Received: 13 September 2022 | Revised: 1 December 2022 | Accepted: 3 December 2022

DOI: 10.1002/pro.4536

FULL-LENGTH PAPER

THE PROTEIN SOCIETY WILEY

Experimental and computational analysis of the ancestry of an evolutionary young enzyme from histidine biosynthesis

Thomas Kinateder | **Lukas Drexler** | **Kristina Straub** | **Rainer Merkl** | **Reinhard Sterner**

Institute of Biophysics and Physical Biochemistry and Regensburg Center for Biochemistry, University of Regensburg, Regensburg, Germany

Correspondence
Reinhard Sterner, Institute of Biophysics and Physical Biochemistry and Regensburg Center for Biochemistry, University of Regensburg, Regensburg, Germany.
Email: reinhard.sterner@ur.de

Funding information
Deutsche Forschungsgemeinschaft, Grant/Award Numbers: ME2259/4-1, STE891/13-1

Review Editor: John Kuriyan

Abstract

The conservation of fold and chemistry of the enzymes associated with histidine biosynthesis suggests that this pathway evolved prior to the diversification of Bacteria, Archaea, and Eukaryotes. The only exception is the histidinol phosphate phosphatase (HolPase). So far, non-homologous HolPases that possess distinct folds and belong to three different protein superfamilies have been identified in various phylogenetic clades. However, their evolution has remained unknown to date. Here, we analyzed the evolutionary history of the HolPase from γ-Proteobacteria (HisB-N). It has been argued that HisB-N and its closest homologue D-*glycero*-D-*manno*-heptose-1,7-bisphosphate 7-phosphatase (GmhB) have emerged from the same promiscuous ancestral phosphatase. GmhB variants catalyze the hydrolysis of the anomeric D-*glycero*-D-*manno*-heptose-1,7-bisphosphate (αHBP or βHBP) with a strong preference for one anomer (αGmhB or βGmhB). We found that HisB-N from *Escherichia coli* shows promiscuous activity for βHBP but not αHBP, while βGmhB from *Crassaminicella sp.* shows promiscuous activity for HolP. Accordingly, a combined phylogenetic tree of αGmhBs, βGmhBs, and HisB-N sequences revealed that HisB-Ns form a compact subcluster derived from βGmhBs. Ancestral sequence reconstruction and in vitro analysis revealed a promiscuous HolPase activity in the resurrected enzymes prior to functional divergence of the successors. The following increase in catalytic efficiency of the HolP turnover is reflected in the shape and electrostatics of the active site predicted by AlphaFold. An analysis of the phylogenetic tree led to a revised evolutionary model that proposes the horizontal gene transfer of a promiscuous βGmhB from δ- to γ-Proteobacteria where it evolved to the modern HisB-N.

Thomas Kinateder and Lukas Drexler contributed equally to this study.

KEYWORDS

AlphaFold, ancestral sequence reconstruction, D-*glycero*-D-*manno*-heptose-1,7-bisphosphate 7-phosphatase, enzyme, enzyme evolution, GmhB, HisB-N, histidinol phosphate phosphatase, horizontal gene transfer, promiscuity

1 | INTRODUCTION

Histidinol phosphate phosphatases (HolPases) catalyze the penultimate step within the biosynthesis of the amino acid histidine (Alifano et al., 1996; Ames, 1957). This biosynthetic pathway can be found in microorganisms and plants and consists of 10 consecutive enzymatic steps (Figure S1) (Alifano et al., 1996; Winkler & Ramos-Montañez, 2009). The chemical reactions and their order are conserved among all organisms which synthesize histidine and it has hence been argued that the pathway was most likely assembled prior to the existence of the last universal common ancestor (Fani et al., 1995; Fani et al., 1998). The route itself has been extensively studied in *Escherichia coli*, where eight different enzymes, including two fusion enzymes, are involved (Figure S1) and where the corresponding genes are organized in a tightly regulated operon (Carlomagno et al., 1988; Winkler & Ramos-Montañez, 2009). Most of the genes and their genomic organization are conserved between phylogenetically diverse species and the corresponding enzymes possess homologous folds (Del Duca et al., 2020; Fani et al., 2005; Winkler & Ramos-Montañez, 2009). However, there is one exception, namely the HolPase (Brilli & Fani, 2004; Kulis-Horn et al., 2017). In *E. coli* and related γ-Proteobacteria, the HolPase function is catalyzed by the N-terminal part of the bi-functional HisB enzyme and was hence termed HisB-N, while the C-terminal part harbors the imidazole glycerol phosphate dehydratase (IGPDH) (Brady & Houston, 1973; Brilli & Fani, 2004; Chiariotti et al., 1986; Chumley & Roth, 1981; Fani et al., 2005). In other organisms, these two enzymes are encoded by two independent genes (Brilli & Fani, 2004). Interestingly, the fold of the IGPDH domain is conserved and is also found in monofunctional enzymes, whereas the fold of the HolPase domain is not conserved. Specifically, HisB-N from *E. coli* (*ec*HisB-N) exhibits a Rossmannoid fold and belongs to the haloacid dehalogenase protein superfamily (HAD) (Rangarajan et al., 2006). In contrast, the monofunctional HolPase from *Lactococcus lactis* exhibits a $(\beta\alpha)_7$-barrel fold and belongs to the polymerase and histidinol phosphatase (PHP) superfamily (Ghodge et al., 2013), and the monofunctional HolPases from *Mycobacterium tuberculosis* and *Medicago truncatula* show the fold of a αβαβα-sandwich and belong to the inositol monophosphatase (IMP) superfamily (Figure S2) (Jha et al., 2018; Ruszkowski & Dauter, 2016). The differences between HolPases also extend to the genomic organization of their respective genes: The gene that encodes *ec*HisB-N is part of the *his* operon, whereas genes of HolPases from the PHP and IMP superfamily are typically located outside of the *his* operon (Brilli & Fani, 2004; Fani et al., 2005; Ghodge et al., 2013; Kulis-Horn et al., 2017; Mormann et al., 2006; Ruszkowski & Dauter, 2016). Taken together, these observations indicate that HisB-N together with other HolPases are evolutionary young compared with the rest of the histidine biosynthetic enzymes.

These observations prompted theoretical studies that aimed to rationalize the evolution of HisB-N in γ-Proteobacteria (Brilli & Fani, 2004; Fani et al., 2005). Relying on multiple sequence alignments, homology, and genomic organization, it has been argued that an ancestor of HisB-N has been recruited into the histidine biosynthetic pathway after the separation of different classes of γ-Proteobacteria. It was further hypothesized that HisB-N and its closest homologue D-*glycero*-D-*manno*-heptose-1,-7-bisphosphate 7-phosphatase (GmhB) were derived from the same ancestral precursor and that a gene duplication event within an ancient γ-Proteobacterium followed by divergent evolution led to the modern HisB-N and GmhB.

Indeed, both HisB-N and GmhB catalyze dephosphorylation reactions. The native substrate of HisB-N is histidinol phosphate (HolP) while GmhB enzymes catalyze the preferential dephosphorylation of one out of two anomeric sugars, namely D-*glycero*-D-*manno*-heptose-1α,-7-bisphosphate (αHBP) or D-*glycero*-D-*manno*-heptose-1β,-7-bisphosphate (βHBP) (Kneidinger et al., 2001; Kneidinger et al., 2002; Valvano et al., 2002). While αHBP is an intermediate in the S-layer biosynthesis, which is often found in gram-positive bacteria, βHBP is an intermediate in the lipopolysaccharide biosynthesis, which is found in gram-negative bacteria. The preference of GmhB enzymes for one anomer over the other typically ranges from 6:1 to 1:150 (α:β) (Valvano et al., 2002; Wang et al., 2010). Here, we refer to enzymes that form part of the S-layer biosynthesis as αGmhB and to enzymes that form part of the lipopolysaccharide biosynthesis as βGmhB.

Intrigued by the above observations, we decided to investigate the hypothesis of a shared evolution between HisB-N and GmhB. Specifically, we intended to test the current model of the HisB-N evolution with a combination of in vivo, in vitro, and in silico approaches. First, we analyzed the extant HisB-N and GmhB enzymes from

E. coli. To this end, we examined the structures of both enzymes and explored the substrate spectrum to check for promiscuous side activities that would indicate common ancestry. To determine the type and degree of evolutionary relationship, we calculated a phylogenetic tree based on sequences representing variants of both enzymes. Based on this tree, we additionally selected GmhB from *Crassaminicella sp*. for functional in vitro and in vivo characterization. The tree was furthermore used to reconstruct several ancestral enzymes that marked the putative functional transition between GmhB and HisB-N. The reconstructed enzymes were functionally characterized in vitro and the observed changes in catalytic efficiencies were rationalized by an analysis of the geometry and electrostatics of the active sites as predicted by the program AlphaFold (Jumper et al., 2021). The experimental data finally allowed for a detailed phylogenetic analysis, which led to a revised model for the evolution of HisB-N and GmhB in γ-Proteobacteria.

2 | RESULTS AND DISCUSSION

2.1 | Sequential and structural comparison of HisB-N and GmhB

It has been postulated that HisB-N and GmhB have evolved from the same ancestral phosphatase that underwent a gene duplication event and subsequent specialization (Brilli & Fani, 2004; Fani et al., 2005). This hypothesis is supported by the fact that both enzymes belong to the HAD superfamily (Wang et al., 2010). Moreover, HisB-N and GmhB from the same organism usually show sequence identities of 26%–31% (Brilli & Fani, 2004). This value is significantly higher than the level of sequence identity that is normally observed for HAD enzymes with different functions (Burroughs et al., 2006) and indicates structural homology (Sander & Schneider, 1991). Thus, the structures of the respective *E. coli* enzymes (*ec*HisB-N and *ec*GmhB) (Nguyen et al., 2010; Rangarajan et al., 2006) exhibit the same Rossmannoid fold that is characterized by an αβα-sandwich being typical for the HAD superfamily (Figure 1a,b).

Moreover, both enzymes exhibit a zinc-binding cap structure consisting of a CxH motif and a CxC motif, which are separated by 5 and 12 residues in *ec*HisB-N and *ec*GmhB, respectively. This zinc-binding cap structure comprises a unique structural feature within the HAD superfamily (Taylor et al., 2010). The active sites of both enzymes are confined by a lid-like loop structure, which is involved in substrate binding. A superposition of the two proteins showed that the structures are highly similar and the RMSD was calculated to be 1.2 Å (Figure 1c). Besides the overall fold, the HAD superfamily is characterized by four catalytically relevant sequence elements, which are highly conserved, namely (i) a DxD motif, (ii) a DxxxD or DD motif, (iii) a conserved lysine, and (iv) a conserved threonine or serine residue. In *ec*GmhB and *ec*HisB-N, all motifs are identical and the respective side chains have the same geometrical orientation in the structures (Figure 1d). Moreover, both HisB-N and GmhB share the same reaction mechanism, which involves substrate binding with coordinate bond formation to a magnesium ion, nucleophilic attack by an aspartate residue on the phosphorus atom, formation of a covalent phospho-aspartate intermediate, and subsequent hydrolysis to yield free phosphate and restore aspartate (Figure 1e) (Burroughs et al., 2006; Rangarajan et al., 2006; Taylor et al., 2010). Taken together, the sequential, structural, and mechanistic similarities indicate a close relationship between HisB-N and GmhB within the superfamily supporting the hypothesis of a common evolution. Depending on the phylogenetic distance and degree of divergence, this could translate to a shared substrate spectrum. To clarify this issue, we tested *ec*HisB-N and *ec*GmhB for promiscuous side activities.

2.2 | Purification and functional characterization of *ec*HisB-N and *ec*GmhB

The genes for *ec*HisB-N and *ec*GmhB were overexpressed in *E. coli* and the proteins were purified by affinity chromatography via N-terminal His_6-tags, followed by size exclusion chromatography (sequences of the constructs are given in Table S1). The purity and structural integrity of both enzymes was assessed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) (Figure S3) and far-UV CD-spectroscopy (Figure S4).

