Helicase-Like Functions in Phosphate Loop Containing Beta-Alpha Polypeptides

Pratik Vyas[1], Olena Trofimyuk[1], Liam M. Longo[2], Fanindra Kumar Deshmukh[1], Michal Sharon[1] and Dan S. Tawfik[1]*

[1]Department of Biomolecular Sciences, Weizmann Institute of Science, Rehovot, Israel

[2]Tokyo Institute of Technology, Earth-Life Science Institute, Tokyo, Japan and Blue Marble Space Institute of Science

*Corresponding author. **E-mail:** dan.tawfik@weizmann.ac.il

Abstract

The P-loop Walker A motif underlies hundreds of essential enzyme families that bind nucleotide triphosphates (NTPs) and mediate phosphoryl transfer (P-loop NTPases), including the earliest DNA/RNA helicases, translocases and recombinases. What were the primordial precursors of these enzymes? Could these large and complex proteins emerge from simple polypeptides? Previously, we showed that P-loops embedded in simple βα repeat proteins bind NTPs, but also, unexpectedly so, ssDNA and RNA. Here, we extend beyond the purely biophysical function of ligand binding to demonstrate rudimentary helicase-like activities. We further constructed simple 40-residue polypeptides comprising just one β-(P-loop)-α element. Despite their simplicity, these P-loop prototypes confer functions such as strand separation and exchange. Foremost, these polypeptides unwind dsDNA, and upon addition of NTPs, or inorganic polyphosphates, release the bound ssDNA strands to allow reformation of dsDNA. Binding kinetics and low-resolution structural analyses indicate that activity is mediated by oligomeric forms spanning from dimers to high-order assemblies. The latter are reminiscent of extant P-loop recombinases such as RecA. Overall, these P-loop prototypes comprise a plausible description of the sequence, structure and function of the earliest P-loop NTPases. They also indicate that multifunctionality and dynamic assembly were key in endowing short polypeptides with elaborate, evolutionarily relevant functions.

Introduction

Protein machines, such as ATP synthetases, helicases or RecA recombinases, perform remarkable actions and are immensely complex [1,2]. Among other factors, complexity is manifested in the concerted action of multiple functional elements, the productive orchestration of which demands large, well-defined structures. This complexity is, however, at odds with the assumption that proteins emerged by duplication and fusion of relatively short and simple polypeptides [3,4]. Albeit, these polypeptides predated the last universal common ancestor (LUCA) and as such cannot be reconstructed by conventional phylogenetic methods nonetheless, despite nearly 4 BY of evolution, and extensive diversification of sequence, structure and function, these ancestral polypeptides are traceable by virtue of being the most conserved and functionally essential motifs of modern-day proteins [5]. Perhaps the most widely spread motif is the Walker-A P-loop [6]. Defined as GxxxxGK(T/S) or GxxGxGK, this motif underlies the most abundant and diverse protein class: the P-loop NTPases [6–9]. Structurally, the P-loop NTPase domain comprises a tandem repeat of at least five β-(loop)-α elements arranged in a αβα 3-layer sandwich architecture [7,8,10]. The loops connecting the C-termini of the β-strands to the N-termini of the following α-helices comprise the active site, while short loops link the β-(loop)-α elements to one another (here, unless stated otherwise, loops refer to the former). The Walker A P-loop is the key element mediating NTP binding and catalysis, and it uniformly resides within the first β-(loop)-α element, henceforth designated as β1-(P-loop)-α1.

P-loop NTPases are among the most ancient enzyme families – if not the most ancient enzyme family itself – and the first P-loop NTPase likely emerged at the transition from the RNA world to the primordial RNA-protein world [5,11–13]. Binding of phospho-ligands, foremost ATP and GTP, is the founding function of P-loop NTPases [14]. In agreement with early emergence, the key to phospho-ligand binding is not only the P-loop, but also a 'crown' of hydrogen bonds realized by the end of the Walker A motif (GK(T/S) that comprises the first turn of α1 (Refs. [14,15]). It has been accordingly hypothesized that polypeptides comprising the P-loop and its flanking secondary structural elements, namely, a β-(P-loop)-α segment, were the seed from which modern P-loop NTPases emerged [5,16–18]. However, that relatively short polypeptides can confer an evolutionarily meaningful function is far from obvious. Polypeptides generally lack structural volume and complexity, and cannot, in principle, align multiple functional elements as do intact protein domains [19]. Hence, identifying functional polypeptides, and especially polypeptides that are assigned as the seeding elements of the earliest enzyme families, is of paramount importance to our understanding of protein evolution.

Previously, we showed that an ancestral β1-(P-loop)-α1 motif, inferred via phylogenetic analysis of all known P-loop NTPase families, could be grafted onto a simple repeat scaffold comprised of four consecutive βα elements [16]. This grafting resulted in simple proteins that contained two P-loops. These bound via the P-loop motif ATP and GTP, but rather unexpectedly, also RNA and ssDNA. It is therefore likely that the primordial P-loop was a multifunctional phospho-ligand binder that could function in the absence of other auxiliary residues [16]. In the contemporary P-loop NTPases, single-stranded (ss) DNA binding is mostly conferred by specialized domains [20–23] that are distinct from the ATP binding P-loop [24]. And yet, in support of the P-loop's multifunctionality, we searched for and revealed examples of ssDNA binding mediated by the Walker A P-loop (detailed in 'Discussion'). Further, many of the P-loop NTPases families that date back to LUCA are involved in RNA/DNA remodeling, including helicases, RecA like recombinases, and translocases [25]. ATP synthetase, for example, is thought to have emerged from such nucleic acid-remodeling P-loop NTPases [26].

That ssDNA binding can be mediated by our P-loop prototypes, and the possible origins of P-loop NTPases as nucleic acid remodelers, guided us to extend our previous study [16] beyond the realm of ligand binding *per se*. We now (*i*) examine the potential of P-loop prototypes to exert, biologically and evolutionarily relevant helicase-like functions, *i.e.*, to mediate unwinding of dsDNA and strand exchanges, and (*ii*) identify the minimal structural context in which the P-loop can confer these relevant functions.

Results

P-loop prototypes mediate strand separation

dsDNA undergoes spontaneous local fluctuations referred to as "DNA breathing," [27,28] thus giving access to ssDNA binding proteins [29]. For instance, T4 helicases bind to and stabilize these transient open regions to unwind dsDNA [28] — a mode of unwinding also referred to as "passive unwinding" [30]. To test the potential of P-loop prototypes to facilitate such passive unwinding, we used a standard fluorescence-based helicase assay dubbed a molecular beacon [31] (**Figure 1A**). This assay uses a dsDNA segment comprised of a "beacon oligo" with a fluorophore at the 5' end and a quencher at the 3' end (the beacon sense strand) hybridized to an "unlabeled" complementary antisense-strand. Upon addition of a protein that promotes strand separation, owing to its complementary ends, the beacon strand collapses into an intramolecular hairpin (stem-loop structure). This brings together the fluorophore and the quencher, thus quenching the beacon's fluorescent signal. As shown later, this stem-loop product is likely stabilized by the fluorophore-quencher interaction [32] thus providing a simple way of monitoring DNA unwinding [31].

For the initial round of experiments, the originally described 110 amino acid protein that presents two P-loops (D-Ploop [16]), hereafter referred to as the 'intact prototype', was used. This intact prototype bound ssDNA preferably over double strand (ds) DNA [16]. As shown later, the P-loop prototypes also exhibit preference for TC-rich ssDNA while GA-rich sequences are barely bound. Accordingly, the dsDNA beacon comprised a TC-rich unlabeled antisense strand to promote binding and strand separation, while the beacon sense strand was GA-rich to avoid interference with the stem-loop formation and non-specific quenching due to protein binding (oligonucleotide sequences are provided in **Supplementary Table S1**). When added to the beacon dsDNA, the intact prototype induced strand separation as indicated by the decrease in fluorescence (**Figure 1B**). The P-loop prototype did not induce quenching of the fluorophore itself as indicated by a negligible change in fluorescence intensity with digested beacon dsDNA (**Figure 1B**; nonetheless, this non-specific quenching was subtracted to give the quenching traces reported below, see Methods). Strand-separation, as reported by this beacon dsDNA, was applied for testing various fragments of the intact P-loop prototype as described in the next section. Subsequently, using the most active fragment, the kinetics and mechanism of strand-separation were further investigated as described below (see *The N-$\alpha\beta\alpha$ P-loop prototype shows avid and cooperative strand separation*).

Structural minimization of the P-loop prototype

The intact P-loop prototypes were based on an 'ideal fold' [33] that reproduces the 3-layered $\alpha\beta\alpha$ sandwich architecture of P-loop NTPases [7,8,10,25]. They are composed of four $\beta\alpha$ elements with the ancestral β-(P-loop)-α element replacing the first and third elements [16]. We asked whether the intact prototype could be truncated while retaining its function. In other words, what is the minimal 'stand-alone' P-loop fragment that possesses phospho-ligand binding and strand-separation activity?

The repeat nature of our prototypes allowed multiple options for single P-loop constructs with varied structural topologies. In the first step, following our original report, the first half of the intact P-loop prototype (N-half) was tested [16]. Other 'half' constructs comprising two strands and two helices were also tested (**Figure 1C**, **Supplementary Figure S1**; DNA and amino acids sequences of all constructs are provided in **Supplementary Tables S2 and S3**). Secondly, we used circular permutation to generate half fragments that possessed a $\alpha\beta\alpha\beta$ architecture (**Figure 1C**). All of these constructs, and especially the circular permutants, showed high expression and purification yields (via a His-tag and Ni-NTA; **Supplementary Figure S2A**). Nonetheless, to maintain solubility at high protein concentrations, an osmolyte such as L-arginine was needed. We then truncated the circularly permuted 'half' constructs further to obtain the minimal constructs described below (**Figure 1C, Supplementary Figure S1**).

All ‘half’ constructs, and most of their truncated versions, mediated strand-separation to some degree (**Supplementary Figure S2B**). However, nucleic acid binding is known to be highly sensitive to ionic strength, and indeed the activity of most constructs was diminished in the presence of 100 mM NaCl (**Figure 1D, Supplementary Figure S2C**). We therefore set our assay conditions at 100 mM NaCl as a benchmark for robust activity. Intriguingly, under this stringent binding condition, a truncated variant dubbed ‘N-αβα’, comprising the ancestral β-(P-loop)-α element with one additional helix that precedes it, showed the most efficient and rapid strand separation (**Figure 1D**, red trace).

The minimal P-loop fragment

Can the N-αβα construct be further shortened? Given its polar nature, the primary role of N-terminal helix might be to enhance protein solubility (**Figure 1C, steps 4-6, Supplementary Table S3**). Indeed, partial truncations of this helix showed very similar activity to the intact N-αβα (N-αβα short_2, **Figure 1D**, green trace), while its complete truncation resulted in loss of the strand-separation activity (βα, **Figure 1D**, blue trace). Nonetheless, this shortest construct, which in effect comprises a β-(P-loop)-α fragment, tended to co-purify with nucleic acids, suggesting that it retained some binding capability. The functional prototype N-αβα short_2 contains only a short N-terminal segment (KRRGV) linked to the β-(P-loop)-α fragment by a short linker (GSG; **Supplementary Table S3**). It therefore appears that given a charged solubility tag, a single β-(P-loop)-α fragment can confer strand-separation activity. Also, worth noting is that N-αβα possesses low sequence complexity – it is made of only 12 amino acid types. Ten of these are considered abiotic amino acids, plus two cationic amino acids, Lys and Arg (excluding the 6xHistag for purification, and a tryptophan introduced for determining concentration by absorbance at 280 nm; **Supplementary Table S3**).

