Structural Biology of the Histidine Biosynthetic Pathway in Plants

Wojciech Witek

Department of Structural Biology of Eukaryotes

Institute of Bioorganic Chemistry

Polish Academy of Sciences

Doctoral Dissertation

Poznań, 2024

READER'S GUIDE

In this article-based dissertation, most figures are contained within the attached publications located at the end of this work. To facilitate navigation, the digital version of this dissertation includes **interactive references to figures** located in the attachments, each of which has a "return" button. Other figures and tables are available within the main text.

Please, look at the example below, in which the blue arrow indicates the reference to the "Figure 3" in the first attached publication, and the orange arrow indicates a return button to the main text.
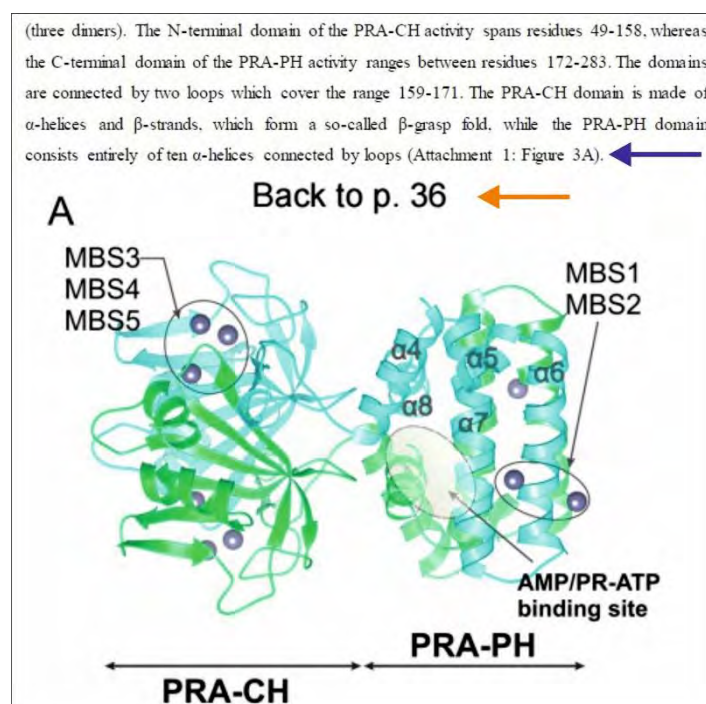
In the case of a few mutants discussed in this work, the terms "Mut1, Mut2" etc. are used, however, the printed version of this dissertation contains an extra page, on which the abbreviations for each mutant are explained for the review purposes. All other abbreviations used in this text are explained at the first use.

Readers who are interested in replicating the enzymatic synthesis, are encouraged to contact Prof. Miłosz Ruszkowski, as all plasmids required for the synthesis are available upon request.

Contents are organized based on the occurrence of enzymes in the pathway.

TABLE OF CONTENTS

1. Publications, conferences, and workshops

Publications (**Ps**) contributing to this dissertation:

P1: Witek W., Śliwiak J., Ruszkowski M. (**2021**) *Structural and mechanistic insights into the bifunctional HISN2 enzyme catalyzing the second and third steps of histidine biosynthesis in plants,* Scientific Reports 11(1):9647

https://doi.org/10.1038/s41598-021-88920-2

P2: Witek W., Imiołczyk B., Ruszkowski M. (**2024**) *Structural, kinetic, and evolutionary peculiarities of HISN3, a plant 5'-ProFAR isomerase,* Plant Physiology and Biochemistry 215:109065

https://doi.org/10.1016/j.plaphy.2024.109065

P3: Witek W., Śliwiak J., Rawski M., Ruszkowski M. (**2024**) *Targeting imidazole-glycerol phosphate dehydratase in plants: novel approach for structural and functional studies, and inhibitor blueprinting,* Frontiers in Plant Science 15;15:1343980

https://doi.org/10.3389/fpls.2024.1343980

P4: Rutkiewicz M., Nogues I., **Witek W.**, Angelaccio S., Contestabile R., Ruszkowski M. (**2023**) *Insights into the substrate specificity, structure, and dynamics of plant histidinol-phosphate aminotransferase (HISN6),* Plant Physiology and Biochemistry 196:759-773

https://doi.org/10.1016/j.plaphy.2023.02.017

Publications not involved in this dissertation:

P5: Marciniak P., **Witek W.**, Szymczak M., Pacholska-Bogalska J., Chowański S., Kuczer M., Rosiński G. (**2020***) FMRFamide-Related Peptides Signaling Is Involved in the Regulation of Muscle Contractions in Two Tenebrionid Beetles,* Frontiers in Physiology 12;11:456

https://doi.org/10.3389/fphys.2020.00456

P6: Witek W., Ragin W., Tran H. L., Połomska E., Podlewski M., Pawłowicz A., Cioch-Biniaś A., Woś A., Andrałojć W., Szachniuk M., Ruszkowski M. (**2024**) *Postępy biologii strukturalnej – jak zobaczyć cząsteczki życia?,* Postępy Biochemii 70 (2)

https://doi.org/10.18388/pb.2021_533

Conferences relevant to this project:

1. 23rd Heart of Europe Bio-Crystallography Meeting (HEC23)
    17.09.2021, Vierzehnheiligen, Germany
    Talk: *Structural and mechanistic insights into the bifunctional HISN2 enzyme catalyzing the second and third steps of histidine biosynthesis in plants*

2. EMBL Conference: Bringing Molecular Structures to Life: 50 Years of the PDB
    20-22.10.2021, Heidelberg, Germany
    Poster: *MtHISN2: A Bifunctional Enzyme in the Plant Histidine Biosynthetic Pathway*

3. Gordon Research Seminar: Enzymes, Coenzymes and Metabolic Pathways
    15-16.07.2023, Waterville Valley, New Hampshire, United States
    Talk: *Histidine Biosynthesis in Plants*

4. Gordon Research Conference: Enzymes, Coenzymes and Metabolic Pathways
    16-21.07.2023, Waterville Valley, New Hampshire, United States
    Poster: *Histidine Biosynthesis in Plants*

List of workshops:

1. Second South-East Asian Crystallographic Overview And Systematic Training (SEA COAST) 2020
    21-29.01.2020, King Mongkut's University of Technology Thonburi, Bangkok, Thailand

2. Autumn Training School 2021 and 2023 on Chemical Biology with a Focus on Screening Technologies, Assay Development, Medicinal Chemistry, Fragment Screening and Chemoproteomics
    15-19.11.2021
    20-22.11.2023
    EU-OPENSCREEN ERIC, Online

3. Virtual Annual CCP4/APS Crystallographic School 2022
    13-24.06.2022
    Advanced Photon Source, Argonne National Laboratory, Chicago, Illinois, USA

2. Protein structures

The protein crystal structures are available in the Protein Data Bank (PDB) and corresponding raw X-ray diffraction datasets can be accessed in the Macromolecular Xtallography Raw Data Repository (MX-RDR). The cryogenic electron microscopy (cryoEM) structures are also available in the Electron Microscopy Data Bank (EMDB).

1. **Witek W.**, Ruszkowski M. (**2021**) *Crystal structure of MtHISN2, a bifunctional enzyme from the histidine biosynthetic pathway*
   PDB: 7BGM
   MX RDR: https://doi.org/10.18150/WRT4WT

2. **Witek W.**, Ruszkowski M. (**2021**) *Crystal structure of MtHISN2-AMP complex, a bifunctional enzyme from the histidine biosynthetic pathway*
   PDB: 7BGN
   MX-RDR: https://doi.org/10.18150/ELDWZ6

3. Ruszkowski M., **Witek W.** (**2021**) *Cryo-EM structure of Medicago truncatula HISN5 protein*
   PDB: 7OJ5
   EMDB: EMD-12938

4. Rutkiewicz M., **Witek W.**, Ruszkowski M. (**2023**) *Crystal structure of Medicago truncatula histidinol-phosphate aminotransferase (HISN6) in apo form*
   PDB: 8BJ4
   MX-RDR: https://doi.org/10.18150/ZEPS7J

5. **Witek W.**, Ruszkowski M. (**2024**) *Medicago truncatula HISN5 (IGPD) in complex with MN and IG2*
   PDB: 8QAV
   EMDB: EMD-18305

6. **Witek W.**, Ruszkowski M. (**2024**) *Medicago truncatula HISN5 (IGPD) in complex with MN, IMD, EDO, FMT, GOL and TRS*
   PDB: 8QAW
   MX-RDR: https://doi.org/10.18150/INUP4Q

7. **Witek W.**, Ruszkowski M. (**2024**) *Medicago truncatula HISN5 (IGPD) in complex with MN, FMT, GOL and TRS*
   PDB: 8QAX
   MX-RDR: https://doi.org/10.18150/XWLKP6

8. **Witek W.**, Ruszkowski M. (**2024**) *Medicago truncatula HISN5 (IGPD) in complex with MN, FMT, ACT, CIT, EDO, SO4*

    PDB: 8QAY

    MX-RDR: https://doi.org/10.18150/FN6VJX

9. **Witek W.**, Imiołczyk B., Ruszkowski M. (**2024**) *Medicago truncatula 5'-ProFAR isomerase (HISN3) D57N mutant in complex with ProFAR*

    PDB: 9FCF

    MX-RDR: https://doi.org/10.18150/DGZKW3

10. **Witek W.**, Imiołczyk B., Ruszkowski M. (**2024**) *Medicago truncatula 5'-ProFAR isomerase (HISN3) D57N mutant in complex with PrFAR*

    PDB: 9FCG

    MX-RDR: https://doi.org/10.18150/LDLSBT

3. Abstract

In this doctoral dissertation, I present my results from the studies of the HISN2, HISN3, HISN5, and HISN6 enzymes catalyzing the histidine biosynthetic pathway (HBP) in plants. The studies were conducted using X-ray crystallography and cryogenic electron microscopy (cryoEM), enriched by bioinformatic and phylogenetic analyses, and enzymatic characterization. The main goals of this research were to (i) determine three-dimensional structures of the HBP enzymes, (ii) characterize their enzymatic activity, (iii) develop protocols for an *in vitro* enzymatic synthesis of substrates, (iv) design scaffolds for future inhibitors, and (v) understand the phylogenetic origin of the enzymes. The structures for HISN2, HISN3, and HISN6 are the first structures of those enzymes from plants.

The histidine biosynthetic pathway in plants comprises eleven reactions catalyzed by eight enzymes, named HISN1-8, regarding their consecutive action. The pathway is an interesting object for studying plant metabolism, due to its interconnection with purine *de novo* biosynthesis and nitrogen metabolism. The alarming need for sustainable sources of food products and soil safety became the foundation for this project. The understanding of the structural peculiarities of enzymes involved in histidine biosynthesis in plants enables development of small molecules, acting as inhibitors or activators, to modulate the efficiency of histidine production. Histidine, due to the properties of its imidazole-containing side chain, is able to chelate metal cations from soil and water, contributing to the process of phytoremediation. This can increase the safety of soil and related food products, which should be free of metal contamination that causes, for instance, prevalent nickel allergies. The food safety might also be assured by the mitigation of the herbicide resistance, which impairs the yield of food production. The approaches presented in this dissertation broad our understanding of the HBP in plants, because such studies in eukaryotes have been neglected for decades, although the histidine biosynthesis was quite well understood in prokaryotes.

Building upon the research conducted by my supervisor, Prof. M. Ruszkowski, who characterized structurally and functionally the HISN1, HISN7, and HISN8 enzymes, I was able to study the structures, activity, and evolution of the HISN2, HISN3, HISN5, and HISN6 enzymes. As a source of the coding sequences for those enzymes, I chose a genome-sequenced barrel medic, *Medicago truncatula* (*Mt*), which is a model for legumes that are environmentally and economically important plants.

The structural studies provided insights into the interactions of *Mt*HISN2 with AMP, which allowed to update the catalytic mechanism. AMP turned out to be an effective inhibitor

of *Mt*HISN2 in physiologically relevant concentrations, therefore, suggesting the existence of a second-tier regulatory mechanism of the pathway flux. The crystal structures of *Mt*HISN3 with its substrate and product contributed to the understanding of differences between plant and bacterial homologs, which may account for the development of kingdom-specific herbicides. Molecular dynamics (MD) simulations of a plant-specific fragment suggested that it facilitates products release, thereby contributing to a high catalytic efficiency. A set of high-resolution crystal and cryoEM structures of *Mt*HISN5 was utilized to identify ligand-binding hot-spots. These results combined with the results of virtual screening (VS) campaigns, served for the proposition of candidate molecules and linkers for the development of future herbicides. The process of a stereospecific enzymatic synthesis of *Mt*HISN5's substrate, resulted in the novel protocol for studying the activity of those enzymes. The crystal structures of *Mt*HISN6 revealed changes of its dynamics, based on its interactions with ligands. The kinetic studies revealed *Mt*HISN6's high selectivity towards the substrate, compared to its bacterial homologs. Structural differences between these homologs stimulated VS campaigns in the regions of the highest possibility for development of the kingdom-specific herbicides.

The phylogenetic studies of the HBP enzymes conducted using sequence similarity networks (SSN) and phylogenetic trees, revealed interesting results about their origin. In this group of enzymes, only HISN5 seems to origin from Cyanobacteria, which is consistent with the endosymbiotic theory. The genes encoding other enzymes, i.e., HISN2, HISN3, and HISN6, were likely acquired early in the evolution through a horizontal gene transfer from the Myxococcota, Bacillota, and Chloroflexota, respectively.

4. Streszczenie

W niniejszej rozprawie doktorskiej prezentuję wyniki moich badań enzymów HISN2, HISN3, HISN5 i HISN6, które katalizują szlak biosyntezy histydyny (SBH) u roślin. Głównymi metodami wykorzystanymi w badaniach były krystalografia rentgenowska i kriomikroskopia elektronowa (cryoEM). Wyniki te były wzbogacone o analizy bioinformatyczne i filogenetyczne, oraz charakteryzację aktywności enzymatycznej. Głównymi celami rozprawy były (i) rozwiązanie trójwymiarowych struktur enzymów SBH, (ii) charakteryzacja ich aktywności enzymatycznej, (iii) opracowanie protokołów do enzymatycznej syntezy *in vitro* substratów, (iv) zaprojektowanie szkieletów molekularnych dla przyszłych inhibitorów, oraz (v) poznanie filogenezy tych enzymów. Struktury HISN2, HISN3 i HISN6 są pierwszymi roślinnymi strukturami tych enzymów.

SBH u roślin składa się z jedenastu reakcji, katalizowanych przez osiem enzymów, nazwanych HISN1-8, zgodnie z kolejnością ich aktywności w szlaku. Szlak ten jest interesującym obiektem badawczym ze względu na jego połączenie z biosyntezą *de novo* puryn oraz metabolizmem azotu. Alarmująca potrzeba posiadania zrównoważonych źródeł pożywienia oraz bezpieczeństwo gleb, stały się fundamentami tego projektu. Zrozumienie właściwości struktur enzymów biorących udział w SBH u roślin umożliwia rozwój małych cząsteczek, które zachowując się jak inhibitory lub aktywatory, mogą modulować wydajność produkcji histydyny. Histydyna, dzięki właściwościom swojego łańcucha bocznego zawierającego pierścień imidazolowy, jest zdolna do chelatacji kationów metali z gleb i wód, umożliwiając proces fitoremediacji. To może potencjalnie zapewnić bezpieczeństwo gleb i żywności, które powinny być wolne od skażenia metalami, powodującymi np. częste alergie na nikiel. Bezpieczeństwo produktów spożywczych może być zapewnione poprzez zmniejszenie herbicydooporności chwastów, która osłabia wydajność produkcji żywności. Podejścia zastosowane w niniejszej rozprawie poszerzają nasze zrozumienie SBH u roślin, ponieważ tego typu badania u eukariontów były zaniedbywane przez dekady, mimo, że biosynteza histydyny została dobrze poznana u prokariontów.

Opierając się na badaniach prowadzonych przez mojego promotora, Prof. Miłosza Ruszkowskiego, który scharakteryzował strukturalnie i funkcjonalnie enzymy HISN1, HISN7 i HISN8, mogłem badać struktury, aktywność i ewolucję enzymów HISN2, HISN3, HISN5 i HISN6. Jako źródło sekwencji kodujących te enzymy wybrałem modelową roślinę strączkową z rodzaju lucerna o zsekwencjonowanym genomie, *Medicago truncatula* (*Mt*), która jest ważnym środowiskowo i ekonomicznie gatunkiem.

Badania strukturalne zapewniły wgląd w interakcje pomiędzy *Mt*HISN2, a AMP, co pozwoliło na zaktualizowanie mechanizmu reakcji. AMP okazało się być efektywnym inhibitorem *Mt*HISN2, w zakresie stężeń fizjologicznych, sugerując istnienie drugorzędowego mechanizmu regulującego przepływ szlaku. Struktury krystaliczne *Mt*HISN3 z substratem i produktem przyczyniły się do zrozumienia różnic pomiędzy roślinnymi i bakteryjnymi homologami, co umożliwia rozwój herbicydów specyficznych na poziomie królestwa. Symulacje dynamiki molekularnej fragmentu specyficznego dla roślin sugerowały jego udział w uwalnianiu produktu, co przyczynia się do wysokiej wydajności katalitycznej. Zestaw wysokorozdzielczych struktur krystalicznych i cryoEM enzymu *Mt*HISN5 został wykorzystany do identyfikacji miejsc wiązania ligandów. Wyniki te, połączone z wynikami wirtualnego skriningu posłużyły do zaproponowania nowych cząsteczek-kandydatów oraz łączników dla przyszłych herbicydów. Proces stereospecyficznej syntezy enzymatycznej substratu *Mt*HISN5 poskutkował opracowaniem nowego protokołu do badania aktywności tego enzymu. Struktury krystaliczne *Mt*HISN6 ujawniły zmiany jego dynamiki, na podstawie interakcji z ligandami. Badania kinetyczne wykazały wysoką selektywność względem substratu, w porównaniu z jego homologami bakteryjnymi. Różnice strukturalne pomiędzy tymi homologami zainspirowały przeprowadzenie kampanii wirtualnego skriningu w regionach o najwyższym prawdopodobieństwie rozwoju herbicydów specyficznych na poziomie królestwa.

Analizy filogenetyczne przeprowadzone z użyciem sieci podobieństwa sekwencyjnego oraz drzew filogenetycznych enzymów SBH wygenerowały interesujące wyniki na temat ich pochodzenia. W tej grupie enzymów, jedynie HISN5 pochodzi od cyjanobakterii, co jest wynikiem zgodnym z teorią endosymbiozy. Geny kodujące pozostałe enzymy, czyli HISN2, HISN3 i HISN6, były najprawdopodobniej pozyskane wcześnie w toku ewolucji, na drodze horyzontalnego transferu genów odpowiednio od Myxococcota, Bacillota i Chloroflexota.

5. Introduction

5.1 Background

L-Histidine (hereafter histidine) was discovered independently by A. Kossel and S. G. Hedin in 1896, with both papers published in the same volume of the *Journal of Physiological Chemistry* (ger.: *Zeitschrift für physiologische Chemie*) (Hedin 1897, Kossel 1897). Histidine is unique among all proteinogenic amino acids (AAs) because of its peculiar biosynthesis and extraordinary properties of a side chain. In general, AAs are biosynthesized from carbohydrate-derived precursors, but portions of histidine's backbone originate from adenosine-5′-triphosphate (ATP). The histidine's side chain is constituted of an imidazole ring, which is already present as a part of the adenine moiety in ATP. However, during the course of histidine biosynthesis, the imidazole ring of the adenine moiety enters purine biosynthesis, and the histidine's imidazole-containing side chain is built *de novo* in subsequent steps (Kirschning 2022) (Figure 1). The imidazole-containing side chain is the reason why histidine plays a crucial role in multiple biological processes, not only as a building block for protein biosynthesis, but mostly as a catalytically active residue. On average, histidine content in proteins has been estimated between 2.0 and 2.5% (Carugo 2008). Interestingly, some proteins contain such a high proportion of histidine, that a whole family of the so-called "histidine-rich proteins" has been distinguished, with some member proteins containing even up to 34% of histidine (Wellems et al. 1986, Oppenheim et al. 1988). Besides the structural role of histidine in proteins, it also plays an important functional role in enzymes, being the most prevalent catalytic AA residue in six out seven Enzyme Commission (EC) classes. This is because when the analysis was conducted, the seventh class did not yet exist, as the EC 7 class was introduced in 2018 (Weber et al. 1981, Holliday et al. 2011, Tipton 2018). The functioning of histidine residues in the acid-base catalysis is possible due to the properties of the side chain's imidazole moiety, which can be protonated around neutral pH values (pKa = 6-7), and is able to coordinate divalent metal cations by either deprotonated Nδ or Nε. An interesting example is a group of proteins called the "histidine-triad protein superfamily" whose members possess the conserved His-φ-His-φ-His-φ-φ ("φ" representing any hydrophobic AA) motif, which binds nucleotides (Seraphin 1992, Brenner et al. 1999). The ability to bind metal cations by histidine is vital for the functioning of metalloproteins, which utilize metal ions as cofactors (Greiner et al. 2007).
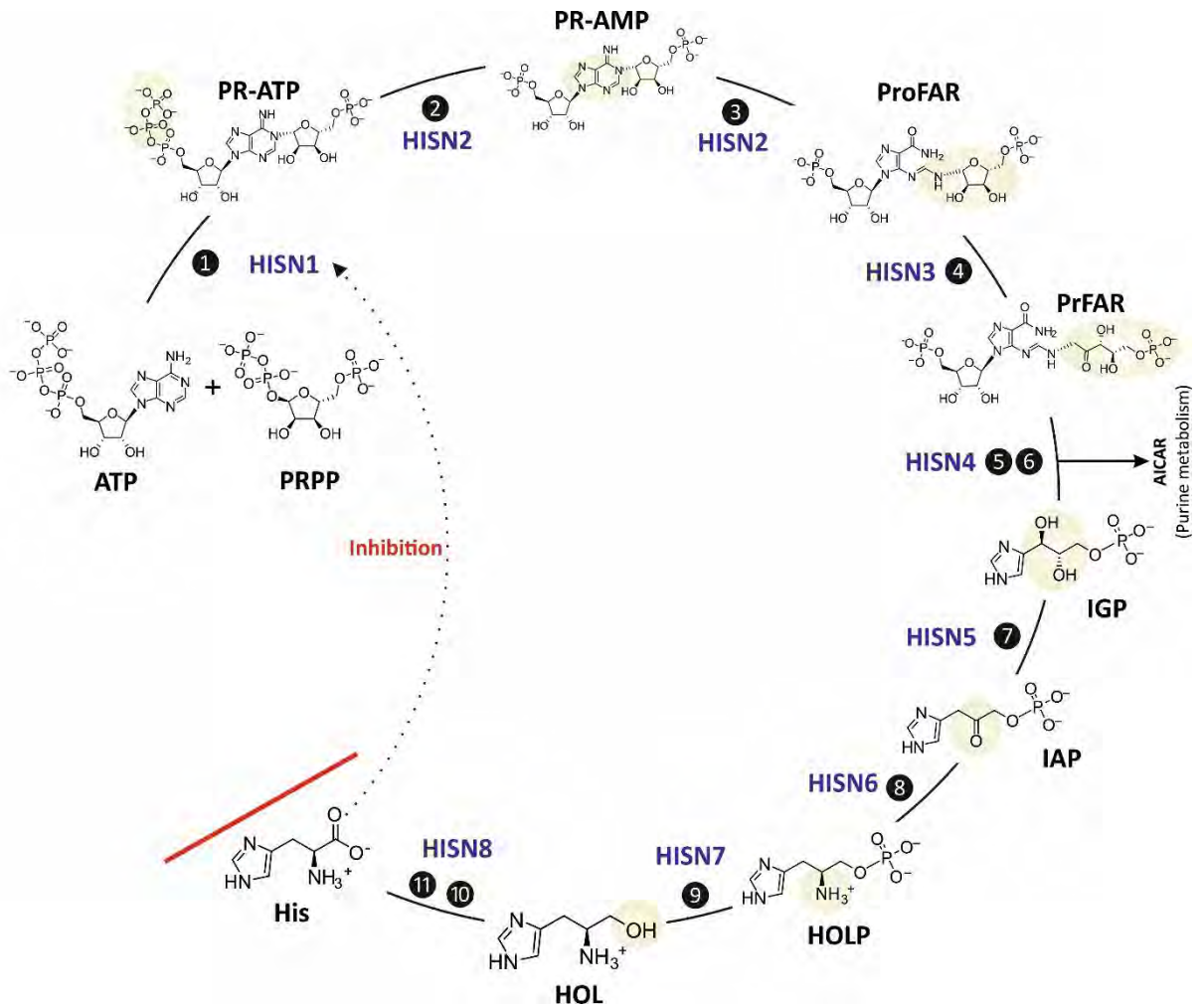
Figure 1. Schematic representation of reactions in histidine biosynthesis. Yellow circles highlight regions undergoing changes in catalysis. Enzyme names are in blue and reactions are in black. ATP – adenosine-5'-triphosphate, PRPP – 5'-phosphoribosyl-1-pyrophosphate, PR-ATP – N'-5'-phosphoribosyl-ATP, PR-AMP – N'-5'-phosphoribosyl-AMP, ProFAR – N'-[(5'-phosphoribosyl)-formimino]-5-aminoimidazole-4-carboxamide ribonucleotide, PrFAR – N'-[(5'-phosphoribulosyl)-formimino]-5-aminoimidazole-4-carboxamide ribonucleotide, AICAR – 5'-amino-4-carboxamide ribonucleotide, IGP - imidazole-glycerol phosphate, IAP – imidazole-acetol phosphate, HOLP – L-histidinol-phosphate, HOL – L-histidinol, His – L-histidine.

The histidine biosynthetic pathway (HBP) occurs in bacteria, archaea, fungi, and plants. The pathway has a similar course in all those groups, which reflects its ancient origin. In plants, the HBP is enabled by eight enzymes, designated HISN1-8, regarding their sequential occurrence (Table 1). The pathway encompasses eleven steps, or ten if the imidazole-glycerol phosphate synthase (HISN4) glutaminase activity is considered auxiliary. The difference between the number of enzymes and reactions is due to the bifunctionality of three enzymes in plants, i.e., HISN2, HISN4, and HISN8.

Table 1. Summary of homologous genes encoding enzymes involved in the HBP in different kingdoms. The genes are listed in italics.

Enzyme	Plants	Fungi	Prokaryotes
ATP-phosphoribosyltransferase (EC 2.4.2.17)	*HISN1*	*HIS1*	*HisG*
Phosphoribosyl-ATP-pyrophosphohydrolase (EC 3.6.1.31)	*HISN2*	*HIS4*	*HisE*
Phosphoribosyl-AMP-cyclohydrolase (EC 3.5.4.19)	*HISN2*	*HIS4*	*HisI*
N'-[(5'-phosphoribosyl)formimino]-5-aminoimidazole-4-carboxamide ribonucleotide isomerase (EC 5.3.1.16)	*HISN3*	*HIS6*	*HisA*
Imidazole-glycerol phosphate synthase (EC 4.3.2.10, EC 3.5.1.2)	*HISN4*	*HIS7*	*HisH, HisF*
Imidazole-glycerol phosphate dehydratase (EC 4.2.1.19)	*HISN5*	*HIS3*	*HisB*
L-Histidinol-phosphate aminotransferase (EC 2.6.1.9)	*HISN6*	*HIS5*	*HisC*
L-Histidinol-phosphate phosphatase (EC 3.1.3.15)	*HISN7*	*HIS2*	*HisB*
L-Histidinol dehydrogenase (EC 1.1.1.23)	*HISN8*	*HIS4*	*HisD*

Unlike the bacterial organization of *his* genes within an operon, the *HISN* genes in plants are encoded separately. For instance, in *Arabidopsis thaliana*, the HISN2, HISN3, HISN4, and HISN8 enzymes are encoded by single-copy genes, while the *HISN1, HISN5, HISN6,* and *HISN7* genes have undergone gene duplication (Muralla et al. 2007, Petersen et al. 2010). Apart from duplication, the *HISN* genes in plants have been shaped by gene fusion and gene elongation (Del Duca et al. 2021). These evolutionary processes have strongly contributed to the structural and functional complexity of the HBP enzymes enabling the pathway to efficiently support histidine biosynthesis across different organisms (Fani et al. 2007). In plants, all *HISN* genes belong to the nuclear genome, and in *A. thaliana* the *HISN2*, *HISN4*, and *HISN5B* genes share a pattern in which the first exon, encoding an N-terminal chloroplast transit peptide, is separated from the ones corresponding to the mature enzyme (Tada et al. 1994, Fujimori et al. 1998b, Fujimori et al. 1998a, Fujimori et al. 1998c, El Malki 2001, Ingle et al. 2005). The transit peptides direct the HBP enzymes into the chloroplast, where the pathway occurs. One of the reasons for the compartmentalization of the histidine biosynthesis in chloroplasts is the proximity to energy supplies. Histidine biosynthesis generates a high metabolic cost, requiring 31-41 ATP molecules per one histidine molecule. Chloroplasts

provide an efficient and readily available source of ATP via the photophosphorylation process to meet this demand (Brenner et al. 1971, Alifano et al. 1996, Akashi et al. 2002, Swire 2007).

The integration of metabolic processes is important for the tight regulation of the cellular homeostasis. The HBP is considered a metabolic crossroad for the above reason but also because it integrates the purine *de novo* biosynthesis with the nitrogen metabolism (Fani et al. 1995). The connection with purine biosynthesis is established at the step catalyzed by imidazole-glycerol phosphate synthase, where a byproduct, 5′-amino-4-carboxamide ribonucleotide (AICAR), is released and then utilized in the purine pathway. The details of this interconnection will be discussed further.

The histidine biosynthesis has been studied since the 1950s, mostly in bacterial model systems, *Escherichia coli* and *Salmonella typhimurium*. These foundational studies have provided significant insights into the enzyme purification methods and regulatory controls of the pathway (Miller et al. 1952, Adams 1954, Ames et al. 1955, Ames 1957b, Ames 1957a). In the following decades, B. N. Ames, M. Brenner, R. G. Martin, and others, continued investigating the bacterial *his* operon, identifying all enzymes, reagents, and byproducts involved in the HBP (Fink et al. 1967, Brenner et al. 1971, Martin 1971). The study of the HBP in plants began in the 1970s, primarily due to the lack of auxotrophic *HISN* mutants, and a complexed biochemistry behind the pathway (Wiater et al. 1971a, Wiater et al. 1971b, Wiater et al. 1971c, Miflin 1980, Fujimori et al. 1998a, Stepansky et al. 2006). These pioneering studies have successfully tackled purification problems, activity measurements, and enzymatic synthesis of substrates. The advent of novel molecular biology tools, genomics, and bioinformatics, eventually addressed the challenges of genomic data availability and protein production. As a result, the HBP was the last AA biosynthetic pathway to be genetically deciphered by the identification of the *HISN7* gene, encoding L-histidinol-phosphate phosphatase (HISN7) (Petersen et al. 2010). The availability of plant *HISN* gene sequences together with isolation and characterization of plant cDNA allowed for the progress in understanding of the HBP.

5.2 Course of the pathway

In plants, the HBP comprises eleven distinct reactions (Figure 1). The pathway begins with the condensation of ATP and 5′-phosphoribosyl-1-pyrophosphate (PRPP) to form N′-5′-phosphoribosyl-ATP (PR-ATP). This initial and crucial reaction is catalyzed by ATP-phosphoribosyltransferase (ATP-PRT or HisG in bacteria, and HISN1 in plants) setting the stage for subsequent steps in the biosynthesis of histidine (Figure 2). HISN1 requires Mg^{2+} ions for its activity and the reaction has to be coupled with inorganic pyrophosphatase (PPa), which directs the reaction towards the product formation by hydrolyzing the emerging inorganic pyrophosphate (PPi).
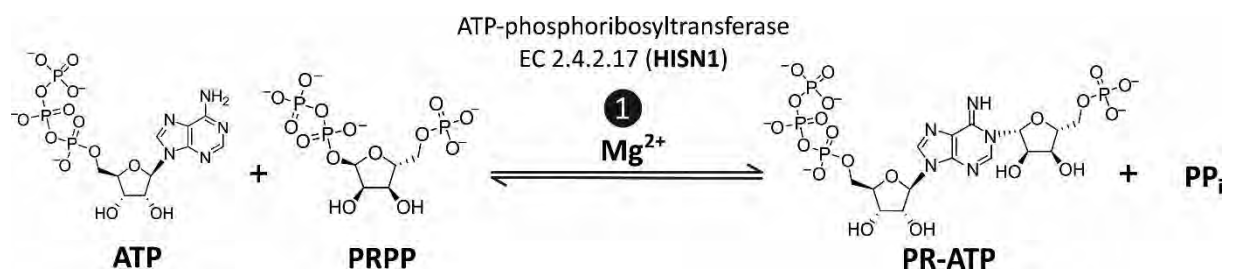
Figure 2. Schematic reaction catalyzed by ATP-phosphoribosyltransferase (HISN1), showing the formation of N′-5′-phosphoribosyl-ATP.

The second and third reactions are catalyzed by a bifunctional HISN2 enzyme in plants (HisIE when bifunctional in some prokaryotes), which displays phosphoribosyl-ATP pyrophosphohydrolase (HisE in prokaryotes, or PRA-PH in plants) and phosphoribosyl-AMP cyclohydrolase (HisI in prokaryotes, or PRA-CH in plants) activities, respectively in the course of the pathway (Figure 3). In details, PR-ATP pyrophosphate is cleaved by the C-terminal domain (PRA-PH) of HISN2, resulting in the formation of N′-5′-phosphoribosyl-AMP (PR-AMP). Next, the N-terminal domain (PRA-CH) of HISN2 hydrolyzes the purine ring, resulting in the formation of N′-[(5′-phosphoribosyl)-formimino]-5-aminoimidazole-4-carboxamide ribonucleotide (ProFAR).
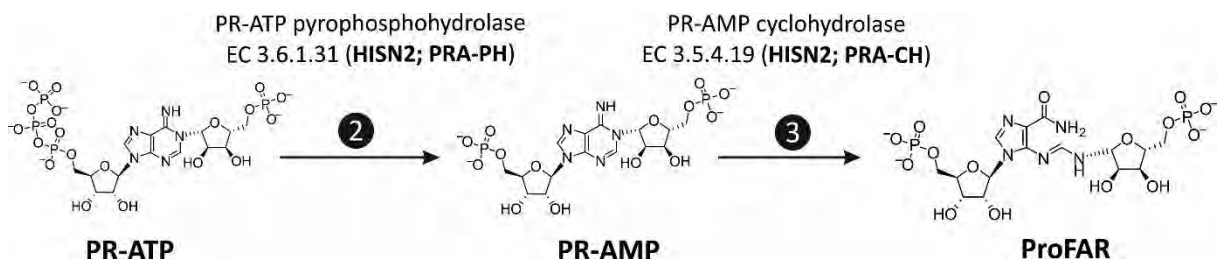
**Figure 3.** Schematic illustration of the reaction mediated by phosphoribosyl-ATP pyrophosphohydrolase (HISN2; PRA-PH), depicting the conversion of phosphoribosyl-ATP to phosphoribosyl-AMP, followed by the adenine ring hydrolysis catalyzed by phosphoribosyl-AMP cyclohydrolase (HISN2; PRA-CH), resulting in the formation of ProFAR.

The fourth reaction is catalyzed by N′-[(5′-phosphoribosyl)formimino]-5-aminoimidazole-4-carboxamide ribonucleotide isomerase (HisA in prokaryotes or HISN3 in plants) (Figure 4). In this reaction, the aminoaldose, N′-[(5′-phosphoribosyl)formimino]-5-aminoimidazole-4-carboxamide ribonucleotide (ProFAR) undergoes isomerization to form the aminoketose, N′-[(5′-phosphoribulosyl)-formimino]-5-aminoimidazole-4-carboxamide ribonucleotide (PrFAR).
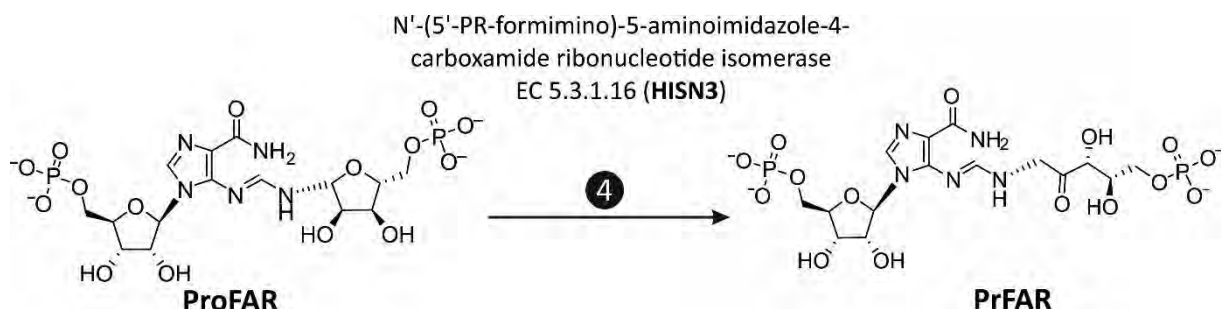
Figure 4. Diagram of the enzymatic isomerization catalyzed by N′-[(5′-phosphoribosyl)formimino]-5-aminoimidazole-4-carboxamide ribonucleotide isomerase (HISN3), illustrating the linearization of the product.

In the fifth and sixth reactions, PrFAR is converted in the presence of L-glutamine (Gln) by imidazole-glycerol phosphate synthase (HisHF in prokaryotes or HISN4 in plants) to imidazole-glycerol phosphate (IGP), AICAR, and L-glutamate (Glu) (Figure 5). After this reaction, AICAR enters *de novo* purine biosynthesis, while the imidazole ring of histidine is newly formed, with the resulting IGP being further processed in the HBP. The connection to the nitrogen metabolism is established by the glutamine molecule, which provides the final nitrogen atom for the imidazole ring in IGP (Fani et al. 2007).
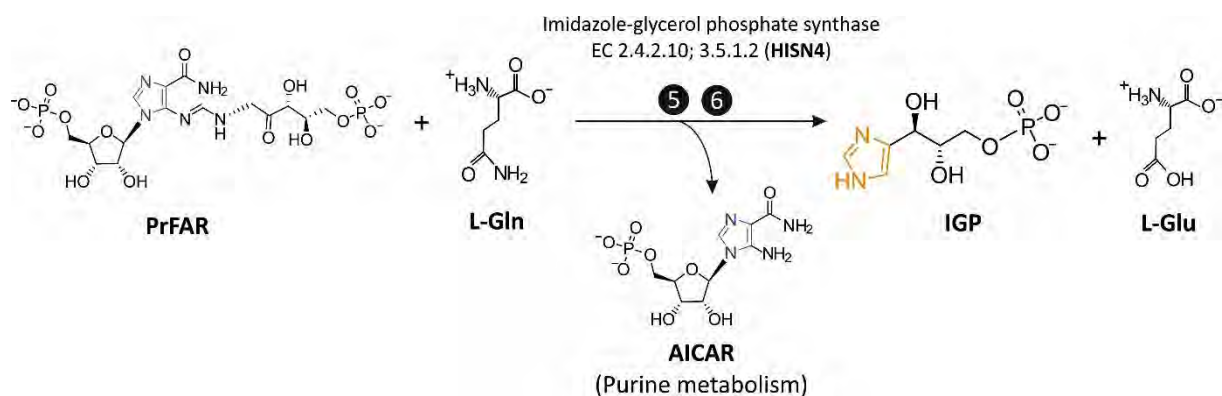
Figure 5. Visual representation of the reaction catalyzed by imidazole-glycerol phosphate synthase (HISN4), showing the re-synthesis of the imidazole moiety (orange) of imidazole glycerol phosphate, as the original imidazole ring (blue) enters purine metabolism in the form of AICAR.

In the seventh reaction, IGP is dehydrated by imidazole-glycerol phosphate dehydratase (HisB in prokaryotes or HISN5 in plants), resulting in the formation of imidazole-acetol phosphate (IAP) (Figure 6).
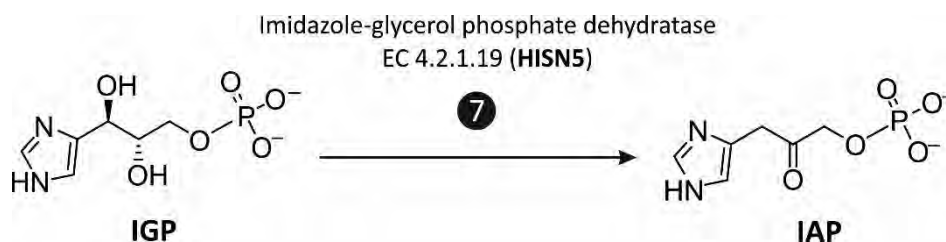
Figure 6. Schematic overview of the dehydration reaction catalyzed by imidazole-glycerol phosphate dehydratase (HISN5).

Next, is a reversible transamination reaction in which the amino group is transferred from Glu to IAP by L-histidinol-phosphate aminotransferase (HAT or HisC in prokaryotes and HISN6 in plants), producing L-histidinol-phosphate (HOLP) and α-ketoglutarate (2OG) (Figure 7).
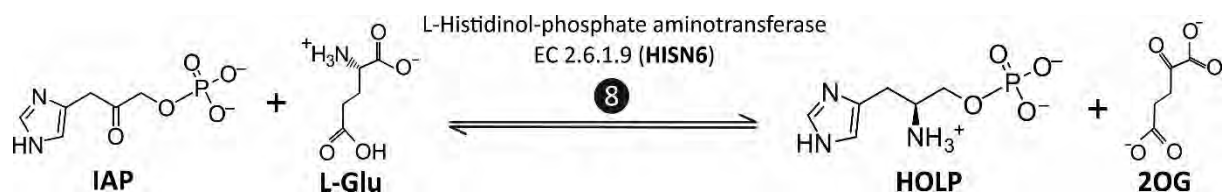
Figure 7. Schematic of the transamination reaction catalyzed by histidinol-phosphate aminotransferase (HISN6).

L-Histidinol-phosphate phosphatase (HisB in prokaryotes or HISN7 in plants) catalyzes the ninth reaction, where HOLP is dephosphorylated to L-histidinol (HOL) (Figure 8).
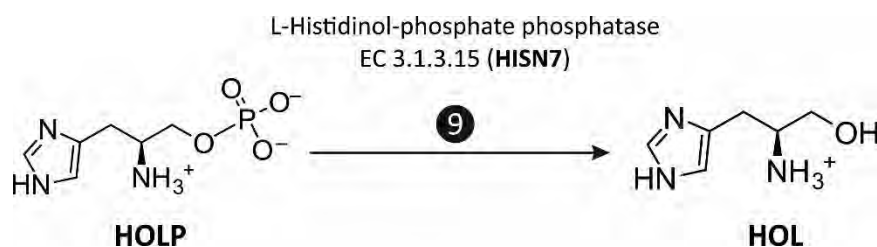
Figure 8. Schematic representation of the penultimate step in the HBP catalyzed by histidinol-phosphate phosphatase (HISN7).

The penultimate and ultimate reactions are the oxidation processes catalyzed by L-histidinol dehydrogenase (HisD in prokaryotes or HISN8 in plants): first, HOL is oxidized to L-histidinal (HAL, not shown), and then HAL is further oxidized to the final product, histidine (Figure 9).
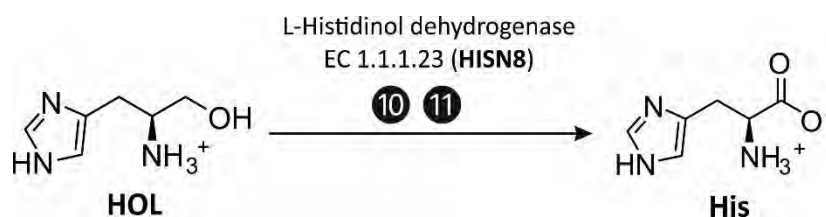
Figure 9. Overview of the double-oxidation reaction catalyzed by histidinol dehydrogenase (HISN8), illustrating the final step in histidine biosynthesis in plants.

5.3 Regulation of histidine biosynthesis

Histidine biosynthesis in plants is primarily regulated at the gene expression and enzymatic levels. Gene expression might be regulated by promoters, enhancers, silencers, and transcription factors, which are associated with specific genes. Although regulatory mechanisms have been reported to control the HBP in bacteria (Winkler 1987, Alifano et al. 1996, Kulis-Horn et al. 2014) and yeast (Ticebaldwin et al. 1989, Hinnebusch 1990, Rébora et al. 2005), relatively little is known about the details of parallel mechanisms in terms of the plant *HISN* genes. The availability of genomic data and auxotrophic mutants is key to understanding the regulation of gene expression in plants. Therefore, most studies refer to results obtained from the model plant *A. thaliana*. In the *A. thaliana* genome, the *HISN1, HISN5, HISN6,* and *HISN7* genes are duplicated, and the remaining genes are present in single copies (Muralla et al. 2007, Petersen et al. 2010). Knock-out mutations of those genes are lethal at the embryonic stage of development. The *HISN* genes are expressed in all plant tissues and throughout a whole development, but generally, the lowest expression levels of all eight genes have been detected in pollen (Muralla et al. 2007). The analysis conducted by R. Muralla et al., revealed several interesting aspects of the *HISN* genes expression in *A. thaliana* tissues: (i) high levels of *HISN1A* and *HISN8* transcripts occur throughout all tissues of the plant, (ii) *HISN1A* expression is much higher than *HISN1B* expression, especially in roots, (iii) *HISN3* shows the lowest level of expression out of all *HISN1-8* genes, (iv) *HISN4* has a moderate expression in seedling and rosette, but low in other tissues, (v) *HISN5A* and *HISN5B* are expressed at moderate levels throughout the plant. In the study conducted by L. Petersen et al., the second isoform of the *HISN7* gene was expressed in all plant tissues and throughout a whole plant development (Petersen et al. 2010). These results indicate the differential expression of the *HISN* genes in plants, which is tissue- and development-dependent. However, such abundance of the *HISN* transcripts in all tissues, especially in roots, seems counterintuitive due to the fact that the HBP occurs in chloroplasts, based on the fact that all eight enzymes contain N-terminal transit peptides. Although bioinformatic tools predict N-terminal sequences as chloroplast transit peptides, it is possible that *in vivo* they could be directing the HBP enzymes to other types of plastids, including those located in non-photosynthetic tissues. That would explain the presence of the *HISN* transcripts throughout a plant. Nonetheless, the last step on the way from transcripts to mature and active HISN enzymes, is the proteolytic processing of the N-terminal transit peptides by stromal and thylakoid peptidases during the protein import (Celedon et al. 2013, Zhu et al. 2022).

The detailed regulation mechanisms of the *HISN* genes transcription in plants remain unknown, however, there are clues connecting transcriptional regulation of the histidine pathway with other AA biosynthetic pathways. A possibility for a cross-pathway regulation has been described in tobacco, *Nicotiana plumbaginifolia,* in which the inhibition of the HBP led to changes in the expression levels of the *HISN* genes (El Malki 2001). A histidine auxotrophic cell line transferred to a histidine-free medium resulted in the induction of *HISN6* expression. Surprisingly, histidine starvation led also to the increase of the dihydrodipicolinate synthase (*dhdps*) transcript. The enzyme encoded by *dhdps* catalyzes the first step of lysine biosynthesis; therefore, the amount of free lysine has also increased. Similarly, inhibition of the HISN5 enzyme in *A. thaliana* by an experimental herbicide, IRL 1803 (3-hydroxy-3-[2H-(1,2,4)triazole-3-yl]-cyclohexyl-phosphonic acid), has increased the expression levels of eight genes involved in the biosynthesis of aromatic AA, histidine, lysine, and purines, as well as decreased the expression of glutamine synthetase (Guyer et al. 1995).

To date, two viable weak alleles of the HBP genes have been identified in *A. thaliana*. The *albino and pale green 10* (*apg10*) mutant carries a V256L mutation in HISN3 and exhibits a pale green phenotype seedlings. In contrast to *HISN3* knock-out mutants, which are embryo-lethal, the *apg10* mutants gradually recover upon histidine supplementation and resemble wild-type (WT) in reproductive tissues. Interestingly, this mutation caused an increased level of AA content, including histidine (Noutoshi 2005). The second example is the *histidinol phosphate aminotransferase 1* (*hpa1*) mutant with A69T substitution in the enzyme encoded by *HISN6A*, which caused the impaired root development at the seedling stage (Mo et al. 2006). Both *apg10* and *hpa1* mutants have different effects on histidine levels in plants. The *apg10* produces similar amount of histidine compared to WT, but causes increased level of all other AAs (Noutoshi 2005). The *hpa1* mutant displays a 30% reduced amount of histidine, but lowers levels of free Asp, Arg, Glu, and Lys (Mo et al. 2006).

A novel His-deficient *A. thaliana* mutant was characterized this year. The *suppressor of dominant-negative LRX1ΔE14* (*sune82*) mutant carries a weak allele of *HISN2* and exhibits impaired development at the seedling and adult stages, with a dwarf phenotype, reduced primary root length, enhanced root hair development, and weaken fertilization. The histidine content in the *sune82* mutant decreased of 60% (Guerin A 2024). The decreased levels of histidine in *sune82* also influenced the Target of Rapamycin (TOR) signaling pathway. The TOR network in plants senses the availability of growth regulators, nutrients, and amino acid levels, which influence the cell wall formation and expansion. Several studies indicated that the decreased levels of free histidine cause the TOR pathway up-regulation in root hairs (Miguel et

al. 2015, Morris et al. 2017), but in other tissues, the TOR pathway is down-regulated (Heinemann et al. 2021). These results suggest the existence of a distinct "meta-mechanism" coordinating the transcriptional regulation of the majority of AA biosynthetic pathways (possibly involving signaling pathways too) in plants, in response to a disrupted homeostasis. However, such a mechanism has not been discovered yet.

The metabolic control of the HBP flux at the enzymatic level resembles that of other biosynthetic pathways and occurs through a negative feedback inhibition. In this mechanism, HISN1 is allosterically inhibited by the final product, histidine. The first studies of this this regulation in plants have been conducted in pea (*Pisum sativum*), barley (*Hordeum vulgare*), and oats (*Avena sativa*) by A. Wiater et al. (Wiater et al. 1971c). Similar results have been obtained for *A. thaliana* (Ohta et al. 2000) and *M. truncatula* (Ruszkowski 2018). However, the study conducted by M. Ruszkowski provided structural insights into HISN1's regulation by AMP, which increases the sensitivity to allosteric inhibition by histidine. This is especially interesting regarding the fact that AMP is a known competitive inhibitor of bacterial homologs of plant HISN1, ATP-PRTs. Before the beginning of this project, the information about the activity and regulation of single HBP enzymes in plants was scarce. Novel results addressing this knowledge gap, are presented further in this dissertation.

5.4 Histidine biosynthesis in the light of evolution

The emergence of biosynthetic pathways during the early evolution of life allowed ancient organisms to become less dependent on exogenous sources of nutrients. As the initial supply of AAs and other compounds in the early environment diminished, a significant selective pressure emerged. This pressure favored early heterotrophic cells that evolved the capability to synthesize these vital molecules themselves. Over time, a variety of forces and molecular mechanisms likely influenced the formation of genomes and the advent of new metabolic functions. In particular, gene (and operon) duplication (Lewis 1951, Ohno 1970) followed by gene fusion (Xie et al. 2003, Brilli et al. 2004) and elongation (Fani 2004) have been the main contributors. These mechanisms became a foundation of the most important hypotheses explaining the development and organization of current metabolic pathways, i.e., the patchwork hypothesis (Ycas 1974, Jensen 1976) and the retrograde hypothesis (Horowitz 1945). The patchwork hypothesis describes how primitive, promiscuous, enzymes could have been recruited to assemble metabolic pathways. In general, relatively slow, non-specific enzymes may have enabled primitive cells to overcome their limited capabilities due to small genomes and catalyze two or more similar reactions of the same pathway. The retrograde hypothesis explains the emergence of metabolic pathways in the retrograde direction, in which the final product of a pathway was the compound used by primordial heterotrophs and was available for the uptake from the environment. If the compound was unavailable, the selective pressure allowed the survival of those cells that were able to synthesize it from a similar compound in the (now) ultimate step of a pathway (Fani 2012). Even though some hypotheses tackle the problem of the origin of metabolic pathways, the precise details remain unknown.

The HBP is considered to be an ancient pathway that emerged long before the origin of the Last Universal Common Ancestor (LUCA) (Fani et al. 1995, Fani et al. 2007), which lived around 4.2 billion years ago (Moody et al. 2024). The evolution of the enzymes catalyzing the HBP is well documented in bacteria (Brilli 2004, Fani et al. 2005, Chioccioli et al. 2020, Del Duca et al. 2021) and archaea (Fondi et al. 2009). However, significantly less data are available regarding eukaryotes, with most information pertaining to fungi (Brilli et al. 2004). It is generally accepted, based on the analysis of the structure and organization of the *HISN* genes, that the HBP has been shaped by the three major events: gene duplication, gene fusion, and gene elongation (Fani et al. 1995, Fani et al. 2009).

5.5 Rationale for choosing this pathway

There are several reasons to justify studying the HBP in plants. First, the understanding of the structure-function relationship in enzymes lays the foundation for developing agents that can modulate the pathway's flux. This modulation can occur through inhibition or activation, each having distinct outcomes. The inhibition of the pathway can lead to histidine starvation, disrupting critical metabolic processes such as protein and purine biosynthesis. It can also result in the accumulation of the pathway intermediates, having additional toxic effects. Conversely, the activation of the pathway can increase the concentration of intermediates and/or histidine, potentially enhancing the plant's growth and stress responses.

The inhibition of the HBP in plants could help address the global issue of herbicide resistance (HR) in weeds. Currently, there are 530 reported cases of the HR, involving 270 weed species that have become resistant to one or more of the 170 herbicides used across 72 countries (Heap 2024). These resistance cases span 21 of the 31 known herbicide modes of action (MoA), such as acetolactate synthase inhibition (affecting over 170 species), photosystem II inhibition (affecting over 80 species), and inhibition of 5-*enol*-pyruvyl-shikimate-3-phosphate (EPSP) synthase (affecting over 50 species) (Heap 2024). Despite the widespread resistance to the EPSP synthase inhibitors like glyphosate, some targets within this MoA, e.g., the HBP, remain unaffected (Koo et al. 2022). Glyphosate, the active ingredient in *Roundup* produced by Monsanto/Bayer, is notable for its effectiveness and has been long considered environmentally friendly, due to its rapid degradation by soil microflora into CO_2, ammonia, and inorganic phosphate (Forlani et al. 1999). However, the increasing resistance to glyphosate among weeds necessitates the development of new herbicides to ensure sustainable agriculture for a growing global human population. Furthermore, the glyphosate's safety has been questioned following the World Health Organization (WHO) 2015 announcement, based on the International Agency for Research on Cancer (IARC) assessment, suggesting that glyphosate might be carcinogenic to humans (Guyton et al. 2015). Despite this, eighteen European Union countries renewed the license for the glyphosate use in 2018 for another five years. As of today, the European Union (EU) renewed the approval of glyphosate until December 2033 (EU 2023). It is crucial to explore alternatives of glyphosate, while continuing to investigate its carcinogenic potential and environmental impact. Recent structural studies, e.g., the EPSP synthase structure from *A. thaliana* (PDB ID: 7PXY) (Ruszkowski et al. 2022), offer valuable insights into the mechanisms of action of various potential herbicides and aid in the rational design of new ones.

The development of novel herbicides might not be the only outcome of the HBP studies. Upregulating the flux of the HBP through activators could be a significant added value. A notable example is the activation of the bacterial homologue of HISN1, ATP-PRT, by 3-(2-thienyl)-L-alanine (TIH) in *Mycobacterium tuberculosis*, as discovered by J. Pisco et al. (Pisco et al. 2017). In that study, *M. tuberculosis* ATP-PRT was allosterically activated by TIH by up to 500%. Whether a similar activation is possible in plants is still not known.

The activation of any enzyme involved in histidine biosynthesis could elevate cellular and free histidine levels. Plants with high histidine levels might chelate divalent metal cations, thereby reducing the soil contamination (Wycisk et al. 2004). The soil contamination with metal ions, for example nickel, can lead to higher metal concentrations in food products, such as chocolate, peanut butter, breakfast cereals, and tea, contributing to nickel allergies, which affect 10-20% of European women (Diepgen et al. 2016, Babaahmadifooladi et al. 2020). Nickel hyperaccumulating species, such as those in the *Alyssum* sp. and *Thlaspi* sp. genera, exemplify plants that can mitigate soil contamination (Persans et al. 1999, Broadhurst et al. 2016). Furthermore, divalent metal cation contamination is not limited to soil but also affects water bodies. For instance, algae with increased free histidine content could be employed to remediate lakes and rivers contaminated with metals such as Cu, Co, Cr, Fe, Hg, Ni, U, and Zn (Singer et al. 2007, Bakatula et al. 2014, Sharma et al. 2021). This approach could provide a bioremediation strategy for both soil and water contamination, ensuring food safety.

In the past two decades, histidine biosynthesis has been extensively studied in bacteria to develop new antibiotics in response to emerging antibiotic resistance. Homologs of plant HISN2 (Gupta et al. 2017), and HISN8 have been targeted in *Brucella melitensis* as part of efforts to treat brucellosis, a zoonotic disease (Abdo et al. 2008, Turtaut et al. 2014). The HISN2 homologs were also targeted in *Shigella flexneri*, the causative agent of shigellosis, a disease responsible for approximately 164,000 annual deaths worldwide (Kotloff et al. 2018, Wang et al. 2019). Inhibitors of a bacterial homolog of HISN5 have been sought to treat human tuberculosis caused by *M. tuberculosis* (Kumar et al. 2022). Similarly, the HBP, along with other amino acid pathways, is now being investigated for the development of new antifungal compounds (Dietl et al. 2016, Kuplinska et al. 2021). To date, only triazole compounds have been described as the compounds of the anti-HBP activity in plants. However, only one of these agents, 3-amino-1,2,4-triazole (Amitrole), is available on the market. Unfortunately, amitrole is non-selective and carcinogenic, highlighting the need for the discovery of new, safer compounds (Furukawa et al. 2010).

5.6 *Medicago truncatula* as a model organism

The choice of a model legume, *Medicago truncatula*, as a source of coding sequences for studying plant histidine biosynthesis was motivated by two main factors. Firstly, the availability of genomic sequence data allows for preparation of genetic constructs for downstream procedures. Such genomic data are now available for *M. truncatula* (Young et al. 2011, Tang et al. 2014, Burks et al. 2018). Secondly, the genetic similarity of a model system to other economically important legumes, enhances the reliability of compared results and their extrapolation to other species, especially when stress-induced mechanisms are studied (Rose 2008, Song et al. 2019).

Legumes (Fabaceae) are one of the most economically and environmentally significant forage crops, such as lucerne, *Medicago sativa*, which belongs to the same genus as *M. truncatula*. (Mueller-Harvey et al. 2019, Hrbácková et al. 2020, Sakiroglu et al. 2021). Legumes also include agriculturally important species such as *Glycine max* (soybean) and *Pisum sativum* (pea). This family includes also weedy pests, like the closely related *Medicago intertexta*, *Melilotus indicus* (Hendawy et al. 2019), and a common vetch (*Vicia sativa*). Notably, in 2014, common vetch developed resistance to acetolactate synthase inhibitors, posing a challenge to agricultural management (Heap 2024). Therefore, most of the results obtained from the studies of *M. truncatula* could be applied to understand the biology of the other family members.

Genomic data availability offers other directions of studies. For instance, studying plant-symbiont relationships is widely conducted in model systems from the Fabaceae family (De Bruijn 2020). In particular, *M. truncatula* is a host to mutualistic, nitrogen-fixing bacteria, from the *Rhizobium* sp. and *Sinorhizobium* sp. genera (Liu et al. 2019, Song et al. 2019). Because the histidine biosynthesis intertwines with nitrogen metabolism, utilization of *M. truncatula* as a model organism is justified, especially in terms of future research.

6. Significance of this work

Histidine biosynthesis has been studied for decades, primarily in prokaryotes and fungi, but for a long time it has been neglected in plants due to the lack of tools that would make such studies feasible. However, given its ancient origin, the HBP has a similar course across all organisms. Extensive research in bacterial systems has resulted in the presence of dozens of HBP enzymes' structures in the PDB. Therefore, due to the availability of bacterial structures, I was particularly interested in studying the structural differences between enzymes catalyzing histidine biosynthesis in plants and prokaryotes. Identification of such peculiarities would allow for design of plant-specific inhibitors which might be further developed into herbicides, minimizing the risk of their non-specific interactions with the HBP enzymes present, for example in human microbiota.

The HBP has been the focus of several doctoral dissertations, predominantly addressing individual enzymes involved in the bacterial systems. The first step of the HBP is catalyzed by ATP-phosphoribosyltransferase (ATP-PRT or HisG), which in prokaryotes exists either as a homo-hexameric long form or a hetero-octameric short form. The long form contains two catalytic domains and a C-terminal regulatory domain, while the short form lacks the regulatory domain (HisZ). Therefore, the long form is catalytically independent and the short form requires a histidyl-tRNA synthetase paralogue (HisZ) for its activity (Sissler et al. 1999). The B. Lohkamp's work on the structures of HisG from *E. coli* and the thermophilic bacterium, *Methanobacterium thermoautotrophicum*, provided insights into the inhibition mechanisms and, therefore, regulation of the HisG enzyme via oligomerization with HisZ (Lohkamp 2003). The C. M. Thomson's dissertation explored the short form of ATP-PRT from the gram-negative bacterium *Psychrobacter arcticus*, focusing on its allosteric regulation and inhibition. Thomson established the kinetic mechanisms of non-competitive inhibition by histidine and competitive inhibition by AMP (Thomson 2019). P. Khajeaian characterized a "super-long" form of ATP-PRT from the lactic acid bacterium, *Leuconostoc mesenteroides,* and a short form of ATP-PRT from the hyperthermophile *Aquifex aeolicus*, further expanding the understanding of these enzymes (Khajeaian 2022). The evolution and regulation of the $(\beta\alpha)_8$-barrels in the HBP, specifically 5′-ProFAR isomerase (HisA) and the cyclase subunit of imidazole-glycerol phosphate synthase (HisF), were elucidated by B. Reisinger. This research demonstrated that HisA and HisF could be engineered to bind and process phosphoribosyl anthranilate from the tryptophan biosynthetic pathway (Reisinger 2013). F. List characterized structural changes and the activation mechanism of the entire imidazole-glycerol phosphate synthase complex (HisHF)

(List 2009). Interactions between the HisH and HisF subunits, including allosteric inhibition and regulation, were studied by A. Holinski (Holinski 2017), while E. Hupfeld described the structural changes during the catalytic reaction and the molecular mechanism of the HisH subunit activation (Hupfeld 2020). All these studies collectively contributed to our understanding of the nature of the HBP and its regulation in bacteria.

One of the first studies of the HBP conducted in a eukaryotic system was carried out in the 1960s by A. C. Minson in the mold *Neurospora crassa* (Minson 1968). This work focused on the protein purification, and genetic and biochemical characterization of HIS4, a trifunctional homolog of plant HISN2, which is able to catalyze the second, third and the last steps of histidine biosynthesis. A fungal system was also used in the studies of the chromatin structure and the regulation of the *HIS7* gene expression in *Saccharomyces cerevisiae* and *Aspergillus nidulans* (Valerius 2001). In plants, the overexpression and activity measurements of HISN2 and its mutants from *A. thaliana* were explored (Campbell 1999). To the best of my knowledge, the most recent dissertation on a plant HBP enzyme was published in 2016, and involved cryoEM studies of HISN5 from *A. thaliana* (Rawson 2016).

When I joined the project in 2019, the structural information was available for only four enzymes from the eukaryotic HBP. The first structurally characterized enzyme was HISN5 from *A. thaliana*, with its mechanism of action updated a decade later (Glynn et al. 2005, Bisson et al. 2015, Rawson et al. 2018). Following this, HISN7 from *M. truncatula* was identified based on its high sequence identity to the homolog from *A. thaliana* (which was identified only six years prior, as the last the enzyme of the HBP in plants), allowing for the determination of four high-resolution structures and biochemical characterization, including the discovery of methylene-bridged cross-links between HISN7 monomers (Ruszkowski et al. 2016). A year later, three crystal structures of the last enzyme in the sequence of histidine biosynthesis, i.e., HISN8, were solved. These structures provided detailed insights into the enzyme's architecture and cofactor binding mode, elucidating the mechanisms of a hydride transfer (Ruszkowski et al. 2017). Finally, the first enzyme of the pathway, HISN1, was characterized last, revealing structural changes between the relaxed and tense states upon histidine binding and highlighting the differences in MD of plant HISN1s and prokaryotic ATP-PRTs (Ruszkowski 2018).

In this dissertation, building upon the studies initiated by Prof. M. Ruszkowski, I present a comprehensive set of the first crystal and cryoEM structures, including HISN2, HISN3, HISN5 (the first structure was published by the other group), and HISN6, which nearly complete the entire pathway, with only HISN4 structures missing. By combining experimentally obtained enzyme structures with computational methods, such as VS

campaigns, I was able to identify small molecules that might serve as future herbicide candidates and/or scaffolds for novel herbicide development. This approach addresses the need for safer and plant-specific herbicides, to ensure the sustainability of food production. These structural findings are complemented by the enzymatic characterization and are analyzed within the evolutionary context, providing novel insights into the function, regulation, and phylogenetic origins of the HBP enzymes in plants.

7. Aim of the dissertation

Typically, scientific studies commence by formulating hypotheses that are subsequently either confirmed or rejected by the study. Such an approach is challenging to employ in structural biology due to the exploratory nature of these studies. Often, unexpected features are discovered after solving the structure, potentially altering the initial premises. Consequently, in this thesis, the exploration of histidine biosynthesis in plants has been initiated by formulating four scientific questions (**Qs**) based on existing literature and preliminary observations:

Q1: What are the structural features of the remaining enzymes that catalyze histidine biosynthesis in plants?

Q2: What are the structural and functional differences between the plant and bacterial enzymes involved in the HBP?

Q3: How can understanding of the structure-function relationships of these enzymes facilitate the design of specific inhibitors and future herbicides?

Q4: How are plant and bacterial genes encoding the enzymes of the HBP evolutionary related?

To answer the scientific questions, I have focused on achieving specific research goals by utilizing methods and tools from structural biology, enzymology, and bioinformatics. The aims of the work were to (i) fill the knowledge gap about histidine biosynthesis in plants, (ii) understand the mechanisms of the pathway's flux regulation, and (iii) present these results in the light of evolution. Within this scope, I have formulated a set of research goals (**Gs**):

G1: Determination of the structures of HISN2, HISN3, HISN4, HISN5, and HISN6 in complexes with natural ligands and/or their analogs.

G2: Assessment of the catalytic parameters of the HBP enzymes through kinetic studies.

G3: Development of protocols for the *in vitro* enzymatic synthesis of substrates.

G4: Utilization of structural and functional data to design scaffolds for future enzyme inhibitors, using computational tools.

G5: Explanation of the phylogenetic relationships between plant HBP enzymes and their homologs.

8. Summary of methods and tools

Studying the HBP in plants in order to achieve the presented goals required a diverse set of methods and techniques from various disciplines. The primary methods were those of structural biology, i.e., crystallography and cryoEM. However, they required a prior preparation of genetic constructs, often mutants, therefore utilizing methods from molecular biology. To study enzymes' properties and interactions, a substrate source was necessary, but many of the HBP intermediates are commercially not available, therefore I utilized *in vitro* reconstruction of the pathway to obtain the molecules of interest. Finally, a set of obtained results could be analyzed by a plethora of bioinformatic tools. A detailed description of the methods can be found in the attached publications. The workflow was divided into five steps:

1) Protein preparation involved the same procedures for both structural and functional studies:

- isolation of RNA using Universal RNA Purification Kit (EurX),
- reverse transcription of RNA to obtain cDNA serving as a template using Super-Script II reverse transcriptase (Life Technologies),
- design of primers complementary to the enzymes' coding sequences, followed by DNA amplification via polymerase chain reaction (PCR) using Platinum SuperFi II PCR Master Mix (Thermo Fisher Scientific), to obtain inserts,
- PCR product purification using PCR/DNA Clean-Up Purification Kit (Eurx),
- incorporation of the inserts into pMCSG53 and/or pMCSG68 vectors using the Ligation Independent Cloning (LIC) reaction (Kim et al. 2011),
- site-directed mutagenesis using the Polymerase Incomplete Primer Extension (PIPE) method (Klock et al. 2009),
- agarose gel electrophoresis,
- transformation of *E. coli* BL21 (DE3) Gold competent cells using the heat shock method,
- DNA sequencing by an external company (Genomed S.A.),
- protein expression in *E. coli* BL21 (DE3) Gold,
- protein purification using His-tag affinity chromatography or diethylaminoethyl (DEAE) cellulose ion-exchange chromatography,
- size-exclusion chromatography using ÄKTA systems,
- SDS-PAGE analysis,
- protein concentration using Amicon Ultra 15 mL centrifugal filters (Merck),
- spectrophotometric determination of protein concentration at $\lambda = 280$ nm.

2) Crystallographic studies:

- protein crystallization, co-crystallization, and crystal soaking with ligands,

- crystal harvesting and cryoprotection,

- data measurements using a synchrotron radiation (P13 Beamline, PETRA Synchrotron in Hamburg) or a domestic diffractometer (Rigaku XtaLAB Synergy-R),

- data processing in *XDS* (Kabsch 2010); molecular replacement in *PHASER* (Mccoy et al. 2007) using experimental or computational [*AlphaFold* (Jumper et al. 2021)] structures,

- standardized placement of models in unit cells in *ACHESYM* (Kowiel et al. 2014),

- model building in *COOT* (Emsley et al. 2010), and refinement in *Phenix* (Liebschner et al. 2019) or *CCP4* (Agirre et al. 2023) packages,

- polder maps generation using *Phenix.polder* (Liebschner et al. 2017),

- model validation with *MolProbity* (Williams et al. 2018).

3) CryoEM studies:

- preparation and vitrification of grids,

- imaging with an electron microscope,

- data processing in *Relion* (Scheres 2012) and *cryoSPARC* (Punjani et al. 2017),

- model building and refinement in *Phenix* package (Liebschner et al. 2019) and *COOT*,

- model validation in the *wwPDB Validation System* (Burley et al. 2019).

4) Functional studies and activity measurements:

- activity, inhibitory assays, and interaction studies using spectrophotometry and isothermal titration calorimetry (ITC),

- enzymatic synthesis of substrates, monitored using spectrophotometry.

5) Computational methods and bioinformatic tools:

- experimentally obtained models served as inputs for downstream processes,

- VS using the *ZINC* (Sterling et al. 2015) database and *AutoDock Vina* (Trott et al. 2010),

- MD using *Gromacs* (Abraham et al. 2015) and *Desmond* (Bowers et al. 2006),

- transit peptide prediction using *TargetP* (Emanuelsson et al. 2000) and *AlphaFold*,

- molecular tunnel analyses in *CAVER* (Chovancova et al. 2012, Jurcik et al. 2018),

- phylogenetic analyses in *Mega-X* (Tamura et al. 2021),

- SSN using *EFI-EST* (Zallot et al. 2019) and *Cytoscape* (Shannon et al. 2003),

- residue conservation in *ConSurf* (Yariv et al. 2023),

- phosphate binding site predictions in *Nucleos* (Parca et al. 2013),

- visualization of protein-ligand interactions in *Discovery Studio* (BIOVIA 2023),

- surface electrostatic potential calculations in *PDB2PQR* (Dolinsky et al. 2004) and *APBS* (Baker et al. 2001),

- exploration of macromolecular interfaces using *PDBePISA* (Krissinel et al. 2007),

- structure visualization and structural alignments in *Chimera* (Pettersen et al. 2004) and *ChimeraX* (Pettersen et al. 2021).

9. Summary of results and discussion

9.1 HISN2

P1: Witek W., Śliwiak J., Ruszkowski M. (**2021**) *Structural and mechanistic insights into the bifunctional HISN2 enzyme catalyzing the second and third steps of histidine biosynthesis in plants*, Scientific Reports 11(1):9647, https://doi.org/10.1038/s41598-021-88920-2

Highlights:

- The architecture of *Mt*HISN2 and its interaction with AMP were described
- AMP inhibits *Mt*HISN2 in a physiologically relevant concentration
- A possible second-tier regulatory mechanism of the HBP flux was proposed
- The catalytic mechanism was updated
- Plant *HISN2* genes derive from the Myxococcota phylum

This publication describes a bifunctional enzyme, HISN2, from *M. truncatula* (further referred to as *Mt*HISN2), which catalyzes the second and third reactions of the HBP in plants (Figure 3). In most bacteria, these steps are catalyzed by separate enzymes, whereas in fungi, particularly yeast, they are catalyzed by a trifunctional enzyme which also catalyzes the last reaction of the histidine biosynthesis. The C-terminal domain of HISN2, which has a PRA-PH activity, catalyzes the hydrolysis of a pyrophosphate moiety in PR-ATP, which leads to the formation of PR-AMP. The product then moves to the N-terminal domain, which opens the adenine ring of PR-AMP leading to the formation of ProFAR. This research paper covers four key aspects: (i) characterization of HISN2 based on crystal structures, (ii) description of similarities and differences with its bacterial counterpart, HisIE, (iii) interactions with AMP which participates in the HBP regulation, and (iv) updates the catalytic mechanism.

MtHISN2 is constituted of two intertwining chains, each of 26.4 kDa. A negatively charged surface has an area of about 20,000 Å^2, and the interface between both chains is approx. 4000 Å^2. A complete sequence contains 283 AA residues, however, the construct used in this study was N-terminally truncated at Val49, in order to remove a 48-AA residues long chloroplast transit peptide.

Phylogenetic analyses using 53,000 protein sequences from four InterPro families (IPR008179, IPR021130, IPR002496, and IPR038019) allowed to assess the sequence similarity between prokaryotic and eukaryotic HISN2 homologs, and trace their evolution. The analyses indicated that plant bifunctional enzymes derive from the Myxoccocales order in the phylum Myxococcota (Attachment 1: Figure 2). Back in 2020/2021 when this publication was

still in progress, Myxococcales belonged to the class of Deltaproteobacteria within the Proteobacteria phylum, according to the valid nomenclature at that time. However, there was a significant change in the nomenclature of prokaryotic taxonomy in 2021 (Waite et al. 2020, Oren et al. 2021), and Deltaproteobacteria were reclassified to a separate phylum – Myxococcota. Regarding this change, the updated nomenclature has been used in my further articles and this dissertation.

The X-ray structures of *Mt*HISN2 were solved by experimental phasing using single wavelength anomalous dispersion (SAD) of zinc cations bound to the protein. The unliganded protein (PDB ID: 7BGM, 1.60 Å), as well as the *Mt*HISN2-AMP complex (PDB ID: 7BGN, 2.70 Å) crystallized in the *C*2 space group. The unliganded protein contained two protein chains in the asymmetric unit (ASU), and the complex with AMP contained six chains in the ASU (three dimers). The N-terminal domain of the PRA-CH activity spans residues 49-158, whereas the C-terminal domain of the PRA-PH activity ranges between residues 172-283. The domains are connected by two loops which cover the range 159-171. The PRA-CH domain is made of α-helices and β-strands, which form a so-called β-grasp fold, while the PRA-PH domain consists entirely of ten α-helices connected by loops (Attachment 1: Figure 3A).

The crystal structures of *Mt*HISN2 revealed a total of five metal binding sites (MBS). In general, bacterial PRA-PH enzymes are Mg^{2+}-dependent, however, the crystals of *Mt*HISN2 could only be grown in the presence of a low concentration of Zn^{2+} in addition to Mg^{2+}. A PRA-CH homologous structure from *M. thermoautotrophicum*, which contained Cd^{2+} ions, is another example of other-than-zinc metal binding pattern. Because zinc often binds non-specifically to proteins, or at sites naturally binding other divalent metal cations, we decided to use a more general term – MBS – to avoid confusion. The PRA-PH domain contains two MBS (Attachment 1: Figure 4A), i.e., MBS1 in which the zinc ion is coordinated by one carboxyl oxygen of Glu217 and two carboxyl oxygen atoms of Glu220. In some subunits present in the ASU, Glu217 participates in the coordination sphere of MBS2, which results in the lack of MBS1. Because of that, MBS1 could be just a crystallization artifact, therefore MBS2 is considered to be catalytically relevant. In MBS2, the zinc ion is tetrahedrally coordinated by the side chains of Glu214, Glu234, Asp237, and a water molecule. It is very likely that *in vivo*, MBS1 and MBS2 are occupied by Mg^{2+} cations, instead of Zn^{2+} cations, as magnesium is required for the PRA-PH activity (Smith et al. 1965). The PRA-CH domain contains two or three MBS (model- and subunit-dependent) (Attachment 1: Figure 4B). The MBS3 coordinates the zinc ion through the side chains of Asp125*, Asp127*, Asp129* (asterisks denote residues from a counterpart chain), and two water molecules, forming a trigonal bipyramid. The zinc ion

in the MBS4 is coordinated tetrahedrally by the Nε of His143, a thiol of Cys126*, and two water molecules. Interestingly, the MBS4 was not present in the *Mt*HISN2-AMP complex. Lastly, the MBS5 contains a zinc ion coordinated by thiols of Cys126*, Cys142 and Cys149. Based on the residues in the metal coordination spheres, MBS1-3 are most likely occupied in vivo by Mg^{2+}, whereas MBS4 and MBS5 by Zn^{2+} (Vallee et al. 1993, D'Ordine et al. 2012).