To ensure functional integrity, we first determined the steady-state enzyme catalytic parameters for *ec*HisB-N and its natural substrate HolP. To this end, we used a coupled photometric assay that allowed for continuous measurement of the formation of free phosphate (Suárez et al., 2012). A hyperbolic substrate saturation curve was obtained, which yielded a turnover number (k_{cat}) of 2.8 s^{-1}, a Michaelis constant (K_M) of 48 μM, and a catalytic efficiency (k_{cat}/K_M) of $58.0 \cdot 10^3$ s^{-1} M^{-1} (Figure 2a). Both the Michaelis constant and the turnover number are in good accordance with previously reported values, which are 54 μM for the K_M of *ec*HisB-N and 1–4 s^{-1} for k_{cat} of different monofunctional HolPases (Jha et al., 2018; Nourbakhsh et al., 2014; Ruszkowski & Dauter, 2016).

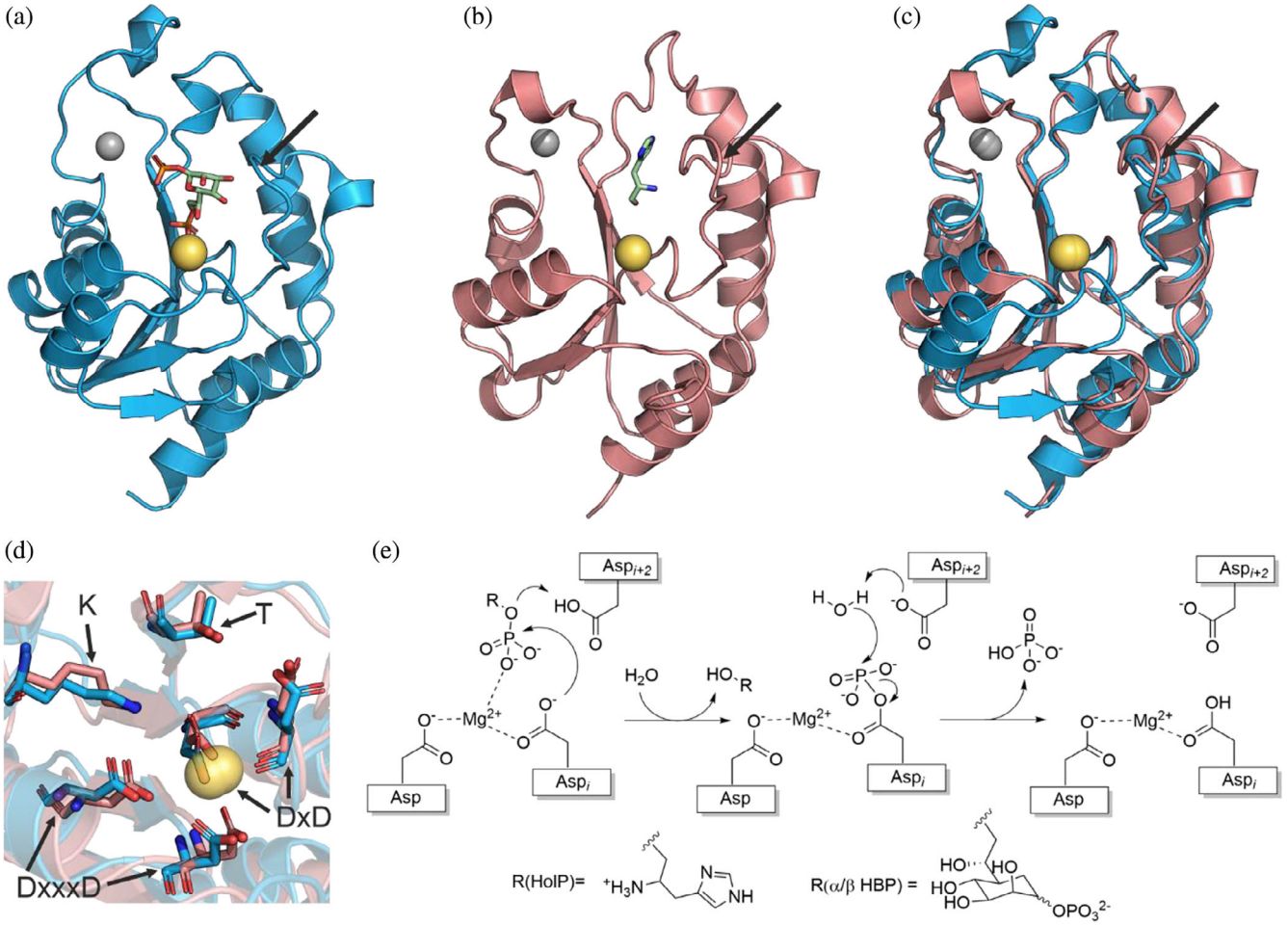

FIGURE 1 (a) Crystal structure of *ec*GmhB with bound substrate D-*glycero*-D-*manno*-heptose-1β,7-bisphosphate (βHBP) (PDB ID: 3L8G) and (b) crystal structure of *ec*HisB-N with bound product L-histidinol (PDB ID: 2FPU) (Nguyen et al., 2010; Rangarajan et al., 2006). Substrates and products are depicted in stick representation, while yellow spheres represent a catalytic magnesium ion and gray spheres represent a zinc ion that is coordinated by a CxH and a CxC motif within a small cap structure, which is unique within the protein superfamily. Black arrows indicate a lid-like loop structure that is involved in substrate binding. (c) Overlay of the unliganded *ec*GmhB (blue) and unliganded *ec*HisB-N (red), yielding an RMSD of 1.2 Å for 91 superimposed C_α-atoms. (d) Zoomed in view of the active sites of *ec*GmhB and *ec*HisB-N with the catalytically relevant residues and residue motifs depicted as sticks. (e) Reaction mechanisms of HisB-N and GmhB for the dephosphorylation of histidinol phosphate (HolP) and of the two anomers of D-*glycero*-D-*manno*-heptose-1,7-bisphosphate (α/β HBP) by the catalytic aspartates from the DxD motif (Asp$_i$, Asp$_{i+2}$) and one aspartate (Asp) from the DxxxD motif (Jumper et al., 2021; Rangarajan et al., 2006; Taylor et al., 2010).

We then checked whether *ec*HisB-N also catalyzes the turnover of any of the two anomeric substrates of GmhB. To this end, we synthesized both αHBP and βHBP according to a revised version of a published protocol (Wang et al., 2010). Then, *ec*HisB-N was incubated with either αHBP or βHBP for up to 20 h, followed by quantification of newly formed free phosphate using the aforementioned photometric assay. As control experiments, αHBP or βHBP was incubated with *ec*GmhB or buffer. The amount of detected phosphate in the buffer control either caused by impurities or by spontaneous hydrolysis was subtracted from all other measurements. Interestingly, a low promiscuous activity of *ec*HisB-N could be measured for βHBP (Figure 2b), resulting in the hydrolysis of approximately 18% of the substrate after 20 h. However, no turnover of the anomeric αHBP by *ec*HisB-N could be detected (Figure 2c). This finding supports the hypothesis of a shared evolution between HisB-N enzymes and GmhB enzymes and suggests that HisB-N is more closely related to βGmhBs than to αGmhBs.

In the next step, we determined the steady-state enzyme catalytic parameters of *ec*GmhB for its native substrate βHBP and the anomeric αHBP. Hyperbolic substrate saturation curves were obtained in both cases. For βHBP, a k_{cat} of 36 s^{-1} and a K_M of 2.3 μM were determined, yielding a k_{cat}/K_M of 15.5 · 10^6 s^{-1} M^{-1}

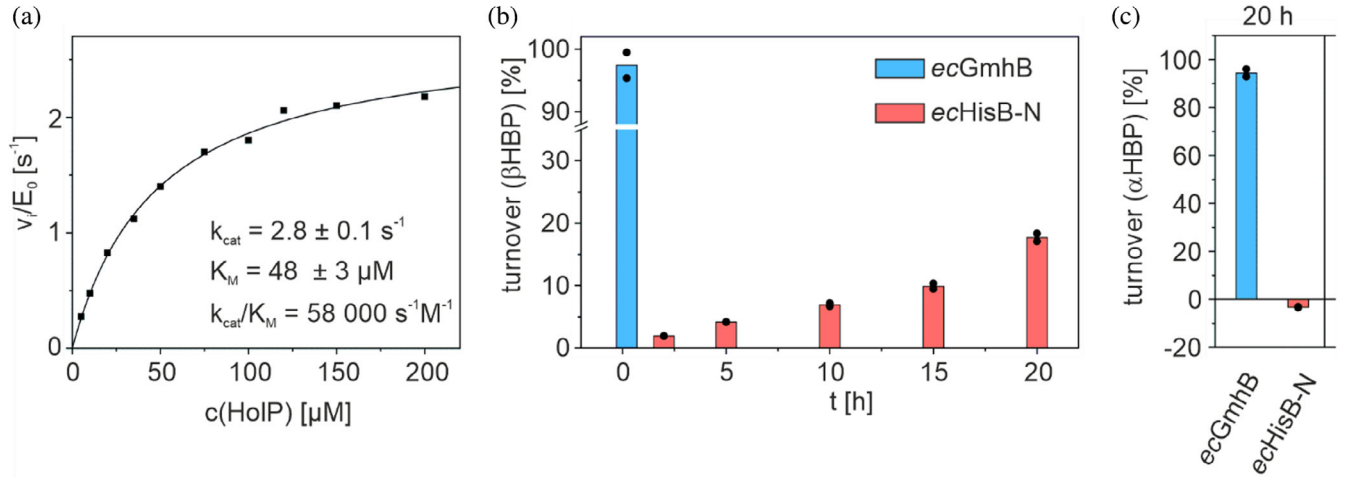

FIGURE 2 Functional characterization of *ec*HisB-N at 25°C. (a) Substrate saturation curve for the turnover of HolP. (b) Average percentage of product formation for the dephosphorylation of D-*glycero*-D-*manno*-heptose-1β,7-bisphosphate (βHBP) as detected by two independent measurements (black dots). Shown is the dephosphorylation of βHBP (30 µM) by *ec*HisB-N (10 µM) or *ec*GmhB (10 nM). For *ec*HisB-N, a promiscuous turnover of 18% substrate was observed within 20 h, whereas for *ec*GmhB, 97% turnover of its preferred anomer was observed within 12 min. (c) Average percentage of product formation for the dephosphorylation of D-*glycero*-D-*manno*-heptose-1α,7-bisphosphate (αHBP) as detected by two independent measurements (black dots). Shown is the turnover of αHBP (30 µM) by *ec*HisB-N (10 µM) or *ec*GmhB (10 µM). For *ec*HisB-N, no promiscuous turnover of αHBP could be detected, whereas for *ec*GmhB, 94% turnover of its nonpreferred anomer was observed within 20 h.