Overall, we found that the core β-(P-loop)-α motif can be placed in a variety of structural contexts, with varying strand topologies, and yet retain, or even show improved biochemical function (**Figure 1D**, **Supplementary Figures S1 and S2**). Given its minimal size and sequence complexity, its high expression, purity, and foremost its high strand separation activity, the N-αβα construct was employed as the representative P-loop prototype for the studies described below.

The N-αβα P-loop prototype shows avid and cooperative strand separation

Strand separation was dependent on the concentration of P-loop prototypes, in terms of both the rate and amplitude of fluorescence decay. At higher concentrations, unwinding was faster and complete (*i.e.*, reached baseline) whereas at lower concentrations unwinding was slower and partial (**Figure 2A, Supplementary Tables S4 and S5**). Isotherms obtained by plotting the end-

point fluorescence-decay values *versus* protein concentration indicated that strand separation is highly cooperative (sigmoidal curves and high Hill's coefficient) and occurs with apparent K_D values in the sub µM range (**Figure 2C, Supplementary Table S6**). The unwinding kinetics are complex since the dependence of the apparent rate constants on protein concentration was not linear (**Figure 2B**). Thus, strand separation appears to be a multiple step process. Specifically, the rate determining step changes non-linearly with protein concentration and at > 0.5 µM becomes essentially concentration independent (**Figure 2B**). As shown later, both of these phenomena (cooperative isotherms and complex kinetics) are in agreement with changes in quaternary structure upon ligand binding. Mutations in the key P-loop residues of N-αβα reduced both the rate and the amplitude of fluorescence decay, and accordingly showed loss of binding to ssDNA by ELISA (**Supplementary Figure S3**).

Unwinding of the beacon dsDNA is likely driven by the binding preference of the P-loop prototypes for ssDNA over dsDNA. This preference was originally observed with the intact prototype using an ELISA-like assay (with immobilized DNA and detection of binding with antibodies to the 6xHis tag) [16]. This result was confirmed here, also with the N-αβα fragment and the same oligonucleotides used in the beacon assay (**Supplementary Figure S4**). However, the ELISA assay exhibited high background. Thus, to more accurately measure binding to ss- versus ds-DNA, and to gain further insight into the mechanism of strand separation, we employed an assay, following a previously described setup [34], to simultaneously measure strand separation by fluorescence quenching (as in the beacon assay, **Figure 1**) and binding to the fluorescently labelled DNA strand(s) by fluorescence anisotropy.

We first tested the beacon dsDNA. Upon addition of the N-αβα prototype, a concentration dependent change in anisotropy was observed that was concomitant with the quenching signal (**Figure 2F**). The change in anisotropy occurs in parallel with quenching suggesting that strand separation and binding occur in a concerted manner, and that the anisotropy signal reports binding to ssDNA; or in other words: the dsDNA bound states do not accumulate. Indeed, under the conditions applied here the P-loop prototypes do not exhibit dsDNA binding, as confirmed with a dsDNA construct in which the quencher resides on the complementary strand. This construct should show an increase in fluorescence upon strand separation (as opposed to the decrease observed with the beacon dsDNA). However, no such increase was observed upon prototype addition (**Figure 2E**) likely due to the to the fluorophore-quencher interaction inducing higher duplex stability [32] (also indicated by strand exchange assays described below). Accordingly, this construct also showed a negligible change in anisotropy (**Figure 2E**).

The second principle governing dsDNA unwinding is the selectivity in binding the TC-rich antisense strand relative to the GA-rich sense strand of the beacon dsDNA. This selectivity was

observed by ELISA (**Supplementary Figure S4**) as well as by anisotropy (**Figure 2F**, **Supplementary Table S7**).

Finally, N-αβα does not quench the fluorescently labeled dsDNA in the absence of the quencher containing complementary strand (**Supplementary Figure S5**). This rules out the posibility that fluorescence quenching is driven by spurious interactions of the P-loop prototype with the fluorophore.

The P-loop prototypes facilitate strand exchange

Next, we assessed the potential of N-αβα to mediate strand-exchange – namely to accelerate the rate of exchange between a duplex bound ssDNA and an identical competing strand added in excess. We started with the dsDNA construct which, due to higher stability, did not show strand-separation (**Figure 2E; Figure 3A**). This construct also showed no exchange when mixed with 100-fold excess of the unlabeled complementary strand, not even after 24 hr. incubation (**Figure 3B**, pink bar). However, addition of N-αβα led to an increase of fluorescence indicating the displacement of the quencher strand by the unlabeled one, in a concentration and time dependent manner (6-24 hr.; **Figure 3B**).

That the fluorophore-quencher interaction stabilizes the duplex DNA and hinders strand separation and exchange was indicated in the swapped setup showing rapid exchange – namely, displacement of an unlabeled strand by a quencher-containing strand (**Figure 3C**). In contrast to the set-up shown in **Figure 3A**, N-αβα mediated strand-exchange at the minute time scale, with only 10-fold excess of the complementary strand (**Figure 3D**; **Supplementary Table S8**). Acceleration of strand exchange by N-αβα was also also observed in the same setup yet with 'linear' ssDNA oligos that do not form a hairpin (**Figure 3E**, **3F**). Here too, no siginficant exchange occurred within the time scale of the experiment in the absence of N-αβα, while addition of ≤ 2.5 μM N-αβα induced complete exchange within minutes (**Figure 3D**, **3F**). Note that the concentration of N-αβα remains in large excess (≥ 0.5 μM) over the excess of competing ssDNA (50 nM). Further, to corroborate that the strand-exchange reactions reached completion, the DNA products from the strand-exchange assays were resolved on native gels, after releasing the bound proteins by adding hexametaphosphate (**Figure 3G**, **3H**).

Overall, as expected, the energy gap between the start and end states dictates whether binding by the P-loop prototypes can shift the equilibrium between ds- and ssDNA, and also how fast the shift is. However, even in a more demanding setup N-αβα accelerated strand exchange (Figure 3B).

ssDNA release upon NTP and polyphosphate binding

Strand-separation as shown so far is, in essence, a shift in equilibrium in favor of P-loop prototype bound ssDNA. However, helicases are enzymes that turnover, namely they also release the bound DNA, typically upon ATP hydrolysis. The P-loop prototypes bind ssDNA and ATP via the same P-loop motif as suggested by mutations in the P-loop residues that abrogated binding of both [16]. Based on this observation, we asked if ATP addition could displace the ssDNA, thus also allowing the DNA to relax to its initial dsDNA state (**Figure 4A**). To this end, the N-αβα prototype was mixed with the beacon dsDNA and strand-separation was allowed to reach equilibrium. Upon subsequent addition of ATP or GTP, fluorescence reverted to its initial state indicating the release of the bound prototype and reversion to the starting dsDNA state (**Figure 4B**). Release by ATP occurred with an apparent K_D value in the range expected for a small ligand (K_D^{App} = 2.8 mM) and GTP was a more potent inducer of reversion with 2-fold tighter K_D^{App} (1.4 mM; **Figure 4C**). It seems, however, that release of the ssDNA is primarily conferred by the triphosphate group of these NTPs, and not the nucleoside moiety, as indicated by even tighter binding of triphosphate (K_D^{App} = 790 μM). Given that inorganic phosphoanhydrides were proposed to have preceded NTPs as life's energy coin, we tested polyphosphate – Kornberg's energy fossil [35] – as well as hexametaphosphate. N-αβα showed a strong preference for both polyphosphates (K_D^{App} = 13.7 μM; calculated for an average of 18 phosphates per molecule, and 5.6 μM for hexametaphosphate; **Figure 4C**). Addition of phosphate at molar concentrations that are >100-fold higher than hexametaphosphate did not release the bound protein (5.6 mM phosphate barely had any effect on releasing the bound ssDNA, whereas 5.6 μM hexametaphosphate, equivalent to 33.6 μM phosphate, induced complete release; **Figure 4C**). Thus, the regain of fluorescence seems to be driven by specific ligand binding rather than by nonspecific effects such as changes in ionic strength. The efficient release of prototype-bound ssDNA by hexametaphosphate was further corroborated by ELISA experiments showing that prior incubation of N-αβα with hexametaphosphate inhibited binding to ssDNA (**Figure 4D**).

Quaternary structural plasticity

The intact P-loop prototypes tend to form dimers [16]. We assumed that their fragments are more likely to do so and possibly form even higher-order oligomers. Despite extensive attempts, we could not obtain crystals of N-αβα, or of any other functional construct described here. However, we performed chemical crosslinking experiments that revealed that N-αβα forms higher order oligomers, spanning from dimers to hexamers (**Supplementary Figure S2D**). Native mass spectrometry (MS) revealed that N-αβα self-assembles to form an ensemble of even higher-order forms, with the predominant and unambiguously assigned assemblies being 10-mer and 30-mer (**Figure 5A**). Furthermore, MS/MS experiments confirmed the identity of these species, as

indicated, for example, by isolating the 30+ charge state of the 30-mer species, and inducing its dissociation into a highly charged monomer and a stripped 29-mer complex (**Figure 5B**). Dynamic light scattering (DLS) measurements also indicated that N-αβα exists in higher oligomeric form. The addition of inorganic polyphosphates to N-αβα results in an increase in the intensity of scattered light which translates to an increase in the average particle diameter (**Supplementary Figure S6**; **Supplementary Table S9**). In agreement with its lower apparent binding affinity (**Figure 4C**), GTP induced a smaller effect and the shift demanded higher concentrations. However, the diameter reported by DLS is only a proxy of the actual size, and the precise nature of the changes in the oligomeric assembly of N-αβα in response to ligand-binding needs to be further investigated.

Discussion

Previous studies have shown that NTP binding and modest catalysis can arise in polypeptides that comprise fragments of modern NTPases[16,36] or of other ancient domains [37]. This study extends our previous report [16] to demonstrate that polypeptides comprising the ancestral β-(P-loop)-α motif, with some minimal sequence additions to facilitate solubility, confer helicase-like functions. Our findings have several implications regarding the primordial P-loop NTPases, as discussed in the sections below.

From a generalist phospho-ligand binder to specialized enzymes

In contemporary helicases, translocases and RecA proteins, the P-loop mediates NTP binding and hydrolysis, while DNA binding is mediated by other surface loops or even by a separate domain [20–23]. In contrast, in our prototypes, the P-loop binds NTPs and inorganic polyphosphates, as well as nucleic acids. This multifunctionality is the key to the prototype's helicase-like action. Specifically, the strand-exchange assays demonstrate the ability of the P-loop prototypes to accelerate the exchange between a free and a duplex-bound DNA strand (**Figure 3**). However, in the unwinding experiments (**Figure 2**), binding of the P-loop prototypes merely shifts the equilibrium toward ssDNA. Nonetheless, the prototypes' action can be reversed by addition of phospho-ligands such as ATP, and foremost by inorganic polyphosphates. These binding-release cycles are not enzymatic turnovers, yet may be a first step toward a *bona fide* helicase.

