On the Emergence of P-Loop NTPase and Rossmann Enzymes from a Beta-Alpha-Beta Ancestral Fragment

- This article is dedicated to the memory of Michael G. Rossmann

Liam M. Longo[1,2,3], Jagoda Jabłońska[1], Pratik Vyas[1], Rachel Kolodny[4,*], Nir Ben-Tal[5,*], Dan S. Tawfik[1,*]

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

[2]Tokyo Institute of Technology, Earth-Life Science Institute, Tokyo, Japan

[3]Blue Marble Space Institute of Science

[4]Unversity of Haifa, Department of Computer Science, Haifa, Israel

[5]Tel Aviv University, George S. Wise Faculty of Life Sciences, Department of Biochemistry and Molecular Biology, Tel Aviv, Israel

[*]To whom correspondence should be addressed: trachel@cs.haifa.ac.il, bental@tauex.tau.ac.il, dan.tawfik@weizmann.ac.il

Abstract

Dating back to the last universal common ancestor (LUCA), the P-loop NTPases and Rossmanns now comprise the most ubiquitous and diverse enzyme lineages. Intriguing similarities in their overall architecture and phosphate binding motifs suggest common ancestry; however, due to a lack of global sequence identity, these families are considered independent emergences. To address this longstanding dichotomy, we systematically searched for 'bridge proteins' with structure and sequence elements shared by both lineages. We detected homologous segments that span the first βαβ segment of both lineages and include two key functional motifs: (*i*) the 'Walker A' phosphate binding loop, the hallmark of P-loop NTPases, and its Rossmann equivalent, both residing at the N-terminus of α1; and (*ii*) an Asp at the tip of β2. The latter comprises the 'Walker B' aspartate that chelates the catalytic metal in P-loop NTPases, or the canonical Rossmann β2-Asp which binds the cofactor's ribose moiety. Tubulin, a Rossmann GTPase, demonstrates the ability of the β2-Asp to take either one of these two roles. We conclude that common P-loops/Rossmann ancestry is plausible, although convergence cannot be completely ruled out. Regardless, we show that both lineages most likely emerged from a polypeptide comprising a βαβ segment carrying the above two functional motifs, a segment that comprises the core of both enzyme families to this very day.

Introduction

In 1970 Michael Rossmann reported the structure of the first αβα sandwich protein, lactate dehydrogenase[1]. This NAD-utilizing enzyme would later become representative of what is now known as the '*Rossmann fold*'[2]. About a decade later, on the basis of a sequence analysis, another major αβα sandwich protein that utilizes phosphorylated nucleosides was proposed[3], which is now known as the P-loop NTPase, or '*P-loop*' for short. The importance of these two protein lineages cannot be overstated: Both have diversified extensively, each yielding more than 200 families associated with more than 120 different enzymatic reactions (see *Methods*). Furthermore, these two lineages are ubiquitous across the tree of life[4]. Accordingly, essentially all studies aimed at unraveling the history of protein evolution concluded that these enzymes emerged well before the last universal common ancestor (LUCA), and were among the very first, if not the first, enzyme families[4–9]. Indeed, both P-loops and Rossmanns are dubbed nucleotide binding domains because they both make use of phosphorylated ribonucleosides such as ATP or NAD, and other pre-LUCA cofactors such as SAM[10].

As elaborated in the next section, P-loop and Rossmann proteins share a number of similar features, but also some distinct differences. Given their pre-LUCA origin, a common P-loop/Rossmann ancestor – even if it did exist at some point – is surely lost to time. Both lineages emerged during the so-called "big bang" of protein evolution, an event that marks the birth of the major protein classes, yet occurred too early to be reconstructed by phylogenetic means[8]. Thus, a fundamental enigma surrounding the birth of the first enzymes is whether the Rossmann and the P-loop lineages diverged from a common ancestor, or perhaps, given that they both make use of phosphorylated ribonucleosides, have converged to similar structural and functional features. The former is a common, evolutionarily appealing scenario, yet the latter is as common and tangible of a scenario[11,12].

To address this longstanding question, we performed a detailed analysis looking for indications of common ancestry with respect to the core elements of these two classes, namely their most conserved and functionally critical segments. Indeed, global sequence homology between these lineages cannot be detected. As such, large-scale analyses of protein homology[7], including SCOPe[13] and the Evolutionary Classifications of Protein Domains (ECOD) database[14] classify P-loop NTPases and Rossmanns as independent evolutionary emergences. However, loss of detectable sequence homology would be expected between lineages that split in the distant past, especially if both have diverged extensively as is the case for P-loops and Rossmanns. Nonetheless, structural anatomy[10] and sophisticated ways of detecting sequence homology may assign common ancestry in highly diverged lineages on the basis of a few common sequence-structure features[11,12,15–17]. Further, parallel evolution may operate, with relics of an ancient common ancestor surfacing sporadically in contemporary proteins, thus resulting in detectable sequence and/or structural homology. Thus, if P-loops and Rossmanns do share common ancestry, we might expect the existence of "bridge proteins"; that is, proteins belonging to one lineage with features that are distinct for the other lineage.

Here, we report the detection and analysis of common features and bridge proteins between P-loops and Rossmanns. The existence of such common features and bridge proteins does not completely rule out convergence, but it strengthens the argument in favor of common ancestry. Foremost, our results suggest what the key features of the ancestor(s) might have been, and indicates that even if these lineages emerged independently, their ancestors shared the very same features. To best frame this analysis, however, we must first dissect the canonical features of Rossmann and P-loop proteins.

P-loop and Rossmann – similar but distinctly different

Both P-loops and Rossmanns adopt the αβα 3-layer sandwich fold (**Figure 1**). This fold, which comprises a parallel β-sheet sandwiched between two layers of α-helices, is among the most ancient, if not the most ancient, protein folds known[5,8,18,19]. In essence, αβα sandwich proteins comprise a tandem repeat of β-loop-α elements, where the loops form the active-site (hereafter referred to as the "functional" or "top" loops; **Figure 1A**). The minimal P-loop or Rossmann domains comprise five β-loop-α elements linked via short "connecting" or "bottom" loops that generally have no functional role. Although many domains have six strands, and sometimes more, we will hereafter consider the minimal 5-stranded core domain for simplicity.

While the overall fold is conserved, the topology – specifically, the strand order of the interior β-sheet – differs between the Rossmanns and P-loops. The Rossmann topology (β3-β2-β1-β4-β5) has a pseudo-2-fold axis of symmetry between β1-β3 and β4-β5 (or β1-β2-β3 and β4-β5-β6; **Figure 1B**). However, in the P-loop topology, at least two strands are swapped (**Figure 1C)**. The core, most common topology is β2-β3-β1-β4-β5 (ref. [4]); although, as discussed below, P-loops can adopt several different strand topologies.

The second shared hallmark is that both P-loops and Rossmans bind phosphorylated ribonucleoside ligands as substrates, co-substrates or cofactors (hereafter, phospho-ligands). While the overall mode binding of phospho-ligands differs, the binding modes of their phosphate moieties share a few similarities (**Figure 2A**,**B**): (i) The phosphate is bound by the first β-loop-α element which resides in the center of the domain (hereafter, β1-(phosphate binding loop)-α1, or β-(phosphate binding loop)-α, for simplicity); (ii) both phosphate binding loops mediate binding via a "nest" of hydrogen bonds formed by backbone amides at the N-terminus of the first canonical α-helix (α1) as well as via residues from the loop itself; and (iii) both phosphate binding loops are glycine-rich sequences with similar patterns: the

canonical Rossmann motif is GxGxxG, while the canonical P-loop motif, dubbed Walker A, is GxxGxGK[T/S].

To avoid confusion, P-loop is used here to refer to the evolutionary lineage of P-loop NTPases only. When referring to the phosphate binding element of a protein, with no relation to a specific protein lineage, phosphate binding loop (or PBL) is used. Hence, P-loop PBL relates to the phosphate binding loop of P-loop NTPases (the Walker A motif), and Rossmann PBL to the Rossmann's phosphate binding loop.

