seven disembodied Walker A-derived peptides were extensively computationally characterized. Although dynamic flickers of nest-like conformations were observed, significant structural similarity between the situated peptide and its disembodied counterpart was not detected. Simulations suggest that phosphate binding is nonspecific, with a preference for GTP over orthophosphate. Control peptides with the same amino acid composition but different sequences and situated conformations behaved similarly to the Walker A peptides, revealing no indication that the Walker A sequence is privileged as a disembodied peptide. We conclude that the evolutionary history of the P-loop NTPase family is unlikely to have started with a disembodied Walker A peptide in an aqueous environment. The limits of evolutionary continuity for this protein family must be reconsidered. Finally, we argue that motifs such as the Walker A motif may represent incomplete or fragmentary molecular fossils—the true nature of which has been eroded by time.

Keywords: Walker A motif, P-loop NTPase, phosphate-binding loop, molecular fossil, primitive proteins

Introduction

P-loop NTPases are the archetypical ancient protein family: Ubiquitously distributed across the tree of life, these domains have been recruited to catalyze phosphorylation reactions and to couple large molecular motions to the synthesis or hydrolysis of condensed phosphates (Aravind et al. 2002; Winstanley et al. 2005; Ma et al. 2008; Bukhari and Caetano-Anollés 2013; Longo et al. 2020a). At the heart of the P-loop fold lies a primitive active site in which the α - and β -phosphate moieties of an NTP molecule rest atop a crown or nest of backbone amides (Watson and Milner-White 2002; Bianchi et al. 2012; Milner-White 2019). This nest is formed by the N-terminus of an α -helix (Longo et al. 2020b) and the preceding loop, a contiguous phosphate-binding fragment dubbed the Walker A motif (Walker et al. 1982). The Walker A motif is highly conserved across P-loop NTPases and is characterized by the glycine-rich sequence GxxGxGK[T/S]. Guided by structural bioinformatics and laboratory-characterized model peptides, we and others

have argued that these core features—a $\beta\alpha$ or $\beta\alpha\beta$ motif bearing a glycine-rich phosphate binding loop—seeded the emergence of P-loop NTPases (Alva et al. 2015; Laurino et al. 2016; Romero Romero et al. 2018; Longo et al. 2020b; Vyas et al. 2021, 2023), as well as several other cofactor-associated domains, including the Rossmann (Lupas et al. 2001; Longo et al. 2020a), HUP (Gruic-Sovulj et al. 2022), and Nat/Ivy (Longo et al. 2022) evolutionary lineages.

The profound conservation and functional importance of the Walker A motif has led some to consider whether the motif may predate the $\beta\alpha$ element within which it is situated, hypothesizing that some aspect of structural and functional continuity within the P-loop NTPase family extends back to a primordial peptide of just six to eight residues in length (Bianchi et al. 2012; Milner-White 2019). The notion that folded domains may have evolved from short peptides is not without precedence: Dayhoff's study of sequence duplications within ferredoxin (Eck and Dayhoff 1966) and Dutton's subsequent demonstration

Received: October 9, 2024. Revised: January 21, 2025. Accepted: February 14, 2025

© The Author(s) 2025. Published by Oxford University Press on behalf of Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

of a metal cluster-supporting peptide maquette (Gibney et al. 1996) have inspired research into the early evolution of cluster-associated protein folds (Timm et al. 2023) that, taken together, supports this view, albeit for a different class of enzymes.

In 2012, Bianchi and coworkers reported that the hexapeptide SGAGKT bound phosphate with microscopic K_D values ranging from about 10 μ M to 1 mM (Bianchi et al. 2012). Although their methodology was unable to probe binding of backbone amides, the authors hypothesized that a stable peptide-phosphate complex could be indicative of a nest-like conformation reminiscent of contemporary enzymes (Bianchi et al. 2012; Milner-White 2019). This result—that a truncated, disembodied Walker A peptide retained significant affinity for phosphate—plus the hypothesized nest formation was an extreme example of functional continuity and would go on to shape thinking about the evolutionary trajectory of P-loop NTPases for over a decade (Gruber et al. 2014; Russell et al. 2014; Romero Romero et al. 2018; Shalaeva et al. 2018; Pal et al. 2020; Wang and Hecht 2020; Fowler et al. 2021; Carvalho et al. 2022; Fried et al. 2022). Yet, the relatively modest interaction energies accessible to phosphate binding, as opposed to the coordinate covalent Fe-S bonds of maquettes, would seem insufficient to drive structuring of a short glycine-rich peptide. Further, the significance of short nucleotide-binding peptides (Barany and Merrifield 1973; Hilpert et al. 2010; Kang et al. 2015; Kroiss et al. 2019)—where interactions to the base may dominate—is unclear. Do Walker A-derived peptides have an intrinsic conformational preference that would support binding as nests? Are the interaction energies associated with nest binding modes sufficient to induce peptide structuring?

The question of whether a Walker A peptide sparked the emergence of P-loop NTPases has significant implications for the evolution of anion-binding domains more generally, several of which host similar loop-nest structures (Watson and Milner-White 2002). If true, (di)nucleotide binding domains would become quintessential examples of molecular “simple beginnings”—complex structures that condensed around low-complexity evolutionary nuclei. Thus, clarifying the extent to which a disembodied Walker A motif adopts a nest-like conformation upon binding to phosphate is fundamental to understanding the early evolutionary history of this fold and others. To this end, we have performed detailed characterization of the dynamic properties of 7 representative Walker A motifs, as well as the previously reported hexapeptide SGAGKT (Bianchi et al. 2012) and a number of control sequences, by Hamiltonian replica exchange molecular dynamics (HREX-MD) simulations, to search for nest formation.

We find that the situated Walker A motif can be surprisingly independent, with some structures forming only a few interactions with the surrounding protein; and rigid, adopting a narrow distribution of conformations and undergoing only minor adjustments upon ligand binding. The disembodied motif, on the other hand, does not adopt nest-like conformations in MD simulations in aqueous solution, though some intrinsic conformational preferences do exist. Uncorrelated preferences for nest-like dihedrals were observed, particularly of the glycine residues, but these preferences were modest. Inclusion of orthophosphate or GTP did not significantly induce nest formation, though GTP had a notably larger effect on the conformation of the peptide than orthophosphate, promoting bent peptide conformations. The GTP-bound bend conformations involved interactions with the entire nucleotide, not just the triphosphate group, suggesting that, from a biophysical

standpoint, nucleotide binding may be a more accessible function than phosphate binding. Further, a series of natural loops with the same amino acid composition but distinct evolutionary histories and situated loop conformations (control loops) were found to behave comparably to Walker A peptides with respect to conformational dynamics and ligand binding. Finally, simulations performed in methanol, which has a lower dielectric constant than water, enhanced the formation of nest-like dihedrals in the presence and absence of ligands.

Despite some flickers of nest-like conformations, we conclude that the Walker A motif—if it ever existed as a *free peptide in an aqueous environment*—is unlikely to have had significant nest-forming potential, even in the presence of phosphate or phospho-ligands. Moreover, the comparison to control loops indicates that a disembodied Walker A peptide is not a uniquely preferable solution to phosphate binding or nest formation under these conditions. Models of early P-loop evolution must be revised. Finally, we argue that great care must be taken when interpreting the significance of highly conserved sequence fragments, which may represent eroded or incomplete molecular fossils.

Results

The Situated Walker A Loop Forms a Rigid, Preorganized Nest for Phosphate Binding

The Walker A nest (Fig. 1a and b) is situated between $\beta 1$ and $\alpha 1$ of a P-loop NTPase domain. Nucleotide binding is dominated by backbone interactions, with up to six backbone amides forming hydrogen bonds to the α and β phosphate moieties of the ligand. The Walker A nest wraps around these phosphate moieties like a hand, rather unlike the phosphate binding loop of the Rossmann fold, which forms a comparatively flat binding surface. As binding occurs at the N-terminus of an α -helix, presumably the α -helix dipole contributes to binding as well (Hol et al. 1978). The primary side chain interaction is a salt bridge between a phosphate moiety and the side chain of Lys7, though the details of this interaction depend on whether a diphosphate or triphosphate is bound. Many P-loop NTPases also bind a Mg^{2+} dication (Fig. 1b, green spheres, supplementary table S1, Supplementary Material online), which is chelated by Ser/Thr8 and forms interactions with the phospho-ligand (but, critically, not the nucleophilic water molecule; i.e. it is not involved directly in the chemical mechanism of catalysis [Kozlova et al. 2022]). Nucleotide binding, however, is not strictly Mg^{2+} dependent and, in structures without a dication, the side chain hydroxyl of Ser/Thr8 can interact directly with the ligand (as in e2gksA2 or e1k6mA1, see also supplementary table S1, Supplementary Material online for a distribution of contacts). The backbone dihedrals of residues 4 to 7 of the Walker A nest are characterized by an alternation between the right-handed α -helix (α_R) and left-handed α -helix (α_L) regions of the Ramachandran plot (Fig. 1c). From residue 7 onward, the structure is a canonical right-handed α -helix. Unsurprisingly, the positions that adopt an α_L conformation are either exclusively (position 6) or mostly (position 4) Gly (Fig. 1d).