The structure of *Mt*HISN2 was compared to its bacterial mono- and bifunctional counterparts, revealing high structural similarities, despite low sequence identities (31-40%). Interestingly, the structural alignment with the only available bacterial structure of a bifunctional *S. flexneri* HisIE (*Sf*HisIE, PDB ID: 6J2L) showed low root mean square deviation (RMSD) values only for individual domains, i.e., 0.90 Å across 72 Cα atoms (PRA-PH and HisE) and 0.84 Å across 92 Cα atoms (PRA-CH and HisI). However, when the PRA-CH and HisE domains were superposed, relative rotation of the PRA-PH and HisI domains differed by approx. 40° along the measured axis. Another difference was that *Sf*HisIE lacks a super-secondary strand-helix-loop motif, which in *Mt*HISN2 encompasses residues 150-172 and is involved in domain swapping (Attachment 1: Figure 5C).

The structure of *Mt*HISN2 was analyzed using *CAVER*, in search of molecular tunnels which may connect active sites of both functional domains to shuttle the PR-AMP intermediate between reactions. No such tunnel was detected, indicating that the intermediate has to be released to the solution and trapped by the PRA-CH domain of another (or the same) enzyme molecule.

AMP in the PRA-PH domain formed hydrogen bonds through its phosphate moiety and the adenine ring. The phosphate moiety was hydrogen bonded to the side chains of Arg183, Ser195, and Thr197, as well as to the backbones of Trp196 and Thr197. The side chain of Arg263 interacted with the N1 atom of the adenine ring. The ring formed π-π stacking interactions the side chain of Tyr240 (Attachment 1: Figure 6A). However, the positioning of the AMP molecule in the PRA-PH domain did not mimic neither PR-ATP (substrate), nor PR-AMP (product). This conclusion was based on the orientation of the AMP phosphate group pointing away from MBS1-2 and interacting with the side chain of Arg183 instead, therefore the adenine ring of AMP was rotated by approx. 180°. Interestingly, AMP binds to *Mt*HISN1 in a similar manner, in which it is rotated by 180° compared to PR-ATP (Ruszkowski 2018).

The second AMP molecule, bound to the PRA-CH domain, formed a network of interactions with surrounding residues. The phosphate was coordinated by the side chains of Trp107, Lys 109, Thr112, Ser113, and a water molecule, as well as by the backbones of Gly110, Thr112, and Ser113. The adenine ring interacted with the backbone of Thr141 and the side

chain of Lys109. The adenine moiety formed edge-to-face interaction with the side chain of Trp107 (Attachment 1: Figure 6C). Experimental results were compared with *in-silico* analyses conducted by other groups in attempts to predict PR-AMP binding to the cyclohydrolase domain (HisI) (D'Ordine et al. 2012, Wang et al. 2019). It was proposed by the authors, based on the *in silico* modelling of PR-AMP to the *M. thermoautotrophicum* HisI (PDB: 1ZPS) crystal structure, that the N1-phosphoribosyl and N9-phosphoribosyl moieties of PR-AMP should interact with the enzyme through regions corresponding to $_{111}ETS_{113}$ and $_{100}SRS_{102}$ in *Mt*HISN2, respectively. However, the N9-phosphoribosyl moiety of AMP in the complex with *Mt*HISN2 interacted with the $_{107}WTKGETS_{113}$ region, and the N1- phosphoribosyl moiety of PR-AMP would interact with $_{100}SRS_{102}$. The position of AMP allowed the catalytic mechanism of the PRA-CH domain proposed by other groups (Reizer et al. 1994, D'Ordine et al. 2012) to be updated. In the updated mechanism, the nucleophilic water molecule in the MBS5 is activated by His143, acting as a general base. The activated hydroxyl anion performs a nucleophilic attack on the purine C6 atom, breaking the N1-C6 bond .

AMP is a known activity regulator of plant HISN1 and its homologs from other kingdoms. It increases HISN1's sensitivity to histidine binding, therefore to the feedback regulation by inhibition (Ruszkowski 2018). However, so far, there have been no indications that other HBP enzymes could be regulated by AMP. In this work, *Mt*HISN2 inhibition by AMP was assayed using enzymatically synthesized PR-ATP. It was observed that AMP caused over 60% inhibition (Attachment 1: Figure 7B) at the 100 µM concentration, which corresponds to the physiological concentration of AMP in maize mesophyll chloroplasts, ranging between 40 and 260 µM (Usuda 1988). This interaction suggests the existence of a second-tier regulatory mechanism of the HBP in plants. The AMP interactions with *Mt*HISN2 WT were further investigated using ITC and resulted in the K_d value of 47 ± 6 µM (Attachment 1: Figure 7D). However, to deduce which domain bound AMP with a given affinity, ITC experiments were performed on seven point mutants: (i) K109A, T112V, S113A, and H143E for the PRA-CH domain, and (ii) R183E, T197V, and Y240T for the PRA-PH domain. The results indicated a lowered AMP binding affinity in the case of the PRA-CH domain mutants, meaning that AMP binding to this domain is physiologically relevant (Attachment 1: Figure 7E). Concluding, as fluctuations of the AMP/ATP ratio reflect the cell metabolic state, this secondary regulatory mechanism of the HBP allows for rational resource consumption by the pathway.

9.2 HISN3

P2: Witek W., Imiołczyk B., Ruszkowski M. (**2024**) *Structural, kinetic, and evolutionary peculiarities of HISN3, a plant 5′-ProFAR isomerase,* Plant Physiology and Biochemistry 215 109065, https://doi.org/10.1016/j.plaphy.2024.109065

Highlights:

- The structure of *Medicago truncatula* 5′-ProFAR isomerase (*Mt*HISN3) was investigated
- *Mt*HISN3 is a highly efficient enzyme
- A β-turn and sodium cations peculiar for plants participate in ligand recognition
- Plant 5′-ProFAR isomerases are likely of a non-cyanobacterial origin
- VS campaigns revealed molecules for herbicide development

This article comprises structural, kinetic, and phylogenetic studies of HISN3, an enzyme from the legume, *M. truncatula* (*Mt*HISN3). HISN3 is an N′-[(5′-phosphoribosyl)formimino]-5-aminoimidazole-4-carboxamide ribonucleotide isomerase (EC 5.3.1.16), also known as 5′-ProFAR isomerase, or BBM II isomerase (Figure 4) (Ingle et al. 2005). *Mt*HISN3, together with its bacterial (HisA) and fungal (His6) homologs, belongs to $(\beta\alpha)_8$-barrel enzymes, which catalyze the reaction of isomerization of ProFAR to PrFAR, utilizing the Amadori rearrangement mechanism (Attachment 2: Figure 1). This mechanism is shared with a similar $(\beta\alpha)_8$-barrel catalyzing the third step in tryptophan biosynthesis, i.e., TrpF (EC 5.3.1.24) of N′-(5′-phosphoribosyl) anthranilate isomerase activity. The similarity between bacterial HisA and TrpF enzymes is so high, that even a single AA substitution can switch from the HisA activity to that of TrpF (Jürgens et al. 2000). However, some bacteria, for example *Streptomyces coelicolor* and *M. tuberculosis,* lack the *trpF* gene, therefore the third reaction in their tryptophan biosynthetic pathway is catalyzed by a bifunctional $(\beta\alpha)_8$-barrel, N′-(5′-phosphoribosyl) anthranilate isomerase A (PriA), of both HisA and TrpF activities (Kuper et al. 2005). This cross-functioning is especially interesting from the evolutionary perspective, because it might bring evidence for the explanation of the origin of metabolic pathways, regarding the patchwork hypothesis.

Another aspect of this study is providing framework for the rational herbicide design as our answer to the emerging HR. The ability to block the HBP pathway at this step might have a dual impact by impairing also the purine *de novo* biosynthesis due to the lack of a key substrate, AICAR, which is produced from PrFAR.

The gene encoding *Mt*HISN3 is located on chromosome 2, and the coding sequence is translated into a 312-residue long polypeptide. *Mt*HISN3, like the other HBP enzymes, contains an N-terminal chloroplast transit peptide, therefore in order to remove this peptide, *Mt*HISN3 was truncated at Ser42 (Nt42). Because attempts to crystallize WT *Mt*HISN3 were unsuccessful possibly due to a high flexibility of loops in the unbound state, an inactive mutant (*Mt*HISN3-Nt42-D57N) was designed. The introduction of the D57N mutation and the co-crystallization of the protein with the substrate were essential, and resulted in two crystal structures of complexes with ProFAR (2.36 Å, PDB ID: 9FCF) and PrFAR (1.54 Å, PDB ID: 9FCG), both solved in the *P*4 space group, with one protein chain in the ASU (Attachment 2: Table 1).

*Mt*HISN3, as a typical monomeric $(\beta\alpha)_8$-barrel, has the interior constituted by eight parallel β-strands surrounded by eight α-helices, and a centrally located active site. Based on the length of loops and location of the N- and C-termini, the catalytic and non-catalytic (basal) faces of the enzyme can be distinguished. The catalytic face contains longer loops, and the basal face contains both termini and shorter loops (Attachment 2: Figure 2). These two sides create a dipole-like charge distribution on the surface as a sum of individual helices dipoles, which expose their positively charged N-termini towards the catalytic face, whereas the negatively charged C-termini make up the basal face. Such polarization likely aids in guiding ProFAR into the active site.

Interestingly, while the *Mt*HISN3-Nt42-D57N crystals harvested after three weeks showed the presence of ProFAR (Attachment 2: Figure 3), the crystals that were kept for four months in a cold room contained mostly PrFAR (Attachment 2: Figure 4). Although the D57N variant was inactive during kinetic measurements (not shown), it apparently retained some residual activity, which led to the processing of ProFAR to PrFAR inside a crystal. This phenomenon has been observed and reported by two other groups, i.e., in the study of a homologous HisA-D7N (PDB ID: 5AB3) from *Salmonella enterica* (Newton et al. 2017), and in a bifunctional homolog PriA-D11N (PDB ID: 2Y88) from *M. tuberculosis* (Due et al. 2011).

To distinguish the highly similar moieties in ProFAR and PrFAR, a reference to N1-phosphoriboses and C5-phosphoribose (or C5-phosphoribulose, exclusively for PrFAR) was introduced (Attachment 2: Figure 3B, C, and Figure 4B, C). Notably, during the refinement of the *Mt*HISN3-Nt42-D57N_PrFAR complex structure, an alternative conformation of the mutated residue Asn57 (refined occupancies: 0.57 and 0.43) was discovered. The minor conformation would clash with the 2′-carbonyl of C5-phosphoribulose at the distance of 1.3 Å, therefore it was deduced that this Asn57 conformer exists together with the water elimination product, PrFAR-E, bound in the active site (Attachment 2: Figure 4D). Although the residual

activity of initially inactive mutants was reported before, to the best of my knowledge, no other case of such an elimination reaction in 5′-ProFAR isomerases has been documented.

The pattern of interactions is almost identical for both ProFAR and PrFAR (Attachment 2: Figure 3B, C, and Figure 4B, C). The N1-phosphoriboses interact through hydrogen bonds with the backbones of Gly138, Thr158, and Ser159, and the side chains of Ser159, Asp187, Ser189, His230, and Trp204. The aminoimidazole moieties interact by H-bonds with the backbones of Gly68 and Arg203, by π-π stacking with the side chain of Trp204, and by weak π-σ contact with the Leu112 side chain. The C5-phosphoribose and C5-phosphoribulose are hydrogen-bonded to the backbones of Gly236, Gly262, Gly285, and Ser286, and to the side chains of Asn57, His108, His230, Glu235, and Ser286. The side chain of Gln65 participates in the binding of the PrFAR-E C5-moiety by hydrogen bonding to C4′ –OH.

In its active site, *Mt*HISN3 contains a sodium cation, relevant for ligand recognition, which has not been observed in bacterial and fungal structures of homologous enzymes. The binding of the negatively charged phosphate moieties of the C5-phosphoribose and C5-phosphoribulose is additionally strengthened by a sodium cation (Na1), which is coordinated by the backbone carbonyls of Ile66, the side chains of Gln65 and Glu235, and by a water molecule only in the complex with PrFAR (Attachment 2: Figure S1A, B). Another sodium cation (Na2) is present in the complex with PrFAR, however it is located outside of the active site, approx. 7 Å from the PrFAR's N1-phosphoribose phosphate moiety. It is coordinated by the backbone carbonyls of Ser159 and Phe162, and four water molecules (Attachment 2: Figure S1B). There are also two chloride ions (Cl1 and Cl2) located on the catalytic face, and they seem to contribute to the structural organization of flexible regions, at least inside the crystals, rather than playing catalytic role.

In general, ions are scarce in the structures of bacterial HISN3 homologs available in the PDB. One Na$^+$ ion was found in *S. enterica* HisA (PDB ID: 5AB3) but it was located on the basal face of the enzyme. The lack of sodium cations in active sites of bacterial homologs was connected with the arginine residue at a position equivalent to a conserved in plants Gln65, which coordinates Na1 in *Mt*HISN3. The arginine's side chain in bacterial HisA enzymes, likely repels metal ions from the active site.

The steady-state kinetic measurements were conducted to test the activity of WT *Mt*HISN3-Nt42 and *Mt*HISN3-Nt42-D57N. Due to a lack of commercially available ProFAR, mostly because of its short half-life (953 min at 37 °C) (Henn-Sax et al. 2002), it had to be synthesized enzymatically. The measurement of the activity of WT *Mt*HISN3-Nt42 (at 50 nM

concentration) was coupled with the excess of *Mt*HISN4-Nt48, according to the procedures used for non-plant HISN3 homologs (Klem et al. 1993, Jürgens et al. 2000, Kuper et al. 2005, Söderholm et al. 2015). As a result, PrFAR produced by *Mt*HISN3 was immediately processed by *Mt*HISN4 to IGP and AICAR, therefore the forward-directionality of the *Mt*HISN3 reaction could be maintained, avoiding the back-processing of PrFAR to ProFAR and mitigating the risk of product inhibition. The reaction catalyzed by *Mt*HISN4-Nt48 could be monitored at $\lambda = 300$ nm, as a decrease in absorption. The obtained data were then fitted to the Michaelis-Menten equation which resulted in $K_M = 4.6 \pm 1.0$ μM, $k_{cat} = 9.7 \pm 0.4$ s^{-1}, giving the catalytic efficiency (k_{cat}/K_M) of 2.1 μM^{-1} s^{-1} (Attachment 2: Figure 5).

The comparisons of WT *Mt*HISN3-Nt42's kinetic parameters with bacterial homologs, as there are no available data for eukaryotic homologs, revealed that the turnover rate for *Mt*HISN3 is slightly higher compared to the bacterial homologs. Despite a moderate K_M value, the *Mt*HISN3's catalytic efficiency is higher than that of most bacterial homologs, except for *E. coli*, whose k_{cat}/K_M is similar to that of *Mt*HISN3 and equals 3.1 μM^{-1} s^{-1} (Henn-Sax et al. 2002).

Although the evolution of bacterial HISN3 homologs, i.e., HisA and PriA has been well documented, little is known about the evolution of eukaryotic HISN3s and His6s. The SSN and phylogenetic trees were utilized to trace the evolution of those enzymes, and the results showed that plant HISN3s and fungal His6s are grouped together with the HisA sequences from Bacillota, suggesting the non-cyanobacterial origin of plant HISN3 sequences. Based on the phylogenetic trees, fungi and plants shared a common, Bacillota-type, ancestor (Attachment 2: Figure S2). The analyses also supported the ancestral role of PriA-like ancestral enzymes in the evolution of the $(\beta\alpha)_8$-fold, which were likely spread by a horizontal gene transfer (HGT) from Actinomycetota to other bacterial groups (Attachment 2: Figure 6).

The residue conservation analyses indicated a region on the basal face of *Mt*HISN3 that was conserved only in plants (Attachment 2: Figure 7A). It was hypothesized that it might play a role in the product release, however, an analysis in *CAVER* did not detect any molecular tunnels that could serve this purpose. Therefore, the function of this patch of residues remains unknown.

The structural alignment between *Mt*HISN3-Nt42 and its only eukaryotic homolog, His6 from *S. cerevisiae* (PDB ID: 2AGK), and bacterial *Se*HisA (PDB ID: 5A5W), yielded the RMSD values of 0.945 Å across 207 Cα pairs, and 1.128 Å across 121 Cα pairs, respectively (Attachment 2: Figure 8). A region comprising residues $_{60}$KGKVKQIVGSTLKDLKDDDGSDPITNFESDKS$_{91}$, which covers the active site of

*Mt*HISN3 and sticks out of the catalytic face was identified. It has a direct contact with ProFAR/PrFAR and bears residues conserved in plants, i.e., Ile66 and Gly68. A similar but shorter (20 residues) motif is present in bacterial homologs. The role of this fragment in catalysis was investigated, hypothesizing that the central part of this motif ($_{74}$LKDDDGS$_{80}$), which contains three aspartate residues, could bind Mg^{2+} ions from the kinetic buffer and mediate in catalysis. The co-occurrence of cofactor magnesium (or other divalent metals) in tandem with sodium in sodium/potassium-binding enzymes has been reported in the literature (Sigel 2016). However, the steady-state kinetics with 10 mM ethylenediaminetetraacetic acid (EDTA) showed no difference compared to *Mt*HISN3-Nt42 assayed in standard conditions. Then, a deletion mutant was produced, *Mt*HISN3-Nt42-Δ74-80, and its activity was measured. The k_{cat} value decreased 12-fold (9.7 s^{-1} for WT vs 0.8 s^{-1} for Δ74-80), whereas K_M increased 2.5 times (4.6 μM vs 11.9 μM). To sum up, the deletion of this fragment resulted in a 30-fold loss in the catalytic efficiency (k_{cat}/K_M = 2.1 vs 0.07) (Attachment 2: Figure S4). The deletion allowed to understand the influence of the $_{74}$LKDDDGS$_{80}$ region on the catalytic efficiency, however, the rationale did not lie in its potential metal-binding capabilities. The Lys60-Ser91 fragment was later subjected to a 2-μs long MD simulation, focused on the dissociation of PrFAR from the enzyme. PrFAR either remained in the active site during a whole event, or was released after approx. 1000 ns (Attachment 2: Figure 9). Around the time frame of 900-1000 ns, PrFAR was forming multiple transient interactions with the Val67-Ile83 fragment, suggesting its role in product release. Because of the structural similarity of PrFAR and ProFAR, it is likely that the recognition of the substrate occurs through the same fragment.

The VS campaign was conducted on a 1.3 mln molecules representing a broad chemical space to find scaffolds of potential inhibitors. Five compounds of the highest binding energy gains, i.e., between -13.6 and -13.1 kcal/mol, were analyzed (Attachment 2: Figure 10B-F). This was a promising result because such high energy gains in primary VS campaigns of the other HBP enzymes have not been observed. A common feature of those five top-hit molecules was the presence of aromatic and polycyclic moieties that enhance π-π stacking interactions with the highly conserved in plants Trp204 and π-σ interactions with Leu112. Considering the high overall conservation of HISN3 and HisA/His6 sequences, the compound N-[2-(3,4-dihydro-1H-isoquinolin-2-yl)ethyl]tetrazolo[1,5-a]pyridine-7-carboxamide (ZINC71752196, binding energy gain of -13.3 kcal/mol) seems particularly interesting as it reaches Gln65, which is substituted with Arg in bacteria. Therefore, ZINC71752196 appears to be the best trade-off between the estimated energy gain and the potential to develop inhibitors specific to plant HISN3s.

9.3 HISN5

P3: Witek W., Śliwiak J., Rawski M., Ruszkowski M. (**2024**) *Targeting imidazole-glycerol phosphate dehydratase in plants: novel approach for structural and functional studies, and inhibitor blueprinting,* Frontiers in Plant Science 15;15:1343980 https://doi.org/10.3389/fpls.2024.1343980

Highlights:

- A set of high-resolution crystal and cryoEM structures was utilized to identify ligand-binding hot-spots

- The binding of small molecule fragments was described

- A VS screening campaign revealed candidate molecules and linkers for future herbicides

- A protocol for the enzymatic synthesis of the substrate was developed

- A new approach to measure HISN5 enzymatic activity by ITC was adopted

- Plant *HISN5* genes derive from Cyanobacteria

This publication presents a comprehensive structural and functional characterization of HISN5 from the legume, *M. truncatula* (*Mt*HISN5), enriched with evolutionary analyses and *in silico* herbicide design, as our response to the need of identification of a novel class of plant-selective and safer herbicides. *Mt*HISN5 catalyzes the sixth step of histidine biosynthesis, in which IGP is dehydrated to IAP (Figure 6). It is a cubic-shaped, 24-meric, homooligomer with a large, empty central void, and a total mass of 540 kDa. Each subunit consists of a four α-helix bundle sandwiched between two β-sheets (Attachment 3: Figure 3). The active site contains two Mn^{2+} ions coordinated by residues of two different subunits. However, upon ligand binding, a third subunit participates in its stabilization by Arg167** and Arg189**, therefore a functional active site is constituted of residues belonging to three subunits. Mn1 is coordinated octahedrally by Nε of His141, Nε of His213, carboxyl of Glu145, Nε of His238*, and a water molecule. The coordination sphere of Mn2 is constituted by Nε of His142, Nε of His115*, Nε of His237*, carboxyl of Glu241*, and a water molecule.

A thorough analysis of three crystal structures and two cryoEM structures allowed to detect and describe ligand-binding hotspots and characterize features of potential novel inhibitors. The ligand-binding hotspots were characterized based on crystal structures which contained small molecules present in crystallization conditions. *Mt*HISN5 crystallized in the *R*3 space group, yielding a structure at 1.55 Å (PDB ID: 8QAW) with imidazole (IMD), sodium

ion, chloride ion, formate (FMT), 1,2-ethanediol (EDO), glycerol (GOL), and tris(hydroxymethyl) aminomethane (TRS). The structure at 1.69 Å (PDB ID: 8QAX) was solved in the *I*4 space group and contained FMT, GOL and TRS. Lastly, the structure at 2.2 Å (PDB ID: 8QAY) was again solved in the *R*3 space group, and contained chloride ion, FMT, EDO, acetate (ACT), sulfate ion, and CIT (Attachment 3: Table 1). The most commonly bound molecules among all structures were EDOs (n = 41) and FMTs (n = 49). EDOs were mostly found on the inner surface of the oligomer and on the interfaces between subunits. FMTs were primary located between manganese ions in the active sites (Attachment 3: Figure 4). Another interesting finding was that IMD was bound not between Mn^{2+} ions in the active sites (like in PDB ID: 4MU1) but approx. 3.8 Å away from the center of the active site. Therefore, we postulated that the carboxylate and imidazole binding sites could be used as pharmacophores for the design of novel inhibitors.

To search for small molecules representing a broad chemical space, a VS campaign was conducted, using 3.3 mln lead-like molecules from the ZINC database (Sterling et al. 2015). The campaign resulted in six molecules of the highest binding energy gain (from -10.0 to -9.4 kcal/mol) (Attachment 3: Figure 5). The top-scoring molecules satisfied the criteria, i.e., the content of heteroatoms, which improve solubility in water and ensure specific binding interactions, as well as, the potential of π-stacking interactions due to the content of aromatic rings.

An analysis using SSN (Attachment 3: Figure 7) and a phylogenetic tree (Attachment 3: Figure S2) revealed a cyanobacterial origin of plant HISN5 enzymes. Plant HISN5 sequences are clearly distant from bacterial and fungal sequences which might be important for reaching herbicidal specificity by potent HISN5 inhibitors. Bioinformatic tools served for identification of regions characteristic to plants, which might bind novel inhibitors, i.e., regions conserved in plants but variable in other kingdoms. A cleft between two active sites was chosen (Attachment 3: Figure 9) due to its high conservation score in plants and high variability among other kingdoms. Another VS campaign was conducted in this region to find suitable linkers between two active sites. A scaffold which would contain competitive inhibitor moieties at the poles of the molecule, connected by aromatic moieties in the molecule center was proposed.

A new approach to measure the enzymatic activity of HISN5 using ITC and enzymatically synthesized substrate, *2R,3S*-IGP, was adopted. The rationale behind the enzymatic substrate synthesis was that the commercially available IGP is a mixture on four diasteroisomers, out of which only *2R,3S*-IGP can be catalyzed to IAP, and the three other bind

competitively to *Mt*HISN5. This is an artifact of a non-stereospecific chemical synthesis (Saika et al. 1993). Moreover, commercial IGP (cIGP) is available only in 1-mg packages and absorption-based methods used previously by other groups required large quantities of IGP (8-12 solutions of different IGP concentration) (Hawkes et al. 1995). By adopting the ITC single-injection method (SIM), approx. 200 µL of the substrate solution was used instead of 8-12 mL to obtain kinetic parameters. Initial ITC experiments with cIGP could not saturate *Mt*HISN5 because of the presence of competitively binding *2S,3S*-IGP (IG2), which could not be converted enzymatically to IAP. To eliminate IG2, IGP (eIGP) was enzymatically synthesized by recreating the first five steps of the HBP in vitro, with a yield of 80%. The data obtained from the ITC-SIM measurements using eIGP, instead of cIGP, were much more informative (Attachment 3: Figure 10) and resulted in calculation of the following parameters: $K_M = 227 \pm 45$ µM, $k_{cat} = 3.4 \pm 0.3$ s^{-1}. The K_M values of plant HISN5 enzymes reported in the literature vary significantly, from 49 µM in common wheat (*Triticum aestivum*), up to 1.7 mM in oat (*Avenia sativa*) (Wiater et al. 1971c). To the best of my knowledge, no k_{cat} values are available for plants. The only k_{cat} value listed in the BRENDA Enzymes database (Chang et al. 2021) is for IGPD from *M. tuberculosis*, which is 1400 s^{-1} (Ahangar et al. 2013). However, referring to previously calculated data is difficult, because the early studies conducted by A. Wiater et al. might have been subjected to errors due to the protein purification from plant extracts, and the impurity of cIGP used by M. Ahangar et. al. Therefore, we postulate to use eIGP for obtaining more reliable and comparable kinetic measurements.

9.4 HISN6

P4: Rutkiewicz M., Nogues I., **Witek W.**, Angelaccio S., Contestabile R., Ruszkowski M. (**2023**) *Insights into the substrate specificity, structure, and dynamics of plant histidinol-phosphate aminotransferase (HISN6),* Plant Physiology and Biochemistry 196:759-773, https://doi.org/10.1016/j.plaphy.2023.02.017

Highlights:

- Plant HISN6 enzymes are closely related to the orthologs from Chloroflexota
- *Mt*HISN6 is more selective that the studied homologs in bacteria
- The crystal structures of *Mt*HISN6 revealed the architecture of plant HISN6 enzymes
- A peculiar and dynamic string-like fragment governs the entrance to the active site
- Virtual screening delineated the design of novel, HISN6-targeted herbicides

This work describes *M. truncatula* HISN6 (*Mt*HISN6), which is a pyridoxal 5′-phosphate (PLP)-dependent enzyme that catalyzes the reversible conversion of IAP into HOLP (Figure 7; Attachment 4, Figure 1). The reaction is concomitant with the conversion of Glu to 2OG.

An analysis using SSN based on 21,000 sequences of *Mt*HISN6 homologs revealed that archaeal and bacterial enzymes evolved rather independently. Land plants and green algae acquired their *HISN6* from Chloroflexota, likely as a result of the HGT, thus contradicting the endosymbiotic theory (Attachment 4: Figure 2). Therefore, HISN3 is the third case of the non-cyanobacterial origin of the plants HBP enzymes, next to HISN2 and HISN3.

Functional assays of *Mt*HISN6 were conducted in the reverse direction (with HOLP as a substrate). The reaction was coupled with bovine glutamate dehydrogenase (GDH), which allowed for measuring NADH formation at 340 nm. *Mt*HISN6 exhibited a relatively high affinity towards HOLP, reflected by the K_M value of 29 µM, which was nearly 15-fold lower than the K_M value reported for *M. tuberculosis*. Another interesting finding was that in contrast to bacterial HISN6 homologs, *Mt*HISN6 showed no measurable activity towards phenylalanine, even at high concentration (25 mM). The catalytic activity of *Mt*HISN6, expressed as the k_{cat} value, was quite slow, at least in the reverse direction, i.e., from HOLP to IAP, and amounted 4 min^{-1} (Attachment 4: Figure 3).

Crystal structures were solved in the monoclinic space group, $P2_1$: (i) as a functional dimer in both open (1.57 Å, PDB ID: 8BJ1) and (ii) closed (1.40 Å, PDB ID: 8BJ2) states, and as (iii) a dimeric complex with HOLP (1.61 Å, PDB ID: 8BJ3). The apo form (1.45 Å, PDB ID: 8BJ4) also crystallized in the $P2_1$ space group but with two dimers in the ASU

47

(Attachment 4: Table 1). In the open conformation structure, the density map for a region Ile41-Leu47 was ambiguous, hinting a possibility of multiple conformations of that region. Both open- and closed-conformation structures contained PLP in their active sites which was bound to Lys244 as a Schiff base (Attachment 4: Figure 4). The *Mt*HISN6_open model contained fifteen 1,2-ethanediol (EDO) molecules and six sulfate ions, among which two were located in each active site (Attachment 4: Figure 4C), and the remaining ones were located in small clefts. *Mt*HISN6_closed contained two acetate ions, and one molecule of 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) bound close to the dimer interface. The complex of *Mt*HISN6-HOLP contained an external aldimine in chain A and geminal diamine intermediate od HOLP in chain B (Attachment 4: Figure 1B). This model contains eight EDO molecules and two sodium ions. The apo*Mt*HISN6 model contained four sodium ions, twelve sulfate ions and 23 EDOs.

*Mt*HISN6 comprises the large domain (Pro72-Asn289) and two smaller ones (Asp25-Pro71, Pro290-Ser384), as per modified terminology introduced by Ford et al. (Ford 1980). The N-terminal part of the small domain exists in open and closed conformations, which vary by the positioning of the Leu35-Pro43 fragment that oscillates like a guitar string, and the fragment Phe44-Ile58 which contains two helices (Attachment 4: Figure 4B). In the open conformation, the Leu35-Pro43 region is further from the active site, opening the entrance for the substrate. In the closed conformation, the string covers the active site entrance. Notably, in the closed form of the *Mt*HISN6-HOLP complex, Tyr38 forms H-bond with the substrate's phosphate oxygen and Nδ atoms. Backbone carbonyls of Lys312, Val314, and Leu317, together with water molecules coordinate sodium cations (Attachment 4: Figure 4E). Sodium cations are bound in the same position in all structures studied in this work, implying a specific binding relevant *in vivo*. This a novelty, compared to bacterial homologs, which did not contain metals, except for a magnesium ion in *Geobacter metallireducens* (PDB ID: 3HDO). To confirm if the presence of Na^+ is plant-specific, more structures need to be solved. Regarding the distance of Na^+ from the active site (approx. 20 Å), its role is rather structural than catalytic.

The active site of *Mt*HISN6 contains a prosthetic group, PLP, in the form of internal aldimine creating a Schiff base (Attachment 4: Figure 5). There are two active sites located at the domain interface of a functional dimer. The bottom of the active site consists of Asp117, Phe141, Met143, Asn190, Asp214, Asp216, Arg240, Gln271*, and Tyr273* (asterisks denote residues from a counterpart chain), whereas the wall is built of Tyr217, Thr241, Ser243, Arg252, and Tyr87* (Attachment 4: Figure 5A). Residues of the oscillating string cover the active site in the form of a lid (Attachment 4: Figure 5B). In the active site, PLP forms an

internal aldimine with Lys244, and a network of non-covalent interactions with the side chains of Tyr87*, Ser243, Arg252, and Thr241, and with the main chain amides of Ala116 and Asp117. The pyridine ring of PLP is sandwiched between Ala216 and Phe141. The N1 of PLP interacts with Asp214, whereas hydroxyl group is stabilized by Tyr217 and Asn190 (Attachment 4: Figure 5A).

Transition from the open to the closed states changes the size of the active site cavity, e.g., the distance between Cα of Ile41 and Cα of His353 decreases from 13.9 Å to 10.1 Å (Attachment 4: Figure 5B). Additionally, the Arg352-Lys360 fragment of the small domain also moves towards the active site, as a result of HOLP binding. The side chains of Arg352 and Tyr354 twist outward from the active site, therefore creating the environment for the phosphate group of HOLP (Attachment 4: Figure 5C). The apo*Mt*HISN6 structure has a poor electron density around the string fragment with a lack of density even for backbones of Ile41 and Leu42. This fragment becomes structured upon the PLP binding.

Sequence conservation analyses of HISN6 homologs revealed high conservation of residues constituting the active site and the dimer interface. The sequence of the Leu35-Pro43 string-like fragment is conserved among plants, except for Ala36, which might be substituted by Ser or Lys. The same fragment among all homologs is rather variable, with only Tyr38 being highly conserved, likely due to its role in HOLP binding (Attachment 4: Figure 6).

The structures of *Mt*HISN6 and *E. coli* HAT (eHspAT) were compared in open and closed states, however the open form in the PDB structure 1GEW did not have a proper electron density. The comparison between closed forms of *Mt*HISN6 and eHspAT (PDB ID: 1FG3) yielded the RMSD value of 1.65 Å. The most significant differences were observed in small N-terminal domains (Attachment 4: Figure 7). The analyses revealed that the transition between the open and closed states in eHspAT occurs due to a hinge-like movement of the small domain towards the active site, instead of the movement of a string-like fragment corresponding to Leu35-Pro43 in *Mt*HISN6. Comparison of *Mt*HISN6 with *M. tuberculosis* HAT (mHspAT, PDB ID: 4R8D) revealed a higher RMSD value (1.9 Å) (Attachment 4: Figure 8). The second analysis was justified by a lack of biochemical description of eHspAT. In the transition from the open to the closed state of mHspAT a long loop, Ser128-Thr137, becomes an α-helix. The observed conformational variability explains why mHspAT has a higher turnover rate (426 s^{-1}) but lower substrate affinity ($K_M = 420 \text{ }\mu\text{M}$) than *Mt*HISN6.

The structures of *Mt*HISN6 were subjected for the *in silico* analyses using MD and VS. The MD simulation showed the Leu35-Pro43 string-like fragment movement towards the active site, and the stabilizing influence of HOLP on the fragment (Attachment 4: Figure 9).

49

Interestingly, the MD simulation indicated that the transition between the opened and closed forms is possible without HOLP binding.

The VS screening campaign of more than 1.35 mln soluble lead-like molecules identified the most potent binding molecules (Attachment 4: Figure S2). The top hits were expected to bind in three, slightly overlapping regions around the string-like fragment, and block the transition to the closed state (Attachment 4: Figure 10). Although inhibitors of plant aminotransferases are available, none of them are HISN6-selective. For example, aminoethoxyvinylglycine can deregulate histidine biosynthesis in plants (Le Deunff et al. 2019). However, it has a strong downside due to its ability to inhibit other aminotransferases in plants and other organisms.

To conclude this work, the structural and functional studies of *Mt*HISN6 provided insights into its evolution and differences between homologs. Plant *HISN6* genes were likely acquired by a HGT from a bacterial phylum, Chloroflexota. The observed high substrate affinity and a low turnover rate of *Mt*HISN6 can be attributed to the architecture of the active site, which is tighter than in bacterial homologs. The substrate binding stabilizes the Leu35-Pro43 string in the closed state, which is necessary to form a fully-functional active site. The high sequence conservation of the string-like fragment among plants emphasizes its importance in the *Mt*HISN6 activity. The regions around the string-like fragment are also promising druggable sites targeted for herbicide development.

10. Conclusions

The HBP follows a similar sequence of reactions and utilizes the same intermediates in prokaryotic and eukaryotic organisms, however, significant differences exist between the homologous enzymes that catalyze histidine biosynthesis inside their cells. A set of high-resolution structures of the *Mt*HISN2, *Mt*HISN3, *Mt*HISN5, and *Mt*HISN6 enzymes provided insights into their architecture, catalytic mechanisms, and ligand-binding regions. The work regarding *Mt*HISN2 contributed to a discovery of a second-tier regulatory mechanism of the HBP at the enzymatic level, in which AMP causes inhibition at physiologically relevant concentrations, reflecting the metabolic status of the cell. This result combined with the kinetic measurement of *Mt*HISN3, which revealed its high catalytic efficiency, suggest tight regulation of the HBP at the enzymatic level occurs before the interconnection of the pathway with the *de novo* purine synthesis (HISN4 step). However, completion of kinetic studies of *Mt*HISN4 is required to support this statement. The kinetic studies also provided detailed information about the catalytic parameters of *Mt*HISN3, *Mt*HISN5, and *Mt*HISN6, allowing for comparisons with their bacterial homologs, and revealing higher catalytic efficiency of *Mt*HISN3, and higher substrate selectivity of *Mt*HISN6. Unfortunately, there is not enough data about the catalytic efficiency of prokaryotic HISN5 homologs.

Due to the lack of commercial availability of most substrates, the protocols for the enzymatic synthesis and purification of various substrates of the HBP have been developed by recreating the pathway's course *in vitro*. The most important outcome was the stereospecific biosynthesis of *2R,3S*-IGP with the 80% yield, overcoming the issues associated with the high content of competitively-binding IGP diastereoisomers in commercial samples. This led to the development of a novel protocol for the enzymatic characterization of HISN5 homologs, using much smaller amounts (0.2 ml vs 8-12 ml) of *2R,3S*-IGP, compared to other studies.

The *in silico* downstream processing of these experimental procedures was facilitated by bioinformatic tools, e.g., SSN and phylogenetic trees, which were instrumental in the analyses of phylogeny. These analyses suggest a non-cyanobacterial genesis of the enzymes *Mt*HISN2 (Myxococcota), *Mt*HISN3 (Bacillota), and *Mt*HISN6 (Chloroflexota), contradicting their alleged endosymbiotic origin. Only *Mt*HISN5 was confirmed to derive from Cyanobacteria. It is likely that the genes encoding these enzymes were acquired early in the evolution of the HBP through a HGT.

Another set of the *in silico* methods used in this work, i.e., MD and VS campaigns, served for better understanding of the enzymes' dynamics and the identification of potential inhibitors, respectively. The MD analyses of *Mt*HISN3 revealed the role of the plant-specific

Lys60-Ser91 fragment in aiding the product release. The MD simulations of *Mt*HISN6 have been consistent with the Leu35-Pro43 string-like fragment movement in the active site formation upon substrate binding. The VS campaigns allowed for the proposal of small molecules with a high binding potential to the studied enzymes. Therefore, by narrowing down a broad chemical space of molecules, we proposed single compounds of the highest binding-energy gain for *Mt*HISN3, *Mt*HISN5, and *Mt*HISN6. All these campaigns were enabled by the structural characterization of enzymes.

To ensure binding specificity of the selected compounds, at least at the kingdom level, I analyzed residue conservation in the studied structures using *Consurf* to identify regions conserved in plants but variable in bacteria and fungi. Considering that mammals host a gut microbiota, which might be susceptible to non-specific binding of the agents to the target enzymes, I focused on identifying the most specific interactions with the plant HBP enzymes to ensure targeted and effective outcomes. This approach, together with the VS campaigns, resulted in the identification of the binding regions with the highest kingdom-specificity.

11. Future perspectives

The understanding of the HBP has been growing over the past few decades, with the most dynamic growth observed in the past twenty years. Yet, there is still a lot to explore. Since the earliest studies, there was a scientific consensus that the only organisms able to synthesize histidine were prokaryotes, fungi, and plants. This consensus was challenged in 2018, when a complete HBP was discovered in the Metazoa kingdom, specifically in robust corals from the Scleractinia order, i.e., *Fungia* sp. and *Goniastrea aspera* (Ying et al. 2018). Initially, a contamination from the phototrophic endosymbiont, *Symbiodinium* sp., has been considered to be the source of the HBP genes identification. However, the coral *HISN* gene sequences were significantly different from those of *Symbiodinium* sp., fungi, and other organisms, supporting

W K H L U   L Q G H S H Q G H Q W   S K \ O R J H Q \   1 R W D E O \   W K H   + % 3

R I   S R W H Q W L D O   I D U F D F R W L R Q   R I   Q R Y H O   K H U E L F L G H V   D Q W L

7 K L V   G L V F R Y H U \   S D Y H H V   M H [S O R Z D D W W R L R Q   X R J I W W K H   + % 3   L Q

D S S O L F D W L R Q   F R P L Q J   I U R P   W K H   P R G X O D W L R Q   R I   W K H

7 K H   U H V H D U F K   U H J D U G L Q J   W K L V   S U R M H F W   L Q   W K

I X Q F W L R Q D O   F K D U D F W H U L] D W W R L R Q   R I   *Mt*HISN4 initially, after unsuccessful crystallizations of WT *Mt*HISN4 with its substrates, PrFAR and L-glutamine, I decided to trim potentially floppy and protruding regions, such as loops that might impede the crystallization process by proteolytic digestion with thermolysin. This approach was promising because it yielded diffracting crystals (Figure 10A)   U H V X O W L Q J   L Q   D + R Z H U X H F W X G X U D L Y R O X W L R Q L F F

W K D G R E W L D O L Q H H G P L Q H D I Obarrel (of the cyclase activity), spanning from Y378 to S533. The half barrel dimerized with itself, recreating a whole   .₈-barrel (Figure 10B).
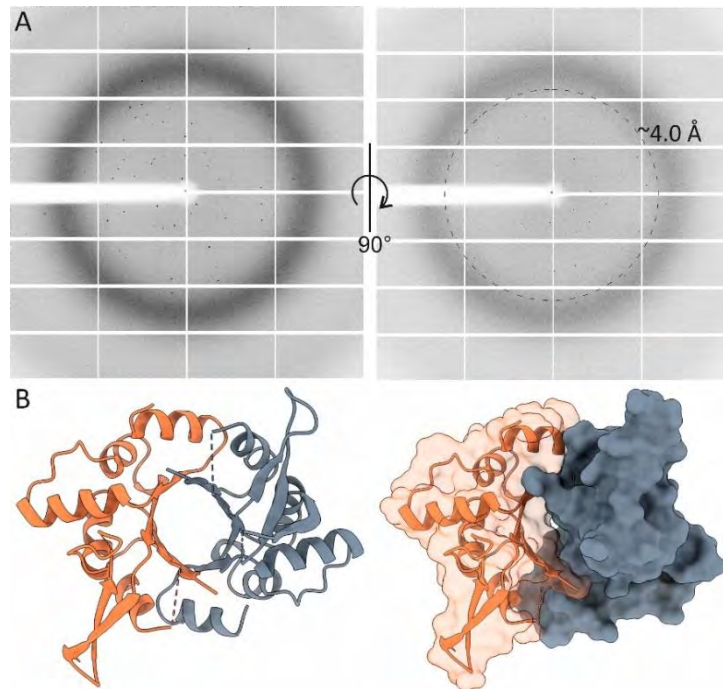
Figure 10. Panel A depicts diffractograms from two *Mt*HISN4-Nt48 crystal orientations. The dashed circle marks a resolution range of approx. 4 Å. Panel B depicts cartoon and surface representations of a $(\beta\alpha)_8$-barrel reconstructed from two C-terminal $(\beta\alpha)_4$-barrels after proteolytic digestion with thermolysin. The orange surface is 80% transparent. Dashed lines indicate unstructured loops.

Although this is a very surprising and interesting result, it has been already reported in the literature that (functional) $(\beta\alpha)_8$-barrels can merge into half-barrels, however not as a result of proteolytic digestion but after a specific construct design (Lang et al. 2000, Hocker et al. 2004, Sperl et al. 2013). Recently, I managed to obtain crystals of the *Mt*HISN4-Mut1-PrFAR complex, however due to a poor diffraction of crystals (~ 6 Å further optimization is required, I have prepared more *Mt*HISN4 mutants, hoping to resolve the structure in near future.
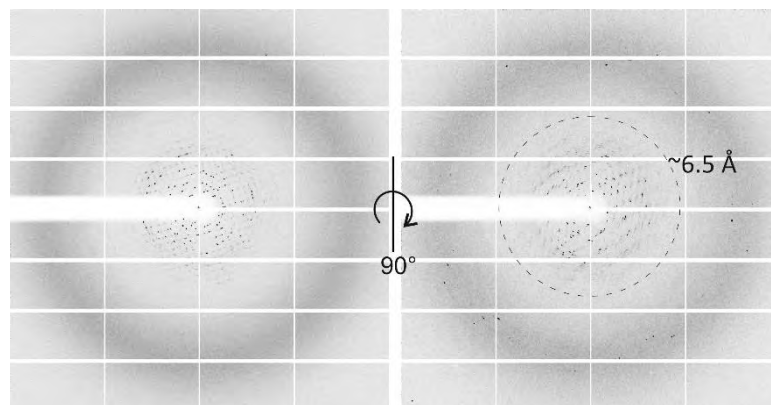
Figure 11. Diffraction pattern of the *Mt*HISN4-Mut1-PrFAR complex. Dashed circle marks resolution range of approx. 6.5 Å.

The need for sustainable food resources raised a question, whether it is possible to retain the enzymatic activity of HISN1 in the presence of histidine? To address this question, I designed a set of four mutants (Mut1-4) insensitive to histidine inhibition. The rationale behind this study was to explore the possibility of increasing free histidine levels in plants, that would allow them to chelate divalent metal ions from the soil. First, I have tested the activity of those mutants, using WT *Mt*HISN1 as a control. The mutants displayed 82-115% of the control setup activity (Figure 12A). Next, I have conducted kinetic measurements of all muteins in the presence of *Mt*HISN1-Mut1 retained 110% of the activity of the control (WT-*Mt*HISN1). *Mt*HISN1-Mut2 and Mut4 retained, respectively, 85 and 98% of the WT activity, proving their insensitivity to inhibition. Mut3 was only 5% more active, therefore is considered to be affected by histidine inhibition (Figure 12B). Such mutant activities could be additionally enhanced by discovering activator compounds, similar to TIH activating bacterial ATP-PRTs. Further studies should be conducted in genetically modified plant systems, by overexpressing the muteins or perhaps focused on identification of varieties that already possess such mutations.

A			B		
Mut1		115%	Mut1 + 100 μM His		110%
Mut2		91%	Mut2 + 100 μM His		85%
Mut3		82%	Mut3 + 100 μM His		5%
Mut4		98%	Mut4 + 100 μM His		98%
WT-*Mt* HISN1		100%	WT-*Mt* HISN1 + 100 μM His		0%
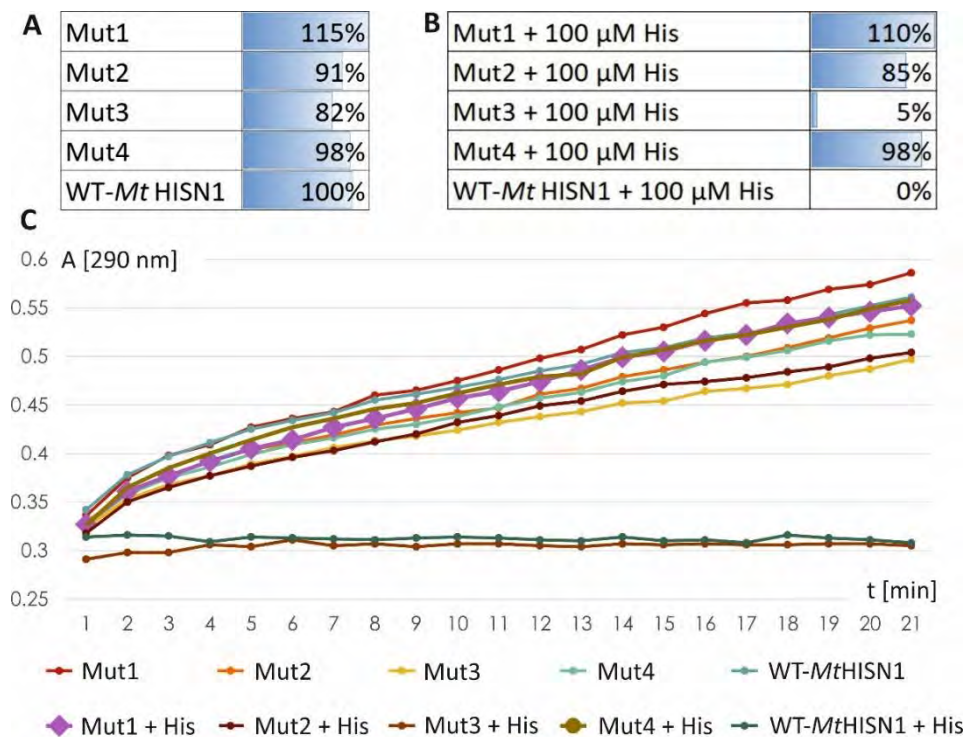
Figure 12. Kinetic measurements of *Mt*HISN1 mutants. Panel A shows activity levels of mutants without presence of histidine. Panel B shows activity levels in the presence of 100 μM histidine. Panel C shows changes in absorption at λ = 290 nm.

The results presented in the publications attached to this dissertation lay a strong foundation for the future herbicide design targeted at the HBP. The characterization of enzyme ¶V FRPSOH[HV ZLWK WKHLU RULEDVJUPHLDQWHVXOSUHRGG LGHQWLILFDWLRQ RI SRWHQWLDO QRYHO ELQGLQJ VLV NHFRPSRVOL EQQWD UURZLQJ GRZQ WKH SYRDWWLQFFKLHDPOLFF DLOR PSVRSX\ DQG WR WLGHERQWSRV\XQGV ZDLOOW KLHQGVLHQQWKLLDSVWUXFRUUHHVFW KRZ RQO\ XS WR WKH VWDJH RI FRPSRXQG LGHQWLILFDWLRQ V\QWKHVLV RI WKRVH FRPSRXQGV WR WHVW*in vitro*WKHLU L plants. Consequently, rational herbicide design could be applied, by a hit-to-lead optimization to characterize the most potent candidates. To successfully apply a potential herbicide to a plant, initial *in cellulo* and *in vivo* assays should be conducted to evaluat H WKH FRPSRXQGV¶ SHQHWLUQDFWHLORRO URSODVWV ZKHUH WKH +%3 RFFXUV

12. References

Abdo, M. R., Joseph, P., Boigegrain, R. A., Montero, J. L., Kohler, S. and Winum, J. Y. (**2008**). "*Brucella suis* histidinol dehydrogenase: synthesis and inhibition studies of substituted N-L-histidinylphenylsulfonyl hydrazide." Journal of Enzyme Inhibition and Medicinal Chemistry 23(3): 357-361.

Abraham, M. J., Murtola, T., Schulz, R., Pall, S., Smith, J. C., Hess, B. and Lindahl, E. (**2015**). "GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers." SoftwareX(1-2): 19-25.

Adams, E. (**1954**). "The Enzymatic Synthesis of Histidine from Histidinol." Journal of Biological Chemistry 209(2): 829-846.

Agirre, J., Atanasova, M., Bagdonas, H., Ballard, C. B., Baslé, A., Beilsten-Edmands, J., Borges, R. J., Brown, D. G., Burgos-Mármol, J. J., Berrisford, J. M., Bond, P. S., Caballero, I., Catapano, L., Chojnowski, G., Cook, A. G., Cowtan, K. D., Croll, T. I., Debreczeni, J. É., Devenish, N. E., Dodson, E. J., Drevon, T. R., Emsley, P., Evans, G., Evans, P. R., Fando, M., Foadi, J., Fuentes-Montero, L., Garman, E. F., Gerstel, M., Gildea, R. J., Hatti, K., Hekkelman, M. L., Heuser, P., Hoh, S. W., Hough, M. A., Jenkins, H. T., Jiménez, E., Joosten, R. P., Keegan, R. M., Keep, N., Krissinel, E. B., Kolenko, P., Kovalevskiy, O., Lamzin, V. S., Lawson, D. M., Lebedev, A. A., Leslie, A. G. W., Lohkamp, B., Long, F., Maly, M., McCoy, A. J., McNicholas, S. J., Medina, A., Millán, C., Murray, J. W., Murshudov, G. N., Nicholls, R. A., Noble, M. E. M., Oeffner, R., Pannu, N. S., Parkhurst, J. M., Pearce, N., Pereira, J., Perrakis, A., Powell, H. R., Read, R. J., Rigden, D. J., Rochira, W., Sammito, M., Rodríguez, F. S., Sheldrick, G. M., Shelley, K. L., Simkovic, F., Simpkin, A. J., Skubak, P., Sobolev, E., Steiner, R. A., Stevenson, K., Tews, I., Thomas, J. M. H., Thorn, A., Valls, J. T., Uski, V., Usón, I., Vagin, A., Velankar, S., Vollmar, M., Walden, H., Waterman, D., Wilson, K. S., Winn, M. D., Winter, G., Wojdyr, M. and Yamashita, K. (**2023**). "The CCP4 suite: integrative software for macromolecular crystallography." Acta Crystallographica Section D: Biological Crystallography 79: 449-461.

Ahangar, M. S., Vyas, R., Nasir, N. and Biswal, B. K. (**2013**). "Structures of native, substrate-bound and inhibited forms of *Mycobacterium tuberculosis* imidazoleglycerol-phosphate dehydratase." Acta Crystallographica Section D: Biological Crystallography 69: 2461-2467.

Akashi, H. and Gojobori, T. (**2002**). "Metabolic efficiency and amino acid composition in the proteomes of *Escherichia coli* and *Bacillus subtilis*." Proceedings of the National Academy of Sciences of the United States of America 99(6): 3695-3700.

Alifano, P., Fani, R., Lio, P., Lazcano, A., Bazzicalupo, M., Carlomagno, M. S. and Bruni, C. B. (**1996**). "Histidine biosynthetic pathway and genes: structure, regulation, and evolution." Microbiological Reviews 60(1): 44-69.

Ames, B. N. (**1957a**). "The Biosynthesis of Histidine - D-Erythro-Imidazoleglycerol Phosphate Dehydrase." Journal of Biological Chemistry 228(1): 131-143.

Ames, B. N. (**1957b**). "The Biosynthesis of Histidine - L-Histidinol Phosphate Phosphatase." Journal of Biological Chemistry 226(2): 583-593.

Ames, B. N. and Mitchell, H. K. (**1955**). "The Biosynthesis of Histidine - Imidazoleglycerol Phosphate, Imidazoleacetol Phosphate, and Histidinol Phosphate." Journal of Biological Chemistry 212(2): 687-696.

Babaahmadifooladi, M., Jacxsens, L., De Meulenaer, B. and Du Laing, G. (**2020**). "Nickel in foods sampled on the Belgian market: identification of potential contamination sources." Food Additives and Contaminants: Part A 37(4): 607-621.

Bakatula, E. N., Cukrowska, E. M., Weiersbye, I. M., Mihaly-Cozmuta, L. and Tutu, H. (**2014**). "Removal of toxic elements from aqueous solution using bentonite modified with L-histidine." Water Science and Technology 70(12): 2022-2030.

Baker, N. A., Sept, D., Joseph, S., Holst, M. J. and McCammon, J. A. (**2001**). "Electrostatics of nanosystems: Application to microtubules and the ribosome." Proceedings of the National Academy of Sciences of the United States of America 98(18): 10037-10041.

BIOVIA, D. S. (**2023**). "Discovery Studio Visualizer v24.1.0.23298."

Bisson, C., Britton, K. L., Sedelnikova, S. E., Rodgers, H. F., Eadsforth, T. C., Viner, R. C., Hawkes, T. R., Baker, P. J. and Rice, D. W. (**2015**). "Crystal Structures Reveal that the Reaction Mechanism of Imidazoleglycerol-Phosphate Dehydratase Is Controlled by Switching Mn(II) Coordination." Structure 23(7): 1236-1245.

Bowers, K. J., Chow, E., Xu, H., Dror, R. O., Eastwood, M. P., Gregersen, B. A., Klepeis, J. L., Kolossvary, I., Moraes, M. A., Sacerdoti, F. D., Salmon, J. K., Shan, Y. and Shaw, D. E. (**2006**). Scalable algorithms for molecular dynamics simulations on commodity clusters. Proceedings of the ACM/IEEE Conference on Supercomputing (SC06), Tampa, Florida, USA.

Brenner, C., Bieganowski, P., Pace, H. C. and Huebner, K. (**1999**). "The histidine triad superfamily of nucleotide-binding proteins." Journal of Cellular Physiology 181(2): 179-187.

Brenner, M. and Ames, B. N. (**1971**). The histidine operon and its regulation. Metabolic Pathways. G. D.M. New York, Academic Press. V**:** 349-387.

Brilli, M. and Fani, R. (**2004**). "The origin and evolution of eucaryal *HIS7* genes: from metabolon to bifunctional proteins?" Gene 339: 149-160.

Brilli, M. F., R. (**2004**). "Molecular Evolution of *hisB* Genes." Journal of Molecular Evolution 58(2): 225-237.

Broadhurst, C. L. and Chaney, R. L. (**2016**). "Growth and Metal Accumulation of an *Alyssum murale* Nickel Hyperaccumulator Ecotype Co-cropped with *Alyssum montanum* and Perennial Ryegrass in Serpentine Soil." Frontiers in Plant Science 7(451).

Burks, D., Azad, R., Wen, J. Q. and Dickstein, R. (**2018**). "The *Medicago truncatula* Genome: Genomic Data Availability." Functional Genomics in Medicago Truncatula: Methods and Protocols 1822: 39-59.

Burley, S. K., Berman, H. M., Bhikadiya, C., Bi, C. X., Chen, L., Di Costanzo, L., Christie, C., Duarte, J. M., Dutta, S., Feng, Z. K., Ghosh, S., Goodsell, D. S., Green, R. K., Guranovic, V., Guzenko, D., Hudson, B. P., Liang, Y. H., Lowe, R., Peisach, E., Periskova, I., Randle, C., Rose, A., Sekharan, M., Shao, C. H., Tao, Y. P., Valasatava, Y., Voigt, M., Westbrook, J., Young, J., Zardecki, C., Zhuravleva, M., Kurisu, G., Nakamura, H., Kengaku, Y., Cho, H., Sato, J., Kim, J. Y., Ikegawa, Y., Nakagawa, A., Yamashita, R., Kudou, T., Bekker, G. J., Suzuki, H., Iwata, T., Yokochi, M., Kobayashi, N., Fujiwara, T., Velankar, S., Kleywegt, G. J., Anyango, S., Armstrong, D. R., Berrisford, J. M., Conroy, M. J., Dana, J. M., Deshpande, M., Gane, P., Gáborová, R., Gupta, D., Gutmanas, A., Koca, J., Mak, L., Mir, S., Mukhopadhyay, A., Nadzirin, N., Nair, S., Patwardhan, A., Paysan-Lafosse, T., Pravda, L., Salih, O., Sehnal, D., Varadi, M., Vareková, R., Markley, J. L., Hoch, J. C., Romero, P. R., Baskaran, K., Maziuk, D., Ulrich, E. L., Wedell, J. R., Yao, H. Y., Livny, M., Ioannidis, Y. E., Consortium, w. and Japan, P. D. B. (**2019**). "Protein Data Bank: the single global archive for 3D macromolecular structure data." Nucleic Acids Research 47(D1): D520-D528.

Campbell, S. A. (**1999**). The second enzyme of Histidine Biosynthesis from Arabidopsis thaliana. PhD, University of Glasgow.

Carugo, O. (**2008**). "Amino acid composition and protein dimension." Protein Science 17(12): 2187-2191.

Celedon, J. M. and Cline, K. (**2013**). "Intra-plastid protein trafficking: How plant cells adapted prokaryotic mechanisms to the eukaryotic condition." Biochimica Et Biophysica Acta 1833(2): 341-351.

Chang, A., Jeske, L., Ulbrich, S., Hofmann, J., Koblitz, J., Schomburg, I., Neumann-Schaal, M., Jahn, D. and Schomburg, D. (**2021**). "BRENDA, the ELIXIR core data resource in 2021: new developments and updates." Nucleic Acids Research 49(D1): D498-D508.

Chioccioli, S., Del Duca, S., Vassallo, A., Castronovo, L. M. and Fani, R. (**2020**). "Exploring the role of the histidine biosynthetic *hisF* gene in cellular metabolism and in the evolution of (ancestral) genes: from LUCA to the extant (micro)organisms." Microbiological Research 240(126555).

Chovancova, E., Pavelka, A., Benes, P., Strnad, O., Brezovsky, J., Kozlikova, B., Gora, A., Sustr, V., Klvana, M., Medek, P., Biedermannova, L., Sochor, J. and Damborsky, J. (**2012**). "CAVER 3.0: a tool for the analysis of transport pathways in dynamic protein structures." PLOS Computational Biology 8(10): e1002708.

D'Ordine, R. L., Linger, R. S., Thai, C. J. and Davisson, V. J. (**2012**). "Catalytic zinc site and mechanism of the metalloenzyme PR-AMP cyclohydrolase." Biochemistry 51(29): 5791-5803.

De Bruijn, F. J. (**2020**). The Model Legume *Medicago truncatula*, Wiley Blackwell.

Del Duca, S., Riccardi, C., Vassallo, A., Fontana, G., Castronovo, L. M., Chioccioli, S. and Fani, R. (**2021**). "The Histidine Biosynthetic Genes in the Superphylum Bacteroidota-Rhodothermota-Balneolota-Chlorobiota: Insights into the Evolution of Gene Structure and Organization." Microorganisms 9(7).

Diepgen, T. L., Ofenloch, R. F., Bruze, M., Bertuccio, P., Cazzaniga, S., Coenraads, P. J., Elsner, P., Goncalo, M., Svensson, A. and Naldi, L. (**2016**). "Prevalence of contact allergy in the general population in different European regions." British Journal of Dermatology 174(2): 319-329.

Dietl, A. M., Amich, J., Leal, S., Beckmann, N., Binder, U., Beilhack, A., Pearlman, E. and Haas, H. (**2016**). "Histidine biosynthesis plays a crucial role in metal homeostasis and virulence of *Aspergillus fumigatus*." Virulence 7(4): 465-476.

Dolinsky, T. J., Nielsen, J. E., McCammon, J. A. and Baker, N. A. (**2004**). "PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations." Nucleic Acids Research 32: W665-W667.

Due, A. V., Kuper, J., Geerlof, A., von Kries, J. P. and Wilmanns, M. (**2011**). "Bisubstrate specificity in histidine/tryptophan biosynthesis isomerase from by active site metamorphosis." Proceedings of the National Academy of Sciences of the United States of America 108(9): 3554-3559.

El Malki, F., Jacobs, M. (**2001**). "Molecular characterization and expression study of a histidine auxotrophic mutant (his1-) of *Nicotiana plumbaginifolia*." Plant Molecular Biology 45(2): 191-199.

Emanuelsson, O., Nielsen, H., Brunak, S. and von Heijne, G. (**2000**). "Predicting subcellular localization of proteins based on their N-terminal amino acid sequence." Journal of Molecular Biology 300(4): 1005-1016.

Emsley, P., Lohkamp, B., Scott, W. G. and Cowtan, K. (**2010**). "Features and development of Coot." Acta Crystallographica Section D: Biological Crystallography 66(Pt 4): 486-501.

EU. (**2023**). "Commission Implementing Regulation (EU) 2023/2660." from https://eur-lex.europa.eu/eli/reg_impl/2023/2660/oj.

Fani, R. (**2004**). Gene duplication and gene loading. Microbial evolution: gene establishment, survival, and exchange. Washington, DC, ASM.

Fani, R. (**2012**). "The Origin and Evolution of Metabolic Pathways: *Why* and *How* did Primordial Cells Construct Metabolic Routes?" Evolution: Education and Outreach 5: 367-381.

Fani, R., Brilli, M., Fondi, M. and Lio, P. (**2007**). "The role of gene fusions in the evolution of metabolic pathways: the histidine biosynthesis case." BMC Ecology and Evolution 7 Suppl 2: S4.

Fani, R., Brilli, M. and Liò, P. (**2005**). "The origin and evolution of operons: The piecewise building of the proteobacterial histidine operon." Journal of Molecular Evolution 60(3): 378-390.

Fani, R. and Fondi, M. (**2009**). "Origin and evolution of metabolic pathways." Physics of Life Reviews 6(1): 23-52.

Fani, R., Lio, P. and Lazcano, A. (**1995**). "Molecular evolution of the histidine biosynthetic pathway." Journal of Molecular Evolution 41(6): 760-774.

Fink, G. R. and Martin, R. G. (**1967**). "Translation and polarity in the histidine operon. II. Polarity in the histidine operon." Journal of Molecular Biology 30(1): 97-107.

Fondi, M., Emiliani, G., Liò, P., Gribaldo, S. and Fani, R. (**2009**). "The Evolution of Histidine Biosynthesis in Archaea: Insights into the Genes Structure and Organization in LUCA." Journal of Molecular Evolution 69(5): 512-526.

Ford, G. C., Eichele, G., Jansonius, J.N. (**1980**). "Three-dimensional structure of a pyridoxal-phosphate-dependent enzyme, mitochondrial aspartate aminotransferase." Proceedings of the National Academy of Sciences of the United States of America 77: 2559-2563.

Forlani, G., Mangiagalli, A., Nielsen, E. and Suardi, C. M. (**1999**). "Degradation of the phosphonate herbicide glyphosate in soil: evidence for a possible involvement of unculturable microorganisms." Soil Biology & Biochemistry 31(7): 991-997.

Fujimori, K. and Ohta, D. (**1998a**). "An Arabidopsis cDNA encoding a bifunctional glutamine amidotransferase/cyclase suppresses the histidine auxotrophy of a Saccharomyces cerevisiae his7 mutant." Febs Letters 428(3): 229-234.

Fujimori, K. and Ohta, D. (**1998b**). "Isolation and characterization of a histidine biosynthetic gene in *Arabidopsis* encoding a polypeptide with two separate domains for phosphoribosyl-ATP pyrophosphohydrolase and phosphoribosyl-AMP cyclohydrolase." Plant Physiology 118(1): 275-283.

Fujimori, K., Tada, S., Kanai, S. and Ohta, D. (**1998c**). "Molecular cloning and characterization of the gene encoding N′-[(5′-phosphoribosyl)-formimino]-5-aminoimidazole-4-carboxamide ribonucleotide (BBM II) isomerase from." Molecular and General Genetics 259(2): 216-223.

Furukawa, A., Oikawa, S., Harada, K., Sugiyama, H., Hiraku, Y., Murata, M., Shimada, A. and Kawanishi, S. (**2010**). "Oxidatively generated DNA damage induced by 3-amino-5-mercapto-1,2,4-triazole, a metabolite of carcinogenic amitrole." Mutation Research 694(1-2): 7-12.

Glynn, S. E., Baker, P. J., Sedelnikova, S. E., Levy, C. W., Rodgers, H. F., Blank, J., Hawkes, T. R. and Rice, D. W. (**2005**). "Purification, crystallization and preliminary crystallographic analysis of *Arabidopsis thaliana* imidazoleglycerol-phosphate dehydratase." Acta Crystallographica Section F: Structural Biology and Crystallization Communications 61(Pt 8): 776-778.

Greiner, B. A., Marshall, N. M., Sarjeant, A. A. N. and McLauchlan, C. C. (**2007**). "Imidazole-based nickel(II) and cobalt(II) coordination complexes for potential use as models for histidine containing metalloproteins." Inorganica Chimica Acta 360(10): 3132-3140.

Guerin A, L. C., Herger A, Renggli D, Sotiropoulos AG, Kadler G, Hou X, Schaufelberger M, Meyer C, Wicker T, Bigler L, Ringli C (**2024**). "Histidine limitation causes alteration in the TOR network and plant development.". bioRxiv.

Gupta, M., Prasad, Y., Sharma, S. K. and Jain, C. K. (**2017**). "Identification of Phosphoribosyl-AMP cyclohydrolase, as drug target and its inhibitors in *Brucella melitensis* bv. 1 16M using metabolic pathway analysis." Journal of Biomolecular Structure & Dynamics 35(2): 287-299.

Guyer, D., Patton, D. and Ward, E. (**1995**). "Evidence for Cross-Pathway Regulation of Metabolic Gene-Expression in Plants." Proceedings of the National Academy of Sciences of the United States of America 92(11): 4997-5000.

Guyton, K. Z., Loomis, D., Grosse, Y., El Ghissassi, F., Benbrahim-Tallaa, L., Guha, N., Scoccianti, C., Mattock, H., Straif, K. and International Agency for Research on Cancer Monograph Working Group, I. (**2015**). "Carcinogenicity of tetrachlorvinphos, parathion, malathion, diazinon, and glyphosate." Lancet Oncology 16(5): 490-491.

Hawkes, T. R., Thomas, P. G., Edwards, L. S., Rayner, S. J., Wilkinson, K. W. and Rice, D. W. (**1995**). "Purification and Characterization of the Imidazoleglycerol-Phosphate Dehydratase of *Saccharomyces cerevisiae* from Recombinant *Escherichia coli*." Biochemical Journal 306: 385-397.

Heap, I. (**2024**). "The international survey of herbicide resistant weeds." from www.weedscience.com.

Hedin, S. G. (**1897**). "Zur Kenntniss der Spaltungsproducte der Proteïnkörper." 22(2): 191-196.

Heinemann, B. and Hildebrandt, T. M. (**2021**). "The role of amino acid metabolism in signaling and metabolic adaptation to stress-induced energy deficiency in plants." Journal of Experimental Botany 72(13): 4634-4645.

Hendawy, S. F., Abouziena, H. F., Abd El-Razik, T. M., Amer, H. M. and Hussein, M. S. (**2019**). "Winter weeds and its control in the medicinal plants in Egypt: a survey study." Egyptian Pharmaceutical Journal 18(1): 16-26.

Henn-Sax, M., Thoma, R., Schmidt, S., Hennig, M., Kirschner, K. and Sterner, R. (**2002**). "Two $(\beta\alpha)_8$-barrel enzymes of histidine and tryptophan biosynthesis have similar reaction mechanisms and common strategies for protecting their labile substrates." Biochemistry 41(40): 12032-12042.

Hinnebusch, A. G. (**1990**). "Transcriptional and Translational Regulation of Gene-Expression in the General Control of Amino-Acid Biosynthesis in *Saccharomyces cerevisiae*." Progress in Nucleic Acid Research and Molecular Biology 38: 195-240.

Hocker, B., Claren, J. and Sterner, R. (**2004**). "Mimicking enzyme evolution by generating new $(\beta\alpha)_8$-barrels from $(\beta\alpha)_4$-half-barrels." Proceedings of the National Academy of Sciences of the United States of America 101(47): 16448-16453.

Holinski, A. (**2017**). Analysis of complex stability and allosteric interaction in the imidazole glycerol phosphate synthase complex. PhD, University of Regensburg.

Holliday, G. L., Fischer, J. D., Mitchell, J. B. O. and Thornton, J. M. (**2011**). "Characterizing the complexity of enzymes on the basis of their mechanisms and structures with a bio-computational analysis." Febs Journal 278(20): 3835-3845.

Horowitz, N. H. (**1945**). "On the Evolution of Biochemical Syntheses." Proceedings of the National Academy of Sciences of the United States of America 31(6): 153-157.

Hrbácková, M., Dvorák, P., Takác, T., Tichá, M., Luptovciak, I., Samajová, O., Ovecka, M. and Samaj, J. (**2020**). "Biotechnological Perspectives of Omics and Genetic Engineering Methods in Alfalfa." Frontiers in Plant Science 11(592).

Hupfeld, E. (**2020**). Linkage between catalytic mechanism and conformational dynamics in $(\beta\alpha)_8$-barrel enzymes at the example of Imidazole Glycerol Phosphate Synthase. PhD, University of Regensburg.

Ingle, R. A., Mugford, S. T., Rees, J. D., Campbell, M. M. and Smith, J. A. C. (**2005**). "Constitutively high expression of the histidine biosynthetic pathway contributes to nickel tolerance in hyperaccumulator plants." Plant Cell 17(7): 2089-2106.

Jensen, R. A. (**1976**). "Enzyme Recruitment in Evolution of New Function." Annual Review of Microbiology 30: 409-425.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Zídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A. W., Kavukcuoglu, K., Kohli, P. and Hassabis, D. (**2021**). "Highly accurate protein structure prediction with AlphaFold." Nature 596(7873): 583-589.

Jurcik, A., Bednar, D., Byska, J., Marques, S. M., Furmanova, K., Daniel, L., Kokkonen, P., Brezovsky, J., Strnad, O., Stourac, J., Pavelka, A., Manak, M., Damborsky, J. and Kozlikova, B. (**2018**). "CAVER Analyst 2.0: analysis and visualization of channels and tunnels in protein structures and molecular dynamics trajectories." Bioinformatics 34(20): 3586-3588.

Jürgens, C., Strom, A., Wegener, D., Hettwer, S., Wilmanns, M. and Sterner, R. (**2000**). "Directed evolution of a $(\beta\alpha)_8$-barrel enzyme to catalyze related reactions in two different metabolic pathways." Proceedings of the National Academy of Sciences of the United States of America 97(18): 9925-9930.

Kabsch, W. (**2010**). "XDS." Acta Crystallographica Section D: Biological Crystallography 66: 125-132.

Khajeaian, P. (**2022**). Exploring the Evolution and Function of the first Enzyme in Histidine Biosynthesis. PhD, Victoria University of Wellington.

Kim, Y., Babnigg, G., Jedrzejczak, R., Eschenfeldt, W. H., Li, H., Maltseva, N., Hatzos-Skintges, C., Gu, M., Makowska-Grzyska, M., Wu, R., An, H., Chhor, G. and Joachimiak, A. (**2011**). "High-throughput protein purification and quality assessment for crystallization." Methods 55(1): 12-28.

Kirschning, A. (**2022**). "On the Evolutionary History of the Twenty Encoded Amino Acids." Chemistry - A European Journal 28(55).

Klem, T. J. and Davisson, V. J. (**1993**). "Imidazole Glycerol Phosphate Synthase - the Glutamine Amidotransferase in Histidine Biosynthesis." Biochemistry 32(19): 5177-5186.

Klock, H. E. and Lesley, S. A. (**2009**). "The Polymerase Incomplete Primer Extension (PIPE) method applied to high-throughput cloning and site-directed mutagenesis." Methods in Molecular Biology 498: 91-103.

Koo, D., Sathishraj, R., Friebe, B. and Gill, B. S. (**2022**). "Deciphering the Mechanism of Glyphosate Resistance in *Amaranthus palmeri* by Cytogenomics." Cytogenetic and Genome Research 161 (12): 578–584.

Kossel, A. (**1897**). "Ueber die basischen Stoffe des Zellkerns." Hoppe-Seyler´s Zeitschrift für physiologische Chemie 22(2): 176-187.

Kotloff, K. L., Riddle, M. S., Platts-Mills, J. A., Pavlinac, P. and Zaidi, A. K. M. (**2018**). "Shigellosis." Lancet 391(10122): 801-812.

Kowiel, M., Jaskolski, M. and Dauter, Z. (**2014**). "ACHESYM: an algorithm and server for standardized placement of macromolecular models in the unit cell." Acta Crystallographica Section D: Biological Crystallography 70: 3290-3298.

Krissinel, E. and Henrick, K. (**2007**). "Inference of macromolecular assemblies from crystalline state." Journal of Molecular Biology 372(3): 774-797.

Kulis-Horn, R. K., Persicke, M. and Kalinowski, J. (**2014**). "Histidine biosynthesis, its regulation and biotechnological application in *Corynebacterium glutamicum*." Microbial Biotechnology 7(1): 5-25.

Kumar, D., Jha, B., Bhatia, I., Ashraf, A., Dwivedy, A. and Biswal, B. K. (**2022**). "Characterization of a triazole scaffold compound as an inhibitor of *Mycobacterium tuberculosis* imidazoleglycerol-phosphate dehydratase." Proteins 90(1): 3-17.

Kuper, J., Doenges, C. and Wilmanns, M. (**2005**). "Two-fold repeated $(\beta\alpha)_4$ half-barrels may provide a molecular tool for dual substrate specificity." Embo Reports 6(2): 134-139.

Kuplinska, A. and Rzad, K. (**2021**). "Molecular targets for antifungals in amino acid and protein biosynthetic pathways." Amino Acids 53(7): 961-991.

Lang, D., Thoma, R., Henn-Sax, M., Sterner, R. and Wilmanns, M. (**2000**). "Structural evidence for evolution of the β/α barrel scaffold by gene duplication and fusion." Science 289(5484): 1546-1550.

Le Deunff, E., Beauclair, P., Deleu, C. and Lecourt, J. (**2019**). "Inhibition of Aminotransferases by Aminoethoxyvinylglycine Triggers a Nitrogen Limitation Condition and Deregulation of Histidine Homeostasis That Impact Root and Shoot Development and Nitrate Uptake." Frontiers in Plant Science 10(1387).

Lewis, E. B. (**1951**). "Pseudoallelism and Gene Evolution." Cold Spring Harbor Symposia on Quantitative Biology 16: 159-174.

Liebschner, D., Afonine, P. V., Baker, M. L., Bunkoczi, G., Chen, V. B., Croll, T. I., Hintze, B., Hung, L. W., Jain, S., McCoy, A. J., Moriarty, N. W., Oeffner, R. D., Poon, B. K., Prisant, M. G., Read, R. J., Richardson, J. S., Richardson, D. C., Sammito, M. D., Sobolev, O. V., Stockwell, D. H., Terwilliger, T. C., Urzhumtsev, A. G., Videau, L. L., Williams, C. J. and Adams, P. D. (**2019**). "Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix." Acta Crystallographica Section D: Biological Crystallography 75: 861-877.

Liebschner, D., Afonine, P. V., Moriarty, N. W., Poon, B. K., Sobolev, O. V., Terwilliger, T. C. and Adams, P. D. (**2017**). "Polder maps: improving OMIT maps by excluding bulk solvent." Acta Crystallographica Section D: Biological Crystallography 73: 148-157.

List, F. (**2009**). Die Imidazolglycerinphosphat-Synthase aus *Thermotoga maritima*: Struktur, Regulation und Evolution einer Glutaminamidotransferase. PhD, University of Regensburg.