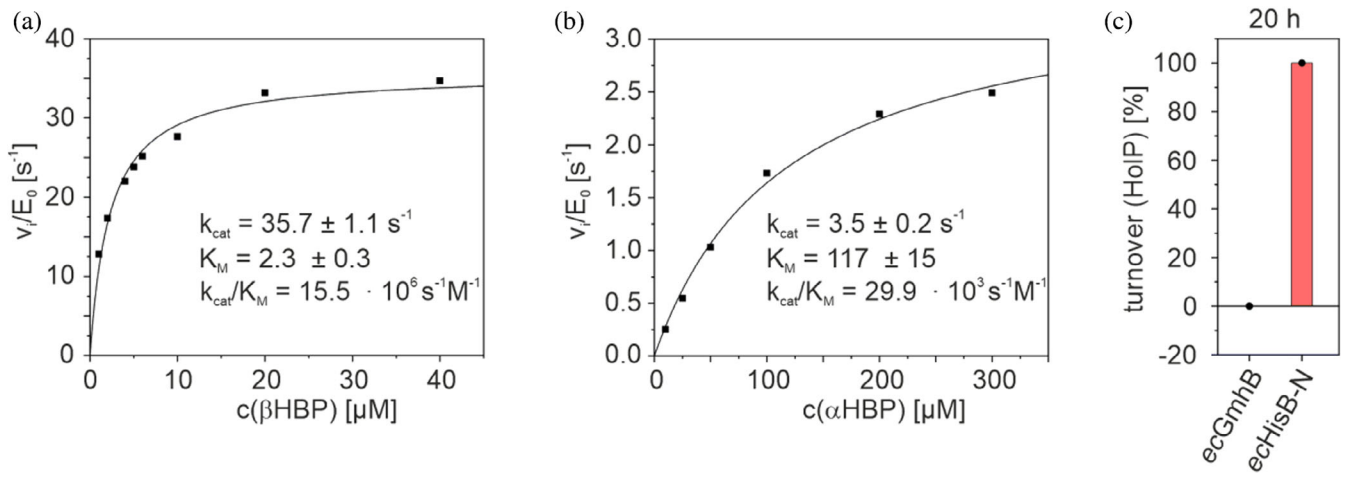

FIGURE 3 Functional characterization of *ec*GmhB at 25°C. (a) Substrate saturation curve for the turnover of D-*glycero*-D-*manno*-heptose-1β,7-bisphosphate (βHBP). (b) Substrate saturation curve for the turnover of D-*glycero*-D-*manno*-heptose-1α,7-bisphosphate (αHBP). (c) Average amount of product formation for the dephosphorylation of histdinol phosphate (HolP) (200 µM) as detected by two independent measurements (black dots) that were performed with *ec*GmhB (10 µM) or *ec*HisB-N (10 nM). For *ec*GmhB, no promiscuous turnover of HolP could be detected, whereas for *ec*HisB-N, 100% turnover was observed within 20 h.

(Figure 3a). For αHBP, a k_{cat} of 3.5 s^{-1} and a K_M of 117 µM were determined, yielding a k_{cat}/K_M of 29.9 · 10^3 s^{-1} M^{-1} (Figure 3b). These values are in good accordance with previously reported results (Wang et al., 2010).

In contrast, no product formation was detectable following the incubation of HolP with *ec*GmhB within 20 h (Figure 3c). In accordance with these findings, an *E. coli* knockout strain lacking the gene for the HolPase could not be rescued by a plasmid-encoded *ec*gmhB gene (Figure S5). The accepted substrates and all experimentally accessible steady-state kinetic parameters of *ec*HisB-N and *ec*GmhB are summarized in Table 1.

From an enzymological point of view, these findings were surprising; first, because βHBP is significantly larger and sterically more demanding than HolP, therefore

Enzyme	Substrate	k_{cat} (s^{-1})	K_M (μM)	k_{cat}/K_M (s^{-1} M^{-1})
*ec*HisB-N	HolP[a]	2.8 ± 0.1	48 ± 3	58,000
	βHBP[b]	18% turnover (10 μM enzyme, 30 μM substrate, 20 h)		
	αHBP[b]	No detectable turnover (10 μM enzyme, 30 μM substrate, 20 h)		
*ec*GmhB	HolP[b]	No detectable turnover (10 μM enzyme, 200 μM substrate, 20 h)		
	βHBP[a]	35.7 ± 1.1	2.3 ± 0.3	15,500,000
	αHBP[a]	3.5 ± 0.2	117 ± 15	29,900
*cs*GmhB	HolP[b]	27% turnover (10 μM enzyme, 200 μM substrate, 20 h)		
	βHBP[a]	0.5 ± 0.03	2.4 ± 0.5	210,000
	αHBP[a]	0.06 ± 0.001	10.2 ± 0.6	5880

TABLE 1 Activity toward HolP, βHBP, and αHBP of *ec*HisB-N, *ec*GmhB, and *cs*GmhB at 25°C

Abbreviations: HolP, histdinol phosphate; αHBP, D-*glycero*-D-*manno*-heptose-1α,7-bisphosphate; βHBP, D-*glycero*-D-*manno*-heptose-1β,7-bisphosphate.
[a]Data from steady-state kinetic experiments.
[b]Data from discontinuous activity assays.

making it unlikely that its turnover by *ec*HisB-N was a fortuitous event. Even more so, as it was shown before that *ec*HisB-N exhibits an unusually high substrate specificity within the HAD superfamily (Huang et al., 2015; Kuznetsova et al., 2006). On the other hand, HolP is much smaller than βHBP and should therefore fit into the binding pocket of *ec*GmhB. Hence, assuming a common promiscuous ancestor, it was unexpected that *ec*GmhB would lose any promiscuous side activity while it would be preserved in *ec*HisB-N. To help with the interpretation of these in vitro data, we decided to perform a phylogenetic analysis of HisB-N and GmhB.

2.3 | Phylogenetic analysis

The promiscuous side activity of *ec*HisB-N for βHBP supports the hypothesis of a close evolutionary relationship between HisB-N and GmhB, whereas the lack of any activity of *ec*GmhB toward HolP suggests a distant relationship. To determine the degree of evolutionary relatedness and to elucidate the nature of the putative common ancestor, a comprehensive phylogenetic tree including HisB-N, αGmhB, and βGmhB sequences was deduced.

In a first step, we retrieved a comprehensive dataset comprising HisB-N, αGmhB, and βGmhB sequences from the KEGG database (Kanehisa & Goto, 2000). The annotation regarding substrate preference was incomplete, which renders it difficult to discriminate between αGmhB and βGmhB. To solve this issue, we checked for the occurrence of either the lipopolysaccharide biosynthesis or S-layer biosynthesis in the corresponding host organism, which then allowed us to classify an enzyme as αGmhB or βGmhB. It is noteworthy that GmhB sequences were found for a wide variety of bacteria, among others in Acidobacteriales, α-, β-, γ-, δ-, and ε-Proteobacteria, Bacilli, Bacteroidetes, Clostridia, Corynebacteriales, Micrococcales, Negativicutes, Sphingobacteria, Streptomycetales, Synechococcales, and Thermodesulfobacteriales, whereas HisB-N sequences were only found in γ-Proteobacteria, ε-Proteobacteria, and Bacteroidetes. The broader phylogenetic distribution of GmhB sequences suggests that this function was established earlier than the HisB-N function. This is in line with the observation that GmhB enzymes are found in gram-positive and gram-negative bacteria, which already suggests that their precursor enzyme originated prior to the separation of these two bacterial groups.

The initially retrieved sequences were filtered to reduce the overrepresentation of preferentially sequenced phyla and to minimize other biases. In this step of the sequence selection process, we generated several diverse sets of sequences, which all yielded the same overall topology. To base the calculation on a broad phylogenetic diversity, we decided to choose the sequence set listed in Supplementary File S2 MSA_extant_sequences.fa. This file contains an MSA of the sequences, which was used to calculate a phylogenetic tree (Figure 4a, Figure S6). Note that we utilized a Bayesian approach and deduced a consensus tree. The Supplementary File S4 Tree_MrBayes.tre contains the full specification of this tree including posterior probabilities for each node.

The calculations gave rise to a topology that clearly separated the three enzyme functions in three distinct subtrees, but some ancestral nodes were only moderately supported. For the subsequent analysis, it is important that only the splits between the three enzyme functions are relevant and these splits are supported by high posterior probabilities. One particular branchpoint, which is supported by a posterior probability of 0.97, discriminates between an isolated cluster of αGmhB enzymes and

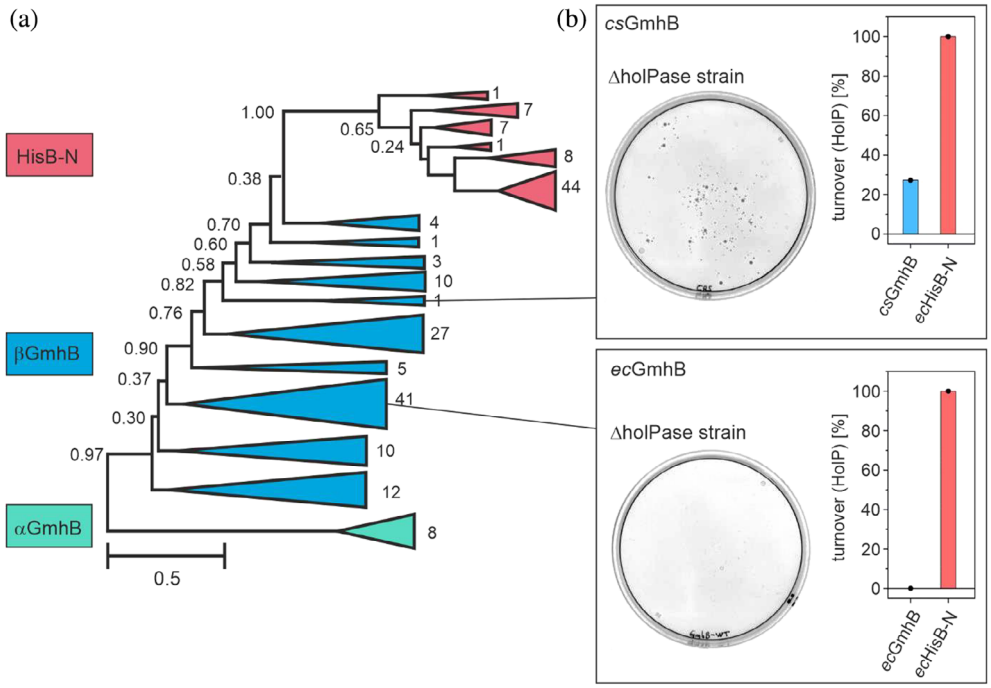

FIGURE 4 Phylogenetic analysis of αGmhB, βGmhB, and HisB-N enzymes and HolPase activities of two βGmhB representatives. (a) Condensed phylogenetic tree deduced from a representative set of αGmhB, βGmhB, and HisB-N sequences using a consensus approach based on Bayesian phylogenetics. The putatively first functional divergence led to a separation of extant αGmhB (cyan) and βGmhB (blue) enzymes, while the second divergence gave rise to the extant HisB-N enzymes (red) in a sub-branch of the βGmhB cluster. Values to the left of each branch indicate the posterior probabilities, values to the right give the number of sequences in each cluster, and the scale bar shows the mean mutation rate per site. The clustering of αGmhB and of HisB-N sequences is both supported by highly significant posterior probabilities. Rooting was performed with the help of the αGmhB group. (b) In vivo and in vitro HolPase activities of GmhB from *Crassaminicella sp*. (*cs*GmhB) (upper panel) and *ec*GmhB (lower panel). A vector encoding *cs*GmhB is able to rescue an *E. coli* ΔholPase knockout strain and *cs*GmhB (10 μM) shows 27% HolP turnover within 20 h. In contrast, *ec*GmhB lacks HolPase activity both in vivo and in vitro. For *ec*HisB-N (10 nM), 100% HolP turnover was observed.

several clusters that include both βGmhB and HisB-N representatives. The clusters closest to this branchpoint are populated by αGmhB and βGmhB enzymes. This probably was the first functional divergence and discriminates between S-layer and lipopolysaccharide biosynthesis. Thus, this separation was used for rooting the tree. Interestingly, the HisB-N enzymes arise as a distinct cluster within the βGmhB branch. This topology indicates a closer evolutionary relationship between βGmhB and HisB-N than between αGmhB and HisB-N. This relationship is in line with the finding that *ec*HisB-N shows promiscuous activity for βHBP but not for αHBP (Figure 2b,c).

Based on these results, we tested whether βGmhBs that are located closer to the HisB-N cluster (i.e., separated by fewer nodes) than *ec*GmhB show detectable HolP hydrolysis. Indeed, we could identify a low promiscuous HolPase activity for GmhB of *Crassaminicella sp*. (*cs*GmhB, Figure 4b, upper panel). The purity and structural integrity of the enzyme was assessed by SDS-PAGE (Figure S7) and far-UV CD-spectroscopy (Figure S8). Moreover, steady-state enzyme kinetics showed that *cs*GmhB hydrolyzed its native substrate βHBP with high catalytic efficiency and exhibited a 36-fold preference for βHBP over αHBP (Figure S9, Table 1).