However, despite similar phosphate binding elements, the mode of phospho-ligand binding by Rossmanns and P-loops is fundamentally different, and relates to important functional differences between the two lineages. Although Rossmann and P-loop proteins both utilize RNA-derived metabolites, the phosphate groups of these metabolites play a fundamentally different role. P-loops primarily catalyze phosphoryl transfer (including to water, *i.e.*, hydrolysis) and thus most often operate on ATP and GTP with the help of a metal dication (mostly Mg^{2+}, but also Ca^{2+}). Rossmanns, on the other hand, primarily use NAD(P), with the phosphate moieties serving only as a handle for binding, while the redox chemistry occurs elsewhere (*e.g.*, the nicotinamide base in NAD^{+}). These functional differences are accompanied by a number of structural differences in the phosphate binding mode: The P-loop Walker A is a relatively long, surface-exposed loop that extends beyond the protein's core and wraps, like the palm of a hand, around the phosphate moieties of the ligand (**Figure 2A**). The Rossmann PBL, however, is short and forms a flat interaction surface (**Figure 2B**); the phosphate groups interact mostly with the N-terminus of α1 via a highly conserved and ordered water molecule[20]. Foremost, the orientation of the phospho-ligand being bound is different: The nucleoside moiety in Rossmanns is oriented "inside", *i.e.*, in the direction of the β-sheet core, whereas in P-loops it points "outside", *i.e.*, away from the protein interior – an approximately 180-degree rotation compared to Rossmanns.

The above difference in orientation relates to differences in interactions that Rossmanns and P-loops make with parts of the ribonucleotide ligands other than their phosphate moieties. In the canonical Rossmann binding site, both the phosphate moiety and the ribose moiety are bound. The phosphate interacts with the Rossmann PBL at the N-terminus of α1 while the ribose moiety is held in place by an Asp/Glu residue at the tip of β2 (**Figure 2B**). This acidic residue forms a unique bidentate interaction with the 2' and 3' hydroxyls of the ribose moiety, and was shown to be already present as aspartate in the earliest Rossmann ancestor (hereafter β2-Asp)[10]. In P-loops, on the other hand, the core of the αβα domain does not interact with the ribose, instead making more extensive, catalytically-oriented interactions with the phosphate moieties (via the Walker A P-loop, **Figure 2A**, as well as other key residues). Foremost, phospho-ligand binding also involves a metal cation, typically Mg^{2+}, ligated by two key conserved residues: the hydroxyl of the canonical serine or threonine of the Walker A motif (GxxGxGK[**S/T**]) and by an Asp/Glu residing on the tip of an adjacent β-strand—the "Walker B" motif, located at the tip of either β3 or β4 (see Ref. [21] for a detailed analysis).

A shared β-(phosphate binding loop)-α evolutionary seed?

Individually, any one of the shared features described above may relate to convergence. The αβα sandwich fold has likely emerged multiple times, independently[22]. The key shared functional feature, namely the phosphate binding site at the N-terminus of an α-helix, and the Gly-rich phosphate binding motifs, were likely favored at the early stages protein evolution because they effectively comprise the only mode of phosphate binding that can be realized with short and simple peptides[23]. Thus, that Rossmanns and P-loops share Gly-rich loops, and the same mode of phosphate biding, may also be the outcome of convergence, especially

because the overall mode of binding of their phospho-ligands fundamentally differs (**Figure 2A**, **B**).

Curiously, however, the phosphate binding site is located in the first β-loop-α element of both Rossmanns and P-loops. In fact, the β1-α1 location of the PBL is seen not only in P-loop and Rossmann proteins, but also in Rossmann-like protein classes such as flavodoxin and HUP. However, a closer examination reveals that, although rare, phosphate binding in αβα sandwich folds can occur at alternative locations, suggesting that there is no inherent, physical constraint on its location. An illustrative example can be found in the HUP lineage (ECOD X-group 2005) – a monophyletic group of 3-layer αβα sandwich, Rossmann-like proteins that includes Class I aminoacyl tRNA synthetases[24]. Most families within this lineage achieve phosphate binding at the tip of α1, as do Rossmann and P-loop proteins. However, two families, the universal stress protein (Usp) family (F-group 2005.1.1.145) and electron transport flavoprotein (ETF; F-group 2005.1.1.132) both use the tip of α4 (**Figure 3**). Intriguingly, α4, resides on the other side of the β-sheet, just opposite to α1. Accordingly, this change in the PBL's location results in a mere flip of ATP's phosphate groups, while preserving other features of the binding site, including the adenine's location and the anchoring of the ribose moiety to β1 and β4 (**Figure 3** mid panel). Thus, from a purely biophysical point of view, α1 and α4 are equivalent locations for phosphate binding. Nonetheless, phosphate binding at α4 is a rare exception, and never seen to our knowledge, in either P-loop or Rossmann proteins. This suggests that the positioning of the PBL at the tip of α1 in both Rossmann and P-loop proteins is a signal of shared ancestry rather than convergence. As a minimum, the identification of α4 as a feasible alternative supports a model of emergence of both lineages from a seed βαβ fragment as outlined further below. By this scenario, α4 only emerged at a later stage, well after phosphate binding had been established at α1.

A shared β2-Asp motif?

As outlined above, the β1-PBL-α1 likely represents a primordial polypeptide that could later be extended to give the modern αβα sandwich domains[7,25]. However, there are several indications that the ancestral, seeding peptide(s) of both P-loops and Rossmanns also contained β2[7]. In the case of the Rossmann, inclusion of one additional element in the seeding primordial peptide has been proposed: an Asp at the tip of β2 that forms a bidentate interaction with the ligand's ribose hydroxyls (**Figure 2B**)[10]. The putative Rossmann common ancestor thus comprises β-PBL-α-β-Asp. Might such a fragment also be the P-loop ancestor?

In fact, both families make use of an Asp residue at the tip of the β-strand just next to β1 – in P-loops this residue is the above-mentioned Walker B motif (**Figure 2A**). Is this feature also a sought-after signature of shared ancestry? In the simplest P-loop topology, the Walker-B-Asp resides at the tip of the β-strand which is adjacent to β1, as it is in Rossmanns. Thus, putting aside the connectivity of strands, both P-loop and Rossmann possess a functional core of two adjacent strands, one from which the phosphate binding loop extends and the other with an Asp at its tip (**Figure 2A**, **B**). However, because in P-loops the strand topology is swapped, in the primary sequence, the Walker B Asp typically resides at the tip of β3 (in the simplest topology described in **Figure 1C**, and at the tip β4 in another common topology as detailed below). As elaborated later, variations in topology of P-loops support a model by which additional β-loop-α elements got inserted into the ancestral β-PBL-α-β-Asp seed fragment such that what was initially β2 became β3 (or even β4 in other P-loop families).

However, even if we put the question of topology aside for the moment, there remains the fundamental functional difference between the P-loop Walker B-Asp (binding of a catalytic dication) and the Rossmann β2-Asp (ribose binding; **Figure 2A** and **2B**, respectively). Can this difference be reconciled? This question might be answered by identifying cases of parallel evolution, or "bridging proteins." Specifically, is there an example of a Rossmann NTPase enzyme? Does it use a catalytic metal – and might that metal be bound by the β2-Asp?