Liu, Y. S., Geng, J. C., Sha, X. Y., Zhao, Y. X., Hu, T. M. and Yang, P. Z. (**2019**). "Effect of Rhizobium Symbiosis on Low-Temperature Tolerance and Antioxidant Response in Alfalfa (*Medicago sativa* L.)." Frontiers in Plant Science 10(538).

Lohkamp, B. (**2003**). Structural Studies on Histidine Metabolism in Microorganisms. PhD, University of Glasgow.

Martin, R. G., Berberich, M.A., Ames, B.N., Davis, W.W., Goldberger, R.F., and Yourno, J.D. (**1971**). Enzymes and intermediates of histidine biosynthesis in *Salmonella typhimurium*. Methods in Enzymology: Metabolism of Amino Acids and Amines Part B. T. C. W. Tabor H. New York, Academic Press**:** 3-44.

Mccoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. and Read, R. J. (**2007**). "Phaser crystallographic software." Journal of Applied Crystallography 40: 658-674.

Miflin, B. (**1980**). Histidine biosynthesis. The biochemistry of plants: a comprehensive treatise, vol 5: Amino acids and derivatives. New York, Academic Press. 5**:** 533–541.

Miguel, M. A., Postma, J. A. and Lynch, J. P. (**2015**). "Phene Synergism between Root Hair Length and Basal Root Growth Angle for Phosphorus Acquisition." Plant Physiology 167(4): 1430-1439.

Miller, L. L. and Bale, W. F. (**1952**). "Factors Affecting Biosynthesis of Plasma Albumin and Globulin Fractions as Studied with the Aid of Lysine-Epsilon and Histidine-2-C-14." Federation Proceedings 11(1): 260-260.

Minson, A. C. (**1968**). Studies of the biosynthesis of histidine in *Neurospora crassa*. PhD, Australian National University

Mo, X. R., Zhu, Q. Y., Li, X., Li, J., Zeng, Q. N., Rong, H. L., Zhang, H. M. and Wu, P. (**2006**). "The hpa1 mutant of *Arabidopsis* reveals a crucial role of histidine homeostasis in root meristem maintenance." Plant Physiology 141(4): 1425-1435.

Moody, E. R. R., Alvarez-Carretero, S., Mahendrarajah, T. A., Clark, J. W., Betts, H. C., Dombrowski, N., Szantho, L. L., Boyle, R. A., Daines, S., Chen, X., Lane, N., Yang, Z., Shields, G. A., Szollosi, G. J., Spang, A., Pisani, D., Williams, T. A., Lenton, T. M. and Donoghue, P. C. J. (**2024**). "The nature of the last universal common ancestor and its impact on the early Earth system." Nature Ecology and Evolution.

Morris, E. C., Griffiths, M., Golebiowska, A., Mairhofer, S., Burr-Hersey, J., Goh, T., von Wangenheim, D., Atkinson, B., Sturrock, C. J., Lynch, J. P., Vissenberg, K., Ritz, K., Wells, D. M., Mooney, S. J. and Bennett, M. J. (**2017**). "Shaping 3D Root System Architecture." Current Biology 27(17): R919-R930.

Mueller-Harvey, I., Bee, G., Dohme-Meier, F., Hoste, H., Karonen, M., Kölliker, R., Lüscher, A., Niderkorn, V., Pellikaan, W. F., Salminen, J. P., Skot, L., Smith, L. M. J., Thamsborg, S. M., Totterdell, P., Wilkinson, I., Williams, A. R., Azuhnwi, B. N., Baert, N., Brinkhaus, A. G., Copani, G., Desrues, O., Drake, C., Engstrom, M., Fryganas, C., Girard, M., Huyen, N. T., Kempf, K., Malisch, C., Mora-Ortiz, M., Quijada, J., Ramsay, A., Ropiak, H. M. and Waghorn, G. C. (**2019**). "Benefits of Condensed Tannins in Forage Legumes Fed to Ruminants: Importance of Structure, Concentration, and Diet Composition." Crop Science 59(3): 861-885.

Muralla, R., Sweeney, C., Stepansky, A., Leustek, T. and Meinke, D. (**2007**). "Genetic dissection of histidine biosynthesis in *Arabidopsis*." Plant Physiology 144(2): 890-903.

Newton, M. S., Guo, X. H., Söderholm, A., Näsvall, J., Lundström, P., Andersson, D. I., Selmer, M. and Patrick, W. M. (**2017**). "Structural and functional innovations in the real-time evolution

of new (βα)$_8$ barrel enzymes." Proceedings of the National Academy of Sciences of the United States of America 114(18): 4727-4732.

Noutoshi, Y. I., T. Shinozaki, K. (**2005**). "ALBINO AND PALE GREEN 10 encodes BBMII isomerase involved in histidine biosynthesis in *Arabidopsis thaliana*." Plant and Cell Physiology 46(7): 1165-1172.

Ohno, S. (**1970**). Evolution by Gene Duplication. New York, Springer-Verlag.

Ohta, D., Fujimori, K., Mizutani, M., Nakayama, Y., Kunpaisal-Hashimoto, R., Munzer, S. and Kozaki, A. (**2000**). "Molecular cloning and characterization of ATP-phosphoribosyl transferase from *Arabidopsis*, a key enzyme in the histidine biosynthetic pathway." Plant Physiology 122(3): 907-914.

Oppenheim, F. G., Xu, T., Mcmillian, F. M., Levitz, S. M., Diamond, R. D., Offner, G. D. and Troxler, R. F. (**1988**). "Histatins, a Novel Family of Histidine-Rich Proteins in Human-Parotid Secretion - Isolation, Characterization, Primary Structure, and Fungistatic Effects on Candida-Albicans." Journal of Biological Chemistry 263(16): 7472-7477.

Oren, A. and Garrity, G. M. (**2021**). "Valid publication of the names of forty-two phyla of prokaryotes." International Journal of Systematic and Evolutionary Microbiology 71(10).

Parca, L., Ferre, F., Ausiello, G. and Helmer-Citterich, M. (**2013**). "Nucleos: a web server for the identification of nucleotide-binding sites in protein structures." Nucleic Acids Research 41(W1): W281-W285.

Persans, M. W., Yan, X. G., Patnoe, J. M. M. L., Krämer, U. and Salt, D. E. (**1999**). "Molecular dissection of the role of histidine in nickel hyperaccumulation in *Thlaspi goesingense* (Halacsy)." Plant Physiology 121(4): 1117-1126.

Petersen, L. N., Marineo, S., Mandala, S., Davids, F., Sewell, B. T. and Ingle, R. A. (**2010**). "The missing link in plant histidine biosynthesis: *Arabidopsis* myoinositol monophosphatase-like2 encodes a functional histidinol-phosphate phosphatase." Plant Physiology 152(3): 1186-1196.

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C. and Ferrin, T. E. (**2004**). "UCSF chimera - A visualization system for exploratory research and analysis." Journal of Computational Chemistry 25(13): 1605-1612.

Pettersen, E. F., Goddard, T. D., Huang, C. R. C., Meng, E. E. C., Couch, G. S., Croll, T. I., Morris, J. H. and Ferrin, T. E. (**2021**). "UCSF ChimeraX: Structure visualization for researchers, educators, and developers." Protein Science 30(1): 70-82.

Pisco, J. P., de Chiara, C., Pacholarz, K. J., Garza-Garcia, A., Ogrodowicz, R. W., Walker, P. A., Barran, P. E., Smerdon, S. J. and de Carvalho, L. P. S. (**2017**). "Uncoupling conformational states from activity in an allosteric enzyme." Nature Communications 8(203).

Punjani, A., Rubinstein, J. L., Fleet, D. J. and Brubaker, M. A. (**2017**). "cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination." Nature Methods 14(3): 290-296.

Rawson, S., Bisson, C., Hurdiss, D. L., Fazal, A., McPhillie, M. J., Sedelnikova, S. E., Baker, P. J., Rice, D. W. and Muench, S. P. (**2018**). "Elucidating the structural basis for differing enzyme inhibitor potency by cryo-EM." Proceedings of the National Academy of Sciences of the United States of America 115(8): 1795-1800.

Rawson, S. D. (**2016**). EM as a tool to study structure and function to guide inhibitor design. PhD, University of Leeds.

Rébora, K., Laloo, B. and Daignan-Fornier, B. (**2005**). "Revisiting purine-histidine cross-pathway regulation in *Saccharomyces cerevisiae*: A central role for a small molecule." Genetics 170(1): 61-70.

Reisinger, B. (**2013**). Mechanistic investigation of artificially designed, light regulation of naturally occurring, and characterization of ancestral (βα)$_8$-barrel enzymes. PhD, University of Regensburg.

Reizer, J., Buskirk, S., Bairoch, A., Reizer, A. and Saier, M. H., Jr. (**1994**). "A novel zinc-binding motif found in two ubiquitous deaminase families." Protein Science 3(5): 853-856.

Rose, R. J. (**2008**). "*Medicago truncatula* as a model for understanding plant interactions with other organisms, plant development and stress biology: past, present and future." Functional Plant Biology 35(4): 253-264.

Ruszkowski, M. (**2018**). "Guarding the gateway to histidine biosynthesis in plants: *Medicago truncatula* ATP-phosphoribosyltransferase in relaxed and tense states." Biochemical Journal 475: 2681-2697.

Ruszkowski, M. and Dauter, Z. (**2016**). "Structural Studies of *Medicago truncatula* Histidinol Phosphate Phosphatase from Inositol Monophosphatase Superfamily Reveal Details of Penultimate Step of Histidine Biosynthesis in Plants." Journal of Biological Chemistry 291(19): 9960-9973.

Ruszkowski, M. and Dauter, Z. (**2017**). "Structures of *Medicago truncatula* L-Histidinol Dehydrogenase Show Rearrangements Required for NAD$^+$ Binding and the Cofactor Positioned to Accept a Hydride." Scientific Reports 7: 10476.

Ruszkowski, M. and Forlani, G. (**2022**). "Deciphering the structure of *Arabidopsis thaliana* 5-*enol*-pyruvyl-shikimate-3-phosphate synthase: An essential step toward the discovery of novel inhibitors to supersede glyphosate." Computational and Structural Biotechnology Journal 20: 1494-1505.

Saika, H., Fruh, T., Iwasaki, G., Koizumi, S., Mori, I. and Hayakawa, K. (**1993**). "Synthesis of (*2r,3r*)-, (*2s,3s*)-, (*2r,3s*)- and (*2s,3r*)-Imidazole Glycerol Phosphates (IGP) - Substrates for IGP-Dehydratase (IGPD)." Bioorganic & Medicinal Chemistry Letters 3(10): 2129-2134.

Sakiroglu, M. and Ilhan, D. (**2021**). "*Medicago sativa* species complex: Revisiting the century-old problem in the light of molecular tools." Crop Science 61(2): 827-838.

Scheres, S. H. W. (**2012**). "RELION: Implementation of a Bayesian approach to cryo-EM structure determination." Journal of Structural Biology 180(3): 519-530.

Seraphin, B. (**1992**). "The HIT protein family: a new family of proteins present in prokaryotes, yeast and mammals." DNA Sequence 3(3): 177-179.

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B. and Ideker, T. (**2003**). "Cytoscape: a software environment for integrated models of biomolecular interaction networks." Genome Research 13(11): 2498-2504.

Sharma, P., Pandey, A. K., Udayan, A. and Kumar, S. (**2021**). "Role of microbial community and metal-binding proteins in phytoremediation of heavy metals from industrial wastewater." Bioresource Technology 326: 124750.

Sigel, A., Sigel, H., Sigel, R. (**2016**). The Alkali Metal Ions: Their Role for Life. Zurich, Springer.

Singer, A. C., Bell, T., Heywood, C. A., Smith, J. A. C. and Thompson, I. P. (**2007**). "Phytoremediation of mixed-contaminated soil using the hyperaccumulator plant *Alyssum*

lesbiacum: Evidence of histidine as a measure of phytoextractable nickel." Environmental Pollution 147(1): 74-82.

Sissler, M., Delorme, C., Bond, J., Ehrlich, S. D., Renault, P. and Francklyn, C. (**1999**). "An aminoacyl-tRNA synthetase paralog with a catalytic role in histidine biosynthesis." Proceedings of the National Academy of Sciences of the United States of America 96(16): 8985-8990.

Smith, D. W. E. and Ames, B. N. (**1965**). "Phosphoribosyladenosine Monophosphate an Intermediate in Histidine Biosynthesis." Journal of Biological Chemistry 240(7): 3056-3063.

Söderholm, A., Guo, X. H., Newton, M. S., Evans, G. B., Näsvall, J., Patrick, W. M. and Selmer, M. (**2015**). "Two-step ligand binding in a $(\beta\alpha)_8$ barrel enzyme: Substrate-bound structures shed new light on the catalytic cycle of HisA." Journal of Biological Chemistry 290(41): 24657-24668.

Song, Y. G., Lv, J., Ma, Z. Q. and Dong, W. (**2019**). "The mechanism of alfalfa (*Medicago sativa* L.) response to abiotic stress." Plant Growth Regulation 89(3): 239-249.

Sperl, J. M., Rohweder, B., Rajendran, C. and Sterner, R. (**2013**). "Establishing catalytic activity on an artificial $(\beta\alpha)_8$-barrel protein designed from identical half-barrels." FEBS Letters 587(17): 2798-2805.

Stepansky, A. and Leustek, T. (**2006**). "Histidine biosynthesis in plants." Amino Acids 30(2): 127-142.

Sterling, T. and Irwin, J. J. (**2015**). "ZINC 15-Ligand Discovery for Everyone." Journal of Chemical Information and Modeling 55(11): 2324-2337.

Swire, J. (**2007**). "Selection on synthesis cost affects interprotein amino acid usage in all three domains of life." Journal of Molecular Evolution 64(5): 558-571.

Tada, S., Volrath, S., Guyer, D., Scheidegger, A., Ryals, J., Ohta, D. and Ward, E. (**1994**). "Isolation and Characterization of cDNAs Encoding Imidazoleglycerolphosphate Dehydratase from *Arabidopsis thaliana*." Plant Physiology 105(2): 579-583.

Tamura, K., Stecher, G. and Kumar, S. (**2021**). "MEGA11 Molecular Evolutionary Genetics Analysis Version 11." Molecular Biology and Evolution 38(7): 3022-3027.

Tang, H. B., Krishnakumar, V., Bidwell, S., Rosen, B., Chan, A. N., Zhou, S. G., Gentzbittel, L., Childs, K. L., Yandell, M., Gundlach, H., Mayer, K. F. X., Schwartz, D. C. and Town, C. D. (**2014**). "An improved genome release (version Mt4.0) for the model legume *Medicago truncatula*." BMC Genomics 15(312).

Thomson, C. M. (**2019**). Allostery and Inhibition of *Psychrobacter arcticus* Adenosine 5'-triphosphate Phosphoribosyltransferase. PhD, University of St Andrews.

Ticebaldwin, K., Fink, G. R. and Arndt, K. T. (**1989**). "Bas1 Has a Myb Motif and Activates His4 Transcription Only in Combination with Bas2." Science 246(4932): 931-935.

Tipton, K. (**2018**). "Translocases (EC 7): A new EC Class." from https://www.enzyme-database.org/news.php.

Trott, O. and Olson, A. J. (**2010**). "AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading." Journal of Computational Chemistry 31(2): 455-461.

Turtaut, F., Lopez, M., Ouahrani-Bettache, S., Kohler, S. and Winum, J. Y. (**2014**). "Oxo- and thiooxo-imidazo[1,5-c]pyrimidine molecule library: beyond their interest in inhibition of

Brucella suis histidinol dehydrogenase, a powerful protection tool in the synthesis of histidine analogues." Bioorganic & Medicinal Chemistry Letters 24(21): 5008-5010.

Usuda, H. (**1988**). "Adenine Nucleotide Levels, the Redox State of the NADP System, and Assimilatory Force in Nonaqueously Purified Mesophyll Chloroplasts from Maize Leaves under Different Light Intensities." Plant Physiology 88(4): 1461-1468.

Valerius, O. (**2001**). Chromatin Structure and Regulation of Gene Expression at the Histidine/Adenine Branch Point in Yeast and Aspergillus. PhD, University of Göttingen.

Vallee, B. L. and Auld, D. S. (**1993**). "Cocatalytic Zinc Motifs in Enzyme Catalysis." Proceedings of the National Academy of Sciences of the United States of America 90(7): 2715-2718.

Waite, D. W., Chuvochina, M., Pelikan, C., Parks, D. H., Yilmaz, P., Wagner, M., Loy, A., Naganuma, T., Nakai, R., Whitman, W. B., Hahn, M. W., Kuever, J. and Hugenholtz, P. (**2020**). "Proposal to reclassify the proteobacterial classes Deltaproteobacteria and Oligoflexia, and the phylum Thermodesulfobacteria into four phyla reflecting major functional capabilities." International Journal of Systematic and Evolutionary Microbiology 70(11): 5972-6016.

Wang, Y. N., Zhang, F., Nie, Y., Shang, G. J. and Zhang, H. Q. (**2019**). "Structural analysis of *Shigella flexneri* bi-functional enzyme HisIE in histidine biosynthesis." Biochemical and Biophysical Research Communications 516(2): 540-545.

Weber, A. L. and Miller, S. L. (**1981**). "Reasons for the occurrence of the twenty coded protein amino acids." Journal of Molecular Evolution 17(5): 273-284.

Wellems, T. E. and Howard, R. J. (**1986**). "Homologous genes encode two distinct histidine-rich proteins in a cloned isolate of *Plasmodium falciparum*." Proceedings of the National Academy of Sciences of the United States of America 83(16): 6065-6069.

Wiater, A., Hulanicka, D. and Klopotowski, T. (**1971a**). "Structural requirements for inhibition of yeast imidazoleglycerol phosphate dehydratase by triazole and anion inhibitors." Acta Biochimica Polonica 18(3): 289-297.

Wiater, A., Klopotowski, T. and Bagdasarian, G. (**1971b**). "Synergistic inhibition of plant imidazoleglycerol phosphate dehydratase by aminotriazole and phosphate." Acta Biochimica Polonica 18(3): 309-314.

Wiater, A., Krajewska-Grynkiewicz, K. and Klopotowski, T. (**1971c**). "Histidine biosynthesis and its regulation in higher plants." Acta Biochimica Polonica 18(3): 299-307.

Williams, C. J., Headd, J. J., Moriarty, N. W., Prisant, M. G., Videau, L. L., Deis, L. N., Verma, V., Keedy, D. A., Hintze, B. J., Chen, V. B., Jain, S., Lewis, S. M., Arendall, W. B., Snoeyink, J., Adams, P. D., Lovell, S. C., Richardson, J. S. and Richardson, D. C. (**2018**). "MolProbity: More and better reference data for improved all-atom structure validation." Protein Science 27(1): 293-315.

Winkler, M. E. (**1987**). Biosynthesis of Histidine *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology. I. J. L. Neidhardt F.C., Low K.B., Magasanik B., Schaechter M., Umbarger H.E. Washington, American Society for Microbiology. 1**:** 395-411.

Wycisk, K., Kim, E. J., Schroeder, J. I. and Krämer, U. (**2004**). "Enhancing the first enzymatic step in the histidine biosynthesis pathway increases the free histidine pool and nickel tolerance in *Arabidopsis thaliana*." Febs Letters 578(1-2): 128-134.

Xie, G., Keyhani, N. O., Bonner, C. A. and Jensen, R. A. (**2003**). "Ancient origin of the tryptophan operon and the dynamics of evolutionary change." Microbiology and Molecular Biology Reviews 67(3): 303-342.

Yariv, B., Yariv, E., Kessel, A., Masrati, G., Chorin, A. B., Martz, E., Mayrose, I., Pupko, T. and Ben-Tal, N. (**2023**). "Using evolutionary data to make sense of macromolecules with a "face-lifted" ConSurf." Protein Science 32(3): e4582.

Ycas, M. (**1974**). "On Earlier States of Biochemical System." Journal of Theoretical Biology 44(1): 145-160.

Ying, H., Cooke, I., Sprungala, S., Wang, W. W., Hayward, D. C., Tang, Y. R., Huttley, G., Ball, E. E., Forêt, S. and Miller, D. J. (**2018**). "Comparative genomics reveals the distinct evolutionary trajectories of the robust and complex coral lineages." Genome Biology 19(175).

Young, N. D., Debellé, F., Oldroyd, G. E. D., Geurts, R., Cannon, S. B., Udvardi, M. K., Benedito, V. A., Mayer, K. F. X., Gouzy, J., Schoof, H., Van de Peer, Y., Proost, S., Cook, D. R., Meyers, B. C., Spannagl, M., Cheung, F., De Mita, S., Krishnakumar, V., Gundlach, H., Zhou, S. G., Mudge, J., Bharti, A. K., Murray, J. D., Naoumkina, M. A., Rosen, B., Silverstein, K. A. T., Tang, H. B., Rombauts, S., Zhao, P. X., Zhou, P., Barbe, V., Bardou, P., Bechner, M., Bellec, A., Berger, A., Bergès, H., Bidwell, S., Bisseling, T., Choisne, N., Couloux, A., Denny, R., Deshpande, S., Dai, X. B., Doyle, J. J., Dudez, A. M., Farmer, A. D., Fouteau, S., Franken, C., Gibelin, C., Gish, J., Goldstein, S., González, A. J., Green, P. J., Hallab, A., Hartog, M., Hua, A., Humphray, S. J., Jeong, D. H., Jing, Y., Jöcker, A., Kenton, S. M., Kim, D. J., Klee, K., Lai, H. S., Lang, C. T., Lin, S. P., Macmil, S. L., Magdelenat, G., Matthews, L., McCorrison, J., Monaghan, E. L., Mun, J. H., Najar, F. Z., Nicholson, C., Noirot, C., O'Bleness, M., Paule, C. R., Poulain, J., Prion, F., Qin, B. F., Qu, C. M., Retzel, E. F., Riddle, C., Sallet, E., Samain, S., Samson, N., Sanders, I., Saurat, O., Scarpelli, C., Schiex, T., Segurens, B., Severin, A. J., Sherrier, D. J., Shi, R. H., Sims, S., Singer, S. R., Sinharoy, S., Sterck, L., Viollet, A., Wang, B. B., Wang, K. Q., Wang, M. Y., Wang, X. H., Warfsmann, J., Weissenbach, J., White, D. D., White, J. D., Wiley, G. B., Wincker, P., Xing, Y. B., Yang, L. M., Yao, Z. Y., Ying, F., Zhai, J. X., Zhou, L. P., Zuber, A., Dénarié, J., Dixon, R. A., May, G. D., Schwartz, D. C., Rogers, J., Quétier, F., Town, C. D. and Roe, B. A. (**2011**). "The *Medicago* genome provides insight into the evolution of rhizobial symbioses." Nature 480(7378): 520-524.

Zallot, R., Oberg, N. and Gerlt, J. A. (**2019**). "The EFI Web Resource for Genomic Enzymology Tools: Leveraging Protein, Genome, and Metagenome Databases to Discover Novel Enzymes and Metabolic Pathways." Biochemistry 58(41): 4169-4182.

Zhu, D., Xiong, H. B., Wu, J. H., Zheng, C. H., Lu, D. D., Zhang, L. X. and Xu, X. M. (**2022**). "Protein Targeting Into the Thylakoid Membrane Through Different Pathways." Frontiers in Physiology 12(802057).

13. Attachments

Attachment 1: P1: Witek W., Śliwiak J., Ruszkowski M. (**2021**) *Structural and mechanistic insights into the bifunctional HISN2 enzyme catalyzing the second and third steps of histidine biosynthesis in plants,* Scientific Reports 11(1):9647, https://doi.org/10.1038/s41598-021-88920-2

Attachment 2: P2: Witek W., Imiołczyk B., Ruszkowski M. (**2024**) *Structural, kinetic, and evolutionary peculiarities of HISN3, a plant 5'-ProFAR isomerase,* Plant Physiology and Biochemistry 215 109065, https://doi.org/10.1016/j.plaphy.2024.109065

Attachment 3: P3: Witek W., Śliwiak J., Rawski M., Ruszkowski M. (**2024**) *Targeting imidazole-glycerol phosphate dehydratase in plants: novel approach for structural and functional studies, and inhibitor blueprinting,* Frontiers in Plant Science 15;15:1343980, https://doi.org/10.3389/fpls.2024.1343980

Attachment 4: P4: Rutkiewicz M., Nogues I., **Witek W.**, Angelaccio S., Contestabile R., Ruszkowski M. (**2023**) *Insights into the substrate specificity, structure, and dynamics of plant histidinol-phosphate aminotransferase (HISN6),* Plant Physiology and Biochemistry 196:759-773**,** https://doi.org/10.1016/j.plaphy.2023.02.017

Attachment 5: Authors' declarations

scientific reports

OPEN

Structural and mechanistic insights into the bifunctional HISN2 enzyme catalyzing the second and third steps of histidine biosynthesis in plants

Wojciech Witek[1], Joanna Sliwiak[1] & Milosz Ruszkowski[1,2✉]

The second and third steps of the histidine biosynthetic pathway (HBP) in plants are catalyzed by a bifunctional enzyme–HISN2. The enzyme consists of two distinct domains, active respectively as a phosphoribosyl-AMP cyclohydrolase (PRA-CH) and phosphoribosyl-ATP pyrophosphatase (PRA-PH). The domains are analogous to single-domain enzymes encoded by bacterial *hisI* and *hisE* genes, respectively. The calculated sequence similarity networks between HISN2 analogs from prokaryotes and eukaryotes suggest that the plant enzymes are closest relatives of those in the class of *Deltaproteobacteria*. In this work, we obtained crystal structures of HISN2 enzyme from *Medicago truncatula* (*Mt*HISN2) and described its architecture and interactions with AMP. The AMP molecule bound to the PRA-PH domain shows positioning of the N1-phosphoribosyl relevant to catalysis. AMP bound to the PRA-CH domain mimics a part of the substrate, giving insights into the reaction mechanism. The latter interaction also arises as a possible second-tier regulatory mechanism of the HBP flux, as indicated by inhibition assays and isothermal titration calorimetry.

Metabolic pathways have been the subject of extensive studies for more than a century. The study of L-histidine (hereafter histidine) biosynthesis in prokaryotes and lower eukaryotes has engaged scientists for nearly 70 years. The histidine biosynthetic pathway (HBP) was first studied on microorganisms, e.g., *Salmonella typhimurium* and *Escherichia coli*, and is well characterized in prokaryotes[1]. It unraveled many mechanisms fundamental to cell biology, e.g., an operon structure and gene expression[2]. Genetic and biochemical analysis of thousands of mutations in *his* operon in *S. typhimurium* showed that, in contrast to the fungus *Neurospora crassa*, the bacterial *his* genes were tightly clustered[3,4]. The observation of coordinated expression of that cluster led to the idea that a group of genes functions as a single unit of expression and regulation, today known as an operon[5,6]. Together with *lac*[7] and *trp* operons[8], *his* operon was used as a model system to study polar mutations[9]. Moreover, studies of the HBP helped to discover the regulation of amino acid biosynthesis by attenuation[10].

The HBP is rather conservative among different domains of life; however, there are differences in the number of genes involved in the pathway and their expression pattern[11]. In bacteria, *his* genes are arranged in a compact operon (*hisGDC [NB] HAF [IE]*), with three of them (*hisD, hisNB* and *hisIE*) sometimes but not always coding for bifunctional enzymes[10,12]. Analysis of the structure of *his* genes revealed three main molecular mechanisms that are important in shaping the HBP, i.e., gene duplication, gene fusion, and gene elongation, which make this pathway a suitable model for understanding general molecular mechanisms behind metabolic routes[2].

In plants, the HBP study was delayed (until the 1980s) due to a lack of genetic approach and complicated biochemistry standing behind the pathway. As a result, the first auxotrophic mutants in higher-plant systems arrived much later than their bacterial or fungal counterparts[13]. Recent progress in molecular biology techniques has revealed that many of the enzymatic steps of the HBP in plants are performed by proteins encoded by single genes, which is in contrast to the extensive gene redundancy found in other amino acid biosynthetic pathways in plants[14]. Genes encoding all eight histidine biosynthetic enzymes (*HISN1-8*) have been identified in *Arabidopsis*

[1]Center for Biocrystallographic Research, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Noskowskiego 12/14, 61-704 Poznan, Poland. [2]Synchrotron Radiation Research Section of MCL, National Cancer Institute, Argonne, IL, USA. ✉email: mruszkowski@ibch.poznan.pl

thaliana[15]. Five of the HBP enzymes in *A. thaliana* are encoded by single-copy genes, with duplications in *HISN1*, *HISN5,* and *HISN6*[16].

The HBP flux regulation at the post-translational level links to the first enzyme (HISN1), an ATP-phosphoribosyl transferase (ATP-PRT, EC 2.4.2.17, Fig. 1). ATP-PRTs catalyze condensation of ATP (adenosine-5′-triphosphate) and PRPP (phosphoribosylpyrophosphate) into PR-ATP (N^1-5′-phosphoribosyl-ATP). ATP-PRTs are allosterically feedback-inhibited by histidine[17]. Furthermore, binding of adenosine-5′-monophosphate (AMP) at the active site increases the enzyme's sensitivity to histidine, also in plants[18]. So far, there have been no implications that any other HBP enzyme could be regulated.

In the second step of the HBP, PR-ATP is hydrolyzed to N^1-5′-phosphoribosyl-AMP (PR-AMP) by phosphoribosyl ATP pyrophosphohydrolase (PRA-PH; EC 3.6.1.31). In the third step, PR-AMP cyclohydrolase (PRA-CH, EC 3.5.4.19) opens the adenine ring of PR-AMP to produce N^1-[(5′-phosphoribosyl)formimino]-5-aminoimidazole-4-carboxamide-ribonucleotide (ProFAR). Then the HBP follows to yield histidine after eight more steps, catalyzed by subsequent enzymes.

Prokaryotes' genomes often contain separate genes, *hisE* and *hisI,* that encode PRA-PH and PRA-CH enzymes, respectively[19]. However, in some bacteria, such as *E. coli* or *S. typhimurium*, the protein product of a fused gene, *hisIE*, has both activities. The gene fusion can go even further as, e.g., in yeast, *Saccharomyces cerevisiae,* a single gene (*HIS4*) encodes a trifunctional enzyme with activities of PRA-PH, PRA-CH, and histidinol dehydrogenase (HDH, EC 1.1.1.23)[20]. These are the second, third, and last reactions of the HBP, respectively (Fig. 1).

In the plant HBP, a single gene (*HISN2*) encodes a HISN2 enzyme that performs two subsequent reactions (Fig. 1). One domain of HISN2 has the PRA-PH activity, whereas the second domain has the PRA-CH activity[21,22]. In this study, we investigated the HISN2 enzyme from a model legume, *Medicago truncatula*, named *Mt*HISN2. The research focused on (i) the enzyme molecular structure, (ii) similarities and differences with bacterial orthologs of known structures, (iii) interactions with AMP, a proposed activity regulator, and (iv) the catalytic mechanism.

Results and discussion

Phylogenetic analysis suggests the evolutionary origin of plant HISN2 sequences. We have analyzed 53 111 available sequences assigned to InterPro families IPR008179, IPR021130, IPR002496, and IPR038019 to assess the sequence similarity between prokaryotic and eukaryotic HISN2-equivalent enzymes and trace the evolution of plant HISN2 proteins. The analysis suggests that plant bifunctional enzymes derive from the *Myxococcales* order in the class of *Deltaproteobacteria* (Fig. 2). Fungal trifunctional proteins (HIS4 in yeast) with PRA-PH, PRA-CH, and HDH (histidinol dehydrogenase) activities also derive from orders close to *Myxococcales*. Moreover, sequences from some *Gammaproteobacteria* and *Spirochaetia* of PRA-PH, PRA-CH, and ProFAR isomerase activities seem to derive from a similar common ancestor. Multifunctional enzymes permit an optimal yield of gene expression without a need for additional transcription regulation, as noted in the genetic history of the HBP[16]. Aside from the multifunctional enzymes, most bacterial classes like *Alpha-, Beta-, Gamma-,* and *Deltaproteobacteria, Actinobacteria, Flavobacteria, Cytophagia,* and *Opitutae* express single-activity enzymes. Monofunctional enzymes are also common in the superkingdom of *Archaea*; however, there is a small group of archaeal species with bifunctional enzymes (Fig. 2).

As recently reported by Del Duca et al.[23], gene elongation was a leading mechanism in the evolution of *hisA*, *hisF*, *hisB,* and *hisD* histidine biosynthetic genes. The hypothesis for their evolution was confirmed by high sequence similarities between two halves of the proteins and by structural and biochemical studies. Since sequences of the four enzymes encoded by those genes are highly conserved in prokaryotic and eukaryotic organisms, it is most likely that the gene elongation occurred in the early stage of HBP evolution, before the Last Universal Common Ancestor[23]. The diversity in *hisI/E* (bacteria), HIS4 (fungi), and HISN2 (plants) may be another example of the importance of the gene elongation and duplication that occurred at the very early stage of the HBP evolution.

The overall structure of *Mt*HISN2: a dimeric enzyme with discrete and directly interacting pyrophosphohydrolase and cyclohydrolase domains. The complete sequence of *Mt*HISN2 contains 283 amino acid residues (UniProt ID[24]: A0A072U2X9; Gene: 25498966). All plant enzymes of the HBP are encoded by the genomic DNA and contain N-terminal chloroplast-targeting signal peptides[22]. In *Mt*HISN2, bioinformatic analysis with TargetP[25] suggested the signal peptide encompasses approx. forty N-terminal residues. In *A. thaliana* HISN2, the target peptide spans fifty residues (UniProt ID: O82768). We designed the construct to include sequence conserved in plant species; hence our final construct starts from Val49, preceded by a linker tripeptide, Ser-Asn-Ala.

The X-ray structure of *Mt*HISN2 was solved by experimental phasing using single-wavelength anomalous dispersion (SAD) on zinc cations bound to the protein. The unliganded protein (with metals) crystallized in the *C*2 space group (Table 1) with two protein chains in the asymmetric unit (ASU). *Mt*HISN2-AMP complex crystallized in the *C*2 space group but with different unit cell parameters (Table 1) and six protein chains (three dimers) in the ASU. The obtained electron density maps allowed us to trace most of the protein chain unambiguously, except for up to eighteen C-terminal residues and fragments between 157–165 and 186–194 (model- and chain-dependent) that were disordered.

*Mt*HISN2 forms a tight dimer of 26.4 kDa subunits (Fig. 3A), sharing a ~ 4000 Å^2 interface, according to PISA analysis[26,27]. The dimeric form is consistent with the size-exclusion elution profile (not shown). The dimer's surface area is ~ 20,000 Å^2 and is negatively charged (Fig. 3B), agreeing with the calculated pI of 5.3. The negative charge suggests that metal cations play an important role in interactions with negatively charged, phosphate-containing substrates, PR-ATP and PR-AMP.
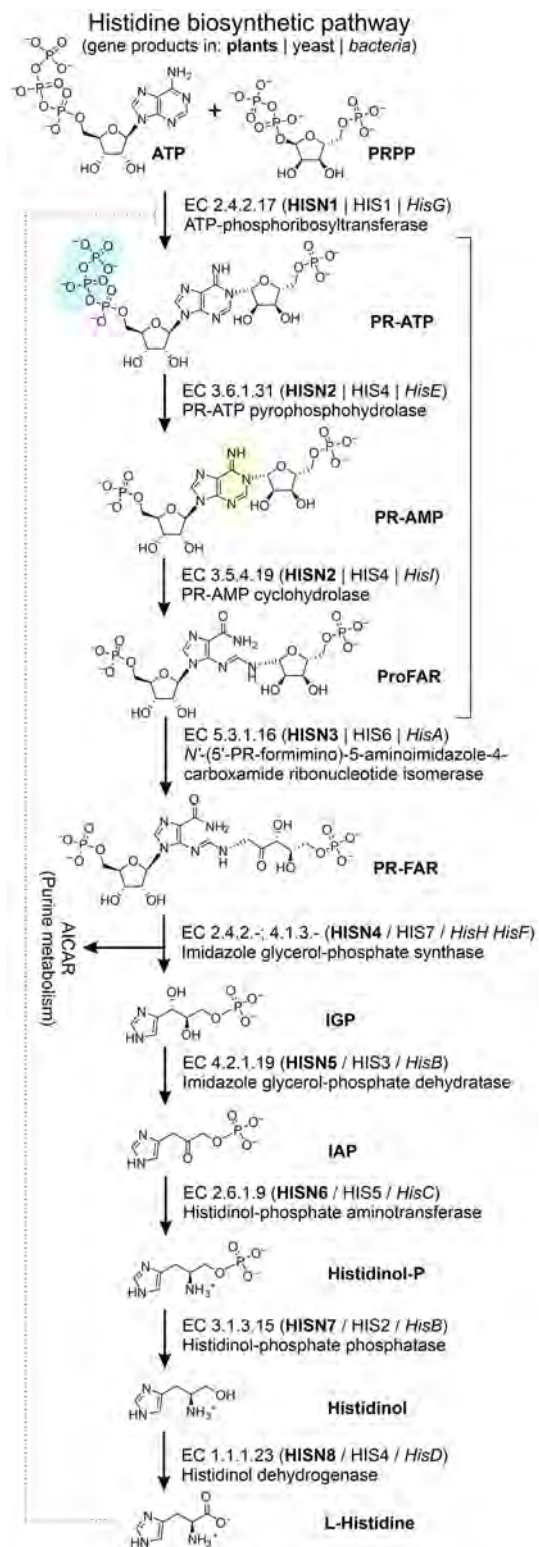
Figure 1. Organization of the histidine biosynthetic pathway (HBP) in plants, yeast, and bacteria. The large bracket marks the pathway fragment catalyzed by the PRA-PH and PRA-CH domains of *Mt*HISN2. Hydrolyzed groups by the PRA-PH and PRA-CH domains are highlighted by blue and yellow, respectively. The dotted line illustrates the feedback inhibition.

Back to p. 35
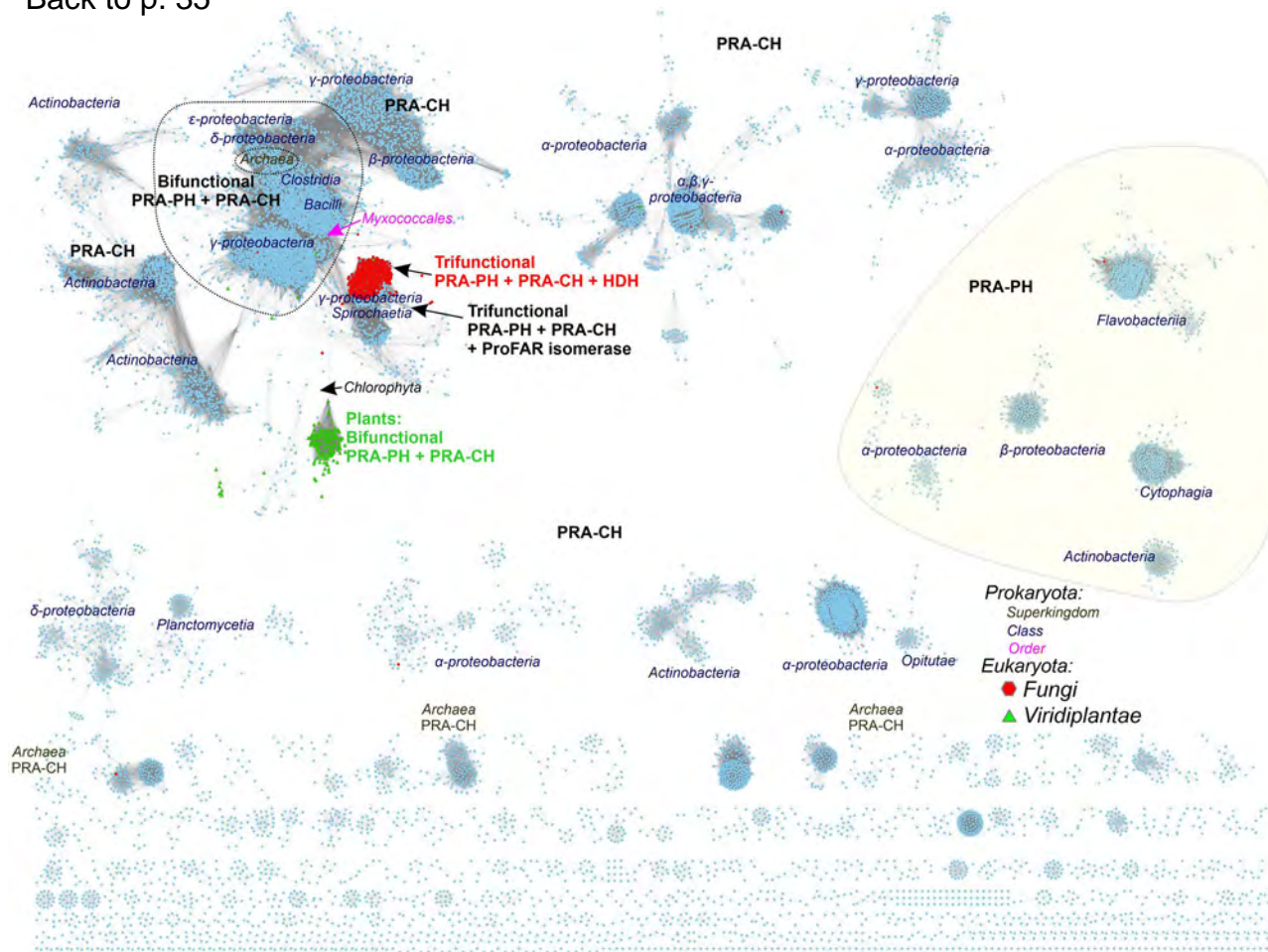
Figure 2. Sequence similarity network of PRA-PH and PRA-CH domains. UniRef90 sequences in InterPro families IPR008179, IPR021130, IPR002496, and IPR038019 were analyzed. 14 933 nodes are presented in the figure after rejecting 6748 outliers from the diagram. Sequences from *Eukaryota* are marked according to the legend. Monofunctional PRA-PH and PRA-CH proteins are most common in bacteria. Bi- and trifunctional enzymes are indicated with their specific activities. HDH, L-histidinol dehydrogenase; ProFAR, N^1-(5′-phosphoribosyl-formimino)-5-aminoimidazole-4-carboxamide ribonucleotide.

The enzyme dimer is formed by two mutually swapped polypeptide chains, forming a bilobial protein—each domain forms one lobe (Fig. 3). Per sequence analogy to corresponding enzymes from bacteria and other plant species, those domains catalyze PRA-PH and PRA-CH reactions (Fig. 1). The PRA-CH domain is located at the N-terminus, spanning residues 49–158 (Fig. 3). The PRA-PH domain lies at the C-terminus, ranging from residues 172–283. It must be noted here that in this article, we treat a domain as a complete and functional dimeric entity—with two active sites. The existence of a monomeric form of either PRA-PH or PRA-CH domain is highly improbable as it would expose vast hydrophobic regions. In *Arabidopsis*, both domains, apparently as dimers, were shown as functionally independent, even when expressed separately[28].

The PRA-PH domain consists of two overlapping and swapped protein chains built entirely of α-helices connected by loops. Each chain of the domain contributes five α-helices (α4–α8, Fig. 3A). Helices α6 and α7 form a four-helix bundle with their counterparts from the dimer mate, α6* and α7* (an asterisk denotes an element from the other subunit within the dimer). Helices α6 and α7 contain the PRA-PH active sites, defined near the metal-binding sites 1 and 2 (MBSs, Fig. 4A). Except for the bundle consisting of the four longest helices, there are short helices α4 and α8 and their counterparts α4*, α8* that overlap on top of each other, creating a tight chain swap. The swap separates the four-helix bundle from the PRA-CH domain.

In general, PRA-PH enzymes are Mg^{2+}-dependent[29]. However, the *Mt*HISN2 crystals could only be grown in the presence of a low concentration of Zn^{2+} in addition to Mg^{2+}. Zinc often binds to proteins at non-specific sites or at sites naturally binding other metals[30], which likely was the case here. Thus, we decided to use a more general term—MBSs—in this work to avoid confusion. There are two unique MBSs in the PRA-PH domain. MBS1 contains Zn^{2+} coordinated by two carboxyl oxygen atoms of Glu220 and one carboxyl oxygen of Glu217 (Fig. 4A). In MBS2, Zn^{2+} is tetrahedrally coordinated by carboxylic groups of Glu214, Glu234, Asp237, and a water molecule. In some subunits, Glu217 also participates in Zn^{2+} coordination in MBS2—resulting in the disappearance of MBS1. Because metal at MBS1 was absent in some subunits in our structures, only MBS2 may

	*Mt*HISN2		*Mt*HISN2-AMP		
Data collection					
Wavelength (Å)	0.9793		1.0000		
Space group	*C*2		*C*2		
Unit cell parameters					
a, b, c (Å)	172.1, 69.3, 52.0		202.9, 68.6, 135.6		
β (°)	94.7		128.8		
Resolution cut-off method	Isotropic		Anisotropic		
Resolution shell	All	Outer	All	Inner	Outer
Resolution (Å)	80–1.60	1.70–1.60	48.45–2.70	48.45–7.83	2.85–2.70
Unique reflections	79,751	12,569	33,411	5848	6725
Multiplicity	3.7	3.8	3.9	3.5	4.1
Ellipsoidal completeness (%)	N/A	94.1	99.2	55.8	
Spherical completeness (%)	98.5	96.3	85.0	99.2	26.8
R_{merge} (%)	4.5	62.4	9.1	3.0	73.2
$<I/\sigma(I)>$	14.6	1.9	11.1	28.7	2.0
CC(1/2)	99.9	69.5	99.7	99.9	67.5
Refinement					
R_{free} reflections	1037		1010		
No. of atoms (non-H)	3701		10,004		
Protein	3336		9658		
Water	356		62		
Other	9		284		
R_{work}/R_{free} (%)	14.3/17.5		18.1/24.6		
RMSD from ideal geometry					
Bond lengths (Å)	0.011		0.006		
Bond angles (°)	1.014		0.771		
Ramachandran statistics (%)					
Favored	97.8		98.9		
Allowed	2.2		1.1		
Outliers	0.0		0.0		
PDB ID	7BGM		7BGN		

Table 1. Diffraction data and refinement statistics.

be catalytically relevant. Also, it is very likely that in vivo Mg^{2+} cations (not Zn^{2+}) occupy MBS2, as magnesium, not zinc, is required for PRA-PH activity[31].

The PRA-CH domain also consists of two overlapping chains but has an entirely different structure (Fig. 3A). The domain connects with the PRA-PH domain via two long loops consisting of twelve residues (159–171), each belonging to one chain. The core of the PRA-CH domain is made of β-strands and α-helices forming the so-called barrelizing β-grasp fold (β-GF), wherein the β-sheet "grasps" an α-helix in a fasciclin-like assemblage[32]. There are many kinds of β-GF, but all of them share a similar topology, where β-strands form a mixed β-sheet surrounding a helix (α2 in *Mt*HISN2). The most characteristic feature of the core four-stranded β-sheet is that the flanking strands are parallel to each other, while the two middle strands are anti-parallel to the flanking strands. This means that the first and the last strands (by sequence) are located in the central part of the sheet with a crossover via an α-helical fragment. Variety of unrelated proteins where the β-GF was found indicates that, despite its relatively small size, the β-GF is a multifunctional scaffold suited for small-molecule binding (PR-AMP in this case). In *Mt*HISN2, a β-strand is followed by a helix and a loop that together form a super-secondary motif responsible for the domain swap. The β-sheet is connected with the motif via a long loop spanning residues 138–149 and contains residues coordinating MBSs 4–5 (see, Fig. 4B).

In our structures, the PRA-CH domains contain two or three (model- and subunit-dependent) MBSs that bind metal cations through conserved aspartate (Asp125*, Asp127*, and Asp129* in *Mt*HISN2), cysteine (Cys126*, Cys142, Cys149), and histidine (His143) residues (Fig. 4B). As noted by D'Ordine et al.[33], corresponding residues are universally conserved in cyclohydrolases. In the PRA-CH structure from *Methanobacterium thermoautotrophicum,* the aspartate residues (Asp85, Asp87, and Asp89) coordinated Cd^{2+} in a site corresponding to the MBS3 of *Mt*HISN2 (Asp125, Asp127, and Asp129, respectively), where Zn^{2+} was bound[33]. The MBS3 is formed by the carboxylic groups of Asp125*, Asp127*, Asp129* and by two water molecules in a shape of a trigonal bipyramid. In the next site, MBS4, Zn^{2+} is coordinated tetrahedrally by two water molecules and Nε of His143 and by the thiol of Cys126*. However, we did not observe a metal cation bound at MBS4 in the *Mt*HISN2-AMP complex, suggesting that a metal bound to the MBS4 can either promote substrate binding or may not be physiologically relevant. Lastly, Zn^{2+} bound in the MBS5 is coordinated by thiols of Cys126*, Cys142, and Cys149

Back to p. 36
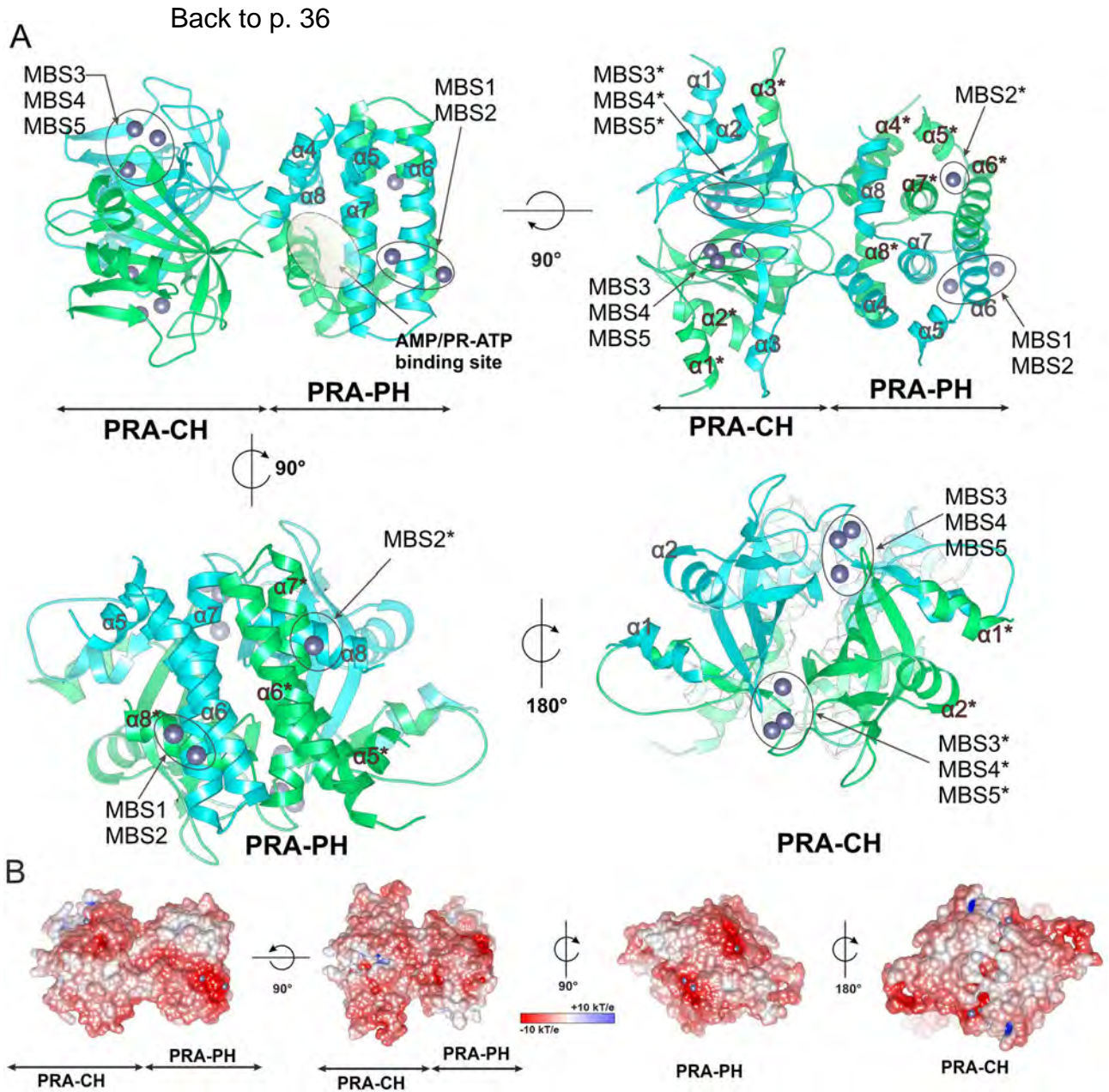
Figure 3. Structure of unliganded *Mt*HISN2. (**A**) Ribbon representation of the *Mt*HISN2 dimer; the metal-binding sites (MBSs) containing Zn^{2+} (dim gray) are marked in elipses; the AMP/PR-ATP binding site in the PRA-PH domain is marked in a dashed elipse. Notice overlapping chain A (light green) and chain B (cyan) that form well-separated domains of PRA-PH and PRA-CH activities. Asterisks (*) represent elements of symmetric subunits. (**B**) Surface electrostatic potential of *Mt*HISN2 is color-coded as shown in a bar. Protein rotations correspond to panel **A**, respectively.

(Fig. 4B). Considering the chemical nature of residues in the metal coordination spheres (metal ligands), it is likely that Zn^{2+} occupies only MBS5 in vivo, while MBS3 and MBS4 may bind a cocatalytic Mg^{2+}, per definition by Valle and Auld[34]. This hypothesis is consistent with the results of chemical probing of *Methanococcus vannielii* PRA-CH enzyme, which showed only one high-affinity Zn^{2+} binding site (corresponding to MBS5) per subunit[33].

Structural alignment of *Mt*HISN2 and its bifunctional bacterial counterpart reveals differences in the enzyme architecture while individual domains are similar. Structural comparisons of bacterial PRA-PH enzymes revealed high structural similarity, despite low sequence identities[29]. For instance, sequence identity as low as 31% between HisE from *Mycobacterium tuberculosis* (*Mtb*HisE, PDB ID: 1Y6X) and *Chromobacterium violaceum* HisE (CvHisE, 2A7W) still results in a very similar three-dimensional structure. BLAST sequence alignment between *Mtb*HisE and *Mt*HISN2 shows no significant similarity; however, both proteins share similar architecture in secondary structure topology, chain swaps, and the four-helix bundle.
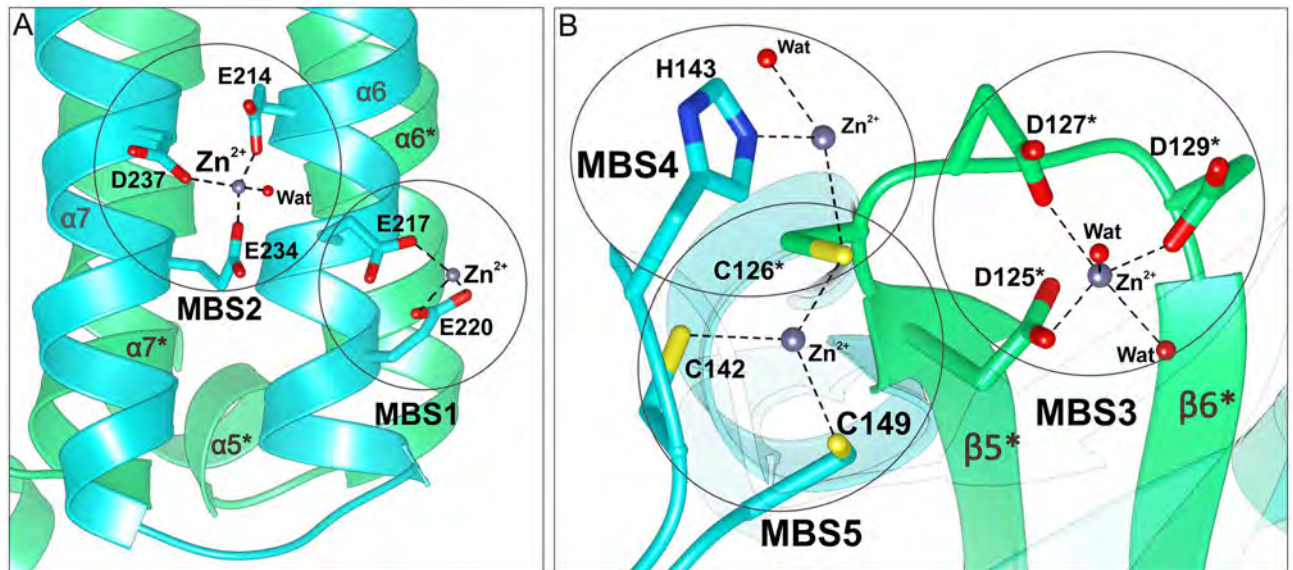
Figure 4. Metal-binding sites (MBSs) of *Mt*HISN2 in the unliganded form. Panel **A** shows zinc (dim grey) coordination by the residues of α6 and α7 and water molecules (red balls) in the PRA-PH domain. Orientation is the same as in the first panel in 3A. Panel **B** shows metal binding sites in the PRA-CH domain; in the AMP complex MBS4 is absent. Zinc in MBS3 is coordinated by residues of the loop connecting strands β5* and β6*. MBS4 and MBS5 are coordinated by residues of the chain B and Cys126* of the chain A.

The structural similarity despite relatively low sequence identity applies to the PRA-CH domain as well, as reflected by the RMSD of 0.68 Å between the *Mt*HISN2 PRA-CH domain and HisI of *M. thermoautotrophicum*, sharing sequence identity of 40%. As pointed by D'Ordine et al., alignment between archaeal, bacterial, and eukaryotic sequences, e.g., *M. thermoautotrophicum*, *E. coli*, *S. cerevisiae*, reveals that some residues are highly conserved among PR-AMP cyclohydrolases[33], which is consistent with their role in metal coordination also in *Mt*HISN2.

So far, the only structure of a bifunctional HisIE enzyme has been determined for *Shigella flexneri Sf*HisIE (PDB ID: 6J2L)[35]. Sequences of *Mt*HISN2 and *Sf*HisIE share 35% identity and 51% similarity, which indicates relatively low conservation. However, *Sf*HisIE has a similar topology to *Mt*HISN2 and lacks only the β7 strand and the α5 helix (in *Mt*HISN2 topology). The *Sf*HisIE sequence has three gaps, corresponding to residues 159–171, 185–188, 223–225 in *Mt*HISN2 (Fig. 5A).

Despite *Mt*HISN2 and *Sf*HisIE are somewhat distant homologs, their structural alignment reveals significant similarity in both individual PRA-PH (RMSD of 0.90 Å) and PRA-CH domains (RMSD of 0.84 Å). For instance, the PRA-CH active site of *Sf*HisIE and *Mt*HISN2 share a very similar architecture (Fig. 5B). However, significant differences arise from the comparison of the entire enzyme molecules. When the PRA-CH domains are superposed, relative rotations of the PRA-PH domains, measured as the axis of the α4 helix, differ by ~ 40° (Fig. 5C). Another major difference is the presence of a super-secondary strand-helix-loop motif near the domain-domain interface in the plant enzyme. It encompasses residues 150–172 of the *Mt*HISN2 sequence, which correspond to 105–110 in *Sf*HisIE. In *Mt*HISN2, it is involved in domain swapping by mutually overlapping corresponding chains, whereas *Sf*HisIE lacks that motif entirely (Fig. 5C). In summary, most differences between *Mt*HISN2 and *Sf*HisIE appear near or at the inter-domain junction.

The architecture of *Mt*HISN2 indicates that PR-AMP intermediate is released between the two catalytic events.

The protein structure was investigated using CAVER 3.0[36] PyMOL Plugin in the context of possible tunnels that may connect active sites of PRA-PH and PRA-CH domains to shuttle the PR-AMP intermediate. Such tunnels are common in hydrolases, including two-domain hydrolases[37–39]. In *Mt*HISN2, none of those tunnels would allow the transport of molecules—even as small as water—between the catalytic sites. We note that in some cases, such tunnels appear after binding of small molecules that change the overall shape of a protein; however, (i) we did not detect any conformational changes in the enzyme, and (ii) the diameter of the narrow fragment between the domains is only ~ 15 Å wide. This excludes the possibility of moving the PR-AMP intermediate between the catalytic sites. Because the catalytic sites of both domains are > 40 Å apart, PR-AMP must diffuse to the solvent (chloroplast stroma) after pyrophosphate cleavage to reach the PRA-CH domain. This also means that after being produced by the PRA-PH domain, PR-AMP molecules may be processed further by a PRA-CH domain in a different enzyme molecule.

AMP binding to the PRA-PH domain: positioning of the PR-ATP N1-phosphoribosyl.

Our *Mt*HISN2-AMP complex showed that the enzyme active sites are adapted to bind nucleotides despite the lack of super-secondary structures typical for such specificity. More precisely, there are no Rossmann-fold motifs, often
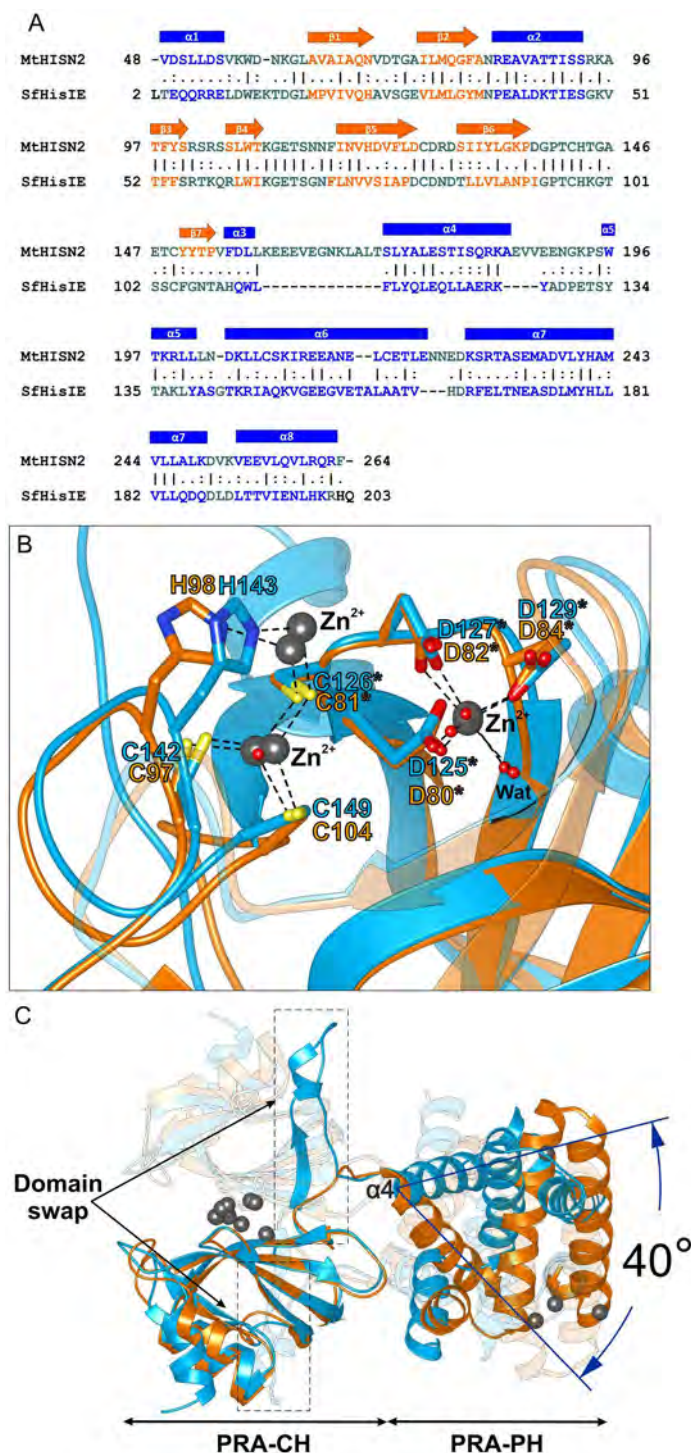
Figure 5. Comparison of a bacterial bifunctional HisIE enzyme and plant *Mt*HISN2. Panel **A** shows a sequence alignment of *Mt*HISN2 and *Sf*HisIE (35% identity; PDB ID: 6J2L). Secondary structure elements are colored in blue (α-helices) and orange (β-strands). **B** Superposition of MBSs in the PRA-CH domains of *Mt*HISN2 (blue) and the *Sf*HisIE (orange), made up by evolutionary conserved residues coordinating Zn²⁺ (dim gray). Superposition of the two structures in panel **C** reveals well-aligning PRA-CH domains, whereas PRA-PH domains are rotated by 40°, measured between the α4 helix in *Mt*HISN2 and its counterpart in *Sf*HisIE. The second subunits of both dimers are transparent for clarity.

Back to p. 37
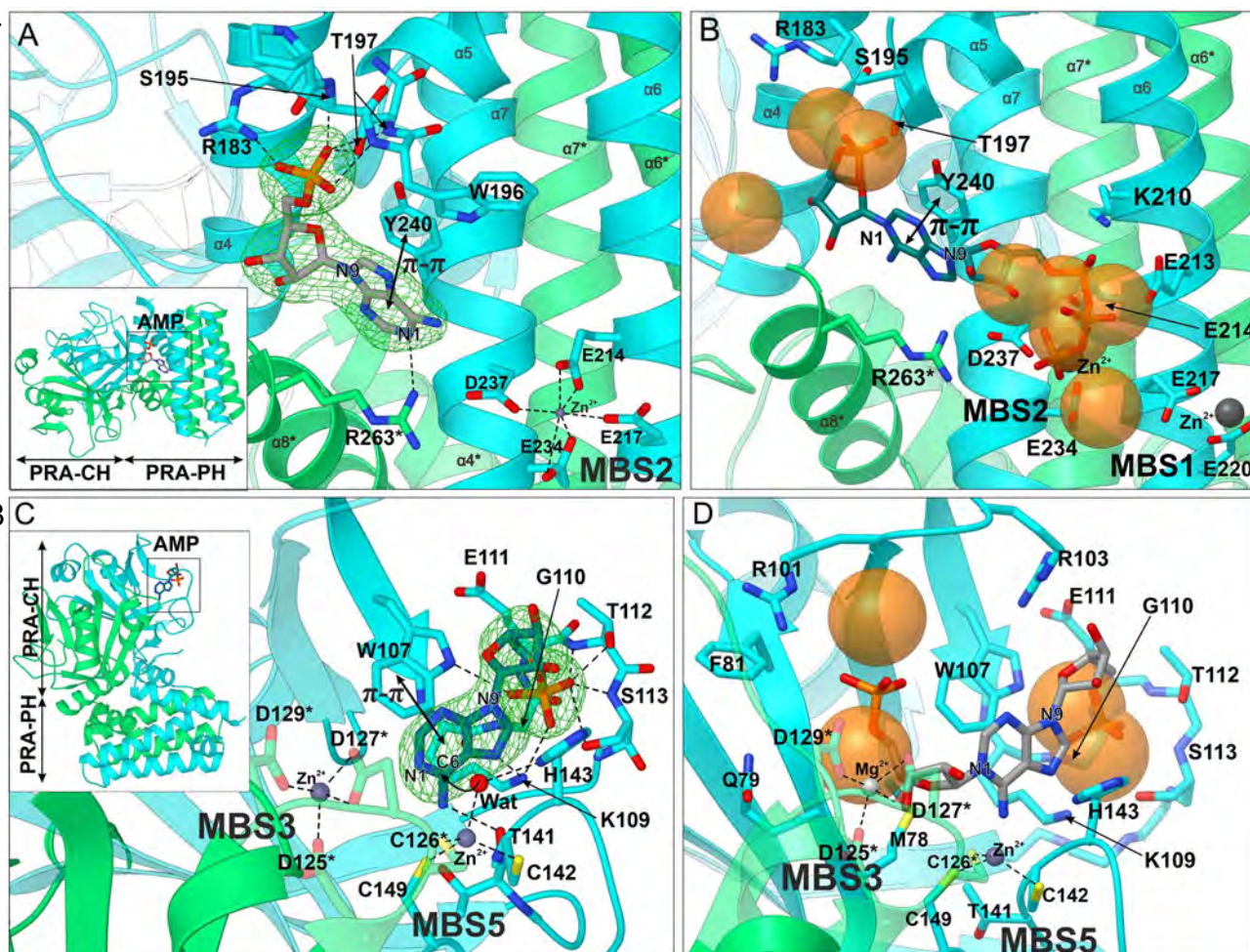
Back to p. 38

Figure 6. Interaction of AMP with *Mt*HISN2 and in silico docking of PR-ATP and PR-AMP. AMP binding to the PRA-PH domain (chain A) is shown in panel **A**. AMP was bound near MBS2 and coordinated by electrostatic interactions and π–π stacking between the side chain of Y240 and adenine ring of AMP. $F_o–F_c$ polder maps (green mesh) are contoured at 5σ level. Panel **B** illustrates in silico prediction of PR-ATP (dark cyan) binding, performed in AutoDock Vina. Sidechains of residues within a 5-Å radius are shown. Transparent orange balls from the analysis by the Nucleos server indicate areas of the highest probability of phosphate binding. Note that the triphosphate of docked PR-ATP aligns well with the prediction of multiple phosphate positions near MBS2. (Panel **C**) shows AMP binding to the PRA-CH domain. AMP was bound near MBS5 by electrostatic interactions, H-bonds, and π–π T-shaped stacking. The catalytic water molecule, activated by Zn^{2+} at MBS5 and His143, performs the nucleophilic attack (curvy arrow). The PR-AMP docking pose (obtained and presented as in panel **B**), compatible with the AMP binding mode, is shown in panel **D**. The metal cation at MBS3 was changed to Mg^{2+} as it appears more relevant in vivo.

associated with cofactors like FAD, NAD^+, and $NADP^+$, or Walker motifs, commonly present in ATP-binding proteins[40,41]. The previous analysis of *Sf*HisIE also did not reveal Rossmann fold and Walker motifs[35]. In the *Mt*HISN2-AMP complex, AMP molecules were found near MBSs in both domains, PRA-PH and PRA-CH. For clarity, representative AMP molecules with the lowest B-factors are described.

AMP bound in the PRA-PH domain formed hydrogen bonds through the phosphate moiety and the adenine ring (Fig. 6A). The guanidine group of Arg183 formed polar hydrogen bonds with one oxygen of the phosphate. The second oxygen of the phosphate group interacted with the hydroxyl groups of Ser195 and Thr197 and with the backbone amide of Thr197. The backbone amide of Trp196 contacted the third phosphate oxygen. The adenine N1 atom interacted with the Arg263 guanidine group. We also observed the π-π stacking between the adenine ring and the Tyr240 side chain; the approximate inter-ring distance was 3.6 Å (Fig. 6A).

In that context, we note that AMP bound to the PRA-PH domain in our structure most likely does not show a part PR-ATP (substrate) or PR-AMP (product). This conclusion is based on the orientation of the AMP phosphate group pointing away from the metal center (MBS1-2) and interacting with the guanidine group of Arg183 instead. In contrast, the ATP fragment of PR-ATP should have its triphosphate group near the metal center for the hydrolysis to occur. To gain more insights, we utilized two in silico methods in parallel. We analyzed putative phosphate-binding regions in the *Mt*HISN2 structure using Nucleos[42]. It indicated that more phosphate groups (e.g., triphosphate) could bind near the MBS1-2 sites rather than near Arg183 (Fig. 6B). Molecular docking of

PR-ATP with AutoDock Vina was consistent with the Nucleos results (Fig. 6B). The proposed orientation of the adenine ring of PR-ATP was rotated by ~ 180° in the ring's plane to the AMP pose in the *Mt*HISN2-AMP complex. This means that the binding of AMP to the PRA-PH domain in our *Mt*HISN2 complex apparently shows the positioning of the N1-phosphoribosyl of PR-ATP and the plane of its adenine ring.

AMP binding to the PRA-CH domain: an update to the catalytic mechanism. The second AMP binding site was located within the PRA-CH domain (Fig. 6C). The phosphate moiety formed an extensive network of hydrogen bonds with surrounding residues. The phosphate O1 atom bound to Nε of Trp107 and the backbone N of Gly110. The O2 atom interacted with the hydroxyl group and the backbone N of Ser113 and the hydroxyl group and backbone N of Thr112. The O3 atom was bound to the hydroxyl group of Ser113, the amine group of Lys109, and a water molecule. Moreover, the adenine N6 atom interacted with the carbonyl of Thr141, whereas N7 H-bonded with the amine group of Lys109. We also observed edge-to-face interaction between the aromatic rings of the adenine and Trp107, with ≈ 3.5 Å distance and angle ω ≈ 45°.

As reported by D'Ordine et al.[33], the in silico docking of PR-AMP to the PRA-CH enzyme from *M. thermo-autotrophicum* indicated that the substrate molecule in the active site is bound mainly by eighteen residues of which sixteen are conserved, and one is preserved in all PR-AMP cyclohydrolases[33,43]. The authors proposed two phosphate-binding regions, (i) Ser60, Thr61 and Ser62 (Ser100, Arg101, Ser102 in *Mt*HISN2) for the N9-phosphoribosyl, and (ii) Glu71, Ser72 and Ser73 for the N1-phosphoribosyl (Glu111, Thr112, Ser113 in *Mt*HISN2). Another interaction predicted by the authors to assist in substrate recognition is edge-to-face interaction between the adenine ring and Trp67 (Trp107 in *Mt*HISN2). The N9 ribosyl group was proposed to interact with Mg^{2+} and Arg15, which has no corresponding residue in *Mt*HISN2. His110 (His143 in *Mt*HISN2) was predicted to have a role in catalysis and π-stacking with the incoming substrate molecule. In terms of N1- and N9-phosphoribosyl orientations, a similar model has been reported by Wang et al.[35], who also used in silico PR-AMP docking.

The AMP position in our *Mt*HISN2-AMP complex does not agree with the previously-presented in silico models. Nevertheless, the *Mt*HISN2-AMP complex is the first experimental structure showing (at least) a part of the PR-AMP substrate in the PRA-CH active site. In the *Mt*HISN2-AMP complex, N9-phosphoribosyl interacts with the region formed by residues $_{107}$WTKGETS$_{113}$, suggesting that the PR-AMP pose would be rotated by ~ 180° in the adenine ring plane, compared to the model by D'Ordine et al.[33]. In consequence, the region formed by residues $_{100}$SRS$_{102}$, likely interacts with the N1-phosphoribosyl. It is also possible that MBS3 plays a role in binding the N1-phosphoribosyl, especially since Mg^{2+} bound to the corresponding site was essential for the activity of other PRA-CH enzymes[33,44]. Our AMP pose with the N6 atom pointing towards the protein core (and not the solvent) agrees with the complexes of adenosine deaminases, a family of Zn^{2+}-dependent hydrolases acting on adenosine-like substrates[45,46]. We must also note that we observed C2′-*endo* ribose in the *Mt*HISN2-AMP complex, meaning that even AMP, lacking the N1-phosphoribosyl, already binds "contracted" to the PRA-CH active site. D'Ordine et al. acknowledged that dealing with the flexibility of ribose rings was a big challenge during docking[33]. In our docking experiments, PR-AMP was bound to the PRA-CH domain (Fig. 6D) in a pose that is compatible with that of AMP in the *Mt*HISN2-AMP (Fig. 6B, D).

Thanks to the conserved three-cysteine active site (Cys142, Cys149, and Cys126*, MBS5), the general PRA-CH mechanism may be adopted from other reports[33,47] and updated by the experimental position of AMP, which mimics a part of PR-AMP (Fig. 6C,D). First, PR-AMP is oriented in the catalytic pocket by the two phosphate-binding regions, namely (i) N1-phosphoribosyl orients towards $_{100}$SRS$_{102}$ and/or Mg^{2+} coordinated by Asp125*, Asp127*, and Asp129*, while (ii) N9-phosphoribosyl attracts to $_{107}$WTKGETS$_{113}$. The adenine moiety is secured by a hydrogen bond between its N7 atom and Nζ amine of Lys109 and by the edge-to-face interaction with Trp107. The nucleophilic water molecule in the Zn^{2+} coordination sphere (MBS5) is activated by His143, acting as a general base. A metal cation (MBS4) may play a role in priming His143; in the unliganded *Mt*HISN2 structure, His143 does not bind a water molecule but instead is in the MBS4 coordination sphere (Fig. 4B). The activated water molecule (or rather a hydroxyl anion) performs a nucleophilic attack on the purine C6 atom, breaking the N1-C6 bond. Distances observed in the *Mt*HISN2-AMP complex, Zn^{2+}…H_2O of 2.4 Å, Nδ of His143…H_2O of 3.0, and H_2O…C6 of 3.1 Å, are consistent with this mechanism. The role of the His143 as the general base is supported by lack of detectable activity of the H143E mutant, while a (weaker) binding of PR-AMP may still occur, as deduced from the K_d for AMP of 68 μM (Fig. 7). Moreover, environment of the active site pocket suggests that the optimal positioning of N1-phosphoribosyl may stretch the substrate, aiding the ring hydrolysis (Fig. 6D).