2.4 | Ancestral sequence reconstruction

The promiscuous HolPase activity of *cs*GmhB could have arisen by chance, but it could also be an inherited property that was preserved over the course of evolution. The latter would lead to the hypothesis that HolPase activity was a common feature of ancient progenitor enzymes. With the aim to test this, we decided to perform ancestral sequence reconstruction based on the phylogenetic tree shown in Figure S6 and to functionally characterize the resurrected enzymes. To conduct a thorough analysis of the evolutionary trajectory that led to the modern HisB-Ns, we reconstructed a set of seven enzymes, which were dubbed Anc1–Anc7 from the oldest to the youngest

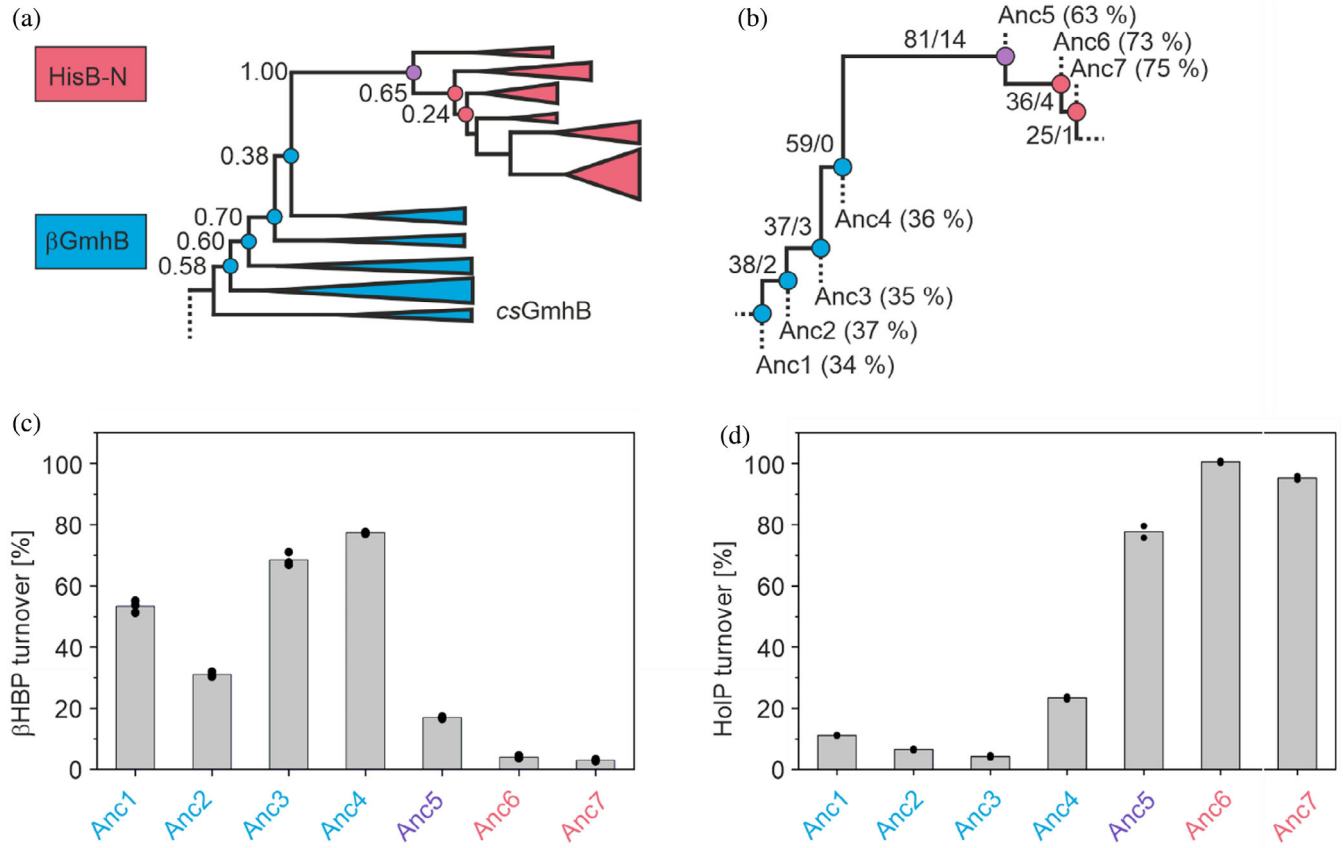

FIGURE 5 Ancestral sequence reconstruction and functional characterization of the resurrected progenitors of βGmhB and HisB-N. (a) Detailed view of the abstracted phylogenetic tree shown in Figure 4a; reconstructed nodes are marked with circles and posterior node probabilities are given next to each edge. (b) Position in the phylogenetic tree of the ancestors Anc1–Anc7 that were reconstructed using a maximum likelihood approach. The first number at each branch indicates the number of mutations and the second number gives the number of insertions or deletions between two adjacent variants. Numbers in brackets give the percent sequence identity of Anc1–Anc7 as compared to *ec*HisB-N. (c) Average amount of product formation within 20 h for the dephosphorylation of βHBP (30 μM) as detected by three independent measurements (black dots) that were performed with Anc1–Anc7 (5 μM). Incubation of βHBP with Anc1–Anc4 results in conversion rates above 30%, while for Anc5–Anc7, the conversion is decreasing to below 20%. (d) Average amount of product formation within 20 h for the dephosphorylation of HolP (55 μM) as detected by two independent measurements (black dots) that were performed with Anc1–Anc7 (10 μM). Incubation of HolP with Anc1–Anc5 results in partial conversion of the substrate, while Anc6 and Anc7 show more than 95% conversion.

variant (Figure 5a,b, sequences of Anc1–Anc7 are given in Table S1; a complete tree and the remaining ancestral sequences are provided in the Supplementary files S5 Tree_ancNodes.tre and S1 MSA_ancestors.csv, respectively).

The posterior probabilities for several associated ancestral nodes are moderate, which is caused by some uncertainty in the positioning of the extant enzymes that are derived from these nodes (Figure 5a). One could in principle increase the posterior probabilities by neglecting extant enzymes and working with a reduced data set. However, this would also reduce the number of ancestral nodes and raise the number of mutations between any pair of ancestral enzymes. Yet in this case, the ancestral enzymes were already separated by a considerable number of mutations (Figure 5b), which implies the vast evolutionary distance that is covered by this phylogenetic tree. To balance the trade-off between robustness and evolutionary detail, we opted against a further reduction of the dataset. Despite the high number of mutations between the different ancestors, the individual sequences are supported by high marginal ancestral probabilities across all residues with median values ≥99.8% for all seven reconstructed ancestors (details are given in Table S3; all data shown in the Supplementary file S1 MSA_ancestors.csv).

The genes for Anc1–Anc7 were codon optimized for the expression in *E. coli* and cloned into plasmids encoding an N-terminal His_6-tag (sequences of the constructs are given in Table S1). The genes were overexpressed,

TABLE 2 Activity toward HolP of Anc1–Anc7 and *ec*HisB-N at 25°C

Enzyme	Substrate	k_{cat} (s^{-1})	K_M (μM)	k_{cat}/K_M ($s^{-1}\ M^{-1}$)
Anc1-4	HolP[a]	4%–24% turnover (5 μM enzyme, 55 μM substrate, 20 h)		
Anc5	HolP[b]	$(15 \pm 1.7) \times 10^{-3}$	698 ± 123	21
Anc6	HolP[b]	1.8 ± 0.15	59 ± 12	30,508
Anc7	HolP[b]	0.8 ± 0.08	122 ± 31	6557
*ec*HisB-N	HolP[b]	2.8 ± 0.07	48 ± 3.3	57,437

[a]Data from discontinuous activity assays.
[b]Data from steady-state kinetic experiments.

and the proteins were purified by affinity chromatography followed by size exclusion chromatography. For Anc6 and Anc7, highly pure protein was obtained (Figure S10). However, for Anc2–Anc4, the protein levels as judged by SDS-PAGE were very low, possibly due to misfolding and subsequent degradation. Similarly, for Anc1 and Anc5, the purity of the target proteins remained limited, which obstructed further characterization. Therefore, Anc1–Anc5 were subcloned into an expression plasmid with an N-terminally fused maltose-binding protein (MBP), which served as a solubility tag. With this tag, Anc1–Anc5 could be obtained with good purity (Figure S10). Moreover, far-UV CD spectra indicated that all resurrected proteins were properly folded (Figure S11).

In the next step, we functionally characterized Anc1–Anc7. For this purpose, proteins were incubated with either βHBP or HolP, followed by quantification of product formation. Remarkably, all reconstructed enzymes were able to catalyze the turnover of both substrates (Figure 5c,d). Regarding βHBP hydrolysis, it is of note, however, that none of the variants showed complete product formation within 20 h. This result may seem surprising at least for Anc1–Anc4, which are the immediate precursors of extant βGmhBs. A plausible explanation for incomplete product formation would be that only a subfraction of the proteins was properly folded and active even in the presence of MBP. The recorded far-UV CD-spectra do however not show any indication for a large fraction of disordered regions. Another possible explanation lies in the moderate posterior probabilities linking Anc1 with Anc4 (Figure 5a) and the finding that the erroneous prediction of only some relevant residues might already lead to a dramatic drop in activity (Bonthron & Jaskólski, 1997; Jacquier et al., 2013). Regarding HolP hydrolysis, Anc5–Anc7 accomplish almost complete product formation within 20 h, which is in accordance with their close phylogenetic proximity to extant HisB-Ns. To further quantify the observed HolP turnover by Anc5–Anc7, steady-state enzyme kinetic measurements were performed (Figures S12–S14). The determined catalytic parameters are listed in Table 2.

While the activities of Anc1–Anc4 were too low for reliable measurements, the kinetic parameters of Anc5–Anc7 could be determined. Anc5 exhibits a k_{cat} value of 0.015 s^{-1} and a K_M of approximately 700 μM, which is a drastic improvement over Anc1–Anc4 but still constitutes a moderate activity as compared to the extant *ec*HisB-N. Anc6 and Anc7 display k_{cat} values of 1.8 and 0.8 s^{-1} and K_M values of 59 and 122 μM. This indicates both a drastic improvement in the turnover number and a significant improvement in affinity for HolP as compared to Anc5. Taken together, Anc6 and Anc7 showed HolPase activities that are similar to the HolPase activity of the extant *ec*HisB-N.

In summary, the in vitro analysis showed that the HolPase activity was already present as a side activity in the ancestors that precede the branchpoint between HisB-N and βGmhB, while high catalytic activity was established only after the separation of the HisB-N cluster in the phylogenetic tree.

2.5 | Structural analysis of predecessors Anc1–Anc7

The functional analysis of the ancestral sequences revealed a drastic increase in the HolPase activity during the evolution from Anc1–Anc4 to Anc5–Anc7. With the aim to rationalize the causes for this observation, the structures of Anc1–Anc7 were predicted with AlphaFold (Jumper et al., 2021) and compared with the structures of *ec*HisB-N and *ec*GmhB (Figure 6).

When comparing the structures of Anc1–Anc7, it became evident that they exhibit the same overall fold with only minor differences in the distal parts of the enzymes (Figure S15). Moreover, a close inspection of the active site revealed that the subsite closest to the phosphate binding site also showed high similarity (Figure 6a, indicated by magnesium ions that are represented as yellow spheres).