Nucleotide binding does not dramatically alter the backbone conformation of the nest, in agreement with the observation that the phosphate-binding nests of P-loop NTPases tend to be relatively rigid (Saraste et al. 1990; Smith and Rayment 1996; Taberner et al. 2008; Pálffy et al. 2020; Crean et al. 2021). Both the structural overlay in Fig. 1b and the Ramachandran plot in Fig. 1c include liganded and

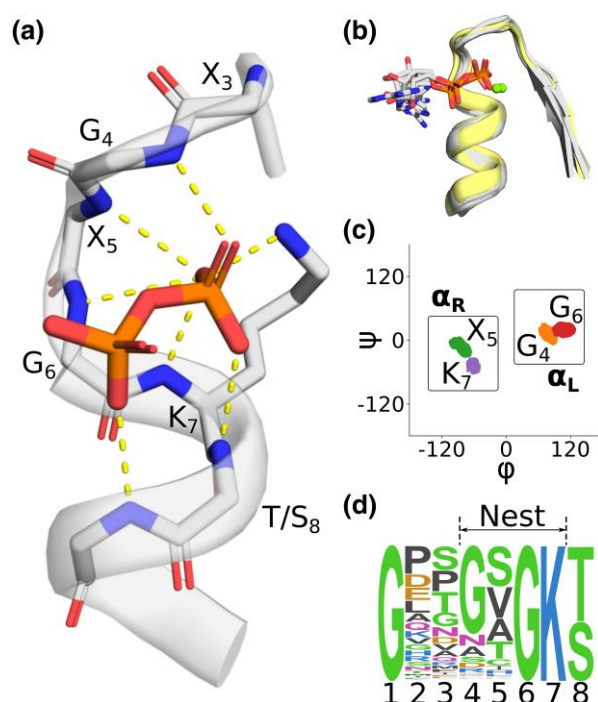


Fig. 1. The phosphate binding nest of P-Loop NTPases. a) A representative phosphate binding nest, with 6 backbone amides and one lysine side chain oriented toward and interacting (dashed lines) with the pyrophosphate moiety of ADP. ECOD domain e3k1jA2. b) Structural context of the Walker A motifs analyzed in this study. Structures with a bound phospho-ligand are colored gray; unbound structures are colored yellow. Mg^{2+} ions are drawn as green spheres and are present in only a subset of structures. Mg^{2+} ions were not included in our simulations, as described in the main text. ECOD domains: e3ievA2, e2it1A8, e3k1jA2, e5jrjA2, e6ojxA1, e2qy9A2. Note that the γ phosphate of the bound ATP was removed for clarity in e6ojxA1. c) Ramachandran plot showing the per residue distribution of ϕ , ψ angles, including both liganded and unliganded structures. Boxes indicate the regions taken to be α_R and α_L in subsequent analyses. Note the alternating α_L - α_R - α_L - α_R character of the nest. d) Sequence logo of the Walker A Motif demonstrating the canonical GxxGxGK(T/S) sequence pattern. P-loop NTPase sequences were gathered from the PDB and filtered for 90% identity at the domain level.

unliganded structures, as well as structures with and without a bound dication. A breakdown of ϕ , ψ angles for liganded and unliganded nests (supplementary fig. S1, Supplementary Material online) shows that the dihedral distributions are overlapping for every residue of the putative nest. The largest adjustments relate to ϕ_6 , which has a difference between the means of 14 degrees. We conclude, then, that the backbone of the Walker A nest adopts a relatively rigid structure with complete or near-complete pre-organization for ligand binding. Stabilization of this conformation is likely due to a combination of intrinsic structural preference and supportive external interactions—the balance of which has implications for the potential adoption of nest-like structures by a disembodied Walker A peptide. An analysis of supportive interactions (supplementary fig. S2, Supplementary Material online) revealed that while many Walker A motifs can make extensive contacts with the surrounding protein, there are also structures that make relatively few supportive interactions (2 to 15 contacts, on average 8; excluding backbone hydrogen bonds propagating the α -helix at the C-terminus and interactions to adjacent residues). Moreover, the specific supportive interactions between the phosphate binding loop and amino

acids that are noncontiguous with the Walker A motif are not strictly conserved. Ultimately, we conclude that the Walker A structure is not likely to be encoded primarily through supportive interactions to interior residues of the motif, and other factors likely maintain the Walker A structure. For example, conformational constraints conferred by the angle and distance between the residues that anchor the Walker A motif to the rest of the protein, and/or enhanced ability of the Walker A motif to adopt the canonical conformation in a more hydrophobic environment.

A Walker A-Derived Hexapeptide Does Not Form Nests in Aqueous Solution

To better understand the conformational dynamics and evolutionary history of the Walker A motif, we analyzed the hexapeptide SGAGKT, which was previously hypothesized to form nests upon interaction with phosphate (Watson and Milner-White 2002; Bianchi et al. 2012; Milner-White 2019), by HREX-MD simulations. Simulations were carried out with and without ligands (HPO_4^{2-} and GTP) for 1 μs and comprising eight replicas, with a temperature range of 300 to 450 K per condition (see the Materials and Methods for detailed information). We note here two important factors in our analysis: First, although primarily an ATP binding sequence (Walker et al. 1982), both the Walker A and Walker B motifs occur in GTP binding proteins, for example, elongation factor Tu and Ras GTPase (Georgiadis et al. 1992; Fry et al. 2022). Indeed, initial studies of P-loop NTPase nests focused on GTPases (Bianchi et al. 2012; Maggiolo et al. 2023). This apparent interchangeability stems from the fact that the situated Walker A peptide does not participate in base binding. Second, although phosphate binding of the Walker A motif typically involves metal dications, these metals are not required for ligand binding and are not a key determinant of nest formation (supplementary fig. S3, Supplementary Material online). Consistent with the first experimental characterization of the Walker A-derived hexapeptide (Bianchi et al. 2012), we have not included metal dications in our analysis.

We find that the end-to-end distance (Fig. 2a) and bend angle (Fig. 2b) distributions of the SGAGKT peptide are both broad, generally indicative of a flexible peptide. Nevertheless, a degree of conformational preference is apparent, with peaks corresponding to a tight hairpin (end-to-end distance of about 5 Å and a bend angle of about 45°). A time-lagged independent component analysis (tICA) of the MD trajectories (Fig. 2c) reveals 6 primary clusters, the highest occupancy of which involves the peptide folding back onto itself, consistent with the end-to-end distance and bend angle distributions.

Although the *global* similarity to the situated conformation is low (Fig. 2a to c), there is a modest preference for uncorrelated nest-like dihedral angles (Fig. 2f and supplementary fig. S4, Supplementary Material online). Position 6 (Gly in the hexapeptide; motif numbering scheme in Fig. 1d) has a greater propensity for α_L backbone dihedrals than do positions 5 and 7, as in the situated conformation. In addition, positions 5 and 7 have a modest preference for the α_R backbone dihedral, once again reflecting the situated conformation. Correlated nest-like conformations between residues (Fig. 2d and supplementary table S2, Supplementary Material online) are rare but not strictly absent from the simulations, with residues 5 to 7 having the greatest propensity to form nest-like regions, albeit with an occupancy of less than 1%.

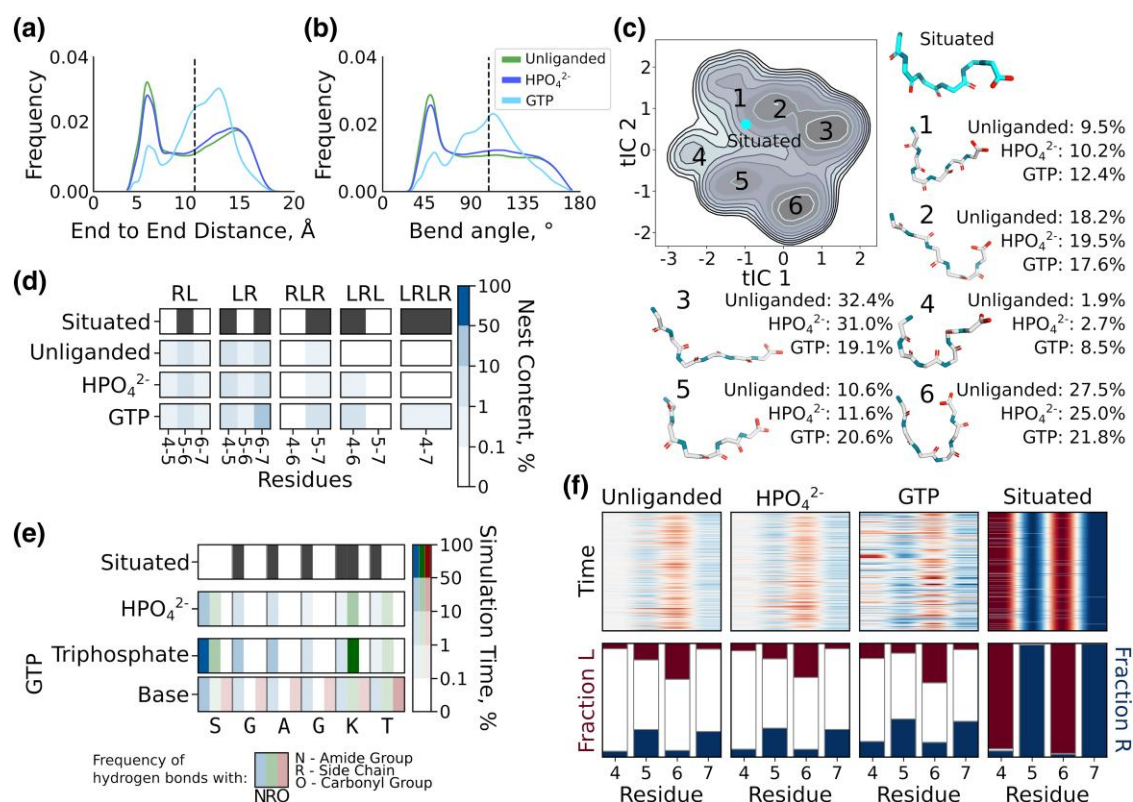


Fig. 2. Conformational dynamics of the Walker A-derived hexapeptide SGAGKT. a) Distribution of $C_{\alpha 1}$ - $C_{\alpha 6}$ distances for the free peptide in the presence and absence of a ligand. b) Distribution of $C_{\alpha 1}$ - N_4 - $C_{\alpha 6}$ angles for the free peptide in the presence and absence of a ligand. The situated conformation is indicated with a dotted line in a and b. c) tICA projection derived from the dihedrals of all hexapeptide conformations, with and without ligand. The occupancy and a representative structure are provided for each of the six main clusters. The situated conformation is indicated by a cyan dot. d) Occurrence of correlated stretches of α_L and α_R conformations in each of our simulations, broken down per residue. e) Interaction profile of the peptide with a ligand. f) Uncorrelated preference of α_R or α_L backbone dihedrals in the presence and absence of phosphate ligands. The raw data for d and e) is provided in [supplementary tables S2 and S3, Supplementary Material](#) online, respectively.