AMP is an inhibitor of the PRA-CH domain of *Mt*HISN2 at physiologically-relevant concentrations. AMP is an activity regulator of plant HISN1 enzymes and their counterparts from other kingdoms of life. Although it has been shown that AMP alone does not exhibit an inhibitory effect on *Mt*HISN1, it significantly increases sensitivity to feedback regulation by free histidine[18]. However, so far, there have been no indications that other HBP enzymes could be regulated by AMP. In this work, *Mt*HISN2 inhibition by AMP was assayed using PR-ATP produced enzymatically, as PR-ATP is commercially unavailable. The PR-ATP production, prior to *Mt*HISN2 measurements, was monitored spectrophotometrically (at 290 nm, Fig. 7A). That mixture was then used to trigger AMP inhibition assays with *Mt*HISN2, in which the PR-ATP concentration was 18 μM, so that absorbance changes (at 300 nm) could be monitored[44]. Since ATP-PRT enzyme was still present in the *Mt*HISN2 reaction mixture, we also cross-validated the assay by including free histidine (at 100 μM), known to inhibit ATP-PRTs. We observed that 100 μM AMP caused over 60% inhibition. It must also be noted here that the AMP concentration, for instance, in maize chloroplasts ranges from 40 μM to 260 μM[48]. This puts *Mt*HISN2-AMP interaction as a possible secondary regulation mechanism of the HBP flux. Unfortunately, to our knowledge there is no data on the PR-ATP concentration in vivo. Notwithstanding, the 18 μM concentration used in our assay may even be exaggerated, as PR-ATP is readily processed by the HBP. Interestingly, when
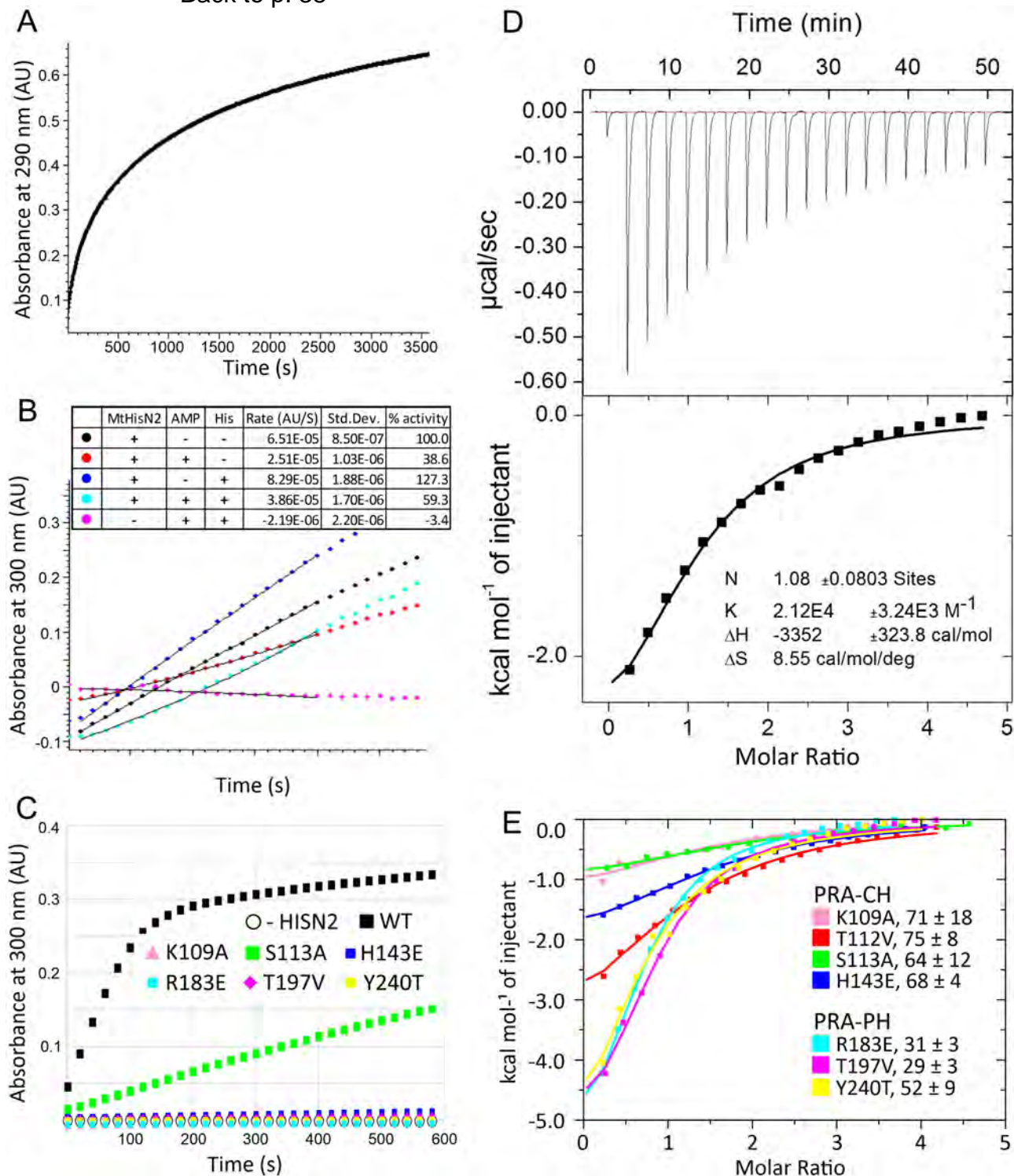
Figure 7. *Mt*HISN2 activity measurements. Panel **A** shows the course of enzymatic PR-ATP production. Results of the AMP (at 100 μM concentration) inhibition assay, in the presence and absence of histidine (100 μM), are presented in panel **B**; ProFAR increase was monitored at 300 nm. Panel **C** illustrates relative activities of the wild-type *Mt*HISN2 and its point mutants. Microcalorimetric study of the interaction between *Mt*HISN2 and AMP is shown in panels **D–E**. Representative ITC results for the wild-type protein is shown in panel **D**; the raw data are in the upper part, while the bottom part shows the best fit of one sets of binding sites model to the integrated peaks. Panel **E** shows AMP binding properties of the *Mt*HISN2 point mutants; K_d values ± errors are shown for each mutant in μM.

Primer name	Sequence
MtHISN2-WT-F	TACTTCCAATCCAATGCCGTAGACTCATTGTTGGACAGTGTAAAATG
MtHISN2-WT-R	TTATCCACTTCCAATGTTATCAATTTTCCACCGATTTCTGGGTTGG
K109A-F	GTTGTGGACCGCGGGAGAGACCTCCAATAATTTCATCAATGTC
K109A-R	GTCTCTCCCGCGGTCCACAACGATGATCGTGACC
T112V-F	GGAGAGGTGTCCAATAATTTCATCAATGTCCATGATGTC
T112V-R	GAAATTATTGGACACCTCTCCTTTGGTCCACAACGATG
S113A-F	GGAGAGACCGCGAATAATTTCATCAATGTCCATGATGTC
S113A-R	GAAATTATTCGCGGTCTCTCCTTTGGTCCACAACG
H143E-F	CCTACCTGCGAGACAGGGGCAGAAACATGCTACTATAC
H143E-R	GCCCCTGTCTCGCAGGTAGGCCCATCAGGTTTC
R183E-F	CAATATCCCAGGAGAAGGCAGAGGTAGTAGAAGAAAATGGAAAG
R183E-R	CTCTGCCTTCTCCTGGGATATTGTTGACTCTAATGCATACAG
T197V-F	CTTCATGGGTCAAGCGGTTATTGCTTAATGATAAGTTGC
T197V-R	CAATAACCGCTTGACCCATGAAGGCTTTCCATTTTCTTCTAC
Y240T-F	GATGTACTCACGCATGCCATGGTTCTGTTGGCACTG
Y240T-R	CATGGCATGCGTGAGTACATCAGCCATCTCTGAAGCAG

Table 2. Primer sequences used in this work.

both AMP and histidine were present, the *Mt*HISN2 inhibition was mitigated to 41% (Fig. 7B). Because ATP-PRT enzymes bind AMP in the presence of histidine, the pool of AMP available to bind to *Mt*HISN2 decreases, providing the most likely explanation to this phenomenon. The control sample, without *Mt*HISN2, excluded the impact of the ATP-PRT reaction on the observed absorbance change at 300 nm at the moment of the HISN2 reactions, which were run simultaneously.

AMP interaction with *Mt*HISN2 in solution was further investigated using isothermal titration calorimetry (ITC). Our data show that AMP binding to *Mt*HISN2 (Fig. 7D) is characterized by the K_d value of 47 ± 6 μM and stoichiometry N = 1. Thermodynamic parameters are ΔH = -3352 ± 324 cal/mol and ΔS = 8.6 cal/mol/deg. To deduce whether the obtained K_d can be attributed to AMP binding to the PRA-PH or to the PRA-CH domain, we performed ITC experiments on point mutants of *Mt*HISN2. Four mutants within the PRA-CH domain (K109A, T112V, S113A, and H143E) and three within the PRA-PH domain (R183E, T197V, and Y240T) were tested and the results are shown in Fig. 7E. The results clearly indicate that the AMP binding affinity is lowered in the case of PRA-CH domain mutants. Moreover, these mutations significantly lower the heat effect of AMP binding in comparison with PRA-PH domain mutants (Fig. 7E). These two observations indicate that AMP binding to the PRA-CH domain is driven by enthalpy, thus can be measured by ITC. We cannot also exclude an auxiliary impact of AMP binding to the PRA-PH domain on the overall *Mt*HISN2 activity.

Conclusions and outlook

This article is the fifth in a series of papers that show the structures of plant HBP enzymes. Previous structures were reported for: HISN1[18], HISN5[49], HISN7[50], and HISN8[51]. In this work, we experimentally solved the structure of the HISN2 enzyme from the model legume, *Medicago truncatula* using X-ray diffraction data. The bifunctional *Mt*HISN2, with distinct PRA-PH and PRA-CH domains, showed significantly different relative orientation of the domains than in bacterial enzymes. Comparing bacterial and plant enzymes shed new light on the possible design of small-molecule inhibitors as potential antibiotics or herbicides. In this perspective, HisI, HisE, (or HisIE), homologs of fungal HIS4, and plant HISN2 enzymes may arise as promising molecular targets. If one wants to target bacterial or plant enzymes specifically, regions other than the conserved active sites appear most auspicious. The proposed insights into the regulation and catalytic mechanism provide groundwork for the design of HISN2 inhibitors, in addition to bringing a deeper comprehension of the plant HBP.

*Mt*HISN2 interacts with AMP, as shown by our complex crystal structure, inhibition assays, and ITC experiments, which indicated that *Mt*HISN2 activity regulation occurs in a physiologically-relevant range of AMP concentration. This way, the HBP flux can be tightly controlled on two steps, catalyzed by HISN1 and HISN2 enzymes. The need to control the HBP flux rises from a high metabolic cost of the pathway, estimated as equivalent to over thirty ATP molecules[52]. The HBP is at the same time the only pathway of amino acid biosynthesis that utilizes carbon and nitrogen directly from ATP. As fluctuations of the AMP/ATP ratio reflect the cell metabolic status, an AMP-based control can regulate resource consumption by the HBP.

Materials and methods

Cloning, expression, and purification. The total RNA was isolated from young *M. truncatula* leaves using the RNeasy Plant Mini Kit (Qiagen). The following reverse transcription with oligo dT_{18} primer yielded the complementary DNA (cDNA). The chloroplast-targeting peptide was recognized using the TargetP 1.1 server[25,53], and the produced construct was N-truncated at Val49. The desired fragment was amplified by polymerase chain reaction; primers used in this work are given in Table 2. The expression plasmid, based on the pMCSG68 backbone (Midwest Center for Structural Genomics), was created by the ligase-independent cloning

method[54]. Mutagenic substitutions were conducted using the Polymerase Incomplete Prime Extension (PIPE) method[55] on the wild-type *Mt*HISN2 expression plasmid as a template and primers listed in Table 2. Correctness of all inserts was confirmed by DNA sequencing.

Overexpression was carried in BL21 Gold *E. coli* cells (Agilent Technologies) in LB media with 150 μg/mL ampicillin. After incubation with shaking at 190 rpm at 37 °C until the A_{600} reached 1.0, the cultures were chilled to 18 °C, and isopropyl-d-thiogalactopyranoside was added at a final concentration of 0.5 mM to start overexpression, which went on for 18 h. The cell pellet from the 2-L culture was centrifuged at 3500×*g* for 20 min at 4 °C and resuspended in 35 mL of binding buffer [50 mM Hepes–NaOH pH 7.5; 500 mM NaCl; 20 mM imidazole; 2 mM tris(2-carboxyethyl)phosphine (TCEP)] and stored at − 80 °C for purification.

The cells were disrupted by sonication (4 min with intervals for cooling), and the cell debris was removed by centrifugation at 25,000×*g* for 30 min at 4 °C. The supernatant was mixed with 3 mL of HisTrap HP resin (GE Healthcare) in a column on the VacMan setup (Promega). The resin-bound protein was washed five times with the binding buffer and eluted with 20 mL of elution buffer (50 mM Hepes–NaOH pH 7.5; 500 mM NaCl; 400 mM imidazole; 2 mM TCEP). The His$_6$-tag was cleaved with TEV protease (at final concentration 0.2 mg/mL) overnight, simultaneously with dialysis to lower the imidazole concentration to 20 mM. The second run through the HisTrap resin resulted in pure *Mt*HISN2 in the flow-through to which ZnCl$_2$ was added at 100 μM final concentration. The sample was concentrated to 2.4 mL and loaded on a HiLoad Superdex 200 16/60 column (GE Healthcare), equilibrated with buffer: 25 mM Hepes–NaOH pH 7.5, 100 mM KCl, 50 mM NaCl, 100 μM ZnCl$_2$, and 1 mM TCEP. The protein was then concentrated and used for crystallization or functional assays.

Crystallization, X-ray data collection, and processing. *Mt*HISN2 was crystallized using the vapor diffusion method. The protein concentration was 10 mg/ml, as determined by A_{280} measurement (molar extinction coefficient, ε of 43,430 M^{-1}·cm^{-1}). The unliganded structure results from crystals (hanging-drop) obtained by mixing 4 μl of the protein solution and 2 μl of 60% Morpheus D1 condition (Molecular Dimensions)[56]. The components of Morpheus D1 are: 0.12 M Alcohols (0.2 M 1,6-Hexanediol; 0.2 M 1-Butanol 0.2 M 1,2-Propanediol; 0.2 M 2-Propanol; 0.2 M 1,4-Butanediol; 0.2 M 1,3-Propanediol) 0.1 M Buffer System 1, pH 6.5 (Imidazole; MES-acid) 30% Precipitant Mix 1 (20% v/v PEG 500* MME; 10% w/v PEG 20,000). The crystals were cryoprotected by adding Morpheus D1 condition supplemented with 20% ethylene glycol. For the *Mt*HISN2-AMP complex structure, 10 mM MgCl$_2$, 0.1 mM ZnCl$_2$, and 20 mM AMP (added in 100 mM Hepes pH 7.5) were included in the solution subjected to crystallization. Then, the PEG/Ion screen (Hampton Research) supplemented with 7.5% glycerol was set up on a sitting-drop crystallization plate (1:1 μl mixtures). The crystals appeared in A11 condition (0.2 M potassium iodide, 20% Polyethylene glycol 3350). Immediately before crystal harvesting, 1 μl of PEG/Ion A11 condition with 50% of glycerol was added to the drop. All crystals were vitrified in liquid nitrogen and stored for synchrotron data collection.

Diffraction data were collected at the SER-CAT beamline 22-ID and SBC 19-ID at the Advanced Photon Source, Argonne National Laboratory, USA. Diffraction data were processed with the XDS package[57]. Anisotropic truncation of X-ray data for the *Mt*HISN2-AMP complex was done using the STARANISO server[58]. Data processing statistics are given in Table 1.

Determination and refinement of the crystal structures. The crystal structure of *Mt*HISN2 was solved by SAD using protein crystallized in the presence of 100 μM ZnCl$_2$, using the same data as for the *Mt*HISN2 unliganded structure refinement (PDB ID: 7BGM). Notably, other *Mt*HISN2 crystals were also soaked with selenourea crystal, as proposed by Luo[59], but no selenourea molecules were found upon inspection of the final electron density maps. The phasing was performed with *Phenix.Autosol*[60]. The initial model was built using *Phenix.AutoBuild*[61], and was placed inside the unit cell with the *ACHESYM* server[62]. *COOT*[63] was used for manual model corrections between rounds of automatic model refinement in *Phenix.Refine*[64]. The nearly finished model of *Mt*HISN2 served to solve the AMP complex by molecular replacement with *PHASER*[65]. The refinement statistics are listed in Table 1.

Kinetic measurements. Steady-state kinetic measurements were performed at 22 °C according to the method developed by Ames et al.[66], with the Agilent 8453 spectrophotometer equipped with 8-cell automatic sample changer. Prior to experiments with *Mt*HISN2, the reaction mixture for PR-ATP production (R1) contained the kinetic buffer (4 mM Mg^{2+}, 25 mM Hepes pH 7.5, 50 mM NaCl, 100 mM KCl, 1 mM TCEP), 1.3 μM *Sc*HIS1 (subunits concentration), 2.5 μM *A. thaliana* inorganic pyrophosphatase[67], 1 mM ATP and 1 mM PRPP; however, PRPP was added immediately after blanking to start the reaction. PR-ATP formation in R1 was monitored at λ = 290 nm during 60-min incubation (Fig. 7A). The PR-ATP concentration reached 180 μM, based on the absorption coefficient, ε290 = 3600 M^{-1} cm^{-1}.

The inhibition assay was performed in five cuvettes simultaneously; their content together with the experiment result is shown in Fig. 7B. Before the reaction, the cuvettes containing 900 μL of the kinetic buffer +/− AMP and/or histidine, both at 100 μM (final concentration) and wild-type MtHISN2 at 19 nM (f.c.) were incubated for 30 min. The control cuvette did not contain *Mt*HISN2. To start the reaction, 100 μl of the R1 mixture (PR-ATP) was added, the initial PR-ATP concentration was ~ 18 μM. The reaction progress was measured by monitoring ProFAR formation at λ = 300 nm[44].

Comparative activity assay of *Mt*HISN2 mutants was performed using 790 μL of kinetic buffer to which 200 μl of the R1 mixture was added. The reactions were started by adding 10 μl of 1 mg/ml solutions of *Mt*HISN2 variants. The control cuvette did not contain *Mt*HISN2. The assay was performed in eight 1-ml cuvettes simultaneously, and the reaction progress was monitored at λ = 300 nm; the result is shown in Fig. 7C.

Microcalorimetric study of the interaction between HISN2 and AMP. ITC measurements were carried out with MicroCal PEAQ-ITC (Malvern) at 298 K. Titrations of AMP (2 mM) against *Mt*HISN2 protein (kept at ≈ 100 μM concentration determined at 280 nm) were done in 25 mM HEPES buffer pH 7.5 (100 mM NaCl, 50 mM KCl, 1 mM TCEP, 4 mM MgCl$_2$, 10 μM ZnCl$_2$). AMP was injected in 19 aliquots of 2 μl. Raw ITC data were analyzed with the *Origin* 7.0 software (Origin-Lab) to obtain thermodynamic parameters like stoichiometry (*N*), dissociation constant (K_d), and the changes in the enthalpy (Δ*H*) and entropy. One set of binding sites model was fitted to the data. Reference power was set to 5. A stirring speed of 750 rpm and spacing of 150 s was used. Experiments were performed in triplicate. To assign the AMP binding to a particular domain, analogical AMP titration measurements were carried on MtHISN2 mutants of the PRA-CH domain (K109A, T112V, S113A, H143E) as well as of the PRA-PH domain (R183E, T197V, Y240T).

In-silico analyses and data presentation. The EFI-ESN web server[68] served to calculate the sequence similarity network. The number of sequences (53 111) in the four included InterPro families: IPR008179, IPR021130, IPR002496, and IPR038019 was limited to the UniRef90 subset, which contained 21 942 sequences. The calculations were based on the alignment score of 50 for sequences between 70 and 1000 residues long. The figure was created in *Cytoscape* 3.3[69]; 6748 outliers were manually excluded from the figure.

Molecular figures were created in *UCSF Chimera*[70], which also served to calculate the RMSD values for Cα atom pairs within 2-Å distance. Molecular docking was performed in *AutoDock Vina*[71]. The ligand and receptor files were prepared in *PyRx*[72] and the *UCSF Chimera DockPrep* tool. The receptor file was based on *Mt*HISN2-AMP complex, with AMP removed. The search box was approx. 30 × 30 × 30 Å, centered at the AMP binding sites.

The *Nucleos* webserver[42] was used to identify putative phosphate binding sites in the *Mt*HISN2 structure. The allowed RMSD for the structural matches between the *Mt*HISN2 structure and the reference mini-structures of nucleobases, carbohydrates, and phosphates was set to a default value of 0.6 Å. The results for nucleobase and carbohydrate predictions were omitted in the presentation.

Caver 3.0.3 PyMol plugin was used to calculate molecular tunnels in the structure of *Mt*HISN2 with following parameters: minimum probe radius = 0.9, shell depth = 10, shell radius = 8, clustering threshold = 3.5.

Research involving plants. Studies complied with local and national regulations for using plants.

Data availability

References

1. Winkler, M. E. & Ramos-Montanez, S. Biosynthesis of histidine. *EcoSal Plus* https://doi.org/10.1128/ecosalplus.3.6.1.9 (2009).
2. Fani, R., Brilli, M., Fondi, M. & Lio, P. The role of gene fusions in the evolution of metabolic pathways: The histidine biosynthesis case. *BMC Evol. Biol.* **7**(Suppl 2), S4. https://doi.org/10.1186/1471-2148-7-S2-S4 (2007).
3. Haas, F., Mitchell, M. B., Ames, B. N. & Mitchell, H. K. A series of histidineless mutants of Neurospora crassa. *Genetics* **37**, 217–226 (1952).
4. Ames, B. N., Garry, B. & Herzenberg, L. A. The genetic control of the enzymes of histidine biosynthesis in *Salmonella typhimurium*. *J. Gen. Microbiol.* **22**, 369–378. https://doi.org/10.1099/00221287-22-2-369 (1960).
5. Ames, B. N. & Garry, B. Coordinate repression of the synthesis of four histidine biosynthetic enzymes by histidine. *Proc. Natl. Acad. Sci. USA* **45**, 1453–1461. https://doi.org/10.1073/pnas.45.10.1453 (1959).
6. Kasai, T. Regulation of the expression of the histidine operon in *Salmonella typhimurium*. *Nature* **249**, 523–527. https://doi.org/10.1038/249523a0 (1974).
7. Newton, W. A., Beckwith, J. R., Zipser, D. & Brenner, S. Nonsense mutants and polarity in the lac operon of *Escherichia coli*. *J. Mol. Biol.* **14**, 290–296. https://doi.org/10.1016/s0022-2836(65)80250-9 (1965).
8. Jacob, F. & Monod, J. Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.* **3**, 318–356. https://doi.org/10.1016/s0022-2836(61)80072-7 (1961).
9. Fink, G. R. & Martin, R. G. Translation and polarity in the histidine operon.II. Polarity in the histidine operon. *J. Mol. Biol.* **30**, 97–107. https://doi.org/10.1016/0022-2836(67)90246-x (1967).
10. Blasi, F. & Bruni, C. B. Regulation of the histidine operon: Translation-controlled transcription termination (a mechanism common to several biosynthetic operons). *Curr. Top. Cell Regul.* **19**, 1–45. https://doi.org/10.1016/b978-0-12-152819-5.50018-x (1981).
11. Alifano, P. *et al.* Histidine biosynthetic pathway and genes: Structure, regulation, and evolution. *Microbiol. Rev.* **60**, 44–69 (1996).
12. Carlomagno, M. S., Chiariotti, L., Alifano, P., Nappo, A. G. & Bruni, C. B. Structure and function of the *Salmonella typhimurium* and *Escherichia coli* K-12 histidine operons. *J. Mol. Biol.* **203**, 585–606. https://doi.org/10.1016/0022-2836(88)90194-5 (1988).
13. Miflin, B. *The Biochemistry of Plants: A Comprehensive Treatise, Vol. 5: Amino Acids and Derivatives* 533–541 (Academic Press, 1980).
14. Radwanski, E. & Last, R. Tryptophan biosynthesis and metabolism: Biochemical and molecular genetics. *Plant Cell* **7**, 921–934. https://doi.org/10.1105/tpc.7.7.921 (1995).
15. Petersen, L. N. *et al.* The missing link in plant histidine biosynthesis: Arabidopsis myoinositol monophosphatase-like2 encodes a functional histidinol-phosphate phosphatase. *Plant Physiol.* **152**, 1186–1196. https://doi.org/10.1104/pp.109.150805 (2010).
16. Muralla, R., Sweeney, C., Stepansky, A., Leustek, T. & Meinke, D. Genetic dissection of histidine biosynthesis in Arabidopsis. *Plant Physiol.* **144**, 890–903. https://doi.org/10.1104/pp.107.096511 (2007).
17. Ohta, D. *et al.* Molecular cloning and characterization of ATP-phosphoribosyl transferase from Arabidopsis, a key enzyme in the histidine biosynthetic pathway. *Plant Physiol.* **122**, 907–914. https://doi.org/10.1104/pp.122.3.907 (2000).
18. Ruszkowski, M. Guarding the gateway to histidine biosynthesis in plants: Medicago truncatula ATP-phosphoribosyltransferase in relaxed and tense states. *Biochem. J.* **475**, 2681–2697. https://doi.org/10.1042/BCJ20180289 (2018).

19. Chiariotti, L., Alifano, P., Carlomagno, M. S. & Bruni, C. B. Nucleotide sequence of the *Escherichia coli* hisD gene and of the *Escherichia coli* and *Salmonella typhimurium* hisIE region. *Mol. Gen. Genet.* **203**, 382–388. https://doi.org/10.1007/bf00422061 (1986).

20. Donahue, T. F., Farabaugh, P. J. & Fink, G. R. The nucleotide sequence of the HIS4 region of yeast. *Gene* **18**, 47–59. https://doi.org/10.1016/0378-1119(82)90055-5 (1982).

21. Ingle, R. A. Histidine biosynthesis. *Arabidopsis Book* **9**, e0141. https://doi.org/10.1199/tab.0141 (2011).

22. Stepansky, A. & Leustek, T. Histidine biosynthesis in plants. *Amino Acids* **30**, 127–142. https://doi.org/10.1007/s00726-005-0247-0 (2006).

23. Del Duca, S., Chioccioli, S., Vassallo, A., Castronovo, L. M. & Fani, R. The role of gene elongation in the evolution of histidine biosynthetic genes. *Microorganisms* https://doi.org/10.3390/microorganisms8050732 (2020).

24. UniProt, C. UniProt: A worldwide hub of protein knowledge. *Nucleic Acids Res.* **47**, D506–D515. https://doi.org/10.1093/nar/gky1049 (2019).

25. Emanuelsson, O., Nielsen, H., Brunak, S. & von Heijne, G. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J. Mol. Biol.* **300**, 1005–1016. https://doi.org/10.1006/jmbi.2000.3903 (2000).

26. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797. https://doi.org/10.1016/j.jmb.2007.05.022 (2007).

27. Wilkins, M. R. *et al.* Protein identification and analysis tools in the ExPASy server. *Methods Mol. Biol.* **112**, 531–552. https://doi.org/10.1385/1-59259-584-7:531 (1999).

28. Fujimori, K. & Ohta, D. Isolation and characterization of a histidine biosynthetic gene in Arabidopsis encoding a polypeptide with two separate domains for phosphoribosyl-ATP pyrophosphohydrolase and phosphoribosyl-AMP cyclohydrolase. *Plant Physiol.* **118**, 275–283. https://doi.org/10.1104/pp.118.1.275 (1998).

29. Javid-Majd, F., Yang, D., Ioerger, T. R. & Sacchettini, J. C. The 1.25 A resolution structure of phosphoribosyl-ATP pyrophosphohydrolase from Mycobacterium tuberculosis. *Acta Crystallogr. D Biol. Crystallogr.* **64**, 627–635. https://doi.org/10.1107/S0907444908007105 (2008).

30. Cha, S. S. *et al.* Experimental phasing using zinc anomalous scattering. *Acta Crystallogr. D Biol. Crystallogr.* **68**, 1253–1258. https://doi.org/10.1107/S0907444912024420 (2012).

31. Smith, D. W. & Ames, B. N. Phosphoribosyladenosine monophosphate, an intermediate in histidine biosynthesis. *J. Biol. Chem.* **240**, 3056–3063 (1965).

32. Burroughs, A. M., Balaji, S., Iyer, L. M. & Aravind, L. Small but versatile: the extraordinary functional and structural diversity of the beta-grasp fold. *Biol. Direct* **2**, 18. https://doi.org/10.1186/1745-6150-2-18 (2007).

33. D'Ordine, R. L., Linger, R. S., Thai, C. J. & Davisson, V. J. Catalytic zinc site and mechanism of the metalloenzyme PR-AMP cyclohydrolase. *Biochemistry* **51**, 5791–5803. https://doi.org/10.1021/bi300391m (2012).

34. Vallee, B. L. & Auld, D. S. Cocatalytic zinc motifs in enzyme catalysis. *Proc. Natl. Acad. Sci. USA* **90**, 2715–2718. https://doi.org/10.1073/pnas.90.7.2715 (1993).

35. Wang, Y., Zhang, F., Nie, Y., Shang, G. & Zhang, H. Structural analysis of Shigella flexneri bi-functional enzyme HisIE in histidine biosynthesis. *Biochem. Biophys. Res. Commun.* **516**, 540–545. https://doi.org/10.1016/j.bbrc.2019.06.099 (2019).

36. Chovancova, E. *et al.* CAVER 3.0: A tool for the analysis of transport pathways in dynamic protein structures. *PLoS Comput. Biol.* https://doi.org/10.1371/journal.pcbi.1002708 (2012).

37. Fontaine, J. X. *et al.* Further insights into the isoenzyme composition and activity of glutamate dehydrogenase in Arabidopsis thaliana. *Plant Signal. Behav.* **8**, e23329. https://doi.org/10.4161/psb.23329 (2013).

38. Svendsen, A. *Understanding Enzymes: Function, Design, Engineering, and Analysis* (PAN STANFORD PUBLISHING, 2016).

39. Marco-Nardini, I. S. R. *et al.* The X-ray structure of epoxide hydrolase from Agrobacterium radiobacter AD1. *J. Biol. Chem.* **274**, 14579–14586. https://doi.org/10.1074/jbc.274.21.14579 (1999).

40. Kleiger, G. & Eisenberg, D. GXXXG and GXXXA motifs stabilize FAD and NAD(P)-binding Rossmann folds through C(alpha)-H···O hydrogen bonds and van der waals interactions. *J. Mol. Biol.* **323**, 69–76. https://doi.org/10.1016/s0022-2836(02)00885-9 (2002).

41. Saraste, M., Sibbald, P. R. & Wittinghofer, A. The P-loop—A common motif in ATP- and GTP-binding proteins. *Trends Biochem. Sci.* **15**, 430–434. https://doi.org/10.1016/0968-0004(90)90281-f (1990).

42. Parca, L., Ferre, F., Ausiello, G. & Helmer-Citterich, M. Nucleos: A web server for the identification of nucleotide-binding sites in protein structures. *Nucleic Acids Res.* **41**, W281–W285. https://doi.org/10.1093/nar/gkt390 (2013).

43. Sivaraman, J. *et al.* Crystal structure of Methanobacterium thermoautotrophicum phosphoribosyl-AMP cyclohydrolase HisI. *Biochemistry* **44**, 10071–10080. https://doi.org/10.1021/bi050472w (2005).

44. D'Ordine, R. L., Klem, T. J. & Davisson, V. J. N1-(5'-phosphoribosyl)adenosine-5'-monophosphate cyclohydrolase: Purification and characterization of a unique metalloenzyme. *Biochemistry* **38**, 1537–1546. https://doi.org/10.1021/bi982475x (1999).

45. Wang, Z. & Quiocho, F. A. Complexes of adenosine deaminase with two potent inhibitors: X-ray structures in four independent molecules at pH of maximum activity. *Biochemistry* **37**, 8314–8324. https://doi.org/10.1021/bi980324o (1998).

46. Zavialov, A. V., Yu, X., Spillmann, D., Lauvau, G. & Zavialov, A. V. Structural basis for the growth factor activity of human adenosine deaminase ADA2. *J. Biol. Chem.* **285**, 12367–12377. https://doi.org/10.1074/jbc.M109.083527 (2010).

47. Reizer, J., Buskirk, S., Bairoch, A., Reizer, A. & Saier, M. H. Jr. A novel zinc-binding motif found in two ubiquitous deaminase families. *Protein Sci.* **3**, 853–856. https://doi.org/10.1002/pro.5560030515 (1994).

48. Usuda, H. Adenine nucleotide levels, the redox state of the NADP system, and assimilatory force in nonaqueously purified mesophyll chloroplasts from maize leaves under different light intensities. *Plant Physiol.* **88**, 1461–1468. https://doi.org/10.1104/pp.88.4.1461 (1988).

49. Glynn, S. E. *et al.* Structure and mechanism of imidazoleglycerol-phosphate dehydratase. *Structure* **13**, 1809–1817. https://doi.org/10.1016/j.str.2005.08.012 (2005).

50. 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.* **291**, 9960–9973. https://doi.org/10.1074/jbc.M115.708727 (2016).

51. Ruszkowski, M. & Dauter, Z. Structures of Medicago truncatula L-histidinol dehydrogenase show rearrangements required for NAD(+) binding and the cofactor positioned to accept a hydride. *Sci. Rep.* **7**, 10476. https://doi.org/10.1038/s41598-017-10859-0 (2017).

52. Swire, J. Selection on synthesis cost affects interprotein amino acid usage in all three domains of life. *J. Mol. Evol.* **64**, 558–571. https://doi.org/10.1007/s00239-006-0206-8 (2007).

53. Nielsen, H., Engelbrecht, J., Brunak, S. & von Heijne, G. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng.* **10**, 1–6 (1997).

54. Kim, Y. *et al.* High-throughput protein purification and quality assessment for crystallization. *Methods* **55**, 12–28. https://doi.org/10.1016/j.ymeth.2011.07.010 (2011).

55. Klock, H. E. & Lesley, S. A. The Polymerase Incomplete Primer Extension (PIPE) method applied to high-throughput cloning and site-directed mutagenesis. *Methods Mol. Biol.* **498**, 91–103. https://doi.org/10.1007/978-1-59745-196-3_6 (2009).

56. Gorrec, F. The MORPHEUS protein crystallization screen. *J. Appl. Crystallogr.* **42**, 1035–1042. https://doi.org/10.1107/S0021889809042022 (2009).

57. Kabsch, W. Xds. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132. https://doi.org/10.1107/S0907444909047337 (2010).
58. Tickle, I. J. *et al.* (Cambridge, United Kingdom: Global Phasing Ltd., 2018).
59. Luo, Z. Selenourea: a convenient phasing vehicle for macromolecular X-ray crystal structures. *Sci. Rep.* **6**, 37123. https://doi.org/10.1038/srep37123 (2016).
60. Terwilliger, T. C. *et al.* Decision-making in structure solution using Bayesian estimates of map quality: The PHENIX AutoSol wizard. *Acta Crystallogr. D Biol. Crystallogr.* **65**, 582–601. https://doi.org/10.1107/S0907444909012098 (2009).
61. Terwilliger, T. C. *et al.* Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta Cryst. D* **64**, 61–69. https://doi.org/10.1107/S090744490705024X (2008).
62. Kowiel, M., Jaskolski, M. & Dauter, Z. ACHESYM: An algorithm and server for standardized placement of macromolecular models in the unit cell. *Acta Cryst. D* **70**, 3290–3298. https://doi.org/10.1107/S1399004714024572 (2014).
63. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Cryst. D* **66**, 486–501. https://doi.org/10.1107/S0907444910007493 (2010).
64. Afonine, P. V. *et al.* Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D Biol. Crystallogr.* **68**, 352–367. https://doi.org/10.1107/S0907444912001308 (2012).
65. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674. https://doi.org/10.1107/S0021889807021206 (2007).
66. Ames, B. N., Martin, R. G. & Garry, B. J. The first step of histidine biosynthesis. *J. Biol. Chem.* **236**, 2019–2026 (1961).
67. Grzechowiak, M. *et al.* Crystal structures of plant inorganic pyrophosphatase, an enzyme with a moonlighting autoproteolytic activity. *Biochem. J.* **476**, 2297–2319. https://doi.org/10.1042/BCJ20190427 (2019).
68. Zallot, R., Oberg, N. & Gerlt, J. A. The EFI web resource for genomic enzymology tools: Leveraging protein, genome, and metagenome databases to discover novel enzymes and metabolic pathways. *Biochemistry* **58**, 4169–4182. https://doi.org/10.1021/acs.biochem.9b00735 (2019).
69. Shannon, P. *et al.* Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504. https://doi.org/10.1101/gr.1239303 (2003).
70. Pettersen, E. F. *et al.* UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612. https://doi.org/10.1002/jcc.20084 (2004).
71. Trott, O. & Olson, A. J. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **31**, 455–461. https://doi.org/10.1002/jcc.21334 (2010).
72. Dallakyan, S. & Olson, A. J. Small-molecule library screening by docking with PyRx. *Methods Mol. Biol.* **1263**, 243–250. https://doi.org/10.1007/978-1-4939-2269-7_19 (2015).

Acknowledgements

Author contributions

W.W. refined the structures, analyzed them, and drafted the manuscript. M.R. designed the experiments, obtained the crystals, solved the structures, and edited the text. J.S. performed and analyzed ITC experiments.

Funding

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to M.R.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Structural, kinetic, and evolutionary peculiarities of HISN3, a plant 5′-ProFAR isomerase

Wojciech Witek, Barbara Imiolczyk, Milosz Ruszkowski [*]

Department of Structural Biology of Eukaryotes, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Noskowskiego 12/14, 61-704, Poznan, Poland

A B S T R A C T

Histidine biosynthesis is essential for the growth and development of plants, where it occurs within chloroplasts. The eleven reactions are catalyzed by eight enzymes, known as HISN1-8, each acting sequentially. Here, we present the crystal structures of a 5′-ProFAR isomerase (HISN3) from the model legume *Medicago truncatula* bound to its enzymatically synthesized substrate (ProFAR) and product (PrFAR). The active site of *Mt*HISN3 contains a sodium cation that participates in ligand recognition, a feature not observed in bacterial and fungal structures of homologous enzymes. The steady-state kinetics of wild-type *Mt*HISN3 revealed a slightly higher turnover rate compared to its bacterial homologs. Plant HISN3 sequences contain an unusually elongated Lys60-Ser91 fragment, while deletion of the 74–80 region resulted in a 30-fold loss in catalytic efficiency compared to the wild-type. Molecular dynamics simulations suggested that the fragment facilitates product release, thereby contributing to a higher k_{cat}. Moreover, conservation analyses suggested a non-cyanobacterial origin for plant HISN3 enzymes, which is another instance of a non-cyanobacterial enzyme in the plant histidine biosynthetic pathway. Finally, a virtual screening campaign yielded five molecules, with the energy gains ranging between −13.6 and −13.1 kcal/mol, which provide new scaffolds for the future development of herbicides.

1. Introduction

The histidine biosynthetic pathway (HBP) plays a crucial role in the growth and development of prokaryotes, fungi, and plants, by providing a building block for protein synthesis and contributing to the *de novo* synthesis of purines. Mammals do not have the HBP and depend on dietary uptake of histidine. Plant mutants with knocked-out alleles of *his* genes either die at the embryonic stage or exhibit severely distorted development, e.g., by displaying a short-root phenotype (Noutoshi and Shinozaki, 2005; Mo et al., 2006). For decades, due to a complicated biochemistry behind the pathway and lack of auxotrophic plant mutants, studies on histidine biosynthesis were limited mostly to bacteria (Haas et al., 1952; Ames, 1957a, 1957b, 1961; Ames and Garry, 1959; Miflin, 1980). Early studies of the plant HBP began in the 1960s and focused not only on the nature and regulation of the pathway, but also on the potential to target HBP with herbicides (Hilton et al., 1965; Klopotowski T., 1965; Wiater et al., 1971a,b,c). The advent of new molecular biology techniques, including genetic engineering, molecular cloning, and DNA sequencing, combined with the development of structural biology and computational methods, has allowed us to tackle

these problems and decipher the HBP in eukaryotes. This article is part of a series dedicated to structural studies of the HBP enzymes in plants. To date, we have published the crystal and/or cryo-EM structures of HISN1 (Ruszkowski, 2018), HISN2 (Witek et al., 2021), HISN5 (Witek et al., 2024), HISN6 (Rutkiewicz et al., 2023), HISN7 (Ruszkowski and Dauter, 2016), and HISN8 (Ruszkowski and Dauter, 2017).

In plants, the HBP occurs in chloroplasts. It is an eleven-step pathway catalyzed by eight enzymes, named HISN1-8, according to their consecutive action in the HBP. The fourth step is catalyzed by HISN3 (in plants), which is an N'-[(5′-phosphoribosyl)formimino]-5-aminoimidazole-4-carboxamide ribonucleotide isomerase (EC 5.3.1.16), also known as 5′-ProFAR isomerase, or BBM II isomerase (Ingle, 2011). Plant HISN3s, along with their homologs (bacterial HisA and fungal His6), belong to a $(\beta\alpha)_8$-barrel fold enzymes and catalyze the reaction of N'-[(5′-phosphoribosyl)formimino]-5-amino-imidazole-4-carboxamide ribonucleotide (ProFAR) isomerization to N'-[(5′- phosphoribulosyl) formimino)]-5-amino-imidazole-4-carboxamide ribonucleotide (PrFAR). The reaction occurs via Amadori rearrangement in which an aminoaldose, ProFAR, is converted to a ketose, PrFAR (Fig. 1).

The Amadori rearrangement mechanism is shared with TrpF (EC

5.3.1.24) of N'-(5′-phosphoribosyl) anthranilate isomerase activity, which catalyzes the third step in the tryptophan biosynthetic pathway (TBP). Interestingly, bacterial HisA and TrpF are so similar that a single amino acid substitution can switch between the HisA and TrpF activity (Jürgens et al., 2000). In some bacteria, e.g., *Streptomyces coelicolor* and *Mycobacterium tuberculosis*, which lack the *trpF* gene, the third reaction of the TBP is actually catalyzed by a bifunctional (βα)₈-barrel enzyme, N'-(5′-phosphoribosyl) anthranilate isomerase A (priA), of both HisA and TrpF activities (Kuper et al., 2005). This cross-functioning indicates a significant evolutionary relationship between HisA homologs, TrpF, and PriA enzymes. Numerous studies have linked HisA and TrpF evolution with a PriA-like ancestral enzyme, mostly based on their fold and the patchwork hypothesis (Jensen, 1976; Jürgens et al., 2000; Gerlt and Babbitt, 2001; Barona-Gómez and Hodgson, 2003; Kuper et al., 2005; Söderholm et al., 2015; Romero-Rivera et al., 2022). Here, we describe the results of phylogenetic analyses which show the plant HISN3's position in an intricate network of evolutionary relationships.

Emerging herbicide resistance (HR) has already begun to threaten global food production. There are now 530 unique cases involving more than 270 weed species that have become resistant to one or more of the 168 herbicides that are in use in 72 countries (Heap, 2024). The HR affects 21 of the 31 known modes of action (MoA) in which herbicides work, with the highest number of resistant species in the inhibition of acetolactate synthase (>170 species), photosystem II inhibition (>80 species), and inhibition of enolpyruvyl shikimate phosphate (EPSP) synthase (>50 species) (Ruszkowski and Forlani, 2022; Heap, 2024). Apart from EPSP synthase, which is severely affected by the HR due to glyphosate overuse (Koo et al., 2022), and glutamine synthase, with only six cases reported, other amino acid biosynthetic pathways are promising targets. The HBP is an example of such a pathway, but preliminary successes regarded only imidazole-glycerol phosphate dehydratase (HISN5), which can be inhibited by 3-amino-1,2,4-triazole (amitrole) (Kishore and Shah, 1988; Rawson et al., 2018). Notably, the HBP intersects with the *de novo* purine synthesis at the HISN4 step. Therefore, inhibition of HISN3 might potentially have a double impact on weed growth by limiting the amount of both histidine and purines. As a consequence, this would slow down the HR in plants, allowing for longer and safer weed management.

To rationally design new inhibitors, it is crucial to understand the detailed atomic structure of the targeted enzyme. Despite the availability of numerous bacterial HisA structures and one fungal His6 structure in the Protein Data Bank (PDB), we cannot extrapolate their properties to plant HISN3. There are significant differences in amino acid composition which may influence, e.g., substrate recognition and product release. In this work, we provide the first crystal structures and enzymatic characterization of a plant HISN3 enzyme. The model legume *Medicago truncatula* was used as the source organism. We obtained the structures of the *Mt*HISN3-D57N mutant in complex with the substrate (ProFAR) and product (PrFAR) to provide the first insights into the HISN3 architecture in plants. Based on the structural data, we inferred

the roles of the additional elements present only in plant HISN3s. In particular, we focused on the function of a flexible β-turn localized near the active site, which is specific to plants. We deduced its role based on molecular dynamics (MD) simulations, supported by comparative steady-state kinetics of wild-type (WT) *Mt*HISN3 versus a deletion mutant (Δ74-80). We also present the results of a large-scale virtual screening (VS) campaign using a library of over 1.3 million compounds, which sets the direction for future herbicide development. Finally, we analyzed the evolutionary relationships between HISN3 homologs across all kingdoms of life to identify elements that could ensure kingdom selectivity, potentially leading to the development of herbicides, antifungals, and antibiotics.

2. Results and discussion

2.1. Overview of MtHISN3 complexes with substrate and product

The gene encoding *Mt*HISN3 is located on chromosome 2 (Gene ID: 11441829; LOC11441829) and is transcribed in one form (XM_003593639.4). The coding sequence is translated into a polypeptide of 312 residues (UniProt ID: G7IFI7). Similar to the other enzymes in the plant HBP, *Mt*HISN3 contains an N-terminal chloroplast transit peptide. According to *TargetP 2.0* (Emanuelsson et al., 2000), *Mt*HISN3 has a 48-residues long thylakoid luminal transit peptide. However, we based our constructs on *AlphaFold* (Jumper et al., 2021) predictions, and N-terminally truncated the construct at Ser42 (Nt42) to reflect the mature enzyme. Nevertheless, attempts to crystallize WT *Mt*HISN3-Nt42 were unsuccessful, possibly because of the high flexibility of the loops at the catalytic site. To rigidify the enzyme, we co-crystallized the D57N mutant (*Mt*HISN3-Nt42-D57N), designed to be inactive, with the enzymatically produced ProFAR. We obtained two crystal structures of *Mt*HISN3-Nt42-D57N in complexes with ProFAR and PrFAR, both solved in the *P*4 space group, containing one protein chain in the asymmetric unit (Table 1). Although the structures are not isomorphous, their structural alignment shows high similarity, expressed as the root-mean-square deviation (RMSD) = 0.23 Å across 255 Cα pairs. In the *Mt*HISN3-Nt42-D57N_ProFAR complex, resolved at 2.36 Å resolution, the polypeptide chain starts at Ser42 and ends at Ala312, missing only the Lys75-Ser80 region. The complex of *Mt*HISN3-Nt42-D57N_PrFAR at 1.54 Å resolution includes residues spanning from Ser49 to Val309, except for the Asp76-Gly79 fragment. The role of this region is discussed in a separate section. *Mt*HISN3, a typical monomeric (βα)₈-barrel, has an interior formed by eight parallel β-strands surrounded by eight α-helices and contains a centrally located active site. The two sides of the barrel are hereafter referred to as the basal and catalytic faces (Fig. 2), according to the nomenclature introduced by others (Höcker et al., 2001; Söderholm et al., 2015). The basal face contains eight short loops and both N-, and C-termini. The catalytic face is characterized by much longer loops, which interconnect α-helices with β-strands. These two sides create a dipole-like charge distribution
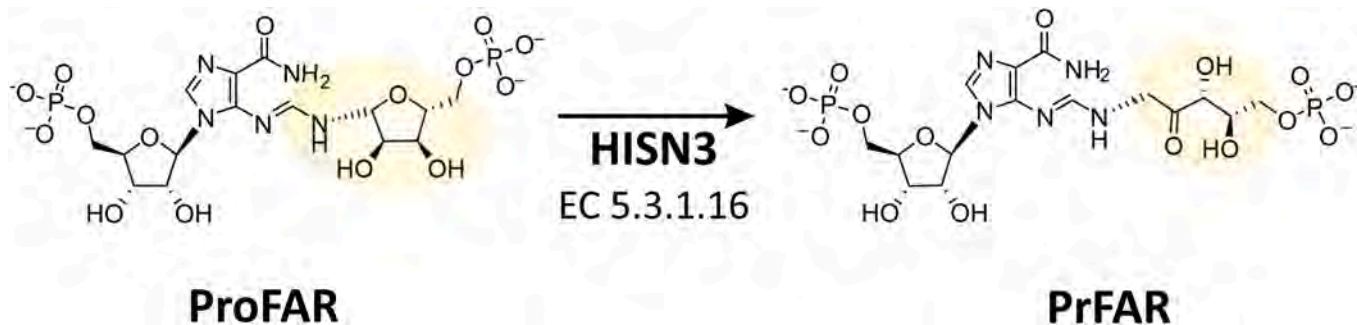
Fig. 1. The reaction catalyzed by HISN3 in plants. ProFAR undergoes isomerization to PrFAR via Amadori rearrangement. Yellow highlights mark the ribose and ribulose moieties in the substrate and product, respectively.

Back to p. 40

Table 1
X-ray data collection and refinement statistics.

	*Mt*HISN3-Nt42-D57N_ProFAR	*Mt*HISN3-Nt42-D57N_PrFAR
Beamline	PETRA III, Beamline P13, DESY Hamburg	PETRA III, Beamline P13, DESY Hamburg
Wavelength (Å)	0.9763	0.9763
Temperature (K)	100	100
Rotation range per image (°)	0.2	0.2
Total rotation range (°)	220	360
Space group	P4	P4
$a = b$, c (Å)	91.4, 35.9	87.9, 35.7
Mosaicity (°)	0.340	0.146
Resolution range (Å)/highest resolution shell	80–2.36/2.50–2.36	80–1.54/1.63–1.54
No. of unique reflections	12538/1978	40420/6090
Completeness (%)	99.9/99.9	98.8/92.9
Redundancy	8.21/8.24	13.06/10.09
I/σ(I)	10.74/2.16	12.26/1.46
R_{meas} (%)	16.0/102.4	14.9/167.7
$CC_{1/2}$ (%)	99.6/84.2	99.8/59.8
No. of reflections: working/test set	12538/1001	40420/1011
R_{work}/R_{free}	0.222/0.254	0.173/0.197
No. of non-H atoms: Protein/Ligand/Water	2023/39/4	1976/101/304
R.m.s. deviations: Bonds (Å)/Angles (°)	0.005/0.821	0.006/0.904
Ramachandran plot: Most favored/allowed/outliers (%)	95.8/4.2/0.0	97.2/2.8/0.0
Average B-factor: Protein/water/ligands (Å²)	47.9/40.1/40.1	18.7/30.2/19.1

on the protein surface (Fig. 2A and B). The basal side is negatively charged, while the catalytic face (including the active site) shows a strong positive charge (Fig. 2B). The helix dipoles additionally contribute to the distribution of the electrostatic potential, as they all expose the positively charged N-termini towards the catalytic face, whereas the negatively charged C-termini make up the basal face. Such a polarized surface likely aids in guiding ProFAR to the active site.

Interestingly, while the crystal harvested after three weeks indeed showed the presence of ProFAR (Fig. 3), the one that was kept for four months contained mostly PrFAR (Fig. 4). Although the *Mt*HISN3-Nt42-D57N variant was inactive during kinetic measurements (not shown), some residual activity was probably retained, which led to PrFAR formation. This phenomenon has been reported by other groups in the study of *Salmonella enterica* HisA-D7N mutant (PDB ID: 5AB3) (Newton et al., 2017) and in *M. tuberculosis* bifunctional PriA-D11N (PDB ID: 2Y88) (Due et al., 2011).

To distinguish the moieties in ProFAR and PrFAR, we refer to N1-phosphoriboses and C5-phosphoribose (or C5-phosphoribulose, exclusively for PrFAR) (Fig. 3B, C, and Fig. 4B and C). Interestingly, during the refinement of the 1.54 Å complex, we discovered an alternative conformation of the mutated residue Asn57 (refined occupancies of 0.57 and 0.43). The minor conformation would clash (distance of 1.3 Å) with the 2′ carbonyl of C5-phosphoribulose. Therefore, we deduced that the second Asn57 conformer exists together with the water elimination product, PrFAR-E, trapped in the active site (Fig. 4D). The structure refined very well with the PrFAR-E molecule at 0.43 occupancy, and owing to the different geometry it was possible to clear all difference map peaks near the ligand. To the best of our knowledge, no other case of such elimination in 5′-ProFAR isomerases has been reported.

ProFAR and PrFAR share almost identical pattern of interactions with the surrounding residues (Fig. 3B and C and 4B, C). The N1-phosphoriboses are hydrogen-bonded to the backbones of Gly138, Thr158 and Ser159, and the side chains of Ser159, Asp187, Ser189, Trp204, and His230. The aminoimidazole moieties interact by hydrogen bonds with the backbones of Gly68 and Arg203, by π-π stacking with the

side chain of Trp204, and by weak π-σ contact with the Leu112 side chain. The C5-phosphoribose and C5-phosphoribulose are hydrogen-bonded to the backbones of Gly236, Gly262, Gly285, and Ser286, and to the side chains of Asn57, His108, His230, Glu235, and Ser286. The side chain of Gln65 participates in the binding of the PrFAR-E C5-moiety by hydrogen bonding to C4′ –OH. The binding of the phosphates of both C5-phosphoribose and C5-phosphoribulose is additionally strengthened by sodium cations (at a position referred to as Na1). On the protein side, Na1 is coordinated by the backbone carbonyl of Ile66, the side chains of Gln65 and Glu235, and by a water molecule only in the complex with PrFAR (Figs. S1A and B). Notably, the presence of sodium has been inferred by the analysis of bond lengths and geometry (deduction of H_2O/Na^+/K^+/Mg^{2+}), and by the size of the peak in the electron density map (Na^+/K^+). The identity of the metal was additionally verified with the CheckMyMetal server (Zheng et al., 2017).

The complex with PrFAR contains an additional sodium cation (Na2) situated on the opposite side of the active site, approximately 7 Å from the PrFAR's N1-phosphoribose phosphate (Fig. S1B). Na2 is coordinated by the backbone carbonyls of Ser159 and Phe162, along with four water molecules. Close to Na2 (approx. 4 Å distance) resides a chloride anion (Cl1), which is hydrogen bonded to the backbone amide of Gly165 and three water molecules, two of which are shared with the sodium Na2 coordination sphere. Cl1 also occurs in the ProFAR complex. However, the complex with PrFAR contains yet another chloride anion (Cl2) co-ordinated by the backbone amide of Asn294 and two water molecules (Fig. S1B).

Ions are scarce in the available structures of HISN3 homologs. One Na^+ was found in *S. enterica* HisA (PDB ID: 5AB3) but was located near the protein termini on the basal side. The lack of sodium cations in the active sites of bacterial structures is probably due to the Arg residue at a position equivalent to Gln65, which coordinates Na^+ in *Mt*HISN3 and is conserved in plants. The Arg side chain likely repels the metal cations from the active site. The sodium cations (Na1) in the active sites of plant HISN3s stabilize phosphate moieties by increasing the positive charge which interacts electrostatically with the phosphate of C5-phosphoribosyl. The role of sodium cations is further discussed in this study.

Chloride anions were identified in *Arthrobacter aurescens* HisAp (PDB ID: 4WD0), where one Cl⁻ is located on the catalytic face and the other is located on the basal face, and in *Saccharomyces cerevisiae* His6 (PDB ID: 2AGK), which contains one Cl⁻ bound on the basal face. Based on the presence of chloride ions in the catalytic face of *Mt*HISN3, we conclude that they contribute to the structural organization of flexible regions, at least inside the crystals; therefore, they may indirectly improve substrate binding. However, it is difficult to deduce the *in vivo* function with current data.

The *Mt*HISN3 complex with PrFAR also contains three 2-methyl-2,4-pentanediol (MPD) molecules (Fig. S1C). MPD1 interacts with the enzyme via two water molecules, mediating the interaction with the side chain of Gln166 and the backbone of Met170. MPD2 is bound by the side chain of Glu218. MPD3 is bound by the side chains of Glu270, Arg273, and Gln307. MPD binding occurs at the crystal contacts and is rather non-specific, based on their occurrence in regions of different surface potentials.

2.2. *MtHISN3 is a highly efficient enzyme*

We performed steady-state kinetic measurements of WT *Mt*HISN3-Nt42 and *Mt*HISN3-Nt42-D57N. ProFAR is an unstable molecule with an estimated half-life of 953 min at 37 °C (Henn-Sax et al., 2002), therefore, it is commercially unavailable. We synthesized ProFAR enzymatically from 5-phosphoribosyl-1-pyrophosphate (PRPP) and adenosine-5′--triphosphate (ATP) using *Mt*HISN1, *Mt*HISN2, and *Escherichia coli* inorganic pyrophosphatase (*Ec*PPA). To measure the activity of WT *Mt*HISN3-Nt42 (at 50 nM concentration), we used a coupled assay with an excess of *Mt*HISN4-Nt48, according to the procedures used for
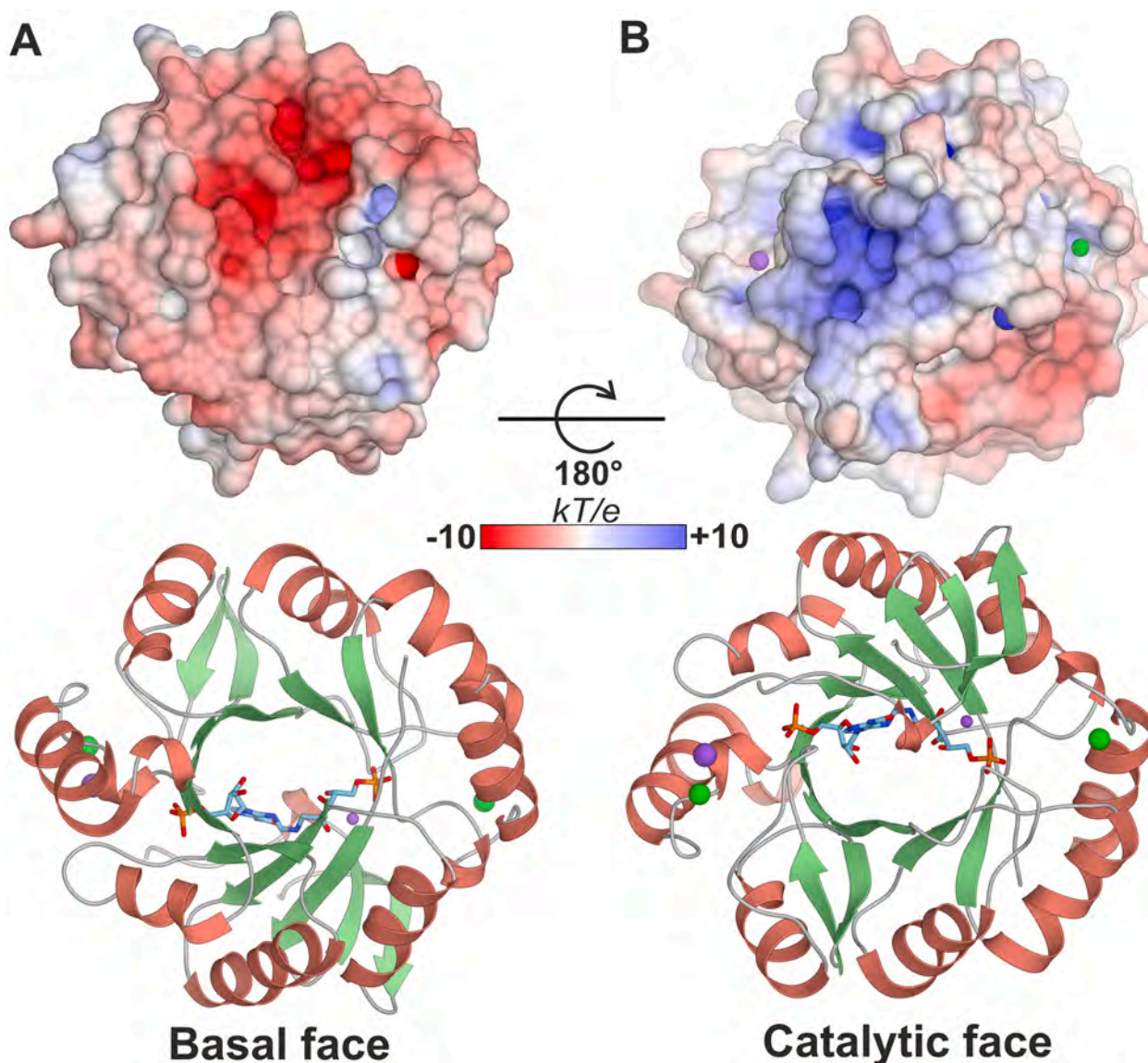
Fig. 2. Structure overview and positioning of PrFAR in the active site. Panels A and B depict the basal and catalytic faces in *Mt*HISN3, respectively. The scale bar represents surface potential values regarding the upper images. Ribbon diagrams are colored by secondary structure elements. Green and purple spheres are chloride and sodium ions, respectively.

non-plant HISN3 homologs (Klem and Davisson, 1993; Jürgens et al., 2000; Kuper et al., 2005; Söderholm et al., 2015). *Mt*HISN4-Nt48 converts PrFAR created by WT *Mt*HISN3-Nt42 into imidazole-glycerol phosphate (IGP) and 5-aminoimidazole-4-carboxamide ribonucleotide (AICAR). The processing of PrFAR by HISN4 is associated with a decrease in absorption at 300 nm. This approach also mitigates the effect of product inhibition on the HISN3 reaction. We fitted the data to the Michaelis-Menten equation, which resulted in $K_M = 4.6 \pm 1.0$ μM, $k_{cat} = 9.7 \pm 0.4$ s^{-1}, yielding the catalytic efficiency (k_{cat}/K_M) of 2.1 μM^{-1} s^{-1} (Fig. 5). The relatively high error for K_M results from low signal to noise ratio, especially at ProFAR concentrations below 5 μM. The k_{cat} value was not affected by this issue since it corresponds to measurements at high concentrations of the substrate.

We also checked the activity of the *Mt*HISN3-Nt42-D57N mutant (at 1 μM concentration), using 200 μM ProFAR, coupled with 10 μM *Mt*HISN4-Nt48. No measurable activity was detected in a 10-min assay. However, based on the electron density spotted in the 1.54 Å complex, we deduced that this mutant may have possessed a residual activity inside a crystal, resulting in the formation of the main product, PrFAR, and the side product PrFAR-E after four months. We excluded the

possibility a of non-enzymatic conversion of ProFAR to PrFAR or PrFAR-E based on the results obtained by Davisson et al., which showed that the decomposition of ProFAR leads to the formation of AICAR and not PrFAR (Davisson et al., 1994). Moreover, blank experiments (including all reaction components except for *Mt*HISN3-Nt42) showed no change in absorbance (not shown).

Literature contains kinetic data about bacterial homologs of *Mt*HISN3 (including bifunctional PriA), but no information is available for plant and fungal homologs. Among the bacterial examples reported in BRENDA (Chang et al., 2021), the highest substrate affinity was reported for *Thermotoga maritima* HisA, $K_M = 0.6$ μM (at 25 °C), and the highest turnover number was measured for *E. coli* HisA, $k_{cat} = 4.9$ s^{-1} (Henn-Sax et al., 2002). Therefore, despite a moderate K_M, the catalytic efficiency of *Mt*HISN3, is higher than that of most bacterial HisA and PriA enzymes, except for *E. coli* ($k_{cat}/K_M = 3.1$ μM^{-1} s^{-1}).

2.3. Genetic background and phylogenetic analyses

Scientists have extensively studied the evolution of bacterial HisA enzymes and their relationship with ancestral-like PriA enzymes that
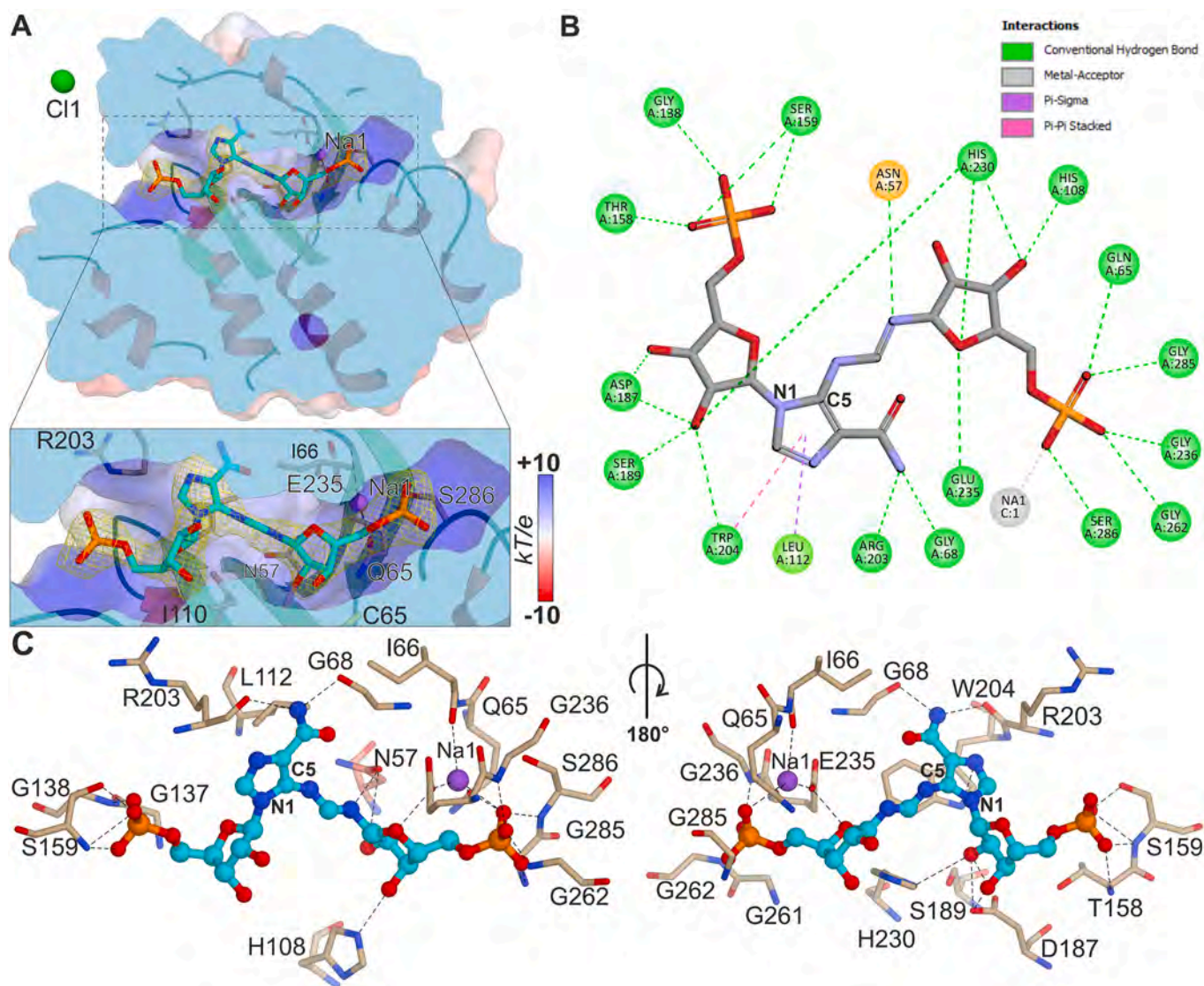
Fig. 3. Molecular interactions between ProFAR and *Mt*HISN3. Panel A shows a cross-section of *Mt*HISN3 and a close-up view of the active site pocket. Surface potential is colored as in the scale bar. The polder map for ProFAR is contoured at σ = 3.7. Panel B is a schematic representation of interactions between ProFAR and surrounding residues. The legend in the top right corner shows the types of interactions. All interactions are within the range of 2.5–4.5 Å. Panel C depicts residues interacting with ProFAR (balls and sticks) within a range of 4 Å. The front-plane residues and water molecules were removed for clarity.

display HisA and TrpF activities (Jürgens et al., 2000; Höcker et al., 2001; Henn-Sax et al., 2002; Barona-Gómez and Hodgson, 2003; Kuper et al., 2005; Lundin et al., 2020). However, little is known about the evolution of eukaryotic HisA homologs, such as His6 and HISN3. To obtain new insights, we combined two computational approaches. The sequence similarity network (SSN) revealed that plant HISN3s and fungal His6s form a large cluster together with HisA sequences from *Bacillota*, suggesting a non-cyanobacterial origin of plant HISN3 enzymes (Fig. 6). To eliminate bias, we also calculated a phylogenetic tree which showed that homologs in *Cyanobacteria* diverged twice from the last common ancestor shared by *Archaea* and *Eukaryota* (Figs. S2 and S3). HISN3 is the third case of plant HBP enzyme of a likely non-cyanobacterial origin. The other two enzymes are HISN2 (Witek et al., 2021) and HISN6 (Rutkiewicz et al., 2023), which could have been acquired by a horizontal gene transfer. The SSN results additionally revealed that some *Actinomycetota* possess bifunctional PriA enzymes, which supports their ancestral role in the evolution of (βα)₈-barrels, possibly spread by a horizontal gene transfer from *Actinomycetota* to other bacterial groups.

The HISN3's (βα)$_8$ barrel structure most presumably emerged from a gene duplication of an ancient, promiscuous, (βα)$_4$ priA-like enzyme with low substrate specificity, and therefore low catalytic efficiency (Fani et al., 1994; Lang et al., 2000). Subsequent diversification in plants resulted in a slightly higher catalytic efficiency (through higher k_{cat}) than that of many bacterial HisA homologs or moonlighting ancestral enzymes. Evolutionary pressure also promoted the HBP regulation in plants beyond HISN1. We described the inhibition of *Mt*HISN2 by adenosine-5′-monophosphate (AMP), which acts as a second-tier regulator of the HBP flux (Witek et al., 2021). This feature enables regulation in response to the metabolic state of the cell, expressed by the AMP/ATP ratio. This represents an evolutionary novelty, as in bacteria, histidine biosynthesis is regulated at the genetic level through the expression of the *his* operon and at the HISN1-equivalent step. Since eukaryotic genomes are not organized into operons, plants have had to develop alternative regulatory mechanisms. It is, therefore, extremely interesting to note the occurrence of a highly efficient HISN3 immediately after the AMP-regulated HISN2. This suggests a very tight control of the available ProFAR pool in the plant cell.
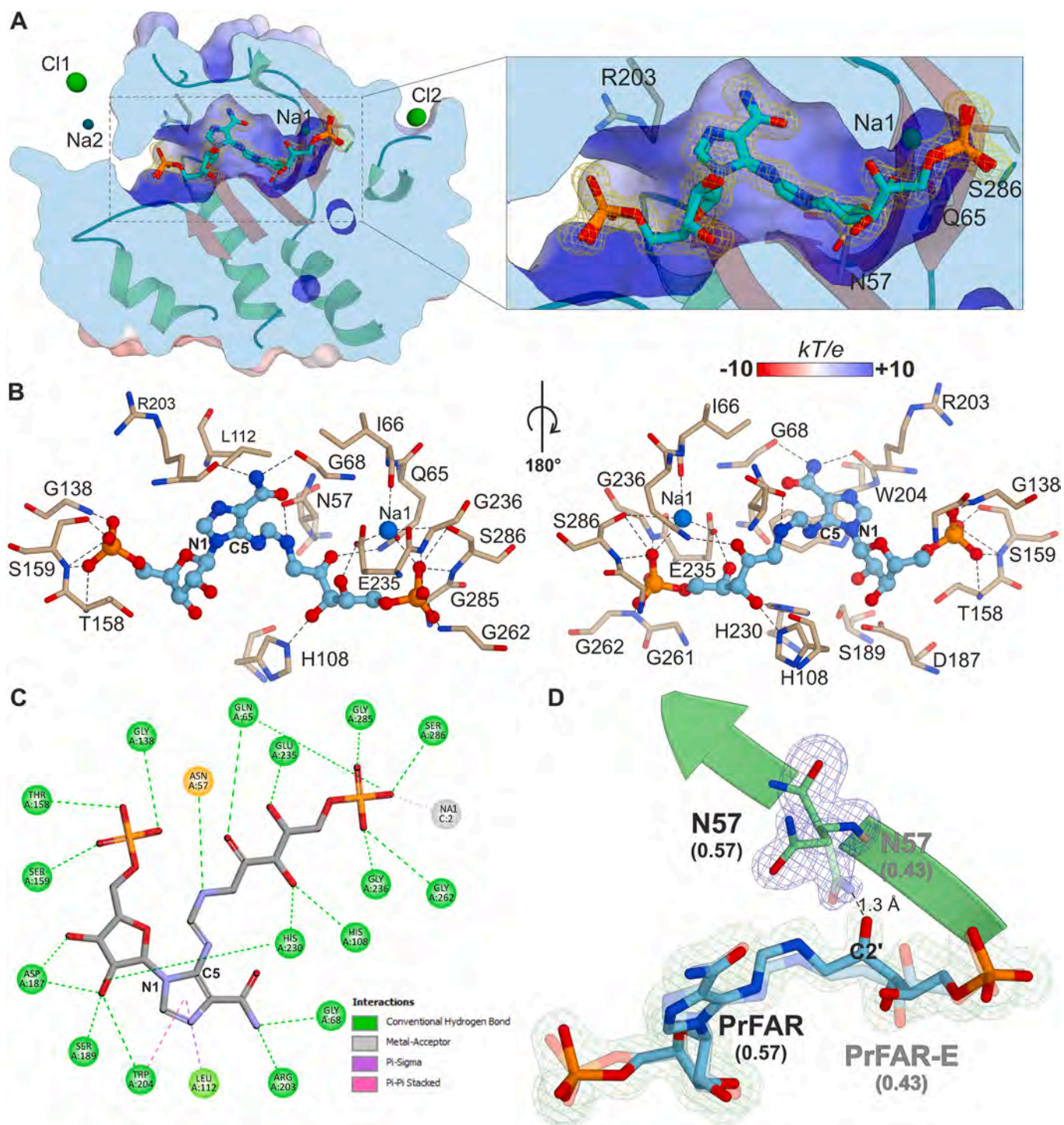
Fig. 4. PrFAR and PrFAR-E interactions with protein residues. Panel A shows a cross-section through the *Mt*HISN3 active site, where both ligands are located. Panel B depicts residues interacting with PrFAR (balls and sticks) within a range of 4 Å. The front-plane residues, PrFAR-E, and water molecules were removed for clarity. Panel C shows a simplified network of interactions with PrFAR. The legend in the bottom right corner describes types of interactions. Panel D shows a relationship between the double conformation of Asn57, PrFAR, and PrFAR-E. The values in brackets indicate atom occupancies. The distance of 1.3 Å shows an initial clash which was visible during the early stages of structure refinement. PrFAR-E and corresponding Asn57 conformation are 60% transparent. Polder maps were contoured at σ = 5.1 for Asn57 and σ = 4.1 for PrFAR and PrFAR-E.

2.4. Residue conservation among HISN3 homologs and structure comparison

We were interested in the structural differences that could explain the differences in the turnover rate of *Mt*HISN3 compared to bacterial counterparts. In addition, the identification of unique elements, variable between kingdoms could guide the design of kingdom-selective inhibitors, i.e., herbicides vs antibiotics. Therefore, we analyzed residue conservation in *Mt*HISN3 by comparing the results obtained for (i) a set of plant and (ii) bacterial, archaeal, plant, and fungal sequences (Fig. 7A and B, respectively). In general, the highest conservation is displayed by the residues constituting the β-strands which participate in ProFAR/
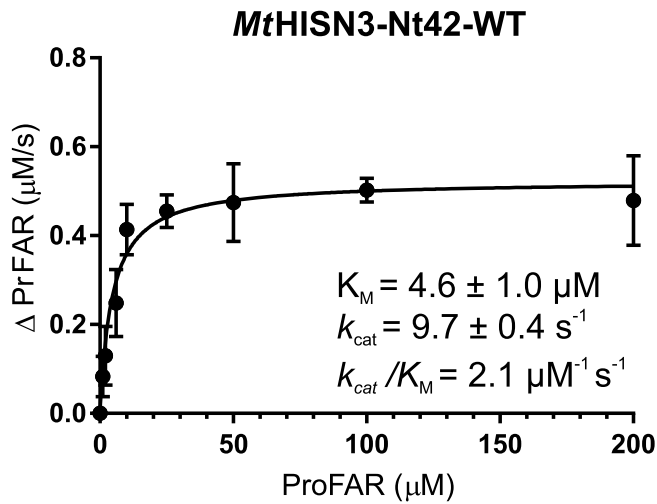
Fig. 5. Michaelis-Menten curve of *Mt*HISN3-Nt42-WT. ProFAR concentrations are placed on the X axis, and changes in the turnover rate are shown on the Y axis. Error bars derive from data fitting to the curve.

PrFAR stabilization in the active site (Fig. 7C). Loops in the catalytic face of the plant enzymes show higher conservation scores than those calculated for all organisms (Fig. 7A and B). In plants, α-helices are conserved mostly at residues that interact with the central β-strands, whereas those exposed to the solvent are more variable. Sequences from all organisms show high variability in these secondary structures, likely as a result of the evolutionary pressure caused by different compositions

of cytosol and chloroplast stroma. Our attention was first brought to a region in the basal face of *Mt*HISN3, which shows much stronger conservation in plants than in other organisms (Fig. 7A). We examined whether this region might serve as an exit from the active site of the enzyme after catalysis. However, the analysis conducted in *CAVER* (Chovancova et al., 2012) revealed a lack of tunnels connecting the catalytic and basal faces, indicating that both substrate binding and product release occur at the catalytic face. Hence, the function of this patch of conserved residues remains elusive.