At the solvent exposed parts, however, the enzymes differed significantly, which allowed for the distinction of three groups, namely Anc1–Anc4, Anc5, and Anc6–

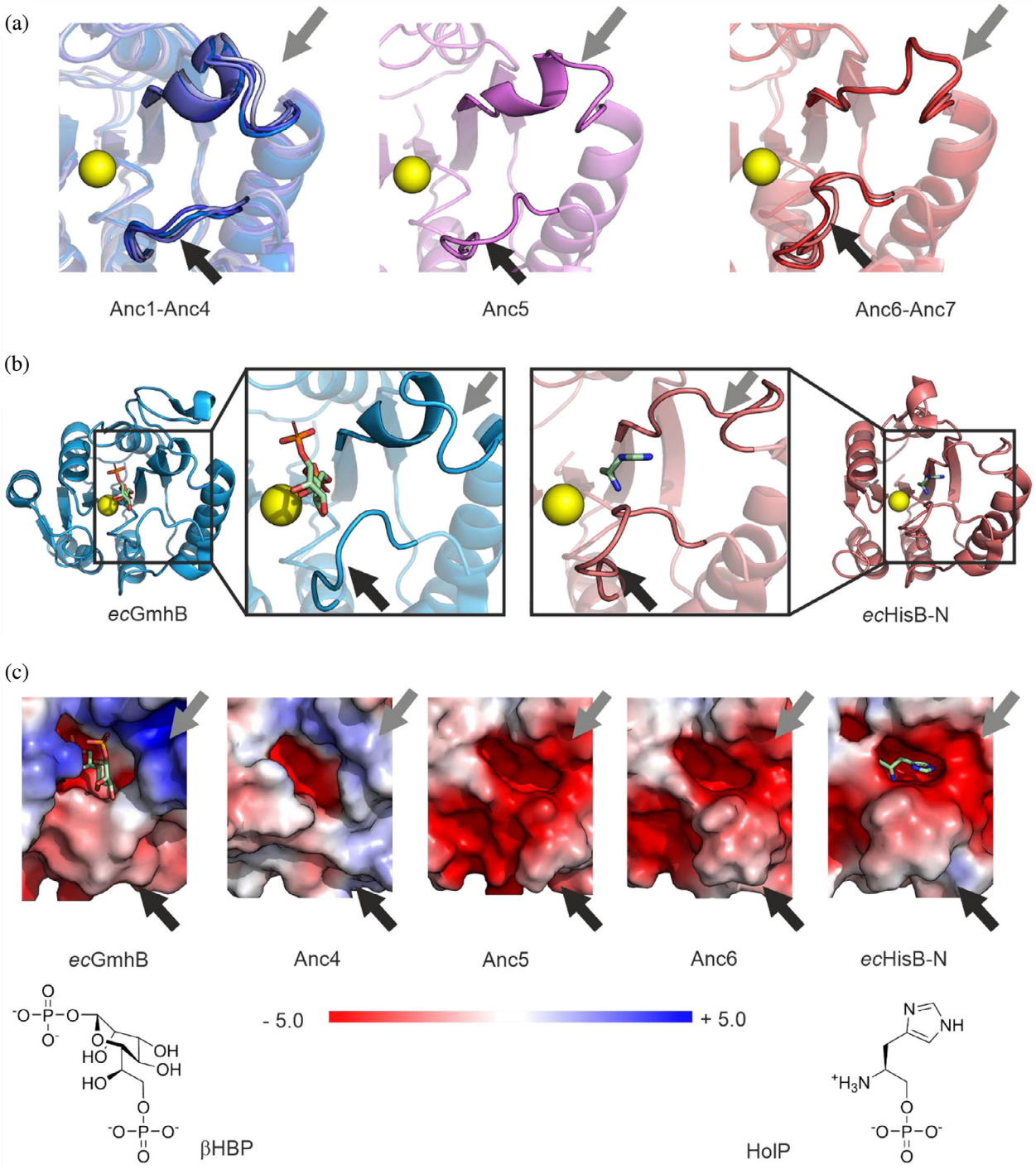

FIGURE 6 Structural analysis of fold, shape, and charge distribution of the active sites of Anc1–Anc7. (a) Active sites of the structures of Anc1–Anc4 (superimposed in the left panel), of Anc5 (middle panel), and of Anc6–Anc7 (superimposed in the right panel) as predicted by AlphaFold (Jumper et al., 2021). The yellow spheres represent magnesium ions that were extracted from a superimposed *ec*HisB-N structure. The black arrows indicate a lid-like loop structure, the length of which is increased from Anc1 to Anc7. The increased length was probably necessary to ensure efficient closure of the active site while the size of the substrate decreased. The gray arrows mark a region, which folds as helix in Anc1–Anc4 and as loop in Anc5–Anc7. (b) Overall structures and zoomed-in active site views of *ec*GmhB (PDB ID: 3L8G) with bound βHBP (green sticks) and *ec*HisB-N (PDB ID: 2FPU) with bound histidinol (green sticks). (c) Surface shape and charge distribution of *ec*GmhB, Anc4, Anc5, Anc6, and *ec*HisB-N and the two substrates βHBP and HolP. In *ec*GmhB and Anc4, a short helix limits the space toward the right (gray arrows), whereas in Anc5, Anc6, and *ec*HisB-N, a loop creates a bigger binding pocket toward the right (gray arrows). Moreover, in *ec*GmhB and Anc4, the surface exposed parts of the active sites are charged positively allowing for the interaction with the second phosphate group and hydroxyl groups of βHBP, while they are negatively charged in Anc5–Anc6 and *ec*HisB-N allowing for the interaction with the amine group and imidazole ring of HolP.

Anc7. In Anc1–Anc4, the lid-like loop region is four residues shorter than in Anc5 and the loop in Anc5 is one residue shorter than in Anc6–Anc7 (Figure 6a, black arrow, loop sequences are given in Table S2). This is in accordance with the observation that in *ec*GmhB the corresponding loop is three residues shorter than in *ec*HisB-N (Figure 6b, Table S2). This finding fits to an earlier argument according to which a bulky substrate requires a short loop, whereas a small substrate requires a long loop that ensures the exclusion of water from the active site (Burroughs et al., 2006).

In addition to that, Anc1–Anc4 differ from Anc5 and Anc6–Anc7 in the top right part of the active site (Figure 6a–c, Figure S16, gray arrows). The secondary structure in Anc1–Anc4 was predicted as a short helix, whereas in Anc5–Anc7, there is a short loop (Figure 6a, gray arrows). In *ec*GmhB, the corresponding short helix fills the space thereby limiting the active site (Figure 6b). In contrast, in *ec*HisB-N, the corresponding loop region creates a small binding pocket, which is occupied by the imidazole ring of the substrate (Figure 6b). A similar observation can be made for the ancestral enzymes. In Anc1–Anc4, the space to the right is occupied by the helix residues, whereas the loop in Anc5 and more so in Anc6–Anc7 takes up less space and creates a cavity for the substrate HolP (Figure 6c).

Finally, the electrostatics of the active sites differ significantly: While the deep end of the active site is negatively charged in all enzymes, the surface-exposed parts of the active site are positively charged in the case of Anc1–Anc4 and *ec*GmhB but negatively charged in Anc5–Anc7 and *ec*HisB-N (Figure 6c). The positive charge in *ec*GmhB and Anc1–Anc4 most likely enables electrostatic interactions with the phosphate and hydroxyl groups of the substrate βHBP. In contrast, the negative charge of the surface-exposed parts in Anc5–Anc7 ensures charge complementarity with the amino group and the imidazole ring of HolP, which is protonated to a significant degree at neutral pH. Interestingly, Anc5 shares most of these properties with *ec*HisB-N, but it is still a catalyst with moderate efficiency. It can therefore be concluded that the significant structural adaptations observed between Anc4 and Anc5 are necessary to allow for a reasonable turnover of HolP, but only the mutations between Anc5 and Anc6 lead to an efficiency boost.

In summary, the analyses showed that the HolPase activity was already present in the early ancestors Anc1–Anc4, which share many of the structural features of modern βGmhBs. Starting with the transition from Anc4 to Anc5 the HolPase activity was enhanced. This enhancement was accompanied by significant adaptations of the shape and electrostatics of the active site.

2.6 | A revised model for the evolution of HisB-N

Previously, it has been hypothesized that HisB-N and GmhB were both derived from the same ancestral HAD phosphatase, which underwent a gene duplication event. One of these copies was integrated into the *his* operon whereupon its gene product evolved toward a modern HisB-N (Brilli & Fani, 2004). Indeed, our results now show that HisB-N sequences form a sub-cluster within a bigger βGmhB cluster (Figure 7). This sequence relationship is additionally supported by function: *ec*HisB-N and its most recent progenitors Anc6 and Anc7 are still able to hydrolyze βHBP. At the same time, the HolPase activity could already be detected in ancestral βGmhB-like enzymes Anc1–Anc4 that existed long before the functional divergence of HisB-N and βGmhB. This refines the previous model, as it suggests that HisB-N is derived from a βGmhB and not an αGmhB.

Next, we were interested to follow the evolution of these two enzymatic functions on an organismal level and to this end reexamined the phylogenetic tree (Figure 7).

The most recent common precursor of HisB-N and βGmhB in this phylogenetic tree is represented by Anc4 (or in a broader sense by Anc1–Anc4). The βGmhB enzymes (blue triangles) that are directly derived from these ancestors are mostly found in Thermodesulfobacteria and δ-Proteobacteria. An ongoing debate proposes to combine δ-Proteobacteria and Thermodesulfobacteria in one phylogenetic group (Waite et al., 2020), which is in agreement with the similarity of their βGmhB sequences. However, neither δ-Proteobacteria nor Thermodesulfobacteria possess HisB-N enzymes. It is hence unlikely that a duplication event of the precursor of HisB-N and βGmhB happened in a δ-Proteobacterium or a Thermodesulfobacterium.

The HisB-N enzymes (red triangles) that are directly derived from Anc4 almost exclusively belong to γ-Proteobacteria (for a detailed discussion see Figure S18). Interestingly, γ-Proteobacterial species form two separate βGmhB clusters and this clustering correlates with the occurrence of HisB-N in an organism. We named the γ-Proteobacteria that possess both a gmhB and a hisB-N gene $\gamma^{+\text{HisB-N}}$-Proteobacteria, and the species that only possess a gmhB gene $\gamma_{\text{-HisB-N}}$-Proteobacteria. The finding that some γ-Proteobacteria lack a HisB-N like HolPase poses the question how histidine is synthesized by these bacteria. In principle, these $\gamma_{\text{-HisB-N}}$-Proteobacteria could either (i) rely on histidine uptake from external sources, (ii) possess a HolPase from the PHP or IMP superfamily, or (iii) possess a HolPase from a different superfamily that is yet to be discovered. Which

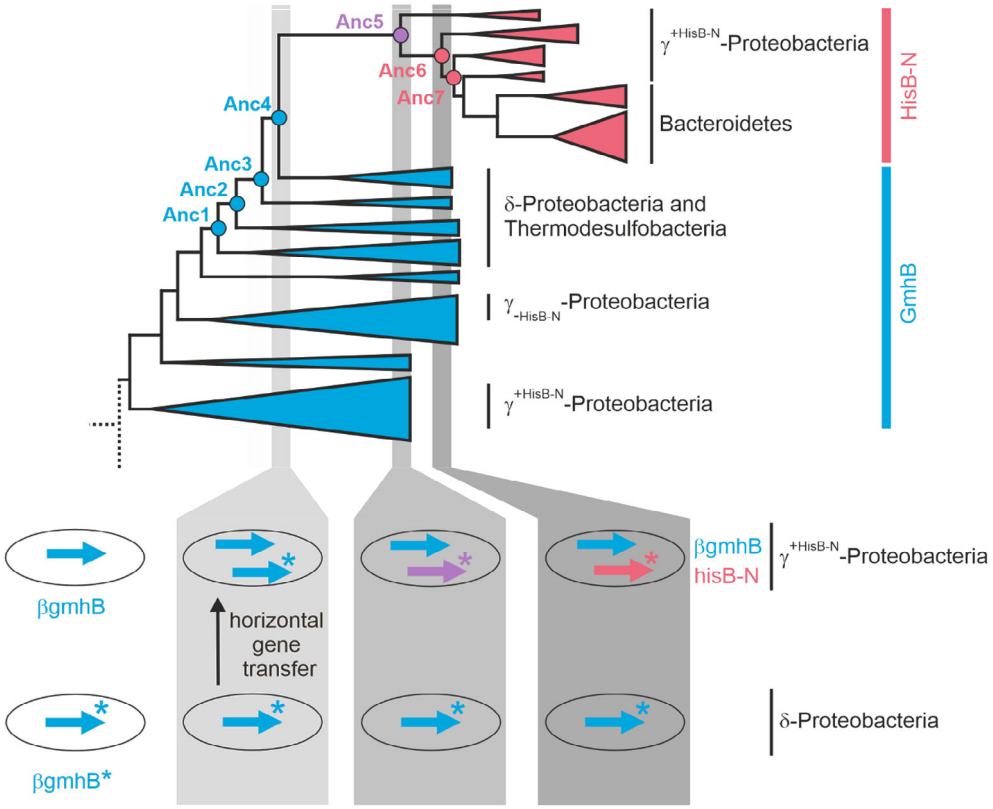

FIGURE 7 Phylogenetic distribution of *hisB-N* and βgmhB genes, and revised evolutionary model. The upper panel represents parts of the condensed phylogenetic tree shown in Figure 4a with clusters colored according to enzyme function (HisB-N in red, βGmhB in blue). Bacterial phyla are given on the right of each cluster. The γ-Proteobacteria can be divided into two groups: $\gamma^{+\text{HisB-N}}$-Proteobacteria possess a βgmhB and a hisB-N gene; in contrast $\gamma_{\text{-HisB-N}}$-Proteobacteria only possess a gmhB gene. The panel below illustrates the proposed evolutionary scenario. At the stage of Anc3, $\gamma^{+\text{HisB-N}}$-Proteobacteria possessed a gene for a βGmhB enzyme (blue arrow), while δ-Proteobacteria had a gene for a promiscuous enzyme βGmhB* (blue arrow with asterisk). At the stage of Anc4, the horizontal gene transfer of βgmhB* resulted in two functional βgmhBs in $\gamma^{+\text{HisB-N}}$-Proteobacteria. At the stage of Anc5, mutations in the βgmhB* gene (violet arrow) led to a significant increase of the originally low HolPase activity. From stage Anc6 onward, further mutations in the βgmhB* gene (red arrow) led to the modern HisB-N enzyme.