This multifunctionality is surprising, as although nucleic acids have phosphate groups, they fundamentally differ from NTPs. Can vestiges of a generalist P-loop, and specifically of ssRNA/DNA binding, be found in extant P-loop NTPases? We searched the PDB for domains that belong to the P-loop NTPase lineage and have ssRNA/DNA bound in proximity to the P-loop (see Methods). This search identified at least two extant P-loop NTPase families: XPD helicase (**Figure 6A**), and polynucleotide kinase (**Figure 6B**), where the P-loop interacts with ssDNA, as it

does in our P-loop prototypes. In the case of XPD helicase, the canonical Walker A motif diverged and phosphate binding is mediated by a short motif, SGR, at the tip of α1. Remarkably, in case of polynucleotide kinase (**Figure 6B**), a canonical P-loop binds GTP/GDP as well as ssDNA (oligonucleotide) by a network of interactions facilitated by the P-loop residues. This, to our knowledge, is the only instance of an extant multifunctional P-loop that not only catalyzes phosphoryl transfer, but also binds ssDNA.

At present, in the absence of structures at atomic resolution, the precise mode of binding by our P-loop prototypes remains unknown. It seems, however, that the Walker A motif is not critical per-se. Specifically, we observed that the first Gly of the Walker A motif of our P-loop prototypes plays no role, and that the last three residues (GK(S/T)) seem most critical; however, even at these three positions, mutations reduce but not completely abolish binding (see Ref. [16] and **Supplementary Figure S3**). Indeed, the most rudimentary forms of phosphate binding make use of backbone amides, of glycine as well as other residues, and most critically, of bidentate interactions involving both backbone and side-chain H-bonds, foremost by Ser/Thr [14]. These elements are seen in XPD helicase (**Figure 6A**) and are likely to be the key to a generalist P-loop.

That the ancient P-loop was a 'generalist' that mediates multiple functions is in line with the hypothesis that the earliest enzymes were multifunctional [38]. A plausible explanation, in accord with the Dayhoff's hypothesis [3,4], is that such multi-functional primordial P-loop fragments were, at a later stage, duplicated and fused. This resulted in a tandem repeat of β-(P-loop)-α elements, where the P-loops comprise the active site. Further divergence allowed one P-loop to retain NTP binding and develop the ability to catalyze phosphoryl transfer, while the other P-loop(s) were repurposed to bind nucleic acids. This scenario is illustrated by XPD helicases wherein a P-loop ATPase domain, the DEAD domain (ECOD domain ID: e6fwrA2), is fused to the ssDNA binding helicase C_2 domain.

A minimal structural context for P-loop function

Bioinformatics analyses have pointed out the β-(P-loop)-α–β fragment as the recurrent structural/functional unit in dinucleotide binding proteins [5,39]. While we found that this fragment mediates strand separation, it was aggregation-prone, likely due to the highly hydrophobic β1 strand (**Supplementary Table S3**), and active only under non-stringent conditions. In contrast, the α–β-(P-loop)-α fragment was soluble and functional in the presence of salt. Having a helix preceding the β-(P-loop)-α fragment seems to present an advantage, primarily by enhancing solubility of this exceedingly hydrophobic βα fragment. We also note that although in most P-loop NTPase domains β1, from which the P-loop extends, comprises the N-terminus, an alpha helix

that precedes β1 is also in some P-loop families, *e.g.* in the F-type ATP synthases (F-groups, ECOD: 2004.1.1.53; Pfam: PF00006), Transcription termination factor Rho (F-groups, ECOD: 2004.1.1.53; Pfam: PF00006), and RecA (F-groups, ECOD: 2004.1.1.237; Pfam: PF00154). Moreover, the helicase-like function is retained despite truncation of most of this N-terminal helix, and even the β-(P-loop)-α fragment that could not be isolated was co-purified with bound nucleic acids. These results, along with the observation that ancestral β-(P-loop)-α motif can be embedded in various structural contexts while retaining function, suggest that β-(P-loop)-α is likely the minimal stand-alone seed from which P-loop NTPases could have emerged.

Dynamic self-assembly is key to function

We hypothesized that self-assembly is a critical bridging step that enables short polypeptides to be functional on their own prior to their duplication and fusion—a step that likely required advanced genetic and protein translation machineries [4]. Shorter polypeptides are unlikely to confer biochemical functions such as small ligands binding [40]. However, self-assembly, even in rudimentary forms such as amyloid fibers [41,42] or coacervates [43], can provide the operative volume and network of interactions necessary for biochemical function [42,44]. Although further studies are required to elucidate the mechanistic and structural basis of these assemblies, the propensity of the P-loop prototypes to self-assemble is evident (**Figure 5**). Foremost, oligomerization confers avidity, *i.e.*, binding via multiple P-loops per functional unit. Otherwise, binding at μM affinity by a solvent exposed loop would have been impossible. Changes between oligomeric states are also indicated by the cooperativity of the strand separation reactions, as manifested in high Hill coefficients and complex kinetics (**Figure 2**). Similar DNA-induced cooperative functions are documented for SSB proteins [45], helicases [46,47] and RecA proteins [48–51], and the latter two are also P-loop NTPases. Further, RecA has been shown to also possess strand-separation activity on short dsDNAs [52], and its recombinase function involves changes in oligomeric states, from monomers-dimers up to long filaments [1,2].

The potential role of P-loop prototypes in the pre-LUCA world

Whether proteins (metabolism) or nucleic acids (information) came first has been extensively debated. However, these scenarios are not necessarily mutually exclusive – both elements had to coexist in some rudimentary form well before the LUCA [53]. It is widely accepted that P-loop NTPases emerged early, possibly within an RNA-protein world [5,11–13,17,54,55]. Accordingly, RNA/DNA remodelers such as SFII helicases [55] and RecA constitute a major class of P-loop NTPases ascribed to the LUCA [7,8]. P-loop polypeptides may have therefore marked the early stages of cooperation between proteins and nucleic acids (other examples include polypeptides that interacted with ribosomal RNA [56–58]). Specifically, in the absence of protein-based

polymerases, nucleic acids must have replicated either non-enzymatically [59,60] and/or in a self-catalytic manner [61]. However, in such scenarios, a fundamental impediment for multiple rounds of replication is the stability of duplex RNA [62] (or DNA for that matter [63]). Heat can separate the strands, yet reannealing would be faster than any abiotic replication [64]. A plausible solution for unwinding RNA duplex products is short "invader" oligonucleotides [65]; however, how strand-displacement would initiate remains unresolved [65]. Thus, for multi-round replications, certainly of templates that are dozens of nucleotides long, an unwinding polypeptide is likely necessary. The duplex unwinding and strand displacement activities of our P-loop prototypes provide a plausible solution to these challenges.

Our results suggest that the strand-separation and exchange mediated by the P-loop prototypes is a thermodynamically driven process. The feasibility and rate of this process are dictated by the energetic gap between the initial and final DNA states. It is therefore likely that prior to the emergence of enzymatic function, primordial helicase-like polypeptides were fundamentaly ssDNA/RNA binding proteins that 'passively' unwound DNA duplexes by relying on transient exposure of ssDNA due to thermal fluctuations or/and DNA fraying [27,28]. Considering that the prototypes' affinity to the GA-rich strand of the beacon is low, it is likely that strand separation is primarily driven by binding to the TC-rich strand (**Figure 2F**). In fact, this is a commonly reported mode of unwinding for hexameric replicative helicases where the helicase-ring encircles one strand while the other strand is excluded [66]. The ability to bind ssDNA likely undelies the strand-exchange process as well (**Figure 3**) thus resembling the mode of action of extant RecA proteins [67]. Thus, although further studies will be required to decipher the precise mechanism of action of these P-loop prototypes, it appears that these rudimentary P-loop helicases hold tangible links to their modern descendants.

Inorganic polyphosphates as energy fossils

Our P-loop prototypes bind phosphate moieties with no apparent interactions with the nucleoside (ATP, GTP and triphosphate bind similarly; **Figure 4**). Foremost, long chain polyphosphate binds with 1000-fold higher affinity than ATP. Indeed, inorganic phosphoanhydrides were proposed to have been the abiotic energy precursors of NTPs [68]. Of particular interest is hexametaphosphate – a cyclic ring of six phosphates that also binds with µM affinity (**Figure 4**). In the primitive biotic world, polyphosphates could have served not only as an energy source, but also as a scaffold that facilitated the assembly and orientation of phospholipids, nucleic acids and proteins [69]. Further, trimetaphosphate has been extensively explored as an abiotic condensation reagent that promotes synthesis of peptides and nucleic acids [70]. Trimetaphosphate binds weakly to our P-loop prototypes, but it coexists in equilibrium with hexametaphosphate [70] that binds at µM affinity.

Thus, the mode of action of these P-loop prototypes is tantalizingly tailored to the requirements of a primordial world.

To conclude, Charles Darwin's immortal statement echoes: "from so simple a beginning endless forms most beautiful and most wonderful have been, and are being, evolved". The P-loop prototypes described here provide a glimpse of how these simple beginnings might have looked – short and simple polypeptides that are nonetheless linked to their modern descents in sequence, structure and function.

Materials and Methods

Protein expression and purification

Synthetic gene fragments encoding P-loop prototypes were purchased from *Twist Bioscience*. Cloning of synthetic genes and generation of mutant constructs, were performed as described previously [16]. All constructs had a C-terminal tag, comprising a Trp reside for concentration determination by measuring absorbance at 280 nm, (the P-loop prototypes have no aromatic residues) followed by 6xHis for purification (DNA and amino acid sequences are provided in **Supplementary Tables S2 and S3**). All proteins were expressed in *Escherichia coli* BL21 (DE3) cells. Cells were grown in LB media at 37 ºC with shaking (220 rpm) until OD600 ~0.6 and protein expression was induced by the addition of ITPG (1 mM final concentration) at 20 ºC. Following overnight induction at 20 ºC with shaking, cells were harvested by centrifugation (4000 g for 15 min at 4ºC) and resuspended in lysis buffer: 50 mM sodium phosphate, 20 mM imidazole, 500 mM NaCl, 6 M guanidine hydrochloride (pH 7.4), supplemented with the EDTA-free antiprotease cocktail supplied by Sigma. Cells were lysed by sonication, the debris was centrifuged down for 30 mins at 13000 rpm, and the supernatant was filtered through 0.22 µm membrane filter (Merck Millipore). Clarified lysates were loaded onto a HisTrapTM FF 5 mL column (GE Healthcare) using an AKTA FPLC. The guanidine hydrochloride that was added to dissociate bound nucleic acids was removed using a linear gradient of 12 column volumes of 50 mM sodium phosphate, 20 mM imidazole, 500 mM NaCl, pH 7.4. Bound proteins were eluted using one-step elution with 50 mM sodium phosphate, 500 mM imidazole, 500 mM NaCl, pH 7.4. Protein yield and purity was assessed by SDS-PAGE. Two rounds of dialysis were carried out to dialyze out the imidazole. The P-loop prototypes generally precipitated during the dialysis to remove imidazole. Precipitated protein samples were centrifuged, the pellet was collected and dissolved in buffer containing 50 mM Tris, 150 mM NaCl and 1 M L-arginine (pH 8). Purified prototypes stored in this buffer at 4 °C, at 100 to 200 µM concentrations, remained mostly soluble and active for periods of seven to ten days.