To date, the only available structure of a eukaryotic HISN3 homolog has been the structure of His6 from *S. cerevisiae* (PDB ID: 2AGK). A structural alignment between *Mt*HISN3-Nt42 and its homologs, *Sc*HIS6 and *Se*HisA (PDB ID: 5A5W), yielded the RMSD values of 0.945 across 207 Cα pairs for 2AGK, and 1.128 across 121 Cα pairs for 5A5W (Fig. 8). Despite the high overall similarity, we found two regions that vary significantly in each kingdom. Notably, the 2AGK model lacks a central segment of a short loop (spanning from His178 to Asp190), which is present in the *Se*HisA model (5A5W) and in *Mt*HISN3, where it comprises residues $_{230}$HGVDVEGKKLGIDE$_{243}$. This region stabilizes the C5-phosphoribose moiety of ProFAR, with two residues directly responsible for substrate binding in *Mt*HISN3, i.e., Glu235 and Gly236 (Fig. 3).

The second region is a β-turn with an insert of an incomplete α-helix, comprising residues of the range ($_{60}$KGKVKQIVGSTLKDLKDDDGSDPITN-FESDKS$_{91}$), which correspond to Asn12-Pro40 in *Sc*HIS6. In the *Mt*HISN3-Nt42-D57N_PrFAR complex, this motif is almost completely visible and lacks only four residues (Asp76-Gly79). Notably, for the conservation analysis, we manually added the residues $_{76}$DDDG$_{79}$ into the model using Coot (Emsley et al., 2010) with respect to proper geometry, followed by energy minimization in Maestro (Schrodinger) (Fig. 7C). The entire
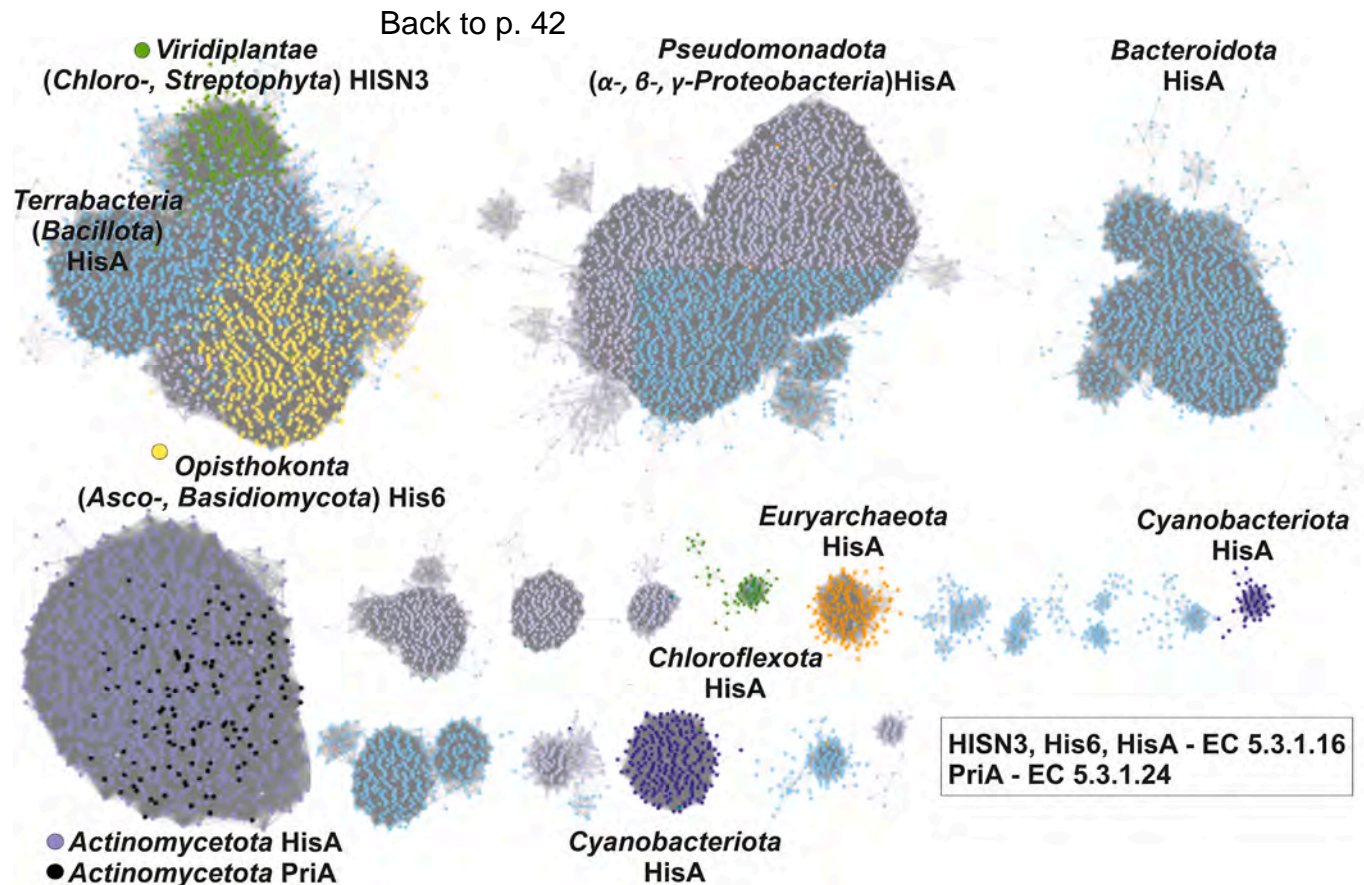
Fig. 6. The SSN comprises sequences of eukaryotic HISN3 and His6 enzymes, archaeal and bacterial monofunctional HisA enzymes, and bacterial bifunctional PriA enzymes. Eukaryotic HISN3 and His6 enzymes are grouped together with *Bacillota* and are separated from *Cyanobacteriota,* which indicates the non-cyanobacterial origin of plant sequences. Bifunctional PriA sequences clustered together with monofunctional HisA sequences indicate an ancestral role of the former. The network is refined to 8716 nodes and the EC numbers of the enzymes are in the bottom right corner.
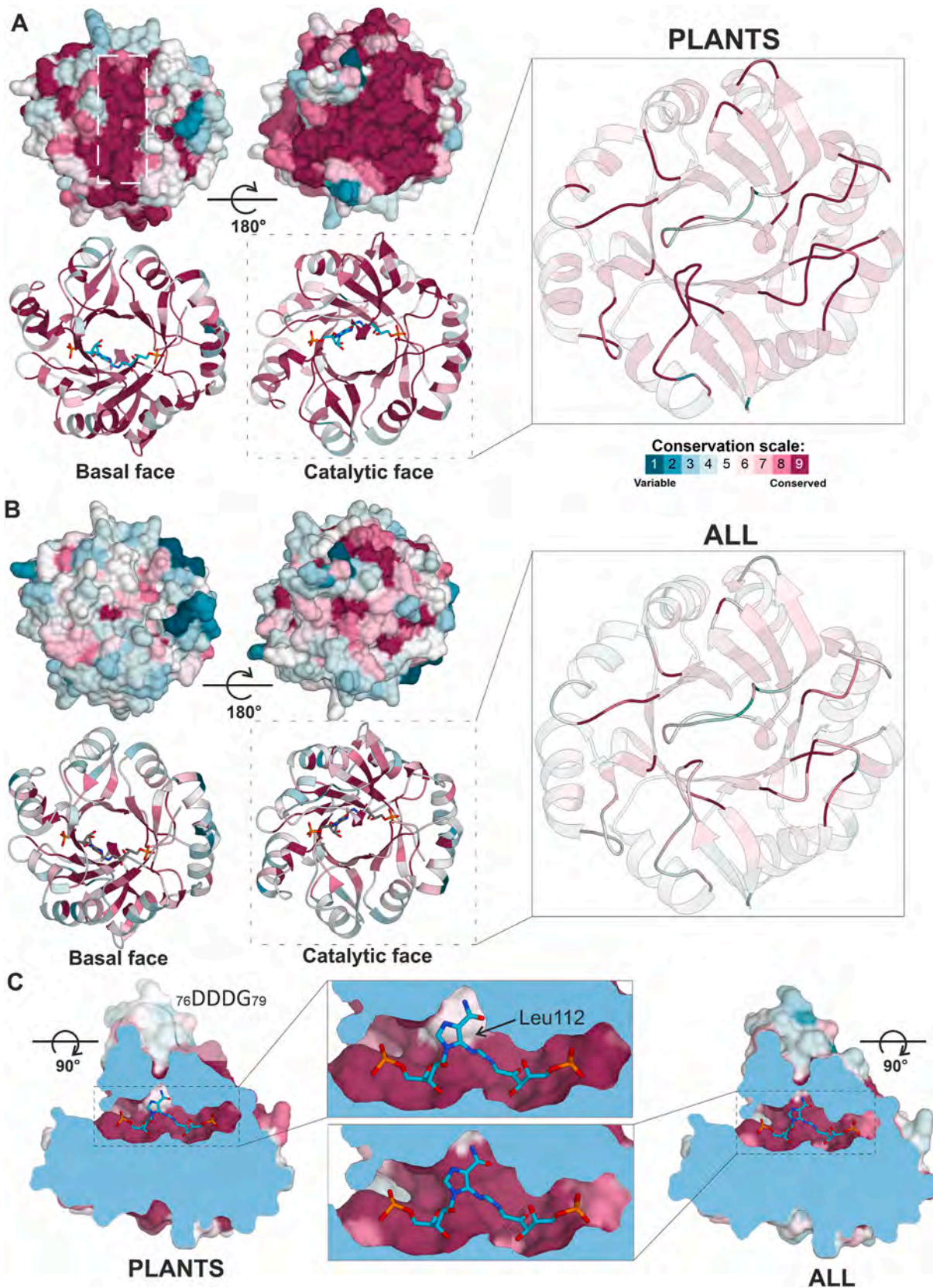
Fig. 7. The residue conservation in HISN3/HisA enzymes based on the MSA. Panels A and B show the residue conservation on the protein's surface and secondary structure elements in plants and all organisms, respectively. The white rectangle in panel A marks a region of unknown function that is highly conserved in plants. Close-ups on the right side focus on differences in loops conservation. Other secondary structure elements are 80% transparent. Panel C shows a cross-section through the active site. The sequence in the corner indicates a missing loop which was added to the model used in this analysis. All structures were colored regarding the conservation scale between the panels A and B. Notably, Leu112 received a lower conservation (compared to surrounding residues) but still is more conserved in plants (score -1.178) than in all species (-0.191).
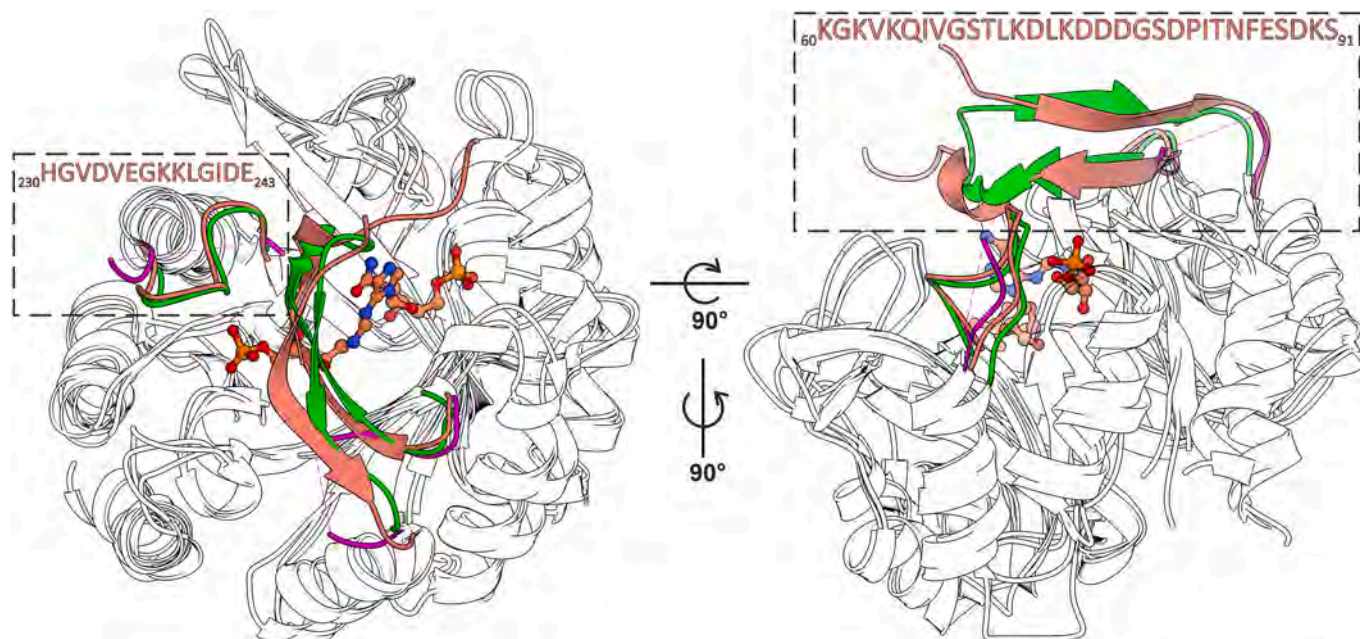
Fig. 8. Superposition of *Mt*HISN3 (salmon) with *Sc*His6 (magenta) and *Se*HisA (green). The major differences are marked with dashed rectangles; sequence fragments from *Mt*HISN3 are given in salmon. PrFAR is shown in balls and sticks. Dashed lines mark breaks in the polypeptide chain due to lack of electron density.

Lys60-Ser91 motif covers the active site and sticks out of the catalytic face. It contains residues that are conserved in plants, e.g., Ile66 and Gly68, and have a direct contact with ProFAR/PrFAR. It is also longer comparing to available bacterial structures, approx. 30 residues in eukaryotes vs 20 residues in bacteria. Comparison of this region to bacterial structures revealed that a similar but shorter (30 vs 20 residues) β-turn is present in *S. enterica* HisA (PDB ID: 5AB3). In structures that contain the substrate/product, this motif bends towards active sites, e.g., PDB IDs: 2Y88, 5A5W. A similar conformation was observed in the structure of *S. coelicolor* bifunctional PriA (PDB ID: 2X30), which contains two phosphate ions, but lacks the substrate/product in the active site. In the structure of apo-HisA from *Campylobacter jejuni* (PDB ID: 4GJ1), the β-turn faces the solvent and acts as an open gate for the active site.

2.5. The molecular function of the plant-specific Lys60-Ser91 fragment

As the Lys60-Ser92 fragment is the most characteristic element in *Mt*HISN3, we further investigated its involvement in catalysis. The central part of this region contains the $_{74}$LKDDDGS$_{80}$ motif, whose three Asp residues could potentially bind metal cations. Because ProFAR is also negatively charged, we asked whether a divalent cation, such as Mg^{2+}, present in our kinetic buffer, could at least transiently mediate the substrate recognition or product release. The co-occurrence of cofactor magnesium (or other divalent metals) in tandem with sodium in sodium/potassium-binding enzymes has been reported in the literature (Sigel et al., 2016). We tested the divalent metal-dependence using 10 mM ethylenediaminetetraacetic acid (EDTA), which did not affect the catalysis (50 μM ProFAR concentration, not shown). We then produced a deletion mutant, *Mt*HISN3-Nt42-Δ74-80, and measured its properties (Fig. S4). Surprisingly, the k_{cat} decreased 12-fold (9.7 s^{-1} for WT vs 0.8 s^{-1} for Δ74-80), whereas K_M increased 2.5 times (4.6 μM vs 11.9 μM). Altogether, deletion of the $_{74}$LKDDDGS$_{80}$ fragment caused a 30-fold loss in the catalytic efficiency (k_{cat}/K_M = 2.1 vs 0.07). Therefore, while the $_{74}$LKDDDGS$_{80}$ fragment clearly contributes to the high catalytic efficiency of *Mt*HISN3, the rationale does not lie in its potential capabilities of binding divalent metal cations. Notably, proper folding of the *Mt*HISN3-Nt42-Δ74-80 variant was inferred from the fact that it retained partial activity and eluted as a single peak in size exclusion chromatography (SEC, not shown), which corresponded to that of *Mt*HISN3-Nt42.

To determine the possible molecular role of the Lys60-Ser91 fragment, we employed 2-μs long MD simulations. Because MD is more powerful for observing dissociation events than association, we focused on the PrFAR complex and analyzed the simulation trajectories. Depending on the MD simulation replica, PrFAR either remained in the active site for the entire duration or dissociated after approximately 1000 ns (Fig. 9). Notably, the replicas were not biased toward either scenario. When PrFAR remained in the active site, after the initial dissociation of N1-phosphoribosyl, the rest of the molecule formed numerous transient bonding interactions (Fig. S5). PrFAR N1-phosphoribose was the first to lose the binding interactions seen in our experimental structures also in the scenario when the product dissociated (Fig. 9A and B). In this case, the Na1 cation dissociates after ~300 ns from its position in our crystal structures. However, between one and three Na$^+$ cations localize nearby until PrFAR leaves, indicating that Na$^+$ indeed participates in ligand recognition as observed in our structures. In the next steps, PrFAR gradually left the active site, which was enabled by opening up of the Val200-Val210 region during the first 800 ns (Fig. 9C). The PrFAR dissociation occurred between 600 and 1000 ns. Strikingly, between 900 and 1000 ns, PrFAR was forming multiple transient interactions with the Val67-Ile83 fragment (Fig. 9D). In addition to changes in the number of protein-ligand contacts and the ligand RMSD, the dissociation of PrFAR from the active site was reflected in fluctuations of (i) radius of gyration, (ii) number of intramolecular H-bonds, and the solvent-accessible surface area (Fig. 9E). Interestingly, the Lys60-Ser91 fragment constitutes the most dynamic part of *Mt*HISN3, as revealed by the analysis of per-residue root-mean-square-fluctuation plot (Fig. 9F). Altogether, data suggest that the role of the Lys60-Ser91 fragment is in aiding the product release. It is also likely that equivalent transient interactions form with ProFAR, and as such would facilitate the substrate recognition.

The multiple sequence alignment of HISN3 sequences from all kingdoms revealed that the fragment Lys60-Ser91 is highly conserved only in plants (not shown), with minor differences in the Lys75–Ser80 fragment. Consequently, the behavior exhibited by the *Mt*HISN3's Lys60-Ser91 equevalents is likely similar in other plants. Interestingly, certain species, such as soybean (*Glycine max*), completely lack this component, whereas other, like the herb *Senna tora*, miss only a single
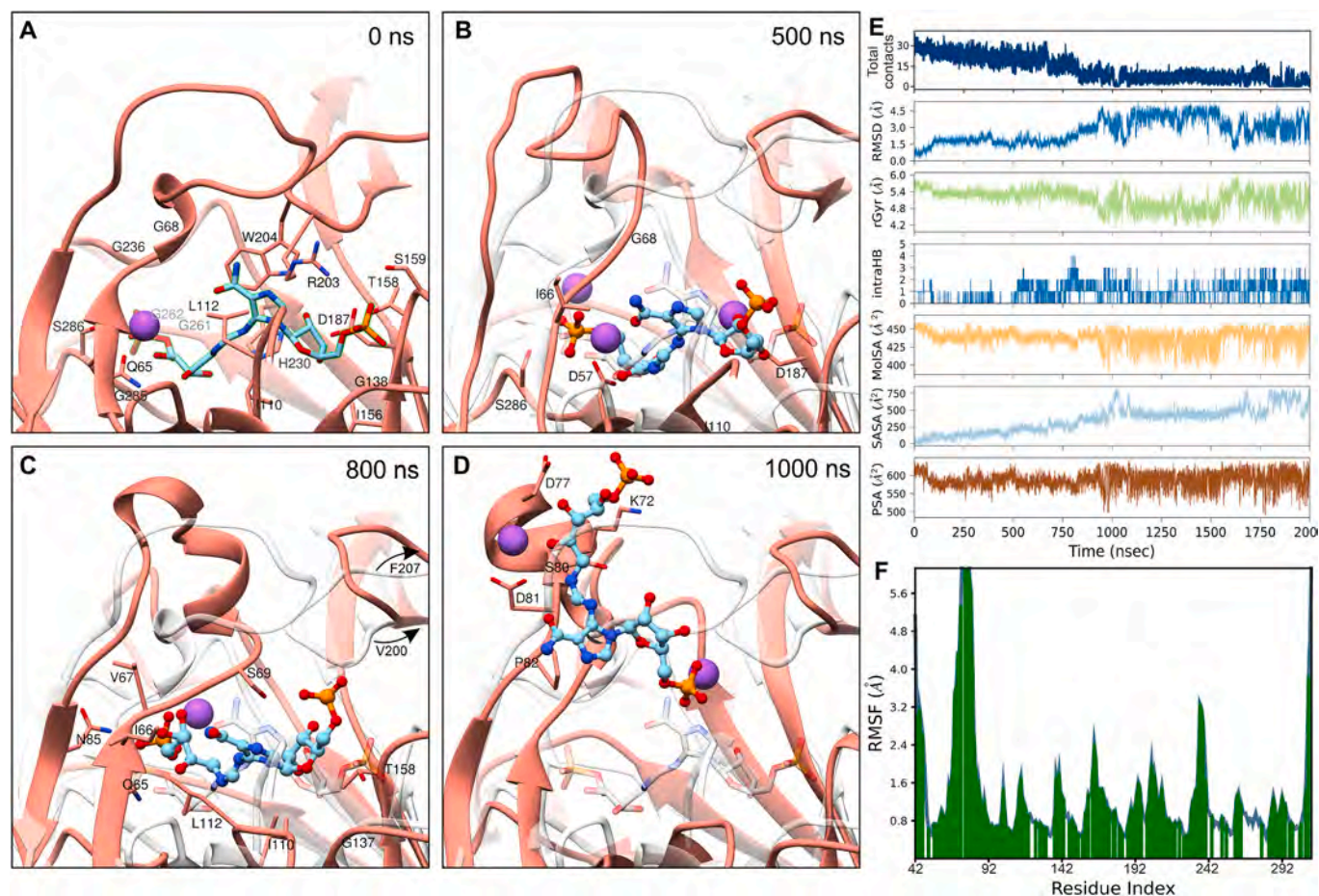
Fig. 9. Snapshots along the MD simulation trajectory in the case when PrFAR dissociates. (A) starting state. (B–D) representative states with the starting conformation superimposed as ghost (semitransparent). Protein residues within 3.5 Å distance from PrFAR are shown and labeled. The morphing of the 200–210 region is indicated in panel C by showing the positional shift of Val200 and Phe207. (E) Graphs showing changes of the following parameters over the simulation course: (i) total number of contacts; (ii) RMSD, root mean square deviation of a ligand with respect to the reference conformation (the first frame at t = 0); (iii) rGyr, radius of Gyration; (iv) intraHB, intramolecular Hydrogen Bonds; (v) MolSA, molecular surface area calculated with 1.4 Å probe radius, equivalent to a van der Waals surface area; (vi) SASA, solvent-accessible surface area; and (vii) PSA, polar surface area, contributed only by oxygen and nitrogen atoms. (F) Per-residue plot of root mean square fluctuations (RMSFs) for protein residues.

residue corresponding to Asp78 in *Mt*HISN3.

The MD simulations supported the product release at the catalytic face, meaning the function of the conserved path on the basal face (Fig. 7A) remains obscure. Moreover, an analysis of the trajectories did not reveal a state with the PrFAR phosphates bound at positions corresponding to Cl⁻ in our structures. This suggests that, if they play any role *in vivo*, the Cl⁻ ions are likely involved in maintaining the protein's structural integrity.

2.6. Virtual screening

To the best of our knowledge, the literature lacks even starting molecules for HISN3 inhibitors that could potentially be developed into herbicides. To search for initial hits and scaffold(s), we conducted a VS campaign. A total of 1,355,457 molecules representing a broad chemical space were obtained from the ZINC15 database (Sterling and Irwin, 2015) and docked into the active site of *Mt*HISN3 (Fig. 10A). Interestingly, the estimated binding energy gains were very high, compared to results obtained usually in AutoDock Vina (Trott and Olson, 2010), with the top five results having scored between −13.6 and −13.1 kcal/mol (Fig. 10B–F). This is promising and suggests strong druggability of HISN3 because such strong energy gains have not been observed in the primary VS on other HBP enzymes, including HISN5 (Witek et al., 2024) and HISN6 (Rutkiewicz et al., 2023).

The common feature of these five top-hit molecules is the presence of aromatic and polycyclic moieties that enable parallel π-π stacking with highly conserved Trp204 and π-σ interactions with Leu112. The identity of other substituents determined whether the molecule pointed towards the active site entrance (Fig. 10B,C,E,F) or towards the protein core (ZINC71752196, Fig. 10D). The compounds positioned at the active site entrance of *Mt*HISN3-WT formed the following hydrogen bonds: (i) ZINC9015982 with the backbone amides of Gly137 and Arg203, and with the side chain of Ser159; (ii) ZINC49510 with the backbone amides of Val157 and Thr158, as well as with the side chains of Thr158 and Trp204; (iii) ZINC9397396 with the backbone amides of Gly137, Thr158, Ser159, and Arg203, along with the side chain of Ser159; (iv) ZINC71903360 with the backbone amides of Gly138, Thr158, and Arg203, along with the side chain of Trp204 (Figs. S6B, C, E, and F, respectively). Meanwhile, ZINC71752196, positioned deeper at the protein's core, formed hydrogen bonds with the backbone amides of Gly262, Val263, and Ser286, as well as with the side chains of Asp57, Gln65, His230, and Ser286 (Fig. S6D). Given the high overall conservation of HISN3 and HisA/His6 sequences, molecule ZINC71752196 seems particularly interesting as it reaches Gln65, which is substituted with Arg in bacteria. Therefore, ZINC71752196 appears to be the best trade-off between the estimated energy gain and the potential to develop inhibitors specific to plant HISN3s.
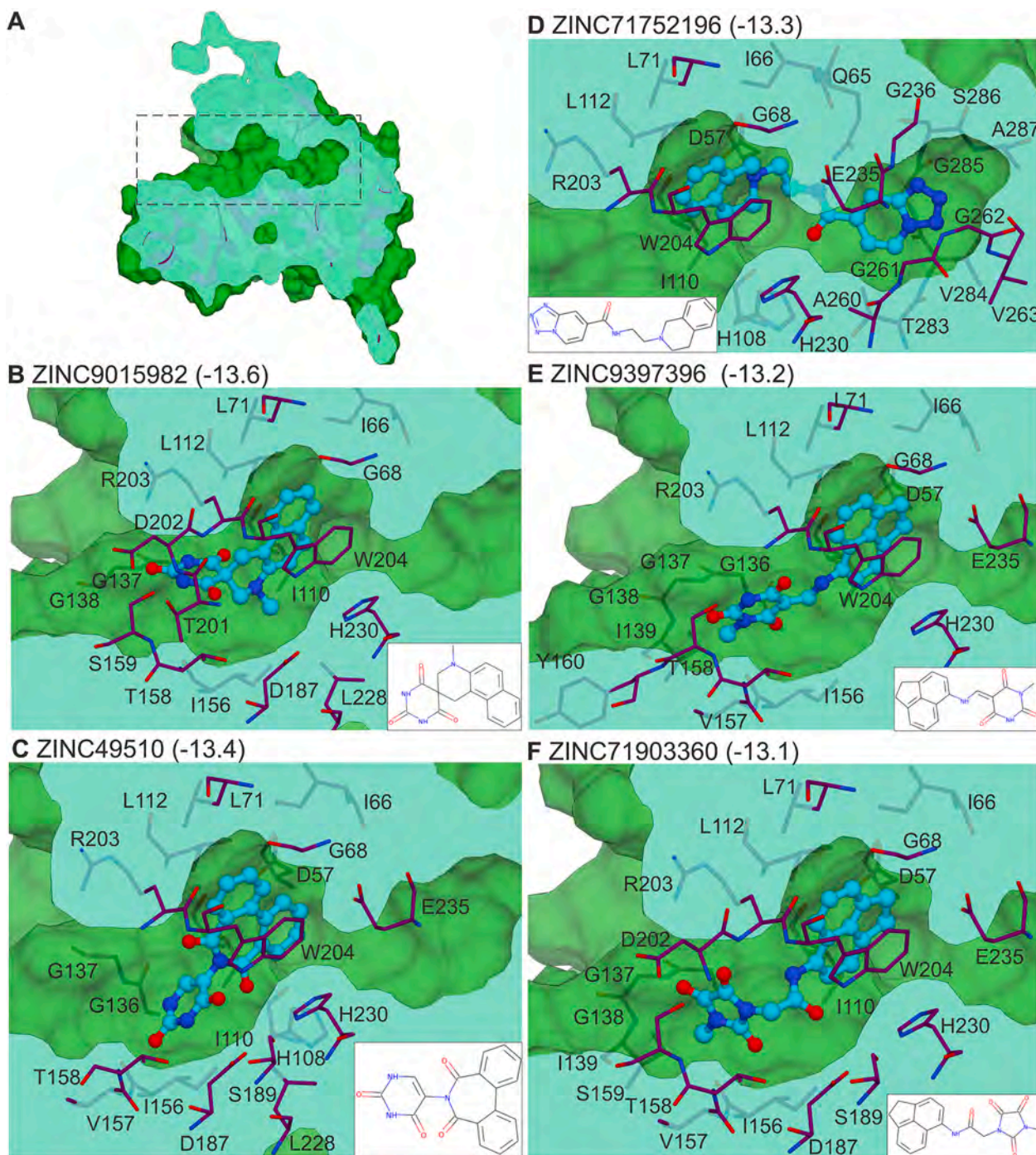
Fig. 10. Top five results of a virtual screening campaign. Panel A shows a cross-section through the *Mt*HISN3 surface. The active site pocket is marked by a dashed rectangle. Panels B–F show ligands of the highest energy gain, between −13.6 and −13.1 kcal/mol. Labeled are residues within a 3.5 Å range. Each panel contains a schematic view of the compound molecule in the corner.

3. Conclusions and outlook

In this work, we present the crystal structures of complexes of *Mt*HISN3-Nt42-D57N_ProFAR and *Mt*HISN3-Nt42-D57N_PrFAR, at 2.36 Å and 1.54 Å resolution, respectively. The crystal structures combined with the results of the activity measurements of WT *Mt*HISN3-Nt42 and the Δ74-80 deletion mutant allowed us to postulate the function of a conserved β-turn in ligand recognition. This fragment appears essential to ensure the high catalytic efficiency, exhibited mostly by the high k_{cat}. Sodium cations, located in the active site, participate in substrate and product binding. Plant HISN3s, unlike bacterial orthologs, possess the

Gln65 residue, which enables metal binding, yet another feature likely key to effective operation. This work brings us only a step away from the structural characterization of the entire plant HBP, missing only the experimental structures of HISN4.

Performed within the scope of his work phylogenetic and conservation analyses implied the non-cyanobacterial origin of plant HISN3 enzymes. Our results suggest that HISN3 is yet another enzyme of the HBP, which seemingly was acquired by the horizontal gene transfer. However, with current data it is very difficult to answer whether the transfer was driven by the cyanobacterial ortholog not satisfying the evolutionary pressure.

Studies of the HBP in plants can help develop new herbicides that target the pathway enzymes. In addition to causing histidine starvation, the inhibition of *Mt*HISN3 could have a strong impact owing to the impaired purine synthesis. The implemented VS campaign indicated five molecules with calculated energy gains stronger than −13.1 kcal/mol. Comparison of their molecular features can now inspire the development of herbicides acting on HISN3 to bring a remedy to the rapidly growing HR in weeds.

4. Materials and methods

4.1. Cloning, mutagenesis, expression and purification

Total RNA was isolated from young *M. truncatula* sprouts using the Universal RNA Purification Kit (E3598 EurX). RNA was reverse-transcribed using Super-Script II reverse transcriptase (Life Technologies). The coding sequence (CDS) of *Mt*HISN3 was obtained from the National Center for Biotechnology Information (NCBI) database (Gene ID: 11441829, XM_003593639.40). The construct was N-terminally truncated at Ser42 to remove the chloroplast transit peptide predicted with *TargetP 2.0* webserver (Emanuelsson et al., 2000) and *AlphaFold* (Jumper et al., 2021). Complementary DNA (cDNA) was amplified using Polymerase Chain Reaction (PCR) with Platinum SuperFi II PCR Master Mix (Thermo Fisher Scientific). The following primers were used for the PCR amplification (i) WT *Mt*HISN3-Nt42-FWD: TACTTCCAATC-CAATGCCTCTCCACCTTCAATTCAATGCTCCGTTCAATTC, (ii) WT *Mt*HISN3-Nt42-REV: TTATCCACTTCCAATGTTAAGCCACTGAGACCTT TTGCTGGTTATGC. The amplified *HISN3* gene was purified (PCR/DNA Clean-Up Purification Kit, Eurx) and embedded into the pMCSG68 (Midwest Center for Structural Genomics) vector using Ligation Independent Cloning (LIC) (Kim et al., 2011). The pMCSG68 plasmid containing the his-tagged WT *Mt*HISN3-Nt42 insert was used to transform *E. coli* BL21 (DE3) Gold competent cells (Agilent Technologies) using the heat shock method. Plasmids were later isolated (Plasmid DNA Purification Kit, Eurx) and sequenced by an external company (Genomed, Poland) to confirm the correctness of the coding sequence.

To obtain *E. coli* template DNA for PCR, we used a thermally lysed (95 °C) bacterial cell pellet. The *Ec*PPA construct was designed using a coding sequence from NCBI (Gene ID: 913899) and amplified using the following primers: (i) *Ec*PPA-FWD: TACTTCCAATCCAATGCCAT-GAGCTTACTCAACGTCCCTGC and (ii) *Ec*PPA-REV: TTATCCACTTC-CAATGTTATTATTTATTCTTTGCGCGCTCGAAGGAG. The PCR product was then inserted into the pMCSG68 vector and sequenced.

The inactive mutant *Mt*HISN3-Nt42-D57N was produced using the Polymerase Incomplete Primer Extension (PIPE) method (Klock and Lesley, 2009). The substitution was introduced into the WT *Mt*HISN3-Nt42 expression plasmid using the following primers: (i) *Mt*HISN3-Nt42-D57N-FWD CCTGCATCAATATTCATAAGGGGAAAGTGAAACAAAT TGTTGG, and (ii) *Mt*HISN3-Nt42-D57N-REV CCTTATGAATATTGA TGCAGGGGCGGAATTGAAC. The correctness of the sequence was confirmed by sequencing. The deletion mutant *Mt*HISN3-Nt42-Δ74-80 was also constructed using PIPE, with the following primers: (i) *Mt*HISN3-Nt42-Δ74-80-FWD CCTTAAAGATGATCCAATCACTAATTTTGA GTCTGATAAATCGGCA, and (ii) *Mt*HISN3-Nt42-Δ74-80-REV GTGATTG GATCATCTTTAAGGGGTTGACCCAACAATTTGTTTCAC.

Overexpression was conducted in *E. coli* BL21 Gold cells cultivated in a lysogeny broth (LB) medium supplemented with ampicillin at 100 μg/ml. The cultures were incubated for 2.5 h at 37 °C and shaken at 180 RPM until OD_{600} reached 0.8. The temperature was lowered to 20 °C, and expression was induced using 0.25 mM isopropyl-D-thiogalactopyranoside (IPTG). The cultures were shaken for 18 h and centrifuged at 4000×*g* for 15 min at 4 °C. The bacterial pellet was resuspended in 35 mL of a binding buffer (Tris (tris(hydroxymethyl) aminomethane) 50 mM, pH 8.0; NaCl 500 mM; imidazole 20 mM;

glycerol 10%; tris(2-carboxyethyl)phosphine (TCEP) 1 mM). The bacterial suspension was deep-frozen at −80 °C and stored for purification.

The frozen bacterial suspension was thawed and sonicated in cycles consisting of 4 s of sonication, 26 s of break, with 5 min of active sonication. Cell debris was pelleted by centrifugation at 20,000×*g* for 30 min at 4 °C. The supernatant was poured into a column containing HisTrap HP resin (GE Healthcare), equilibrated with the binding buffer, and incubated for 3 min. The solution was then removed from the column using a VacMan pump (Promega). The resin with bound *Mt*HISN3-Nt42 was washed eight times with 60 mL of the cold binding buffer and eluted with 15 mL of elution buffer (50 mM Tris pH 8.0; NaCl 500 mM; imidazole 500 mM; glycerol 10%; and TCEP 1 mM). The solution was transferred into a dialysis tubing (SnakeSkin™ Dialysis Tubing, Thermo Fisher) with the 10-kDa membrane cutoff and supplemented with the final concentration of 0.2 mg/mL of tobacco etch virus (TEV) protease to remove the his-tag. The cleavage was conducted overnight at 4 °C in dialysis buffer (Tris 50 mM, pH 8.0; NaCl 500 mM and TCEP 1 mM). The sample was run on a column with equilibrated HisTrap HP resin, and *Mt*HISN3-Nt42 in the flow-through was collected from the column and concentrated to 2.0 mL using an Amicon Ultra 15 mL centrifugal filter (Merck) with a cutoff of 10 kDa. The concentrated protein solution was loaded onto a HiLoad Superdex75 column (GE Healthcare) previously equilibrated with the SEC buffer (Tris 50 mM, pH 8.0; NaCl 200 mM and TCEP 1 mM). Fractions containing *Mt*HISN3-Nt42 were collected, combined, and concentrated to 13 mg/mL using Amicon filters. The presence of the protein was confirmed by SDS-PAGE, and the concentration was assessed spectrophotometrically at 280 nm using a molecular weight of 29400 Da and an extinction coefficient (ε) of 23170.

4.2. Enzymatic synthesis of ProFAR

The enzymatic synthesis of ProFAR required two subsequent reactions (referred to as R1 and R2) according to the method developed by Ames et al. (1961), and modified by us. The synthesis was conducted in a UV microcuvette at 22 °C in an Agilent 8453 spectrophotometer, and both reactions were monitored at $\lambda = 290$ nm in a total volume of 1000 μL. Prior to the experiment, the kinetic buffer (50 mM Tris-HCl, pH 8.0; 5 mM $MgCl_2$; 100 mM KCl, 50 mM NaCl, and 1 mM TCEP) was brought to the ambient temperature. All enzymes used in synthesis were N-terminally His-tagged to enable their removal after synthesis. R1 mixture included 2 μM *Mt*HISN1 (Ruszkowski, 2018), 80 μM *Ec*PPa, and 4 mM ATP. The mixture was blanked, and the reaction was initiated by adding 4 mM PRPP (Sigma). After approximately 40 min, when R1 reached a plateau ($A_{290} = 0.8$), the mixture was blanked again, and the R2 reaction was started by adding 2 μM *Mt*HISN2 (Witek et al., 2021). R2 was conducted for approximately 25 min, and after reaching a plateau ($A_{290} = 1.4$), the mixture was transferred into an Eppendorf tube containing 300 μL HisTrap HP resin equilibrated in the binding buffer. The sample was incubated with the resin on ice for 5 min. After incubation, the Eppendorf tube was centrifuged at 1000×*g* for 2 min. The supernatant with ProFAR was transferred into a new, cooled on ice, UV-microcuvette, and the ProFAR concentration was measured spectrophotometrically at $\lambda = 290$ nm resulting in a 2.8 mM ProFAR ($\varepsilon_{290} = 7700$ M^{-1} cm^{-1}).

The enzymatic synthesis of ProFAR for steady-state kinetics was similar. R1 was conducted in kinetic buffer containing 40 μM *Ec*PPA, 1 μM *Mt*HISN1, and 2 mM ATP. The mixture was blanked and the first reaction was initiated by adding 2 mM PRPP. R1 was monitored at $\lambda = 290$ nm until it reached a plateau (50 min, A = 0.65) and the mixture was blanked again. To start the second reaction, 1 μM *Mt*HISN2 was added. R2 was conducted for approximately 20 min until a plateau at A = 0.8, after which *Mt*HISN1, *Mt*HISN2, and *Ec*PPA were eliminated, and ProFAR concentration was assessed as 0.81 mM.

4.3. Crystallization, X-ray data collection, and processing

The MtHISN3-Nt42-D57N_ProFAR and MtHISN3-Nt42-D57N_PrFAR complexes were crystallized using the vapor diffusion method in a hanging-drop setup. The complex of *Mt*HISN3-Nt42-D57N_ProFAR was obtained by a 1-h incubation (on ice) of *Mt*HISN3-Nt42-D57N at 12.0 mg/mL (414 μM) with ProFAR at 1.2 mM. The mixture was concentrated and diluted with SEC buffer to decrease the concentration of residual Mg^{2+} (final conc. approx. 2.5 mM). The final concentration of *Mt*HISN3 for crystallization was 8.5 mg/ml. Furthermore, the complex was mixed with the C11 condition from the ShotGun screen (Molecular Dimensions) (Fazio et al., 2014) at a 2.8:1.2 μL volume ratio. The C11 condition contained 0.2 M sodium acetate, 0.1 M sodium cacodylate pH 6.5, 15% PEG 8000. The reservoir solution was made up of 500 μL of 1.5 M NaCl. The crystallization was performed at 18 °C.

The MtHISN3-Nt42-D57N_PrFAR complex was prepared in the same way as the complex with ProFAR; however, crystallization occurred at 4 °C. The complex was mixed with the PEG/ION H10 (Hampton Research) solution at a 1:1.5 ratio. The PEG/ION H10 condition contained 0.2 M sodium bromide and 20% PEG 3350. The reservoir contained 500 μL of 1.5 M NaCl.

The crystals of the *Mt*HISN3-Nt42-D57N_ProFAR and *Mt*HISN3-Nt42-D57N_PrFAR complexes were harvested after three weeks and four months, respectively, from the crystallizations setup. The crystals of the *Mt*HISN3-Nt42-D57N_ProFAR complex were cryoprotected by mixing a crystallization condition solution with 1,2-ethanediol (EDO) which was added to a crystallization drop to a final concentration of 10% EDO. The crystals of the *Mt*HISN3-Nt42-D57N_PrFAR complex were prepared similarly, resulting in a final concentration of 20% MPD. Both groups of crystals were vitrified in liquid nitrogen and stored for diffraction measurements.

X-ray data were collected at the P13 Beamline of the PETRA Synchrotron in Hamburg, Germany. The datasets were processed using *XDS* (Kabsch, 2010); Table 1 presents the statistics.

4.4. Determination and refinement of the crystal structures

Structures of *Mt*HISN3-Nt42-D57N_ProFAR and *Mt*HISN3-Nt42-D57N_PrFAR were solved by molecular replacement (MR) using *Phaser* (McCoy et al., 2007). As a model, we used a prediction from *ColabFold* (Mirdita et al., 2022) that was based on the amino acid sequence of *Mt*HISN3-Nt42. Models were built in *Coot* (Emsley et al., 2010) and refined in *Phenix.refine* (Afonine et al., 2012). Ligand restraints were generated in *eLBOW* (Moriarty et al., 2009). Geometric parameters and model-map correlation were assessed using *Molprobity* (Williams et al., 2018). The atomic displacement parameters were refined isotropically using translation-libration-screw rotation parameters (TLS).

4.5. Steady-state kinetics

Steady-state kinetic measurements were performed at 22 °C using an Agilent 8453 UV–visible spectroscophotometer. The activity of WT *Mt*HISN3-Nt42, *Mt*HISN3-Nt42-D57N, and *Mt*HISN3-Nt42-Δ74-80 was measured at λ = 300 nm in a total volume of 800 μL in a UV-cuvette. The assay required an excess of *Mt*HISN4-Nt48 (Witek et al., 2024) to convert PrFAR into IGP (and AICAR), which was monitored as a decrease in absorbance at λ = 300 nm; therefore, *Mt*HISN3s and *Mt*HISN4 were used at 50 nM and 5 μM, respectively. ProFAR was diluted in kinetic buffer to the following concentrations: 200, 100, 50, 25, 10, 6, 2, and 1 μM, and *Mt*HISN4-Nt48, together with 5 mM L-glutamine, were added to the UV-cuvette. After a 5 min incubation, the solution was blanked, and the reaction was started by adding *Mt*HISN3. All measurements were conducted in at least duplicate. The

initial velocities were converted into a change in the PrFAR concentration increase (μM/s) using its extinction coefficient ($\varepsilon_{PrFAR} = 6069$ M^{-1} cm^{-1}) (Klem and Davisson, 1993). Kinetic parameters were calculated using nonlinear regression in *GraphPad Prism 6.07* (GraphPad Software), which also produced kinetic curves.

4.6. Phylogenetic tree

The phylogenetic tree was constructed using *Mega-X* (Tamura et al., 2021). The MtHISN3-Nt42 sequence was used as a query in the *BLASTP* (Altschul et al., 1997) search of UniProtKB/Swiss-Prot (UniProt, 2019) HISN3 sequences. Similarly, we used reviewed sequences of *E. coli* HisA (Uniprot P10371), *S. cerevisiae* His6 (Uniprot P40545), and *S. coelicolor* PriA (Uniprot P16250) to find respective homologous sequences. The percent identities of the retrieved sequences to the query sequences ranged 22–100%. The search resulted in 97 sequences that were aligned in *Mega-X* using the MUSCLE algorithm (Edgar, 2004) at default settings and 24 iterations. The multiple sequence alignment was used to create a maximum likelihood tree using the Jones-Taylor-Thornton (JTT) model. We used a hundred protein sequences from plants with 84–100 %ID and a hundred sequences from all organisms with 25–81 %ID. The phylogeny was tested using the bootstrap method, with 250 iterations. The *Sc*PriA sequence was chosen as the outgroup root (%ID to *Mt*HISN3-Nt42 = 29.8).

4.7. Sequence similarity network

The SSN was calculated using the EFI-EST web server (Zallot et al., 2019) from 42680 sequences combined into 18,961 UniRef90 clusters from the InterPro IPR044524 family. The sequences ranged from 200 to 320 residues with an alignment score of 90. Outlier sequences were rejected, resulting in the SSN being refined to 8716 nodes. The SSN was visualized using *Cytoscape* (Shannon et al., 2003).

4.8. Residue conservation

Residue conservation was calculated using the *Consurf* server (Ashkenazy et al., 2016; Yariv et al., 2023). MSA for all organisms was calculated from 148 sequences and the MSA for plant sequences was conducted on 38 sequences with %ID to the sequence of *Mt*HISN3-Nt42 between 24 and 97% for all organisms, and 82–100% for plants.

4.9. MD simulations

The MD simulations were performed in *Desmond 7.6.135* (Bowers et al., 2006) using the *Maestro* 2023-4 package (Schrodinger). The missing Asp76 to Gly79 fragment was built in the residual electron density maps in *Coot* before importing the PDB file into *Maestro*. The structures were prepared using default settings, and the ions and water molecules were retained for the analysis. The MD simulation system was set up under the OPLS4 (Lu et al., 2021) force field using TIP4P solvent model, and the charges were neutralized by adding Na^+ ions. The system was further supplemented with the equivalent of 150 mM NaCl. The system was relaxed using the default protocol before the final simulation runs. The production MD simulations were carried for 2000 ns in 300 K under the NPT ensemble class. The recording interval was set to 100 ps, resulting in 20,000 frames. The simulations were run in duplicate, with random numbers for velocities. The trajectories and molecular interactions were analyzed using the *Simulation Interaction Diagram* tool, embedded in *Maestro*.

4.10. Virtual screening

The docking campaign was performed in *AutoDock Vina* (Trott and Olson, 2010), using custom Python scripts to automate and parallelize the work (exhaustiveness = 8). The receptor file was prepared with *UCSF Chimera DockPrep* (Pettersen et al., 2004) using the 1.54 Å structure (mutated back to Asp57). The search box of dimensions of 22 × 25 × 29 Å was centered at 11, 26, −5 Å (x, y, z). As the *Mt*HISN3 active site is highly polar, the subset from the lead-like library in the ZINC15 database (Sterling and Irwin, 2015), limited to compounds with logP ≤2 (1,355,457 molecules, downloaded in March 2022), was used. All the results were scored based on the calculated binding energy gain.

4.11. Other software

Molecular tunnels were analyzed using *CAVER 3.0.3* (Chovancova et al., 2012). The structure at 1.54 Å served as an input and the following settings were used: minimum probe radius = 0.9, shell depth = 4, shell radius = 3, clustering threshold = 3.5, number of approximating balls = 12, maximum distance = 3, desired radius = 5, starting point coordinates: x = 15.121, y = 25.872, z = −1.007. The interactions between the proteins and ligands were visualized using *Discovery Studio 2024, Biovia Dassault Systèmes*). Protein models were placed inside unit cells using *ACHESYM* (Kowiel et al., 2014). Protein models were visualized in *Chimera 1.15* (Pettersen et al., 2004) and *ChimeraX 1.6.1* (Pettersen et al., 2021). The electrostatic potential of the surface was calculated using *APBS-PDB2PQR* (Jurrus et al., 2018). The electrostatic potential of the surface is expressed as $kT/e = k_B Te_c^{-1} = 25.6$ mV at 298 K, where k_B is Boltzmann's constant, T is temperature at 298K, and e_c is the elementary charge (APBS, 2024). Schematic views of the molecules obtained from the VS campaign were downloaded from PubChem (Kim et al., 2023). Polder maps were made using *Phenix.polder* (Liebschner et al., 2017).

Author contributions

WW produced the proteins, measured enzymatic activities, prepared ProFAR, participated in crystallization, solved and refined the crystal structures, drafted the manuscript (both the first and the revised versions). BI optimized protein production and obtained the best diffracting crystals. MR obtained funding, conceptualized and supervised all the work, and edited the manuscript (the first and revised versions).

Databases

The crystal structures have been deposited in the Protein Data Bank (PDB; www.rcsb.org). The complex of MtHISN3-Nt42-D57N_ProFAR is accessible under the PDB ID: 9FCF, and the complex of MtHISN3-Nt42-D57N_PrFAR is accessible under the PDB ID: 9FCG.

The crystallographic datasets for the structures presented in this study are accessible the Macromolecular Xtallography Raw Data Repository (MX-RDR; www.mxrdr.icm.edu.pl): MtHISN3-Nt42-D47N_ProFAR complex, https://doi.org/10.18150/DGZKW3; MtHISN3-Nt42-D47N_PrFAR complex, https://doi.org/10.18150/LDLSBT.

Funding

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.plaphy.2024.109065.

References

Afonine, P.V., Grosse-Kunstleve, R.W., Echols, N., Headd, J.J., Moriarty, N.W., Mustyakimov, M., Terwilliger, T.C., Urzhumtsev, A., Zwart, P.H., Adams, P.D., 2012. Towards automated crystallographic structure refinement with phenix.refine. Acta Crystallogr D Biol Crystallogr 68, 352–367. https://doi.org/10.1107/S0907444912001308.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J.H., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389–3402. https://doi.org/10.1093/nar/25.17.3389.

Ames, B.N., 1957a. The biosynthesis of histidine - D-erythro-imidazoleglycerol phosphate dehydrase. J. Biol. Chem. 228, 131–143.

Ames, B.N., 1957b. The biosynthesis of histidine - L-histidinol phosphate phosphatase. J. Biol. Chem. 226, 583–593.

Ames, B.N., 1961. First step of histidine biosynthesis. J. Biol. Chem. 236, 2019.

Ames, B.N., Garry, B., 1959. Coordinate repression of the synthesis of four histidine biosynthetic enzymes by histidine. Proc Natl Acad Sci U S A 45, 1453–1461. https://doi.org/10.1073/pnas.45.10.1453.

Ames, B.N., Martin, R.G., Garry, B.J., 1961. The first step of histidine biosynthesis. J. Biol. Chem. 236, 2019–2026. Apbs (2024). Available: https://apbs.readthedocs.io/en/latest/using/units.html.

Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T., Ben-Tal, N., 2016. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. Nucleic Acids Res. 44, W344–W350. https://doi.org/10.1093/nar/gkw408.

Barona-Gómez, F., Hodgson, D.A., 2003. Occurrence of a putative ancient-like isomerase involved in histidine and tryptophan biosynthesis. EMBO Rep. 4, 296–300. https://doi.org/10.1038/sj.embor.embor771.

Chang, A., Jeske, L., Ulbrich, S., Hofmann, J., Koblitz, J., Schomburg, I., Neumann-Schaal, M., Jahn, D., Schomburg, D., 2021. BRENDA, the ELIXIR core data resource in 2021: new developments and updates. Nucleic Acids Res. 49, D498–D508. https://doi.org/10.1093/nar/gkaa1025.

Chovancova, E., Pavelka, A., Benes, P., Strnad, O., Brezovsky, J., Kozlikova, B., Gora, A., Sustr, V., Klvana, M., Medek, P., Biedermannova, L., Sochor, J., Damborsky, J., 2012. Caver 3.0: a tool for the analysis of transport pathways in dynamic protein structures. PLoS Comput. Biol. 8, e1002708 https://doi.org/10.1371/journal.pcbi.1002708.

Davisson, V.J., Deras, I.L., Hamilton, S.E., Moore, L.L., 1994. A plasmid-based approach for the synthesis of a histidine biosynthetic intermediate. J. Org. Chem. 59, 137–143. https://doi.org/10.1021/jo00080a022.

Due, A.V., Kuper, J., Geerlof, A., Von Kries, J.P., Wilmanns, M., 2011. Bisubstrate specificity in histidine/tryptophan biosynthesis isomerase from by active site metamorphosis. Proc Natl Acad Sci U S A. 108, 3554–3559. https://doi.org/10.1073/pnas.1015996108.

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797. https://doi.org/10.1093/nar/gkh340.

Emanuelsson, O., Nielsen, H., Brunak, S., Von Heijne, G., 2000. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. J. Mol. Biol. 300, 1005–1016. https://doi.org/10.1006/jmbi.2000.3903.

Emsley, P., Lohkamp, B., Scott, W.G., Cowtan, K., 2010. Features and development of Coot. Acta Crystallogr. D 66, 486–501. https://doi.org/10.1107/S0907444910007493.

Fani, R., Lio, P., Chiarelli, I., Bazzicalupo, M., 1994. The evolution of the histidine biosynthetic genes in prokaryotes - a common ancestor for the hisa and hisf genes. J. Mol. Evol. 38, 489–495. https://doi.org/10.1007/Bf00178849.

Fazio, V.J., Peat, T.S., Newman, J., 2014. A drunken search in crystallization space. Acta Crystallographica Section F-Structural Biology Communications 70, 1303–1311. https://doi.org/10.1107/S2053230x1401841x.

Gerlt, J.A., Babbitt, P.C., 2001. Divergent evolution of enzymatic function: mechanistically diverse superfamilies and functionally distinct suprafamilies. Annu. Rev. Biochem. 70, 209–246. https://doi.org/10.1146/annurev.biochem.70.1.209.

Haas, F., Mitchell, M.B., Ames, B.N., Mitchell, H.K., 1952. A series of histidineless mutants of neurospora crassa. Genetics 37, 217–226 doi:).

Heap, I., 2024. The international survey of herbicide resistant weeds [Online]. www.weedscience.com.

Henn-Sax, M., Thoma, R., Schmidt, S., Hennig, M., Kirschner, K., Sterner, R., 2002. Two (βα)₈-barrel enzymes of histidine and tryptophan biosynthesis have similar reaction mechanisms and common strategies for protecting their labile substrates. Biochemistry 41, 12032–12042. https://doi.org/10.1021/bi026092h.

Hilton, J.L., Kearney, P.C., Ames, B.N., 1965. Mode of action of herbicide 3-amino-1,2,4-triazole(amitrole) - inhibition of an enzyme of histidine biosynthesis. Arch. Biochem. Biophys. 112, 544–547. https://doi.org/10.1016/0003-9861(65)90093-7.

Höcker, B., Jürgens, C., Wilmanns, M., Sterner, R., 2001. Stability, catalytic versatility and evolution of the (βα)₈-barrel fold. Curr. Opin. Biotechnol. 12, 376–381. https://doi.org/10.1016/S0958-1669(00)00230-5.

Ingle, R.A., 2011. Histidine biosynthesis. In: Arabidopsis Book, vol. 9, e0141. https://doi.org/10.1199/tab.0141.

Jensen, R.A., 1976. Enzyme recruitment in evolution of new function. Annu. Rev. Microbiol. 30, 409–425. https://doi.org/10.1146/annurev.mi.30.100176.002205.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S.a.A., Ballard, A.J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A.W., Kavukcuoglu, K., Kohli, P., Hassabis, D., 2021. Highly accurate protein structure prediction with AlphaFold. Nature 596, 583–589. https://doi.org/10.1038/s41586-021-03819-2.

Jürgens, C., Strom, A., Wegener, D., Hettwer, S., Wilmanns, M., Sterner, R., 2000. Directed evolution of a (βα)₈-barrel enzyme to catalyze related reactions in two different metabolic pathways. Proc Natl Acad Sci U S A 97, 9925–9930. https://doi.org/10.1073/pnas.160255397.

Jurrus, E., Engel, D., Star, K., Monson, K., Brandi, J., Felberg, L.E., Brookes, D.H., Wilson, L., Chen, J.H., Liles, K., Chun, M.J., Li, P., Gohara, D.W., Dolinsky, T., Konecny, R., Koes, D.R., Nielsen, J.E., Head-Gordon, T., Geng, W.H., Krasny, R., Wei, G.W., Holst, M.J., Mccammon, J.A., Baker, N.A., 2018. Improvements to the APBS biomolecular solvation software suite. Protein Sci. 27, 112–128. https://doi.org/10.1002/pro.3280.

Kabsch, W., 2010. Xds. Acta Crystallographica Section D-Biological Crystallography 66, 125–132. https://doi.org/10.1107/S0907444909047337.

Kim, S., Chen, J., Cheng, T.J., Gindulyte, A., He, J., He, S.Q., Li, Q.L., Shoemaker, B.A., Thiessen, P.A., Yu, B., Zaslavsky, L., Zhang, J., Bolton, E.E., 2023. PubChem 2023 update. Nucleic Acids Res. 51, D1373–D1380. https://doi.org/10.1093/nar/gkac956.

Kim, Y., Babnigg, G., Jedrzejczak, R., Eschenfeldt, W.H., Li, H., Maltseva, N., Hatzos-Skintges, C., Gu, M., Makowska-Grzyska, M., Wu, R., An, H., Chhor, G., Joachimiak, A., 2011. High-throughput protein purification and quality assessment for crystallization. Methods 55, 12–28. https://doi.org/10.1016/j.ymeth.2011.07.010.

Kishore, G.M., Shah, D.M., 1988. Amino acid biosynthesis inhibitors as herbicides. Annu. Rev. Biochem. 57, 627–663. https://doi.org/10.1146/annurev.bi.57.070188.003211.

Klem, T.J., Davisson, V.J., 1993. Imidazole glycerol phosphate synthase - the glutamine amidotransferase in histidine biosynthesis. Biochemistry 32, 5177–5186. https://doi.org/10.1021/bi00070a029.

Klock, H.E., Lesley, S.A., 2009. The Polymerase Incomplete Primer Extension (PIPE) method applied to high-throughput cloning and site-directed mutagenesis. Methods Mol. Biol. 498, 91–103. https://doi.org/10.1007/978-1-59745-196-3_6.

Klopotowski, T.W.A., 1965. Synergism of aminotriazole and phosphate on inhibition of yeast imidazole glycerol phosphate dehydratase. Arch. Biochem. Biophys. 112, 562–566. https://doi.org/10.1016/0003-9861(65)90096-2.

Koo, D., Sathishraj, R., Friebe, B., Gill, B.S., 2022. Deciphering the mechanism of glyphosate resistance in Amaranthus palmeri by cytogenomics. Cytogenet. Genome Res. https://doi.org/10.1159/000521409.

Kowiel, M., Jaskolski, M., Dauter, Z., 2014. ACHESYM: an algorithm and server for standardized placement of macromolecular models in the unit cell. Acta Crystallographica Section D-Biological Crystallography 70, 3290–3298. https://doi.org/10.1107/S1399004714024572.

Kuper, J., Doenges, C., Wilmanns, M., 2005. Two-fold repeated (βα)₄ half-barrels may provide a molecular tool for dual substrate specificity. EMBO Rep. 6, 134–139. https://doi.org/10.1038/sj.embor.7400330.

Lang, D., Thoma, R., Henn-Sax, M., Sterner, R., Wilmanns, M., 2000. Structural evidence for evolution of the β/α barrel scaffold by gene duplication and fusion. Science 289, 1546–1550. https://doi.org/10.1126/science.289.5484.1546.

Liebschner, D., Afonine, P.V., Moriarty, N.W., Poon, B.K., Sobolev, O.V., Terwilliger, T.C., Adams, P.D., 2017. Polder maps: improving OMIT maps by excluding bulk solvent. Acta Crystallographica Section D-Structural Biology 73, 148–157. https://doi.org/10.1107/S2059798316018210.

Lu, C., Wu, C.J., Ghoreishi, D., Chen, W., Wang, L.L., Damm, W., Ross, G.A., Dahlgren, M.K., Russell, E., Von Bargen, C.D., Abel, R., Friesner, R.A., Harder, E.D., 2021. OPLS4: improving force field accuracy on challenging regimes of chemical space. J. Chem. Theor. Comput. 17, 4291–4300. https://doi.org/10.1021/acs.jctc.1c00302.

Lundin, E., Näsvall, J., Andersson, D.I., 2020. Mutational pathways and trade-offs between HisA and TrpF functions: implications for evolution via gene duplication and divergence. Front. Microbiol. 11, 588235 https://doi.org/10.3389/fmicb.2020.588235.

Mccoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., Read, R.J., 2007. Phaser crystallographic software. J. Appl. Crystallogr. 40, 658–674. https://doi.org/10.1107/S0021889807021206.

Miflin, B., 1980. Histidine biosynthesis. In: The Biochemistry of Plants: a Comprehensive Treatise, Vol 5: Amino Acids and Derivatives. Academic Press, New York, 533–541.

Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., Steinegger, M., 2022. ColabFold: making protein folding accessible to all. Nat. Methods 19, 679. https://doi.org/10.1038/s41592-022-01488-1.

Mo, X.R., Zhu, Q.Y., Li, X., Li, J., Zeng, Q.N., Rong, H.L., Zhang, H.M., Wu, P., 2006. The hpa1 mutant of Arabidopsis reveals a crucial role of histidine homeostasis in root meristem maintenance. Plant Physiol. 141, 1425–1435. https://doi.org/10.1104/pp.106.084178.

Moriarty, N.W., Grosse-Kunstleve, R.W., Adams, P.D., 2009. Electronic Ligand Builder and Optimization Workbench (eLBOW): a tool for ligand coordinate and restraint generation. Acta Crystallographica Section D-Structural Biology 65, 1074–1080. https://doi.org/10.1107/S0907444909029436.

Newton, M.S., Guo, X.H., Söderholm, A., Näsvall, J., Lundström, P., Andersson, D.I., Selmer, M., Patrick, W.M., 2017. Structural and functional innovations in the real-time evolution of new (βα)₈ barrel enzymes. Proc Natl Acad Sci U S A 114, 4727–4732. https://doi.org/10.1073/pnas.1618552114.

Noutoshi, Y.I., Shinozaki, T., K, 2005. ALBINO AND PALE GREEN 10 encodes BBMII isomerase involved in histidine biosynthesis in Arabidopsis thaliana. Plant Cell Physiol. 46, 1165–1172. https://doi.org/10.1093/pcp/pci119.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., Ferrin, T.E., 2004. UCSF chimera - a visualization system for exploratory research and analysis. J. Comput. Chem. 25, 1605–1612. https://doi.org/10.1002/jcc.20084.

Pettersen, E.F., Goddard, T.D., Huang, C.R.C., Meng, E.E.C., Couch, G.S., Croll, T.I., Morris, J.H., Ferrin, T.E., 2021. UCSF ChimeraX: structure visualization for researchers, educators, and developers. Protein Sci. 30, 70–82. https://doi.org/10.1002/pro.3943.

Rawson, S., Bisson, C., Hurdiss, D.L., Fazal, A., Mcphillie, M.J., Sedelnikova, S.E., Baker, P.J., Rice, D.W., Muench, S.P., 2018. Elucidating the structural basis for differing enzyme inhibitor potency by cryo-EM. Proc Natl Acad Sci U S A. 115, 1795–1800. https://doi.org/10.1073/pnas.1708839115.

Romero-Rivera, A., Corbella, M., Parracino, A., Patrick, W.M., Kamerlin, S.C.L., 2022. Complex loop dynamics underpin activity, specificity, and evolvability in the (betaalpha)(8) barrel enzymes of histidine and tryptophan biosynthesis. JACS Au 2, 943–960. https://doi.org/10.1021/jacsau.2c00063.

Ruszkowski, M., 2018. Guarding the gateway to histidine biosynthesis in plants: Medicago truncatula ATP-phosphoribosyltransferase in relaxed and tense states. Biochem. J. 475, 2681–2697. https://doi.org/10.1042/BCJ20180289.

Ruszkowski, M., Dauter, Z., 2016. Structural studies of histidinol phosphate phosphatase from inositol monophosphatase superfamily reveal details of penultimate step of histidine biosynthesis in plants. J. Biol. Chem. 291, 9960–9973. https://doi.org/10.1074/jbc.M115.708727.

Ruszkowski, M., Dauter, Z., 2017. Structures of Medicago truncatula L-histidinol dehydrogenase show rearrangements required for NAD(+) binding and the cofactor positioned to accept a hydride. Sci. Rep. 7, 10476 https://doi.org/10.1038/s41598-017-10859-0.

Ruszkowski, M., Forlani, G., 2022. Deciphering the structure of Arabidopsis thaliana 5-enol-pyruvyl-shikimate-3-phosphate synthase: an essential step toward the discovery of novel inhibitors to supersede glyphosate. Comput. Struct. Biotechnol. J. 20, 1494–1505. https://doi.org/10.1016/j.csbj.2022.03.020.

Rutkiewicz, M., Nogues, I., Witek, W., Angelaccio, S., Contestabile, R., Ruszkowski, M., 2023. Insights into the substrate specificity, structure, and dynamics of plant histidinol-phosphate aminotransferase (HISN6). Plant Physiol. Biochem. 196, 759–773. https://doi.org/10.1016/j.plaphy.2023.02.017.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., Ideker, T., 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 13, 2498–2504. https://doi.org/10.1101/gr.1239303.

Sigel, A., Sigel, H., Sigel, R.K.O., 2016. The alkali metal ions: their role for life. In: Metal Ions in Life Sciences, first ed. Springer International Publishing : Imprint: Springer, Cham.

Söderholm, A., Guo, X.H., Newton, M.S., Evans, G.B., Näsvall, J., Patrick, W.M., Selmer, M., 2015. Two-step ligand binding in a (βα)₈ barrel enzyme: substrate-bound structures shed new light on the catalytic cycle of HisA. J. Biol. Chem. 290, 24657–24668. https://doi.org/10.1074/jbc.M115.678086.

Sterling, T., Irwin, J.J., 2015. ZINC 15-ligand discovery for everyone. J. Chem. Inf. Model. 55, 2324–2337. https://doi.org/10.1021/acs.jcim.5b00559.

Tamura, K., Stecher, G., Kumar, S., 2021. MEGA11 molecular evolutionary genetics analysis version 11. Mol. Biol. Evol. 38, 3022–3027. https://doi.org/10.1093/molbev/msab120.

Trott, O., Olson, A.J., 2010. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J. Comput. Chem. 31, 455–461. https://doi.org/10.1002/jcc.21334.

Uniprot, C., 2019. UniProt: a worldwide hub of protein knowledge. Nucleic Acids Res. 47, D506–D515. https://doi.org/10.1093/nar/gky1049.

Williams, C.J., Headd, J.J., Moriarty, N.W., Prisant, M.G., Videau, L.L., Deis, L.N., Verma, V., Keedy, D.A., Hintze, B.J., Chen, V.B., Jain, S., Lewis, S.M., Arendall, W.B., Snoeyink, J., Adams, P.D., Lovell, S.C., Richardson, J.S., Richardson, D.C., 2018. MolProbity: more and better reference data for improved all-atom structure validation. Protein Sci. 27, 293–315. https://doi.org/10.1002/pro.3330.

Witek, W., Sliwiak, J., Rawski, M., Ruszkowski, M., 2024. Targeting imidazole-glycerol phosphate dehydratase in plants: novel approach for structural and functional studies, and inhibitor blueprinting. Front. Plant Sci. 15, 1343980 https://doi.org/10.3389/fpls.2024.1343980.

Witek, W., Sliwiak, J., Ruszkowski, M., 2021. Structural and mechanistic insights into the bifunctional HISN2 enzyme catalyzing the second and third steps of histidine

biosynthesis in plants. Sci. Rep. 11, 9647. https://doi.org/10.1038/s41598-021-88920-2.

Yariv, B., Yariv, E., Kessel, A., Masrati, G., Chorin, A.B., Martz, E., Mayrose, I., Pupko, T., Ben-Tal, N., 2023. Using evolutionary data to make sense of macromolecules with a "face-lifted" ConSurf. Protein Sci. 32, e4582 https://doi.org/10.1002/pro.4582.

Zallot, R., Oberg, N., Gerlt, J.A., 2019. The EFI web resource for genomic enzymology tools: leveraging protein, genome, and metagenome databases to discover novel enzymes and metabolic pathways. Biochemistry 58, 4169–4182. https://doi.org/10.1021/acs.biochem.9b00735.

Zheng, H., Cooper, D.R., Porebski, P.J., Shabalin, I.G., Handing, K.B., Minor, W., 2017. CheckMyMetal: a macromolecular metal-binding validation tool. Acta Crystallogr. D 73, 223–233. https://doi.org/10.1107/S2059798317001061.

Bowers, K.J., Chow, E., Xu, H., Dror, R.O., Eastwood, M.P., Gregersen, B.A., Klepeis, J.L., Kolossvary, I., Moraes, M.A., Sacerdoti, F.D., Salmon, J.K., Shan, Y., Shaw, D.E., 2006. Scalable algorithms for molecular dynamics simulations on commodity clusters. In: *Proceedings of the ACM/IEEE Conference on Supercomputing (SC06)*.