Adding orthophosphate has a minimal effect on the conformation of the peptide (Fig. 2a to f), with only a small increase in the nest-like conformation of residues 4 to 6. Addition of GTP yields a more pronounced effect—making the end-to-end distance, bend angle, and uncorrelated backbone dihedrals more *nest-like* but without inducing a global nest conformation (Fig. 2a to f). The peptide-phosphate group interactions formed between the hexapeptide and either ligand (Fig. 2e) are mediated largely by salt bridges between charged moieties (either the N-terminus or the side chain of Lys7) with a more modest contribution from backbone amides, the hallmark feature of a nest binding mode. In contrast, a broad range of interactions involving the sugar and base of GTP were observed ([supplementary fig. S5 and tables S3 and S4, Supplementary Material](#) online). Thus, while flickers of nest-like conformations in the free Walker A peptide do occur, we do not observe global nest formation, nor do we detect persistent, organized phosphate binding mediated by the peptide backbone.

Six Representative Octapeptides Do Not Form Nests in Aqueous Solution

Six ubiquitously distributed P-loop NTPase families were identified ([supplementary table S5, Supplementary Material](#) online). For each family, a representative structure was selected from the PDB (Fig. 1b) and the conformational dynamics of its free Walker A motif analyzed as above. The selected

Walker A octapeptides are identical at 3 to 7 of the 8 positions, owing to the diversity of the X positions, the occasional substitution of Gly4, and the use of Thr or Ser at position 8.

Like the hexapeptide, the octapeptides are highly dynamic but with flickers of nest-like conformations (Fig. 3a and b, top rows). As observed for the hexapeptide, the octapeptides have uncorrelated preferences for the situated backbone dihedral angles (Fig. 3a)—with positions 5 and 7 generally adopting more α_R dihedrals and positions 4 and 6 generally adopting more α_L dihedrals. Surprisingly, even when position 4 is not Gly, a slight preference for α_L can be retained (as in GKPNVGKS), though this is not always the case (as in GPESGKT). The correlated nest propensities (Fig. 3b; [supplementary table S2, Supplementary Material](#) online) show some notable differences between the peptides: First, a greater propensity for correlated nest-like structures is observed for almost all of the octapeptides relative to the hexapeptide. Second, clear differences are observed between the octapeptides, suggesting that the identity of the X residues also augments the nest-forming potential. We note that the peptide with the greatest correlated nest-like conformations (GEPGTGKS) has a proline before Gly4, which may play a key role in restricting the conformational space of the peptide and promoting nest-like conformations that involve three residues (residues 4 to 6, 7.1% occupancy; compared to 0.5% to 1.7% occupancy for all other octapeptides).

The inclusion of phosphate has only a minor effect on the properties of the octapeptides (Fig. 3a to b, middle rows;

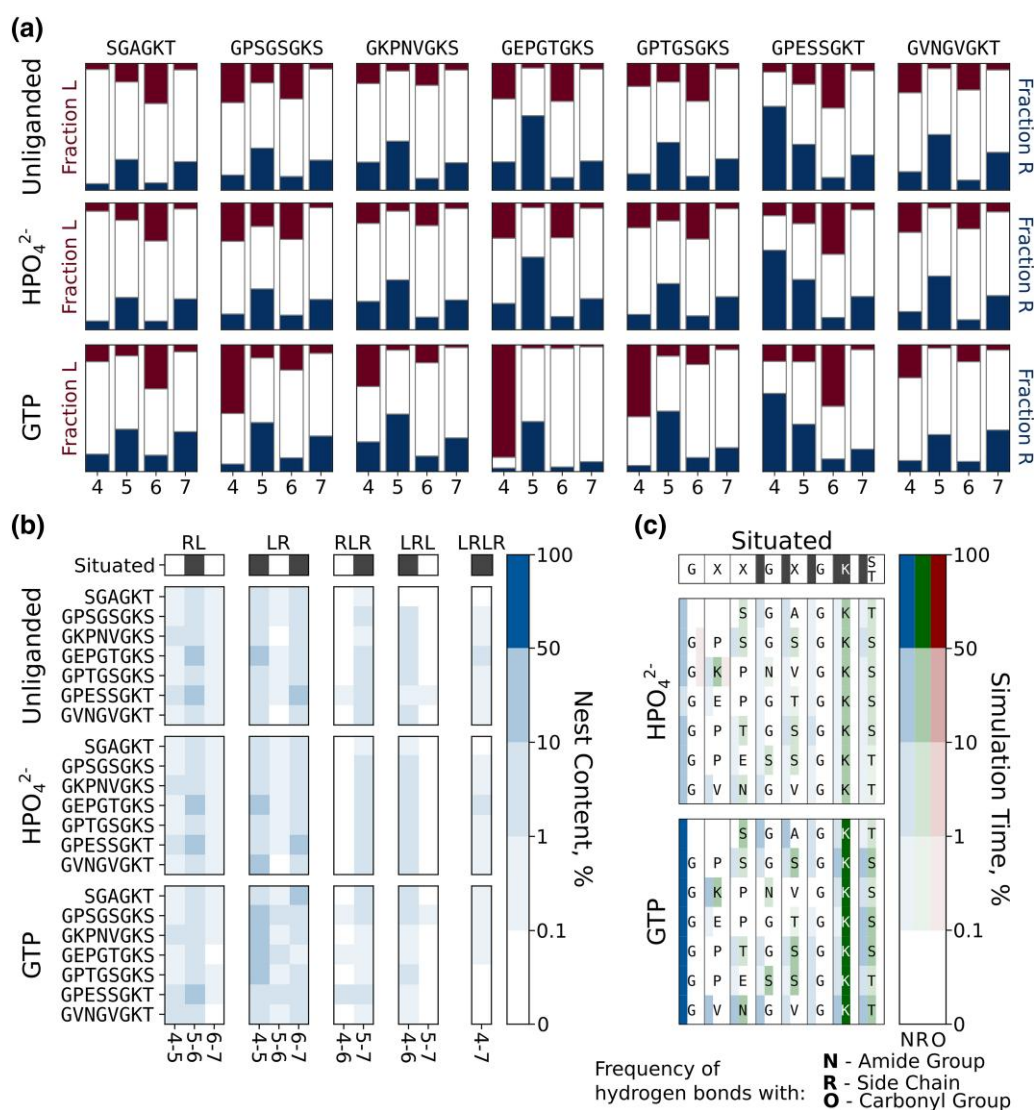


Fig. 3. Conformational dynamics of Walker A-derived octapeptides. a) Uncorrelated preference for α_L and α_R backbone dihedrals. b) Occurrence of correlated stretches of α_L and α_R conformations. c) Interaction profile of the peptides with a ligand. Plot for contacts with GTP shows cumulative probability for interaction with all three phosphate groups of GTP. The raw data for b and c) is provided in [supplementary tables S2 and S3, Supplementary Material online](#).

[supplementary table S2, Supplementary Material online](#)), consistent with the hexapeptide results. GTP (Fig. 3a to b, bottom rows; [supplementary table S2, Supplementary Material online](#)) inclusion had a more dramatic effect, increasing nest-like dihedrals at positions 4 to 5 by ~15% to 30% occupancy for three octapeptides. This increase in nest character, however, does not translate to longer stretches of nest-like residues. As before, (tri)phosphate binding is mediated largely by interactions involving the charged N-terminus and Lys7, with a comparatively modest contribution from backbone amides (Fig. 3c, [supplementary table S3, Supplementary Material online](#)). Interactions with the GTP sugar and base are more varied—with participation of the N-terminus, backbone carbonyls, polar/charged side chains, and backbone amides ([supplementary fig. S5, Supplementary Material online](#))—and GTP binding was significantly more persistent than binding of orthophosphate (>99% vs. ~30% simulation time, respectively; [supplementary table S4, Supplementary Material online](#)). In short, robust adoption of nest-like conformations was not observed for the octapeptides, even in the presence

of phospho-ligands. Simulations performed with a different force field echo these results ([supplementary fig. S6, table S6, Supplementary Material online](#)). Even adding a distance restraint that penalizes deviation of the N- and C-termini from the end-to-end distance of the situated motif (a crude mimic of the effect of the protein scaffold on P-loop structure) did not induce global nest formation or nest-like binding, though it did increase the partial nest character of the peptides ([supplementary fig. S7, tables S7 and S8, Supplementary Material online](#)).