Strand separation and exchange assays

Synthetic DNA oligonucleotides for unwinding and strand-exchange assays were obtained from Integrated DNA Technologies, Inc, and their sequences, and which experiments they were used, are listed in **Supplementary Table S1**. The beacon dsDNA was prepared by hybridizing the beacon oligo (GA-beacon sense strand) and the unlabeled antisense oligo each at 10 µM in 50 mM Tris (pH 8). Hybridization was done by heating the oligo mixture at 95 °C for 5 mins followed by slow cooling overnight. Strand separation reactions were performed in a buffer containing 50 mM Tris (pH 8), and at the stringent binding conditions in 50 mM Tris (pH 8) with 100 mM NaCl. The reaction was initiated by adding P-loop prototypes (1 to 0.1 µM) to beacon dsDNA (5 nM) in 100 µl reaction volumes, at 24 °C, in Nunc-flat black 96-well plates. Fluorescence was monitored using Infinite M Plex microplate reader (TECAN) with an excitation/emission wavelength of 495/540 nm. Fluorescence decay values were normalized to the initial fluorescence before protein addition (F_0) to obtain F/F_0 values. The non-specific quenching values observed with 5 nM of beacon DNA digested overnight with Benzonase (Merck Millipore) was subtracted from the protein derived signals. Rate constants were determined by fitting the time dependent F/F_0 values to one-phase (Eq. 1).

$$\textbf{(1)}\quad F/F_{0\ (t)} = (F/F_0)^{\infty} + (1 - (F/F_0)^{\infty}) \times e^{-kapp*t}$$

where t is time, $(F/F_0)^{\infty}$ is the end-point value of F/F_0, $kapp$ is the apparent first-order rate constant.

For determining apparent binding affinities, the end-point F/F_0 values from the strand-separation experiments were derived for each protein concentration (P), and normalized (0 = the starting, fluorescent beacon substrate; 1 = the assay baseline, i.e., the fully quenched hairpin state). Normalized F/F_0 values were then plotted against protein concentration. The resulting data were fitted to a binding isotherm equation (Eq. 2; GraphPad Prism) with cooperativity, to derive the apparent dissociation constant (K_D^{App}) and the Hill coefficient (h).

$$\textbf{(2)}\quad F/F_{0\ (P)} = [P]^h/(K_D^{App} + [P]^h)$$

Strand exchange assays were carried out in 50 mM Tris (pH 8) with 100 mM NaCl. Stock solutions of dsDNA were prepared by hybridizing the FAM sense strand (FAM-GA sense strand or FAM-GA linear sense strand; **Supplementary Table S1**) with the BHQ-1 labeled or unlabeled TC-rich antisense strands (hairpin forming or linear; **Supplementary Table S1**) in 50 mM Tris pH 8, at 10 µM each, by heating and slow cooling as above. A mixture of this hybridized dsDNA (5 nM) and BHQ-1 or unlabeled antisense strands (50 nM) (hairpin forming or linear; **Supplementary Table S1**) was then used as the DNA substrate and reactions were initiated by

adding P-loop prototypes (0.5 to 10 μM). Changes in fluorescence were normalized to the initial state (prior to protein addition) and data were fit to Eq. 1 as described above.

Once the strand-exchange reactions described above have reached completion, 750 μM hexametaphosphate was added to release the bound proteins. The samples were immediately mixed with loading dye and the DNA products were loaded onto 10% (29:1 acrylamide:bisacrylamide) Tris-Borate EDTA (TBE) gels. The gels were run at 150 V for 40 min, at room temperature, in 1xTBE buffer and scanned using Amersham Typhoon scanner using the 488/525 nm laser-filter set at 400 volt PMT (photo multiplier tube) setting.

For determining apparent binding affinities of phospho-ligands, the fraction of released ssDNA (**Figure 4B**), for each ligand concentration (P), was normalized (1 = complete release and reversion of ssDNA to the starting, fluorescent beacon dsDNA substrate; 0 = no release, i.e., without added ligand). Normalized F/F_0 values were then plotted against the ligand concentration and the resulting data were fitted to a binding isotherm equation with cooperativity as described above (Eq. 2), to derive the apparent dissociation constant (K_D^{App}) and the Hill coefficient (h).

Enzyme-linked immunosorbent assay (ELISA)

ELISA experiments were performed as previously described [16] with some modifications. Stock solutions of 24 bp biotinylated dsDNA were prepared by hybridizing (see above) equimolar concentrations (100 μM) of biotinylated-24nt-ssDNA with the unlabeled (antisense) strand in buffer containing 50 mM sodium phosphate and 100 mM NaCl; pH 7 (oligo sequences are listed in **Supplementary Table S1**). Streptavidin coated plates (*StreptaWell, Roche*) were coated with 100 µl of 0.05 µM of biotinylated 24 bp dsDNA, or biotinylated-24nt-ssDNA, diluted in ELISA buffer (50 mM sodium phosphate and 100 mM NaCl; pH 7 and 1 % non-ionic detergent Nonidet P40 and 0.1% BSA, pH 7) for 1 hr at ambient temperature. In parallel, wells for background subtraction were left uncoated (incubated with 100 µl ELISA buffer). Following coating, wells were washed with ELISA buffer and 100 µl of P-loop prototypes solutions were added (at 0.25 to 0.06 µM) to both DNA-coated and background wells, and incubated for 1 hr. The unbound proteins were washed extensively with ELISA buffer (10 times) and 100 µl of 0.1 µg/ml HRP labeled mouse anti-His antibody (200 µg/ml, Santa Cruz Biotechnology) were added and incubated for 40 mins. The wells were washed extensively, 100 µl of substrate solution containing 1.25 mMol/L TMB and 2.21 mMol/L hydrogen peroxide, <1% dimethyl sulfoxide in 0.08 Mol/L acetate buffer at pH 4.9 (3, 3', 5, 5'-tetramethybenzidine; TMB; ES001, Millipore) was added and absorbance at 650 nm was monitored for 30 mins. The results are presented as endpoint values of OD^{650nm} (A_{650}) after subtraction of the background signal.

Fluorescence anisotropy measurements

Fluorescence anisotropy measurements were carried out using Cytation 5 (BioTek) instrument at 24 °C. Here, 5 nM of DNA (6-FAM labeled, or dually labeled (6-FAM and BHQ-1), single or double strand DNA constructs, as listed in **Supplementary Table S1**) were titrated with varying concentrations of N-αβα (0.1 to 2 µM) in 100 µL reactions. Fluorescence polarization was measured using the green filter set (485/20 nm for excitation, and 528/20 nM for emission) and top optic probe with a 510 nm dichroic mirror. The reactions were monitored for two hours, and the data processed with the Gen5.0 (BioTek) software. The resultant parallel (Ivv) and perpendicular (Ivh) emission intensity values were then used to calculate the samples' anisotropy with the standard equation (Eq. 3):

(3) $$r_{(P)} = (\mathrm{Ivv} - (G * \mathrm{Ivh}))/(\mathrm{Ivv} + (2 * G * \mathrm{Ivh}))$$

Where $r(P)$ is the steady-state anisotropy (after two hrs incubation) at a given protein concentration, Ivv and Ivh are the parallel and perpendicular emission intensities, respectively, and G is the grating factor that corrects for instrumental variation in detection of Ivh and Ivv. G is calculated as $\mathrm{Ivh}/\mathrm{Ivv}$.

The endpoint anisotropy values as calculated from eq. 3 were normalized to the anisotropy of the free DNA (the latter being equal to 0). The normalized values were then fitted to a binding isotherm equation with cooperativity (Eq. 4; GraphPad Prism), to derive the apparent dissociation constant (K_D) and Hill coefficient (h).

(4) $$r_{(P)} = r_{max} * [P]^h/(K_D + [P]^h)$$

r_{max} was set to a constant value of 0.2 and corresponds to the maximum observed anisotropy in our experimental setup; K_D is dissociation constant; h is Hill's coefficient.

Fluorescence anisotropy-quenching dual assays were performed using the same instrument and reaction conditions described above. For each protein concentration, the absolute level of fluorescence, and fluorescence polarization, were measured for a duration of two hours. F/F_0, and change in anisotropy, were calculated as described above, and then applied to equations 2 and 4, respectively, using GraphPad (Prism) as described above.

Chemical crosslinking

Chemical crosslinking of P-loop prototypes was performed with EDC (1-ethyl-3-(3-dimethylaminopropyl) carbodiimide, hydrochloride salt) [71]. Reactions were carried out as follows: purified P-loop prototypes were dialyzed against 10 mM MES buffer (pH 4.5), and diluted to 30

µM concentration in the same buffer. In 20 µl reaction volumes, 30 µM of protein were mixed with EDC (10 to 60 mM final concentration) dissolved in the above buffer. Reactions were incubated for 30 mins at room temperature and quenched with adding 50-fold molar excess (over cross-linker concentration) of Tris buffer (pH 8). Samples were mixed with SDS loading dye and analyzed on 16% SDS-PAGE gels run with Tris-glycine buffer.

Dynamic light scattering (DLS)

DLS analysis of N-αβα construct was performed using Malvern Zetasizer Nano-ZS particle analyzer (*Malvern Panalytical*). The P-loop prototypes were diluted to 1 µM in 50 mM Tris, 100 mM NaCl (pH 8). Following an equilibration time of 2 min, light scattering was monitored in a disposable cuvette in 1 ml volumes at 25 °C. The measurement duration was set to 'automatic', resulting in 11 to 17 runs for each measurement and three measurements were performed for each sample. The data were fit to a correlation function and analyzed by the method of cumulants [72] to obtain the z-average diameter (nm) (i.e. mean particle size). To test the effect of phospho-ligands, the prototype was mixed with different phosphate ligands, and light scattering was monitored as described above to obtain the z-average diameter.

Native mass spectrometry analysis

Nanoflow electrospray ionization MS and tandem MS experiments were conducted under nondenaturing conditions on a Q Exactive UHMR Hybrid Quadrupole-Orbitrap mass spectrometer (Thermo Fisher Scientific). Prior to MS analysis, the purified N-αβα was dialyzed against 150 mM ammonium acetate (pH 7.5) and diluted to a final concertation of 10 µM. Spectra were collected under positive ion mode, and conditions were optimized to enable the ionization and removal of adducts while retaining the noncovalent interactions of the proteins tested. In MS/MS experiments, the relevant m/z values were isolated and argon gas was admitted to the collision cell. Typically, aliquots of 2 µL of sample were electrosprayed from gold-coated borosilicate capillaries prepared in-house [73]. Spectra are shown using UniDec [74] for mass detection, and without smoothing or background subtraction.

The following experimental parameters were used on the UHMR Hybrid Quadrupole-Orbitrap platform: spray voltage 1.35 kV, inlet capillary temperature 250 °C. The trapping gas pressure was set to 5 corresponding to HV pressures of 1.5x10-4 mbar. To facilitate efficient desolvation of proteins, in source trapping was set to -100 V and high-energy collision dissociation was set to -10 eV. Mass spectra were recorded at a resolution of 3,125. Injection flatapole DC bias was set to 5 V. For tandem MS the precursor (6,540 m/z) was isolated and activated using the HCD cell at

100 eV. Instruments were externally mass-calibrated using a cesium iodide solution at a concentration of 2 mg/mL.