Supplementary Material

Structural, kinetic, and evolutionary peculiarities of HISN3, a plant 5'-ProFAR isomerase

Wojciech Witek[1], Barbara Imiolczyk[1], Milosz Ruszkowski[1*]

[1]Department of Structural Biology of Eukaryotes, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznan, Poland

* Corresponding author:
Milosz Ruszkowski
Department of Structural Biology of Eukaryotes
Institute of Bioorganic Chemistry,
Polish Academy of Sciences, Poznan, Poland
mruszkowski@ibch.poznan.pl

Supplementary Figures

Figure. S1. Interactions between *Mt*HISN3 and sodium cations, chloride ions and MPD molecules**.** (A,B) Binding of sodium cations and chloride ions in the complex with ProFAR (Panel A) and PrFAR (Panel B). Sodium Na1 in both structures is located in the active site. Distance proportions are distorted due to perspective. Water coordinating Na1 is in the back plane, behind the sodium cation. Panel C shows a non-specific MPD binding in surface regions of different electrostatic potential (colored regarding the scale bar). Polder maps were contoured at σ = 3.5.
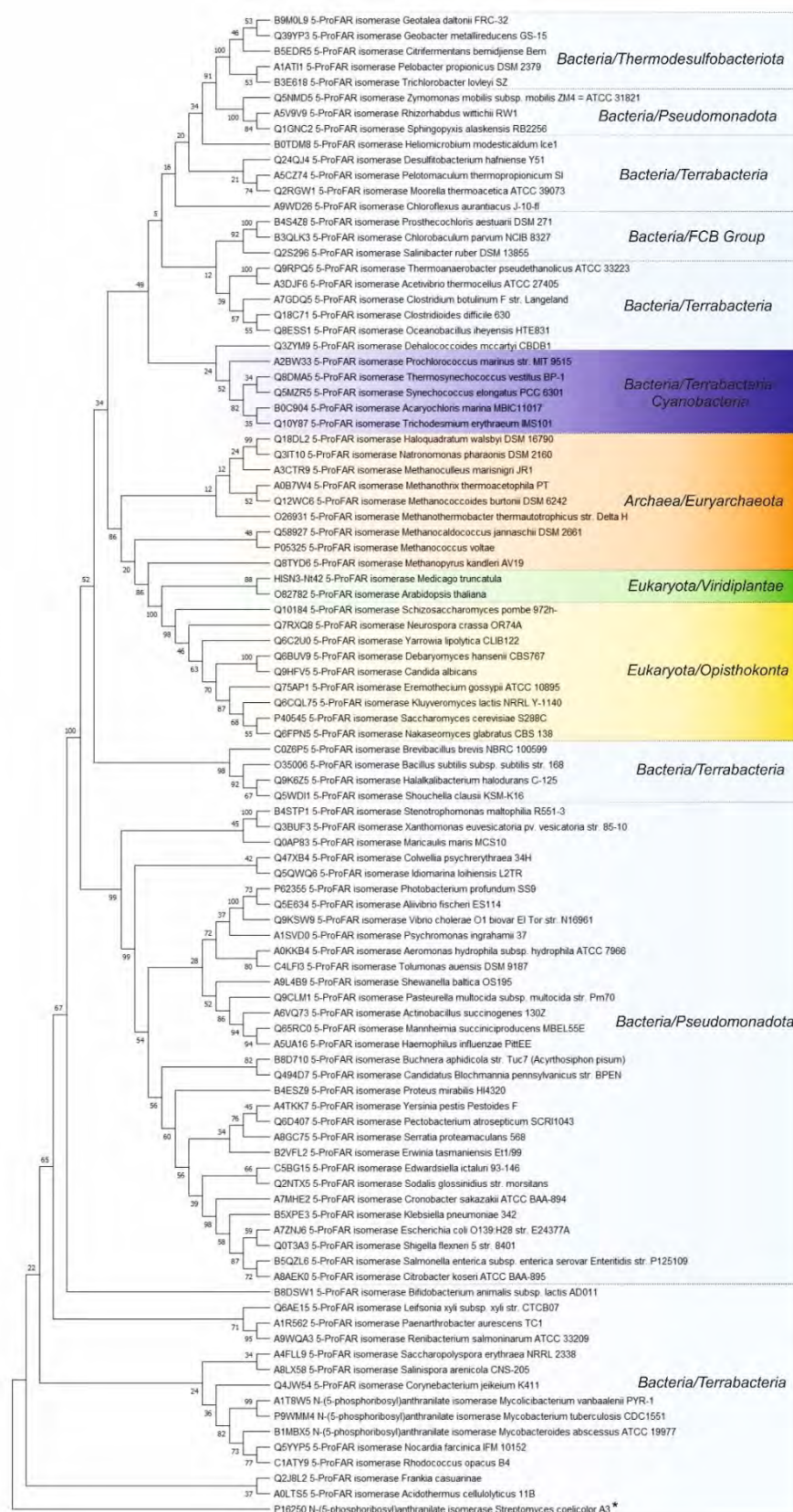
Figure S2. Bootstrap consensus phylogenetic tree of *Mt*HISN3 homologs. Archaeal and eukaryotic sequences are highlighted in orange and green, respectively. *Streptomyces coelicolor* PriA sequence was used as an outgroup root. Values on the branches indicate bootstrap percentage after 250 iterations in constructing the tree.
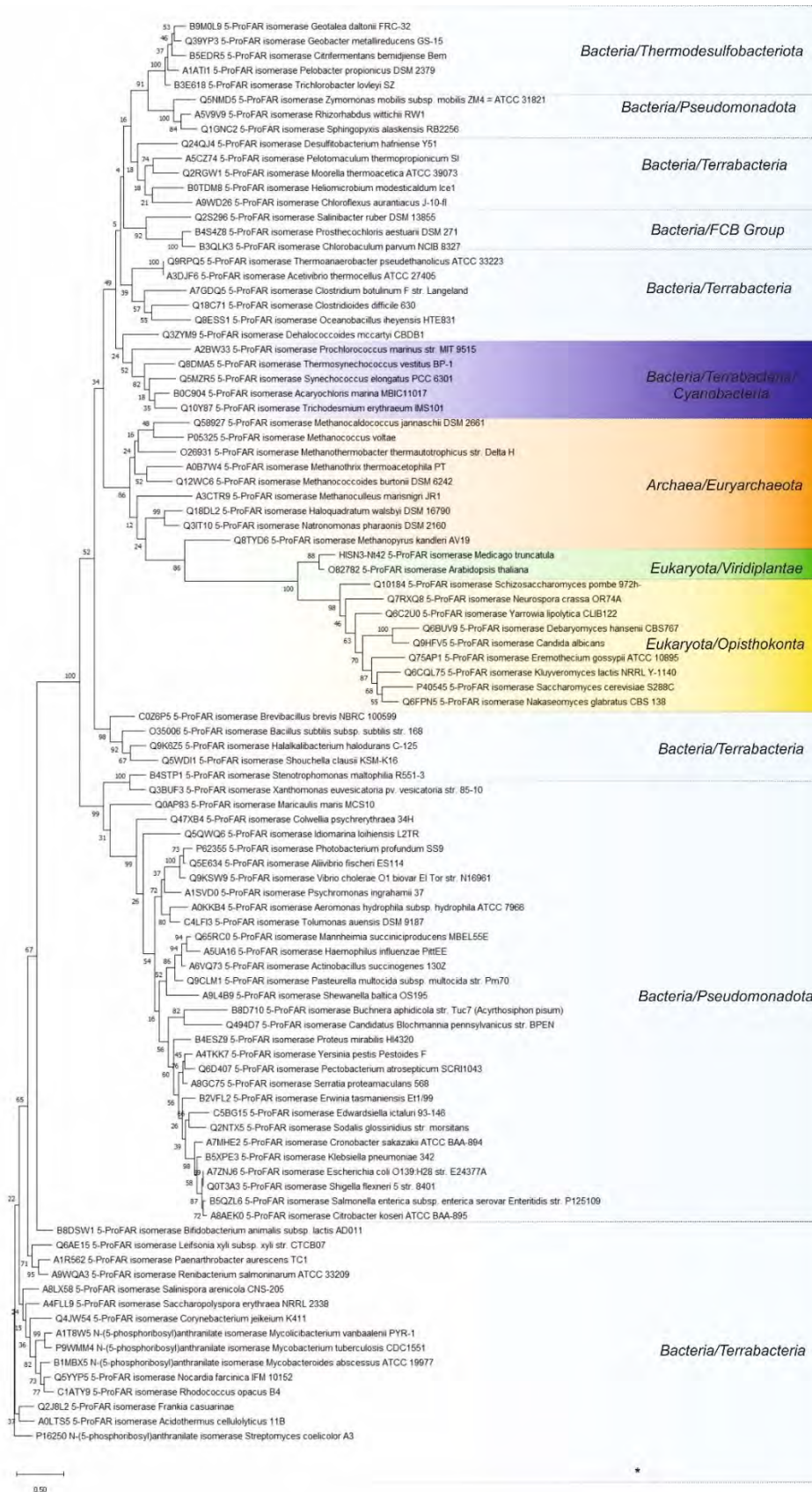
Figure S3. Original phylogenetic tree of *Mt*HISN3 homologs. Archaeal and eukaryotic sequences are highlighted in orange and green, respectively. *Streptomyces coelicolor* PriA sequence was used as an outgroup root. Scale bar refers to a phylogenetic distance of 0.5 substitutions per site. Values on the branches indicate bootstrap percentage after 250 iterations in constructing the tree.
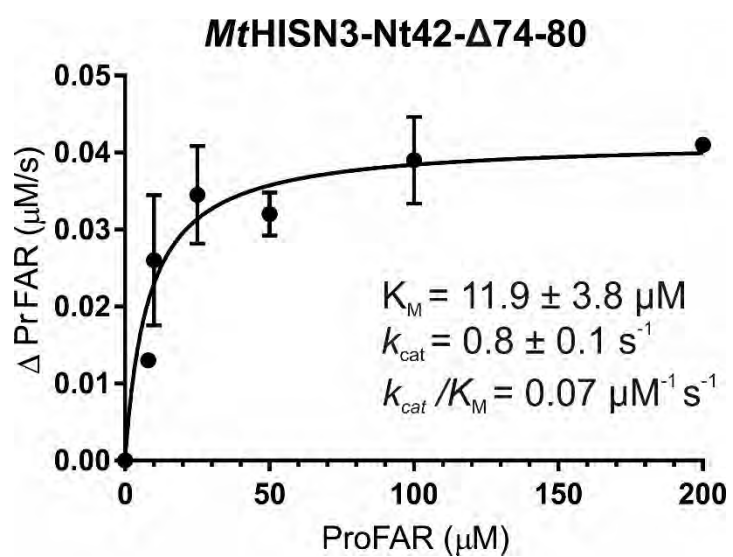
Figure S4. Michaelis-Menten curve for *Mt*HISN3-Nt42-Δ74-80. ProFAR concentrations are on the horizontal axis, whereas initial velocities are plotted on the vertical axis. The error bars correspond to the data fitting to the curve.
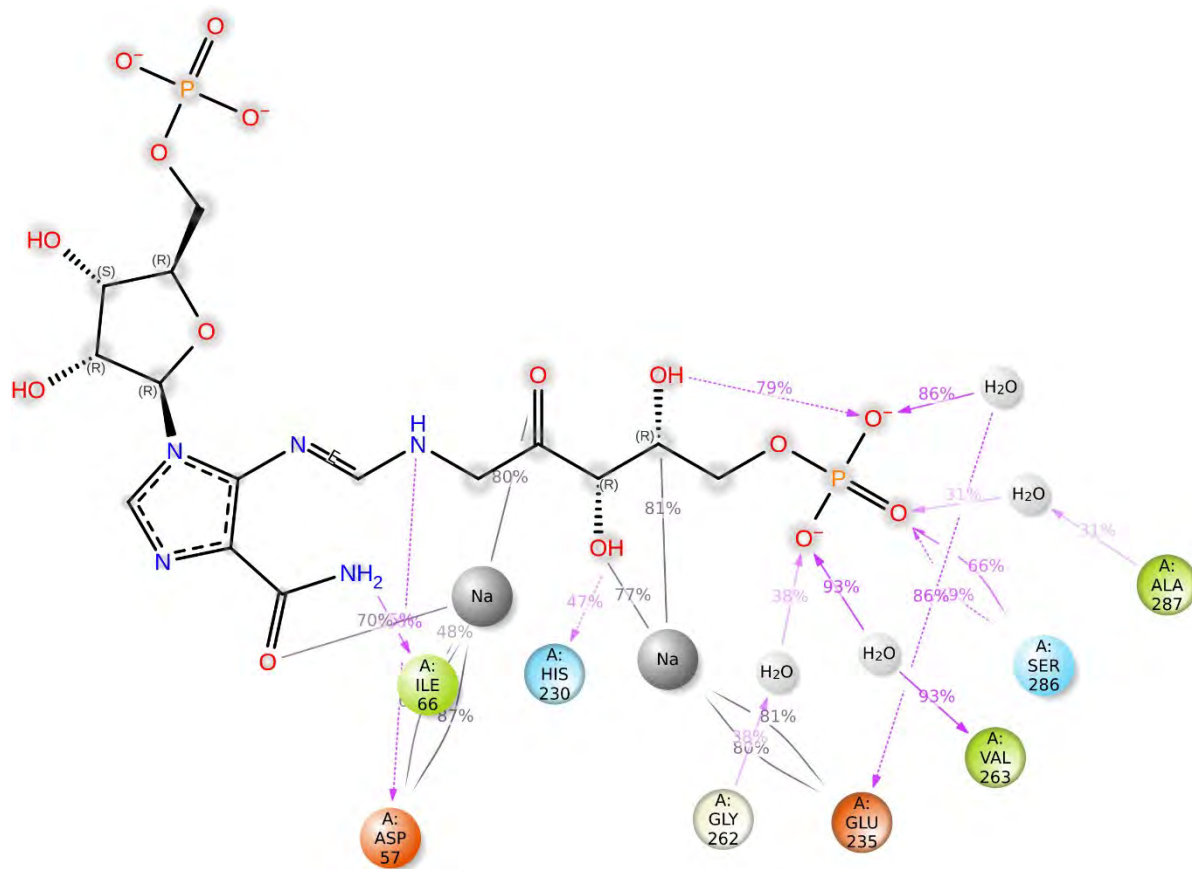
Figure S5. A scheme of ligand-protein interactions in the MD scenario where PrFAR remains in the active site. Interactions that occur more than 30.0% of the simulation time in the trajectory (0 through 2000 ns), are shown; the N1-phosphoribosyl interactions do not satisfy this threshold.
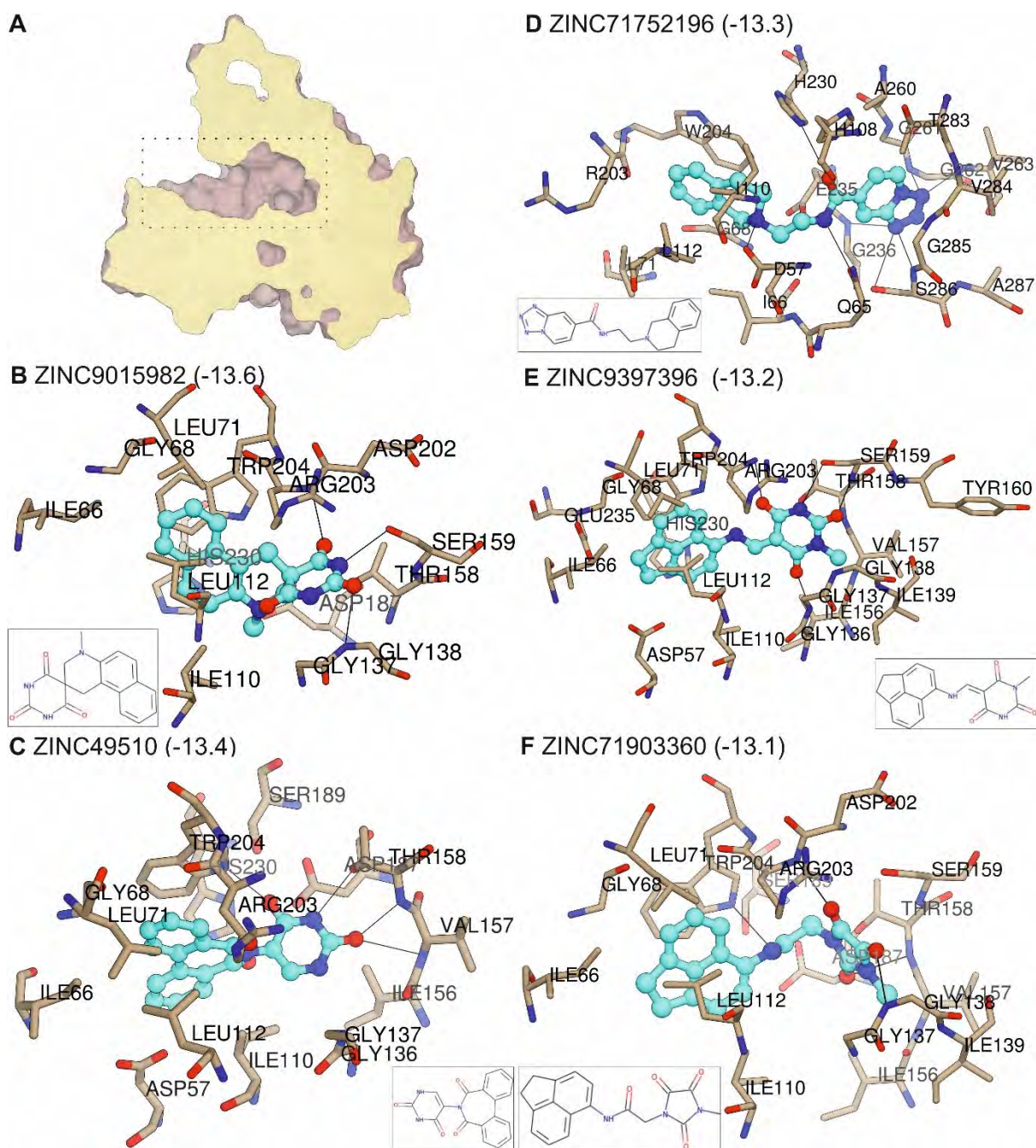
Figure S6. Hydrogen bonding patterns between *Mt*HISN3-Wt residues with top five hits obtained in the VS campaign. Panel A shows a cross-section through the protein. The active site area is marked with a dashed rectangle. B-F are the VS campaign hits, ordered by the energy gain in kcal/mol. Residue labels from the front are marked in black, and residues in the back plane are in grey. Hydrogen bonds are marked with black lines.

Targeting imidazole-glycerol phosphate dehydratase in plants: novel approach for structural and functional studies, and inhibitor blueprinting

Wojciech Witek[1], Joanna Sliwiak[1], Michal Rawski[2] and Milosz Ruszkowski[1*]

[1]Department of Structural Biology of Eukaryotes, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznan, Poland, [2]Cryo-EM Facility, SOLARIS National Synchrotron Radiation Centre, Krakow, Poland

The histidine biosynthetic pathway (HBP) is targeted for herbicide design with preliminary success only regarding imidazole-glycerol phosphate dehydratase (IGPD, EC 4.2.1.19), or HISN5, as referred to in plants. HISN5 catalyzes the sixth step of the HBP, in which imidazole-glycerol phosphate (IGP) is dehydrated to imidazole-acetol phosphate. In this work, we present high-resolution cryoEM and crystal structures of *Medicago truncatula* HISN5 (*Mt*HISN5) in complexes with an inactive IGP diastereoisomer and with various other ligands. *Mt*HISN5 can serve as a new model for plant HISN5 structural studies, as it enables resolving protein-ligand interactions at high (2.2 Å) resolution using cryoEM. We identified ligand-binding hotspots and characterized the features of plant HISN5 enzymes in the context of the HISN5-targeted inhibitor design. Virtual screening performed against millions of small molecules not only revealed candidate molecules but also identified linkers for fragments that were experimentally confirmed to bind. Based on experimental and computational approaches, this study provides guidelines for designing symmetric HISN5 inhibitors that can reach two neighboring active sites. Finally, we conducted analyses of sequence similarity networks revealing that plant HISN5 enzymes derive from cyanobacteria. We also adopted a new approach to measure *Mt*HISN5 enzymatic activity using isothermal titration calorimetry and enzymatically synthesized IGP.

KEYWORDS

IGPD, HISN5, herbicide design, enzyme kinetics, histidine biosynthesis

Introduction

Since the 1960s, more than 250 weed species have become resistant to over 150 herbicides, mostly because of repeated use (Gould et al., 2018; Beckie et al., 2021; Gaines et al., 2021). Current herbicides also raise safety concerns and have negative impacts on the environment. Recent reports have delivered evidence that the most common herbicide, glyphosate, is harmful to honeybee broods, impairing their sensory and cognitive abilities and gut microbiome (Motta et al., 2018; Farina et al., 2019; Vázquez et al., 2020). Other herbicides, such as triazines (atrazine, hexazinone), anilides (acetochlor, alachlor), and carbamates, enter aquatic environments and accumulate in coral reefs (Tyohemba et al., 2022). They cause acute toxicity, leading to reduced zooxanthellar photosynthetic efficiency (Negri et al., 2005), resulting in bleaching, reduced reproductive output, and partial or full-colony mortality (Cantin et al., 2007). These factors incite the development of new herbicides to ensure that eight billion people on the planet can be fed sustainably. In this view, the histidine biosynthetic pathway (HBP) has become a promising new target for the development of herbicides (Hall et al., 2020).

The HBP occurs in bacteria, archaea, plants, and other lower eukaryotes, such as yeast or protozoans, but is absent in animals. It has been intensively studied since the 1950s (Miller and Bale, 1952; Adams, 1954; Ames and Mitchell, 1955; Ames, 1957a, Ames, 1957b), mostly in *Escherichia coli* and *Salmonella typhimurium*. The research was later continued by Ames, Brenner, and Martin, who identified all enzymes, metabolic intermediates, and by-products (Brenner and Ames, 1971; Martin et al., 1971). Studies of the HBP in plants started later due to the lack of auxotrophic mutants and the complicated biochemistry behind the pathway (Wiater et al., 1971c; Ingle, 2011). In fact, the plant HBP was genetically deciphered in 2010, as the last amino acid biosynthetic pathway (Petersen et al., 2010).

The overall organization of the HBP is conserved across kingdoms, but there are significant differences between homologous enzymes. These differences were caused by genetic events during evolution, such as gene duplications, elongations, horizontal gene transfers (HGT), and gene fusions resulting in the emergence of bi- or even trifunctional enzymes (Brilli and Fani, 2004; Stepansky and Leustek, 2006; Reyes-Prieto and Moustafa, 2012; Del Duca et al., 2020; Rutkiewicz et al., 2023). The HBP consists of ten steps (eleven reactions considering the glutaminase activity of HISN4 auxiliary), catalyzed in plants by eight enzymes that are named HISN1-8 by their action in the HBP sequence. Each of the eight enzymes in plants is encoded by nuclear DNA and contains an N-terminal chloroplast transit peptide (Fujimori and Ohta, 1998).

This work focuses on D-*erythro*-imidazole-glycerol phosphate dehydratase (IGPD, EC 4.2.1.19) or HISN5 as it is referred to in plants. Notably, IGPD-encoding genes are named inconsistently between kingdoms, for example, *HISN5* in plants, *HIS3* in yeast, or *HisB* in bacteria (Muralla et al., 2007). Interestingly, in most species, IGPDs are monofunctional enzymes; however, in some bacterial phyla, they perform two HBP reactions as a result of gene fusion between genes encoding IGPD and histidinol-phosphate phosphatase (HPP, EC 3.1.3.15), which are HISN5 and HISN7 counterparts in plants (Brilli and Fani, 2004). Plant HISN5 catalyzes the sixth step of the HBP, in which imidazole-glycerol phosphate (IGP) is dehydrated to form imidazole-acetol phosphate (IAP, Figure 1). HISN5 activity in plants was first described in 1971 (Wiater et al., 1971c) but the first plant enzyme was purified from wheat germ in 1993 (Mano et al., 1993). HISN5 has been considered for about fifty years as a potential target for triazole compounds, e.g., amitrole, 3-amino-1,2,4-triazole, and 2-hydroxy-3-(1,2,4-triazol-1-yl) (C348) (Hilton et al., 1965; Wiater et al., 1971a, Wiater et al., 1971b; Kishore and Shah, 1988; Rawson et al., 2018). However, amitrole is a non-selective herbicide with off-target effects. As reported by Furukawa et al., amitrole shows carcinogenic activity in rats, mice, and humans (Furukawa et al., 2010). Therefore, there is great demand for more selective HISN5 inhibitors that exhibit fewer side effects.

Arabidopsis thaliana was the first model for structural studies of plant HISN5 and provided the groundwork for structure-based inhibitor design. *At*HISN5 occurs in two isoforms, A and B, which are 270 and 272 amino acid residues long, respectively. Its fold resembles a sandwich constituted of a bundle of four, centrally located, α-helices which are surrounded by two β-sheets from both sides. This metalloenzyme utilizes Mn^{2+} for proper folding and catalysis (Glynn et al., 2005b). The plant HISN5, with its 24-meric structure, could potentially offer a plethora of druggable sites, not only in the active site but also at unique channels, clefts, and inter-subunit interfaces. Such hot-spots often make a major contribution to the protein-ligand binding free energy gain and are therefore an important factor in the discovery of bioactive compounds (Zerbe et al., 2012). The first crystal structure of *At*HISN5, at a 3 Å resolution, was obtained in 2005, providing insights into manganese cations coordination and reaction mechanism (Glynn et al., 2005b). A decade later, a series of *At*HISN5 crystal structures at highly improved resolution (1.1 – 1.5 Å) offered the first details about substrate and inhibitor positioning (Bisson et al., 2015, Bisson et al., 2016). The advent of cryogenic electron microscopy (cryoEM) brought about the first microscopic structure of a complex of *At*HISN5 with a triazole inhibitor, obtained at 3.1 Å (Rawson et al., 2018). At that resolution, it was only possible to confirm the presence of the inhibitor in the EM map without providing
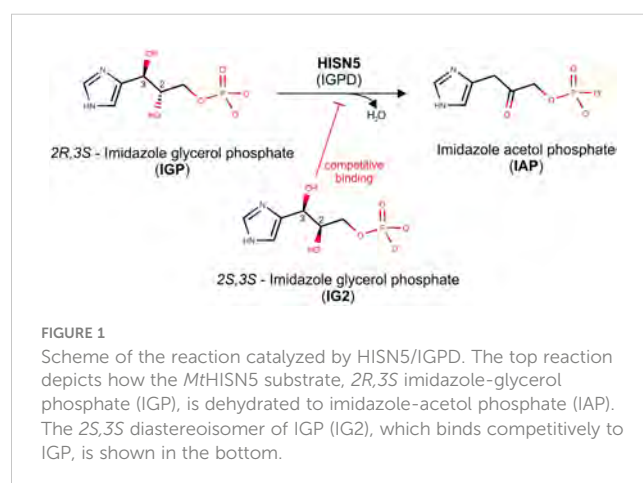
FIGURE 1
Scheme of the reaction catalyzed by HISN5/IGPD. The top reaction depicts how the *Mt*HISN5 substrate, *2R,3S* imidazole-glycerol phosphate (IGP), is dehydrated to imidazole-acetol phosphate (IAP). The *2S,3S* diastereoisomer of IGP (IG2), which binds competitively to IGP, is shown in the bottom.

details about its binding mode. However, rational herbicide design requires high-quality structures of enzymes and an understanding of the interactions that occur at the molecular level. Hence, there is a need for a better model of the plant HISN5 enzyme to study its interactions with small molecules using cryoEM.

The HISN5 enzyme from a model legume, *Medicago truncatula*, was selected for this study for various reasons. Structural studies require a prior preparation of expression constructs, e.g., cleavage of signal peptides, testing different ranges, etc. The availability of *M. truncatula* genomic sequence makes such modifications feasible (Young et al., 2011; Burks et al., 2018). Furthermore, this study is a continuation of work to provide a complete picture of the HBP in legumes. So far, we published structures of HISN1 (Ruszkowski, 2018), HISN2 (Witek et al., 2021), HISN6 (Rutkiewicz et al., 2023), HISN7 (Ruszkowski and Dauter, 2016), and HISN8 (Ruszkowski and Dauter, 2017). The model *M. truncatula* is closely related to *Medicago sativa* (lucerne or alfalfa), an economically and environmentally important forage crop (Mueller-Harvey et al., 2019; Hrbáčková et al., 2020; Sakiroglu and Ilhan, 2021).

Although there are available *At*HISN5 structures, our research was motivated by the lack of high-resolution cryoEM structures of plant HISN5 enzymes, which could allow to study protein interactions with small molecules. Therefore, we established a pipeline for HISN5 cryoEM research that yields maps at a resolution allowing to study interactions with ligands. Our experimental results were combined with computational approach to describe ligand binding hot-spots and potential pharmacophores for future design of novel inhibitors. So far, only a few molecules have been the subject of interest in terms of plant HISN5 inhibition. The similarity of plant, fungal, and bacterial IGPD enzymes in the active site poses a potential threat for off-target (antimicrobial) activities introduced by HISN5-targeted herbicides. To provide a background for reaching selectivity at the kingdom level, we analyzed and compared residue conservation for plants and other organisms to identify a region near the active site which is specific to plants. Finally, we developed a new approach to measure catalytic properties of *Mt*HISN5. This was motivated by two factors, i.e., poor availability of IGP on the market and its contamination with an inactive diastereoisomer (Saika et al., 1993; Bisson et al., 2015), able to bind competitively instead of the *bona fide* substrate.

Results and discussion

CryoEM and crystal structures — an overview of *Mt*HISN5 structural features

This work describes results obtained from five experimental structures of *Mt*HISN5, including three crystal structures at 1.55, 1.69, and 2.2 Å resolutions (Supplementary Figures S1A–C) and two cryoEM structures resolved at 2.4 Å (*Mt*HISN5-unliganded, PDB ID: 7OJ5) and 2.2 Å (complex *Mt*HISN5 with *2S,3S*-IGP (referred to as IG2), PDB ID: 8QAV) (Figure 2). The latter was obtained using the commercially available IGP (cIGP).

The crystal structure at 1.55 Å was obtained in space group *R*3 and comprises eight chains in the asymmetric unit (ASU). The crystal structure at 2.2 Å is isomorphous to the 1.55 Å structure but contains different ligands, e.g., citrate (CIT, Supplementary Figure S1D, see below). We also obtained the structure in space group *I*4 at 1.69 Å resolution with six chains in the ASU. The crystal structures cover the *Mt*HISN5 sequence starting from Gly77 or Ala78 through Arg260/261 (chain and structure-dependent). Both cryoEM structures (unliganded and IG2 complex) have been reconstructed from EM maps with octahedral symmetry, and therefore contain 24 identical protein chains, spanning from Ala78 to Arg261 (IG2 complex) and Arg262 (unliganded).

The *Mt*HISN5 subunit contains a four α-helix bundle sandwiched between two β-sheets whose four strands are almost perpendicular to each other (Figure 3A). In solution, *Mt*HISN5 forms a homo 24-mer with octahedral (432) symmetry (Figures 2A, F, 3B) of approx. 110 Å diameter and a total mass of 540 kDa. It remains unclear why natural selection has promoted 24-merization for IGPDs/HISN5s. Possible drivers may be (i) minimization of energy cost and amino acid usage (Akashi and Gojobori, 2002; Seligmann, 2003), (ii) proteome stability and efficiency of translation (Kepp, 2020), and (iii) cost of gene expression (Frumkin et al., 2017). Another evolutionary driver of oligomerization may be the benefit of cooperative regulation of the enzyme activity. However, to the best of our knowledge, such a property has never been reported for any IGPD enzyme. The dimensions of the *Mt*HISN5 oligomer did not change significantly after the binding of IG2, indicating a lack of major conformational rearrangements. The active site of *Mt*HISN5 (see below) contains two Mn^{2+} cations (Mn1 and Mn2, Figure 3C) bound by residues belonging to different subunits. In other species, it has been shown that withdrawal of Mn^{2+} causes HISN5 dissociation to inactive trimers while re-addition of Mn^{2+} or other divalent metal cations (e.g., Co^{2+}, Cd^{2+}, Ni^{2+}, Fe^{2+}, and Zn^{2+}) reassembles the enzyme (Sinha et al., 2004; Glynn et al., 2005b). However, we did not observe *Mt*HISN5 trimers at any stage of the purification process, even when 40 mM EDTA was used (not shown).

Characteristics of the *Mt*HISN5 active site

As mentioned above, each subunit of *Mt*HISN5 contains two Mn^{2+} cations (Mn1 and Mn2). The cryoEM and crystal structures share the same pattern of Mn^{2+} coordination (Figures 3C, D). Mn1 is coordinated octahedrally by Nε of His141, Nε of His213, carboxyl of Glu145, Nε of His238* from a neighboring subunit (*), and a water molecule. The Mn2 cations are complexed by Nε of His142, Nε of His115*, Nε of His237*, carboxyl of Glu241*, and a water molecule. The Mn2 coordination sphere is incomplete in the *Mt*HISN5-unliganded cryoEM structure, owing to the lower map resolution. The crystal structures were obtained in the presence of formate (FMT) between Mn1 and Mn2, which completes their coordination spheres (Figure 3C). The corresponding position is occupied by a water molecule in the unliganded cryoEM structure (Figure 2E) or by the imidazole moiety in the *Mt*HISN5-IG2 complex (Figure 2J).

The cryoEM *Mt*HISN5-IG2 complex structure reveals detailed information about the most likely substrate positioning before
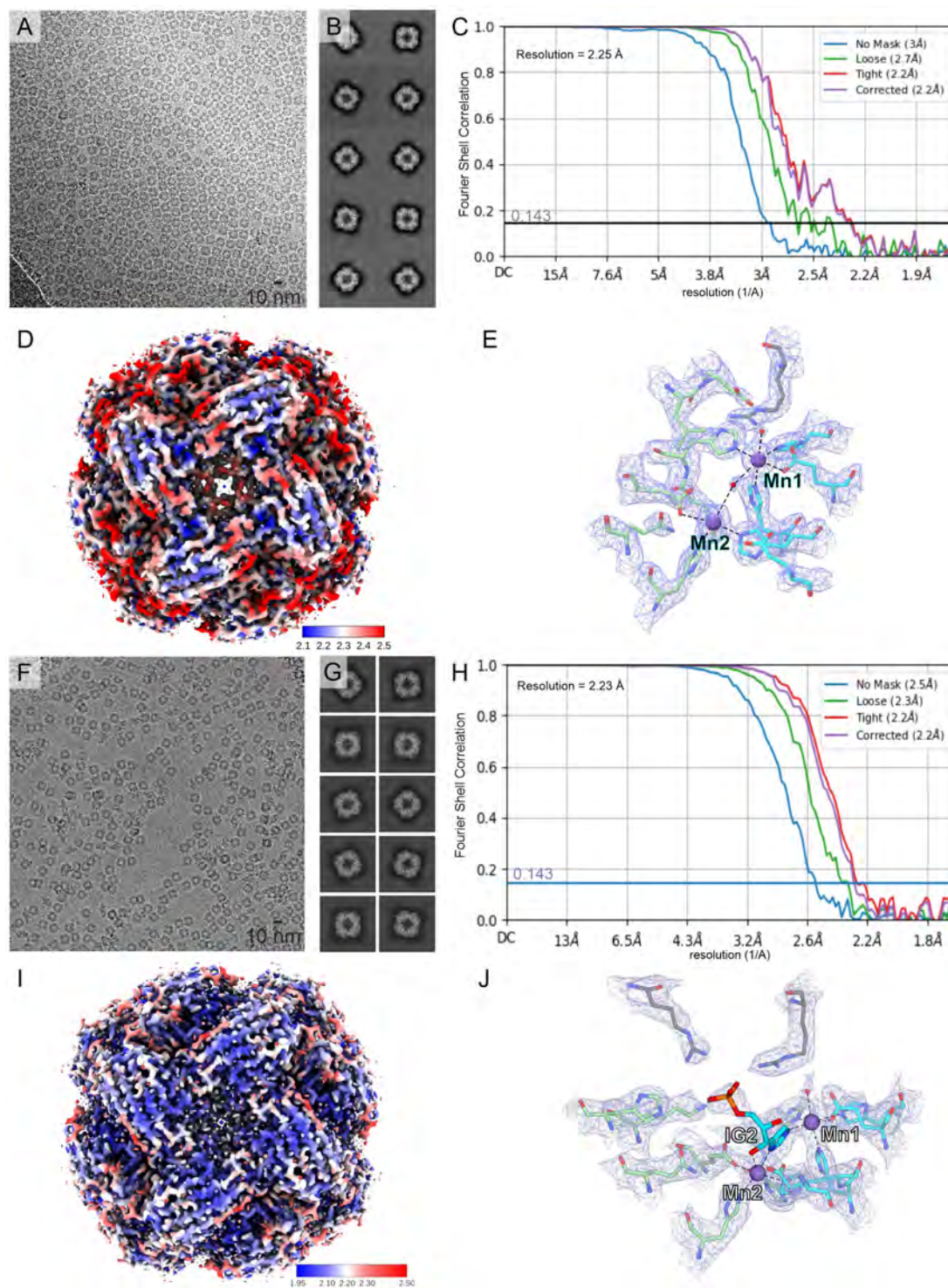
FIGURE 2
CryoEM structures of *Mt*HISN5. **(A–E)** present the *Mt*HISN5-unliganded structure, whereas **(F–J)** refer to the *Mt*HISN5-IG2 complex;
(A, F) representative micrographs; **(B, G)** example 2D classes; **(C, H)** Fourier-shell correlation curves; **(D, I)** maps colored by local resolution; and
(E, J) map fragments within the active site.

catalysis. The imidazole ring is trapped between manganese ions, facing its Nδ towards Mn2 and Nε towards Mn1. While the formation of the active site with the bi-Mn^{2+} cluster requires a contribution of residues from two subunits, a third subunit (**) participates in substrate/product binding by contributing

guanidines of Arg167** and Arg189** that bind the phosphate group (Figure 3D). However, based on our complex with IG2, these polar H-bonds are rather weak, with the donor-acceptor distances of 3.5 – 3.8 Å. Participation of the corresponding Arg121 of *Mycobacterium tuberculosis* HisB in closing the active site has

FIGURE 3

The *Mt*HISN5 structure. **(A)** The subunit of *Mt*HISN5 in two orientations: α-helices are in light green, β-strands are in forest green, and loops are in cyan. Four α-helices form a core bundle surrounded by four β-strands on each side. The N-terminal β-sheet forms an external surface and the C-terminal β-sheet forms the internal surface of the oligomer. **(B)** *Mt*HISN5 24-mer, shown along the three-fold axis. The black, dashed rectangle indicates the regions shown in **(C)**. **(C)** Octahedral coordination of the manganese cations constituting the active site. Formate (FMT, semitransparent) was found in the active sites of crystal structures. Chains are colored differently (light green and turquoise) to emphasize their contribution to the active site formation. The red balls represent the water molecules. Asterisks (*) indicate residues from the counterpart molecule. **(D)** The architecture of the active site in the cryoEM *Mt*HISN5-IG2 complex. The complete substrate binding requires residues from the third (**) subunit. The imidazole ring of IG2 is bound between Mn1 and Mn2 and its phosphate group is held by side chain guanidines of R167** and R189**; semitransparent R167** and R189** present conformation with no ligand in the active site. The cryoEM potential map is contoured 2 Å around IG2 (σ = 0.26) and colored in green. The inset shows the map in semitransparent surface representation, clearly revealing the *2S,3S* configuration in IG2.

been pointed out recently by Kumar and coworkers (Kumar et al., 2022).

It must be emphasized that our *Mt*HISN5-IG2 complex structure is a spectacular example where the cryoEM maps are of such a high quality that they permit resolving stereoisomers of ligands bound to a protein. In fact, we expected the reaction product, IAP as the enzyme was incubated with cIGP (at 2 mM concentration) for 2 days prior to the cryoEM grid preparation. We clearly recognized IG2 based on EM maps (Figure 3D, inset). The improvement in map resolution is an important advancement compared to *At*HISN5, which yielded a 3.1-Å EM map (PDB ID: 6EZJ), making it difficult to determine the positioning of ligands and water molecules as well as distinguishing between the *R*- and *S*-isomers (IGP vs IG2) (Rawson et al., 2018). In this context, using *Mt*HISN5 as a model and our pipeline for the cryoEM structure-based development of novel herbicides will be a significant improvement.

*Mt*HISN5 possesses a variety of ligand-binding hot-spots

In addition to the IG2 observed in our cryoEM complex, we identified several types of molecules bound to *Mt*HISN5, suggesting hot-spots prone to bind certain chemical moieties. The crystal structure at 1.55 Å resolution contains imidazole (IMD), sodium ion, chloride ion, formate (FMT), 1,2-ethanediol (EDO), glycerol (GOL), and tris(hydroxymethyl) aminomethane (TRS). The structure at 1.69 contains FMT, GOL, and TRS. The structure at 2.2 Å contains a chloride ion, FMT, EDO, acetate (ACT), sulfate ion, and CIT. The most abundant binders among these structures are EDOs (n = 41) and FMTs (n = 49). EDOs bind mostly to the inner surface and interfaces between subunits (Figure 4A), but a few are also found on the outer surface and in the vicinity of the active site (4-6 Å, Figure 4B). FMTs primary location was in the active sites, between Mn1 and Mn2 (Figures 3C, 4B). However, it was very
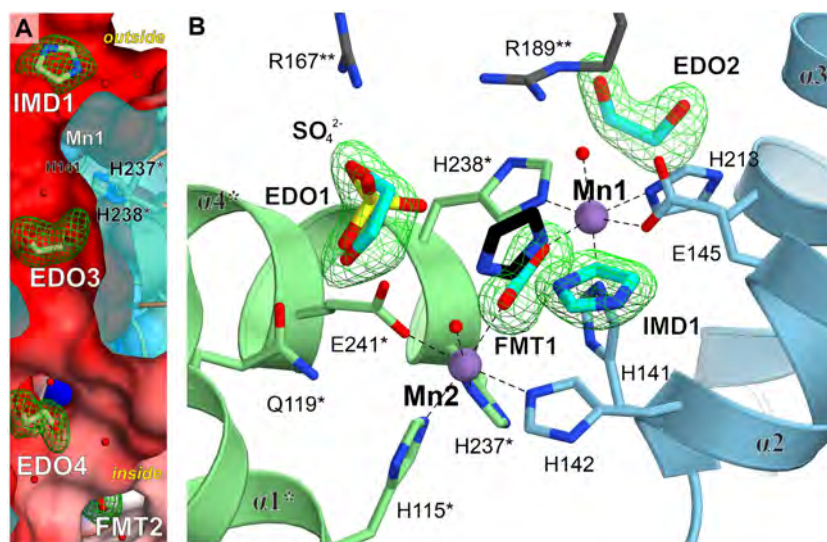
FIGURE 4

Ligand-binding hot-spots in and near the *Mt*HISN5 active site. **(A)** A molecular tunnel along a two-fold axis with ligands trapped inside. Mn1 is located ~6.9 Å from EDO3 which is in the middle of the tunnel. Polder maps around the ligands are contoured at the 6.8 σ level for IMD, 5.6 for EDO3, 5.0 for EDO4 and 7.3 for FMT2. The *outside/inside* labels indicate protein surfaces. The clipped surface is semitransparent. **(B)** Comparison of small-molecule binding positions. IMD1 in *Mt*HISN5 binds to a different site than in the *At*HISN5 structure (black, PDB ID: 4MU1). Polder maps are contoured at 6.7-10.2 σ levels for this work structures.

interesting to find that IMD (IMD1, Figure 4B) bound not between the Mn^{2+} ions in the active site, but instead was positioned approximately 3.8 Å from the C atom of FMT. This is in contrast to other reported structures (*At*HISN5, PDB ID: 4MU1 (Bisson et al., 2015)), where imidazole mimicked part of the substrate/ product (between Mn1 and Mn2). In *Mt*HISN5, IMD1 forms hydrogen bonds with FMT and water. We postulate that the carboxylate and imidazole binding sites can be used as pharmacophores for the design of selective inhibitors of plant HISN5. The fact that we see FMT between Mn^{2+} and imidazole

positioned differently suggests that the bi-Mn^{2+} cluster has a high affinity for carboxylate, in addition to the imidazole of IGP or IAP.

To search for small molecules representing a broader chemical space, we performed virtual screening (VS) by *in silico* docking 3.3 mln lead-like molecules from the ZINC database (Sterling and Irwin, 2015) in the neighborhood of the *Mt*HISN5 active site. Six molecules scoring the highest binding energy gain (between -10.0 and -9.4 kcal/mol) are shown in Figure 5. The top-scoring molecules satisfy the following criteria: (i) the content of heteroatoms that improve water solubility and can potentially
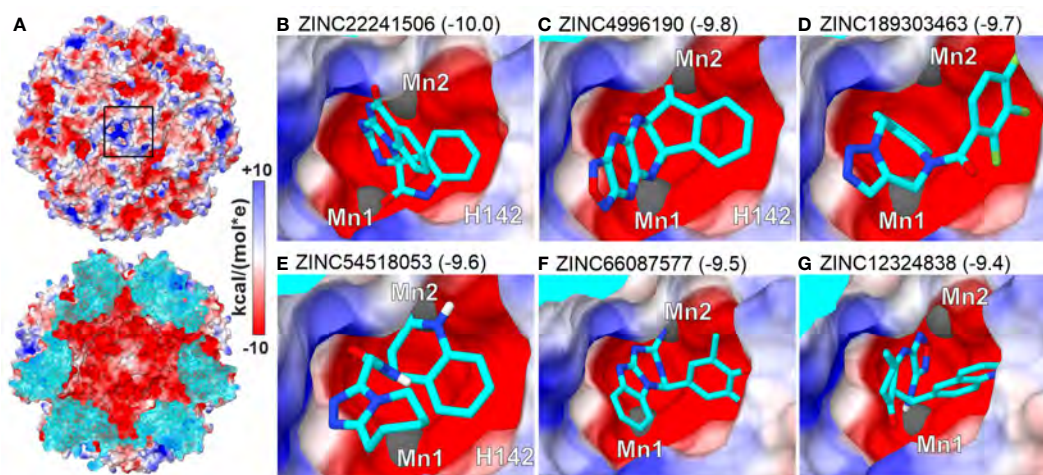
FIGURE 5

Top-scoring results of virtual screening (VS) at the active site. **(A)** *Mt*HISN5 24-mer surface potential (Coulombic); the cross-section (bottom) of the oligomer reveals a negatively charged inner surface. **(B−G)** (same surface color scheme) show molecules with the highest calculated binding energy gain (kcal/mol).

ensure specific binding to the protein, and (ii) the potential for parallel or T-shaped π-stacking with surrounding residues.

Intrigued by the highly symmetric and porous structure of *Mt*HISN5, we analyzed the molecular tunnels (in addition to the one presented in Figure 4A) that could let small molecules penetrate inside the enzyme. To this end, we studied the cryoEM *Mt*HISN5-unliganded structure in the *Caver Analyst 2.0* and *Caver 3.0.3 PyMol plugin* (Chovancova et al., 2012; Jurcik et al., 2018). Most of the tunnels are distributed along the 2-, 3-, and 4-fold axes (Figure 6). The average diameter of the tunnels along the 2-fold axes is large enough to allow infiltration only by very small molecules (< 60 Da), such as EDO, FMT, and water. Larger molecules, however, could permeate the inner cavity of *Mt*HISN5 through the tunnels along the 3-fold and 4-fold axes, which are approximately 2.5 times wider in diameter. Consistently, we identified TRS (121 Da) and CIT (192 Da, Supplementary Figure S1D) in the 3- and 4-fold tunnels, respectively.

Genetic background of *Mt*HISN5 and phylogenetic relationships with its homologs

*Mt*HISN5 (Uniprot ID: I3SDM5) is encoded by the *HISN5* gene (Ensembl: MTR_1g103820, Gene database: LOC25485215) located on chromosome 1. According to the Gene Database (Benson et al., 2013) and the *TargetP 2.0* webserver (Emanuelsson et al., 2000), *Mt*HISN5 contains an exon corresponding to a 76 amino acid residues long chloroplast transit peptide. The following intron separates the sequence of the genuine *Mt*HISN5 enzyme, which complies with previous observations for a canonical isoform in *A. thaliana*, HISN5B (Gene database: AT4G14910). This pattern suggests the fusion of the exon encoding the transit peptide with the exon encoding the enzyme sequence during evolution, which has been observed for several nuclear genes encoding chloroplast proteins (Wolter et al., 1988; Gantt et al., 1991; Stepansky and

Leustek, 2006). A likely reason for the compartmentalization of the HBP is the interconnection with *de novo* purine metabolism that occurs in chloroplasts (and mitochondria). The HBP shares a precursor, 5-phosphoribosyl-1-pyrophosphate (PRPP) and an intermediate, aminoimidazolecarboximide ribonucleotide (AICAR), with *de novo* purine metabolism (Smith and Atkins, 2002; Witte and Herde, 2020). The obtained *Mt*HISN5 structures also allowed us to investigate the two transcript isoforms, X1 which is 1300 nt long, and X2 (1227 nt) in *M. truncatula*. The isoform X2 lacks the 5th exon, which would result in a protein missing the loop-β7-loop fragment (residues Asp185 to Gln209). Therefore, it is very unlikely that the isoform X2 is expressed as a functional enzyme.

To assess the similarity between prokaryotic and eukaryotic IGPD/HISN5 enzymes, we analyzed 12 710 sequences from the InterPro family IPR000807 by calculating a sequence similarity network (SSN, Figure 7). The result showed a close relationship between plants (*Viridiplantae*), green algae (*Chlorophyta*), and *Cyanobacteria* suggesting that plant HISN5s derive from cyanobacterial IGPDs. This is consistent with the endosymbiotic theory. To verify this observation, we generated a phylogenetic tree based on homologous sequences of *Mt*HISN5 (Supplementary Figure S2). The tree also showed a close relationship between the plant and cyanobacterial IGPDs, supporting the cyanobacterial origin of plant HISN5 enzymes. In contrast, we have recently shown that plant *HISN2* and *HISN6* are distant homologs of their cyanobacterial counterparts and are likely to have been acquired by HGT (Witek et al., 2021; Rutkiewicz et al., 2023).

The analysis also revealed a relatively large group of bacterial bifunctional enzymes presenting IGPD and HPP activity (Figure 7). Previous analyses of the phylogenetic origin of bacterial bifunctional *hisB*, i.e., *hisNB*, genes were limited to the classes of γ- and ε-*Proteobacteria* (Brilli and Fani, 2004; Kinateder et al., 2023). Our results suggest that the presence of fused genes occurs also in other bacterial phyla, providing new models for *his* genes' evolution, especially interesting when combined with novel
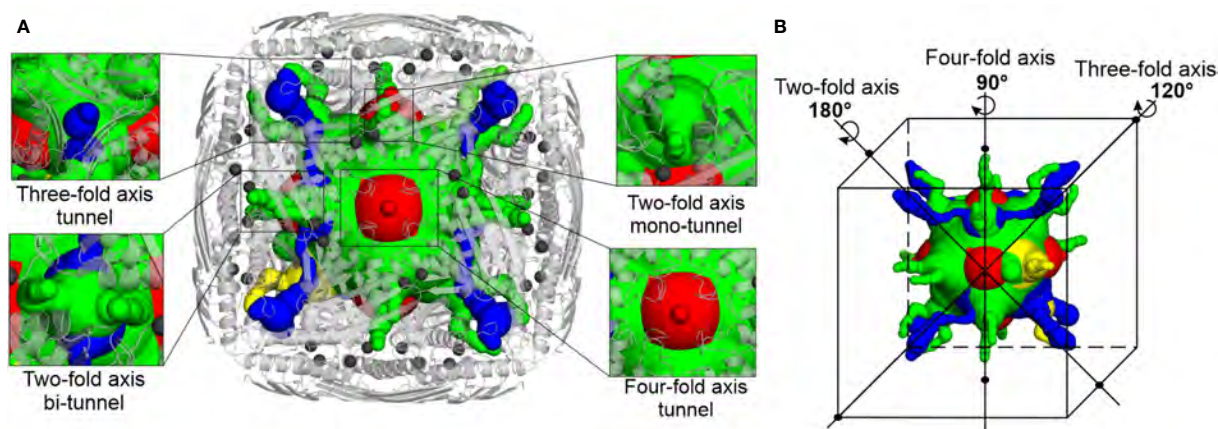
FIGURE 6
Molecular tunnels in the *Mt*HISN5 structure. **(A)** Tunnel distribution along symmetry axes; the tunnels along two-fold axes are subdivided into mono- and bi-tunnels. Two-fold axis mono-tunnels are depicted in yellow and bi-tunnels are depicted in green. Three-fold axis tunnels are blue and four-axis tunnels are red. Protein oligomer is shown as light gray cartoons and manganese ions are depicted as dim gray spheres. Note the proximity of the tunnel to the manganese-containing active sites. **(B)** Positioning of the tunnels along the symmetry axes; colors are the same as in **(A)**.

FIGURE 7

Sequence similarity network for the InterPro Family IPR000807. Plant and green algae HISN5 enzymes derive directly from *Cyanobacteria*. Red algae (*Rhodophyta*) constitute one group with *Cyanobacteria*. Fungal sequences (*Ascomycota* and *Basidiomycota*) are separated from other eukaryotic sequences. Bifunctional IGPD + HPP sequences from bacteria form a large group. Most bacterial and archaeal sequences are scattered in the diagram. Dots represent prokaryotes, diamonds represent eukaryotes, and the edges between them represent similarities.

approaches for enzyme functional annotation (Kinateder et al., 2024). These bifunctional enzymes exist in *Acidobacteriota* (*Acidobacteria*), *Actinomycetota* (*Actinobacteria*)*, Bacteroidota* (*Bacterioidetes*), *Calditrichota* (*Calditrichaeota*)*, Pseudomonadota* (*γ-Proteobacteria*) and *Spirochaetota* (*Spirochaetes*); named according to the recent nomenclature update by the International Code of Nomenclature of Prokaryotes (Oren and Garrity, 2021) (former names are in parentheses).

In general, the majority of bacterial sequences present high sequence variability between phyla, which is much more significant than, for example, within plants. Moreover, fungal IGPD sequences are disconnected from other groups, suggesting that they differentiated early and have evolved in parallel. The conservation of HISN5 sequences in plant species, clearly distant from other kingdoms, suggests the possibility of reaching general-purpose herbicidal activity by potent HISN5 inhibitors. Such a potential is discussed in detail in the next chapter.

Distinct features of plant HISN5 enzymes near the active site provide guidelines for inhibitor design

The overall fold of IGPD enzymes, including their 24-meric assembly, is highly conserved between kingdoms of life, even though sequence conservation varies strongly. More precisely, *Mt*HISN5 sequence shares 88% identity with *A. thaliana* HISN5B (*At*HISN5B; 89% with *At*HISN5A), 45% with *Acanthamoeba castellani* (*Ac*IGPD), 40% with *Saccharomyces cerevisiae* (*Sc*IGPD), and 40% with *Staphylococcus aureus* (*Sa*IGPD). Therefore, we decided to perform a detailed analysis to pinpoint differences that could be exploited to ensure the selectivity of future inhibitors of plant HISN5 versus bacterial IGPD homologs. To obtain a perspective on the conserved and variable regions, we analyzed residue conservation using the *ConSurf* web server (Ashkenazy et al., 2016). Results of the Multiple Sequence
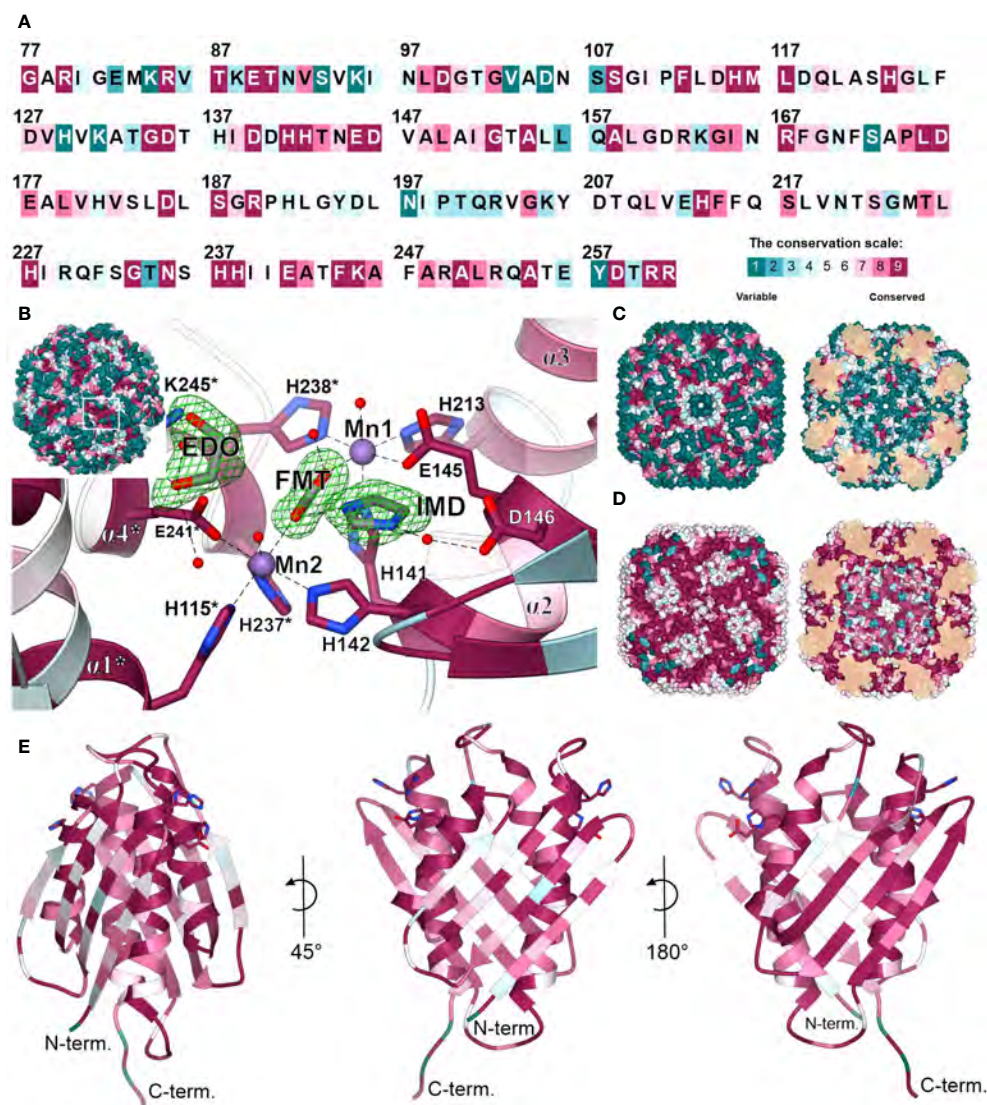
FIGURE 8
Residue conservation mapped on *Mt*HISN5. **(A)** Color-coded residue conservation of *Mt*HISN5 residues compared to all kingdoms based on the multiple sequence alignment (MSA); the color key given in the bottom-right corresponds to all panels in this Figure. **(B)** A close-up view of the conserved active site. Residues with the highest conservation score (all kingdoms) are involved in Mn^{2+} coordination and in the binding of small molecules in our structures. Red balls represent water molecules; formic acid (FMT), imidazole (IMD), and ethylene glycol (EDO) are in stick representation and contoured with polder electron density maps at the 10.0, 7.7, and 7.0 σ levels, respectively. **(C)** The outer surface (left) and inner surface (right) residue conservation based on the MSA calculated for all kingdoms. **(D)** Residue conservation among plants. Notice patterns at the interfaces on the outer surface of the protein. **(E)** The *Mt*HISN5 subunit is colored according to residue conservation in plants. The exposed side chains are residues that coordinate manganese ions.

Alignment (MSA) amongst all kingdoms reveal that the highest conservation score was assigned to residues forming HISN5 active sites and coordinating manganese ions: (i) Mn1 by His141, Glu145, His213, His238*, and (ii) Mn2 by His115*, His237*, Glu241*, His142 (Figure 8A). There were other residues located close to the active site area that were assigned the highest conservation rank and are also conserved in the aforementioned species, e.g., Asp146 or Lys245 that take part in ligand binding (Figure 8B) or Arg167 and Arg189 that stabilize the substrate's phosphate group by weak hydrogen bonding (Figure 3D). Interestingly, residues comprising both β-sheets that are located on the outer and inner surfaces of the oligomer are rather variable across sequences from all kingdoms

(Figure 8C). It is possible that this variability stems from the evolutionary pressure caused by operating in various environments, such as the bacterial and fungal cytosol and chloroplast stroma of plants.

When only protein sequences within the plant kingdom are considered, residues forming β-sheets on the outer and inner HISN5 surfaces are more conserved (Figure 8D). Still, more variability is observed at the outer surface compared to the inner (Figure 8E), suggesting some evolutionary pressure to maintain the environment in the hollow core of HISN5. In addition to the active site residues, the highest conservation was observed at the inter-subunit interfaces. Residues forming the central α-helical bundle

FIGURE 9
Depiction of the cleft that connects two active sites and is highly conserved in plants. All the panels show the same perspective. Panels **(A)** and **(B)** show the residue conservation scores for all organisms and plants, respectively. **(A)** contains the ligands bound in the crystal structures. The same residues in **(B)** are dark pink, meaning they are highly conserved in plants. The distance between Mn1 and Mn1* is indicated in **(B)**. **(C)** depicts hydrophobicity in the region of interest. **(D)** presents coulombic potential that indicates a majority of negatively charged residues. The proposed scaffold of symmetric inhibitors reaching two active sites includes competitive inhibitor (Comp. Inh.), aromatic (Ar), and hydrogen bond donating (H-donor) moieties.

are also highly conserved in plants (Figure 8E). Interestingly, several loops are also conserved, which is common not only for loops involved in substrate recognition, but also for those shaping internal tunnels, channels, or voids (Kress et al., 2018).

We then compared the *ConSurf* data obtained for all kingdoms and exclusively for plants to identify a surface region in the vicinity of the active site that would be conserved in plants but vary in other kingdoms (Figures 9A, B). The goal was to propose a development pathway that would ensure both selectivity for plant HISN5 sequences and high potency, which will be crucial for designing HISN5 inhibitors and the subsequent development of herbicides. In this context, the cleft near Thr153, Ser187, His191, Asn220, and Thr221 (Figure 9C) is the most interesting, being variable in other species (Figure 9A) and highly conserved in plants (Figure 9B). This cleft connects active sites of two *Mt*HISN5 subunits (the Mn1-Mn1* distance is ~27 Å). The cleft is long and intrinsically symmetric, as it is crossed by one of the 2-fold axes. It is also rather hydrophilic and negatively charged (Figures 9C, D, respectively).

A common compound optimization method involves linking molecules that bind to a target at separate sites (Kirsch et al., 2019). The most versatile linkers are oligoethylene glycol chains. However, neither ethylene glycol molecules nor polyethylene glycol (PEG) fragments were identified in our structures, despite their use at high concentrations. This is consistent with the negative charge in the cleft center, which makes it predisposed to bind H-bond donors. However, PEG is an H-bond acceptor, except for the terminal

hydroxyl groups. Therefore, we conducted another VS campaign focused on the cleft to find more suitable linkers between the two active sites. This time, however, we narrowed the screening library to include only more polar compounds (logP ≤ 2, ~1.3 mln molecules). The estimated binding energies were weaker than those obtained by VS in the active site (-8.1 vs. -10 kcal/mol, Supplementary Figure S3). Nonetheless, interesting common features became apparent in the top-scoring molecules. First, the fragment near the 2-fold axis between Pro190 and Pro190* appears prone to bind hydrophilic, preferably aliphatic, and six-membered-ring moieties, whereby secondary amines are H-bond donors. In contrast, more hydrophobic moieties tend to bind between the apolar parts of the Arg189 and Thr221* side chains, which are closer to the active sites. It is very important to note that targeting this cleft will enable the design of symmetric inhibitors that reach two active sites simultaneously. Such a multivalency can enhance the overall binding affinity and selectivity, reducing the likelihood of interactions with off-target proteins and minimizing side effects. Based on our experimental, comparative, and computational approaches, the scaffold which would best correspond to the HISN5 pharmacophore, would include competitive inhibitor (Comp. Inh.) moieties, at the poles of the molecule (Figure 9D). Next, aromatic moieties would connect to the molecule center that would be created by a symmetric and polar moiety with H-bond donors. Importantly, the competitive inhibitor moiety does not need to be a substrate/product or a transition state analog but could

also contain carboxylate and imidazole moieties, as based on our crystal structures.

MtHISN5 activity measurements using isothermal titration calorimetry

The supply of IGP is limited, and only 1-mg packages are currently available. Previously used absorption-based methods require large quantities of IGP as one needs to prepare 8 to 12 separate solutions, each containing a different substrate concentration, to obtain a single set of experimental data (Hawkes et al., 1995). Therefore, we decided to adopt and adjust the isothermal titration calorimetry single-injection method (ITC-SIM (Wang et al., 2020)). ITC-SIM consumes only ~200 µL of substrate at saturating concentration (instead of 8-12 mL) to obtain the enzyme kinetics graph showing rate of reaction as a function of substrate concentration. Our first ITC trials were performed on the same cIGP that was used for the cryoEM experiments. Although other IGPD enzymes were assayed using cIGP from the same source, we were not able to determine kinetic parameters. We could only confirm that the enzyme was active, as the curves clearly showed an exothermic event when compared to blank experiments (Figures 10A, B). Nonetheless, these measurements provided interesting insights into the MtHISN5 behavior with cIGP. The exothermic reaction occurred after a lag phase, whose length was directly proportional to the cIGP concentration. During the lag phase, the heat production was very low, whereas theoretically, one should observe saturation with the substrate at the beginning of the experiment (Figure 10A). When using 24, 65, 253, and 800 µM of

cIGP, this lag lasted for about 5, 7, 12, and 25 minutes, respectively. Subsequently, the highest heat production was observed, reflecting the achievement of V_{max} under the given conditions. Interestingly, with 2 mM cIGP, the initial lag reached the maximum time that the PEAQ-ITC apparatus offers (170 min). This peculiar behavior can likely be attributed to the contamination of cIGP with IG2, as revealed by our cryoEM experiments. The presence of IG2 (which cannot be enzymatically converted to IAP) in cIGP was also observed by others (Saika et al., 1993; Bisson et al., 2015). Hence, it was impossible to saturate MtHISN5 with the cIGP substrate because an increase of the 2R,3S diastereoisomer concentration always elevated the amount of 2S,3S (IG2), which is a competitive binder.

To eliminate IG2 in the substrate sample, we synthesized IGP enzymatically (hereafter referred to as eIGP). We utilized MtHISN1-4 and E. coli inorganic pyrophosphatase, each purified separately by Ni²⁺ affinity chromatography. Each reaction was monitored by absorbance (Supplementary Figure S4). After the reaction with MtHISN4 was completed, the mixture was run through the Ni²⁺ resin to eliminate the proteins. This approach yields stereochemically pure IGP owing to the stereoselectivity of the HBP enzymes. The obtained yield of eIGP synthesis was 80%.

The ITC-SIM data obtained with eIGP were significantly more informative than those obtained with cIGP (Figure 10C). Despite the strong buffer mismatch, deriving from the high salt content in the vacuum-concentrated eIGP sample (in SpeedVac) and a likely heat effect of protein-substrate initial interaction, it was possible to observe the V_{max} plateau and the substrate depletion curve. Those data allowed us to fit the Michaelis-Menten equation and calculate the kinetic parameters: $K_M = 227 \pm 45$ µM, $k_{cat} = 3.4 \pm 0.3$ s⁻¹

FIGURE 10

ITC measurements. **(A)** ITC-SIM superimposed raw data plots obtained after the blank experiment; injection of 30 µL of 0.5 mM commercial IGP (cIGP) into the cell containing buffer (black), 10-µL injection of 0.5 mM cIGP (final concentration 24 µM) into the cell with 10 nM MtHISN5 (violet), 30-µL injection of 0.5 mM cIGP (final 65 µM) into the cell with 10 nM MtHISN5 (green), 35 µL injection of 1.7 mM cIGP (final 253 µM) into the cell with 5 nM MtHISN5 (orange), injection of the MtHISN5 to the final concentration of 12 nM into the cell with 800 µM of cIGP (gray). **(B)** Injection of MtHISN5 at a final concentration of 15 nM into the cell with 2 mM cIGP (magenta) and the blank (buffer to the substrate) experiment (black). **(C)** Representative raw ITC-SIM data (green line) after injection of 3.8 µL of 48 µM enzyme (to the final concentration in the cell of 901 nM) into eIGP at 3.4 mM concentration. DP values in the area marked with the red, dashed lines were converted into the rates and plotted against the substrate concentration; the black curve shows raw data from the blank experiment (buffer to reaction mixture). **(D)** The rates plotted against the substrate concentration were fitted with the Michaelis-Menten equation (red line), and the given parameters were calculated as an average of the values obtained from three separate experiments.

(Figure 10D). The K_M value reported for *At*HISN5 was 170 μM (Tada et al., 1995; Bisson et al., 2015), while HISN5 enzymes in crops exhibit a wide K_M range of 49 and 83 μM in *Triticum aestivum* (germ wheat and mature wheat, respectively), 600 μM in barley (*Hordeum vulgare*), up to 1.7 mM in oat, *Avenia sativa* (Wiater et al., 1971c). As for the k_{cat} value, according to our best knowledge and the BRENDA Enzymes database (Chang et al., 2021), only k_{cat} = 1400 for *M. tuberculosis* IGPD has been reported (Ahangar et al., 2013). Importantly, when comparing kinetic parameters from different studies, one should bear in mind different measurement conditions. Concentration of eIGP in the reaction buffer (optimal for enzymes synthesizing the substrate) results in relatively high salt content in the *Mt*HISN5 reaction (200 mM KCl and 100 mM NaCl), which may lower k_{cat}. Nonetheless, the presented ITC-SIM enabled us to measure the kinetics for *Mt*HISN5 and obtain K_M and k_{cat} values consuming ~200 μL of saturating (3 mM) eIGP per one replicate. Such amounts are significantly lower than those required by former methods (Hawkes et al., 1995).

Conclusions and outlook

For the past two decades, there has been an increase of interest in deciphering plant HBP, first from genetic, and then structural aspects. Previous methodological limitations in genetics, molecular, and structural biology have been overcome. The advent of new molecular tools, e.g., CRISPR-Cas9 (clustered regularly interspaced palindromic repeats)/Cas9-mediated genome editing), has proved that the lack of auxotrophic mutants is no longer a problem. Recently, a model liverworth, *Marchantia polymorpha*, was established as a eukaryotic *his* auxotrophic system for studying biocontainment and transformant selection without the need for antibiotics (Fukushima and Kodama, 2022). For many years, crystallography has been the only way to study the structures of plant HBP enzymes in detail (Ruszkowski and Dauter, 2016; Ruszkowski and Dauter, 2017; Ruszkowski, 2018; Witek et al., 2021; Rutkiewicz et al., 2023). The same was true for HISN5 (Glynn et al., 2005a; Bisson et al., 2015, Bisson et al., 2016) but now cryoEM will help to solve the experimental structures of plant HISN5 complexes with small molecules, including herbicide candidates.

This work also provides insights into *Mt*HISN5 phylogenetic relations with its prokaryotic and eukaryotic homologs, indicating that plant HISN5 sequences derive from *Cyanobacteria*, which is consistent with the endosymbiotic theory. Computational tools helped us point out highly conserved residues in plant HISN5s and map them onto the protein structure obtained experimentally. The highest conservation scores referred to the residues contributing to substrate binding and to those located at the intersubunit interfaces. Using both experimental and computational tools, we identified hot-spots that can interact with small molecules. Crystal and cryoEM structures allowed the identification of numerous ligands bound to *Mt*HISN5, i.e., FMT, EDO, IMD, TRS, CIT, GOL, ACT, PEG, Na^+, Cl^-, and SO_4^{2-} (waters and Mn^{2+} excluded). None of the ligands significantly affected the

global conformation of the enzyme. VS indicated molecules that could potentially bind to the *Mt*HISN5 surface. *In silico* analyses allowed us to characterize tunnels of different lengths and diameters, through which small molecules can permeate to the HISN5 central void and bind to the inner surface. Although with current data it is impossible to determine whether these tunnels are relevant for catalysis, novel plant HISN5 inhibitors could partially bind there, resulting in increased potency and selectivity.

HISN5 has been a target for herbicide design for decades, but no HISN5-specific herbicide is currently available (Klopotowski and Wiater, 1965; Wiater et al., 1971a, Wiater et al., 1971b). In the past years, there has been a noticeable growth of interest in targeting HISN5 (Bisson et al., 2015, Bisson et al., 2016; Rawson et al., 2018; Wang et al., 2021a, Wang et al., 2021b). The most common candidates for HISN5 inhibitors are triazole compounds such as amitrole or 2-hydroxy-3-(1,2,4-triazol-1-yl) propylphosphonate (C348) (Bisson et al., 2016). However, significant impediments have been affecting the evaluation of HISN5 activity owing to the scarcity of IGP on the market and the presence of competitively binding IGP diastereoisomer, IG2. With that in mind, we have adopted and adjusted ITC-SIM to measure HISN5 activity using enzymatically synthesized IGP. Therefore, this study sheds not only new light on plant HISN5, provides novel fragments of potential HISN5 inhibitors but also presents new methodologies for the rational design of HISN5 inhibitors with herbicidal activity.

Materials and methods

Sequence similarity network

Sequence similarity networks were calculated using the EFI-EST web server (Zallot et al., 2019). Input data consisted of 36 356 sequences belonging to the InterPro Family IPR000807 and were later reduced to 12 710 UniRef90 sequences. The calculations were based on sequences of 180 – 390 residues long and an alignment score of 80. Output results were visualized in *CytoScape 3.3* (Shannon et al., 2003).

Cloning, expression, and purification

The coding sequence (CDS) of the *Mt*HISN5 gene was retrieved from the NCBI database (entry XP_013469848.1). The CDS was PCR-amplified using the primers *Mt*HISN5-Nt77-F (TACTTCCAATCCAATGCCGGTGCTAGAATTGGAG AGATGAAAAGG) and *Mt*HISN5-CtFL-R (TTATCCAC TTCCAATGTTAACTACGCGACAGAACCCCTTTTGAA). The PCR product was purified and cloned into the expression plasmid pMCSG68 using the ligase-independent cloning (LIC) method (Kim et al., 2011). The final *Mt*HISN5 construct in this work was N-terminally truncated at Ser70 to obtain a higher yield of expression and enzyme stability compared to other tested constructs, truncated at residues 15, 30, 45, and 56. Because of the problematic purification of the enzyme fused with the His-tag, we deleted the tag entirely and instead inserted the 70-76 region of

the original *Mt*HISN5 sequence using the polymerase incomplete primer extension (PIPE) method (Klock et al., 2008). The primers used for the PIPE method were *Mt*HISN5-Nt70-delHT-F (GAAGGAGATATACATATGCAACTTTCCCAT ATTGACTCAGGTGC) and *Mt*HISN5-Nt70-delHT-R (CATATT GTTATATCTCCTTCTTAAAGTTAAACAAATATTAT TTTCTAGAGGGG). The cloning correctness was confirmed by DNA sequencing. Overexpression was carried out in BL21 Gold *E. coli* cells (Agilent Technologies) in an LB medium containing 150 µg/mL ampicillin. The cultures were grown at 37°C and shaken at 180 rpm. When OD_{600} reached 1.0, the temperature was lowered to 18°C, and the overexpression was induced using 0.5 mM isopropyl-D-thiogalactopyranoside (IPTG) and went on for 18 h; $MnCl_2$ at 10 mM final concentration was added after IPTG. The cultures were centrifuged at 5000 × g for 15 min at 4°C and the sediment was suspended in 30-35 mL of purification buffer (40mM Tris–HCl pH 8.0, 40 mM NaCl, 4 mM Mn^{2+}, 0.4 mM EDTA) and frozen at -75°C for purification.

The cells were disrupted by sonication (5 min with intervals for cooling), and cell debris was removed by centrifugation at 25,000 × g for 30 min at 4°C. We then followed a method of purification with DEAE-cellulose and NaCl gradient described by Glynn (Glynn et al., 2005b), however, *Mt*HISN5 was not present in the increasing NaCl gradient but in the first flow-through. Therefore, we decided to skip the gradient procedure and modify the method. After centrifugation at 25,000 × g, *Mt*HISN5 was precipitated by the addition of 1.7 M ammonium sulfate. The precipitate was collected by centrifugation at 20,000 × g for 15 min at 4°C and dissolved in 2 mL of purification buffer. The protein solution was filtered through a 0.45 µm syringe filter and 2.0 mL were loaded onto a Superose 6 column previously equilibrated with the purification buffer for size-exclusion chromatography (SEC). After elution, fractions containing *Mt*HISN5 were combined, concentrated to ~ 2.0 mL using Amicon Ultra Centrifugal Filters (Merck), and loaded onto a Superdex 200 column, also equilibrated with the purification buffer. Eluted fractions containing pure *Mt*HISN5 (based on SDS-PAGE) were pooled and concentrated.

The concentration was assessed by two methods because of the low extinction coefficient (ϵ = 4470 M^{-1} cm^{-1}) and possible contamination with small molecules absorbing at 280 nm, thus creating a false-positive result of a higher than actual protein concentration (>50 mg/mL). The initial measurements were conducted at λ = 280 nm. The second spectrophotometric measurement was performed using the Bradford method (Bradford, 1976). Crystal structures were obtained from concentrations of 11-15 mg/mL, whereas cryoEM structures were obtained from 1 mg/mL.

*Mt*HISN1 and *Mt*HISN2 for enzymatic synthesis of IGP were produced as described previously (Ruszkowski, 2018; Witek et al., 2021), omitting TEV cleavage and dialysis. *Mt*HISN3 (N-terminally truncated at residue 42), *Mt*HISN4 (truncated at residue 48), and *E. coli* inorganic pyrophosphatase were obtained following the procedure described for *Mt*HISN1, with SEC directly after elution from Ni-NTA resin. The following primers were used to amplify *Mt*HISN3 and *Mt*HISN4 CDSs: MtHISN3-Nt42-F TACTTCCAATCCAATGCCTCTCCACCTTCAATT

TCAATGCTCCGTTCAATTC, MtHISN3-CtFL-R TTATCCACT TTCCAATGTTAAGCCACTGAGACCTTTTGCTGGTTATGC, MtHISN4-Nt48-F TACTTCCAATCCAATGCCACTTCTAT ATGATTCTGTTGTGACTTTGCTTGATTATGGTG, and MtHISN4-CtFL-R TTATCCACTTCCAATGTTAGATTCGG ACTTCTATGCCTTCATTCAACAAATGTTCTTT.

Crystallization, X-ray data collection, and processing

*Mt*HISN5 was crystallized using the vapor diffusion method (hanging drop) and all crystallizations were set up manually at 20°C. The structure at 1.55 Å was obtained from the Morpheus screen (Gorrec, 2009) in Molecular Dimensions (MD1-46), condition 2-27 (G3) containing 0.1 M carboxylic acids (0.2 M sodium formate, 0.2 M ammonium acetate, 0.2 M sodium citrate tribasic dihydrate, 0.2 M potassium sodium tartrate tetrahydrate, 0.2 M sodium oxamate), 0.1 M imidazole, 4-morpholine-ethane-sulfonic acid (MES) adjusted by ratio to pH 6.5 and 30% v/v precipitant mix (40% v/v glycerol and 20% v/v PEG 4000).

The ShotGun Screen (SG-1 MD1-89-ECO), from Molecular Dimensions (Fazio et al., 2014) supplemented with 15% glycerol yielded the structure at 1.69 Å resolution. The measured crystals were obtained by mixing 2.0 µL of the *Mt*HISN5 solution with 2.0 µL of the condition 2-11 (E11) containing 2.0 M sodium formate, 0.1 M sodium acetate pH 4.6.

The structure at 2.2 Å was obtained based on the Morpheus screen (Gorrec, 2009) (MD1-46) in condition 2-31 (G7) containing 0.1 M carboxylic acids (0.2 M sodium formate, 0.2 M ammonium acetate, 0.2 M sodium citrate tribasic dihydrate, 0.2 M potassium sodium tartrate tetrahydrate, 0.2 M sodium oxamate), sodium 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), 3-(*N*-morpholino) propane sulfonic acid (MOPS) at pH 7.5, and 30% v/v precipitant mix (40% v/v glycerol and 20% v/v PEG 4000). The crystals were cryoprotected with 15% glycerol, vitrified in liquid nitrogen, and stored for data measurement. Diffraction data for structures at 1.55 and 1.69 Å were measured at the P13 beamline at the PETRA III synchrotron in Hamburg, Germany. The diffraction data for the structure at 2.2 Å were measured using an in-house X-ray diffractometer, Rigaku XtaLAB Synergy-R. All datasets were processed using the *XDS Package* (Kabsch, 2010). Data statistics are summarized in Table 1.

Determination and refinement of the crystal structures

The crystal structure of *Mt*HISN5 was solved using molecular replacement based on the structure of *A. thaliana* HISN5 (PDB ID: 4MU0) in *PHASER* (Mccoy et al., 2007). The initial model was built using *Phenix.Autobuild* (Terwilliger et al., 2008). The *ACHESYM* server was used to rearrange the model within the unit cell (Kowiel et al., 2014). Automatic model refinement was performed in *Phenix.Refine* (Afonine et al., 2018) and manual corrections were conducted in *COOT* (Emsley et al., 2010). For generation of ligand restraints, *Phenix.eLBOW* was used (Moriarty et al., 2009). The

TABLE 1 Diffraction data and refinement statistics.

	*Mt*HISN5 (1.55Å) PDB ID: 8QAW	*Mt*HISN5 (1.69 Å) PDB ID: 8QAX	*Mt*HISN5 (2.2 Å) PDB ID: 8QAY
Diffraction source	PETRA III, Beamline P13, DESY Hamburg	PETRA III, Beamline P13 DESY Hamburg	Rigaku XtaLAB Synergy-R IBCH PAS Poznan
Wavelength (Å)	0.9762	0.9763	1.5418
Temperature (K)	100	100	100
Rotation range per image (°)	0.1	0.1	0.2
Total rotation range (°)	360	240	135
Space group	*R3*	*I4*	*R3*
a, b, c (Å)	137.5, 137.5, 265.6	120.5, 120.5, 183.0	137.6, 137.6, 265.8
Mosaicity (°)	0.056	0.116	0.169
Resolution range (Å)/ highest resolution shell	58 – 1.55/ 1.65-1.55	80 – 1.69/1.79 – 1.69	80 – 2.20/2.26 – 2.20
No. of unique reflections	272024	145061	93797
Completeness (%)	99.9/99.5	99.9/99.2	98.8/95.0
Redundancy	10.29	9.13	2.99
$I/\sigma(I)$	11.9/1.8	21.6/1.1	5.7/1.2
R_{meas} (%)	11.0/102.1	5.6/215.1	23.4/134.7
CC1/2 (%)	99.7/83.3	99.9/57.3	98.6/47.0
No. of reflections: working/test set	272 024/1089	145 061/1001	93797/932
R_{work}/R_{free}	0.129/0.161	0.174/0.198	0.189/0.231
No. of non-H atoms: Protein/Ligand/Water	11518/ 263/1334	8572/88/498	11536/ 176/777
R.m.s. deviations: Bonds (Å)/Angles (°)	0.005/0.783	0.006/0.808	0.007/0.931
Ramachandran plot: Most favored/allowed/ outliers (%)	96.2/3.8/0.0	95.8/4.2/0.0	95.6/4.4/0.0
Average B-factor: Protein/water/ ligands (Å²)	29.3/45.8/50.9	41.5/47.2/44.8	36.7/40.4/47.8

structure at 1.55 Å was refined anisotropically, and structures at 1.69 Å and 2.2 Å were refined isotropically with translation-libration-screw (TLS) parameters. The refinement statistics are included in Table 1.

CryoEM data collection

Preparation of the cryoEM samples and the data collection were performed at the SOLARIS CryoEM Facility (Kraków, Poland).

Quantifoil TEM grids (300 mesh R1.2/1.3 copper) were glow-discharged on EM ACE200 (Leica Microsystems). The grids were then placed inside the FEI Vitrobot Mark IV chamber set to 100% humidity at 4°C and a total of 2.5 µl of the protein solution was applied (blotting parameters: blot time, 4 s; wait time, 10 s; drain time, 0 s; blot force, 0; blot total, 1). The grids were plunge-vitrified in liquid ethane and clipped in liquid nitrogen. The data were collected on a Titan Krios microscope (Thermo Fisher Scientific) operated at 300 kV, equipped with an FEI Falcon III (4k x 4k) direct electron detector. The detector operated in counting mode at 96 000× magnification, resulting in a calibrated physical pixel size of 0.86 Å px^{-1}. The micrographs were acquired as 40-frame movies (the total dose of 40 e$^-$ Å^{-2}) at under-focus with a defocus range of −3.0 to −0.9 µm and 0.3 µm defocus step. A total of 954 and 3330 micrographs were collected for the *Mt*HISN5-unliganded and *Mt*HISN5-IG2 structures, respectively.

CryoEM data processing

For the *Mt*HISN5-unliganded dataset, reconstruction of the map was performed in *Relion 3.1* within the *CCP-EM* package (Scheres, 2012; Wood et al., 2015). The contrast transfer function (CTF) was estimated in *CTFFIND4* (Rohou and Grigorieff, 2015); 871 micrographs passed the curation step. 2D references for particle picking were obtained after automatic picking of 8707 particles from 31 micrographs. Template-based picking identified 456759 particles which were extracted as 352-pixel boxes. Based on the 2D classification, 424472 particles (41 classes out of 50) were selected and subjected to 3D classification. Finally, 229366 good particles were used for high-resolution refinement with the *O* symmetry imposed. Per-particle CTF refinement and Bayesian particle polishing after the first refinement improved the resolution from 3.2 to 2.4 Å (gold-standard Fourier-shell correlation, GSFSC; without masking and postprocessing). The "shiny" particles were then imported into *Cryosparc 4.4* and used in non-uniform refinement which further improved the resolution to 2.25 Å.