Walker A Peptides Behave Similarly to Control Peptides

Six control loops were selected that have the same amino acid composition as a Walker A octapeptide but different sequences and situated conformations (Fig. 4a and b; [supplementary table S9, Supplementary Material online](#)). In most cases, the control loops are not associated with phosphate binding or a highly conserved sequence motif, and are

thus unlikely to have undergone an evolutionary trajectory from a short peptide to a folded domain, as has been proposed for P-loop NTPases. Analyzing the properties of these loops, then, can serve as a useful point of comparison with which to judge the properties of the octapeptides. Foremost, and particularly in the absence of ligand (Fig. 4) or with orthophosphate (supplementary fig. S8, Supplementary Material online), the control loops and the Walker A octapeptides have essentially equivalent dynamic properties: Broad distributions of end-to-end distances (Fig. 4c) and bend angles (Fig. 4d), though often with some conformational preference, and a similar degree of structural similarity between the situated and the free peptide (Fig. 4e). Although the patterns of correlated and uncorrelated nest-like conformations (supplementary fig. S9, Supplementary Material online) were different from the Walker A peptides, their magnitude was roughly the same—suggesting that this degree of structuring is common among free peptides. As with the other loops analyzed, the control loops did not bind persistently to orthophosphate (25.2% to 53.0% of simulation time; slightly higher than for the Walker A octapeptides, which ranged from 24.7% to 39.9% of simulation times) and interacted predominantly via charged groups and the N-terminus of the peptide (despite the different locations of these residues, supplementary fig. S9, Supplementary Material online). Both the Walker A octapeptides and the control peptides adopt similar secondary structure distributions with or without added ligands (supplementary fig. S10, Supplementary Material online).

Once again, interactions with GTP caused larger shifts in the conformations of the peptides than orthophosphate, no doubt due to its comparatively large interaction surface (Fig. 4). Intriguingly, whereas the Walker A peptides generally become more similar to their situated structure (Fig. 4d), the control peptides become less similar to their situated structures, despite having similar overall interactions with different parts of the nucleotide (supplementary fig. S9 and tables S9 to S11, Supplementary Material online, although note that some interactions are stronger in the Walker A than in the control sequences, see supplementary fig. S5, Supplementary Material online). This result is likely because GTP induces a bend in the peptide that is similar to the bend angle of the situated Walker A loop but rather dissimilar to the bend angles of several situated control loops (Fig. 4a and b).

Nest Formation is Enhanced in a Lower Dielectric Medium

Our simulations strongly suggest that nest formation, if observed at all, is at best transient for the Walker A peptides studied in this work when they are simulated in water solvent. As pointed out by a Reviewer, it is possible that in an early evolutionary context, life's building blocks (which include primitive peptides such as those studied here) may have encountered less polar or even hydrophobic settings that could favor intramolecular hydrogen bonding (e.g. lipid membranes or interfacial environments). These environments could allow for more stable nest formation in the Walker A peptides and increased phospho-ligand binding due to enhanced stabilizing interactions.

To explore these possibilities, we have simulated the dynamics of the liganded forms of the SGAGKT peptide, the GPSGSGKS Walker A octapeptide (which is most similar to the Walker A consensus sequence), and the KSPGGSSG

control sequence (which is a scrambled version of the simulated Walker A sequence) in methanol. We chose methanol as solvent for a number of reasons: First, in very low dielectric solvents (e.g. hexane), the phosphate groups of the phospholigands would be fully protonated, and thus not in a biologically relevant state. Methanol has a significantly lower dielectric constant (32.7) than water (80) but is sufficiently polar that phosphate groups can be plausibly unprotonated. Second, experimental data on GTP hydrolysis in methanol by elongation factor Ts show that GTP hydrolysis is strongly stimulated by the alcohol, whereas aminoacyl-tRNA binding to the ribosomes is not affected (Ballesta and Vazquez 1972). These considerations make methanol functionally interesting as a solvent, even if it is not generally considered relevant for early peptide evolution.

Our simulations (Fig. 5; supplementary fig. S11, Supplementary Material online) show that lowering the dielectric constant of the medium does indeed lead to (i) more frequent uncorrelated L/R mini-nest formation, (ii) slightly more frequent correlated LRLR next formation, (iii) more frequent peptide contacts with the phosphate groups, with polar contacts still dominating, (iv) more structuring of the peptide (based on the end-to-end distance and bend angle of the peptide), and (v) increased peptide-ligand interactions for both HPO_4^{2-} and GTP. Further, when comparing the structural behavior of the GPSGSGKS Walker A octapeptide and the KSPGGSSG control shuffle peptide, we see that in both cases, moving from water to methanol as solvent increases the prominence of short L/R nest segments, but only the Walker A sequence shows a meaningful increase in the sampling of contiguous LRLR nests. This observation emphasizes how methanol appears to reveal some intrinsic structural preferences inherent to these sequences that were not observed in our simulations in aqueous solution. We note that despite the apparent formation of more contiguous nest structures, the contact maps shown in Fig. 5 indicate that the ligand and binding patterns do not follow a canonical nest (for example through involvement of the carbonyl groups in interacting with HPO_4^{2-}). Overall, however, medium effects appear to play an important role in enhancing the nest formation and phospholigand binding ability of these sequences.

Discussion

The limits of evolutionary continuity, particularly of P-loop NTPases, is a topic of intense interest (Fry et al. 1988; Chuang et al. 1995; Bianchi et al. 2012; Alva et al. 2015; Romero Romero et al. 2018; Milner-White 2019; Longo et al. 2020a, 2020b; Vyas et al. 2021, 2023). There is a broad consensus that the origin of this family lies somewhere between a short Walker A peptide that forms a nest upon phospho-ligand binding (Bianchi et al. 2012; Milner-White 2019) and a short $\beta\alpha$ or $\beta\alpha\beta$ peptide that can either fold independently as a monomer or upon oligomerization (Alva et al. 2015; Romero Romero et al. 2018; Longo et al. 2020a; Vyas et al. 2021, 2023). Whereas the potential for fold evolution from an oligomerizing peptide composed of several secondary structure elements has been demonstrated for multiple folds (Lee and Blaber 2011; Smock et al. 2016; Romero Romero et al. 2018; Yagi et al. 2021; Seal et al. 2022)—making this a viable degree of complexity for an evolutionary starting point—evolutionary continuity down to a simple, short peptide of ~10 residues is less well established, with FeS maquettes being the primary example (Gibney et al. 1996; Kim et al. 2018). Nevertheless, short nucleotide-binding peptides

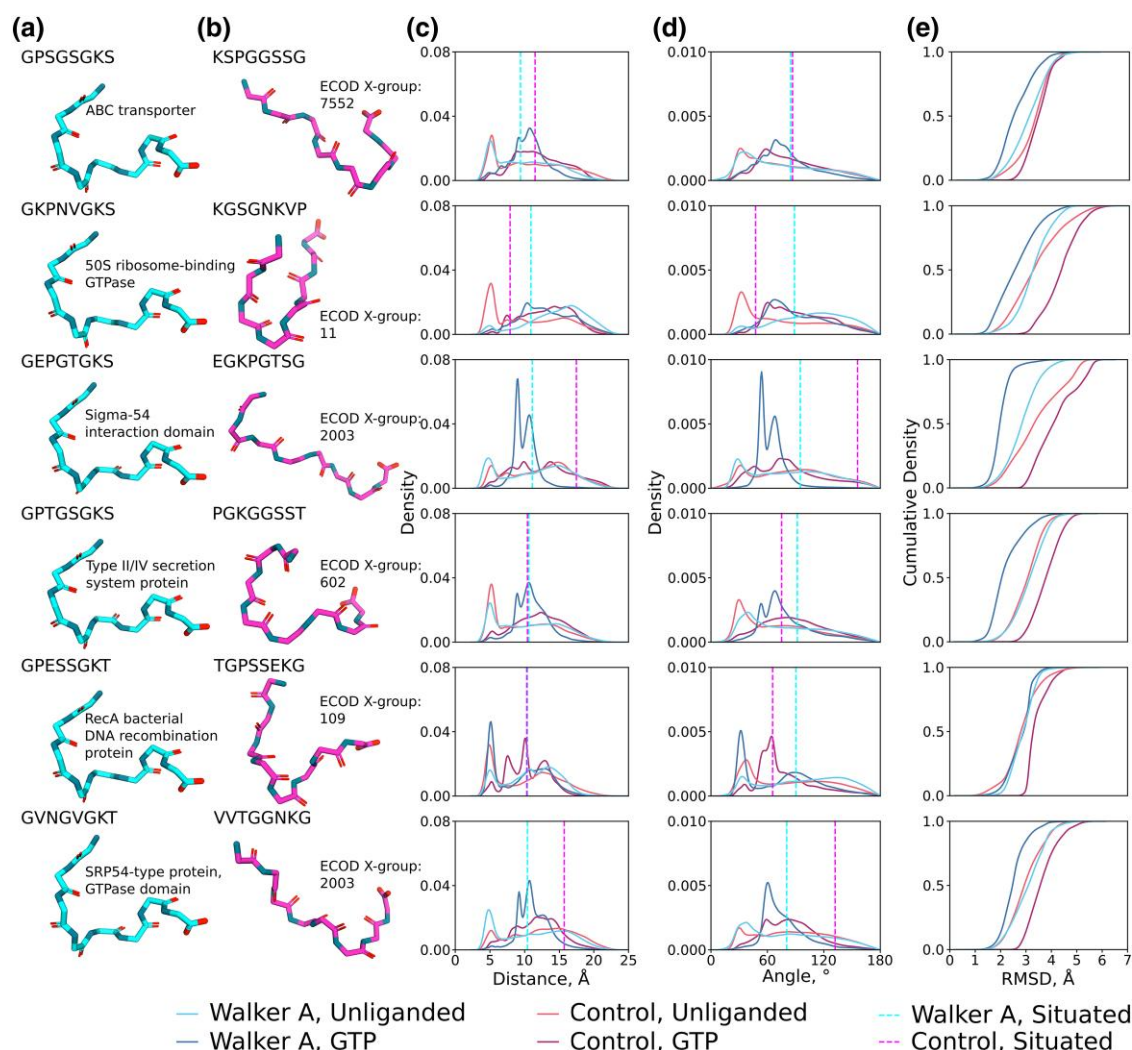


Fig. 4. Conformational dynamics of the control octapeptides. a) Situated structures of the Walker A-derived octapeptides. b) Situated structures of control octapeptides. c) Distribution of $C_{\alpha 1}$ - $C_{\alpha B}$ distances for the free peptide in the presence and absence of GTP. d) Distribution of $C_{\alpha 1}$ - N_5 - $C_{\alpha B}$ angles for the free peptide in the presence and absence of GTP. The situated conformation for panels C and D is indicated with a dotted line. e) Cumulative distribution of the RMSDs of each set of simulations toward the situated structure of the respective peptide in the presence and absence of GTP.