Identification of ssDNA binding by P-loop NTPases

All structures annotated by the ECOD database (version develop275) as containing a P-Loop domain (X-group 2004) were downloaded from the Protein Data Bank – 5,818 structures in total. Structures in which a P-Loop domain and ssDNA come within 3.5 Å of each other were identified using a short script implemented in Python. Hits were validated by visual inspection.

Acknowledgments

This research is funded by a Minerva Foundation grant for scientific cooperation between Germany and Israel. We are grateful to Prof. Ita Gruic-Sovulj for her insightful suggestion to simultaneously monitor quenching and anisotropy. We thank Dr. Guy Shmul for help with DLS analysis and for providing size standards and Dr. Gili Ben-Nissan for help with native mass-spectrometry experiments. Schematics for strand-separation, strand exchange and release assays were created with BioRender.com.

References

1. Bell JC, Kowalczykowski SC. RecA: Regulation and Mechanism of a Molecular Search Engine. *Trends Biochem Sci*. 2016;41(6):491-507. doi:10.1016/j.tibs.2016.04.002

2. Bleichert F, Botchan MR, Berger JM. Mechanisms for initiating cellular DNA replication. *Science (80-)*. 2017;355(6327):eaah6317. doi:10.1126/science.aah6317

3. Eck R V, Dayhoff MO. Evolution of the structure of ferredoxin based on living relics of primitive amino Acid sequences. *Science*. 1966;152(3720):363-366. doi:10.1126/science.152.3720.363

4. Romero MLR, Rabin A, Tawfik DS. Functional Proteins from Short Peptides : Dayhoff s Hypothesis Turns 50. *Angew Chem Int Ed Engl*. 2016;1980:15966-15971. doi:10.1002/anie.201609977

5. Alva V, Söding J, Lupas AN. A vocabulary of ancient peptides at the origin of folded proteins. *Elife*. 2015;4(DECEMBER2015):1-19. doi:10.7554/eLife.09410.001

6. 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.

7. Leipe DD, Wolf YI, Koonin E V., Aravind L. Classification and evolution of P-loop GTPases and related ATPases. *J Mol Biol*. 2002;317(1):41-72. doi:10.1006/jmbi.2001.5378

8. Leipe DD, Koonin E V., Aravind L. Evolution and classification of P-loop kinases and related proteins. *J Mol Biol*. 2003;333(4):781-815. doi:10.1016/j.jmb.2003.08.040

9. 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. doi:https://doi.org/10.1016/0968-0004(90)90281-F

10. Ogura T, Wilkinson AJ. AAA+ superfamily ATPases: common structure–diverse function. *Genes to Cells*. 2001;6(7):575-597. doi:10.1046/j.1365-2443.2001.00447.x

11. Koonin E V, Wolf YI, Aravind L. Protein fold recognition using sequence profiles and its application in structural genomics. *Adv Protein Chem*. 2000;54:245-275.

12. Söding J, Lupas AN. More than the sum of their parts: On the evolution of proteins from peptides. *BioEssays*. 2003;25(9):837-846. doi:10.1002/bies.10321

13. Ma B, Chen L, Ji H, et al. Characters of very ancient proteins. *Biochem Biophys Res*

Commun. 2008;366(3):607-611. doi:10.1016/j.bbrc.2007.12.014

14. 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*. 2020;117(10):5310 LP - 5318. doi:10.1073/pnas.1911742117

15. Watson JD, Milner-White EJ. A novel main-chain anion-binding site in proteins: the nest. A particular combination of phi,psi 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. doi:10.1006/jmbi.2001.5227

16. Romero Romero ML, Yang F, Lin Y-R, et al. Simple yet functional phosphate-loop proteins. *Proc Natl Acad Sci*. 2018;115(51):E11943 LP-E11950. doi:10.1073/pnas.1812400115

17. Berezovsky IN. Towards descriptor of elementary functions for protein design. *Curr Opin Struct Biol*. 2019;58:159-165. doi:https://doi.org/10.1016/j.sbi.2019.06.010

18. Laurino P, Tóth-Petróczy Á, Meana-Pañeda R, Lin W, Truhlar DG, Tawfik DS. An Ancient Fingerprint Indicates the Common Ancestry of Rossmann-Fold Enzymes Utilizing Different Ribose-Based Cofactors. *PLOS Biol*. 2016;14(3):e1002396. https://doi.org/10.1371/journal.pbio.1002396.

19. Matthews BW, Craik CS, Neurath H. Can small cyclic peptides have the activity and specificity of proteolytic enzymes? *Proc Natl Acad Sci U S A*. 1994;91(10):4103-4105. doi:10.1073/pnas.91.10.4103

20. Weinstock GM, McEntee K, Lehman IR. ATP-dependent renaturation of DNA catalyzed by the recA protein of Escherichia coli. *Proc Natl Acad Sci*. 1979;76(1):126 LP - 130. doi:10.1073/pnas.76.1.126

21. Menetski JP, Bear DG, Kowalczykowski SC. Stable DNA heteroduplex formation catalyzed by the Escherichia coli RecA protein in the absence of ATP hydrolysis. *Proc Natl Acad Sci*. 1990;87(1):21-25. doi:10.1073/PNAS.87.1.21

22. Gorbalenya AE, Koonin E V. Helicases: amino acid sequence comparisons and structure-function relationships. *Curr Opin Struct Biol*. 1993;3(3):419-429. doi:10.1016/S0959-440X(05)80116-2

23. Prentiss M, Prévost C, Danilowicz C. Structure/function relationships in RecA protein-mediated homology recognition and strand exchange. *Crit Rev Biochem Mol Biol*. 2015;50(6):453-476. doi:10.3109/10409238.2015.1092943

24. Saraste M, Sibbald PR, Wittinghofer A. The P-loop - a common motif in ATP- and GTP-binding proteins. *Trends Biochem Sci*. 2017;15(11):430-434. doi:10.1016/0968-0004(90)90281-F

25. Iyer LM, Leipe DD, Koonin E V, Aravind L. Evolutionary history and higher order classification of AAA+ ATPases. *J Struct Biol*. 2004;146(1-2):11-31. doi:10.1016/j.jsb.2003.10.010

26. Mulkidjanian AY, Makarova KS, Galperin MY, Koonin E V. Inventing the dynamo machine: the evolution of the F-type and V-type ATPases. *Nat Rev Microbiol*. 2007;5(11):892-899. doi:10.1038/nrmicro1767

27. von Hippel PH, Johnson NP, Marcus AH. Fifty years of DNA “breathing”: Reflections on

old and new approaches. *Biopolymers*. 2013;99(12):923-954. doi:10.1002/bip.22347

28. Phelps C, Lee W, Jose D, von Hippel PH, Marcus AH. Single-molecule FRET and linear dichroism studies of DNA breathing and helicase binding at replication fork junctions. *Proc Natl Acad Sci*. 2013;110(43):17320 LP - 17325. doi:10.1073/pnas.1314862110

29. Fei J, Ha T. Watching DNA breath one molecule at a time. *Proc Natl Acad Sci*. 2013;110(43):17173 LP - 17174. doi:10.1073/pnas.1316493110

30. Lionnet T, Spiering MM, Benkovic SJ, Bensimon D, Croquette V. Real-time observation of bacteriophage T4 gp41 helicase reveals an unwinding mechanism. *Proc Natl Acad Sci*. 2007;104(50):19790 LP - 19795. doi:10.1073/pnas.0709793104

31. Belon CA, Frick DN. Monitoring Helicase Activity With Molecular Beacons. *Biotechniques*. 2009;45(4):433-442. doi:10.2144/000112834.Monitoring

32. Zimmers ZA, Adams NM, Gabella WE, Haselton FR. Fluorophore-quencher interactions effect on hybridization characteristics of complementary oligonucleotides. *Anal Methods*. 2019;11(22):2862-2867. doi:10.1039/c9ay00584f

33. Koga N, Tatsumi-Koga R, Liu G, et al. Principles for designing ideal protein structures. *Nature*. 2012;491(7423):222-227. doi:10.1038/nature11600

34. Hiller DA, Fogg JM, Martin AM, Beechem JM, Reich NO, Perona JJ. Simultaneous DNA Binding and Bending by EcoRV Endonuclease Observed by Real-Time Fluorescence. *Biochemistry*. 2003;42(49):14375-14385. doi:10.1021/bi035520w

35. Kornberg A. Inorganic polyphosphate: Toward making a forgotten polymer unforgettable. *J Bacteriol*. 1995;177(3):491-496. doi:10.1128/jb.177.3.491-496.1995

36. Chuang WJ, Abeygunawardana C, Gittis AG, Pedersen PL, Mildvan AS. Solution Structure and Function in Trifluoroethanol of PP-50, an ATP-Binding Peptide from F1ATPase. *Arch Biochem Biophys*. 1995;319(1):110-122. doi:https://doi.org/10.1006/abbi.1995.1272

37. Pham Y, Kuhlman B, Butterfoss GL, Hu H, Weinreb V, Carter CWJ. Tryptophanyl-tRNA synthetase Urzyme: a model to recapitulate molecular evolution and investigate intramolecular complementation. *J Biol Chem*. 2010;285(49):38590-38601. doi:10.1074/jbc.M110.136911

38. Roy A. Jensen. Enzyme Recruitment in Evolution of New Function. *Annu Rev Microbiol*. 1976;(30):409-425.

39. Wierenga RK, De Maeyer MCH, Hoi WGJ. Interaction of Pyrophosphate Moieties with α-Helixes in Dinucleotide Binding Proteins. *Biochemistry*. 1985;24(6):1346-1357. doi:10.1021/bi00327a012

40. Corey DR, Phillips MA. Cyclic peptides as proteases: a reevaluation. *Proc Natl Acad Sci*. 1994;91(10):4106 LP - 4109. doi:10.1073/pnas.91.10.4106

41. Greenwald J, Riek R. On the Possible Amyloid Origin of Protein Folds. *J Mol Biol*. 2012;421(4):417-426. doi:https://doi.org/10.1016/j.jmb.2012.04.015

42. Rufo CM, Moroz YS, Moroz O V, et al. Short peptides self-assemble to produce catalytic amyloids. *Nat Chem*. 2014;6(4):303-309. doi:10.1038/nchem.1894

43. Longo LM, Despotović D, Weil-Ktorza O, et al. Primordial emergence of a nucleic acid binding protein via phase separation and statistical ornithine to arginine conversion. *Proc Natl Acad Sci*. January 2020:2020.01.18.911073. doi:10.1101/2020.01.18.911073

44. Makhlynets O V, Gosavi PM, Korendovych I V. Short Self-Assembling Peptides Are Able to Bind to Copper and Activate Oxygen. *Angew Chem Int Ed Engl*. 2016;55(31):9017-9020. doi:10.1002/anie.201602480

45. Dubiel K, Myers AR, Kozlov AG, et al. Structural Mechanisms of Cooperative DNA Binding by Bacterial Single-Stranded DNA-Binding Proteins. *J Mol Biol*. 2019;431(2):178-195. doi:10.1016/j.jmb.2018.11.019

46. Lee N-R, Kwon H-M, Park K, Oh S, Jeong Y-J, Kim D-E. Cooperative translocation enhances the unwinding of duplex DNA by SARS coronavirus helicase nsP13. *Nucleic Acids Res*. 2010;38(21):7626-7636. doi:10.1093/nar/gkq647

47. Byrd AK, Raney KD. Increasing the Length of the Single-Stranded Overhang Enhances Unwinding of Duplex DNA by Bacteriophage T4 Dda Helicase. *Biochemistry*. 2005;44(39):12990-12997. doi:10.1021/bi050703z

48. Lehman MC and IR. recA Protein-promoted DNA Strand Exchange. *J Biol Chem*. 1982;257(14):8523-8532.

49. Cox MM. Motoring along with the bacterial RecA protein. *Nat Rev Mol Cell Biol*. 2007;8(2):127-138. doi:10.1038/nrm2099

50. Kowalczykowski SC. Structural biology: Snapshots of DNA repair. *Nature*. 2008;453(7194):463-466. doi:10.1038/453463a

51. Ragunathan K, Joo C, Ha T. Real-time observation of strand exchange reaction with high spatiotemporal resolution. *Structure*. 2011;19(8):1064-1073. doi:10.1016/j.str.2011.06.009

52. Bianchi M, Riboli B, Magni G. E. coli recA protein possesses a strand separating activity on short duplex DNAs. *EMBO J*. 1985;4(11):3025-3030. https://pubmed.ncbi.nlm.nih.gov/3905387.