The *Mt*HISN5-IG2 cryoEM data were processed in *Cryosparc 4.1* (Punjani et al., 2017). 2990 micrographs were selected based on the CTF fit resolution, ice thickness, and accumulated motion. Blob picking, followed by 2D classification was used to generate four templates for automatic picking. After inspection of the picks, 661163 particles were retained and extracted in 300-pixel boxes for 2D classification (100 classes). The good 25 classes contained 605614 particles which were used to build the initial map and for high-resolution refinement. The best resolution (2.2 Å, GSFSC) was obtained using the non-uniform refinement protocol.

The molecular model of *At*HISN5 was placed into the map in *UCSF Chimera 1.15* (Pettersen et al., 2004) and the sequence was fitted using *Phenix.Autobuild* (Terwilliger et al., 2008). Manual corrections to the models were done in *Coot* (Emsley et al., 2010), between iterative rounds of automatic real-space model refinements in *Phenix.Refine* (Afonine et al., 2012). The latter also validated the model geometry and model-to-map correlation; details are listed in Table 2.

	HISN5-unliganded	HISN5-IG2
PDB ID	7OJ5	8QAV
EMDB	EMD-12938	EMD-18305
Magnification (×)	96 000	96 000
Voltage (kV)	300	300
Electron exposure (e–/Å^2)	40	40
Defocus range (µm)	−3.0 to −0.9	−3.0 to −0.9
Pixel size (Å)	0.86	0.86
Initial particle images (no.)	456759	661163
Final particle images (no.)	229366	605614
Resolution (gold-standard, Å)	2.25	2.23
FSC threshold	0.143	0.143
Map resolution range (Å)	2.10-3.25	1.95-3.50
B-factor for map sharpening	−110	113
Composition:		
Atoms	35587	35883
Protein residues	4440	4416
Water	931	1131
R. m. s. deviations:		
Bond (Å) (# > 4σ)	0.008 (0)	0.008 (0)
Angles (°)(# > 4σ)	1.004 (48)	1.188 (0)
Ramachandran plot (%):		
Outliers	0.55	0.55
Allowed	8.20	7.14
Favored	91.26	92.31
Rotamer outliers (%)	2.58	3.25
Cβ outliers (%)	0.00	0.00
Mean ADP (B-factors)		
Protein	30.58	19.58
Ligand	33.28	37.63
Water	26.70	20.45

Virtual screening

All docking experiments were performed in *AutoDock Vina* (Trott and Olson, 2010) with the exhaustiveness = 8, using Python scripts to automate and parallelize the work. The receptor files were prepared with the *UCSF Chimera DockPrep* tool (Pettersen et al., 2004). VS in the active site was run using the library of lead-like molecules (3,344,603 in-stock compounds; 300-350 Da, logP ≤ 3.5) downloaded from the ZINC15 database (Sterling and Irwin, 2015) in December 2021. The structure in the *I*4 space group (1.69 Å resolution) was used as the receptor. The search box was centered

at -2, 44, 63 Å (x, y, z) and measured 31 × 23 × 27 Å. To propose potential binders in the cleft conserved in plants but variable in other kingdoms, we selected a subset from the lead-like library in the ZINC15 database (Sterling and Irwin, 2015) containing more soluble molecules of logP ≤2 (1,355,624 docking-ready files downloaded in March 2022). The unliganded cryoEM structure (PDB ID: 7OJ5) was used as the receptor. The search box with the dimensions of 20 × 20 × 23 Å was centered at 114, 115, 155 Å (x, y, z). All results were scored based on the calculated binding energy gain.

Other software

Multiple sequence alignment and the analysis of residue variability were performed using the *ConSurf Server* (Ashkenazy et al., 2016). The multiple sequence alignment for both high (> 95%) and low (> 35%) sequence identity for homologs was built using the MAFFT algorithm. The homologs were collected from UNIREF90 by the CS-BLAST search algorithm. The calculations resulted in 54 unique sequences of high percent identity and 150 unique sequences of low identity. The conservation scores were assigned using the Bayesian method of calculation and the best-fit model of substitution for proteins.

The phylogenetic tree was constructed using BLAST (Altschul et al., 1997) pairwise alignments. Distances were calculated using Kimura's method (Kimura, 1983), and the tree was built using the Neighbor-Joining method (Saitou and Nei, 1987). BLASTP search using *Mt*HISN5-Nt70 sequence as a query, resulted in 477 Swissprot records with sequence identity between 30 and 90%. Sequences with the lowest percent identity were chosen as an outgroup root (*Thermotoga* sp.). The tree was visualized in *IcyTree* (Vaughan, 2017).

Molecular tunnel analysis was conducted using *CAVER Analyst 2.0* (Jurcik et al., 2018) and *CAVER 3.0.3* (Chovancova et al., 2012) plugin for *PyMOL 2.4.0* software. To detect the molecular tunnels, we used the following starting point coordinates 129.834, 136.402, 130.468 (x, y, z), probe radius 1.4, shell radius 6.0, shell depth 4, clustering threshold 3.5, frame weighting coefficient = 1.0, frame clustering threshold = 1.0 and number of iterations = 12. Received tunnels and their distance from water molecules within active sites were depicted and calculated using *UCSF Chimera 1.15* (Pettersen et al., 2004) and *PyMOL* (Schrodinger).

All protein models and structural alignment using the Needleman – Wunsch algorithm and the BLOSUM-62 matrix were visualized using *UCSF Chimera 1.15* (Pettersen et al., 2004) and *UCSF ChimeraX 1.5* (Pettersen et al., 2021).

Enzymatic synthesis of IGP

Enzymatic synthesis required *in vitro* reconstitution of five steps of the HBP leading to the formation of *2R,3S*-IGP. The reaction was conducted in a total volume of 2.75 mL at 295 K and absorbance was measured at λ = 290 nm. All enzymes used in the synthesis were His-tagged at the N-termini. All steps were conducted in the kinetic buffer (Tris-HCl 50 mM, pH 8.0; MgCl$_2$ 4 mM; KCl 100 mM; NaCl 50 mM,

TCEP 1 mM). The first step required 1 μM MtHISN1, 40 μM E. coli pyrophosphatase, and 2 mM ATP. The mixture was blanked and the reaction was initiated using 2 mM PRPP. The reaction was conducted until a plateau was reached (approximately 50 min, absorbance reached 1.1, Supplementary Figure S4A). The mixture was blanked once again, and 1 μM MtHISN2 was added. After 20 min when A = 2.0 (Supplementary Figure S4B), the mixture was blanked and 1 μM MtHISN3, 5 μM MtHISN4, and 5 mM L-glutamine were added simultaneously to start the last steps of enzymatic synthesis. After approximately 150 s, a decrease in absorbance was observed at 300 nm, indicating that PR-FAR was converted by MtHISN4 into 2R,3S-IGP (Supplementary Figure S4C). The mixture was incubated for 5 minutes on ice with 300 μL of Ni-NTA resin (GE Healthcare) and centrifuged at 2000 × g for 3 minutes. Notably, we tried purification using membrane filters to remove enzymes and potential intermediates, but we did not detect eIGP in the flowthrough, suggesting that eIGP was captured by the membrane. The resin was pre-equilibrated in binding buffer (Tris-HCl 50 mM, pH 8.0; NaCl 500 mM, imidazole 20 mM, TCEP 1 mM, 10% glycerol). The supernatant containing IGP was aspirated, transferred into a fresh Eppendorf tube, and kept on ice. The concentration of 2R,3S-IGP was assessed using a glutamate assay kit (Sigma Aldrich). Glutamate was formed from glutamine by MtHISN4, therefore its concentration of 1.6 mM stoichiometrically corresponded to a concentration of 2R,3S-IGP, indicating 80% yield of eIGP synthesis.

Isothermal titration calorimetry

The kinetics of MtHISN5 were monitored using microcalorimetry (MicroCal PEAQ-ITC and MicroCal iTC200, Malvern). ITC-SIM has been used for this purpose (Wang et al., 2020). The reaction was conducted in SEC buffer (Tris-HCl 40 mM pH 7.8, NaCl 40 mM, MnCl₂ 4 mM, EDTA 0.4 mM) at 30°C. The differential power (DP) of all ITC-SIM experiments was set to 10 μcal/mol and the stirring speed at 650 rpm. For experiments with cIGP (Santa Cruz Biotechnology, sc-218019), the protein in the cell was maintained at a low nanomolar concentration (5-10 nM), and the substrate in the syringe was added in one injection to the final concentration varying from 24 to 235 μM. Measurements were stopped when the baseline returned to the initial state, signaling substrate depletion. For one of the tested setups of the ITC-SIM experiment, a blank experiment was performed consisting of one injection of 30 μL of 0.5 mM cIGP into the cell containing the buffer, and the observation time was set to the maximum for the apparatus (10 000 s) to check the possibility of heat effect derived from the non-enzymatic substrate degradation. Additionally, an inverted ITC-SIM system was introduced to achieve a higher cIGP concentration (by circumventing the problem of the high heat of substrate dilution), where the substrate (in the cell) was kept at 0.8- and 2-mM concentrations and concentrated enzyme in the syringe (capped with 3 μL of the buffer to prevent the early leakage of the enzyme) was injected in 1 portion into the cell at a final concentration of 12 and 15 nM, respectively.

For the ITC-SIM experiments on eIGP, the substrate was concentrated using a SpeedVac. We maintained the concentration of eIGP in the cell in the range of 1.9-3.4 mM and injected the

concentrated MtHISN5 to its final concentration of 792-901 nM (a 3.8 μL aliquot of 42-48 μM enzyme, subunit concentration). Measurements were stopped ~10 min after the baseline returned to near the initial state, signaling total substrate depletion. The DP baseline was analyzed after the initial DP drop (coming from the buffer mismatch and initial protein-substrate interaction) and fitted to the 'Enzyme Kinetics – Single Injection' model within the MicroCal PEAQ-ITC analysis software. Briefly, the raw data were transformed into reaction rates and IGP concentrations and fitted to the Michaelis-Menten equation. The final kinetic parameters were calculated by averaging the values obtained from the three separate experiments.

Data availability statement

Author contributions

WW: Data curation, Investigation, Methodology, Visualization, Writing – original draft. JS: Conceptualization, Investigation, Methodology, Writing – original draft. MRa: Methodology, Writing – original draft. MRu: Conceptualization, Funding acquisition, Methodology, Project administration, Supervision, Writing – review & editing.

Funding

Acknowledgments

Publisher's note

electron microscope, where the measurements were performed. We acknowledge the MCB Structural Biology Core Facility (supported by the TEAM TECH CORE FACILITY/2017-4/6 grant from the Foundation for Polish Science) for valuable support. The diffraction data were collected at beamline P13 operated by EMBL Hamburg at the PETRA III storage ring (DESY, Hamburg, Germany). The plasmids for the production of enzymes necessary for eIGP synthesis will be made available upon request.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2024.1343980/full#supplementary-material

References

Adams, E. (1954). The enzymatic synthesis of histidine from histidinol. *J. Biol. Chem.* 209, 829–846. doi: 10.1016/S0021-9258(18)65512-7

Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., et al. (2012). Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D Biol. Crystallogr.* 68, 352–367. doi: 10.1107/S0907444912001308

Afonine, P. V., Poon, B. K., Read, R. J., Sobolev, O. V., Terwilliger, T. C., Urzhumtsev, A., et al. (2018). Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Cryst. D* 74, 531–544. doi: 10.1107/S2059798318006551

Ahangar, M. S., Vyas, R., Nasir, N., and Biswal, B. K. (2013). Structures of native, substrate-bound and inhibited forms of Mycobacterium tuberculosis imidazoleglycerol-phosphate dehydratase. *Acta Crystallogr. D Biol. Crystallogr.* 69, 2461–2467. doi: 10.1107/S0907444913022579

Akashi, H., and Gojobori, T. (2002). Metabolic efficiency and amino acid composition in the proteomes of Escherichia coli and Bacillus subtilis. *PNAS* 99, 3695–3700. doi: 10.1073/pnas.062526999

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J. H., Zhang, Z., Miller, W., et al. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402. doi: 10.1093/nar/25.17.3389

Ames, B. N. (1957a). The biosynthesis of histidine - D-erythro-imidazoleglycerol phosphate dehydrase. *J. Biol. Chem.* 228, 131–143. doi: 10.1016/S0021-9258(18)70696-0

Ames, B. N. (1957b). The biosynthesis of histidine - L-histidinol phosphate phosphatase. *J. Biol. Chem.* 226, 583–593. doi: 10.1016/S0021-9258(18)70840-5

Ames, B. N., and Mitchell, H. K. (1955). The biosynthesis of histidine - imidazoleglycerol phosphate, imidazoleacetol phosphate, and histidinol phosphate. *J. Biol. Chem.* 212, 687–696. doi: 10.1016/S0021-9258(18)71007-7

Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T., et al. (2016). ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res.* 44, W344–W350. doi: 10.1093/nar/gkw408

Beckie, H. J., Busi, R., Lopez-Ruiz, F. J., and Umina, P. A. (2021). Herbicide resistance management strategies: how do they compare with those for insecticides, fungicides and antibiotics? *Pest Manage. Sci.* 77, 3049–3056. doi: 10.1002/ps.6395

Benson, D. A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., et al. (2013). GenBank. *Nucleic Acids Res.* 41, D36–D42. doi: 10.1093/nar/gks1195

Bisson, C., Britton, K. L., Sedelnikova, S. E., Rodgers, H. F., Eadsforth, T. C., Viner, R. C., et al. (2015). Crystal structures reveal that the reaction mechanism of imidazoleglycerol-phosphate dehydratase is controlled by switching mn(II) coordination. *Structure* 23, 1236–1245. doi: 10.1016/j.str.2015.05.012

Bisson, C., Britton, K. L., Sedelnikova, S. E., Rodgers, H. F., Eadsforth, T. C., Viner, R. C., et al. (2016). Mirror-image packing provides a molecular basis for the nanomolar equipotency of enantiomers of an experimental herbicide. *Angew Chem. Int. Edit* 55, 13485–13489. doi: 10.1002/anie.201607185

Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* 72, 248–254. doi: 10.1016/0003-2697(76)90527-3

Brenner, M., and Ames, B. N. (1971). "The histidine operon and its regulation," in *Metabolic Pathways*, 3 ed. Ed. D. M. G. (Academic Press, New York).

Brilli, M., and Fani, R. (2004). Molecular evolution of *hisB* genes. *J. Mol. Evol.* 58, 225–237. doi: 10.1007/s00239-003-2547-x

Burks, D., Azad, R., Wen, J. Q., and Dickstein, R. (2018). The *medicago truncatula* genome: Genomic data availability. *Funct. Genomics Medicago Truncatula: Methods Protoc.* 1822, 39–59. doi: 10.1007/978-1-4939-8633-0_3

Cantin, N. E., Negri, A. P., and Willis, B. L. (2007). Photoinhibition from chronic herbicide exposure reduces reproductive output of reef-building corals. *Mar. Ecol. Prog. Ser.* 344, 81–93. doi: 10.3354/meps07059

Chang, A., Jeske, L., Ulbrich, S., Hofmann, J., Koblitz, J., Schomburg, I., et al. (2021). BRENDA, the ELIXIR core data resource in 2021: new developments and updates. *Nucleic Acids Res.* 49, D498–D508. doi: 10.1093/nar/gkaa1025

Chovancova, E., Pavelka, A., Benes, P., Strnad, O., Brezovsky, J., Kozlikova, B., et al. (2012). CAVER 3.0: a tool for the analysis of transport pathways in dynamic protein structures. *PLoS Comput. Biol.* 8, e1002708. doi: 10.1371/journal.pcbi.1002708

Del Duca, S., Chioccioli, S., Vassallo, A., Castronovo, L. M., and Fani, R. (2020). The role of gene elongation in the evolution of histidine biosynthetic genes. *Microorganisms* 8, ARTN 732. doi: 10.3390/microorganisms8050732

Emanuelsson, O., Nielsen, H., Brunak, S., and Von Heijne, G. (2000). Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J. Mol. Biol.* 300, 1005–1016. doi: 10.1006/jmbi.2000.3903

Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010). Features and development of coot. *Acta Crystallogr. D Biol. Crystallogr.* 66, 486–501. doi: 10.1107/S0907444910007493

Farina, W. M., Balbuena, M. S., Herbert, L. T., Mengoni Gonalons, C., and Vazquez, D. E. (2019). Effects of the herbicide glyphosate on honey bee sensory and cognitive abilities: Individual impairments with implications for the hive. *Insects* 10, 354. doi: 10.3390/insects10100354

Fazio, V. J., Peat, T. S., and Newman, J. (2014). A drunken search in crystallization space. *Acta Cryst. F* 70, 1303–1311. doi: 10.1107/S2053230X1401841X

Frumkin, I., Schirman, D., Rotman, A., Li, F., Zahavi, L., Mordret, E., et al. (2017). Gene architectures that minimize cost of gene expression. *Mol. Cell* 65, 142–153. doi: 10.1016/j.molcel.2016.11.007

Fujimori, K., and Ohta, D. (1998). An Arabidopsis cDNA encoding a bifunctional glutamine amidotransferase/cyclase suppresses the histidine auxotrophy of a Saccharomyces cerevisiae his7 mutant. *FEBS Lett.* 428, 229–234. doi: 10.1016/S0014-5793(98)00535-3

Fukushima, T., and Kodama, Y. (2022). Selection of a histidine auxotrophicMarchantia polymorpha strain with an auxotrophic selective marker. *Plant Biotechnol.* 39, 345–354. doi: 10.5511/plantbiotechnology.22.0810a

Furukawa, A., Oikawa, S., Harada, K., Sugiyama, H., Hiraku, Y., Murata, M., et al. (2010). Oxidatively generated DNA damage induced by 3-amino-5-mercapto-1,2,4-triazole, a metabolite of carcinogenic amitrole. *Mutat. Res.* 694, 7–12. doi: 10.1016/j.mrfmmm.2010.08.004

Gaines, T. A., Busi, R., and Kupper, A. (2021). Can new herbicide discovery allow weed management to outpace resistance evolution? *Pest Manage. Sci.* 77, 3036–3041. doi: 10.1002/ps.6457

Gantt, J. S., Baldauf, S. L., Calie, P. J., Weeden, N. F., and Palmer, J. D. (1991). Transfer of rpl22 to the nucleus greatly preceded its loss from the chloroplast and involved the gain of an intron. *EMBO J.* 10, 3073–3078. doi: 10.1002/embj.1991.10.issue-10

Glynn, S. E., Baker, P. J., Sedelnikova, S. E., Davies, C. L., Eadsforth, T. C., Levy, C. W., et al. (2005a). Structure and mechanism of imidazoleglycerol-phosphate dehydratase. *Structure* 13, 1809–1817. doi: 10.1016/j.str.2005.08.012

Glynn, S. E., Baker, P. J., Sedelnikova, S. E., Levy, C. W., Rodgers, H. F., Blank, J., et al. (2005b). Purification, crystallization and preliminary crystallographic analysis of Arabidopsis thaliana imidazoleglycerol-phosphate dehydratase. *Acta Cryst. F* 61, 776–778. doi: 10.1107/S1744309105022451

Gorrec, F. (2009). The MORPHEUS protein crystallization screen. *J. Appl. Crystallogr.* 42, 1035–1042. doi: 10.1107/S0021889809042022

Gould, F., Brown, Z. S., and Kuzma, J. (2018). Wicked evolution: Can we address the sociobiological dilemma of pesticide resistance? *Science* 360, 728–732. doi: 10.1126/science.aar3780

Hall, C. J., Mackie, E. R., Gendall, A. R., Perugini, M. A., and Soares Da Costa, T. P. (2020). Review: Amino acid biosynthesis as a target for herbicide development. *Pest Manag Sci.* 76, 3896–3904. doi: 10.1002/ps.5943

Hawkes, T. R., Thomas, P. G., Edwards, L. S., Rayner, S. J., Wilkinson, K. W., and Rice, D. W. (1995). Purification and characterization of the imidazoleglycerol-phosphate dehydratase of saccharomyces-cerevisiae from recombinant escherichia-coli. *Biochem. J.* 306, 385–397. doi: 10.1042/bj3060385

Hilton, J. L., Kearney, P. C., and Ames, B. N. (1965). Mode of action of herbicide 3-amino-1,2,4-triazole(Amitrole) - inhibition of an enzyme of histidine biosynthesis. *Arch. Biochem. Biophys.* 112, 544–&. doi: 10.1016/0003-9861(65)90093-7

Hrbáčková, M., Dvořák, P., Takáč, T., Tichá, M., Luptovciak, I., Samajová, O., et al. (2020). Biotechnological perspectives of omics and genetic engineering methods in alfalfa. *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.00592

Ingle, R. A. (2011). Histidine biosynthesis. *Arabidopsis Book* 9, e0141. doi: 10.1199/tab.0141

Jurcik, A., Bednar, D., Byska, J., Marques, S. M., Furmanova, K., Daniel, L., et al. (2018). CAVER Analyst 2.0: analysis and visualization of channels and tunnels in protein structures and molecular dynamics trajectories. *Bioinformatics* 34, 3586–3588. doi: 10.1093/bioinformatics/bty386

Kabsch, W. (2010). Xds. *Acta Crystallographica Section D-Biological Crystallogr.* 66, 125–132. doi: 10.1107/S0907444909047337

Kepp, K. P. (2020). Survival of the cheapest: how proteome cost minimization drives evolution. *Q Rev. Biophys.* 53, e7. doi: 10.1017/S0033583520000037

Kim, Y., Babnigg, G., Jedrzejczak, R., Eschenfeldt, W. H., Li, H., Maltseva, N., et al. (2011). High-throughput protein purification and quality assessment for crystallization. *Methods* 55, 12–28. doi: 10.1016/j.ymeth.2011.07.010

Kimura, M. (1983). *The Neutral Theory of Molecular Evolution* (Cambridge: Cambridge University Press). doi: 10.1017/CBO9780511623486

Kinateder, T., Drexler, L., Straub, K., Merkl, R., and Sterner, R. (2023). Experimental and computational analysis of the ancestry of an evolutionary young enzyme from histidine biosynthesis. *Protein Sci.* 32, e4536. doi: 10.1002/pro.4536

Kinateder, T., Mayer, C., Nazet, J., and Sterner, R. (2024). Improving enzyme functional annotation by integrating in *vitro* and in silico approaches: The example of histidinol phosphate phosphatases. *Protein Sci.* 33, e4899. doi: 10.1002/pro.4899

Kirsch, P., Hartman, A. M., Hirsch, A. K. H., and Empting, M. (2019). Concepts and core principles of fragment-based drug design. *Molecules* 24, 4309. doi: 10.3390/molecules24234309

Kishore, G. M., and Shah, D. M. (1988). Amino acid biosynthesis inhibitors as herbicides. *Annu. Rev. Biochem.* 57, 627–663. doi: 10.1146/annurev.bi.57.070188.003211

Klock, H. E., Koesema, E. J., Knuth, M. W., and Lesley, S. A. (2008). Combining the polymerase incomplete primer extension method for cloning and mutagenesis with microscreening to accelerate structural genomics efforts. *Proteins* 71, 982–994. doi: 10.1002/prot.21786

Klopotowski, T., and Wiater, A. (1965). Synergism of aminotriazole and phosphate on inhibition of yeast imidazole glycerol phosphate dehydratase. *Arch. Biochem. Biophys.* 112, 562–566. doi: 10.1016/0003-9861(65)90096-2

Kowiel, M., Jaskolski, M., and Dauter, Z. (2014). ACHESYM: an algorithm and server for standardized placement of macromolecular models in the unit cell. *Acta Crystallographica Section D-Biological Crystallogr.* 70, 3290–3298. doi: 10.1107/S1399004714024572

Kress, N., Halder, J. M., Rapp, L. R., and Hauer, B. (2018). Unlocked potential of dynamic elements in protein structures: channels and loops. *Curr. Opin. Chem. Biol.* 47, 109–116. doi: 10.1016/j.cbpa.2018.09.010

Kumar, D., Jha, B., Bhatia, I., Ashraf, A., Dwivedy, A., and Biswal, B. K. (2022). Characterization of a triazole scaffold compound as an inhibitor of imidazoleglycerol-phosphate dehydratase. *Proteins* 90, 3–17. doi: 10.1002/prot.26181

Mano, J., Hatano, M., Koizumi, S., Tada, S., Hashimoto, M., and Scheidegger, A. (1993). Purification and properties of a monofunctional imidazoleglycerol-phosphate dehydratase from wheat. *Plant Physiol.* 103, 733–739. doi: 10.1104/pp.103.3.733

Martin, R. G., Berberich, M. A., Ames, B. N., Davis, W. W., Goldberger, R. F., and Yourno, J. D. (1971). "Enzymes and intermediates of histidine biosynthesis in Salmonella typhimurium," in *Methods in Enzymology: Metabolism of Amino Acids and Amines Part B.* Eds. C. W. Tabor and H. Tabor (Academic Press, New York), 3–44.

Miller, L. L., and Bale, W. F. (1952). Factors affecting biosynthesis of plasma albumin and globulin fractions as studied with the aid of lysine-epsilon and histidine-2-C-14. *Fed Proc.* 11, 260–260.

Moriarty, N. W., Grosse-Kunstleve, R. W., and Adams, P. D. (2009). electronic Ligand Builder and Optimization Workbench (eLBOW): a tool for ligand coordinate and restraint generation. *Acta Cryst. D* 65, 1074–1080. doi: 10.1107/S0907444909029436

Motta, E. V. S., Raymann, K., and Moran, N. A. (2018). Glyphosate perturbs the gut microbiota of honey bees. *Proc. Natl. Acad. Sci. U.S.A.* 115, 10305–10310. doi: 10.1073/pnas.1803880115

Mueller-Harvey, I., Bee, G., Dohme-Meier, F., Hoste, H., Karonen, M., Kölliker, R., et al. (2019). Benefits of condensed tannins in forage legumes fed to ruminants: Importance of structure, concentration, and diet composition. *Crop Sci.* 59, 861–885. doi: 10.2135/cropsci2017.06.0369

Muralla, R., Sweeney, C., Stepansky, A., Leustek, T., and Meinke, D. (2007). Genetic dissection of histidine biosynthesis in Arabidopsis. *Plant Physiol.* 144, 890–903. doi: 10.1104/pp.107.096511

Negri, A., Vollhardt, C., Humphrey, C., Heyward, A., Jones, R., Eaglesham, G., et al. (2005). Effects of the herbicide diuron on the early life history stages of coral. *Mar. pollut. Bull.* 51, 370–383. doi: 10.1016/j.marpolbul.2004.10.053

Oren, A., and Garrity, G. M. (2021). Valid publication of the names of forty-two phyla of prokaryotes. *Int. J. Syst. Evol. Microbiol.* 71, 5056. doi: 10.1099/ijsem.0.005056

Petersen, L. N., Marineo, S., Mandala, S., Davids, F., Sewell, B. T., and Ingle, R. A. (2010). The missing link in plant histidine biosynthesis: Arabidopsis myoinositol monophosphatase-like2 encodes a functional histidinol-phosphate phosphatase. *Plant Physiol.* 152, 1186–1196. doi: 10.1104/pp.109.150805

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., et al. (2004). UCSF chimera - A visualization system for exploratory research and analysis. *J. Comput. Chem.* 25, 1605–1612. doi: 10.1002/jcc.20084

Pettersen, E. F., Goddard, T. D., Huang, C. R. C., Meng, E. E. C., Couch, G. S., Croll, T. I., et al. (2021). UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci.* 30, 70–82. doi: 10.1002/pro.3943

Punjani, A., Rubinstein, J. L., Fleet, D. J., and Brubaker, M. A. (2017). cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* 14, 290–296. doi: 10.1038/nmeth.4169

Rawson, S., Bisson, C., Hurdiss, D. L., Fazal, A., Mcphillie, M. J., Sedelnikova, S. E., et al. (2018). Elucidating the structural basis for differing enzyme inhibitor potency by cryo-EM. *PNAS* 115, 1795–1800. doi: 10.1073/pnas.1708839115

Reyes-Prieto, A., and Moustafa, A. (2012). Plastid-localized amino acid biosynthetic pathways of Plantae are predominantly composed of non-cyanobacterial enzymes. *Sci. Rep.* 2, 955. doi: 10.1038/srep00955

Rohou, A., and Grigorieff, N. (2015). CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* 192, 216–221. doi: 10.1016/j.jsb.2015.08.008

Ruszkowski, M. (2018). Guarding the gateway to histidine biosynthesis in plants: Medicago truncatula ATP-phosphoribosyltransferase in relaxed and tense states. *Biochem. J.* 475, 2681–2697. doi: 10.1042/BCJ20180289

Ruszkowski, M., and Dauter, Z. (2016). 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.* 291, 9960–9973. doi: 10.1074/jbc.M115.708727

Ruszkowski, M., and Dauter, Z. (2017). Structures of medicago truncatula L-histidinol dehydrogenase show rearrangements required for NAD(+) binding and the cofactor positioned to accept a hydride. *Sci. Rep.* 7, 10476. doi: 10.1038/s41598-017-10859-0

Rutkiewicz, M., Nogues, I., Witek, W., Angelaccio, S., Contestabile, R., and Ruszkowski, M. (2023). Insights into the substrate specificity, structure, and dynamics of plant histidinol-phosphate aminotransferase (HISN6). *Plant Physiol. Biochem.* 196, 759–773. doi: 10.1016/j.plaphy.2023.02.017

Saika, H., Fruh, T., Iwasaki, G., Koizumi, S., Mori, I., and Hayakawa, K. (1993). Synthesis of (2r,3r)-, (2s,3s)-, (2r,3s)- and (2s,3r)-imidazole glycerol phosphates (Igp) - substrates for igp-dehydratase (Igpd). *Bioorganic Medicinal Chem. Lett.* 3, 2129–2134. doi: 10.1016/S0960-894X(01)81031-3

Saitou, N., and Nei, M. (1987). The neighbor-joining method - a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4, 406–425. doi: 10.1093/oxfordjournals.molbev.a040454

Sakiroglu, M., and Ilhan, D. (2021). *Medicago sativa* species complex: Revisiting the century-old problem in the light of molecular tools. *Crop Sci.* 61, 827–838. doi: 10.1002/csc2.20316

Scheres, S. H. (2012). RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* 180, 519–530. doi: 10.1016/j.jsb.2012.09.006

Seligmann, H. (2003). Cost-minimization of amino acid usage. *J. Mol. Evol.* 56, 151–161. doi: 10.1007/s00239-002-2388-z

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. doi: 10.1101/gr.1239303

Sinha, S. C., Chaudhuri, B. N., Burgner, J. W., Yakovleva, G., Davisson, V. J., and Smith, J. L. (2004). Crystal structure of imidazole glycerol-phosphate dehydratase - Duplication of an unusual fold. *J. Biol. Chem.* 279, 15491–15498. doi: 10.1074/jbc.M312733200

Smith, P. M., and Atkins, C. A. (2002). Purine biosynthesis. Big in cell division, even bigger in nitrogen assimilation. *Plant Physiol.* 128, 793–802. doi: 10.1104/pp.010912

Stepansky, A., and Leustek, T. (2006). Histidine biosynthesis in plants. *Amino Acids* 30, 127–142. doi: 10.1007/s00726-005-0247-0

Sterling, T., and Irwin, J. J. (2015). ZINC 15–Ligand discovery for everyone. *J. Chem. Inf Model.* 55, 2324–2337. doi: 10.1021/acs.jcim.5b00559

Tada, S., Hatano, M., Nakayama, Y., Volrath, S., Guyer, D., Ward, E., et al. (1995). Insect-cell expression of recombinant imidazoleglycerolphosphate dehydratase of arabidopsis and wheat and inhibition by triazole herbicides. *Plant Physiol.* 109, 153–159. doi: 10.1104/pp.109.1.153

Terwilliger, T. C., Grosse-Kunstleve, R. W., Afonine, P. V., Moriarty, N. W., Zwart, P. H., Hung, L. W., et al. (2008). Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta Cryst. D* 64, 61–69. doi: 10.1107/S090744490705024X

Trott, O., and Olson, A. J. (2010). AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* 31, 455–461. doi: 10.1002/jcc.21334

Tyohemba, R. L., Humphries, M. S., Schleyer, M. H., and Porter, S. N. (2022). Accumulation of commonly used agricultural herbicides in coral reef organisms from iSimangaliso Wetland Park, South Africa*. *Environ. pollut.* 294, 118665. doi: 10.1016/j.envpol.2021.118665

Vaughan, T. G. (2017). IcyTree: rapid browser-based visualization for phylogenetic trees and networks. *Bioinformatics* 33, 2392–2394. doi: 10.1093/bioinformatics/btx155

Vázquez, D. E., Balbuena, M. S., Chaves, F., Gora, J., Menzel, R., and Farina, W. M. (2020). Sleep in honey bees is affected by the herbicide glyphosate. *Sci. Rep.* 10, 10516. doi: 10.1038/s41598-020-67477-6

Wang, L., Liu, R. Y., Li, F., Meng, Y., and Lu, H. Z. (2021b). Unveiling the novel characteristics of IGPD polymer and inhibitors binding affinities using 12-6-4 LJ-type nonbonded Mn2+ model. *J. Mol. Liq* 322, 114992. doi: 10.1016/j.molliq.2020.114992

Wang, L., Liu, R., Meng, Y., Li, F., and Lu, H. (2021a). Structure and function of the refined C-terminal loop in imidazole glycerol phosphate dehydratase from different homologs. *J. Agric. Food Chem.* 69, 13871–13880. doi: 10.1021/acs.jafc.1c04282

Wang, Y., Wang, G., Moitessier, N., and Mittermaier, A. K. (2020). Enzyme kinetics by isothermal titration calorimetry: Allostery, inhibition, and dynamics. *Front. Mol. Biosci.* 7. doi: 10.3389/fmolb.2020.583826

Wiater, A., Hulanicka, D., and Klopotowski, T. (1971a). Structural requirements for inhibition of yeast imidazoleglycerol phosphate dehydratase by triazole and anion inhibitors. *Acta Biochim. Pol.* 18, 289–297.

Wiater, A., Klopotowski, T., and Bagdasarian, G. (1971b). Synergistic inhibition of plant imidazoleglycerol phosphate dehydratase by aminotriazole and phosphate. *Acta Biochim. Pol.* 18, 309–314.

Wiater, A., Krajewska-Grynkiewicz, K., and Klopotowski, T. (1971c). Histidine biosynthesis and its regulation in higher plants. *Acta Biochim. Pol.* 18, 299–307.

Witek, W., Sliwiak, J., and Ruszkowski, M. (2021). Structural and mechanistic insights into the bifunctional HISN2 enzyme catalyzing the second and third steps of histidine biosynthesis in plants. *Sci. Rep.* 11, 9647. doi: 10.1038/s41598-021-88920-2

Witte, C. P., and Herde, M. (2020). Nucleotide metabolism in plants. *Plant Physiol.* 182, 63–78. doi: 10.1104/pp.19.00955

Wolter, F. P., Fritz, C. C., Willmitzer, L., Schell, J., and Schreier, P. H. (1988). Rbcs genes in solanum-tuberosum - conservation of transit peptide and exon shuffling during evolution. *PNAS* 85, 846–850. doi: 10.1073/pnas.85.3.846

Wood, C., Burnley, T., Patwardhan, A., Scheres, S., Topf, M., Roseman, A., et al. (2015). Collaborative computational project for electron cryo-microscopy. *Acta Cryst. D* 71, 123–126. doi: 10.1107/S1399004714018070

Young, N. D., Debellé, F., Oldroyd, G. E. D., Geurts, R., Cannon, S. B., Udvardi, M. K., et al. (2011). The *Medicago* genome provides insight into the evolution of rhizobial symbioses. *Nature* 480, 520–524. doi: 10.1038/nature10625

Zallot, R., Oberg, N., and Gerlt, J. A. (2019). The EFI web resource for genomic enzymology tools: Leveraging protein, genome, and metagenome databases to discover novel enzymes and metabolic pathways. *Biochemistry* 58, 4169–4182. doi: 10.1021/acs.biochem.9b00735

Zerbe, B. S., Hall, D. R., Vajda, S., Whitty, A., and Kozakov, D. (2012). Relationship between hot spot residues and ligand binding hot spots in protein-protein interfaces. *J. Chem. Inf Model.* 52, 2236–2244. doi: 10.1021/ci300175u

SUPPLEMENTARY MATERIAL

Targeting imidazole-glycerol phosphate dehydratase in plants: novel approach for structural and functional studies, and inhibitor blueprinting

Wojciech Witek[1], Joanna Sliwiak[1], Michal Rawski[2], Milosz Ruszkowski[1*]

[1]Department of Structural Biology of Eukaryotes, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznan, Poland

[2]SOLARIS National Synchrotron Radiation Centre, Krakow, Poland

*Correspondence:
Milosz Ruszkowski
Department of Structural Biology of Eukaryotes,
Institute of Bioorganic Chemistry,
Polish Academy of Sciences,
Noskowskiego 12/14,
Poznan 61-704, Poland
E-mail: mruszkowski@ibch.poznan.pl

Supplementary Figure S1. Electron density maps
Supplementary Figure S2. Phylogenetic tree.
Supplementary Figure S3. Highest-scoring results of the virtual screening in the cleft between two active sites.
Supplementary Figure S4. Spectral absorption plots for imidazole-glycerol phosphate (IGP) enzymatic synthesis
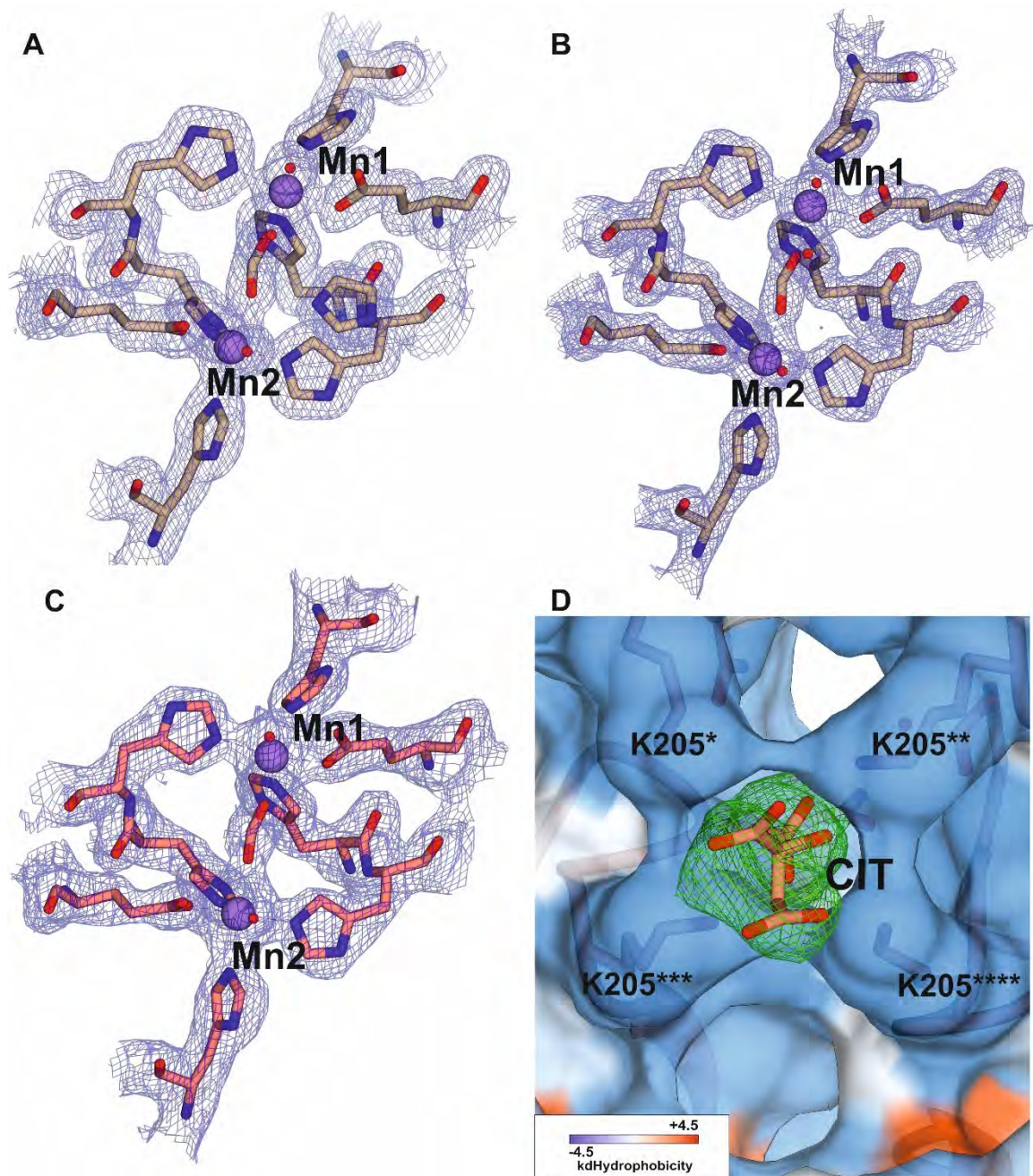
Supplementary Figure S1.
Electron density maps. Panels A-C show $2F_o$-F_c electron density maps around active sites in crystal structures of *Mt*HISN5: 1.55 Å, PDB ID: 8QAW ($\sigma = 1.0$); 1.69 Å, 8QAX ($\sigma = 1.7$); and 2.20 Å, 8QAY ($\sigma = 1.0$), respectively. Panel D represents a polder map around a citrate molecule bound in 8QAY ($\sigma = 4.6$). Citrate is bound at the tunnel along a non-crystallographic four-fold symmetry axis and is stabilized by side chains of Lys205 and its counterparts from other chains which are indicated by asterisks. *Mt*HISN5 surface is 20% transparent and coloring is based on hydrophobicity.
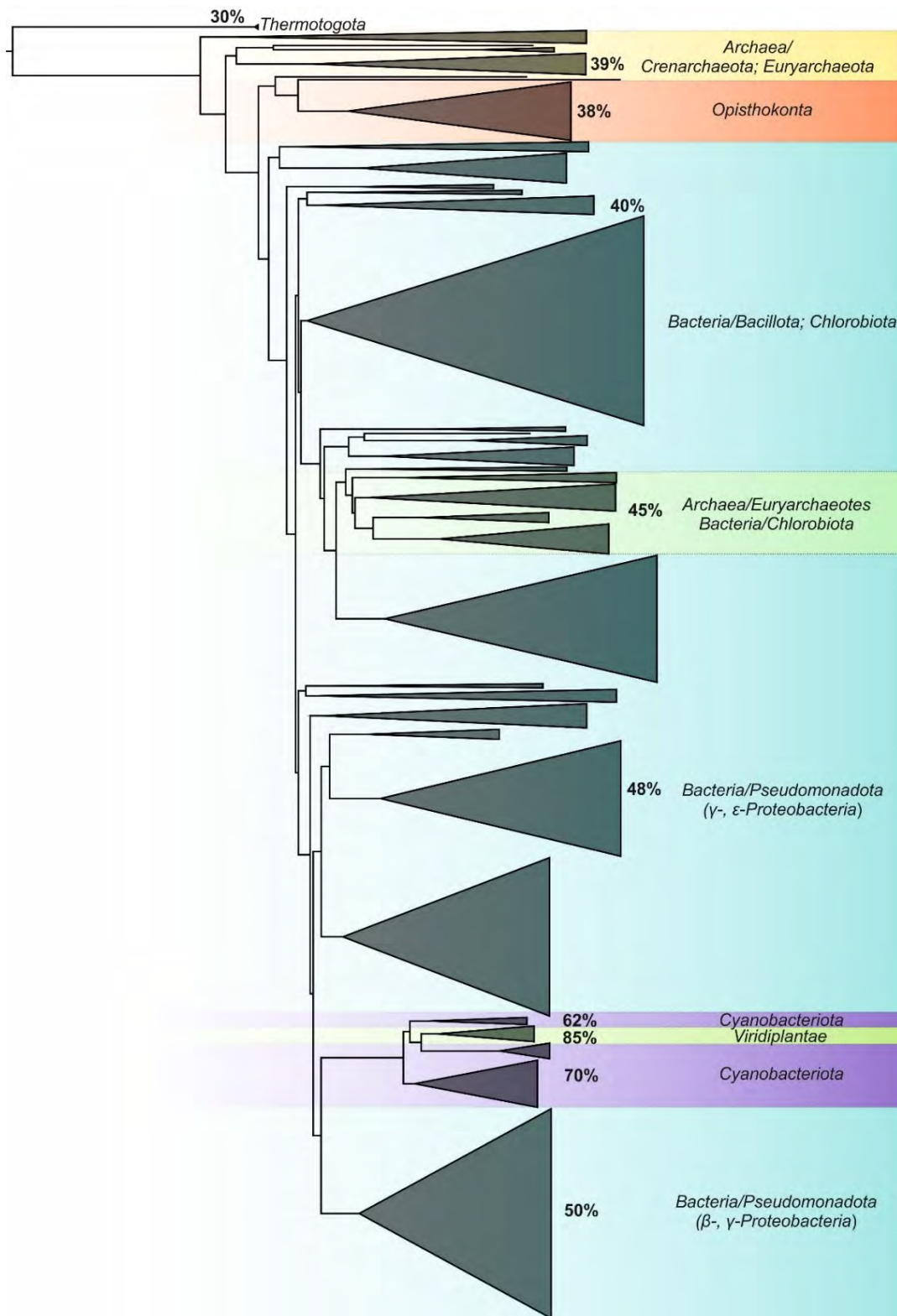
Supplementary Figure S2.

Outgroup rooted (*Termotogota*) phylogenetic tree of *Mt*HISN5 homologs. The tree shows close relationship of cyanobacterial and plant (*Viridiplantae*) sequences. Other eukaryotic sequences, i.e. *Opisthokonta*, are closer to *Archaea* and *Bacillota/Chlorobiota*. Total number of 478 protein sequences were analyzed and grouped to enhance data clarity. Percent values indicate amino acid sequence identity to the sequence of *Mt*HISN5.
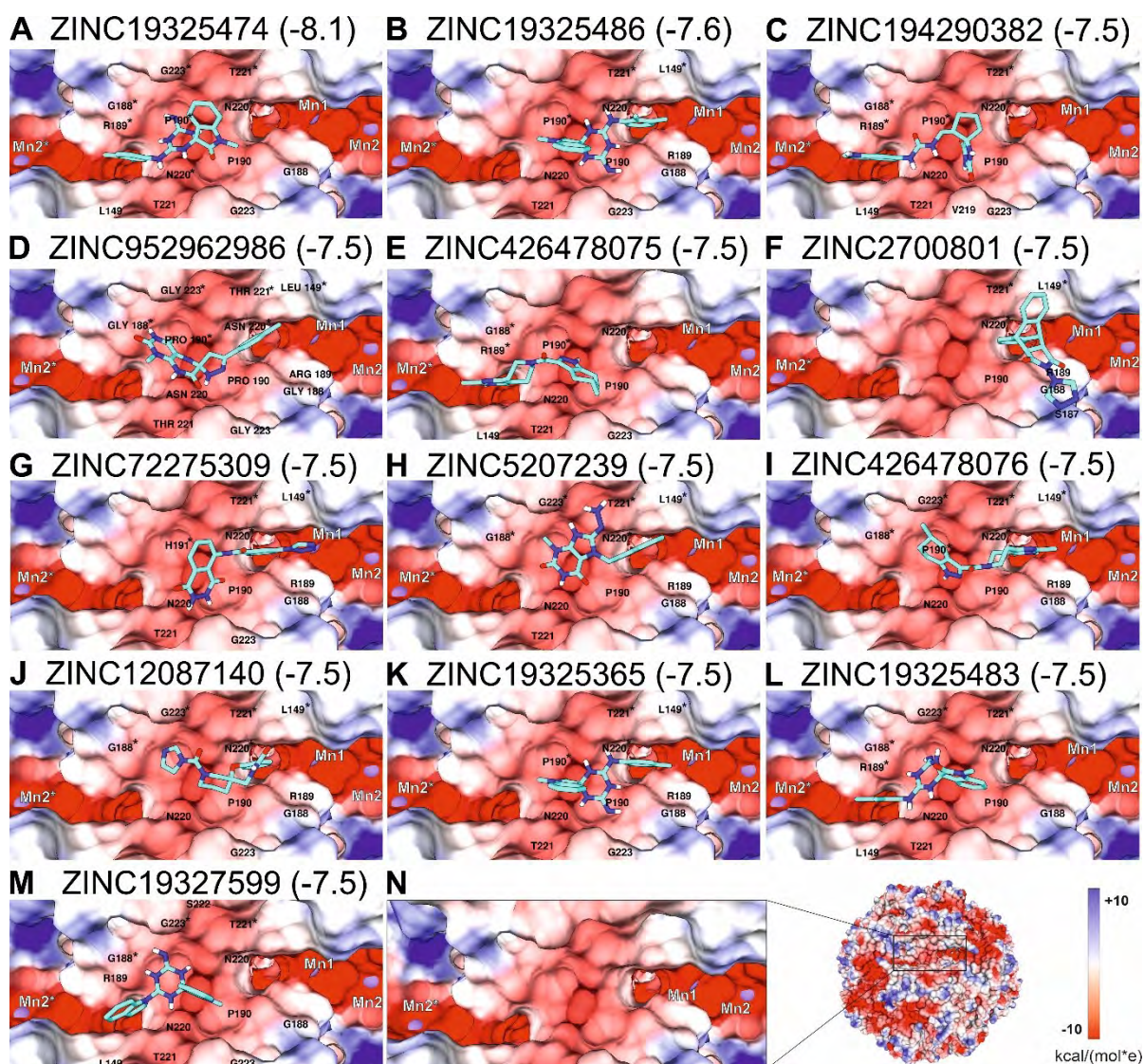
Supplementary Figure S3.

Highest-scoring results of the virtual screening calculated for the cleft between two active sites of adjacent *Mt*HISN5 subunits. (A-M) ZINC IDs are given together with the estimated energy gain (in parentheses, kcal/mol). (N) Fragment of the *Mt*HISN5 surface. *Mt*HISN5 surface is colored by coulombic potential according to the color key (bottom-right corner). The active sites are marked by Mn1, Mn2 and Mn2* from the adjacent subunit.
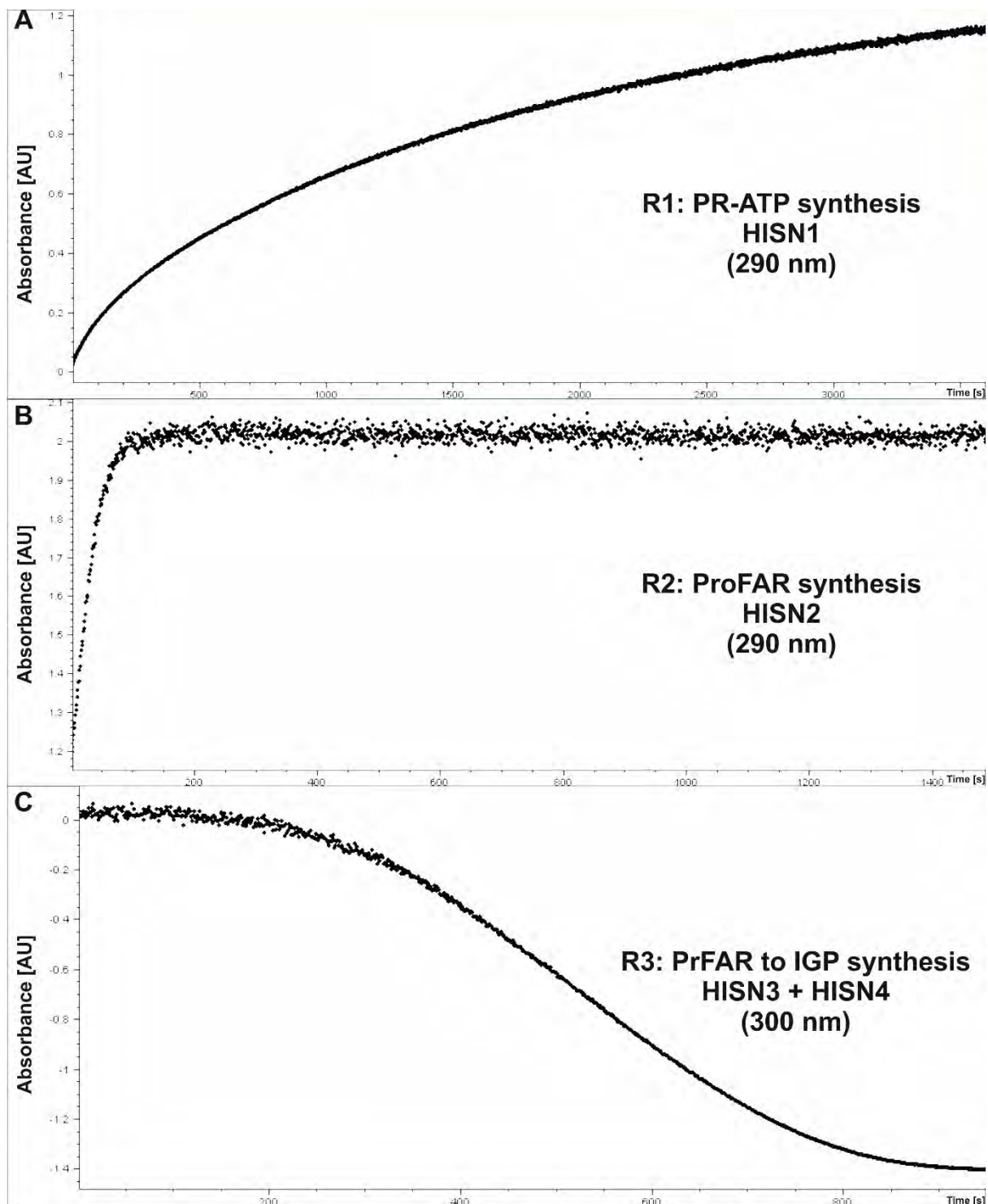
Supplementary Figure S4. Absorption spectra for IGP synthesis. Panel A shows absorption increase at 290 nm owed to production of PR-ATP. Panel B represents PR-ATP conversion to ProFAR, monitored at 290 nm. Panel C shows ProFAR isomerization to PrFAR by HISN3 and immediate PrFAR conversion to IGP, which starts as an absorption decrease at 300 nm after approx. 150 sec.

PPB

Check for updates

Insights into the substrate specificity, structure, and dynamics of plant histidinol-phosphate aminotransferase (HISN6)

Maria Rutkiewicz [a], Isabel Nogues [b], Wojciech Witek [a], Sebastiana Angelaccio [c], Roberto Contestabile [c], Milosz Ruszkowski [a,*]

[a] *Department of Structural Biology of Eukaryotes, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznan, Poland*
[b] *Research Institute on Terrestrial Ecosystems, National Research Council, Monterotondo Scalo, Rome, Italy*
[c] *Istituto Pasteur Italia-Fondazione Cenci Bolognetti, Dipartimento di Scienze Biochimiche "A. Rossi Fanelli", Sapienza Università di Roma, Rome, Italy*

ARTICLE INFO

ABSTRACT

Histidinol-phosphate aminotransferase is the sixth protein (hence HISN6) in the histidine biosynthetic pathway in plants. HISN6 is a pyridoxal 5′-phosphate (PLP)-dependent enzyme that catalyzes the reversible conversion of imidazole acetol phosphate into L-histidinol phosphate (HOLP). Here, we show that plant HISN6 enzymes are closely related to the orthologs from *Chloroflexota*. The studied example, HISN6 from *Medicago truncatula* (*Mt*HISN6), exhibits a surprisingly high affinity for HOLP, which is much higher than reported for bacterial homologs. Moreover, unlike the latter, *Mt*HISN6 does not transaminate phenylalanine. High-resolution crystal structures of *Mt*HISN6 in the open and closed states, as well as the complex with HOLP and the apo structure without PLP, bring new insights into the enzyme dynamics, pointing at a particular role of a string-like fragment that oscillates near the active site and participates in the HOLP binding. When *Mt*HISN6 is compared to bacterial orthologs with known structures, significant differences arise in or near the string region. The high affinity of *Mt*HISN6 appears linked to the particularly tight active site cavity. Finally, a virtual screening against a library of over 1.3 mln compounds revealed three sites in the *Mt*HISN6 structure with the potential to bind small molecules. Such compounds could be developed into herbicides inhibiting plant HISN6 enzymes absent in animals, which makes them a potential target for weed control agents.

1. Introduction

The cultivation of plant species important to agriculture depends on protection from weeds and pests. Humans have developed many ways to stop weeds from overgrowing crop fields. A modern example is herbicides, which are biologically-active chemical compounds that disturb weed growth and development through various modes of action. However, widespread herbicide usage has led to herbicide resistance (HR) in many weed species. The past six decades have brought more than 260 weed species resistant to over 160 herbicides due to their repeated use (Gould et al., 2018; Beckie et al., 2021; Gaines et al., 2021). Hence, we desperately need new herbicides to secure efficient agriculture and feed nearly eight billion people sustainably.

The most commonly used herbicide, glyphosate, inhibits 5-*enol*-pyruvyl-shikimate-3-phosphate synthase (EPSPS, EC 2.5.1.19) (Duke, 2018). EPSPS catalyzes the sixth step of the shikimate pathway that yields aromatic amino acids, *i.e.*, L-tryptophan (Trp), L-phenylalanine

(Phe), and L-tyrosine (Tyr) (Maeda and Dudareva, 2012). For a long time, glyphosate was considered environmentally friendly because soil microflora can degrade it to CO_2, ammonia, and inorganic phosphate (P_i) (Forlani et al., 1999). Unfortunately, more and more weed species have developed resistance to glyphosate. Moreover, in 2015 the World Health Organization (WHO) announced that glyphosate might be carcinogenic according to the assessment by the International Agency for Research on Cancer (IARC) (Guyton et al., 2015). Despite the WHO declaration, in 2018, eighteen countries of the European Union renewed licenses for glyphosate use for the next five years. Recent findings have also demonstrated that glyphosate negatively affects pollinators, harming their cognitive and sensory abilities, thermoregulation, and gut microbiome (Motta et al., 2018; Farina et al., 2019; Vazquez et al., 2020; Weidenmuller et al., 2022). Therefore, searching for new herbicides and targets is now a high priority that should be conducted in parallel to the research on herbicide side effects.

The inhibition of the L-histidine (hereafter histidine) biosynthetic

---

pathway (HBP) in plants is one of the promising solutions. The rationale behind targeting the HBP is threefold: (i) it occurs in prokaryotes, fungi, and plants but is absent in animals, (ii) loss-of-function mutants are lethal (Mo et al., 2006; Muralla et al., 2007), and (iii) most enzymes in the HBP are encoded by single genes (Stepansky and Leustek, 2006). The HBP is also integrated with other metabolic pathways, e.g., biosynthesis of nucleotides, other amino acids, and folates. All that suggests that inhibition of the HBP will disturb vital cellular processes directly by histidine depletion and indirectly by perturbing other metabolic pathways.

The plant HBP consists of eleven reactions catalyzed by eight enzymes named consecutively HISN1-8 that localize to the chloroplast stroma (Stepansky and Leustek, 2006). This article focuses on the HISN6 enzyme, which functionally is an L-histidinol phosphate aminotransferase (HAT) (EC 2.6.1.9). HISN6 is a pyridoxal 5′-phosphate (PLP)-dependent enzyme, classified into the Iβ subfamily (Jensen and Gu, 1996). It reversibly catalyzes the seventh step of the HBP; the interconversion of imidazole acetol phosphate (IAP) to L-histidinol phosphate (HOLP). The reaction is concomitant with the conversion of equimolar amounts of L-glutamate into 2-oxoglutarate (2OG) for the formation of pyridoxamine 5′-phosphate (PMP) (Haruyama et al., 2001; Fernandez et al., 2004) (Fig. 1).

In *Arabidopsis thaliana*, the knockout of the *AtHPA1* gene (encoding the *At*HISN6A enzyme) is embryo-lethal (Mo et al., 2006). Interestingly, the effect was not due to His starvation, but it made the plants unable to

sustain primary root growth 2 days after germination. The exact role of *AtHPA1* in His signaling mechanisms in plants is yet to be understood. However, *A. thaliana* is one of the unique plant species that has the second (HISN6B) isoform, possibly able to take over in the HBP.

This article is a part of a series that provides experimental structures of plant HBP enzymes; the structures of HISN1 (Ruszkowski, 2018), HISN2 (Witek et al., 2021), HISN7 (Ruszkowski and Dauter, 2016), and HISN8 (Ruszkowski and Dauter, 2017) were published by us, and that of HISN5 by others (Glynn et al., 2005). Until now, our comprehension of plant HISN6 enzymes has been limited to what can be deduced from structures of bacterial homologs, as the structures were unavailable for any plant species. Here, we report a study of the structure and function of HISN6 from the model legume plant *Medicago truncatula* (*Mt*HISN6). We traced the evolution of the plant HISN6 enzymes and showed that they have evolved toward higher selectivity than bacterial homologs. We also compared *Mt*HISN6 with homologous structures from prokaryotes. We analyzed the closed and open forms of the *Mt*HISN6 structure, which vary by the positioning of a dynamic segment that either restricts or provides access to the active site. Finally, we determined the potential target regions for designing novel inhibitors of plant HISN6 that could be exploited in the herbicide design.
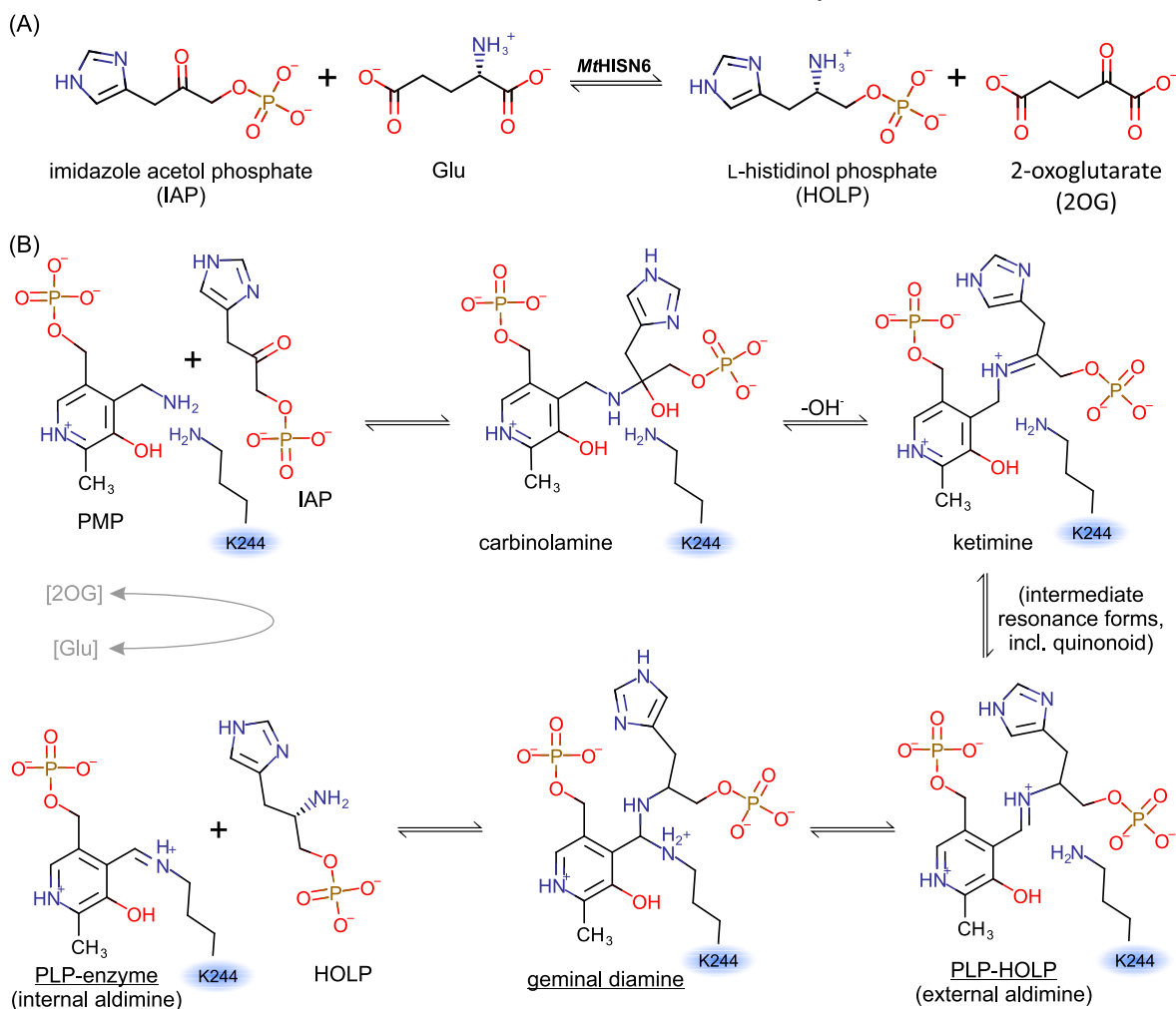
Fig. 1. (A) Simplified scheme of the reaction catalyzed by HISN6. The amino group is transferred from glutamate to imidazole acetol phosphate (IAP), creating histidinol phosphate (HOLP) and 2-oxoglutarate products. (B) The catalytic mechanism of the transamination reaction in the direction of the histidine biosynthetic pathway. Snapshots captured in this work crystal structures are underlined. The second half-reaction, *i.e.*, recharging PMP by glutamate, is indicated in gray.

2. Results

2.1. Evolution of plant HISN6 enzymes

Structural similarity network (SSN) analysis was conducted using 21,142 UniRef90 (Suzek et al., 2015) sequences from the InterPro (Blum et al., 2021) family of L-histidinol phosphate aminotransferases (IPR005861). The study revealed the scattered origin of bacterial HATs. *Archaea* share little similarity with surrounding groups suggesting that their HATs might have evolved independently (Fig. 2). Although *Opisthokonta* (in this case, primarily fungi) seems to be separated, it has a few connections with the phylum of *Bacteroidetes*. HAT enzymes from *Bacteroidetes* were likely ancestral to orthologs found in *Oomycota* (water molds) and *Bacillariophyta* (diatoms). Interestingly, phototrophic *Viridiplantae* (land plants) and *Chlorophyta* (green algae) are closely related to the bacterial *Chloroflexota* phylum (previously referred to as *Chloroflexi*) (Oren and Garrity, 2021). Unlike most chloroplast proteins that evolved from *Cyanobacteria* (Raven and Allen, 2003), a gene encoding HISN6 in plants seems to have been acquired from *Chloroflexota* by a horizontal gene transfer (HGT) independently from chloroplasts formation that occurred by absorbing *Cyanobacteria* (Yuzawa et al., 2012).

2.2. Functional assays of MtHISN6

*Mt*HISN6 orthologs have been studied in other species, but no report regarding substrate specificity is available for any eukaryotic source. IAP is the HISN6 substrate in the HBP, but it is commercially unavailable. Therefore, we assayed *Mt*HISN6 in the reverse direction (with HOLP as substrate) in a coupled assay with bovine glutamate dehydrogenase (GDH), which allows measuring NADH formation at 340 nm. Others have used a similar approach (Nasir et al., 2016). It was astounding to observe that *Mt*HISN6 exhibits a relatively high affinity towards HOLP, as reflected by a low K_M value of 29 μM (Fig. 3), which is nearly 15-fold lower than the K_M of *Mycobacterium tuberculosis* HAT (Nasir et al., 2016).

Moreover, in contrast to *M. tuberculosis* HAT, *Mt*HISN6 shows no measurable activity with Phe (up to 25 mM). We then tested whether the external aldimine forms in the presence of 1 mM Phe. The absorbance at 355 nm increased slowly, with the rate $k = 1.6$ min^{-1} (Supplementary Fig. S1). Notably, in the same experimental setup, the external aldimine formation with HOLP was too quick to measure, suggesting this step is not limiting in our coupled assay. On the other hand, *Mt*HISN6 exhibits a low catalytic activity, at least in the HOLP → IAP direction, with a turnover rate (k_{cat}) of 4 min^{-1}.

Fig. 2. Sequence similarity network of *Mt*HISN6 and its homologs. 21,142 UniRef90 clusters were analyzed. Higher plants (*Viridiplantae*) and green algae (*Chlorophyta*) are marked in green. Fungal sequences (*Opisthokonta*) are marked in red. Archaea are purple. Plant HISN6s group together with HATs from anoxygenic photoautotrophs, *Chloroflexota*.

K_M = 28.9 ± 4.2 μM
k_{cat} = 4.1 ± 0.2 min^{-1}

Fig. 3. Determination of kinetic parameters for the conversion of HOLP into IAP. The reaction was followed in a coupled assay as detailed in the text by monitoring the increase of NADH at 340 nm. Nonlinear least square analysis was used to fit the experimental data with the Michaelis-Menten equation, obtaining estimates of K_M and k_{cat}. Error bars on the average experimental points are standard deviations calculated from two independent replications.

2.3. Overall properties of the MtHISN6 structures

Crystal structures of *Mt*HISN6 were solved in the monoclinic space group, $P2_1$: (i) as a functional dimer in both open (1.57 Å resolution) and (ii) closed (1.40 Å) states and as (iii) a dimeric complex with HOLP (1.61 Å). The apo*Mt*HISN6 (without PLP, 1.45 Å) also crystallized in the $P2_1$ space group but with two dimers in the asymmetric unit (Table 1). The high-resolution electron density maps allowed us to trace residues from D25 until the C-termini (Fig. 4A). For the open conformation structure, the map for the I41-L47 fragment of chain B exhibits certain ambiguity, hinting at multiple possible conformations (one conformation was modeled). PLP in the active sites of closed and open forms of mature *Mt*HISN6 (with the PLP-prosthetic group) is bound as a Schiff-base to K244 (internal aldimine, Fig. 4B). The *Mt*HISN6_open model contains 842 water residues and buffer components, such as two surface-bound sodium ions (one per chain), fifteen 1,2-ethanediol molecules, and six sulfate ions. Two sulfate ions are located in each active site,

while the remaining ones bind in small clefts (Fig. 4C). *Mt*HISN6_closed contains 1041 water residues and two acetate ions. Additionally, the map allowed us to model HEPES molecules (with partial occupancy) bound close to the dimer interface (Fig. 4D). The structure of the *Mt*HISN6-HOLP complex permitted the modeling of external aldimine in chain A and geminal diamine intermediate of HOLP in chain B; the side-chain conformations (except for K244) in the active sites are nearly identical. The model contains 823 water residues, two sodium ions, and eight 1,2-ethanediol molecules. Finally, the apo*Mt*HISN6 model was solved from crystals grown in ammonium sulfate but without adding an excess PLP. It contains 1413 water molecules, four sodium atoms, 12 sulfate ions, and 23 1,2-ethanediol molecules.

*Mt*HISN6 comprises the large (P72–N289) and the small domains (D25-P71, P290–S384), as per the terminology proposed by Ford et al. (1980) with modifications (Fig. 4A). The N-terminal part of the small domain is positioned in a shallow cleft on the dimer interface along the non-crystallographic 2-fold axis. It exists in two conformations: open and closed, which vary primarily by the positioning of the L35-P43 fragment that oscillates like a guitar string and the fragment F44–I58 containing two helices (Fig. 4B). The L35-P43 string is further from the active site in the open conformation, allowing the substrate to enter. Contrary, in the closed conformation, the L35-P43 string is drawn towards the active site entrance, restricting access and providing a proper environment for the catalysis. In the *Mt*HISN6-HOLP complex, which is also in the closed form, Y38 (from the string) H-bonds the substrate's phosphate oxygen and Nδ atoms (see below). Moreover, a minor rotation of the small domain in regard to the large domain occurs; the domains are closer in the *Mt*HISN6_closed structure (Fig. 4B).

The sodium cations are coordinated by the backbone carbonyls of K312, V314, and L317 (Fig. 4E). Furthermore, P315 imposes structural rigidity on the fragment, enforcing the positioning of the carbonyls to create the metal coordination sphere. Na$^+$ is bound to the same site in all ten subunits in this work structures, implying a specific binding relevant *in vivo*. This is a novelty, as previous structures of HATs from other sources did not contain metal cations except a magnesium ion bound to the *Geobacter metallireducens* HAT (Protein Data Bank, PDB, https://www.rcsb.org/; PDB ID: 3hdo, unpublished). However, in that case, the metal was bound at the crystal contact only in one of the two subunits. We cannot unambiguously state whether Na$^+$ binding is common in plant HISN6 enzymes with current data. However, the $_{312}$KAVPFL$_{317}$ fragment is similar in *Arabidopsis thaliana* HISN6A

Table 1

Data reduction and refinement statistics of MtHISN6 crystal structures. Values in parentheses are for the high resolution shell.

	*Mt*HISN6_open PDB ID: 8bj1	*Mt*HISN6_closed PDB ID: 8bj2	*Mt*HISN6_HOLP PDB ID: 8bj3	apo*Mt*HISN6 PDB ID: 8bj4
Diffraction source	APS 22-ID, Argonne, USA	APS 22-ID, Argonne, USA	APS 22-ID, Argonne, USA	P13 PETRA, Hamburg, Germany
Wavelength (Å)	1.0000	1.0000	1.0000	0.9763
Temperature (K)	100 K	100 K	100 K	100 K
Rotation range per image (°)	0.25	0.5	0.5	0.1
Total rotation range (°)	200	200	200	360
Space group	$P2_1$	$P2_1$	$P2_1$	$P2_1$
a, b, c (Å)	66.7, 87.4, 74.2	56.9, 105.8, 66.5	56.7, 66.3, 109.1	73.9, 93.0, 110.5
α, β, γ (°)	90, 95, 90	90, 108.9, 90	90, 106.5, 90	90, 73.5, 90
Mosaicity (°)	0.087	0.164	0.181	0.118
Resolution range (Å)	56.4–1.6 (1.66–1.57)	53.9–1.4 (1.48–1.40)	54.3–1.6 (1.71–1.61)	999.0–1.5 (1.54–1.45)
No. of unique reflections	116,740	145,641	97,979	252,775
Completeness (%)	98.6 (98.6)	99.2 (95.5)	99.5 (97.7)	99.6 (99.0)
Redundancy	3.9 (3.8)	4.0 (3.7)	4.1 (3.9)	6.77 (6.59)
I/σ(I)	18.02 (1.86)	13.89 (1.89)	18.61 (2.06)	13.71 (1.3)
R_{meas} (%)	5.2 (77.7)	5.8 (77.0)	5.8 (78.8)	8.6 (146.5)
CC1/2	99.9 (65.6)	99.9 (73.2)	99.9 (77.3)	99.9 (62.8)
Overall *B* factor: Wilson plot/all atoms (Å2)	20.8/28.3	16.8/21.3	19.2/22.1	18.5/26.5
No. of reflections: working/test set	116,740/1168	145,641/1164	97,979/1177	252,664/2527
R/ R_{free}	0.127/0.170	0.134/0.164	0.153/0.189	0.148/0.181
No. of non-H atoms: Protein/Ligand/Water	5936/92/887	5971/40/1087	5841/30/823	11,601/186/1630
R.m.s. deviations: Bonds (Å)/Angles (°)	0.008/1.011	0.008/0.930	0.008/0.920	0.008/0.944
Ramachandran plot: Most favored/allowed (%)	98.9/1.1	98.0/2.0	98.7/1.3	98.1/1.9
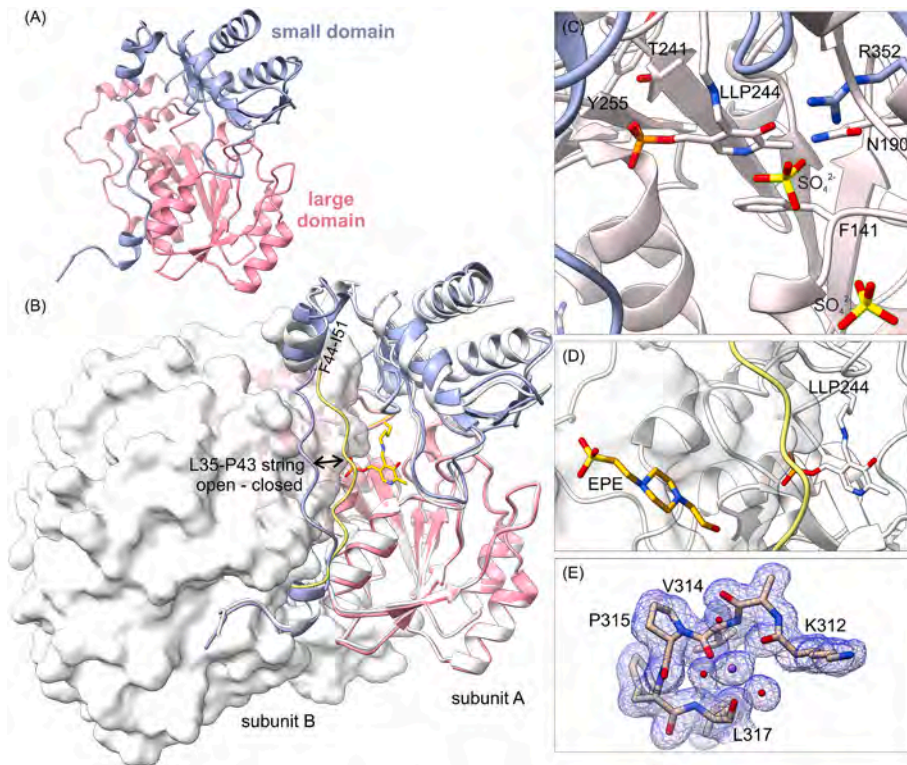
Fig. 4. Crystal structures of *Mt*HISN6 were determined in open and closed conformations. The most striking differences can be observed in the small domain, especially in the L35-P43 string. The subunit A of the open conformation structure is shown as a color-coded ribbon, where the small domain is blue and the large one is pink (A). (B) The dimer, where the subunit B is presented as a surface. *Mt*HISN6_-closed is superimposed and shown as white ribbons, with the L35-P43 string colored green for easier identification. The stick representation of PLP bound in the form of internal aldimine marks the active site (B). The sulfate ions bind in the open conformation structure, mimicking the phosphate group of HOLP. Furthermore, two sulfate ions bind in small clefts between large and small domains (C). The buffer molecule, HEPES (EPE), binds on the dimer interface creating a steric hindrance to the string's opening (D). Panel E illustrates the Na$^+$ binding site; the 2Fo-Fc electron density map (blue mesh) is contoured at 1σ level.
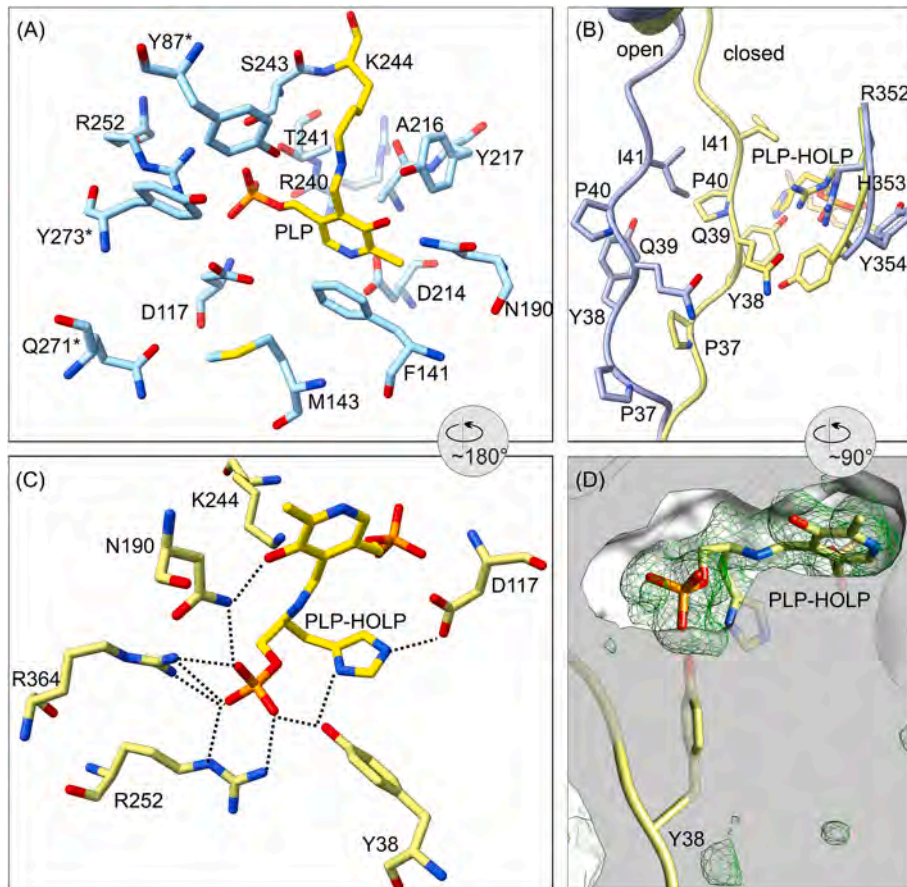
Fig. 5. The active site of *Mt*HISN6. PLP is bound in the form of internal aldimine creating a Schiff base. The side chains of residues constituting the active site are presented as sticks (A). (B) The movement of the string reduces the active site volume. The superposition of open (blue) and liganded (yellow) form structures shows how the closing of the loop restricts access to the active site. Binding of the ligand, HOLP, induces a conformational change of PLP, whose ring rotates by ~20° to attain position almost parallel to the phenyl ring of F141 (C). Panel D shows the Polder map (green mesh), contoured at 3σ level around PLP-HOLP. The protein surface is clipped to visualize the binding pocket.