have been reported (Barany and Merrifield 1973; Hilpert et al. 2010; Kang et al. 2015; Kroiss et al. 2019), speaking to the chemical plausibility of this evolutionary scenario for phospho-ligands. The present evolutionary scenario depends on two key points: (i) The disembodied Walker A motif binds phosphate or a phospho-ligand and (ii) the binding mode of the Walker A-derived peptide is nest-like in structure. These two points, which concern different aspects of continuity—one related to function and the other to structure—will be discussed in turn.

Initial studies on the hexapeptide reported binding of orthophosphate using potentiometric titrations (Bianchi et al. 2012) that are sensitive to shifts in pK_a , and thus cannot probe binding by backbone amides. These results are in rough agreement with the MD simulations, where orthophosphate binding was observed in roughly 30% of frames and was mediated largely by charged groups. However, the Walker A octapeptides and the control peptides were highly similar with respect to both binding patterns and binding persistence, suggesting that interactions observed by potentiometric titration may have been largely nonspecific. Although binding to nucleotides was more persistent (>99% of simulation time in all cases)

and induced larger structural changes, the Walker A peptides and the control peptides once again exhibited similar properties. Taken together, we conclude that a disembodied Walker A peptide is unlikely to be a privileged phospho-ligand binder; instead, a diversity of peptides with equivalent or superior binding properties likely exist.

The sustained GTP binding (supplementary table S4, Supplementary Material online) in our simulations may be noteworthy, given that a variety of short peptides bind nucleotides (usually ATP), including peptides composed of prebiotic amino acids (Hilpert et al. 2010; Kang et al. 2015; Kroiss et al. 2019). One caveat to this observation is that these peptides were typically engineered with sequences specifically selected for their ability to bind ATP (Hilpert et al. 2010). The short ATP binding octapeptide TACGQKSP, on the other hand, was extracted from a proposed, and ultimately incorrect, sequence of the ATP binding site of “frog actomyosin.” Despite this sequence being different from any known frog actomyosin ATP binding site, the peptide was observed to bind MgATP with a K_D of 0.22 mM (Barany and Merrifield 1973). Similar to our GTP simulations (supplementary fig. S5 and tables S3 and S4, Supplementary Material online),

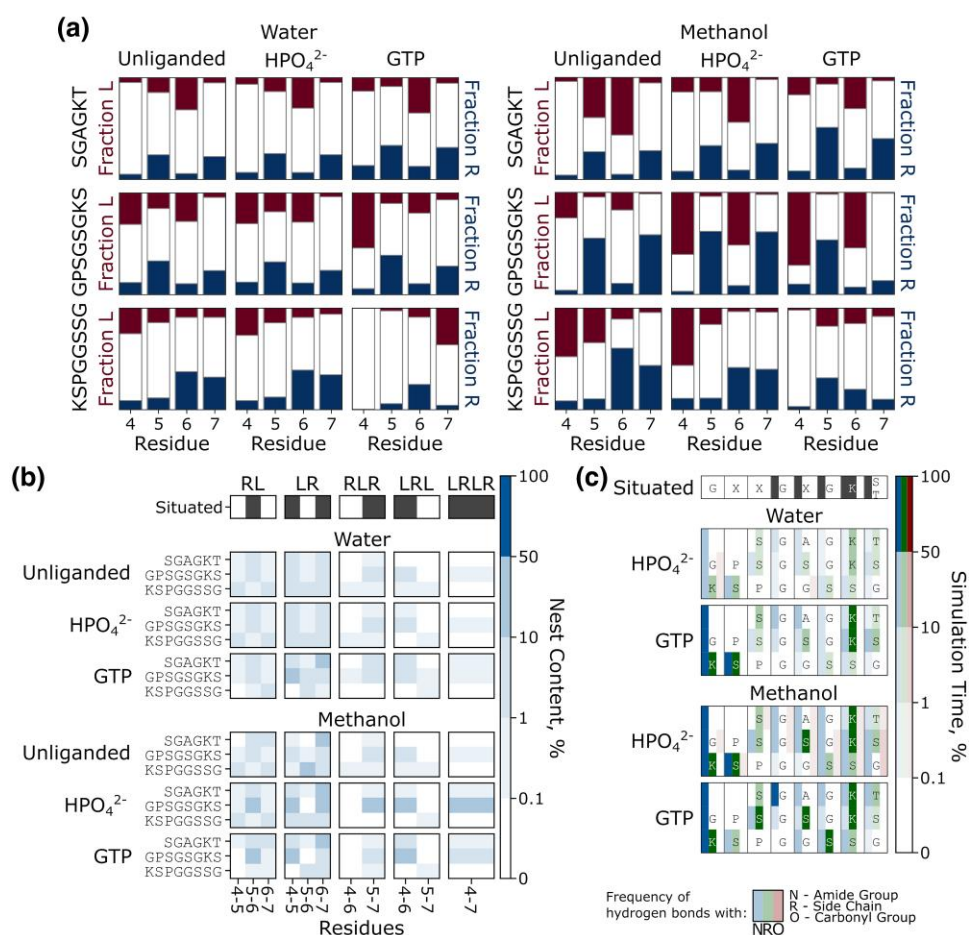


Fig. 5. The impact of solvent on peptide dynamics. a) Uncorrelated preference for α_L and α_R backbone dihedrals. b) Occurrence of correlated stretches of α_L and α_R conformations in water and methanol and in the absence and presence of ligands (HPO_4^{2-} or the triphosphate group of GTP). c) Interaction profile of the peptides with the indicated ligand. Contact plot with GTP shows cumulative probability for interaction with all three phosphate groups of GTP. The raw data for panels d and e) is provided in [supplementary tables S12 and S13, Supplementary Material](#) online.

this peptide binds ATP *via* interactions with all parts of the ligand, not just the triphosphate, and discriminated between ATP, GTP, and CTP binding, with ATP binding preferred (Barany and Merrifield 1973). These data suggest that it is possible to achieve ligand binding from short, disembodied peptides, particularly of nucleotides where supportive interactions with the base can be formed. Indeed, peptide-based interactions could have played an important role in substrate discrimination in an ATP-rich prebiotic world (Pinna et al. 2022).

While some aspect of functional continuity seems at least plausible, the question of structural continuity is more unambiguous: Free Walker A-derived peptides do not have a significant preference for nest-like conformations, even in the presence of phospho-ligands, although the propensity for nest-like dihedrals seems to be enhanced in media with a lower dielectric than water (methanol in the case of our simulations), albeit with a noncanonical binding pattern. The current data suggest that the Walker A motif and the $\beta\alpha\beta$ peptide within which it was situated co-evolved—with the loop and the surrounding peptide reciprocally adjusting until the canonical binding mode was found. Although flickers of nest-like conformations are observed, these are weak and more reasonably interpreted as residual preference for nest-like conformations imparted by evolution operating on the situated peptide. The high content of Gly within the peptide may be a signature of

this co-evolution, as high Gly content in a free peptide is expected to disfavor folding and collapse (though it must also be noted that Gly is the simplest amino acid and likely enjoyed high prebiotic availability). Overall, if a strong propensity for nest formation had been observed, then the framing of this structural element as an evolutionary nucleus would, in our view, have been supported.

We note that there are couple of caveats to our simulations. First, as in any MD simulation study, there is always a risk of force field artifacts, especially in the case of simulations of disordered peptides (García and Sanbonmatsu 2002; Best and Hummer 2009). We mitigate this risk by comparing results from both the Amber ff99SB-ILDN (Lindorff-Larsen et al. 2010) and the CHARMM36m (Huang et al. 2017) force fields (supplementary fig. S6, Supplementary Material online). Second, polarizable force field that can accurately capture the polarization of the peptide bonds relative to the phosphate may show more nest formation.

The field of protein evolution continues to grapple with how to interpret profound sequence conservation—is it evidence of an evolutionary nucleus, the earliest traces of a fold contained within a short peptide? Or, are these signals fragmentary—motifs that resisted erosion over time but nonetheless existed within a specific context that is no longer recoverable, and thus not the “full,” evolutionary seed. Disentangling these

scenarios is difficult. Lupas' work on searching for a "primordial peptide vocabulary," has sought to account for this, in part, by finding conserved sequence-structure elements within multiple, distantly related contexts (Alva et al. 2015). Likewise, Kolodny et al. proposed the concept of "bridging themes," which takes a related approach, except that structure similarity is not required, allowing the inclusion of metamorphic fragments as well (Kolodny et al. 2021). The latter approach was instrumental in uncovering the potential deep evolutionary connection between P-loop NTPases and Rossmann enzymes (Longo et al. 2020a; Kolodny et al. 2021). Although reasonable, these approaches are imperfect, and the identified fragments may have emerged within a fold, as an assembling oligomer, or even folded independently. As the interfaces of the conserved fragments were overwritten to adapt to changing contexts, information about the originating context can be lost or distorted. For the Walker A-derived peptides, the presence of Mg^{2+} (which was not included in the current work, but is present in many Walker A P-loop containing proteins), or the ability to assemble, may reveal a yet-unappreciated aspect of evolutionary continuity between a Walker A peptide and the NTPase evolutionary lineage. It is only through extensive characterization that the relative likelihood of these scenarios can be assessed. At present, the best evidence for the evolutionary seed of P-loop NTPases points to oligomerizing $\beta\alpha\beta$ or $\beta\alpha$ peptides (Romero Romero et al. 2018).