Tubulin - a parallelly evolved Rossmann NTPase

As explained above, Rossmanns typically use the ligand's phosphate moiety as a binding handle, whereas P-loops perform chemistry on the ligands' phosphate groups. Thus, to discover bridging proteins, we looked at the minority of Rossmann families that do act as NTPases. In all but one of these, the NTP is bound in the canonical Rossmann mode, namely with the NTP's ribose moiety bound to the β2-Asp. However, one family, tubulin, is an outlier. Tubulin is a GTPase first discovered in eukaryotes. With time, bacterial and archaeal homologues tubulin were discovered, indicating that this lineage originated in the LUCA[26,27]. Tubulin has undisputable hallmarks of a Rossmann[28], as noted originally[29], and is categorized as such (ECOD family: 2003.1.6.1). The strand topology is distinctly Rossmann (3-2-1-4-5), with a phosphate binding loop located between β1 and α1. We further note that binding of GTP's phosphate groups is mediated by a water molecule bound to the N-terminus of α1 (**Figure 2C**), as in canonical Rossmanns[20] (**Figure 2B**) but not in P-loops. However, as noticed by those who solved the first tubulin structures, GTP is oriented differently compared to the nucleotide cofactors bound by other Rossmanns[28]. Our examination reveals that tubulin binds GTP in the P-loop NTPase mode – namely, with the nucleoside pointing away from the domain's core (**Figure 2C**). Indeed, tubulin's phosphate binding loop is truncated relative to

other Rossmanns and adopts a conformation akin to a tight hairpin (**Figure 2**, **Figure S1A**). In fact, tubulin has a second phosphate binding loop that resides at the tip of α4 and has a critical role in catalysis, indicating that α4 can readily take the role of phosphate binding as seen in the HUP families described above (**Figure 3**).

Foremost, in accordance with GTP binding occurring in a P-loop-like orientation, the β2-Asp interaction with the ribose, a hallmark of Rossmanns, is absent in tubulin (**Figure 2C**). Furthermore, as in P-loops, the triphosphate moiety of GTP is hydrolyzed *via* a water activated by a ligated metal dication. The mode of binding of the catalytic dication of tubulin (often Ca^{2+}) is particularly intriguing: In P-loops, the catalytic dication is ligated by an Asp/Glu on a β-strand adjacent to the P-loop ("Walker B motif"; **Figure 2A**). In tubulin, the canonical Asp on the tip of β2, which usually binds ribose, takes a Walker B role by interacting with the dication (**Figure 2C**). This structural interpretation is supported by the positioning of the Asp residues across many tubulin structures (**Figure S1B**; **Table S1**) and by previous results indicating that the β2-Asp is essential for tubulin's catalytic activity[30]. Note that in many tubulins Asn is seen at the β2 tip position, though this β2-Asn also ligates the dication, either directly or via a water molecule (**Table S1**).

Tubulin therefore comprises an intriguing case of a Rossmann that evolved an NTPase function by reorienting the NTP substrate to bind in the P-loop NTPases mode, thereby repurposing the canonical Rossmann β2-Asp to ligate a catalytic metal. Put differently, tubulin shows that the functional differences between the P-loop Walker B and the Rossmann β2-Asp can be reconciled – the Rossmann β2-Asp of tubulins can function as a Walker B-Asp.

Shared *theme*s between Rossmanns and P-loops

Encouraged by tubulin, we endeavored to look for additional evidence for bridging proteins between Rossmanns and P-loops, ideally with respect to not only structure but also sequence. To this end, we employed the concept of an *agile theme* – short stretches for which alignments are statistically significant ((≥ 20 residues; HHSearch E-value $<10^{-3}$) yet with the flanking regions showing no detectable sequence homology. In the context of this work, we specifically searched for shared themes found in structures that belong to Rossmanns (X-Group 2003) on the one hand and P-loops (X-group 2004) on the other. By focusing the sequence homology search on evolutionarily-distinct domains, and by using bait sequences derived from validated sequence themes[17], the sensitivity and accuracy of this approach exceeds that of standard HMM-based searches (further details, and themes detected between other X-groups are described in a forthcoming manuscript). Given this stringent statistical threshold, only a few shared themes were detected, all involving the P-loop enzyme HPr kinase/phosphatase (F-Group 2004.1.2.1; PDB: 1ko7) (**Figure 4A**). Few different Rossmann F-groups share a theme with this P-loop, with sorbitol dehydrogenase (F-Group 2003.1.1.417; PDB: 1k2w) and short chain dehydrogenase (F-Group 2003.1.1.332; PDB: 3tjr) showing the highest overlap (**Figure 4**).

HPr kinase/phosphatase is a bifunctional bacterial enzyme that catalyzes the phosphorylation of a signaling protein (HPr) and its dephosphorylation[31]. The P-loop domain comprises its C-terminal domain and carries the kinase function (hereafter Hpr kinase). Remarkably, the Walker B aspartate of Hpr kinase resides at the tip of β2, rather than at β3 or β4 as in the canonical P-loops. Consequently, although no such constraint or steering was applied to the search algorithm, the detected shared theme encompasses an intact β1-PBL-α1-β2-Asp element in the Rossmann proteins (where this element in canonical) as well as in this unique P-loop family (**Figure 4A**). As expected, this element is conserved in both the P-loop Hpr kinase and in the Rossmann families, with the Gly residues of the phosphate

binding motifs, and the β2-Asp's being almost entirely conserved (**Figure 4B**). Curiously, also entirely conserved is a Gly residue at the C-terminus of α1 that has a purely structural role (Figure **4C**-**E**). This result underscores the significance of the β1-PBL-α1-β2-Asp motif as the shared evolutionary seed of both Rossmanns and P-loops (detailed in the next section).

Consistent with the idea of parallel evolution, these bridging P-loop and Rossmann proteins seem to be at the fringes of their respective lineages. In the case of HPr kinase, the active site is characterized by a canonical Walker A motif, though the Walker B-Asp seems uncharacteristically situated at the tip of β2 (**Figure 4D**). Further, in P-loop families with the simplest topology, the Walker B-Asp resides at the tip of a β-strand that structurally resides next to β1 (β3, **Figure 2A**). However, in most P-loops, another strand, typically β4, is inserted between β1 and the strand carrying the Walker B Asp (**Figure S2**; Ref. [4]). Hpr kinase belongs to this second category; however, its intervening strand is highly unusual – an anti-parallel β-strand inserted between β1 and β2. In the primary sequence, the intervening strand is an N-terminal extension, and thus upstream to β1 (**Figure 4C**; annotated as β-1). Indeed, HPr kinase is classified as an outlier with respect to the greater space of P-loop proteins. The P-loop X-group in ECOD (X-group 2004) is split into two topology groups (T-groups): *P-loop containing nucleoside triphosphate hydrolases*, which includes 196 F-groups that represent the abundant, canonical P-loop proteins, and *PEP carboxykinase catalytic C-terminal domain*, which is comprised of just three F-groups. HPr kinase is classified as part of the latter. As discussed below, the variation in topology of HPr also highlights the structural plasticity of the P-loop fold with respect to insertions.

The sorbitol dehydrogenase and short chain dehydrogenase both have a canonical Rossmann strand topology and β2-Asp. Homology modeling of the enzyme-NAD^+ complex, and inspection of closely related structures, suggest that binding of the NAD^+ cofactor is also canonical (**Figure 4D**). However, the phosphate binding loop of three of these Rossmann

proteins is nonstandard (GxxxGxG instead of the canonical Rossmann which is GxGxxG; **Figure 4A**). Further, although the structural positioning of the last two glycine residues is rather similar to that in canonical Rossmann proteins, the GxxxGxG motif results in the loop connecting β1 and α1 being extended, thereby bearing higher resemblance to the P-loop PBL (**Figure 4F**). Indeed, the sequence alignment reveals that Hpr's Walker A P-loop (which is canonical) and these nonstandard Rossmann phosphate binding loops are only few mutations away from each other (**Figure 4A**, and **4F**, overlay in right panel).

An ancestral βαβ seed of both Rossmanns and P-loops

The above findings support the notion of a common Rossmann/P-loop ancestor, the minimal structure of which is βαβ. This ancestral polypeptide includes just two functional motifs: a phosphate binding loop, and an Asp, which could play the dual roles of either binding the ribose moiety of phosphoribonucleosides, or of ligating a dication such as Mg^{2+} or Ca^{2+}. Previously, such a polypeptide (*i.e.*, β-PBL-α-β-Asp) has been proposed as the seed from which Rossmann enzymes emerged (Refs. [7,32] and references therein). In contrast, a βα element was assigned as the P-loop ancestral seed (i.e., a β-P-loop-α segment lacking β2; Ref. 7, Ref. 32 and references therein). Here, we argue that ancestral peptide(s) comprising a βαβ element gave rise to both lineages, and possibly that a single peptide served as a common ancestor of both lineages.