53. Preiner M, Asche S, Becker S, et al. The future of origin of life research: Bridging decades-old divisions. *Life*. 2020;10(3). doi:10.3390/life10030020

54. Goncearenco A, Berezovsky IN. Protein function from its emergence to diversity in contemporary proteins. *Phys Biol*. 2015;12(4):45002. doi:10.1088/1478-3975/12/4/045002

55. Anantharaman V, Koonin E. V AL. Comparative genomics and evolution of proteins involved in RNA metabolism. *Nucleic Acids Res*. 2002;30(7):1427-1464. doi:10.1093/nar/30.7.1427

56. Lupas AN, Alva V. Ribosomal proteins as documents of the transition from unstructured (poly)peptides to folded proteins. *J Struct Biol*. 2017;198(2):74-81. doi:https://doi.org/10.1016/j.jsb.2017.04.007

57. Lanier KA, Roy P, Schneider DM, Williams LD. Ancestral Interactions of Ribosomal RNA and Ribosomal Proteins. *Biophys J*. 2017;113(2):268-276. doi:https://doi.org/10.1016/j.bpj.2017.04.007

58. Frenkel-Pinter M, Haynes JW, Mohyeldin AM, et al. Mutually stabilizing interactions between proto-peptides and RNA. *Nat Commun*. 2020;11(1):3137. doi:10.1038/s41467-

020-16891-5

59. Deck C, Jauker M, Richert C. Efficient enzyme-free copying of all four nucleobases templated by immobilized RNA. *Nat Chem*. 2011;3(8):603-608. doi:10.1038/nchem.1086

60. Adamala K, Szostak JW. Nonenzymatic template-directed RNA synthesis inside model protocells. *Science*. 2013;342(6162):1098-1100. doi:10.1126/science.1241888

61. Bartel DP, Doudna JA, Usman N, Szostak JW. Template-directed primer extension catalyzed by the Tetrahymena ribozyme. *Mol Cell Biol*. 1991;11(6):3390-3394. doi:10.1128/mcb.11.6.3390

62. Ross PD, Sturtevant JM. THE KINETICS OF DOUBLE HELIX FORMATION FROM POLYRIBOADENYLIC ACID AND POLYRIBOURIDYLIC ACID. *Proc Natl Acad Sci U S A*. 1960;46(10):1360-1365. doi:10.1073/pnas.46.10.1360

63. Teichert JS, Kruse FM, Trapp O. Direct Prebiotic Pathway to DNA Nucleosides. *Angew Chemie Int Ed*. 2019;58(29):9944-9947. doi:10.1002/anie.201903400

64. Szostak JW. The eightfold path to non-enzymatic RNA replication. *J Syst Chem*. 2012;3(1):2. doi:10.1186/1759-2208-3-2

65. Zhou L, Kim SC, Ho KH, et al. Non-enzymatic primer extension with strand displacement. Nilsen TW, Marletta MA, Nilsen TW, Chaput JC, eds. *Elife*. 2019;8:e51888. doi:10.7554/eLife.51888

66. Yuan Z, Georgescu R, Bai L, Zhang D, Li H, O'Donnell ME. DNA unwinding mechanism of a eukaryotic replicative CMG helicase. *Nat Commun*. 2020;11(1):688. doi:10.1038/s41467-020-14577-6

67. van der Heijden T, Modesti M, Hage S, Kanaar R, Wyman C, Dekker C. Homologous Recombination in Real Time: DNA Strand Exchange by RecA. *Mol Cell*. 2008;30(4):530-538. doi:10.1016/j.molcel.2008.03.010

68. Kornberg A, Rao NN, Ault-riché D. Inorganic Polyphosphate : A MOLECULE OF MANY FUNCTIONS. 1999:89-125.

69. Brown MRW, Kornberg A. Inorganic polyphosphate in the origin and survival of species. *Proc Natl Acad Sci*. 2004;101(46):16085-16087.

70. Baccolini G. The Possible Role of Cyclic Pentacoordinate Phosphorus Intermediates in the Origin and Evolution of Life. are Phosphoric Anhydride and Trimetaphosphates Prebiotic Reagents? *Phosphorus Sulfur Silicon Relat Elem*. 2015;190(12):2173-2186. doi:10.1080/10426507.2015.1083569

71. Lepvrier E, Doigneaux C, Moullintraffort L, Nazabal A, Garnier C. Optimized protocol for protein macrocomplexes stabilization using the EDC, 1-ethyl-3-(3-(dimethylamino)propyl)carbodiimide, zero-length cross-linker. *Anal Chem*. 2014;86(21):10524-10530. doi:10.1021/ac502561e

72. Koppel DE. Analysis of Macromolecular Polydispersity in Intensity Correlation Spectroscopy: The Method of Cumulants. *J Chem Phys*. 1972;57(11):4814-4820. doi:10.1063/1.1678153

73. Kirshenbaum N, Michaelevski I, Sharon M. Analyzing large protein complexes by structural mass spectrometry. *J Vis Exp*. 2010;(40). doi:10.3791/1954

74. Marty MT, Baldwin AJ, Marklund EG, Hochberg GKA, Benesch JLP, Robinson C V. Bayesian Deconvolution of Mass and Ion Mobility Spectra: From Binary Interactions to Polydisperse Ensembles. *Anal Chem*. 2015;87(8):4370-4376. doi:10.1021/acs.analchem.5b00140

Figures

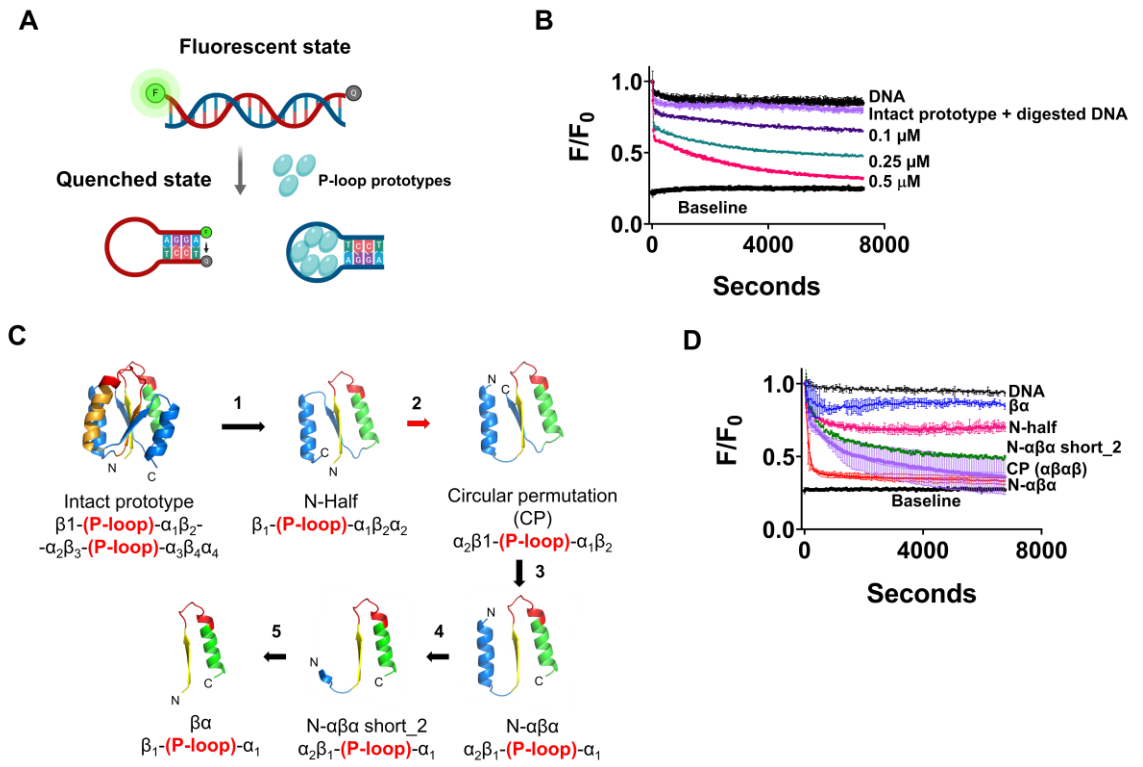

Figure 1. The molecular beacon assay reports the strand separation of P-loop prototypes.

A. Simplified schematic of the strand-separation molecular beacon assay [31]**.** In the initial dsDNA state, the quencher of the beacon strand is held apart from the fluorophore resulting in energy transfer (high fluorescence). The preferential binding of P-loop prototypes to ssDNA induces strand separation, allowing the beacon strand to assume a hairpin state wherein the fluorophore is quenched. The oligos used in this assay, and the assays described in the subsequent figures are listed in **Supplementary Table S1**.

B. A representative strand separation experiment using the intact 110 residue P-loop prototype (D-Ploop; Ref [16]). Strand separation is reported by the change in FRET signal (i.e. fluorescence quenching in our experimental set-up) upon addition of the P-loop prototype at increasing protein concentrations. All assays were performed with 5 nM beacon dsDNA, in 50 mM Tris (pH 8), at 24 °C. Shown are normalized F/F_0 values, whereby the initial fluorescence of beacon dsDNA prior to protein addition takes the value of 1. Baseline represents the signal of the fully quenched hairpin beacon. Digested DNA represents the signal upon addition of 0.5 μM P-loop prototype to the beacon dsDNA pretreated with Benzonase nuclease. Traces were fitted to a biphasic exponential decay model (described in '**Supplementary**

Information' section) and the apparent rate constants are given in **Supplementary Table S4**.

C. New P-loop prototypes were constructed by systematic truncation and circular permutation of the intact prototype. The numbered arrows indicate sequential steps in the engineering of new constructs as follows: *1.* Truncation of the intact 110 residue prototype into half [16] (N-half indicates N-terminal half). *2.* Circular permutation (red arrow) of N-half to a construct with '$\alpha\beta\alpha\beta$' architecture. *3.* Truncation of the C-terminal β-strand to give N-$\alpha\beta\alpha$. *4-5.* Incremental truncations of N-terminal helix of N-$\alpha\beta\alpha$ down to a $\beta\alpha$ fragment. The structural models indicate the ancestral P-loop element in yellow (β1), red (the Walker P-loop) and green (α1), while the remaining parts are in blue.