Materials and Methods

Situated Loop Analysis

Representative P-loop NTPase structures were collected from the Evolutionary Classification of Domains (ECOD) database, version develop291 (Schaeffer et al. 2017). Domain sequences were clustered at 90% identity using CD-HIT (Li and Godzik 2006), and structures with a resolution worse than 2.5 Å were excluded from further analysis. For each cluster, the liganded and unliganded structures with the highest resolution were taken to be the cluster representatives. Per residue distributions of dihedral angles and an analysis of supporting contacts were calculated using custom scripts (Zenodo DOI: <https://doi.org/10.5281/zenodo.13685585>) based on the PyMOL API (pymol.org). Residues that form supporting interactions were identified by a heavy atom distance cutoff of 3.5 Å between the situated loop and the surrounding protein. Residues that are covalently attached to a loop residue, as well as the $i+4$ α -helix hydrogen bonding interactions at the C-terminal end of the situated loop, were excluded from the list of interacting residues as they are largely invariant and reflect anchor points between the loop and the surrounding fold.

Selecting Walker A Sequences for Detailed Characterization

The P-loop NTPase evolutionary lineage is associated with more than 120 distinct families (Longo et al. 2020a). To identify families that are likely to have been present in the last universal common ancestor, we calculated the phyletic distribution of each family across the microbial tree of life (supplementary table S14, Supplementary Material online). Hidden Markov Models (HMMs) for all known P-loop NTPase families were downloaded from the ECOD database (version develop279) (Cheng et al. 2014; Schaeffer et al. 2017). These HMM profiles were then searched against the

bacterial and archaeal genomes in the curated Genome Taxonomy Database (Parks et al. 2022) (GTDB; release 95) using hmmsearch. A positive identification of a P-loop NTPase domain, or hit, was defined as an independent expectation value (E-value) of $\leq 10^{-4}$ and an HMM profile coverage of $\geq 70\%$. The hmmsearch search space parameter, which is used in the calculation of independent E-values, was set to 106,052,079 sequences (taken from Pfam [Finn et al. 2014]) to allow for facile comparison of E-values between runs. 15 families were present in $\geq 90\%$ of phyla and $\geq 80\%$ of species in both Archaea and Bacteria, of which 7 were chosen for detailed analysis (supplementary table S5, Supplementary Material online). These seven families were chosen such that both common strand topologies within the P-loop NTPase evolutionary lineage were represented. For each of these seven families, a representative structure with a resolution ≤ 2.0 Å was identified. Control sequences were chosen such that they have the same amino acid composition as one of the representative Walker A sequences but are from an unrelated protein lineage (supplementary table S9, Supplementary Material online).

Starting Conformations for the Simulations

The simulation starting conformation for the hexapeptide SGAGKT (Bianchi et al. 2012) was generated by mutating the Walker A motif from G-protein p21^{ras}, which has the sequence GGVGKS, in PyMOL (Schrödinger, LLC) (ECOD domain e5p21A1; note that an ECOD domain identifier is constructed from the PDB ID (in this case 5p21 [Pai et al. 1990]), the chain identifier [A] and an ECOD domain identifier [1]). Starting conformations for the Walker A-derived and control octapeptides were taken from their respective crystal structures (supplementary table S5, Supplementary Material online). For simulations with GTP, the position of the ligand was either taken from the crystal structure directly or positioned by superimposition with a liganded structure. For simulations with HPO_4^{2-} , the position of the ligand was obtained by superimposition with the β -phosphate of the previously positioned GTP.

HREX-MD Simulations

HREX-MD simulations (Bussi 2014) were performed on each Walker A-derived and control octapeptide in the presence and absence of the ligands orthophosphate or GTP. All Hamiltonian HREX-MD simulations used the Amber ff99SB-ILDN force field (Lindorff-Larsen et al. 2010) and either the TIP3P water model (Jorgensen et al. 1983) or methanol (parameters obtained from Caldwell and Kollmann 1995), as implemented in the GROMACS 2019.4 simulation package (van der Spoel et al. 2005). The PLUMED v2.6. interface was used (Tribello et al. 2014). Additional control simulations of the unliganded Walker A-derived octapeptides were performed using the CHARMM36 m force field (Huang et al. 2017) to account for secondary structure bias of different force fields (Rauscher et al. 2015). These simulations were prepared using the CHARMM-GUI (Jo et al. 2008). All ligand parameters are provided in supplementary tables S15 and S16, Supplementary Material online. Following initial system preparation and equilibration (see the supplementary materials, Supplementary Material online), 1 μ s HREX-MD simulations were performed on each system, with all solute atoms included in the hot region. HREX-MD simulations were performed using a total of 8 replicas per system, with λ

values scaled exponentially between 0.67 to 1. Exchanges were attempted every 4 ps, achieving an average exchange rate of ~43% across all systems. Subsequent analyses, presented below, were performed only on the neutral replica ($\lambda=1$). Further details of the HREX-MD simulations are provided in the [supplementary materials](#), [Supplementary Material](#) online.

Simulation Analysis

Distance and angle analyses of the simulated peptides were performed using custom scripts (Zenodo DOI <https://doi.org/10.5281/zenodo.13685585>). Analysis of ϕ , ψ backbone dihedrals to detect α_R (R) and α_L (L) conformations of nest-forming residues was based on snapshots taken every 10 ps. The SGAGKT hexapeptide was aligned to the longer sequences shown in [supplementary table S5](#), [Supplementary Material](#) online and renumbered accordingly for presentation purposes. Dihedral ranges were defined as $-20^\circ < \phi < -140^\circ$ and $40^\circ < \psi < -90^\circ$ for α_R and $20^\circ < \phi < 140^\circ$ and $-40^\circ < \psi < 90^\circ$ for α_L (Watson and Milner-White 2002). For each frame, correlated nest-like conformations were assigned when consecutive residues adopted LR, RL, RLR, LRL, or LRLR conformations. Hydrogen bonds between the peptide and ligand were identified by calculating donor-acceptor distances and angles using GROMACS (van der Spoel et al. 2005). Distance and angle cutoffs of 3.5 Å and 135° were used for the donor-acceptor distance and the donor-hydrogen-acceptor angle, respectively.

Dimensionality reduction was performed using tICA. Cosines of the ϕ , ψ backbone dihedrals along the entire peptide chain were used as input with a lag time of 1 ps. The resulting 2D maps were then subject to K-means clustering. Both the tICA analysis and K-means clustering were performed using the MSMBuilder Python package (Beauchamp et al. 2011). Liganded and unliganded simulation data for a given peptide were concatenated prior to dimensionality reduction and clustering. For each trajectory, the root mean square deviation (RMSD) of the backbone heavy atoms with respect to the situated structure was calculated using the GROMACS internal RMSD function (van der Spoel et al. 2005). Secondary structure analysis was performed on snapshots taken every 10 ps using MDTraj (McGibbon et al. 2015) and assigned using the Define Secondary Structure of Proteins nomenclature (Kabsch and Sander 1983). For each residue, the percentage of time spent in each secondary structure conformation was calculated and averaged over the entire length of the peptide to obtain an estimate of the secondary structure content of the entire peptide.

Supplementary Material

[Supplementary material](#) is available at *Molecular Biology and Evolution* online.

Acknowledgments

We thank Dan Tawfik for inspiring this work and for helpful discussion. L.M.L. is grateful to Eric Smith for insightful discussions.

Funding

This work was supported by the Knut and Alice Wallenberg Foundation (grant numbers 2018.0140 and 2019.0431).

P.L. was supported by the Okinawa Institute of Science and Technology Graduate University (OIST) with subsidy funding from the Cabinet Office, Government of Japan. The simulations were enabled by resources provided by the National Academic Infrastructure for Supercomputing in Sweden (NAISS), partially funded by the Swedish Research Council through grant agreement no. 2022-06725.

Data Availability

All simulation parameters necessary for reproducing our simulations are provided on Zenodo, DOI: <https://doi.org/10.5281/zenodo.13685585>. This data package includes simulation starting structures, input files, nonstandard parameters, and any custom analysis scripts used in this work.