From an ancestral seed to intact domains

The above seed fragment was subsequently expanded by addition of β-loop-α elements. Expansion has also enabled a functional split, or sub-specialization, of the two separate lineages – Rossmann and P-loop, that further evolved and massively diversified. In essence,

this split regards two key elements – phospho-ligand binding and β-strand topology. Our analysis indicates the feasibility of both.

The plausibility of common descent of the P-loop Walker A and the Rossmann phosphate binding loops is indicated by the detected shared theme described above (**Figure 4**). Although the canonical motifs of both lineages differ, there still exists – particularly among Rossmann proteins –alternative motifs that could diverge via a few mutations to a Walker A P-loop. Other Rossmanns possess a GxGGxG motif that also represents a potential jumping board to a Walker A P-loop[25]. The shared themes also indicate that the β2-Asp of the presumed ancestral fragment could not only bind the ribose moiety as in Rossmanns, but also serve as a Walker B, as in P-loops. Tubulin lends further support, indicating that a β2-Asp can indeed play a dual role. Repositioning of the phospho-ligand within a Rossmann protein allowed the β2-Asp to be repurposed for dication binding, as in the Walker B motif.

Expansion of the ancestral βαβ fragment would enable not only the sub-functionalization of the two functional elements described above, but also the fixation of two separate β-strand topologies – the sequential Rossmann topology *versus* the swapped P-loop one. Both folds are in essence a tandem repeat of β-loop-α elements (**Figure 1**). The evolutionary history of other repeat folds tells us that expansion would typically occur by duplication of the ancestral fragment[33–36], or parts of it, but also from fusion of independently emerging fragments[37,38]. Regardless of the origin of the extending fragment(s), given a βαβ ancestor, sequential fusions of α/β elements (or larger elements) could give rise to either one of these two folds. As summarized in **Figure 5**, a newly added β-strand can align at the edge, next to the existing two strands, leading to a Rossmann like topology. Alternatively, a βα element inserted in between the two ancestral strands would result in P-loop topology (*Step 2*; resulting in the ancestral β2 that carries what now becomes the Walker B Asp becoming β3). In next extension, the newly added 4$_{th}$ strand aligns at the other side of β1

(*Step 3*; β4 added with its preceding helix, α3). The subsequent strand(s) could in principle be added sequentially, one next to the other, to yield the intact domains as we know them today.

In support of the above scenario, transitions in the topology of β-sheets that result from strand swaps, or strand invasions, have been documented[37]. In particular, the P-loop lineage seems to have undergone various strand swaps and insertions that gave rise to a variety of topologies[4], including the noncanonical one, with an antiparallel strand, seen in Hpr kinases (**Figure 4C**). Indeed, a survey of P-loop F-groups reveals multiple strand topologies (**Figure S2**; see also Ref. [4]). Specifically, families that catalyze phosphoryl transfer, namely kinases, such as thymidylate kinase (F-group 2004.1.1.166), but also GTPases such as elongation factor Tu (F-group 2004.1.1.258), tend to have the simplest 2-3-1-4-5 topology illustrated in **Figure 1C** (see Ref. 4). In these proteins, the Walker A P-loop and the Walker B-Asp reside on adjacent strands, with the Walker B motif on the tip of β3 (as illustrated in **Figure 2A**). On the other hand, "motor proteins", in which ATPase activity drives a large conformational change that turns into some further action, such as helicases or the ATP cassette of ABC transporters, tend to have a strand inserted between β1 and β3, to yield a 2-3-4-1-5 topology. Here, the Walker B-Asp is also situated on the tip of β3, yet with an intervening strand (β4) between the Walker A and Walker B motifs. The split between these topologies is ancient, likely predating the LUCA[4], and supports the hypothesis that multiple events of fusions and insertions were associated with functional radiation.

In contrast to the P-loops variable topologies, the pseudo-symmetrical Rossmann topology seems highly conserved (in a previous analysis of the Rossmann fold, we did not detect a single structure annotated as Rossmann with swapped strand topology[10]). The conservation of this topology in Rossmann-like, or Rossmannoid folds (Flavodoxin, 2-1-3-4-5, and HUP 3-2-1-4-5) may be due, in part, to the generally higher stability of this strand

topology. This stability may also be indicated by higher success rates in the design of Rossmann-like proteins compared to P-loops-like with swapped strand topology[34]. Nonetheless, in support of our proposed emergence scheme (**Figure 5**) extensions at the edges of the core β-sheet do happen in Rossmanns. This is evident by many Rossmann families having just 5-strands (*versus* the common 6-strand) but also by circular permutations that are common in Rossmanns and by some non-canonical additions of a 7th strand at either end of the β-sheet[37].

Conclusions

Protein evolution spans nearly 4 billion years, with the founding events occurring pre-LUCA. As such, for many protein families, definitive assignment of homologous versus analogous relationships (shared ancestry versus convergent evolution) may never be possible[8]. Confounding matters further, early constraints on protein sequence and structure have further limited the number of possible solutions to a subset of structures and binding motifs[23], making convergence a more likely scenario, particularly in the most ancient proteins. Thus, whether the P-Loops and the Rossmann lineages diverged or converged remains an open question, even several decades after their discovery. The availability of thousands of structures, highly curated databases that catalogue them[13,14], and sensitive search methods[16] and algorithms[39] allows this question to be reexamined. Here, evidence in favor of common ancestry between these lineages is provided, though convergence cannot and should not be entirely ruled out. Whether it was convergence or divergence, our analysis suggests that both lineages emerged from a polypeptide comprising a β-PBL-α-β-Asp fragment. Such a fragment was likely the ancestor of both P-loops and Rossmanns–-be it the same fragment, or two (or more) independently emerged ones.

Methods

The functional diversity of the P-loops and Rossmann lineages

F-groups belonging to Rossmann-like (X-group 2003; 211 F-groups), P-loop-like (X-group 2004; 173 F-groups) and Rossmann-like structures with the crossover (X-group 2111; 385 F-groups) were mapped to 599 corresponding Pfam families. Each Pfam family was then mapped to EC classes using the ECDomainMiner resource[40]. In total 1264 4-digit EC classes were associated with 599 Pfam families. Since many Pfam families represent more than one enzymatic activity, we chose only the most common activity within a given F-group (based on the first EC digit). If two or more EC classes were equally common within a Pfam family, we kept all of them. Within all three X-groups, the majority of families exhibit transferase activity (EC 2.-.-.-). As expected, within Rossmann-like X-group, oxidoreductases (EC 1.-.-.-) seem to be almost equally common, whereas for both P-loops and Rossmann-like structures with the crossover second most common enzyme class is a hydrolase (EC 3.-.-.-). For Rossmann-like X-groups hydrolytic activity is the least common.

Identification of shared themes between P-loops and Rossmanns

Briefly, we used HHSearch[15] to compare a set of previously curated themes[17] to a 70% NR set of ECOD domains (version develop210). Using an E-value threshold of 10^{-3}, and a coverage threshold of 85%, we identified all significant hits between proteins belonging to ECOD X-groups 2003.1 (Rossmann-like) and 2004.1 (P-loop domains-like). Because there is extensive redundancy among the themes, we identified the representative meaningful examples. For each domain pair, we calculated the Smith-Waterman (SW) alignments of the sequence segments before and after the shared theme, and the SW or Needleman-Wunsch (NW) alignment for the regions that encompass the shared theme. Overall, we found 50 pairs, all involving Hpr kinase (PDB: 1ko7) and a Rossmann protein, that gave a statistically

meaningful the SW local alignment, and 7 that were also significantly aligned with the NW global alignment. Further details regarding the theme analysis will be described in a forthcoming paper.