D. Strand separation by truncated P-loop prototypes, at 1 μM protein concentration, and under the stringent condition (plus 100 mM NaCl; other assay conditions as in panel A). The lines represent the average from two to six independent experiments and the bars represent the S.D. values.

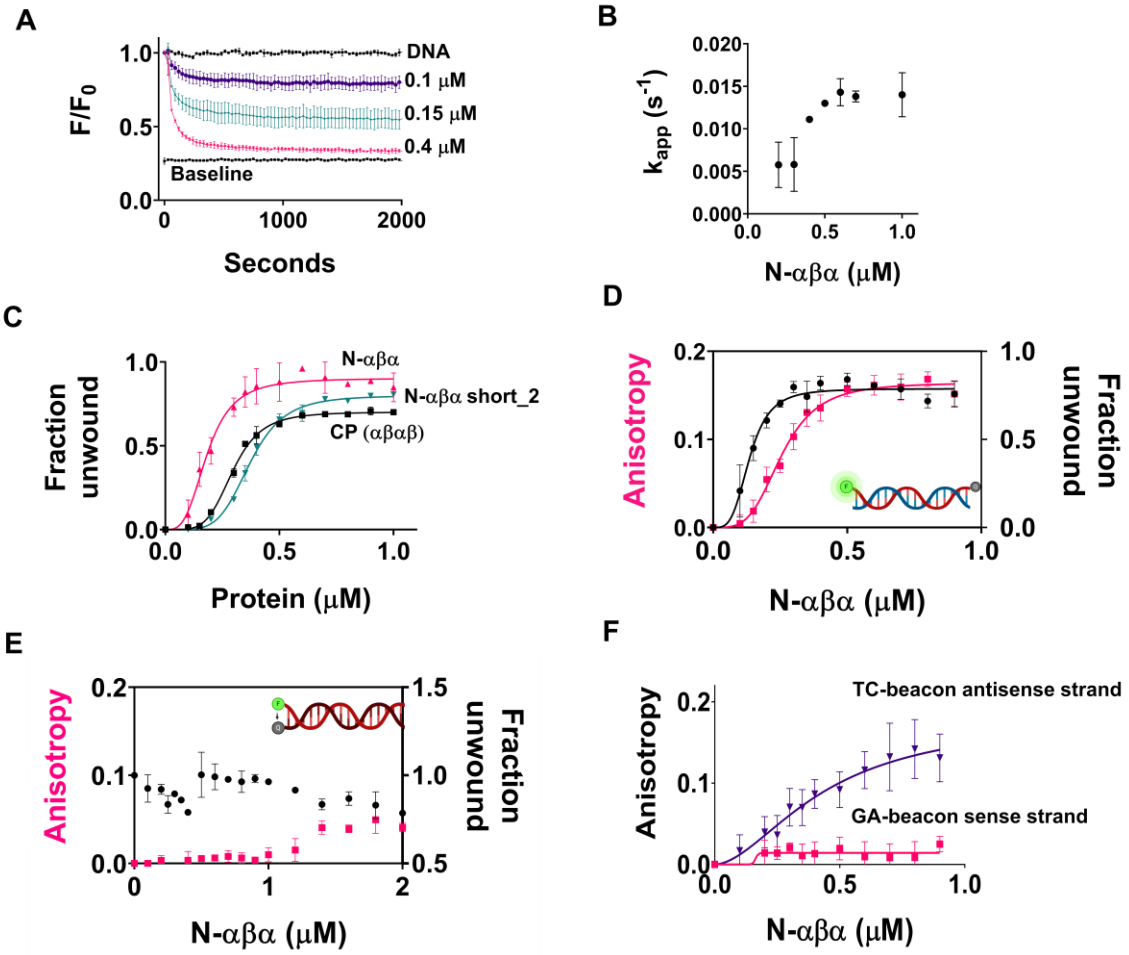

Figure 2. Binding affinities and kinetics of strand-separation by P-loop prototypes.

A. A representative strand-separation experiment with varying concentrations of N-αβα (the beacon dsDNA is described in **Figure 1A**, and assay conditions are as in **Figure 1D**). The values shown are the average from 2 to 6 independent experiments and the bars represent the S.D. values (the number of experiments in the subsequent panels is denoted as *n*).

B. Traces shown in panel *A* were fitted to a one phase exponential decay model (Eq. 1, Methods) and the apparent rate constants were plotted against protein concentration (**Supplementary Table S5)**.

C. Binding isotherms of P-loop prototypes (for their topology, see **Figure 1C**). End-point F/F_0 values from strand separation experiments (*e.g.* panel A) were normalized (fluorescence of the starting beacon dsDNA equals 0, and of the fully quenched ssDNA hairpin equals 1) to derive the relative fraction of unwound dsDNA, and were then plotted vs. protein concentration (n = 2 to 6; error bars represent SD values).

D. Simultaneous monitoring of the changes in fluorescence intensity and in anisotropy of the beacon dsDNA. Changes in anisotropy (pink trace) and in the fraction unwound (black trace; as in panel *C*) were measured as described in Methods (n = 4 to 8; error bars represent SD values).

E. As above but with a dsDNA construct having the fluorophore and quencher on the opposite strands (here, strand separation should result in an increase in fluorescence rather than decrease as in panel *D* and in all other beacon assays; n = 4 to 8; error bars represent SD values).

F. Binding to the individual strands of the dsDNA beacon as monitored by fluorescence anisotropy assay. Changes in fluorescence anisotropy were monitored upon titrating the GA-beacon sense strand (pink trace) or the TC-beacon antisense strand (purple trace) with varying concentrations of N-$\alpha\beta\alpha$. Assay conditions were as described in *D* and *E* (n = 4 to 8; error bars represent SD values).

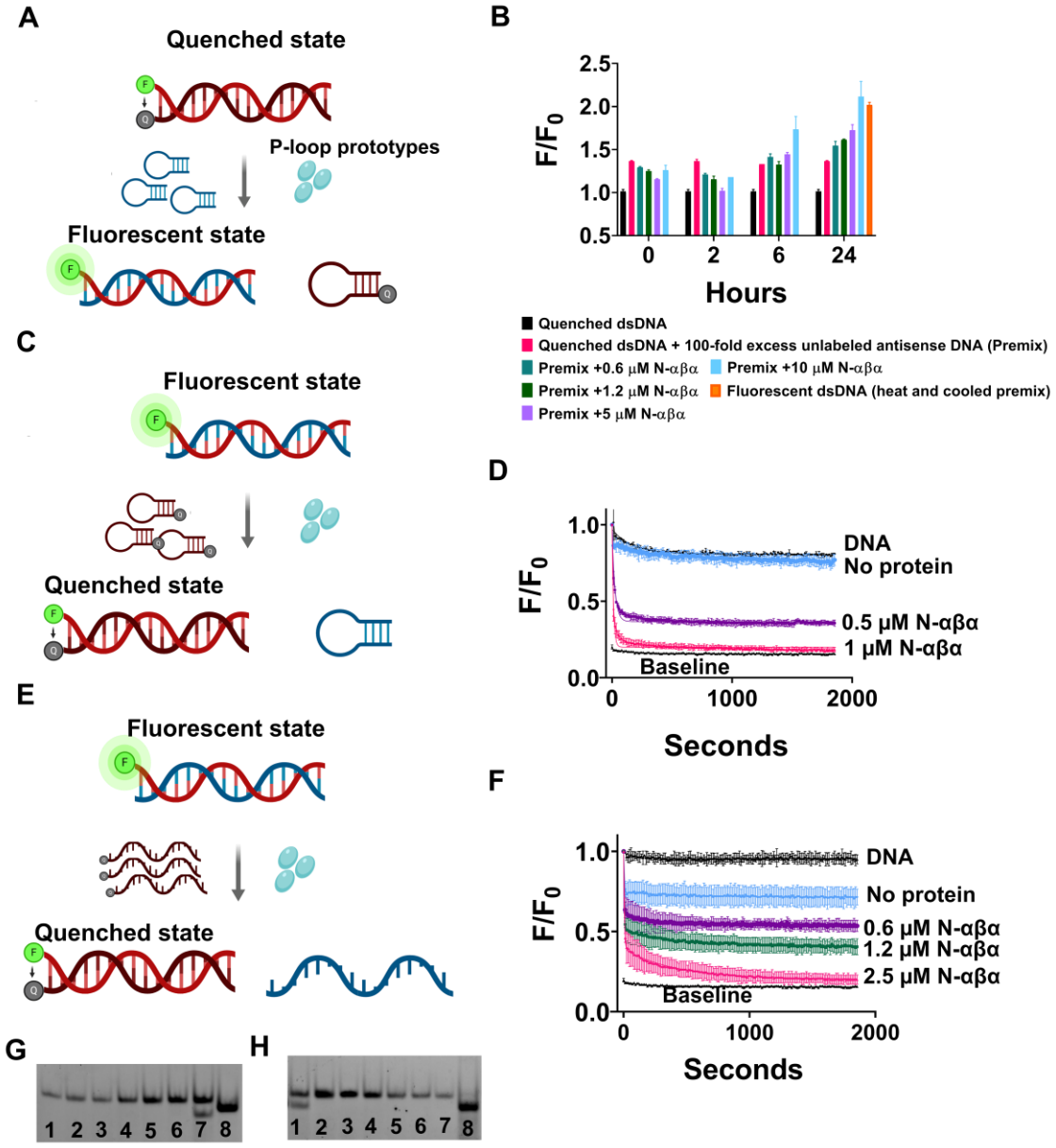

Figure 3. The N-αβα P-loop prototype mediates strand exchanges.

A. A schematic of the strand exchange reaction between a quenched dsDNA and a non-labelled complementary strand (hairpin-forming).

B. The exchange reaction was monitored by increase in fluorescence upon addition of N-αβα to a premix of quenched dsDNA and 100-fold excess of unlabeled complementary antisense

strand. Fluorescence was normalized (F/F_0) to quenched dsDNA (equals 1) and plotted vs. time (n = 2 to 4; error bars represent SD values).

C. A schematic of the strand exchange reaction between a fluorescent dsDNA and a quencher containing complementary strand (hairpin forming).

D. Fluorescence quenching (F/F_0) was monitored upon addition of N-αβα to a premix of fluorescently labeled dsDNA and 10-fold excess of a competing strand (n = 2 to 4; error bars represent SD values). Data were fit to a one-phase exponential decay (**Supplementary Table S8**).

E. A schematic of the strand exchange reaction between a fluorescent dsDNA and a quencher containing complementary strand (linear).

F. Fluorescence quenching (F/F_0) was monitored upon addition of N-αβα to a premix of fluorescently labeled dsDNA and 10-fold excess of a competing strand (n = 2; error bars represent SD values). Data were fit to a two-phase exponential decay.