of the scenarios is true for which individual species cannot be answered comprehensively with the currently available data but requires further investigation. However, a phylogenetic tree of γ-proteobacterial 16S rRNA sequences revealed that the $\gamma^{+\text{HisB-N}}$-species form a distinct cluster and are thus closely related (Figure S17), which confirmed the separation of γ-Proteobacteria into two groups. In summary, this strongly suggests that the first co-occurrence of HisB-N and βGmhB took place in $\gamma^{+\text{HisB-N}}$-Proteobacteria. However, the cluster of βGmhB sequences from $\gamma^{+\text{HisB-N}}$-Proteobacteria is located far apart from the HisB-N cluster. This separation of βGmhB and HisB-N from $\gamma^{+\text{HisB-N}}$-Proteobacteria is supported by a number of interjacent nodes with intermediate to high posterior probabilities. This large distance in the tree renders the previously postulated duplication event in an ancient $\gamma^{+\text{HisB-N}}$-Proteobacterium (Brilli & Fani, 2004) unlikely.

Therefore, we propose an alternative evolutionary scenario, which is outlined in the lower part of Figure 7. This scenario assumes that at the timepoint presented by Anc3, βGmhBs were already present in the first $\gamma^{+\text{HisB-N}}$- and δ-Proteobacteria. The βGmhBs of these ancestral δ-Proteobacteria (βGmhB*) were promiscuous for HolP either because βGmhB* supported two metabolic pathways or due to the lack of substrate specificity. Whether this promiscuity represents a favorable degree of substrate promiscuity or if it is an evolutionary irrelevant degree of sloppiness, cannot be decided with certainty. The causes and consequences of enzyme promiscuity in general are a matter of ongoing debate (Babtie et al., 2010; Copley, 2020; Pandya et al., 2014; Peracchi, 2018). The gene of such a promiscuous βGmhB* was acquired by an ancestral $\gamma^{+\text{HisB-N}}$-Proteobacterium via horizontal gene transfer probably from an ancient Thermodesulfobacterium or

δ-Proteobacterium. According to the phylogenetic tree, an ancient $\gamma_{\text{-HisB-N}}$-, or β-Proteobacterium would also be a possible donor species. Since all modern HisB-N enzymes are N-terminally fused to IGPDH and the respective bi-functional genes are located within the *his* operon, one can assume that after the horizontal gene transfer βgmhB* was incorporated into the *his* operon upstream of the gene that encodes for the IGPDH, as argued previously (Brilli & Fani, 2004; Del Duca et al., 2020). The receiving $\gamma^{+\text{HisB-N}}$-Proteobacterium already possessed a βGmhB enzyme, so the selective pressure on the GmhB function of the newly transferred βgmhB* gene was likely reduced. However, the promiscuous side activity for HolP probably conferred some evolutionary advantage, which led to an increase of the primordial HolPase activity that is visible in Anc5–Anc7.

This evolutionary scenario is also supported by the phylogenetic distribution of the properties of the zinc-binding cap structure (Figure S18, MSA: residue positions 97–115, see Supplementary file MSA_extant_sequences.fa S1). On the N-terminal end, this structure is defined by a CxH motif and on the C-terminal end by a CxC motif. In the βGmhB variants from $\gamma^{+\text{HisB-N}}$-Proteobacteria, Negativicutes, Aquificales and most α-Proteobacteria, the CxH and CxC motifs are separated by 12 or 13 residues. This long cap structure is predominately found in phylogenetically diverse βGmhB enzymes that are close to the branchpoint with αGmhBs. The length of this cap structure and the chemical properties of many residues are conserved, including the strictly conserved tyrosine at residue position 109 (Figure S18). Taken together, conservation of this cap structure and its broad phylogenetic distribution suggests that these βGmhBs correspond to the ancient state. Conversely, in the βGmhB variants from β-, $\gamma_{\text{-HisB-N}}$-, δ-Proteobacteria, Thermodesulfobacteria, and some α-Proteobacteria, the CxH and CxC motifs are separated by merely five to six residues. The same pattern also occurs in HisB-N variants from $\gamma^{+\text{HisB-N}}$-Proteobacteria, which again underscores the relationship between these HisB-N sequences and βGmhBs from β-, $\gamma_{\text{-HisB-N}}$-, δ-Proteobacteria, and Thermodesulfobacteria. Interestingly, the CxH and CxC motifs are missing in HisB-N from Bacteroidetes and few $\gamma^{+\text{HisB-N}}$-Proteobacteria, suggesting that they were lost during the course of evolution.

Overall, all sequences generated by our ancestral sequence reconstruction (ASR) approach possess high marginal probabilities, which suggests high prediction accuracy (Thornton, 2004). Nevertheless, errors in sequence reconstruction can lead to erroneous estimates of structural or functional properties of the deduced ancestors. It has been shown that alignment errors can more severely affect ASR accuracy than phylogenetic uncertainty (Aadland & Kolaczkowski, 2020). In order to minimize alignment errors, we utilized airas (Aadland & Kolaczkowski, 2020), which reduces uncertainty by combining information from different MSAs compiled for the same sequence set. Although phylogenetic uncertainty is commonly considered a less severe problem, we cannot exclude respective ASR errors for individual sequences (Garcia & Kaçar, 2019), mainly because some ancestral nodes possess moderate statistical support values. This is why we did not focus on one specific ancestral sequence for these critical cases but analyzed them and their properties as a set of putative ancestors (Gaucher, 2007). The sequences Anc1–Anc4 and Anc5–Anc7 can be considered as two ensembles, each representing one enzyme function. Importantly, these two ensembles are divided by a strongly supported split and in both ensembles, the length of the zinc binding loop is the same as in the corresponding extant sequences. Thus, we can rule out a frequently observed ASR bias toward the consensus sequence (Bershtein et al., 2008) and assume that this critical element of enzyme function has been adequately reconstructed by our ASR protocol.

Altogether, according to the phylogenetic tree, the experimental data, and the discussed structural features, the proposed horizontal gene transfer seems to be the most convincing evolutionary scenario. However, one has to take into account the generally high sequence divergence of the HAD superfamily (Burroughs et al., 2006). Accordingly, HisB-N and GmhB sequences, but also different GmhB sequences share not more than 30%–40% identical residues. Therefore, a gene duplication event cannot be ruled out with certainty.

In conclusion, our findings nicely fit the innovation amplification and divergence (IAD) model of evolution (Bergthorsson et al., 2007). This model assumes that a (fortuitous) promiscuous side activity of an enzyme can become relevant when posed under selective pressure, as in this case the promiscuous HolPase activity of Anc1–Anc4. It has also been shown that genes that were acquired by horizontal gene transfer from distantly related donor species make up for a significant fraction of genes in bacteria (Ochman et al., 2000). Moreover, it has been argued that horizontal gene transfer plays an underestimated role in the evolution of new enzymes, as a side activity of a newly transferred gene that contributes to the recipients fitness would be a perfect starting point for evolution (Glasner et al., 2020). The herein observed boost in HolPase activity after the horizontal transfer of Anc4 from δ-Proteobacteria to $\gamma^{+\text{HisB-N}}$-Proteobacteria is in line with this rationale.

3 | MATERIALS AND METHODS

3.1 | Cloning

Genes encoding target proteins were codon optimized for *E. coli* and *BsaI* cleavage sites were added. Then, the genes were subcloned into isopropyl-β-D-1-thiogalactopyranoside (IPTG) inducible expression plasmids with N-terminal His_6-tags (pUR23) or a combination of N-terminal His_6-tag and N-terminal maltose binding protein (pMal_BsaI) for overexpression. For complementation experiments, genes were cloned into pTNA_BsaI which ensured low constitutive expression through a tryptophanase operator (Rohweder et al., 2018).

3.2 | Purification of HisB-N and GmhB enzymes

Genes encoding target proteins were expressed in a Δ*his*B knockout strain that was derived from *E. coli* strain BW25113. Transformed cells were grown at 37°C to an OD_{600} of 0.6–0.8. Then, gene expression was induced by the addition of IPTG (0.5 mM) and cells were grown over night at 20°C. Afterward, cells were harvested by centrifugation (rcf 4000 g, 20 min, 4°C), resuspended in buffer (50 mM Tris/HCl, pH 7.5, 400 mM NaCl, 2 mM $MgCl_2$, 10 mM imidazole) and disrupted by sonication (Heinemann Branson sonifier, 60% amplitude, 2 min 30 s sonication time, 2 s on, 2 s off). Cell debris was removed by centrifugation (rcf 23,700 g, 40–45 min, 4°C) and the target proteins were purified from the soluble fraction by affinity chromatography (HisTrap FF crude column, 5 ml, GE Healthcare; buffer: 50 mM Tris/HCl, pH 7.5, 400 mM NaCl, 2 mM $MgCl_2$ with linear gradient of imidazole from 10 to 500 mM) and subsequent size exclusion chromatography (HiLoad™ 26/600 Superdex™ S75 or S200, GE Healthcare, buffer: 50 mM Tris/HCl, pH 7.5, 400 mM NaCl, 2 mM $MgCl_2$). Purified proteins were dripped into liquid nitrogen and stored at −80°C. Protein purity was assessed by SDS-PAGE and protein concentrations were determined by absorbance spectroscopy using molar extinction coefficients at 280 nm that were calculated with ProtParam (Wilkins et al., 1999).

3.3 | Enzymatic synthesis and purification of αHBP and βHBP

The two anomeric substrates βHBP and αHBP were synthesized enzymatically similar to a previously reported protocol (Wang et al., 2010). The synthesis required several auxiliary enzymes, namely transketolase (*ec*Tkt1) from *E. coli*, D-sedoheptulose-7-phosphate isomerase (*ec*GmhA) from *E. coli*, a β-C(1)OH specific D-*glycero*-D-*manno*-heptose-7-phosphate kinase from *Bordetella bronchiseptica* (*bb*HldE), an α-C(1)OH specific D-*glycero*-D-*manno*-heptose-7-phosphate kinase from *Aneurinibacillus thermoaerophilus* (*at*HddA), and adenylate kinase from *E. coli* (*ec*ADK). The corresponding genes were codon optimized for *E. coli*, cloned into expression plasmids, and overexpressed by the addition of IPTG. The target proteins were then purified by affinity chromatography and the purity was assessed by SDS-PAGE. The commercially available substrates fructose-6-phosphate (97 mg), ribose-5-phosphate (110 mg), ATP (300 mg), thiamine pyrophosphate (50 mg), and D-glyceraldehyde-3-phosphate (80 µl, 10 mM) were incubated with the enzymes *ec*Tkt1, *ec*GmhA, and *bb*HldE or *at*HddA at 37°C for 24 h. Afterward, *ec*ADK was added to reduce the amount of ADP and thereby increase the purity of the final product. All enzymes were subsequently removed by filtration, the pH was adjusted to 9.0 and the final product was isolated by anion exchange chromatography (MonoQ HR 16/10, equilibration buffer: 50 mM NH_4HCO_3, pH 9.0; elution buffer: 1 M NH_4HCO_3, pH 7.8). The amount of formed product was quantified by total turnover by *ec*GmhB. For βHBP, the identity of the product was additionally verified by mass spectrometry (Figure S19).