Estimating ligand placement in unliganded structures

The protein structures identified as sharing a theme had no bound ligands. Models of the liganded structures (**Figure 4B**) were generated by overlaying the most closely related structure with bound NTP, or NTP analogue. In the case of HrP kinase (PDB: 1ko7), the conformation of the Walker P-Loop was perfectly canonical. Thus, despite no structure from the same F-group having a relevant ligand, the overall positioning of the ligand is unambiguous. The ligand-containing structure that was used to model the NTP ligand for HrP kinase was PDB 2ixf, an ABC transporter. For sorbitol dehydrogenase (PDB: 1k2w) and short chain dehydrogenase (PDB 3tjr), the structure of the phosphate biding loop is extended and characterized by a noncanonical sequence motif. Yet in this case, a closely related protein from the same F-group was identified and could serve to model the NTP ligand: PDB 2wsb, a short chain dehydrogenase that also bears an extended phosphate binding loop and a noncanonical sequence motif. Despite being unliganded, the rotamers of the β2-Asp residues in both structures (PDBs 1k2w and 3tjr) were well-positioned to interact with the ligands.

Calculating Consensus Sequences and Residue Conservation Scores

The relevant ECOD F-groups (**Figure 4**) were mapped to the corresponding Pfam families. Since 2003.1.1.417 and 2003.1.1.332 are associated with one Pfam family, they were analyzed jointly. Seed alignments were extracted from Pfam, clustered at 70% sequence redundancy using CD-HIT[41], and the consensus sequence and conservation scores were calculated for the shared region (theme) using JalView.

Figure Legends

Figure 1. The 3-layer αβα sandwich. **A**. The αβα sandwich is a modular fold comprised of repeating β-loop-α elements. In this side-view of two tandem βα elements with a Rossmann topology, the functional loops are situated on the "top" of the fold (thick lines) and the β-loop-α element are linked via short, bottom loops (thin, dashed lines). Shown here are the first two elements, beginning with β1 at the N-terminus, and the first two helices (α1, α2) that in this cartoon comprise one external layer of the sandwich. **B**. A view from the top reveals the αβα sandwich architecture with three layers of secondary structure: a parallel β-sheet flanked on both sides by α-helices. The top active site loops face the reader and the N- and C-termini and bottom connecting loop are at the back. The order of the β-strands in the interior β-sheet shown here is the canonical Rossmann topology. **C**. The most common, core P-loop NTPase (P-loops) topology. Noted in red are the differences from the Rossmann topology—migration of β3 from the edge to the center, and of α2 and α5 from one external layer to another.

Figure 2. The Ligand Binding Modes of Rossman and P-Loop Proteins. The phosphate binding loop of both lineages stems from β1 to the N-terminus of α1 (conserved glycine residues are colored magenta). The Rossmann β2-Asp, and the P-loop Walker B Asp, are in green sticks. Water molecules are denoted by red spheres and metal dications by green spheres. **A**. The canonical P-loop NTPase binding mode. The phosphate binding loop (the P-loop Walker A motif; GxxxxGK(S/T)) begins with the first conserved glycine residue at the tip of β1 and ends with T/S within α1. The Walker B's Asp, located at the tip of β3, interacts with the catalytic Mg^{2+}, either directly, or via a water molecule as seen here. **B**. The canonical Rossmann binding mode. The phosphate binding site includes a canonical water molecule (α1 has been rendered transparent so that the conserved water is visible). The Asp sidechain

at the tip of β2 (β2-Asp) forms a bidentate interaction with both hydroxyls of the ribose. Note the different directionality of the ligand in Rossmann and P-loops, reflected by the opposite directions of the ribose and adenine moieties. **C**. Tubulin is a GTPase that possesses the Rossmann's strand topology, phosphate binding loop (including the mediating water), and the canonical β2-Asp. However, the ligand, GTP is bound in the P-loop NTPase mode (as in **A**), and accordingly, the β2-Asp makes a water mediated interaction with the catalytic metal cation (Ca^{2+}or Mg^{2+}) thus acting in effect as a Walker B aspartate. ECOD domains used in this figure, from left to right, are e1yrbA1, e1lssA1, and e5j2tB1.

Figure 3. Alternative phosphate binding sites in αβα sandwich enzymes. HUP proteins are αβα sandwich proteins with Rossmann-like strand topology. The canonical HUP phosphate binding mode is located at the tip of α1 (left panels; ECOD F-group 2005.1.1.13; show is domain e1xngA1) as in Rossmann and P-loop NTPases (**Figure 2**). However, Usp (universal stress proteins) is a HUP family that exhibits ATPase activity wherein phosphate binding migrated to the tip of α4 (right panels; ECOD F-group 2005.1.1.145; shown is domain e2z08A1). As shown in the overlay (middle panels), despite the migration of the phosphate binding site, the ribose and adenine binding modes are conserved.

Figure 4. Theme sharing between Rossmann and P-loop enzymes. A. Sequence alignment of the shared themes. PDB codes are shown on the right, and the ECOD F-group to which they belong are on the left. The identified themes involve a segment of a single P-loop NTPase, Hpr kinase (top line, ECOD domain e1ko7A1), that aligns to a variety of Rossmanns that belong to four F-groups (representatives shown here). **B**. The consensus sequence of each F-group, derived from ECOD with a 70% sequence identity cutoff, shaded according to the degree of conservation. The individual sequences identified by the theme

search show higher similarity by default, yet nonetheless, the family consensus sequences also align well, and the identical residues tend to be highly conserved. **C-D**. Although detection of the shared theme was based on sequence only, structurally, the shared theme encompasses the β1-PBL-α1-β2-Asp element in both the P-loop protein (panel C; Hpr kinase, ECOD domain e1ko7A1) and the theme-related Rossmanns (panel D; ECOD domains e3gedA1, e3tjrA1, e1kvtA1, e3ondA1). The conserved phosphate binding loop glycine residues are colored magenta and the β2-Asp is colored green. Models of liganded structures were created by overlaying the functional loops with closely related homologs (see *Methods*). **E**. An overlay of the β1-PBL-α1-β2-Asp element of the Hpr Kinase (cyan; ECOD domain e1ko7B1) and one of the theme-related Rossmann dehydrogenases (yellow; ECOD domain e3tjrA1). **F**. Structural details of the phosphate binding loops: The Walker A binding loop of Hrp kinase (left panel; ECOD domain e1ko7A1); the phosphate binding loop of sorbitol dehydrogenase (middle panel; ECOD domain e1k2wA1); and an overlay of both loops (right panel).

Figure 5. Divergence of Rossmanns and P-loop NTPases from the presumed ancestral polypeptide. Emergence begins with the presumed β-PBL-α-β-Asp ancestor, that can act as either Rossmann, or a P-loop NTPase, depending on how the phospho-ligands bind and the role taken by the β2-Asp (illustrated in **Figure 2A** and **C**). In the second step, the ancestral fragment is either extended at its C-terminus by fusion of an αβ fragment to generate a Rossmann-like domain (top row); or, a βα fragment is inserted between α1 and β2 to yield a P-loop-like domain (bottom row). Note that insertion results in the ancestral β2 that carries the Walker B Asp becoming β3. Note also that the location of the added helix, α2, differs. It may locate next to α1, namely on the same layer of what will become a Rossmann αβα sandwich, or, on the opposite side, as in P-loop NTPases.

Acknowledgments

This research has been supported by Grant 94747 by the Volkswagen Foundation. N.B.-T.'s research is supported in part by the Abraham E. Kazan Chair in Structural Biology, Tel Aviv University. D.S.T. is the Nella and Leon Benoziyo Professor of Biochemistry. We are grateful to Ita Gruic-Sovulj for her role in the analysis of the HUP domain that led to Figure 3, and to Andrei Lupas for insightful and critical comments.