G. Native EMSA gel of the strand-exchange reaction shown in panel *B*. Bands shown in the lanes are as follows: *1.* Quenched dsDNA (FAM-GA sense strand + BHQ-1 antisense strand); *2.* Premix (quenched dsDNA + 100-fold excess of unlabeled antisense strand); 3-6. Premix + N-αβα (0.6, 1.2, 5 and 10 μM respectively); *7.* Fluorescent beacon dsDNA; *8.* Fluorescent ssDNA (FAM-GA sense strand) (**Supplementary Table S1**).

H. Native EMSA gel of strand-exchange reaction shown in panel *F*. Bands shown in the lanes are as follows: *1.* Fluorescent dsDNA (FAM-GA-linear sense strand + unlabeled linear antisense strand); *2.* Premix (fluorescent dsDNA + 10-fold excess of BHQ-1-linear antisense strand); *3-6*. Premix + N-αβα (0.6, 1.2, 2.5 and 5 μM respectively); *7.* Quenched dsDNA (FAM-GA-linear sense strand + BHQ-1-linear antisense strand); *8.* Fluorescent ssDNA (FAM-GA-linear sense strand) (**Supplementary Table S1**).

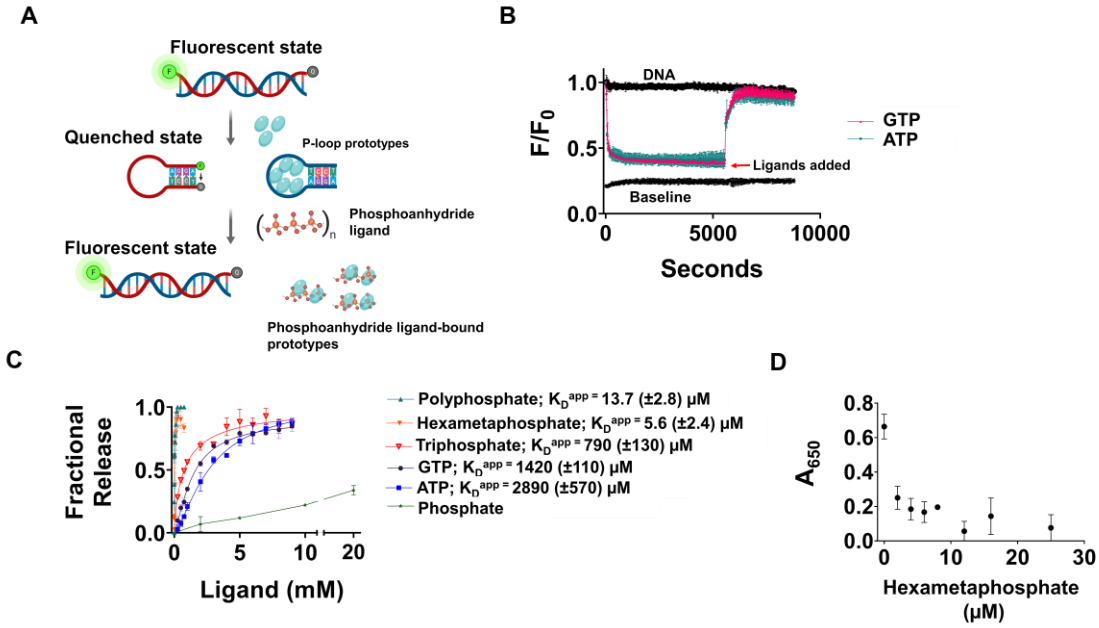

Figure 4: Release of bound ssDNA by phosphoranhydride ligands.

A. A schematic description of the induction of strand separation by a P-loop prototype (1st step) followed by the displacement of the bound ssDNA by various phosphoanhydrides (P_n) and relaxation to the initial dsDNA state (2nd step).

B. Representative strand separation experiments with 0.4 μM N-αβα added at time zero. Subsequent addition of ATP (turquoise) and GTP (pink, both at 9 mM concentration) leads to the release of the bound ssDNA and its relaxation to the initial dsDNA state.

C. The apparent affinities (K_D^{App}) of binding of various phosphoanhydride ligands to P-loop prototype N-αβα. The fraction of released ssDNA upon addition of phospho-ligands was calculated by normalizing the F/F_0 values from the plot in panel B. Complete release (=1) corresponds to the initial unbound dsDNA state and no release (=0) corresponds to the steady-state value of F/F_0 prior to ligand addition. The apparent binding affinities (K_D^{app}) were calculated by Eq. 2 (see Methods). Vertical error bars, and values in parenthesis, represent standard deviation from two independent experiments.

D. Inhibition of ssDNA binding. Preincubation of 0.125 μM N-αβα with increasing concentrations of hexametaphosphate shows abrogation of binding to 24 base biotinylated ssDNA (**Supplementary Table S1**) in an ELISA-format (n = 2 to 4; error bars represent SD values).

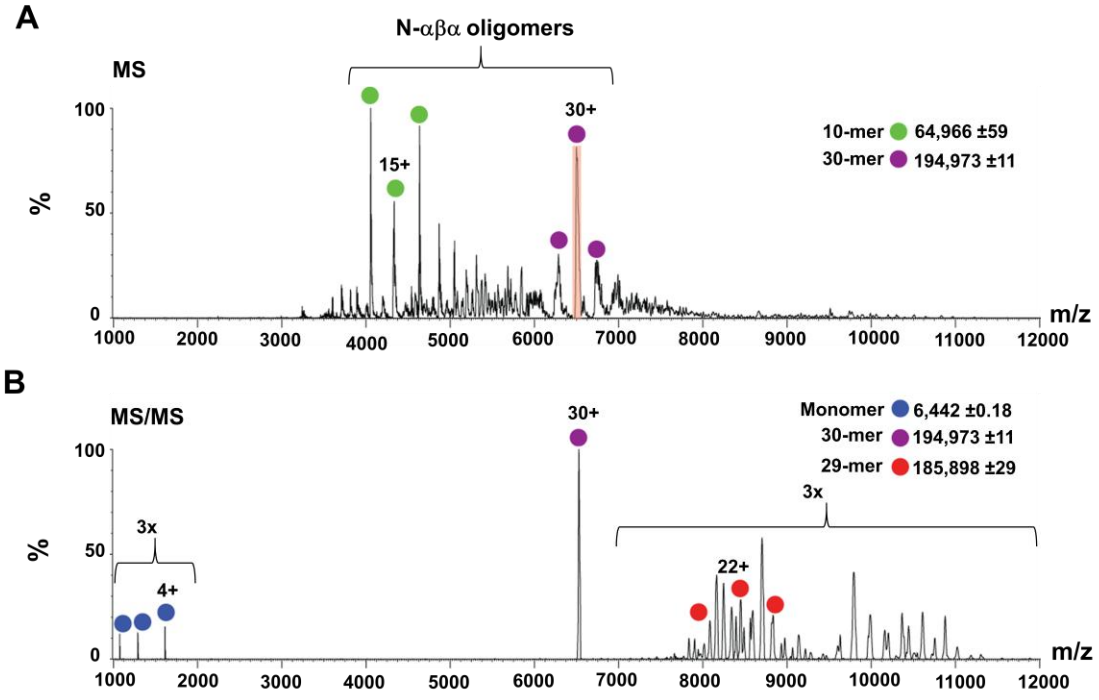

Figure 5. Quaternary structural characterization of N-αβα.

A. Native MS analysis of N-αβα prototype. Under non-dissociative conditions, charge series corresponding to 10 and 30 N-αβα oligomers were unambiguously assigned in the MS spectrum (green and purple dots, respectively). The 30+ charge state that was selected for tandem MS analysis is shadowed in red.

B. Tandem MS of the N-αβα 30-mer oligomer releases a highly charged monomer (blue; at the lower m/z range) and a stripped 29-mer oligomer (red), confirming the oligomer stoichiometry. The spectra are magnified 3-fold above 7,000 and below 2,000 *m/z*.

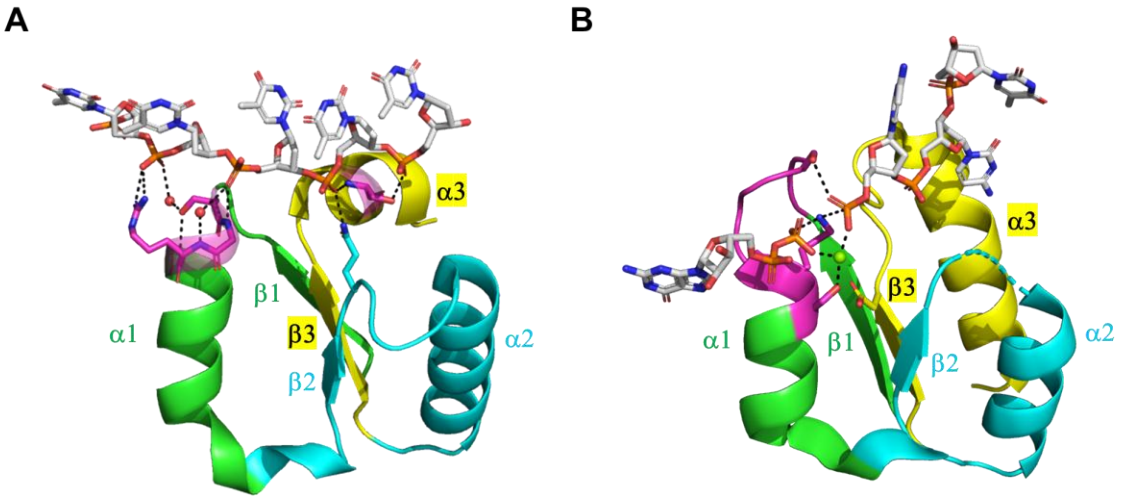

Figure 6: Instances of P-loops binding to ssDNA in extant P-loop NTPases.

A. The Helicase_C_2 domain of XPD helicases (F-groups, ECOD: 2004.1.1.106, Pfam: PF13307). Shown here is a fragment taken from a representative structure (ATP dependent DinG helicase; ECOD domain ID: e6fwrA1, residues 448-703; the fragment shown, residues 535-599, spans from β1 to α3). The strand topology of this domain follows the simplest P-loop NTPase topology (2-3-1-4-5). Its P-loop resides at the tip of α1, in the very same location as the Walker A, but its sequence is non-canonical (SGR, in magenta). Shown are direct as well as water-mediated interactions between the residues of the P-loop and phosphate groups of the ssDNA oligonucleotide (waters shown as red spheres). A glutamine at the tip of β2 (in cyan), and a serine at the N-terminus of α3 (in magenta) provide additional anchoring points for the ssDNA.

B. Bacterial polynucleotide kinase (F-groups, ECOD: 2004.1.1.32; Pfam: PF13671; strand topology: 2-3-1-4-5). Shown here is a fragment of the P-loop NTPase domain from a representative structure (*Clostridium thermocellum* polynucleotide kinase; ECOD domain ID: e4mdeA1, residues 1-170; the shown fragment, residues 8-37 and 62-98, spans from β1 to α3, with residues 38-61 truncated for clarity). The canonical Walker A motif resides between β1 and α1 (in magenta, GSSGSGKS) and binds GTP. This representative structure shows the product complex of GDP•Mg^{+2} and the phosphorylated ssDNA. The lysine and serine residues of the P-loop motif at the tip of α1 provide bridging interactions, coordinated by an Mg^{2+} ion (green sphere), between the β phosphate of the bound GDP and the phosphate group at the 5'-OH of the ssDNA.