References

- Alva V, Söding J, Lupas A. A vocabulary of ancient peptides at the origin of folded proteins. *eLife*. 2015;4:e09410. <https://doi.org/10.7554/eLife.09410>.
- Aravind L, Mazumder R, Vasudevan S, Koonin EV. Trends in protein evolution inferred from sequence and structure analysis. *Curr Opin Struct Biol*. 2002;12(3):392–399. [https://doi.org/10.1016/S0959-440X\(02\)00334-2](https://doi.org/10.1016/S0959-440X(02)00334-2).
- Ballesta JP, Vazquez D. Elongation factor T-dependent hydrolysis of guanosine triphosphate resistant to thiostrepton. *Proc Natl Acad Sci U S A*. 1972;69(10):3058–3062. <https://doi.org/10.1073/pnas.69.10.3058>.
- Barany G, Merrifield RB. An ATP-binding peptide. *Cold Spring Harb Symp Quant Biol*. 1973;37(0):121–125. <https://doi.org/10.1101/SQB.1973.037.01.019>.
- Beauchamp KA, Bowman GR, Lane TJ, Maibaum L, Haque IS, Pande VS. MSMBuilder2: modeling conformational dynamics on the picosecond to millisecond scale. *J Chem Theory Comput*. 2011;7(10):3412–3419. <https://doi.org/10.1021/ct200463m>.
- Best RB, Hummer G. Optimized molecular dynamics force fields applied to the Helix-coil transition of polypeptides. *J Phys Chem B*. 2009;113(26):9004–9015. <https://doi.org/10.1021/jp901540t>.
- Bianchi A, Giorgi C, Ruzza P, Toniolo C, Milner-White EJ. A synthetic hexapeptide designed to resemble a proteinaceous P-loop nest is shown to bind inorganic phosphate. *Proteins*. 2012;80(5):1418–1424. <https://doi.org/10.1002/prot.24038>.
- Bukhari SA, Caetano-Anollés G. Origin and evolution of protein fold designs inferred from phylogenomic analysis of CATH domain structures in proteomes. *PLoS Comput Biol*. 2013;9(3):e1003009. <https://doi.org/10.1371/journal.pcbi.1003009>.
- Bussi G. Hamiltonian replica exchange in GROMACS: a flexible implementation. *Mol Phys*. 2014;112(3-4):379–384. <https://doi.org/10.1080/00268976.2013.824126>.
- Caldwell JW, Kollmann PA. Structure and properties of neat liquids using nonadditive molecular dynamics: water, methanol, and N-methylacetamide. *J Phys Chem*. 1995;99(16):6208–6219. <https://doi.org/10.1021/j100016a067>.
- Carvalho S, Peralta Reis DQ, Pereira SV, Kalafatovic D, Pina AS. Catalytic peptides: the challenge between simplicity and functionality. *Israel J Chem*. 2022;62(9-10):e202200029. <https://doi.org/10.1002/ijch.202200029>.
- Cheng H, Schaeffer RD, Liao Y, Kinch LN, Pei J, Shi S, Kim B-H, Grishin NV. ECOD: an evolutionary classification of protein domains. *PLoS Comput Biol*. 2014;10(12):e1003926. <https://doi.org/10.1371/journal.pcbi.1003926>.
- Chuang WJ, Abeygunawardana C, Gittis AG, Pedersen PL, Mildvan AS. Solution structure and function in trifluoroethanol of PP-50, an ATP-binding peptide from FIATPase. *Arch Biochem Biophys*. 1995;319(1):110–122. <https://doi.org/10.1006/abbi.1995.1272>.
- Crean RM, Biler M, van der Kamp MW, Hengge AC, Kamerlin SCL. Loop dynamics and enzyme catalysis in protein tyrosine

- phosphatases. *J Am Chem Soc.* 2021;143(10):3830–3845. <https://doi.org/10.1021/jacs.0c11806>.
- Eck RV, Dayhoff MO. Evolution of the structure of ferredoxin based on living relics of primitive amino acid sequences. *Science.* 1966; 152(3720):363–366. <https://doi.org/10.1126/science.152.3720.363>.
- Finn RD, Bateman A, Clements J, Coghill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Misty J et al. The Pfam protein families database. *Nucleic Acids Res.* 2014;40(Database issue): D290–D301. <https://doi.org/10.1093/nar/gkt1223>
- Fowler WC, Deng C, Griffen GM, Teodora OT, Guo AZ, Zaiden M, Gottlieb M, de Pablo JJ, Tirrell MV. Harnessing peptide binding to capture and reclaim phosphate. *J Am Chem Soc.* 2021;143(11): 4440–4450. <https://doi.org/10.1021/jacs.1c01241>.
- 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. Roy. Soc. Interface.* 2022;19(187):20210641. <https://doi.org/10.1098/rsif.2021.0641>.
- Fry DC, Byler M, Susi H, Brown EM, Kuby SA, Mildvan AS. Solution structure of the 45-residue magnesium ATP-binding peptide of adenylate kinase as examined by 2-D NMR, FTIR, and CD spectroscopy. *Biochemistry.* 1988;27(10):3588–3598. <https://doi.org/10.1021/bi00410a009>.
- Fry MY, Najdarová V, Maggiolo AO, Saladi SM, Doležal P, Clemons WM. Structurally derived universal mechanism for the catalytic cycle of the tail-anchored targeting factor get3. *Nat Struct Mol Biol.* 2022;29(8):820–830. <https://doi.org/10.1038/s41594-022-00798-4>.
- García AE, Sanbonmatsu KY. Alpha-helical stabilization by Side chain shielding of backbone hydrogen bonds. *Proc Natl Acad Sci U S A.* 2002;99(5):2782–2787. <https://doi.org/10.1073/pnas.042496899>.
- Georgiadis MM, Komiya H, Chakrabarti P, Woo D, Kornuc JJ, Rees DC. Crystallographic structure of the nitrogenase iron protein from *Azotobacter vinelandii*. *Science.* 1992;257(5077):1653–1659. <https://doi.org/10.1126/science.1529353>.
- Gibney BR, Mulholland SE, Rabanal F, Dutton PL. Ferredoxin and ferredoxin-Heme Maquettes. *Proc Natl Acad Sci U S A.* 1996;93(26):15041–15046. <https://doi.org/10.1073/pnas.93.26.15041>.
- Gruber M, Greisen P Jr., Junker CM, Hélix-Neilsen C. Phosphorus binding sites in proteins: structural preorganization and coordination. *J Phys Chem B.* 2014;118(5):1207–1215. <https://doi.org/10.1021/jp408689x>.
- Gruic-Sovulj I, Longo LM, Jabłońska J, Tawfik DS. The evolutionary history of the HUP domain. *Crit Rev Biochem Mol Biol.* 2022;57(1):1–15. <https://doi.org/10.1080/10409238.2021.1957764>.
- Hilpert K, McLeod B, Yu J, Elliott MR, Rautenbach M, Ruden S, Bürck J, Muhle-Goll C, Ulrich AS, Keller S et al. Short cationic antimicrobial peptides interact with ATP. *Antimicrob Agents Chemother.* 2010;54(10):4480–4483. <https://doi.org/10.1128/AAC.01664-09>.
- Hol WGJ, van Duynen PT, Berendsen HJC. The α -Helix dipole and the properties of proteins. *Nature.* 1978;273(5662):443–446. <https://doi.org/10.1038/273443a0>.
- Huang J, Rauscher S, Nawrocki G, Ran T, Feig M, de Groot BL, Grubmüller H, MacKerell Jr. AD. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat Methods.* 2017;14(1):71–73. <https://doi.org/10.1038/nmeth.4067>.
- Jo S, Kim T, Iyer VG, Im W. CHARMM-GUI: a web-based graphical user interface for CHARMM. *J Comput Chem.* 2008;29(11): 1859–1865. <https://doi.org/10.1002/jcc.20945>.
- Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. *J Chem Phys.* 1983;79(2):926–935. <https://doi.org/10.1063/1.445869>.
- Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers.* 1983;22(12):2577–2637. <https://doi.org/10.1002/bip.360221211>.
- Kang S-K, Chen B-X, Tian T, Jia X-S, Chu X-I, Liu R, Dong P-F, Yang Q-Y, Zhang H-Y. ATP selection in a random peptide library consisting of prebiotic amino acids. *Biochem Biophys Res Commun.* 2015;466(3):400–405. <https://doi.org/10.1016/j.bbrc.2015.09.038>.
- Kim JD, Pike DH, Tyryshkin AM, Swapna GVT, Raanan H, Montelione GT, Nanda V, Falkowski PG. Minimal heterochiral *de Novo* designed 4Fe–4S binding peptide capable of robust electron transfer. *J Am Chem Soc.* 2018;140(36):11210–11213. <https://doi.org/10.1021/jacs.8b07553>.
- Kolodny R, Nepomnyachiy S, Tawfik DS, Ben-Tal N. Bridging themes: short protein segments found in different architectures. *Mol Biol Evol.* 2021;38(6):2191–2208. <https://doi.org/10.1093/molbev/msab017>.
- Kozlova MI, Shalaeva DN, Dibrova DV, Mulikidjanian AY. Common mechanism of activated catalysis in P-loop fold nucleoside triphosphatases—united in diversity. *Biomolecules.* 2022;12(10):1346. <https://doi.org/10.3390/biom12101346>.
- Kroiss D, Aramini JM, McPhee SA, Tuttle T, Ulijn RV. Unbiased discovery of dynamic peptide-ATP complexes. *ChemSystemsChem.* 2019;1(1-2):7–11. <https://doi.org/10.1002/syst.201900013>.
- Laurino P, Tóth-Petróczy Á, Meana-Pañeda R, Lin W, Truhlar DG, Tawfik DS. An ancient fingerprint indicates the common ancestry of rosmann-fold enzymes utilizing different ribose-based cofactors. *PLoS Biol.* 2016;14(3):e1002396. <https://doi.org/10.1371/journal.pbio.1002396>.
- Lee J, Blaber M. Experimental support for the evolution of symmetric protein architecture from a simple peptide Motif. *Proc Natl Acad Sci U S A.* 2011;108(1):126–130. <https://doi.org/10.1073/pnas.1015032108>.
- Li W, Godzik A. CD-HIT: a fast program for clustering and comparing large sets of proteins or nucleotide sequences. *Bioinformatics.* 2006;22(13):1658–1659. <https://doi.org/10.1093/bioinformatics/btl158>.
- Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis JL, Dror RO, Shaw DE. Improved Side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins Struct. Func. Bioinform.* 2010;78(8):1950–1958. <https://doi.org/10.1002/prot.22711>.
- Longo LM, Hirai H, McGlynn SE. An evolutionary history of the CoA-binding protein Nat/Ivy. *Prot. Sci.* 2022;31(12):e4463. <https://doi.org/10.1002/pro.4463>.
- Longo LM, Jabłońska J, Vyas P, Kanada M, Kolodny R, Ben-Tal N, Tawfik DS. On the emergence of P-loop NTPase and rosmann enzymes from a Beta-alpha-Beta ancestral fragment. *eLife.* 2020a;9: e64415. <https://doi.org/10.7554/eLife.64415>.
- Longo LM, Petrovic 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 U S A.* 2020b; 117(10):5310–5318. <https://doi.org/10.1073/pnas.1911742117>.
- Lupas AN, Ponting CP, Russel 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(2-3):191–203. <https://doi.org/10.1006/jsbi.2001.4393>.
- Ma BG, Chen L, Ji HF, Chen ZH, Yang FR, Wang L, Qu G, Jiang YY, Ji C, Zhang HY. Characters of very ancient proteins. *Biochem Biophys Res Commun.* 2008;366(3):607–611. <https://doi.org/10.1016/j.bbrc.2007.12.014>.
- Maggiolo AO, Mahajan S, Rees DC, Clemons WM. Intradimeric walker A ATPases: conserved features of A functionally diverse family. *J Mol Biol.* 2023;435(11):167965. <https://doi.org/10.1016/j.jmb.2023.167965>.
- McGibbon RT, Beauchamp KA, Harrigan MP, Klein C, Swails JM, Hernández CX, Schwantes CR, Wang L-P, Lane TJ, Pande VS. MDTraj: a modern open library for the analysis of molecular dynamics trajectories. *Biophys J.* 2015;109(8):1528–1532. <https://doi.org/10.1016/j.bpj.2015.08.015>.
- Milner-White EJ. Protein three-dimensional structures at the origin of life. *Interface Focus.* 2019;9(6):20190057. <https://doi.org/10.1098/rsfs.2019.0057>.
- Pai EF, Kregel U, Petsko GA, Goody RS, Kabsch W, Wittinghofer A. Refined crystal structure of the triphosphate conformation of