References

1. Adams, M. J. *et al.* Structure of lactate dehydrogenase at 2.8 Å resolution. *Nature* (1970). doi:10.1038/2271098a0
2. Rossmann, M. G., Moras, D. & Olsen, K. W. Chemical and biological evolution of a nucleotide-binding protein. *Nature* (1974). doi:10.1038/250194a0
3. Walker, J. E., Saraste, M., Runswick, M. J. & Gay, N. J. 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). doi:10.1002/j.1460-2075.1982.tb01276.x
4. Leipe, D. D., Koonin, E. V. & Aravind, L. Evolution and classification of P-loop kinases and related proteins. *J. Mol. Biol.* (2003). doi:10.1016/j.jmb.2003.08.040
5. Ma, B. G. *et al.* Characters of very ancient proteins. *Biochemical and Biophysical Research Communications* (2008). doi:10.1016/j.bbrc.2007.12.014
6. Edwards, H., Abeln, S. & Deane, C. M. Exploring Fold Space Preferences of New-born and Ancient Protein Superfamilies. *PLoS Comput. Biol.* (2013). doi:10.1371/journal.pcbi.1003325
7. Alva, V., Söding, J. & Lupas, A. N. A vocabulary of ancient peptides at the origin of folded proteins. *Elife* **4**, e09410 (2015).

8. Aravind, L., Mazumder, R., Vasudevan, S. & Koonin, E. V. Trends in protein evolution inferred from sequence and structure analysis. *Current Opinion in Structural Biology* (2002). doi:10.1016/S0959-440X(02)00334-2

9. Goncearenco, A. & Berezovsky, I. N. Protein function from its emergence to diversity in contemporary proteins. *Phys. Biol.* (2015). doi:10.1088/1478-3975/12/4/045002

10. Laurino, P. *et al.* An ancient fingerprint indicates the common ancestry of Rossmann fold enzymes utilizing different ribose based cofactors. *PLOS Biol.* (2016). doi:10.1371/journal.pbio.1002396

11. Galperin, M. Y. & Koonin, E. V. Divergence and convergence in enzyme evolution. *Journal of Biological Chemistry* (2012). doi:10.1074/jbc.R111.241976

12. Elias, M. & Tawfik, D. S. Divergence and convergence in enzyme evolution: Parallel evolution of paraoxonases from quorum-quenching lactonases. *Journal of Biological Chemistry* (2012). doi:10.1074/jbc.R111.257329

13. Chandonia, J. M., Fox, N. K. & Brenner, S. E. SCOPe: Manual Curation and Artifact Removal in the Structural Classification of Proteins – extended Database. *J. Mol. Biol.* (2017). doi:10.1016/j.jmb.2016.11.023

14. Cheng, H. *et al.* ECOD: An Evolutionary Classification of Protein Domains. *PLoS Comput. Biol.* (2014). doi:10.1371/journal.pcbi.1003926

15. Hildebrand, A., Remmert, M., Biegert, A. & Söding, J. Fast and accurate automatic structure prediction with HHpred. *Proteins Struct. Funct. Bioinforma.* (2009). doi:10.1002/prot.22499

16. Potter, S. C. *et al.* HMMER web server: 2018 update. *Nucleic Acids Res.* (2018). doi:10.1093/nar/gky448

17. Nepomnyachiy, S., Ben-Tal, N. & Kolodny, R. Complex evolutionary footprints revealed in an analysis of reused protein segments of diverse lengths. *Proc. Natl. Acad.*

Sci. U. S. A. (2017). doi:10.1073/pnas.1707642114

18. Bukhari, S. A. & Caetano-Anollés, G. Origin and Evolution of Protein Fold Designs Inferred from Phylogenomic Analysis of CATH Domain Structures in Proteomes. *PLoS Comput. Biol.* (2013). doi:10.1371/journal.pcbi.1003009

19. Winstanley, H. F., Abeln, S. & Deane, C. M. How old is your fold? *Bioinformatics* (2005). doi:10.1093/bioinformatics/bti1008

20. Bottoms, C. A., Smith, P. E. & Tanner, J. J. A structurally conserved water molecule in Rossmann dinucleotide-binding domains. *Protein Sci.* (2002). doi:10.1110/ps.0213502

21. Shalaeva, D. N., Cherepanov, D. A., Galperin, M. Y., Golovin, A. V. & Mulkidjanian, A. Y. Evolution of cation binding in the active sites of P-loop nucleoside triphosphatases in relation to the basic catalytic mechanism. *Elife* (2018). doi:10.7554/eLife.37373

22. Medvedev, K. E., Kinch, L. N., Schaeffer, R. D. & Grishin, N. V. Functional analysis of Rossmann-like domains reveals convergent evolution of topology and reaction pathways. *PLoS Comput. Biol.* (2019). doi:10.1371/journal.pcbi.1007569

23. Longo LM, Petrović D, Kamerlin SCL, T. D. 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* (2020).

24. Aravind, L., Anantharaman, V. & Koonin, E. V. Monophyly of Class I aminoacyl tRNA synthetase, USPA, ETFP, photolyase, and PP-ATPase nucleotide-binding domains: Implications for protein evolution in the RNA world. *Proteins Struct. Funct. Genet.* (2002). doi:10.1002/prot.10064

25. Zheng, Z., Goncearenco, A. & Berezovsky, I. N. Nucleotide binding database NBDB - A collection of sequence motifs with specific protein-ligand interactions. *Nucleic*

Acids Res. (2016). doi:10.1093/nar/gkv1124

26. Yutin, N. & Koonin, E. V. Archaeal origin of tubulin. *Biol. Direct* (2012). doi:10.1186/1745-6150-7-10

27. Margolin, W., Wang, R. & Kumar, M. Isolation of an ftsZ homolog from the archaebacterium Halobacterium salinarium: Implications for the evolution of FtsZ and tubulin. *J. Bacteriol.* (1996). doi:10.1128/jb.178.5.1320-1327.1996

28. Nogales, E., Downing, K. H., Amos, L. A. & Löwe, J. Tubulin and FtsZ form a distinct family of GTPases. *Nat. Struct. Biol.* (1998). doi:10.1038/nsb0698-451

29. Nogales, E., Wolf, S. G. & Downing, K. H. Structure of the αβ tubulin dimer by electron crystallography. *Nature* (1998). doi:10.1038/34465

30. Farr, G. W. & Sternlicht, H. Site-directed mutagenesis of the GTP-binding domain of β-tubulin. *J. Mol. Biol.* (1992). doi:10.1016/0022-2836(92)90700-T

31. Márquez, J. A. *et al.* Structure of the full-length HPr kinase/phosphatase from Staphylococcus xylosus at 1.95 Å resolution: Mimicking the product/substrate of the phospho transfer reactions. *Proc. Natl. Acad. Sci. U. S. A.* (2002). doi:10.1073/pnas.052461499

32. Laurino, P. *et al.* An Ancient Fingerprint Indicates the Common Ancestry of Rossmann-Fold Enzymes Utilizing Different Ribose-Based Cofactors. *PLoS Biol.* (2016). doi:10.1371/journal.pbio.1002396

33. Eck, R. V. & Dayhoff, M. O. Evolution of the structure of ferredoxin based on living relics of primitive amino acid sequences. *Science (80-.).* (1966). doi:10.1126/science.152.3720.363

34. Romero Romero, M. L. *et al.* Simple yet functional phosphate-loop proteins. *Proc. Natl. Acad. Sci.* (2018). doi:10.1073/pnas.1812400115

35. Zhu, H. *et al.* Origin of a folded repeat protein from an intrinsically disordered

ancestor. *Elife* (2016). doi:10.7554/eLife.16761

36. Longo, L. M. *et al.* Primordial emergence of a nucleic acid binding protein via phase separation and statistical ornithine to arginine conversion. *bioRxiv* (2020). doi:10.1101/2020.01.18.911073