3.4 | Steady-state kinetics and discontinuous activity assays

The activities of different enzymes under steady-state conditions were determined photometrically by quantification of phosphate release using a coupled assay at 25°C (Suárez et al., 2012). Reaction conditions included 100 mM Tris/HCl (pH 7.8), 5 mM $MgCl_2$, 0.5 mM inosine, 0.25 U/ml purine nucleoside phosphorylase (Sigma-Aldrich), 2.5 U/ml xanthine oxidase (Sigma-Aldrich), and variable concentrations of HolP (Sigma-Aldrich), αHBP, or βHBP. The reactions were started by addition of the respective enzyme. Kinetic parameters were calculated from the initial ascending slopes by curve fitting with the Michaelis–Menten equation (Origin 2019 and Origin 2021, OriginLab).

For discontinuous activity assays, the respective substrate was incubated with the enzyme of interest (10 nM to 10 µM) in buffered solution (100 mM Tris/HCl, pH 7.8; 5 mM $MgCl_2$) for up to 20 h at 37°C. As negative control, the respective substrate was incubated with buffer. As positive control, HolP was incubated with *ec*HisB-N while αHBP and βHBP were incubated with *ec*GmhB.

The reactions were stopped by removing the enzyme by filtration (pore size: 0.22 μm). Afterward, the amount of free phosphate of the filtered reaction mixture was determined photometrically by the addition of inosine, purine nucleoside phosphorylase, and xanthine oxidase, following the same principle as in the steady-state kinetic measurements. The amount of free phosphate in each sample was corrected for the amount of phosphate detected in the negative control caused by impurities or spontaneous hydrolysis.

In the case of Anc1–Anc4, we observed side activity toward the contaminations of the synthesized substrate βHBP. Thus, quantification of product formation via quantification of free phosphate was prone to overestimate the amount of hydrolyzed substrate. Therefore, we instead determined the amount of formed product by quantification of the residual βHBP. To this end, we incubated substrate and Anc1–Anc7 overnight and then removed the enzymes by filtration. The amount of free phosphate was determined as described above by the addition of inosine, purine nucleoside phosphorylase, and xanthine oxidase. When a plateau was reached, *ec*GmhB was added which would hydrolyze all remaining βHBP. The amount of remaining βHBP was calculated and subtracted from the starting amount, which would give the amount that was hydrolyzed by Anc1–Anc7.

3.5 | Circular dichroism spectroscopy

Structural integrity of the proteins was assessed by far-UV CD-spectroscopy. Prior to measurements, buffer was exchanged to 20 mM KP pH 7.5, to avoid absorption of Tris in the far UV region. Spectra were then recorded with a CD-spectrometer (J-815; JASCO) between 260 and 190 nm using a quartz cuvette (0.2 mm) and measuring five to eight replicas. All spectra were corrected for buffer absorption and smoothed using the Savitzky–Golay algorithm (Arnold et al., 2003) with a convolution width of 7. The mean molar ellipticity per residue θ_{MRW} (deg cm^2 dmol^{-1}) was calculated from the observed ellipticity θ_{obs} (mdeg), the width of the cuvette d (cm), the protein concentration c (μM), and number of residues N_A, according to the following equation:

$$\theta_{MRW} = \frac{\theta_{obs} \cdot 10^5}{c \cdot d \cdot N_A}$$

3.6 | Complementation of a HolPase-deficient *E. coli* knockout strain

Complementation experiments were conducted using a *hisB*-deficient *E. coli* strain with a kanamycin resistance gene (Δ*hisB*::*KanR*), which was constructed following the protocol by Datsenko and Wanner (Datsenko & Wanner, 2000). To complement for the missing IGP dehydratase function of the bifunctional HisB enzyme, Δ*hisB* strain was transformed with a plasmid for constitutive expression of the monofunctional IGP dehydratase from *Bacillus subtilis* and a chloramphenicol resistance gene (pTNA_*igpdh_CamR*). Consequently, only the HolPase function is missing, and different GmhB variants could be tested for their HolPase activity. The corresponding genes were subcloned into a vector for low constitutive expression and an ampicillin resistance gene (e.g., pTNA_*gmhB_AmpR*). The HolPase-deficient cells were transformed with genes of interest, streaked on M9$^-$ minimal medium plates equipped with ampicillin, chloramphenicol, and kanamycin, and plates were incubated at 37°C.

3.7 | Phylogenetic analyses and ancestral sequence reconstruction

GmhB (both with αHBP and βHBP as native substrate) and HisB-N sequences were obtained from the KEGG database(Kanehisa & Goto, 2000). GmhB variants missing the annotation of the preferred substrate were classified using the occurrence of a specific nucleotidyltransferase (either D-*glycero*-α-D-*manno*-heptose 1-phosphate guanylyltransferase [KEGG ID K15669] or D-*glycero*-β-D-*manno*-heptose 1-phosphate adenylyltransferase [KEGG ID K21345]) in the respective genome. The occurrence of either enzyme was used as indication for the existence of either the D-*glycero*-D-*manno*-heptose-1α-GDP biosynthesis (for S-layer biosynthesis) or the L-*glycero*-D-*manno*-heptose-1β-ADP biosynthesis (for lipopolysaccharide biosynthesis) in an organism and was used to identify αHBP or βHBP as native substrate of a GmhB enzyme (Kneidinger et al., 2001; Kneidinger et al., 2002; Taylor et al., 2010; Valvano et al., 2002; Wang et al., 2010).

The retrieved and annotated sequences were then filtered, and phylogenetic trees were calculated. The semi-supervised protocol for sequence selection implemented with FitSS4ASR (Straub et al., 2019) was used, utilizing MrBayes (Huelsenbeck & Ronquist, 2001) as phylogenetic program and the heatmap filtering approach. Below, deviations from default parameters are given for each method. Due to the low sequence identity between GmhB and HisB-N, the alignment program MAFFT (Katoh et al., 2002) was used with the parameters *BLOSUM45 matrix* and *a gap open penalty* of 3. In order to calculate a phylogenetic tree from meaningful alignment columns, the alignment was filtered with Gblocks (*b*3 = 5000, *b*4 = 2, *b*5 = a) (Castresana, 2000). The

phylogenetic analysis was performed with MrBayes calculating $5 \cdot 10^7$ tree generations (six independent chains, *gtr* as evolutionary model and *invgamma distribution*). A consensus tree was calculated with the option *sumt* where 25% of the tree generations acted as burn-in. Iteration 1 led to the most promising starting point for subsequent analysis. All possible combinations of the three filtering methods (*heatmap approach*, *rogue taxa* and *RogueNaRok*; Goloboff & Szumik, 2015) were applied manually to generate a data set with the best suitable selected sequences. The final data set with 192 sequences was obtained from the combination of all three filtering methods and a phylogenetic tree was calculated with MrBayes ($1 \cdot 10^7$ tree generations, six independent chains, *gtr* as evolutionary model and *invgamma* distribution). Rooting was performed with the help of the αGmhB group.

Previous studies have demonstrated that phylogenetic uncertainty has only a very weak effect on ASR (Aadland & Kolaczkowski, 2020; Hanson-Smith et al., 2010). As we consider a Bayesian approach as gold standard for phylogenetic analysis, we used the corresponding implementation of MrBayes (Huelsenbeck & Ronquist, 2001) to deduce a consensus tree. Unlike uncertainty of the true phylogeny, errors in sequence alignments can have a severe effect in ASR (Aadland & Kolaczkowski, 2020). We saw the need to minimize alignment errors, because our analysis of the enzymes' 3D structures indicated that the length of the functionally important substrate-binding loop and a structurally relevant zinc-binding loop varied in the extant sequences. Thus, we chose airas for sequence reconstruction, which has been shown to significantly improve the accuracy of ASR by combining information from several MSAs. These alignments are determined for the same sequence set by using different methods of MSA compilation (Aadland & Kolaczkowski, 2020). Alignments from MAFFT (Katoh et al., 2002), MAFFT ginsi, msaprobs (Liu et al., 2010; Liu & Schmidt, 2014), MUSCLE (Edgar, 2004), Probalign (Roshan & Livesay, 2006), CLUSTAL W (Thompson et al., 1994), and T-Coffee (Di Tommaso et al., 2011), each with default parameters, served as input. The alignment with MAFFT was created with the options as chosen for the phylogenetic analysis; the MSA is shown in the additional file. Ancestral sequences were obtained for the ancestral states deduced from the integrated alignment and the rooted phylogenetic tree. In accordance with the parameters used for the calculation of the phylogenetic tree the GAMMA model of rate heterogeneity with proportion of invariable sites, GTR as substitution model and a ML estimation of the base frequencies were selected.

MEGA11 (Tamura et al., 2021) utilized with default parameters and 500 bootstrap replications was used to deduce a tree based on the 16S-rRNA sequences of γ-Proteobacteria.

3.8 | Structure prediction with AlphaFold

The 3D-models were created with a locally installed version of Alphafold 2.1 (Jumper et al., 2021). The full databases were used and all parameters were left at default values.

3.9 | Conservation analysis with WebLogo

Sequence conservation of the zinc-binding cap structure (residue position 97–115) was analyzed by using WebLogo 3.7.12 (Crooks et al., 2004; Schneider & Stephens, 1990). The sequence logos (Figure S18) were created by defining the input sequences as amino acid, using the chemistry color code, and showing no error bars. All other parameters were set to default.

AUTHOR CONTRIBUTIONS

Thomas Kinateder: Conceptualization (equal); investigation (equal); methodology (equal); writing – original draft (equal). **Lukas Drexler:** Investigation (equal); methodology (equal); writing – review and editing (equal). **Kristina Straub:** Investigation (equal); methodology (equal); software (equal); writing – review and editing (equal). **Rainer Merkl:** Conceptualization (equal); funding acquisition (equal); project administration (equal); supervision (equal); writing – review and editing (equal). **Reinhard Sterner:** Conceptualization (equal); funding acquisition (equal); project administration (equal); supervision (equal); writing – review and editing (equal).

ACKNOWLEDGMENTS

The authors thank Sonja Fuchs, Sabine Laberer, and Jeannette Ückert for expert technical assistance. Moreover, the authors thank Julian Nazet for helping us to predict structures with AlphaFold and to assess the KEGG database, and Thomas Klein for providing the adenylate kinase from *E. coli*. This work was supported by grants of the Deutsche Forschungsgemeinschaft to Rainer Merkl (ME 2259/4-1) and Reinhard Sterner (STE 891/13-1). Open Access funding enabled and organized by Projekt DEAL.

ORCID

Thomas Kinateder https://orcid.org/0000-0001-5739-044X
Reinhard Sterner https://orcid.org/0000-0001-8177-8460

REFERENCES

Aadland K, Kolaczkowski B. Alignment-integrated reconstruction of ancestral sequences improves accuracy. Genome Biol Evol. 2020;12:1549–65.

Alifano P, Fani R, Liò P, Lazcano A, Bazzicalupo M, Carlomagno MS, et al. Histidine biosynthetic pathway and genes: structure, regulation, and evolution. Microbiol Rev. 1996;60:44–69.

Ames BN. The biosynthesis of histidine; L-histidinol phosphate phosphatase. J Biol Chem. 1957;226:583–93.