- H-ras p21 at 1.35 Å resolution: implications for the mechanism of GTP hydrolysis. *EMBO J.* 1990;9(8):2351–2359. <https://doi.org/10.1002/j.1460-2075.1990.tb07409.x>.
- Pal D, Sahu S, Banerjee R. New facets of larger nest motifs in proteins. *Prot. Struct. Func. Bioinform.* 2020;88(11):1413–1422. <https://doi.org/10.1002/prot.25961>.
- Pálffy G, Menyhárd DK, Perczel A. Dynamically encoded reactivity of ras enzymes: opening new frontiers for drug discovery. *Cancer Metastasis Rev.* 2020;39(4):1075–1089. <https://doi.org/10.1007/s10555-020-09917-3>.
- Parks DH, Chuvochina M, Rinke C, Mussig AJ, Chaumeil P-A, Hugenholtz P. GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. *Nucleic Acids Res.* 2022;50(D1):D785–D794. <https://doi.org/10.1093/nar/gkab776>.
- Pinna S, Kunz C, Halpern A, Harrison SA, Jordan SF, Ward J, Werner F, Lane N. A prebiotic basis for ATP as the universal energy currency. *PLoS Biol.* 2022;20(10):e3001437. <https://doi.org/10.1371/journal.pbio.3001437>.
- Rauscher S, Gapsys V, Gajda MJ, Zweckstetter M, de Groot BL, Grubmüller H. Structural ensembles of intrinsically disordered proteins depend strongly on force field: a comparison to experiment. *J Chem Theory Comput.* 2015;11(11):5513–5524. <https://doi.org/10.1021/acs.jctc.5b00736>.
- Romero Romero ML, Yang F, Lin Y-R, Toth-Petroczy A, Berezovsky IN, Goncarenko A, Yang W, Wellner A, Kumar-Deshmukh F, Sharon M et al. Simple yet functional phosphate-loop proteins. *Proc Natl Acad Sci U S A.* 2018;115(51):E11943–E11950. <https://doi.org/10.1073/pnas.1812400115>.
- Russell MJ, Barge LM, Bhartia R, Bocanegra D, Bracher PJ, Branscomb E, Kidd R, McGlynn S, Meier DH, Nitschke W et al. The drive to life on wet and icy worlds. *Astrobiology.* 2014;14(4):308–343. <https://doi.org/10.1089/ast.2013.1110>.
- Saraste M, Sibbald PR, Wittinghofer A. The P-loop—a common motif in ATP- and GTP-binding proteins. *Trends Biochem Sci.* 1990;15(11):430–434. [https://doi.org/10.1016/0968-0004\(90\)90281-F](https://doi.org/10.1016/0968-0004(90)90281-F).
- Schaeffer RD, Liao Y, Cheng H, Grishin NV. ECOD: new developments in the evolutionary classification of domains. *Nucleic Acids Res.* 2017;45(D1):D296–D302. <https://doi.org/10.1093/nar/gkw1137>.
- Schrodinger, LLC. The PyMOL molecular graphics system. Version 3.0.
- Seal M, Weil-Ktorza O, Despotović D, Tawfik DS, Levy Y, Metanis N, Longo LM, Goldfarb D. Peptide-RNA coacervates as a cradle for the evolution of folded domains. *J Am Chem Soc.* 2022;144(31):14150–14160. <https://doi.org/10.1021/jacs.2c03819>.
- Shalaeva DN, Cherepanov DA, Glaperin MY, Golovin AY, Mulikidjanian AY. Evolution of cation binding in the active sites of P-loop nucleoside triphosphatases in relation to the basic catalytic mechanism. *eLife.* 2018;7:e37373. <https://doi.org/10.7554/eLife.37373>.
- Smith CA, Rayment I. Active site comparisons highlight structural similarities between myosin and other P-loop proteins. *Biophys J.* 1996;70(4):1590–1602. [https://doi.org/10.1016/S0006-3495\(96\)79745-X](https://doi.org/10.1016/S0006-3495(96)79745-X).
- Smock RG, Yadid I, Dym O, Clarke J, Tawfik DS. *De Novo* evolutionary emergence of a symmetrical protein is shaped by folding constraints. *Cell.* 2016;164(3):476–486. <https://doi.org/10.1016/j.cell.2015.12.024>.
- Taberner L, Aricescu AR, Jones EY, Szdlacsek SE. Protein tyrosine phosphatases: structure-function relationships. *FEBS J.* 2008;275(5):867–882. <https://doi.org/10.1111/j.1742-4658.2008.06251.x>.
- Timm J, Pike DH, Mancini JA, Tryshkin AM, Poudel S, Siess JA, Molinaro PM, McCann JJ, Waldie KM, Koder RL et al. Design of a minimal di-nickel hydrogenase peptide. *Sci Adv.* 2023;9(10):eabq1990. <https://doi.org/10.1126/sciadv.abq1990>.
- Tribello GA, Bonomi M, Branduardi D, Camilloni C, Bussi G. PLUMED 2: new feathers for an old bird. *Comput Phys Commun.* 2014;185(2):604–613. <https://doi.org/10.1016/j.cpc.2013.09.018>.
- van der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJC. GROMACS: fast, flexible and free. *J. Comp. Chem.* 2005;26(16):1701–1718. <https://doi.org/10.1002/jcc.20291>.
- Vyas P, Malitsky S, Itkin M, Tawfik DS. On the origins of enzymes: phosphate-binding polypeptides mediate phosphoryl transfer to synthesize adenosine triphosphate. *J Am Chem Soc.* 2023;145(15):8344–8354. <https://doi.org/10.1021/jacs.2c08636>.
- Vyas P, Trofimuk O, Longo LM, Deshmukh FK, Sharon M, Tawfik DS. Helicase-like functions in phosphate loop containing Beta-alpha polypeptides. *Proc Natl Acad Sci U S A.* 2021;118(16):e2016131118. <https://doi.org/10.1073/pnas.2016131118>.
- Walker JE, Saraste M, Runswick MJ, Gay NJ. Distantly related sequences in the alpha- and Beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *EMBO J.* 1982;1(8):945–951. <https://doi.org/10.1002/j.1460-2075.1982.tb01276.x>.
- Wang MS, Hecht MH. A completely *De Novo* ATPase from combinatorial protein design. *J Am Chem Soc.* 2020;142(36):15230–15234. <https://doi.org/10.1021/jacs.0c02954>.
- Watson JD, Milner-White EJ. A novel main-chain anion-binding site in proteins: the nest. A particular combination of ϕ , ψ values in successive residues gives rise to anion-binding sites that occur commonly and are found often at functionally important regions. *J Mol Biol.* 2002;315(2):171–182. <https://doi.org/10.1006/jmbi.2001.5227>.
- Winstanley HF, Abeln S, Deane CM. How old if your fold? *Bioinformatics.* 2005;21(Suppl 1):i449–i458. <https://doi.org/10.1093/bioinformatics/bti1008>.
- Yagi S, Padhi AK, Vucinic J, Barbe S, Shiex T, Nakagawa R, Simoncini D, Zhang KYJ, Tagami S. Seven amino acid types suffice to create the core fold of RNA polymerase. *J Am Chem Soc.* 2021;143(39):15998–16006. <https://doi.org/10.1021/jacs.1c05367>.