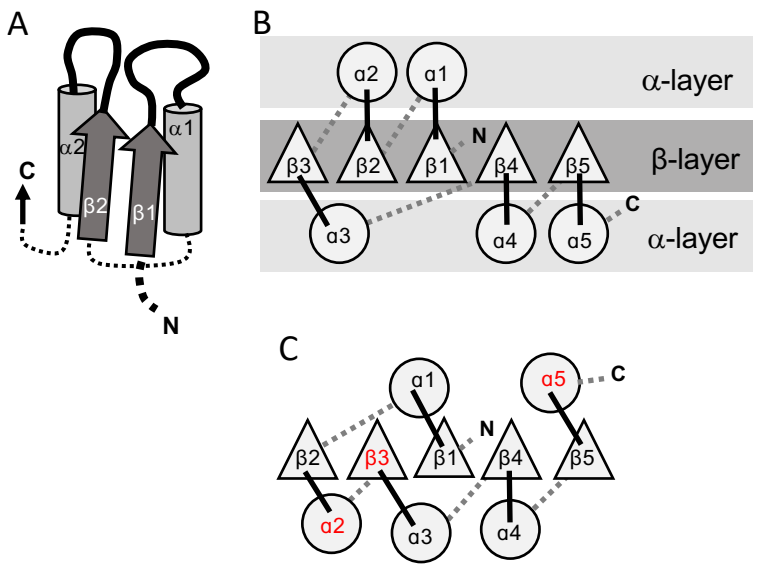
37. Grishin, N. V. Fold change in evolution of protein structures. *J. Struct. Biol.* (2001). doi:10.1006/jsbi.2001.4335

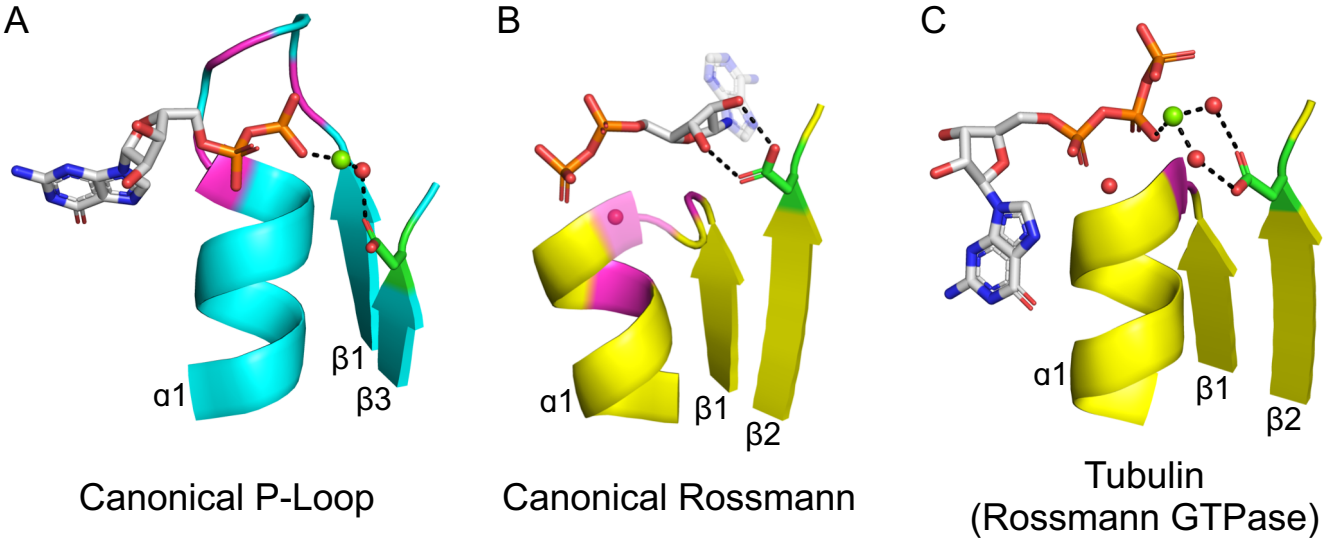
38. Setiyaputra, S., MacKay, J. P. & Patrick, W. M. The structure of a truncated phosphoribosylanthranilate isomerase suggests a unified model for evolution of the (βα)8 barrel fold. *J. Mol. Biol.* (2011). doi:10.1016/j.jmb.2011.02.048

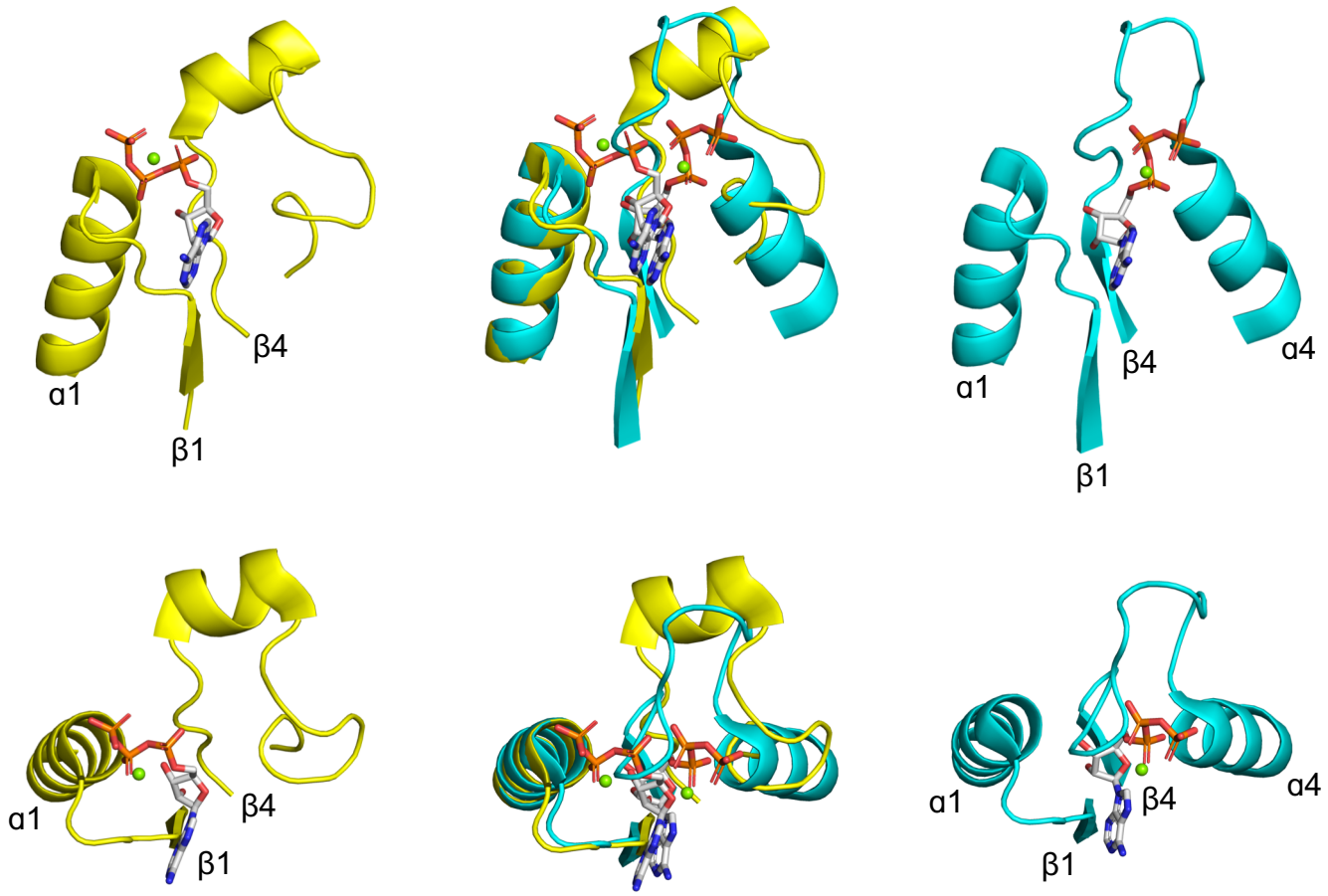
39. Nepomnyachiy, S., Ben-Tal, N. & Kolodny, R. Global view of the protein universe. *Proc. Natl. Acad. Sci.* (2014). doi:10.1073/pnas.1403395111

40. Alborzi, S. Z., Devignes, M. D. & Ritchie, D. W. ECDomainMiner: Discovering hidden associations between enzyme commission numbers and Pfam domains. *BMC Bioinformatics* (2017). doi:10.1186/s12859-017-1519-x

41. Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: Accelerated for clustering the next-generation sequencing data. *Bioinformatics* (2012). doi:10.1093/bioinformatics/bts565