Arnold DN, Santosa F, Rosenthal J, Gilliam DS, editors. Mathematical systems theory in biology, communications, computation, and finance. New York, NY: Springer; 2003.

Babtie A, Tokuriki N, Hollfelder F. What makes an enzyme promiscuous? Curr Opin Chem Biol. 2010;14:200–7.

Bergthorsson U, Andersson DI, Roth JR. Ohno's dilemma: evolution of new genes under continuous selection. Proc Natl Acad Sci U S A. 2007;104:17004–9.

Bershtein S, Goldin K, Tawfik DS. Intense neutral drifts yield robust and evolvable consensus proteins. J Mol Biol. 2008;379:1029–44.

Bonthron DT, Jaskólski M. Why a "benign" mutation kills enzyme activity. Structure-based analysis of the A176V mutant of *Saccharomyces cerevisiae* L-asparaginase I. Acta Biochim Pol. 1997;44:491–504.

Brady DR, Houston LL. Some properties of the catalytic sites of imidazoleglycerol phosphate dehydratase-histidinol phosphate phosphatase, a bifunctional enzyme from *Salmonella typhimurium*. J Biol Chem. 1973;248:2588–92.

Brilli M, Fani R. Molecular evolution of hisB genes. J Mol Evol. 2004;58:225–37.

Burroughs AM, Allen KN, Dunaway-Mariano D, Aravind L. Evolutionary genomics of the HAD superfamily: understanding the structural adaptations and catalytic diversity in a superfamily of phosphoesterases and allied enzymes. J Mol Biol. 2006;361:1003–34.

Carlomagno MS, Chiariotti L, Alifano P, Nappo AG, Bruni CB. Structure and function of the *Salmonella typhimurium* and *Escherichia coli* K-12 histidine operons. J Mol Biol. 1988;203:585–606.

Castresana J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol. 2000;17:540–52.

Chiariotti L, Nappo AG, Carlomagno MS, Bruni CB. Gene structure in the histidine operon of *Escherichia coli*. Identification and nucleotide sequence of the hisB gene. Mol Gen Genet. 1986;202:42–7.

Chumley FG, Roth JR. Genetic fusions that place the lactose genes under histidine operon control. J Mol Biol. 1981;145:697–712.

Copley SD. The physical basis and practical consequences of biological promiscuity. Phys Biol. 2020;17:51001.

Crooks GE, Hon G, Chandonia J-M, Brenner SE. WebLogo: a sequence logo generator. Genome Res. 2004;14:1188–90.

Datsenko KA, Wanner BL. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. Proc Natl Acad Sci U S A. 2000;97:6640–5.

Del Duca S, Chioccioli S, Vassallo A, Castronovo LM, Fani R. The role of gene elongation in the evolution of histidine biosynthetic genes. Microorganisms. 2020;8:732.

Di Tommaso P, Moretti S, Xenarios I, Orobitg M, Montanyola A, Chang J-M, et al. T-coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. Nucleic Acids Res. 2011;39:W13–7.

Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32:1792–7.

Fani R, Brilli M, Liò P. The origin and evolution of operons: the piecewise building of the proteobacterial histidine operon. J Mol Evol. 2005;60:378–90.

Fani R, Liò P, Lazcano A. Molecular evolution of the histidine biosynthetic pathway. J Mol Evol. 1995;41:760–74.

Fani R, Mori E, Tamburini E, Lazcano A. Evolution of the structure and chromosomal distribution of histidine biosynthetic genes. Origins Life Evol Biospheres. 1998;28:555–70.

Garcia AK, Kaçar B. How to resurrect ancestral proteins as proxies for ancient biogeochemistry. Free Radic Biol Med. 2019;140:260–9.

Gaucher EA. Ancestral sequence reconstruction as a tool to understand natural history and guide synthetic biology: realizing and extending the vision of Zuckerkandl and Pauling. Ancestral sequence reconstruction. Oxford: Oxford University Press; 2007.

Ghodge SV, Fedorov AA, Fedorov EV, Hillerich B, Seidel R, Almo SC, et al. Structural and mechanistic characterization of L-histidinol phosphate phosphatase from the polymerase and histidinol phosphatase family of proteins. Biochemistry. 2013;52:1101–12.

Glasner ME, Truong DP, Morse BC. How enzyme promiscuity and horizontal gene transfer contribute to metabolic innovation. FEBS J. 2020;287:1323–42.

Goloboff PA, Szumik CA. Identifying unstable taxa: efficient implementation of triplet-based measures of stability, and comparison with Phyutility and RogueNaRok. Mol Phylogenet Evol. 2015;88:93–104.

Hanson-Smith V, Kolaczkowski B, Thornton JW. Robustness of ancestral sequence reconstruction to phylogenetic uncertainty. Mol Biol Evol. 2010;27:1988–99.

Huang H, Pandya C, Liu C, al-Obaidi NF, Wang M, Zheng L, et al. Panoramic view of a superfamily of phosphatases through substrate profiling. Proc Natl Acad Sci U S A. 2015;112:E1974–83.

Huelsenbeck JP, Ronquist F. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics. 2001;17:754–5.

Jacquier H, Birgy A, le Nagard H, Mechulam Y, Schmitt E, Glodt J, et al. Capturing the mutational landscape of the beta-lactamase TEM-1. Proc Natl Acad Sci U S A. 2013;110:13067–72.

Jha B, Kumar D, Sharma A, Dwivedy A, Singh R, Biswal BK. Identification and structural characterization of a histidinol phosphate phosphatase from *Mycobacterium tuberculosis*. J Biol Chem. 2018;293:10102–18.

Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596:583–9.

Kanehisa M, Goto S. KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000;28:27–30.

Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 2002;30:3059–66.

Kneidinger B, Graninger M, Puchberger M, Kosma P, Messner P. Biosynthesis of nucleotide-activated D-glycero-D-manno-heptose. J Biol Chem. 2001;276:20935–44.

Kneidinger B, Marolda C, Graninger M, Zamyatina A, McArthur F, Kosma P, et al. Biosynthesis pathway of ADP-L-glycero-beta-D-manno-heptose in *Escherichia coli*. J Bacteriol. 2002;184:363–9.

Kulis-Horn RK, Rückert C, Kalinowski J, Persicke M. Sequence-based identification of inositol monophosphatase-like histidinol-phosphate phosphatases (HisN) in *Corynebacterium glutamicum*, Actinobacteria, and beyond. BMC Microbiol. 2017;17:161.

Kuznetsova E, Proudfoot M, Gonzalez CF, Brown G, Omelchenko MV, Borozan I, et al. Genome-wide analysis of substrate specificities of the *Escherichia coli* haloacid dehalogenase-like phosphatase family. J Biol Chem. 2006;281:36149–61.

Liu Y, Schmidt B. Multiple protein sequence alignment with MSA-Probs. Methods Mol Biol (N. Y., NY, U. S.). 2014;1079:211–8.

Liu Y, Schmidt B, Maskell DL. MSAProbs: multiple sequence alignment based on pair hidden Markov models and partition function posterior probabilities. Bioinformatics. 2010;26:1958–64.

Mormann S, Lömker A, Rückert C, Gaigalat L, Tauch A, Pühler A, et al. Random mutagenesis in *Corynebacterium glutamicum* ATCC 13032 using an IS6100-based transposon vector identified the last unknown gene in the histidine biosynthesis pathway. BMC Genomics. 2006;7:205.

Nguyen HH, Wang L, Huang H, Peisach E, Dunaway-Mariano D, Allen KN. Structural determinants of substrate recognition in the HAD superfamily member D-glycero-D-manno-heptose-1,-7-bisphosphate phosphatase (GmhB). Biochemistry. 2010;49:1082–92.

Nourbakhsh A, Collakova E, Gillaspy GE. Characterization of the inositol monophosphatase gene family in *Arabidopsis*. Front Plant Sci. 2014;5:725.

Ochman H, Lawrence JG, Groisman EA. Lateral gene transfer and the nature of bacterial innovation. Nature. 2000;405:299–304.

Pandya C, Farelli JD, Dunaway-Mariano D, Allen KN. Enzyme promiscuity: engine of evolutionary innovation. J Biol Chem. 2014;289:30229–36.

Peracchi A. The limits of enzyme specificity and the evolution of metabolism. Trends Biochem Sci. 2018;43:984–96.

Rangarajan ES, Proteau A, Wagner J, Hung M-N, Matte A, Cygler M. Structural snapshots of *Escherichia coli* histidinol phosphate phosphatase along the reaction pathway. J Biol Chem. 2006;281:37930–41.

Rohweder B, Semmelmann F, Endres C, Sterner R. Standardized cloning vectors for protein production and generation of large gene libraries in *Escherichia coli*. Biotechniques. 2018;64:24–6.

Roshan U, Livesay DR. Probalign: multiple sequence alignment using partition function posterior probabilities. Bioinformatics. 2006;22:2715–21.

Ruszkowski M, Dauter Z. Structural studies of *Medicago truncatula* histidinol phosphate phosphatase from inositol monophosphatase superfamily reveal details of penultimate step of histidine biosynthesis in plants. J Biol Chem. 2016;291:9960–73.

Sander C, Schneider R. Database of homology-derived protein structures and the structural meaning of sequence alignment. Proteins. 1991;9:56–68.

Schneider TD, Stephens RM. Sequence logos: a new way to display consensus sequences. Nucleic Acids Res. 1990;18:6097–100.

Straub K, Linde M, Kropp C, Blanquart S, Babinger P, Merkl R. Sequence selection by FitSS4ASR alleviates ancestral sequence reconstruction as exemplified for geranylgeranylglyceryl phosphate synthase. Biol Chem. 2019;400:367–81.

Suárez ASG, Stefan A, Lemma S, Conte E, Hochkoeppler A. Continuous enzyme-coupled assay of phosphate- or pyrophosphate-releasing enzymes. Biotechniques. 2012;53:99–103.

Tamura K, Stecher G, Kumar S. MEGA11: molecular evolutionary genetics analysis version 11. Mol Biol Evol. 2021;38:3022–7.

Taylor PL, Sugiman-Marangos S, Zhang K, Valvano MA, Wright GD, Junop MS. Structural and kinetic characterization of the LPS biosynthetic enzyme D-alpha, beta-D-heptose-1,-7-bisphosphate phosphatase (GmhB) from *Escherichia coli*. Biochemistry. 2010;49:1033–41.

Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 1994;22:4673–80.

Thornton JW. Resurrecting ancient genes: experimental analysis of extinct molecules. Nat Rev Genet. 2004;5:366–75.

Valvano MA, Messner P, Kosma P. Novel pathways for biosynthesis of nucleotide-activated glycero-manno-heptose precursors of bacterial glycoproteins and cell surface polysaccharides. Microbiology (Reading, Engl.). 2002;148:1979–89.

Waite DW, Chuvochina M, Pelikan C, Parks DH, Yilmaz P, Wagner M, et al. Proposal to reclassify the proteobacterial classes deltaproteobacteria and Oligoflexia, and the phylum Thermodesulfobacteria into four phyla reflecting major functional capabilities. Int J Syst Evol Microbiol. 2020;70:5972–6016.

Wang L, Huang H, Nguyen HH, Allen KN, Mariano PS, Dunaway-Mariano D. Divergence of biochemical function in the HAD superfamily: D-glycero-D-manno-heptose-1,7-bisphosphate phosphatase (GmhB). Biochemistry. 2010;49:1072–81.

Wilkins MR, Gasteiger E, Bairoch A, Sanchez JC, Williams KL, Appel RD, et al. Protein identification and analysis tools in the ExPASy server. Methods Mol Biol (N. Y.). 1999;112:531–52.

Winkler ME, Ramos-Montañez S. Biosynthesis of histidine. EcoSal Plus 3.6.1.9. 2009.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Kinateder T, Drexler L, Straub K, Merkl R, Sterner R. Experimental and computational analysis of the ancestry of an evolutionary young enzyme from histidine biosynthesis. Protein Science. 2023;32(1):e4536. https://doi.org/10.1002/pro.4536