A
α2
α1
β2
β1
C
N
B
α2
α1
α-layer
N
β3
β2
β1
β4
β5
β-layer
C
α3
α4
α5
α-layer
C
α1
α5
C
N
β2
β3
β1
β4
β5
α2
α3
α4

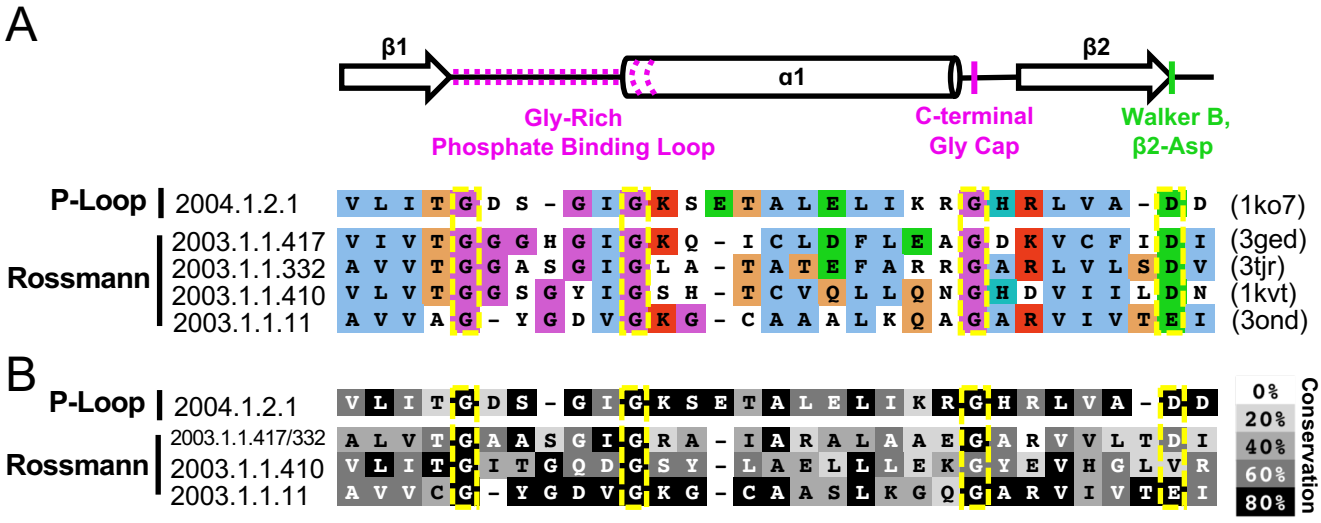

A
α1
β1
β3
Canonical P-Loop
B
α1
β1
β2
Canonical Rossmann
C
α1
β1
β2
Tubulin
(Rossmann GTPase)

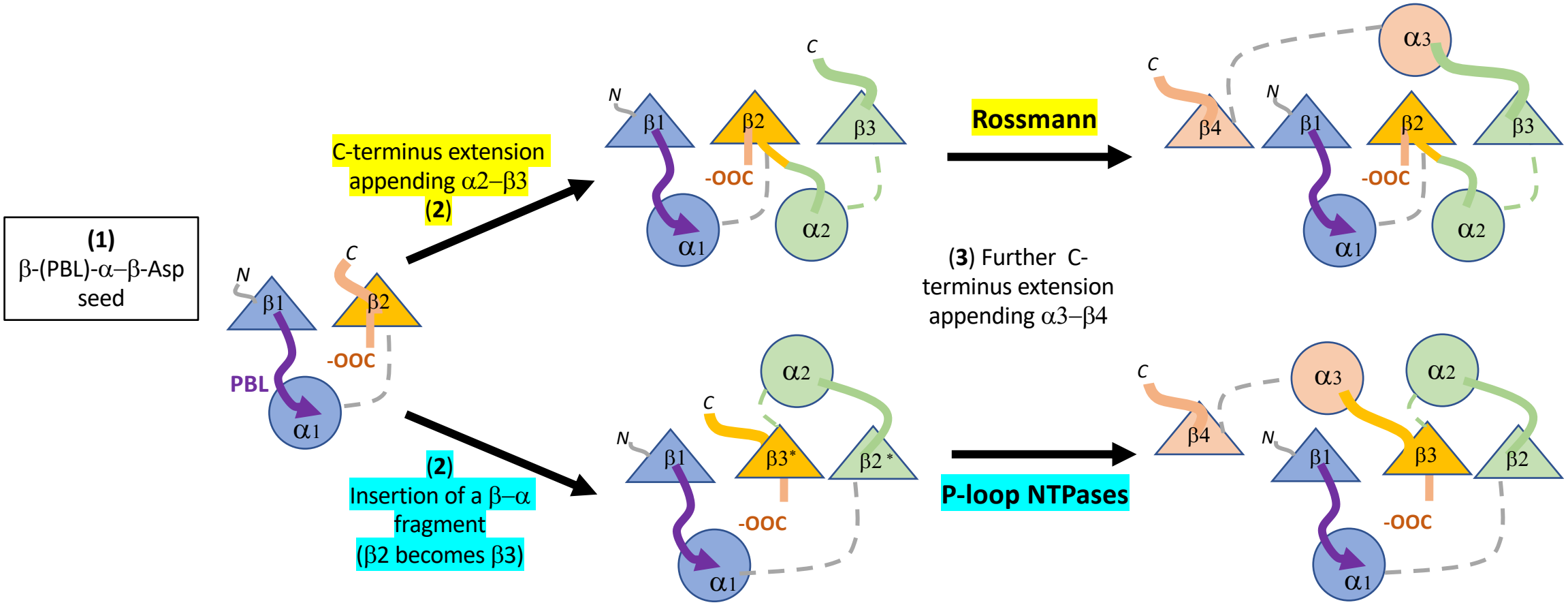


NAD+ synthase,
Canonical α1
Binding Mode

Overlay

Usp,
Non-canonical α4
Binding Mode


A
β1
α1
β2
Gly-Rich
Phosphate Binding Loop
C-terminal
Gly Cap
Walker B,
β2-Asp
P-Loop
2004.1.2.1 V L I T G D S - G I G K S E T A L E L I K R G H R L V A - D D (1ko7)
Rossmann
2003.1.1.417 V I V T G G G H G I G K Q - I C L D F L E A G D K V C F I D I (3ged)
2003.1.1.332 A V V T G G A S G I G L A - T A T E F A R R G A R L V L S D V (3tjr)
2003.1.1.410 V L V T G G S G Y I G S H - T C V Q L L Q N G H D V I I L D N (1kvt)
2003.1.1.11 A V V A G - Y G D V G K G - C A A A L K Q A G A R V I V T E I (3ond)
B
P-Loop
2004.1.2.1 V L I T G D S - G I G K S E T A L E L I K R G H R L V A - D D
Rossmann
2003.1.1.417/332 A L V T G A A S G I G R A - I A R A L A A E G A R V V L T D I
2003.1.1.410 V L I T G I T G Q D G S Y - L A E L L L E K G Y E V H G L V R
2003.1.1.11 A V V C G - Y G D V G K G - C A A S L K G Q G A R V I V T E I
0%
20%
40%
60%
80%
Conservation
C
Walker
B
α1
β2
β-1
β1
Gly Cap
P-Loop Theme
2004.1.2.1
D
β2-Asp
β1
β2
α1
Gly Cap
Overlay of
Rossmann Themes
E
α1
Gly Cap
Conservation of the
α1-β2 linker
F
α1
β1
P-Loop
Phosphate Binding Loop
2004.1.2.1
α1
β1
Rossmann
Phosphate Binding Loop
2003.1.1.417
α1
β1
Overlay of
Phosphate Binding Loops


(1)
β-(PBL)-α–β-Asp
seed
N
C
β1
β2
PBL
-OOC
α1
C-terminus extension
appending α2–β3
(2)
β3
α2
Rossmann
α3
β4
(3) Further C-
terminus extension
appending α3–β4
(2)
Insertion of a β–α
fragment
(β2 becomes β3)
β3*
β2 *
P-loop NTPases