(UniProt: B9DHD3; $_{345}$KEVPFL$_{350}$ sequence). Notwithstanding, the active site is > 20 Å away from Na$^+$, suggesting that the metal stabilizes the protein structure rather than directly participates in substrate binding or catalysis.

2.4. The active site of MtHISN6

The active site of *Mt*HISN6 contains a prosthetic group, PLP (Fig. 5A). PLP formed internal aldimines in our open and closed state structures, well defined by the electron density. The functional dimer has two active-site cavities located at the domain interface around the non-crystallographic 2-fold axis. Customarily, the residues forming the active site pocket are divided into bottom-forming and wall-forming ones (Haruyama et al., 2001). *Mt*HISN6 active site bottom consists of D117, F141, M143, N190, D214, A216, R240, Q271*, Y273* (asterisk denotes a residue from the dimer mate) whereas Y217, T241, S243,

R252, and Y87* build the wall (Fig. 5A). Residues of the small domain, in particular P37, Y38, Q39, P40, and I41 (part of the movable string), make up a lid covering the active site (Fig. 5B).

PLP binds to the active site by forming an internal aldimine with K244 and an extensive net of non-covalent interactions. The PLP phosphate acts as an anchor by interacting with Y87*, S243, R252, and T241 side chains and A116-D117 main-chain amides. The methyl group of A216 and the phenyl ring of F141 sandwich the pyridine ring. The PLP N1 interacts with D214, whereas the hydroxy group, O3, is stabilized by Y217 (Oη) and N190 (Nδ, Fig. 5A). Two additional water-mediated H-bonds further stabilize PLP with respect to the residues at the bottom and side of the active center cavity: R364, D117, and Y87*.

The change from the open to the closed state reduces the size of the active site cavity, e.g., the distance between I41 Cα and H353 Cα decreases from 13.9 Å to 10.1 Å. This change is attained by moving the L35-P43 string, which restricts access to the active site cavity
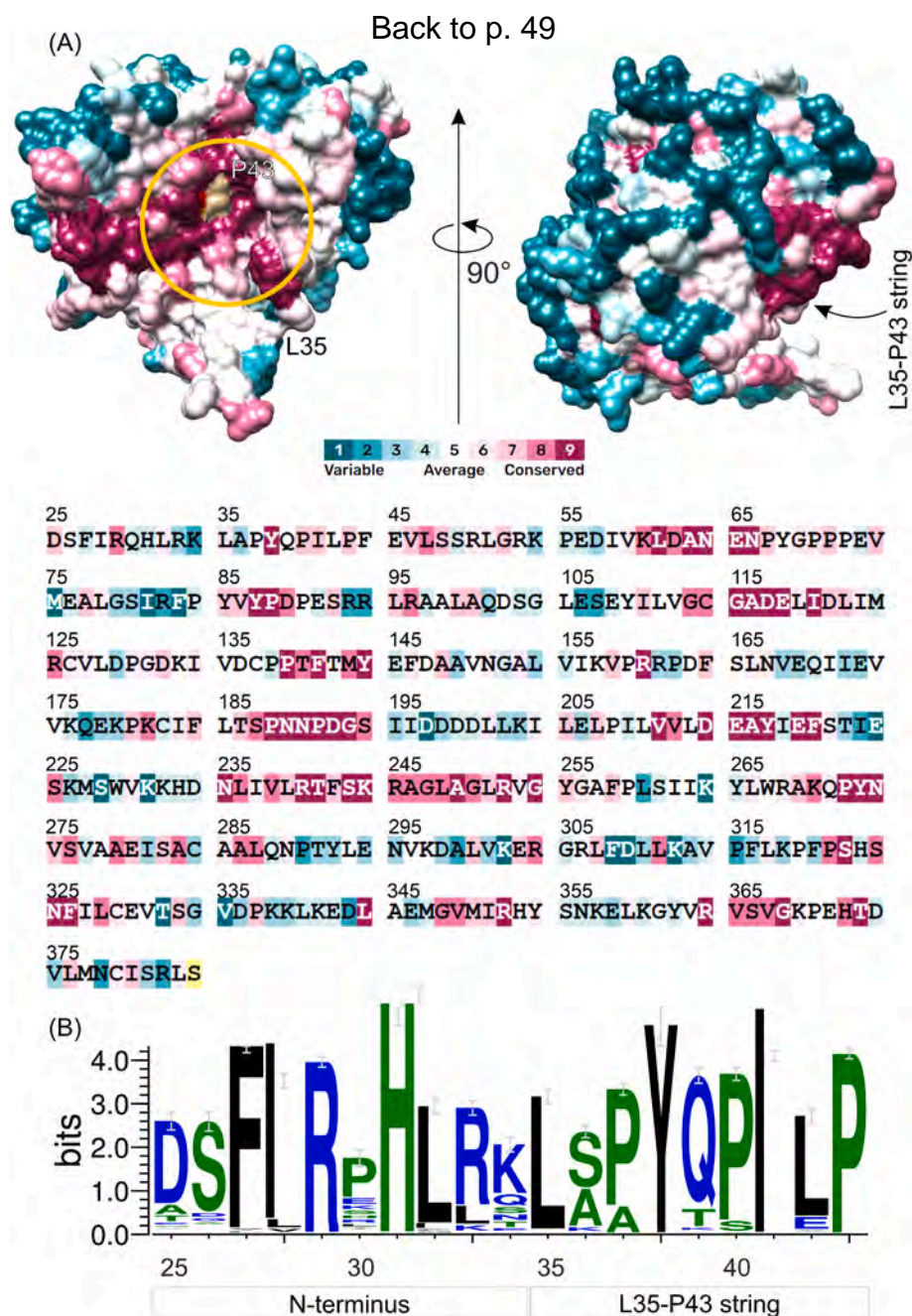
Fig. 6. Sequence conservation in *Mt*HISN6. The visual presentation of conservation scores (according to the color-key) obtained by analysis of 500 homologous sequences by *Consurf* is shown in (A). The active site region is marked by the yellow circle. The khaki surface region in the top-left panel marks the position of PLP; pale yellow highlight of S384 in the sequence indicates "insufficient data" for comparisons. Panel (B) shows sequence conservation in the N-terminus and the L35-P43 string among 349 HISN6 homologs in plants.

(Fig. 5B–D). The small domain (especially the fragment R352-K360) also moves towards the active site upon HOLP binding. We also observed adaptations in the residue conformations triggered by HOLP binding (Fig. 5C). Namely, the side chains of Y354 and R352 twist outward from the active site. This creates the environment for a firm locking of the negatively charged phosphate group of HOLP within the active site (Fig. 5B–D). Furthermore, the pyridine ring of PLP turns ~20° to attain a position almost parallel to the phenyl ring of F141.

Our apo*Mt*HISN6 structure allows us to infer rearrangements that accompany the binding of PLP. The electron density of the fragment K34–S48 (that contains the string region) is partial; the map does not cover most side chains, whereas I41 and L42 lack the density even for the backbone. The apo*Mt*HISN6 structure is most similar to the open form, as indicated by the reciprocal position between small and large domains and the residual electron density of the fragment K34–S48 corresponding with the open conformation of the string. This region becomes structured upon the PLP binding thanks to an extended net of water-mediated contacts.

2.5. Sequence conservation of HATs

We analyzed 500 sequences (261 of which contained the N-terminal part) that sampled 2247 homologous entries from the Uniref90 database using Consurf (Suzek et al., 2015; Ashkenazy et al., 2016) to gain more insights into residue conservation in HAT enzymes. The visual inspection of color-coded results mapped on the structure of HAT revealed that the core of the enzyme, the active site, and the dimerization interface are highly conserved (Fig. 6A). Interestingly, only Y38 (H-bonding the

phosphate and imidazole of HOLP) is conserved among almost all analyzed sequences that contain the N-terminal part, with two exceptions where Phe substitutes it. Other solvent-exposed side chains vary among HATs of different origins.

Using 349 sequences of plant HISN6 homologs retrieved from the InterPro database, we analyzed the conservation of the L35-P43 string and the preceding N-terminal sequence (Fig. 6B). Interestingly, despite its flexible nature, the sequence of the L35-P43 string is conserved among plant homologs, except for A36, which is substituted by Ser or Lys. Moreover, the string in plants contains three Pro residues (P37, P40, and P43 in *Mt*HISN6) that make the string more rigid. While some variability is noted, the sequence of *Mt*HISN6 is dominant among plants. Also, the N-terminus in *Mt*HISN6 (residues 25–34) contains the highest-consensus sequence in plants except for a rarely occurring Q30 (Fig. 6B). Altogether, apparent differences between *Mt*HISN6 and bacterial homologs as well as high sequence conservation of HISN6 enzymes in plants make *Mt*HISN6 an excellent model for the research on the plant HBP.

2.6. Structural comparison with homologs of bacterial origin

Escherichia coli HAT (eHspAT) and *Mt*HISN6 share 31% sequence identity. The open states superpose with the RMSD 1.61 Å calculated for Cα atoms (PDB ID: 1gew (Haruyama et al., 2001)). However, due to the lack of convincing electron density, the string was not modeled in open eHspAT. The string was also absent in the electron density in the complex of eHspAT with PLP-Glu external aldimine (PDB ID: 1gey (Haruyama et al., 2001),). For this reason, we compared the closed states of
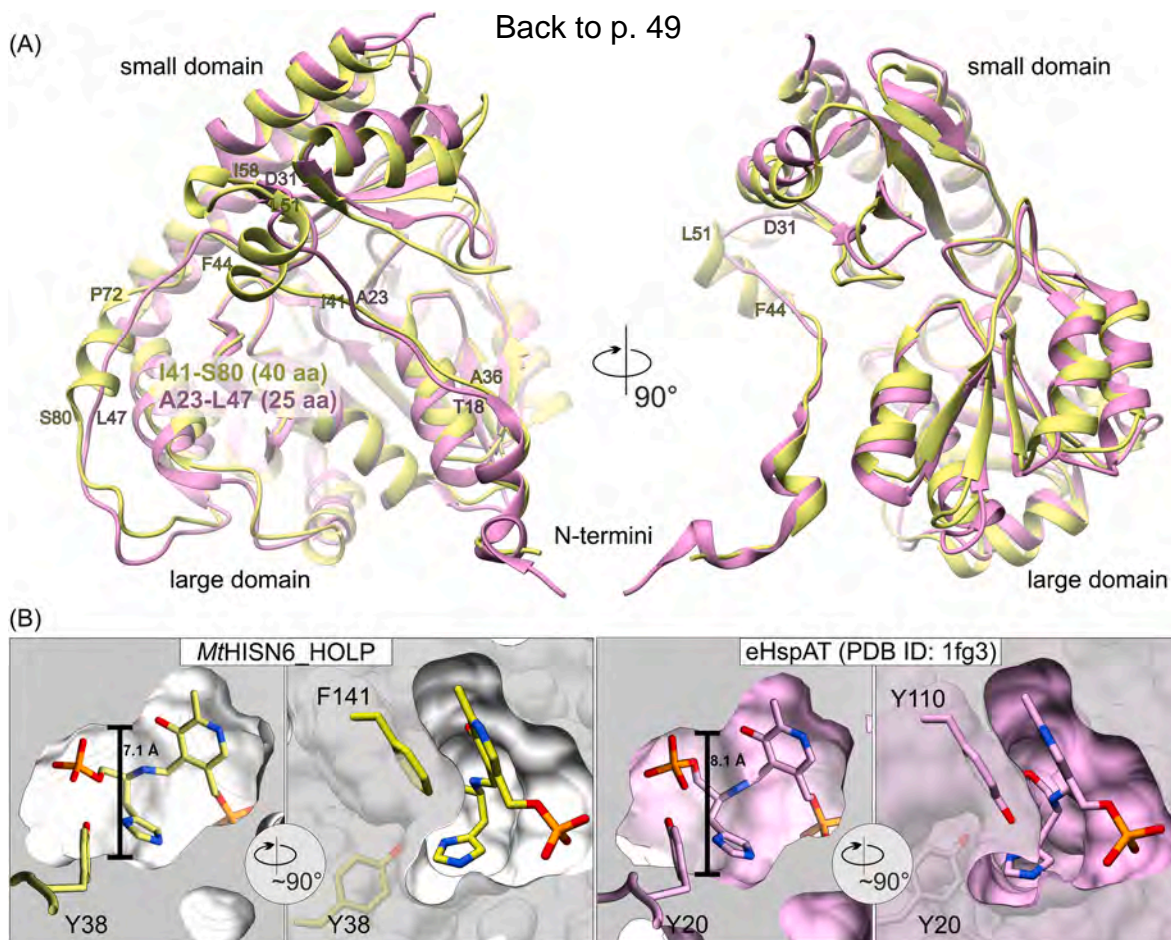
Fig. 7. Comparison of *Mt*HISN6 (green) with HAT from *E. coli* (eHspAT) (violet). Panel A shows the superposition of both structures with bound HOLP (PDB ID: 1fg3 for eHspAT). Cross-sections of the HOLP binding sites are presented in panel B; both surfaces were clipped at the same plane.

both enzymes in complexes with HOLP (RMSD 1.65 Å), using the structure by Sivaraman and coworkers (2001), who modeled the whole length of the chains starting from T3 (PDB ID: 1fg3). While the overall folds of *Mt*HISN6 and eHspAT are analogous, significant variations are evident in their secondary structure elements (Fig. 7A). The most striking differences occur in the small domains. The two enzymes differ significantly in how their N-terminal parts are affixed to the small domain. In *Mt*HISN6, the string is followed by an α-helix (F44-L51), a sharp turn, a β-strand (V59-L61), a linker, and an α-helix (P72–S80) that starts the large domain (Fig. 7A). The discussed range spans 40 residues in *Mt*HISN6 (I41–S80). The equivalent fragment in eHspAT is only 25 residues long (A23-L47) and contains a short loop (T18-D31) followed by a β-strand and a loop (W33-T38). Since the range encompasses the end of the string, its structure may play a role in maintaining the "tension" of the string so that its oscillation "stays in tune" with the reaction. In *Mt*HISN6, the fragment also fills in the indent between the subunits, thereby providing rigidity to the functional dimer. Comparing open (PDB IDs: 1fg7, 1iji) and closed (PDB IDs: 1fg3) states of available fully modeled eHspAT, the string region (L17-G27) in eHspAT adopts the same conformation with no hint of movement accompanying the change from the open to the closed state. The N-terminal fragment of the string (L17-S22) positions analogically to the *Mt*HISN6 string in the closed state, while A23-G27 point away from the active site entrance. Hence, in bacterial eHspAT the open/closed conformational change is based on the hinge-like movement of the small domain towards the active site rather than the movement of the string.

The sequence of the catalytic site is, unsurprisingly, one of the most conserved regions of HATs (Fig. 6A). The single difference we observe in *E. coli* is the presence of Y110 residue instead of F141 in *Mt*HISN6; these residues stack with the pyridine ring of PLP (Fig. 7B). The angle between the PLP and F/Y rings varies from 12° (PDB ID: 1gew) to 20° among different species and counts 17° for *Mt*HISN6. However, a comparison of the external aldimine complexes with HOLP shows that the HOLP binding site of *Mt*HISN6 is tighter (Fig. 7B). The narrow section of the active site measures 7.1 Å in *Mt*HISN6 compared to 8.1 Å in eHspAT. In effect, there is significantly less space near the HOLP imidazole moiety. It should be more difficult to accommodate an even bulkier phenyl of Phe to fit in the *Mt*HISN6 active site, which explains why *Mt*HISN6 did not transaminate Phe.

Since eHspAT lacks a biochemical description, we also compared *Mt*HISN6 to HAT from *Mycobacterium tuberculosis* (mHspAT) (Nasir et al., 2016). Superposition of the structures reveals higher RMSD values: (i) open onto open (PDB ID: 4rae, RMSD 2.7 Å, Fig. 8A) and (ii) closed onto closed (PDB 4r8d, 1.9 Å, Fig. 8B). In general, the differences between *Mt*HISN6 and mHspAT are similar to those noted in comparison with eHspAT. Moreover, in open mHspAT, the dimer mate slips into the active site of a subunit, displacing the whole N-terminal part, including the anchoring region (D25-R33). A long loop (S128-T137) within the large domain becomes an α-helix in the closed state. For *Mt*HISN6, the conformational changes are limited to the string open/closed transition
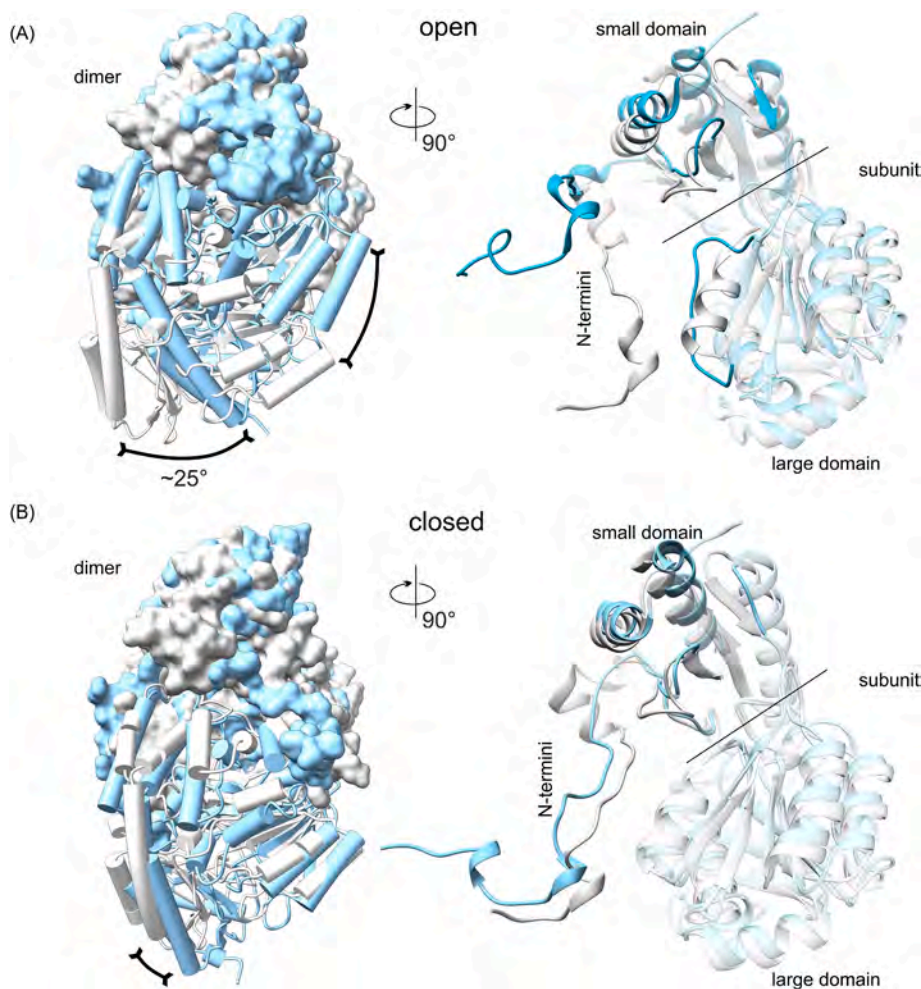
Fig. 8. Superposition of open (A) and close (B) structures of *Mt*HISN6 (white) and mHspAT (blue; PDB IDs: 4rae (A) and 4r8d (B)). The open-form structures differ significantly, especially in the small domain region, while the backbone of the closed forms superposes quite well, with most striking differences noticeable in the N-terminal regions. The approximate border between the small and large domains is marked by the black lines.

and a minor rotation of the small domain (Fig. 4B). This conformational flexibility provides a viable explanation why mHspAT has a significantly higher turnover rate (426 s^{-1}) but lower substrate affinity (K_M 420 μM). In the case of *Mt*HISN6, upon the substrate binding, the string moves further into the active site (by 3.7 Å), reducing its volume significantly. Notably, the high K_M values are a universal feature among studied bacterial homologs.

2.7. Dynamics of MtHISN6

To investigate the dynamics of *Mt*HISN6, we superposed our crystal structures in the open and closed states at their large domains (Fig. 4B; RMSD 0.79 Å). The rotational closing movement of the small domain displaces T291-F316, K348-D336, and K369–S384 helices and the linking fragments. Nonetheless, this movement is less pronounced than in mHspAT or eHspAT. As stated above, the most striking changes occur in the N-terminal part, including the P35-L43 string. The average displacement of the string is ~6.5 Å, reaching as far as 8.9 Å at P40 Cα. The following α-helical fragment, F44-D57, bends toward the active site, with an average displacement of 2 Å at Cα during structure closing. The open/close movement of the string is also accompanied by the side-chain movement of T265, R268, and a slight displacement of the main chain on the I263*-N274* and R352-R364 fragments. At the same time, the fragment D25-K34 stays anchored to the dimer mate's large domain, and the relative positions of two chains forming the dimer remain almost unaffected.

To evaluate the effect of HOLP binding on *Mt*HISN6, we performed 100-ns molecular dynamics (MD) simulations. The simulations were performed with two *Mt*HISN6 complexes with free K244, *i.e.*, complexes with PMP (modeled) and PLP-HOLP external aldimine to minimize bias and limit restrictions to the protein atoms. The results of MD simulations are consistent with the L35-P43 string participating in the active site formation upon the substrate binding (Fig. 9A and B). It is also clear that the binding of HOLP stabilizes the string, as the root-mean-square fluctuations (RMSFs) for Cα atoms decrease by half (Fig. 9C). On the other hand, we observe both open and close states in the absence of HOLP. Our crystal structures and the MD simulations are consistent with this regard. Uncoupling the dynamics of the *Mt*HISN6 string from ligand binding is a novelty. Such behavior may apply to other HATs, or at least be a universal feature of plant HISN6 enzymes, given the high sequence conservation of the string (Fig. 6B).

2.8. Virtual screening of soluble lead-like molecules

The HBP is absent in animals but is critical for plant growth. Thus, targeting plant HAT enzymes is a promising approach to the design of novel herbicides. The dynamics of *Mt*HISN6, especially the movement of the L35-P43 string, prompted us to exploit this region in the search for HISN6 inhibitors. Such molecules could lock the L35-P43 string in the open conformation, thus preventing it from closing, which is indispensable for providing proper reaction conditions.

We performed *in silico* docking of more than 1.35 mln soluble lead-like molecules downloaded from the ZINC database (Sterling and Irwin, 2015). The search box covered the area around the L35-P43 string of the *Mt*HISN6_open crystal structure. This way, we mapped druggable sites near the string while screening a vast chemical space. The chemical formulas and binding energies of the most potent 95 binders are provided in Supplementary Fig. S2.

The group of top hits are expected to bind in three distinct regions (Fig. 10A), with estimated binding energies up to −10.8 kcal/mol. They all have the potential to prevent the string from closing. Region 1 is located near the intersubunit interface between the small domains. ZINC9009723 (Fig. 10B) tightly fills the void between the outer ridge of the string and the small domain, while ZINC5013760 hints at the possibilities of linking Region 1 bound molecules with Region 2 bound molecules.
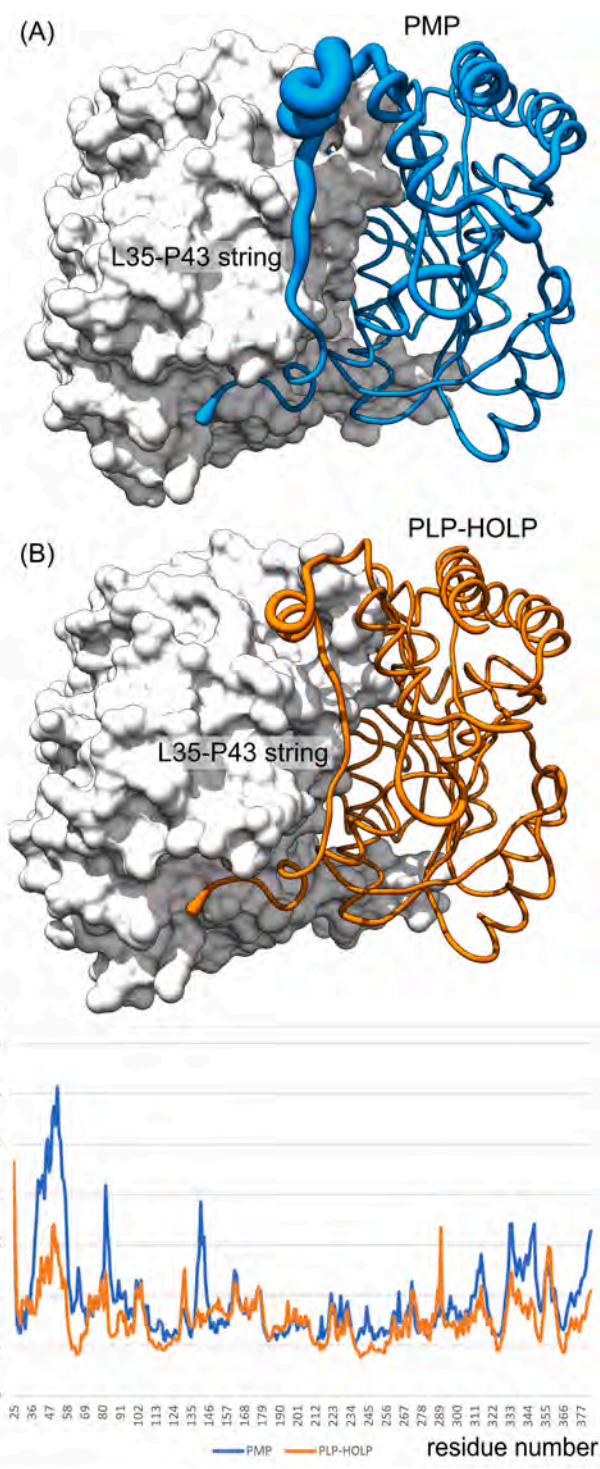
Fig. 9. Molecular dynamics simulations on *Mt*HISN6. Modeled complexes with PMP and PLP-HOLP (both external aldimines) are presented in panels A and B; one subunit is depicted as a white surface for clarity. The illustrations were created in UCSF Chimera (Pettersen et al., 2004) by presenting the B-factors calculated in Gromacs (Abraham et al., 2015) as widths of the "worms". The per residue root-mean-square-fluctuation (RMSF) of Cα atoms is shown in panel C.

Region 2 starts "behind" P43 and stretches towards the active site. The top-scoring molecules have a tetrazolopyridazine ring, locked in the site recognizing the phosphate group of HOLP, and interacting with R364 and PLP (ZINC95393759 (Fig. 10C), ZINC11085217, ZINC95350149). Interestingly, the substitution of the 1,2,4-triazole ring, situated in the middle of the molecule (ZINC95393759), by the
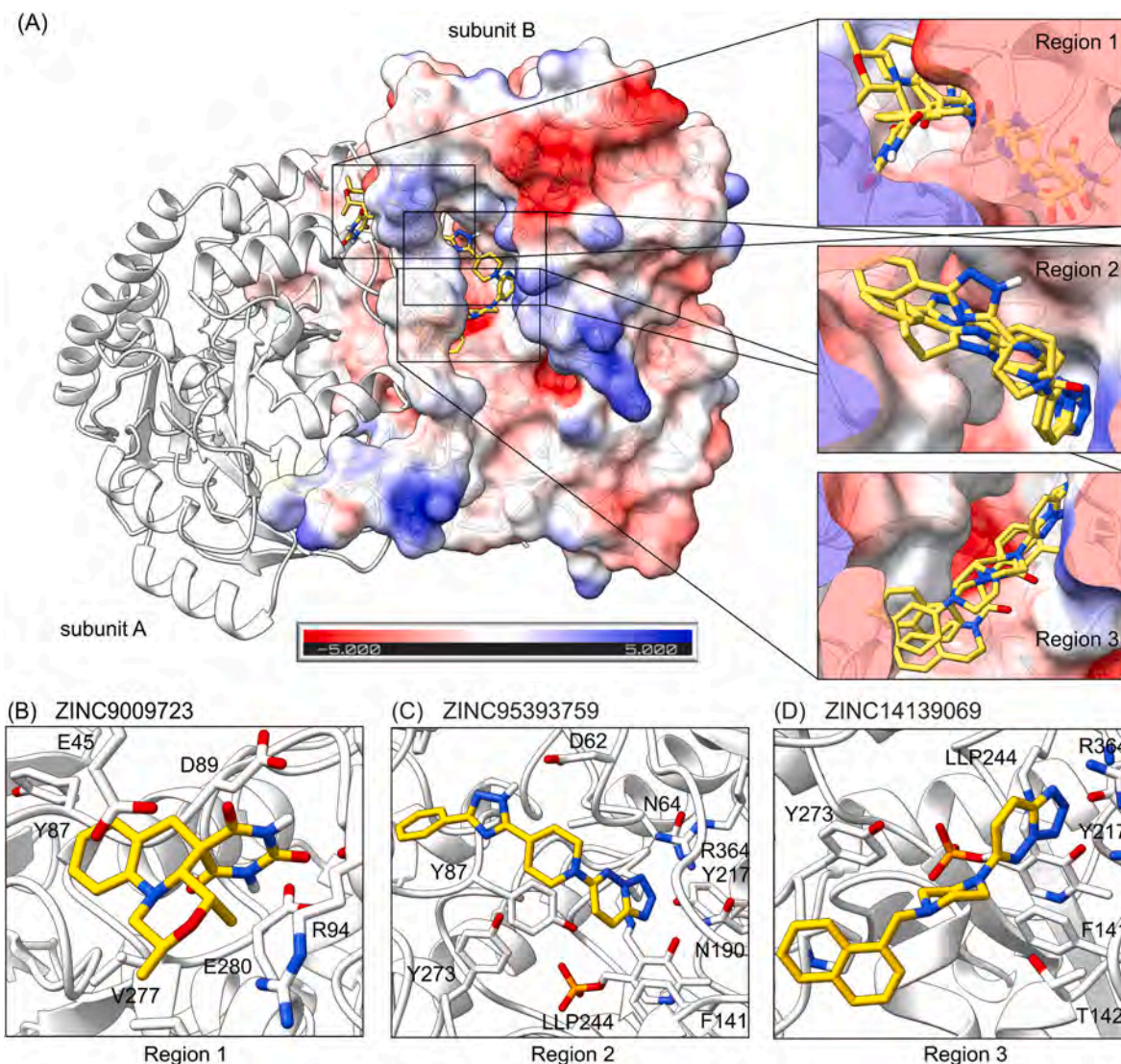
Fig. 10. The search region for the virtual screening campaign was oriented around the L35-P43 string. The molecules are predicted to bind in three slightly overlapping regions around the top part of the string. The three best scoring molecules in each region are presented as sticks with the *Mt*HISN6 surface electrostatic potential in the background (A). Detailed views over the three binding Regions (1–3) and their top-scoring binders are shown in panels B–D, respectively.

1,2,4-oxydiazole ring (ZINC95350149) decreases the estimated binding energy by 0.3 kcal/mol. The phenyl group on the other end of the molecule binds in a small tunnel between the top of the string and the small domain (ZINC95393759, ZINC12323188, ZINC95350149).

Region 3 starts at the heart of the active site, near PLP, and stretches along the large domain edge (T142-F146) towards A36 in the string. The top scoring molecules bind at the phosphate-binding site either using tetrazolopyridazine ring (ZINC14139069 (Fig. 10D)) or tetrazole group (ZINC65566055, ZINC33238693, ZINC5673503, ZINC72140159). The presence of negatively charged groups within the center of the binder has a beneficial effect on the binding energy (ZINC65566055 *vs.* ZINC5673503). Obtained results suggest that the best binders to Region 3 consist of the bicyclic aromatic moiety at the other end of the molecule (ZINC14139069, ZINC65566055, ZINC5673503, ZINC67674282).

3. Discussion

The evolution of plant HISN6 enzymes towards higher substrate selectivity is a phenomenon that was never described. Results obtained by SSN indicate that plant HISN6 evolved from a bacterial group, *Chloroflexota,* unlike many other chloroplast proteins that originate from

Cyanobacteria. Interestingly, both *Chloroflexota* and *Cyanobacteria* are able to photosynthesize. However, most *Chloroflexota* conduct anoxygenic photosynthesis, a more primitive form than oxygenic photosynthesis (Shiha et al., 2017; West-Roberts et al., 2021). The origin of plant HISN6 enzymes from *Chloroflexota* is an unusual indication of plant evolution via horizontal gene transfer (HGT). A few other examples of HGT from bacteria to higher plants include genes essential for survival on the land (Yue et al., 2012, 2013). One such example is phenylalanine ammonia lyase (PAL) which was transferred from soil bacteria or fungi (Emiliani et al., 2009). PAL catalyzes the first step of the phenylpropanoid pathway that is responsible for the production of compounds such as flavonoids and lignin that are essential for the protection against UV light and the formation of xylem, respectively. Genes encoding for proteins associated with vascular tissue formation were also acquired from bacteria, e.g., the vein patterning 1 protein (VAP1) or TAL-type transaldolase (TAL) (Jun et al., 2002; Yang et al., 2015). In previous works, we found out that another HBP enzyme, *Mt*HISN2, is most similar to homologs from the class of *Deltaproteobacteria* (now referred to as *Myxococcota* (Oren and Garrity, 2021; Witek et al., 2021). Hence, there is increasing evidence that many genes associated with cellular metabolism and survival on land were likely obtained from bacteria and later

evolved in plants. There is a pending question about the driver of the HGT in the case of plant HISN6 enzymes. Perhaps HATs from *Cyanobacteria* were not selective enough to ensure optimal metabolism inside the chloroplasts? HGT from *Chloroflexota* might have given plants evolutionary benefits, thereby satisfying the selective pressure. According to the Brenda database (Chang et al., 2021), *Bacillus subtilis* HAT exhibits the lowest K_M with HOLP (150 μM) among the studied homologs (Weigent and Nester, 1976). This still is five-fold higher than K_M of *Mt*HISN6. Currently, no data for any other plant species or *Cyanobacteria* are available.

The N-terminal part of the small domain has previously been indicated as responsible for the substrate specificity of bacterial HATs; it is built by hydrophilic residues that allow the binding of HOLP (Nasir et al., 2012, 2016). The comparison between the open and closed forms (including the HOLP complex) reveals that the binding of HOLP induces conformational changes. Given the structural similarity, the binding of IAP, which is the substrate in the forward direction of the HBP, is most likely similar to that of HOLP. Analysis of the *Mt*HISN6_HOLP structure allowed us to determine the Y38, D117, N190, R352, and R364 as the critical residues in HOLP recognition.

For broad-specificity aminotransferases, a certain ambiguity occurs in substrate preference (Vernal et al., 1998, 2003; Matsui et al., 2000; Fernandez et al., 2004). In such cases, the active site adapts its volume to fit smaller and bulkier substrates by modifying the extent of the small domain movement and conformational change of the so-called arginine switch (Wen et al., 2015; Bujacz et al., 2021). In *Mt*HISN6, Y38 plays a similar role, firmly closing the active site and limiting its size. Furthermore, the three conserved Pro residues (P37, P40, and P43) likely brought about the structural integrity of the closed active site in plant HISN6 enzymes. The contribution of the N-terminal part of the small domain in the regulation of size and accessibility of the active site has been indicated in mHspAT. However, when we compare the substrate profile of the two HATs, mHspAT can process both HOLP and Phe, while *Mt*HISN6 is HOLP-selective. This is consistent with the higher rigidity of the *Mt*HISN6 string, which makes the enzyme unable to accommodate other substrates. Furthermore, the P35-L43 string of *Mt*HISN6 moves further into the active site upon the substrate binding than the homologous region of mHspAT.

Our data suggest that the low turnover rate is the cost that *Mt*HISN6 pays for high substrate affinity. However, the high substrate affinity allows efficacy at low substrate concentrations *in vivo*. Unfortunately, we were not able to measure the activity in the HBP forward direction (IAP → HOLP) because IAP is not available. It must also be remembered that the subsequent reaction in the HBP, *i.e.*, the dephosphorylation of HOLP into L-histidinol by HISN7, is irreversible and therefore shifts the equilibrium towards His production (Petersen et al., 2010; Ruszkowski and Dauter, 2016).

Some inhibitors of plant aminotransferases are available, but none of them is HISN6-selective. For instance, aminoethoxyvinylglycine can deregulate histidine homeostasis in plants (Le Deunff et al., 2019). However, not only does it impact other aminotransferases in plants, it affects other organisms as well. Hence, the closing of the string L35-P43 has caught our eye, as it is clearly indispensable for catalysis. Targeting the string region, which is highly conserved in plant HISN6 enzymes, would restrict the string rearrangement to build the active site and could be an efficient approach toward HISN6 inhibition. We performed a virtual screening to get the first glimpse of specific regions near the string and chemical moieties predestined for binding. The results mapped three hot spots in the string neighborhood and indicated the development direction for HISN6 inhibitors that could become herbicides. Using the top-scoring molecules within each region, one may envision a molecule with a linker between moieties docking to different regions. This could anchor the molecule in the interdomain area, entwine the string from behind, and occupy vast space at the entrance of the catalytic site – preventing its closing. Certain redundancy within the group of top binders may provide a starting point for establishing a

structure-activity relationship (SAR).

Moreover, despite the high concentration of MES buffer (100 mM) in our crystallization trials, we do not observe MES molecules binding to *Mt*HISN6. This is in contrast to the case of mHspAT, where MES molecules bound in the active site and mimicked HOLP (Nasir et al., 2016). The sulfate ions present in our open-state structures could also serve as placeholders for negatively charged groups of new HISN6 inhibitors. Two sulfates bound in the active sites, while others occupy small clefts between the large and small domains, possibly locking the enzyme in the open state. Consistently, no sulfates are bound to the protein in our *Mt*HISN6_HOLP complex, which is the closest mimic of the *Mt*HISN6 active state.

4. Conclusions and outlook

As a result of the functional and structural study of *Mt*HISN6, here we provide a detailed insight into the evolution and characteristics of plant HISN6 enzymes. The calculated SSN suggests that plant HISN6 enzymes had been acquired by an HGT from *Chloroflexota* to have undergone further evolution within plants. The observed high substrate affinity and low turnover rate of *Mt*HISN6 can be attributed to the architecture of the active site that is tighter than that of bacterial counterparts with known structures. The closing of the highly-conserved in plants string region (residues L35-P43 in *Mt*HISN6) occurs dynamically – regardless of whether the substrate is bound in the active site or not. However, the substrate binding stabilizes the string in the closed state. On the other hand, the string must be closed to build a fully-functional active site. The high sequence conservation emphasizes the importance of the string for plant HISN6 activity, making the string dynamics a druggable site for HISN6-targeted herbicide development. Many plant species, unlike *A. thaliana*, possess a single HISN6 enzyme. Future research will show whether inhibition of HISN6 results in histidine starvation, accumulation of toxic levels of IAP, or impact signaling pathways linked to primary root growth (Mo et al., 2006). A phenotypic synergy may, in this case, enhance the efficacy of HISN6 inhibitors. Therefore, this work, which is a part of the project on deciphering the structures of the plant HBP enzymes, not only brings advancement in our understanding of the pathway at the molecular level but also provides the groundwork for the rationalized discovery of HISN6-targeted herbicides.

5. Materials and methods

5.1. Cloning, overexpression and purification of MtHISN6

The *Mt*HISN6 expression construct was produced using the protocol described for *Mt*HISN2 (Witek et al., 2021). In brief, total RNA isolated from *M. truncatula* leaves using the RNeasy Plant Mini Kit (Qiagen), was reverse-transcribed into the complementary DNA (cDNA) with Super-Script II reverse transcriptase (Life Technologies). The open reading frame coding for *Mt*HISN6 was amplified by polymerase chain reaction (PCR). The N-terminal signal peptide was excluded from the construct based on the prediction of the mature sequence based on TargetP 1.1 (Emanuelsson et al., 2000) prediction and by comparison of homologous sequences from other plant species. Based on this, the final construct was N-terminally truncated at D25 (UniProt ID: A0A072U7F9), whereas the C-terminus was not modified. The following primers were used for PCR: Forward, TACTTCCAATCCAATGCCGACTCCTTTATCAGACAACATCTCAGGA; Reverse: TTATCCACTTCCAATGTTACTACGATAGTCTGCTTATGCAATTCATCA. The pMCSG68 vector (Midwest Center for Structural Genomics, USA) served to create the expression plasmids by the ligase-independent cloning method (Kim et al., 2011). The plasmids were used to transform *E. coli* BL21 (DE3) Gold competent cells for protein expression. The correctness of the coding sequence was verified by DNA sequencing.

Overexpression was conducted in BL21 Gold *E. coli* cells (Agilent Technologies). Lysogeny broth (LB) culture medium was supplemented

with 150 μg/mL ampicillin. The cultures were incubated at 37 °C and shaken at 180 RPM until A_{600} reached the value of 1.0. The cultures were cooled down to 18 °C and induced using isopropyl-D-thiogalactopyranoside (IPTG) at final concentration of 0.5 mM. The cultures were then shaken for another 18 h. After that time, the cultures were centrifuged at 4000×*g* for 15 min at 4 °C and resuspended in 30.0 mL of binding buffer [50 mM Hepes-NaOH, pH 7.5; NaCl 500 mM; imidazole 20 mM; 2 mM tris(2-carboxyethyl)phosphine (TCEP)]. The resuspended bacterial cell pellet was deep frozen at −80 °C and stored for purification.

The bacterial cell pellet was thawed and disrupted by sonication (5 min) with intervals for cooling. Cell debris was removed by centrifugation at 25,000×*g* for 30 min at 4 °C. The supernatant was mixed with 3 mL of HisTrap HP resin (GE Healthcare) and incubated for 3 min in a column on the VacMan pump (Promega). The resin with bound protein was thoroughly washed with 400 mL of binding buffer and eluted with 15 mL of elution buffer (50 mM Hepes–NaOH pH 7.5; 500 mM NaCl; 400 mM imidazole; 2 mM TCEP). In order to remove the His₆-tag, tobacco etch virus (TEV) protease was added at a final concentration of 0.2 mg/mL. Cleavage was conducted overnight at 4 °C, simultaneously with dialysis (SnakeSkin™ Dialysis Tubing, Thermo Fisher) to lower the imidazole concentration down to 20 mM. Except for the apo*Mt*HISN6, PLP at 10 μM concentration was added to the external dialysis solution. The sample was loaded again onto a column containing HisTrap HP resin, and *Mt*HISN6 was eluted from the column. The sample was concentrated to ca. 2.0 mL using Amicon Ultra 15 mL centrifugal filters (Merck) and loaded onto a HiLoad Superdex 200 16/60 column (GE Healthcare). The column was previously equilibrated with size exclusion buffer (25 mM Hepes–NaOH pH 7.5; 100 mM KCl; 50 mM NaCl; 1 mM TCEP). After elution, the protein was concentrated using Amicon Ultra 15 mL centrifugal filter.

5.2. Activity assays

The aminotransferase activity of *Mt*HISN6 was measured using a coupled assay with bovine GDH. HOLP (lithium salt, Merck 41,486) was used to measure the activity in the reverse reaction (with respect to the HBP) as IAP is not commercially available. Concentrations and amounts of the constant components of the reaction mixtures were established in our preliminary experiments to be as follows: (i) forty units of bovine GDH (sigma G2626, sufficient to prevent the coupled reaction from being rate-limiting); (ii) *Mt*HISN6, 2 μM; (iii) NAD⁺, 1 mM; (iv) 2OG, 2 mM; (v) PLP, 2 μM and (vi) HOLP (or Phe) at varied concentrations. We also tested *A. thaliana* GDH1 (Grzechowiak et al., 2020), but this isoform was not applicable due to the high K_M. The reaction buffer was 25 mM HEPES-NaOH pH 7.5, 100 mM KCl, 50 mM NaCl. Measurements were performed on a Hewlett-Packard 8453 diode-array spectrophotometer (Agilent Technologies, Santa Clara, CA) in 800 μL volumes in duplicate. The formation of NADH was monitored continuously (2 Hz) through absorbance at 340 nm wavelength. Data were fit to the Michaelis-Menten equation using the GraphPad Prism 6.07 software. No detectable activity was present for Phe up to 25 mM concentration.

The binding of HOLP and Phe was compared by measuring the formation of the external aldimine, monitored by absorbance increase at around 355 nm for 4 min. The assay was performed in the mixture containing 25 mM HEPES-NaOH pH 7.5, 100 mM KCl, 50 mM NaCl, 20 μM PLP, and 20 μM *Mt*HISN6. Either Phe or HOLP was added at 1 mM final concentration. Results were analyzed in Spectragryph 1.2 (Spectroscopy Ninja).

5.3. Crystallization and diffraction data collection

MtHISN6 was crystallized using the vapor diffusion method in a sitting drop setup. The protein was concentrated to 18.8 mg/mL. The concentration was determined spectrophotometrically at A_{280} using the molar extinction coefficient of 30,870. *Mt*HISN6_open form crystals

grew in 90% of the C10 solution from the BCS screen (Molecular Dimensions), *i.e.*, 0.18 M ammonium sulfate, 0.09 M sodium acetate pH 4.6, 22.5% v/v PEG Smear Broad. The MtHISN6_closed form crystals were obtained from Index Screen G6 conditions (Hampton, USA): 0.2 M ammonium acetate; 0.1 M BIS-TRIS pH 5.5; 25% w/v polyethylene glycol 3350. The protein sample was supplemented with 1 mM PLP. As a cryoprotection, the mixture of the crystallization conditions with 20% ethylene glycol was used for both aforementioned forms. The latter crystals were used for soaking to obtain the *Mt*HISN6-HOLP complex. The cryoprotecting solution was supplemented with 0.2 μL of 200 mM HOLP added to 2 μL cryo drop. After a 10 min incubation, the soaked crystals were flash frozen. The apo*Mt*HISN6 crystals were obtained when no PLP was added to the protein sample before crystallization using the ShotGun (SG-1, Molecular Dimensions) in the B4 condition that contained the following components: 0.2 M ammonium sulfate; 0.1 M MES pH 6.5; 30% w/v PEG 5000 MME. The crystals were cryoprotected using SG-1 condition supplemented with 20% ethylene glycol. The crystals were vitrified in liquid nitrogen and stored for synchrotron data measurements.

Data were collected at the 22-ID beamline at the Advanced Photon Source, Argonne, USA or at the P13 beamline, PETRA, Hamburg, Germany. The diffraction images were processed *XDS* (Kabsch, 2010). The statistics of the data collection and processing are summarized in Table 1.

5.4. Determination and refinement of the crystal structures

The structure of *Mt*HISN6_open was solved by molecular replacement in *Phaser* (McCoy et al., 2007), using coordinates of *Cupriavidus pinatubonensis* JMP134 HAT (PDB ID: 3euc, unpublished) as a model; the sequences share 34% identity. The initial model of *Mt*HISN6 was obtained by sequence fitting in *Phenix.AutoBuild* (Afonine et al., 2012). The *Mt*HISN6_closed, *Mt*HISN6_HOLP and apo*Mt*HISN6_structures were solved by molecular replacement using *Mt*HISN6_open as a model. All four structures were refined by iterative cycles of manual rebuilding using *Coot* (Emsley P et al., 2010) and automated refinement in *Phenix. refine* (Liebschner et al., 2019). The geometric parameters and model-map correlation were validated in *MolProbity* (Williams et al., 2018) and the validation server of the PDB (Berman et al., 2000). The refinement statistics are listed in Table 1.

5.5. Virtual screening

A subset containing molecules of molecular weight 300–350 Da and logP ≤2 was derived from the lead-like library; 1,355,624 docking-ready files were downloaded from the ZINC15 database (Sterling and Irwin, 2015) in March 2022. Docking was performed in *AutoDock Vina* (Trott and Olson, 2010) using custom-made Python scripts, with the exhaustiveness = 8. *Mt*HISN6_open was prepared as the receptor for docking with the UCSF *Chimera* DockPrep tool (Pettersen et al., 2004). The search box with the dimensions of 33.9 × 28.0 × 33.3 Å was centered at 78.3, 30.7, 13.65 (x, y, z). The results were scored based on the calculated binding energy.

5.6. Molecular dynamics simulations

MD simulations were performed in *Gromacs 2021.5* (Abraham et al., 2015) with the AMBER99SB-ILDN force field applied (Lindorff-Larsen et al., 2010). *Mt*HISN6 complexes (created in *Coot* (Emsley P et al., 2010)) with either PMP or PLP-HOLP external aldimine were subjected to MD simulations independently to minimize bias. The model was solvated using TIP3P water model in a cubic box. Sodium cations were used to neutralize the system. Energy minimization was followed by NVT and NPT equilibrations, each for 50,000 of 2 fs steps. The production MD was run for a total of 100 ns (50,000,000 of 2-fs steps) at 300 K. The dynamics of the protein were evaluated based on

root-mean-square-fluctuation (RMSF) for Cα atoms, converted to B-factors using *Gromacs* internal tools and presented in UCSF *Chimera* (Pettersen et al., 2004) as worms.

5.7. Other software used

Molecular illustrations were created with UCSF *Chimera X* (Pettersen et al., 2021) and UCSF *Chimera* (Pettersen et al., 2004). RMSD values for the whole PDB search were taken from PDB-Fold (Krissinel and Henrick, 2004). A set of 500 sequences that samples 2247 unique homologs from UniProt (The UniProt, 2017) between 35 and 95% sequence identity to *Mt*HISN6 was created using Multiple Sequence Alignment using MAFFT and HMMER algorithm with an E-value of 0.0001. The conservation was analyzed with *ConSurf* (Ashkenazy et al., 2016). The 349 plant enzyme sequences were retrieved from InterPro database (Blum et al., 2021). Their MSA was prepared using the *Clustal Omega* webserver (https://www.ebi.ac.uk/Tools/msa/clustalo/) (Sievers et al., 2011). The distribution matrix was prepared using *WebLogo3* application (Crooks et al., 2004). The distribution of the surface electrostatic potential was calculated using *PDB2PQR* and *APBS* servers (Baker et al., 2001; Dolinsky et al., 2004). SSN analysis was calculated using the EFI-EST webserver (Zallot et al., 2019) from 21,142 UniRef90 (Suzek et al., 2015) sequences retrieved from the InterPro (Blum et al., 2021) family of L-histidinol phosphate aminotransferases (IPR005861). The alignment score was set to 100, while the sequence lengths were restricted to 320–440 residues. The graph was created in *Cytoscape 3.3* (Shannon et al., 2003).

Author contributions

M. Rutkiewicz refined the structures, analyzed them, and wrote the manuscript. IN performed the kinetic measurements with the help of SA; the data were analyzed by IN and RC. WW obtained the apoMtHISN6 structure and analyzed the HISN6 evolution. M. Ruszkowski obtained the other structures, wrote parts of the manuscript, and edited the text.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

PDB IDs: crystal structure of *Medicago truncatula* histidinol-phosphate aminotransferase (HISN6) in the open state, 8bj1; crystal structure of *Medicago truncatula* histidinol-phosphate aminotransferase (HISN6) in the closed state, 8bj2; crystal structure of *Medicago truncatula* histidinol-phosphate aminotransferase (HISN6) in complex with histidinol-phosphate, 8bj3; crystal structure of *Medicago truncatula* histidinol-phosphate aminotransferase (HISN6) in apo form, 8bj4. Raw diffraction datasets are available from the MX-RDR (https://mxrdr.icm.edu.pl/) under the following DOIs (respectively): https://doi.org/10.18150/VHEKMI; https://doi.org/10.18150/BTMRJV; https://doi.org/10.18150/ZMHOTT; https://doi.org/10.18150/ZEPS7J. All other data, including the results of the virtual screening, will be shared upon request.

Acknowledgments

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.plaphy.2023.02.017.

References

Abraham, M., Murtola, T., Schulz, R., Páll, S., Smith, J., Hess, B., Lindahl, E., 2015. GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. Software 1, 19–25. https://doi.org/10.1016/j.softx.2015.06.001.

Afonine, P.V., Grosse-Kunstleve, R.W., Echols, N., Headd, J.J., Moriarty, N.W., Mustyakimov, M., Terwilliger, T.C., Urzhumtsev, A., Zwart, P.H., Adams, P.D., 2012. Towards automated crystallographic structure refinement with phenix.refine. Acta Crystallogr. D 68, 352–367. https://doi.org/10.1107/S0907444912001308.

Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T., Ben-Tal, N., 2016. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. Nucleic Acids Res. 44, W344–W350. https://doi.org/10.1093/nar/gkw408.

Baker, N.A., Sept, D., Joseph, S., Holst, M.J., Mccammon, J.A., 2001. Electrostatics of nanosystems: application to microtubules and the ribosome. Proc. Natl. Acad. Sci. U. S. A. 98, 10037–10041. https://doi.org/10.1073/pnas.181342398.

Beckie, H.J., Busi, R., Lopez-Ruiz, F.J., Umina, P.A., 2021. Herbicide resistance management strategies: how do they compare with those for insecticides, fungicides and antibiotics? Pest Manag. Sci. 77, 3049–3056. https://doi.org/10.1002/ps.6395.

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., Bourne, P.E., 2000. The protein Data Bank. Nucleic Acids Res. 28, 235–242. https://doi.org/10.1093/nar/28.1.235.

Blum, M., Chang, H.Y., Chuguransky, S., Grego, T., Kandasaamy, S., Mitchell, A., Nuka, G., Paysan-Lafosse, T., Qureshi, M., Raj, S., Richardson, L., Salazar, G.A., Williams, L., Bork, P., Bridge, A., Gough, J., Haft, D.H., Letunic, I., Marchler-Bauer, A., Mi, H.Y., Natale, D.A., Necci, M., Orengo, C.A., Pandurangan, A.P., Rivoire, C., Sigrist, C.J.A., Sillitoe, I., Thanki, N., Thomas, P.D., Tosatto, S.C.E., Wu, C.H., Bateman, A., Finn, R.D., 2021. The InterPro protein families and domains database: 20 years on. Nucleic Acids Res. 49, D344–D354. https://doi.org/10.1093/nar/gkaa977.

Bujacz, A., Rum, J., Rutkiewicz, M., Pietrzyk-Brzezinska, A.J., Bujacz, G., 2021. Structural evidence of active site adaptability towards different sized substrates of aromatic amino acid aminotransferase from psychrobacter. Sp. B6. Materials (Basel) 14. https://doi.org/10.3390/ma14123351.

Chang, A., Jeske, L., Ulbrich, S., Hofmann, J., Koblitz, J., Schomburg, I., Neumann-Schaal, M., Jahn, D., Schomburg, D., 2021. BRENDA, the ELIXIR core data resource in 2021: new developments and updates. Nucleic Acids Res. 49, D498–D508. https://doi.org/10.1093/nar/gkaa1025.

Cianci, M., Bourenkov, G., Pompidor, G., Karpics, I., Kallio, J., Bento, I., Roessle, M., Cipriani, F., Fiedler, S., Schneider, T.R., 2017. P13, the EMBL macromolecular crystallography beamline at the low-emittance PETRA III ring for high- and low-energy phasing with variable beam focusing. J. Synchrotron Radiat. 24, 323–332. https://doi.org/10.1107/S1600577516016465.

Crooks, G.E., Hon, G., Chandonia, J.M., Brenner, S.E., 2004. WebLogo: a sequence logo generator. Genome Res. 14, 1188–1190. https://doi.org/10.1101/gr.849004.

Dolinsky, T.J., Nielsen, J.E., Mccammon, J.A., Baker, N.A., 2004. PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. Nucleic Acids Res. 32, W665–W667. https://doi.org/10.1093/nar/gkh381.

Duke, S.O., 2018. The history and current status of glyphosate. Pest Manag. Sci. 74, 1027–1034. https://doi.org/10.1002/ps.4652.

Emanuelsson, O., Nielsen, H., Brunak, S., Von Heijne, G., 2000. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. J. Mol. Biol. 300, 1005–1016. https://doi.org/10.1006/jmbi.2000.3903.

Emiliani, G., Fondi, M., Fani, R., Gribaldo, S., 2009. A horizontal gene transfer at the origin of phenylpropanoid metabolism: a key adaptation of plants to land. Biol. Direct 4, 7. https://doi.org/10.1186/1745-6150-4-7.

Emsley, P., Lohkamp, B., Scott, W.G., K, C., 2010. Features and development of Coot. Acta Crystallogr. D66, 486–501. https://doi.org/10.1107/S0907444910007493.

Farina, W.M., Balbuena, M.S., Herbert, L.T., Gonalons, C.M., Vazquez, D.E., 2019. Effects of the herbicide glyphosate on honey bee sensory and cognitive abilities: individual impairments with implications for the hive. Insects 10, 354. https://doi.org/10.3390/insects10100354.

Fernandez, F.J., Vega, M.C., Lehmann, F., Sandmeier, E., Gehring, H., Christen, P., Wilmanns, M., 2004. Structural studies of the catalytic reaction pathway of a hyperthermophilic histidinol-phosphate aminotransferase. J. Biol. Chem. 279, 21478–21488. https://doi.org/10.1074/jbc.M400291200.

Ford, G.C., Eichele, G., Jansonius, J.N., 1980. Three-dimensional structure of a pyridoxal-phosphate-dependent enzyme, mitochondrial aspartate aminotransferase. Proc. Natl. Acad. Sci. USA 77, 2559–2563. https://doi.org/10.1073/pnas.77.5.2559.

Forlani, G., Mangiagalli, A., Nielsen, E., Suardi, C.M., 1999. Degradation of the phosphonate herbicide glyphosate in soil: evidence for a possible involvement of unculturable microorganisms. Soil Biol. Biochem. 31, 991–997. https://doi.org/10.1016/S0038-0717(99)00010-3.

Gaines, T.A., Busi, R., Kupper, A., 2021. Can new herbicide discovery allow weed management to outpace resistance evolution? Pest Manag. Sci. 77, 3036–3041. https://doi.org/10.1002/ps.6457.

Glynn, S.E., Baker, P.J., Sedelnikova, S.E., Davies, C.L., Eadsforth, T.C., Levy, C.W., Rodgers, H.F., Blackburn, G.M., Hawkes, T.R., Viner, R., Rice, D.W., 2005. Structure and mechanism of imidazoleglycerol-phosphate dehydratase. Structure 13, 1809–1817. https://doi.org/10.1016/j.str.2005.08.012.

Gould, F., Brown, Z.S., Kuzma, J., 2018. Wicked evolution: can we address the sociobiological dilemma of pesticide resistance? Science 360, 728–732. https://doi.org/10.1126/science.aar3780.

Grzechowiak, M., Sliwiak, J., Jaskolski, M., Ruszkowski, M., 2020. Structural studies of glutamate dehydrogenase (isoform 1) from Arabidopsis thaliana, an important enzyme at the branch-point between carbon and nitrogen metabolism. Front. Plant Sci. 11, 754. https://doi.org/10.3389/fpls.2020.00754.

Guyton, K.Z., Loomis, D., Grosse, Y., El Ghissassi, F., Benbrahim-Tallaa, L., Guha, N., Scoccianti, C., Mattock, H., Straif, K., Working, I.a.R.C.M., 2015. Carcinogenicity of tetrachlorvinphos, parathion, malathion, diazinon, and glyphosate. Lancet Oncol. 16, 490–491. https://doi.org/10.1016/S1470-2045(15)70134-8.

Haruyama, K., Nakai, T., Miyahara, I., Hirotsu, K., Mizuguchi, H., Hayashi, H., Kagamiyama, H., 2001. Structures of Escherichia coli histidinol-phosphate aminotransferase and its complexes with histidinol-phosphate and N-(5′-Phosphopyridoxyl)-L-Glutamate: double substrate recognition of the enzyme. Biochemistry 40, 4633–4644. https://doi.org/10.1021/bi002769u.

Jensen, R.A., Gu, W., 1996. Evolutionary recruitment of biochemically specialized subdivisions of family I within the protein superfamily of aminotransferases. J. Bacteriol. 178, 2161–2171. https://doi.org/10.1128/jb.178.8.2161-2171.1996.

Jun, J.H., Ha, C.M., Nam, H.G., 2002. Involvement of the VEP1 gene in vascular strand development in Arabidopsis thaliana. Plant Cell Physiol. 43, 323–330. https://doi.org/10.1093/pcp/pcf042.

Kabsch, W., 2010. Xds. Acta Cryst. D 66, 125–132. https://doi.org/10.1107/S0907444909047337.

Kim, Y., Babnigg, G., Jedrzejczak, R., Eschenfeldt, W.H., Li, H., Maltseva, N., Hatzos-Skintges, C., Gu, M.Y., Makowska-Grzyska, M., Wu, R.Y., An, H., Chhor, G., Joachimiak, A., 2011. High-throughput protein purification and quality assessment for crystallization. Methods 55, 12–28. https://doi.org/10.1016/j.ymeth.2011.07.010.

Krissinel, E., Henrick, K., 2004. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. Acta Crystallogr. D 60, 2256–2268. https://doi.org/10.1107/S0907444904026460.

Le Deunff, E., Beauclair, P., Deleu, C., Lecourt, J., 2019. Inhibition of aminotransferases by aminoethoxyvinylglycine triggers a nitrogen limitation condition and deregulation of histidine homeostasis that impact root and shoot development and nitrate uptake. Front. Plant Sci. 10, 1387. https://doi.org/10.3389/fpls.2019.01387.

Liebschner, D., Afonine, P.V., Baker, M.L., Bunkoczi, G., Chen, V.B., Croll, T.I., Hintze, B., Hung, L.-W., Jain, S., Mccoy, A.J., Moriarty, N.W., Oeffner, R.D., Poon, B.K., Prisant, M.G., Read, R.J., Richardson, J.S., Richardson, D.C., Sammito, M.D., Sobolev, O.V., Stockwell, D.H., Terwilliger, T.C., Urzhumtsev, A.G., Videau, L.L., Williams, C.J., Adams, P.D., 2019. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. Acta Crystallogr. D 75, 861–877. https://doi.org/10.1107/S2059798319011471.

Lindorff-Larsen, K., Piana, S., Palmo, K., Maragakis, P., Klepeis, J.L., Dror, R.O., Shaw, D.E., 2010. Improved side-chain torsion potentials for the Amber ff99SB protein force field. Proteins 78, 1950–1958. https://doi.org/10.1002/prot.22711.

Maeda, H., Dudareva, N., 2012. The shikimate pathway and aromatic amino acid biosynthesis in plants. Annu. Rev. Plant Biol. 63 (63), 73–105. https://doi.org/10.1146/annurev-arplant-042811-105439.

Matsui, I., Matsui, E., Sakai, Y., Kikuchi, H., Kawarabayasi, Y., Ura, H., Kawaguchi, S.-I., Kuramitsu, S., Harata, K., 2000. The molecular structure of hyperthermostable aromatic aminotransferase with novel substrate specificity from <em>Pyrococcus horikoshii</em&gt. J. Biol. Chem. 275, 4871–4879. https://doi.org/10.1074/jbc.275.7.4871.

Mccoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., Read, R.J., 2007. Phaser crystallographic software. J. Appl. Crystallogr. 40, 658–674. https://doi.org/10.1107/S0021889807021206.

Mo, X.R., Zhu, Q.Y., Li, X., Li, J., Zeng, Q.N., Rong, H.L., Zhang, H.M., Wu, P., 2006. The hpa1 mutant of Arabidopsis reveals a crucial role of histidine homeostasis in root meristem maintenance. Plant Physiol. 141, 1425–1435. https://doi.org/10.1104/pp.106.084178.

Motta, E.V.S., Raymann, K., Moran, N.A., 2018. Glyphosate perturbs the gut microbiota of honey bees. Proc. Natl. Acad. Sci. USA 115, 10305–10310. https://doi.org/10.1073/pnas.1803880115.

Muralla, R., Sweeney, C., Stepansky, A., Leustek, T., Meinke, D., 2007. Genetic dissection of histidine biosynthesis in Arabidopsis. Plant Physiol. 144, 890–903. https://doi.org/10.1104/pp.107.096511.

Nasir, N., Anant, A., Vyas, R., Biswal, B.K., 2016. Crystal structures of Mycobacterium tuberculosis HspAT and ArAT reveal structural basis of their distinct substrate specificities. Sci. Rep. 6, 18880 https://doi.org/10.1038/srep18880.

Nasir, N., Vyas, R., Chugh, C., Ahangar, M.S., Biswal, B.K., 2012. Molecular cloning, overexpression, purification, crystallization and preliminary X-ray diffraction studies of histidinol phosphate aminotransferase (HisC2) from Mycobacterium tuberculosis. Acta Crystallogr. F 68, 32–36. https://doi.org/10.1107/S1744309111045386.

Oren, A., Garrity, G.M., 2021. Valid publication of the names of forty-two phyla of prokaryotes. Int. J. Syst. Evol. Microbiol. 71 https://doi.org/10.1099/ijsem.0.005056.

Petersen, L.N., Marineo, S., Mandala, S., Davids, F., Sewell, B.T., Ingle, R.A., 2010. The missing link in plant histidine biosynthesis: Arabidopsis myoinositol monophosphatase-like2 encodes a functional histidinol-phosphate phosphatase. Plant Physiol. 152, 1186–1196. https://doi.org/10.1104/pp.109.150805.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., Ferrin, T.E., 2004. UCSF Chimera-a visualization system for exploratory research and analysis. J. Comput. Chem. 25, 1605–1612. https://doi.org/10.1002/jcc.20084.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Meng, E.C., Couch, G.S., Croll, T.I., Morris, J.H., Ferrin, T.E., 2021. UCSF ChimeraX: structure visualization for researchers, educators, and developers. Protein Sci. 30, 70–82. https://doi.org/10.1002/pro.3943.

Raven, J.A., Allen, J.F., 2003. Genomics and chloroplast evolution: what did cyanobacteria do for plants? Genome Biol. 4, 209. https://doi.org/10.1186/gb-2003-4-3-209.

Ruszkowski, M., 2018. Guarding the gateway to histidine biosynthesis in plants: Medicago truncatula ATP-phosphoribosyltransferase in relaxed and tense states. Biochem. J. 475, 2681–2697. https://doi.org/10.1042/BCJ20180289.

Ruszkowski, M., Dauter, Z., 2016. 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. 291, 9960–9973. https://doi.org/10.1074/jbc.M115.708727.

Ruszkowski, M., Dauter, Z., 2017. Structures of Medicago truncatula L-histidinol dehydrogenase show rearrangements required for NAD(+) binding and the cofactor positioned to accept a hydride. Sci. Rep. 7, 10476 https://doi.org/10.1038/s41598-017-10859-0.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., Ideker, T., 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 13, 2498–2504. https://doi.org/10.1101/gr.1239303.

Shiha, P.M., Ward, L.M., Fischer, W.W., 2017. Evolution of the 3-hydroxypropionate bicycle and recent transfer of anoxygenic photosynthesis into the Chloroflexi. Proc. Natl. Acad. Sci. USA 114, 10749–10754. https://doi.org/10.1073/pnas.1710798114.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W.Z., Lopez, R., Mcwilliam, H., Remmert, M., Soding, J., Thompson, J.D., Higgins, D.G., 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. 7 https://doi.org/10.1038/msb.2011.75.

Sivaraman, J., Li, Y., Larocque, R., Schrag, J.D., Cygler, M., Matte, A., 2001. Crystal structure of histidinol phosphate aminotransferase (HisC) from Escherichia coli, and its covalent complex with pyridoxal-5'-phosphate and l-histidinol phosphate. J. Mol. Biol. 311, 761–776. https://doi.org/10.1006/jmbi.2001.4882.

Stepansky, A., Leustek, T., 2006. Histidine biosynthesis in plants. Amino Acids 30, 127–142. https://doi.org/10.1007/s00726-005-0247-0.

Sterling, T., Irwin, J.J., 2015. ZINC 15–Ligand discovery for everyone. J. Chem. Inf. Model. 55, 2324–2337. https://doi.org/10.1021/acs.jcim.5b00559.

Suzek, B.E., Wang, Y., Huang, H., Mcgarvey, P.B., Wu, C.H., Uniprot, C., 2015. UniRef clusters: a comprehensive and scalable alternative for improving sequence similarity searches. Bioinformatics 31, 926–932. https://doi.org/10.1093/bioinformatics/btu739.

The Uniprot, C., 2017. UniProt: the universal protein knowledgebase. Nucleic Acids Res. 45, D158–D169. https://doi.org/10.1093/nar/gkw1099.

Trott, O., Olson, A.J., 2010. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J. Comput. Chem. 31, 455–461. https://doi.org/10.1002/jcc.21334.

Vazquez, D.E., Balbuena, M.S., Chaves, F., Gora, J., Menzel, R., Farina, W.M., 2020. Sleep in honey bees is affected by the herbicide glyphosate. Sci. Rep. 10, 10516 https://doi.org/10.1038/s41598-020-67477-6.

Vernal, J., Cazzulo, J.J., Nowicki, C., 1998. Isolation and partial characterization of a broad specificity aminotransferase from Leishmania mexicana promastigotes. Mol. Biochem. Parasitol. 96, 83–92. https://doi.org/10.1016/S0166-6851(98)00117-0.

Vernal, J., Cazzulo, J.J., Nowicki, C., 2003. Cloning and heterologous expression of a broad specificity aminotransferase of Leishmania mexicana promastigotes. FEMS (Fed. Eur. Microbiol. Soc.) Microbiol. Lett. 229, 217–222. https://doi.org/10.1016/s0378-1097(03)00824-3.

Weidenmuller, A., Meltzer, A., Neupert, S., Schwarz, A., Kleineidam, C., 2022. Glyphosate impairs collective thermoregulation in bumblebees. Science 376, 1122–1126. https://doi.org/10.1126/science.abf7482.

Weigent, D.A., Nester, E.W., 1976. Purification and properties of two aromatic aminotransferases in Bacillus subtilis. J. Biol. Chem. 251, 6974–6980.

Wen, J., Nowicki, C., Blankenfeldt, W., 2015. Structural basis for the relaxed substrate selectivity of Leishmania mexicana broad specificity aminotransferase. Mol. Biochem. Parasitol. 202, 34–37. https://doi.org/10.1016/j.molbiopara.2015.09.007.

West-Roberts, J.A., Matheus-Carnevali, P.B., Schoelmerich, M.C., Al-Shayeb, B., Thomas, A.D., Sharrar, A., He, C., Chen, L.-X., Lavy, A., Keren, R., Amano, Y., Banfield, J.F., 2021. The Chloroflexi supergroup is metabolically diverse and representatives have novel genes for non-photosynthesis based CO2 fixation. bioRxiv. https://doi.org/10.1101/2021.08.23.457424.

Williams, C.J., Headd, J.J., Moriarty, N.W., Prisant, M.G., Videau, L.L., Deis, L.N., Verma, V., Keedy, D.A., Hintze, B.J., Chen, V.B., Jain, S., Lewis, S.M., Arendall 3rd, W.B., Snoeyink, J., Adams, P.D., Lovell, S.C., Richardson, J.S., Richardson, D.C., 2018. MolProbity: more and better reference data for improved all-atom structure validation. Protein Sci. 27, 293–315. https://doi.org/10.1002/pro.3330.

Witek, W., Sliwiak, J., Ruszkowski, M., 2021. Structural and mechanistic insights into the bifunctional HISN2 enzyme catalyzing the second and third steps of histidine biosynthesis in plants. Sci. Rep. 11, 9647. https://doi.org/10.1038/s41598-021-88920-2.

Yang, Z.F., Zhou, Y., Huang, J.L., Hu, Y.Y., Zhang, E.Y., Xie, Z.W., Ma, S.J., Gao, Y., Song, S., Xu, C.W., Liang, G.H., 2015. Ancient horizontal transfer of transaldolase-like protein gene and its role in plant vascular development. New Phytol. 206, 807–816. https://doi.org/10.1111/nph.13183.

Yue, J., Hu, X., Huang, J., 2013. Horizontal gene transfer in the innovation and adaptation of land plants. Plant Signal. Behav. 8, e24130 https://doi.org/10.4161/psb.24130.

Yue, J.P., Hu, X.Y., Sun, H., Yang, Y.P., Huang, J.L., 2012. Widespread impact of horizontal gene transfer on plant colonization of land. Nat. Commun. 3, 1152. https://doi.org/10.1038/ncomms2148.

Yuzawa, Y., Nishihara, H., Haraguchi, T., Masuda, S., Shimojima, M., Shimoyama, A., Yuasa, H., Okada, N., Ohta, H., 2012. Phylogeny of galactolipid synthase homologs together with their enzymatic analyses revealed a possible origin and divergence time for photosynthetic membrane biogenesis. DNA Res. 19, 91–102. https://doi.org/10.1093/dnares/dsr044.

Zallot, R., Oberg, N., Gerlt, J.A., 2019. The EFI web resource for genomic enzymology tools: leveraging protein, genome, and metagenome databases to discover novel enzymes and metabolic pathways. Biochemistry 58, 4169–4182. https://doi.org/10.1021/acs.biochem.9b00735.

Supplementary Material

Insights into the Substrate Specificity, Structure, and Dynamics of Plant Histidinol-Phosphate Aminotransferase (HISN6)

Maria Rutkiewicz[1], Isabel Nogues[2], Wojciech Witek[1], Sebastiana Angelaccio[3], Roberto Contestabile[3], Miłosz Ruszkowski[1]*

[1]Department of Structural Biology of Eukaryotes, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznan, Poland
[2]Research Institute on Terrestrial Ecosystems, National Research Council, Monterotondo Scalo, Rome, Italy
[3]Istituto Pasteur Italia-Fondazione Cenci Bolognetti and Dipartimento di Scienze Biochimiche "A. Rossi Fanelli", Sapienza Università di Roma, Rome, Italy

*To whom correspondence should be addressed:
Milosz Ruszkowski, Department of Structural Biology of Eukaryotes, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznan, Poland; mruszkowski@ibch.poznan.pl

Supplementary Figures:

S1 Kinetic analysis of phenylalanine binding by MtHISN6.
S2 Chemical formula and binding energy of top 95 results of Virtual Screening.
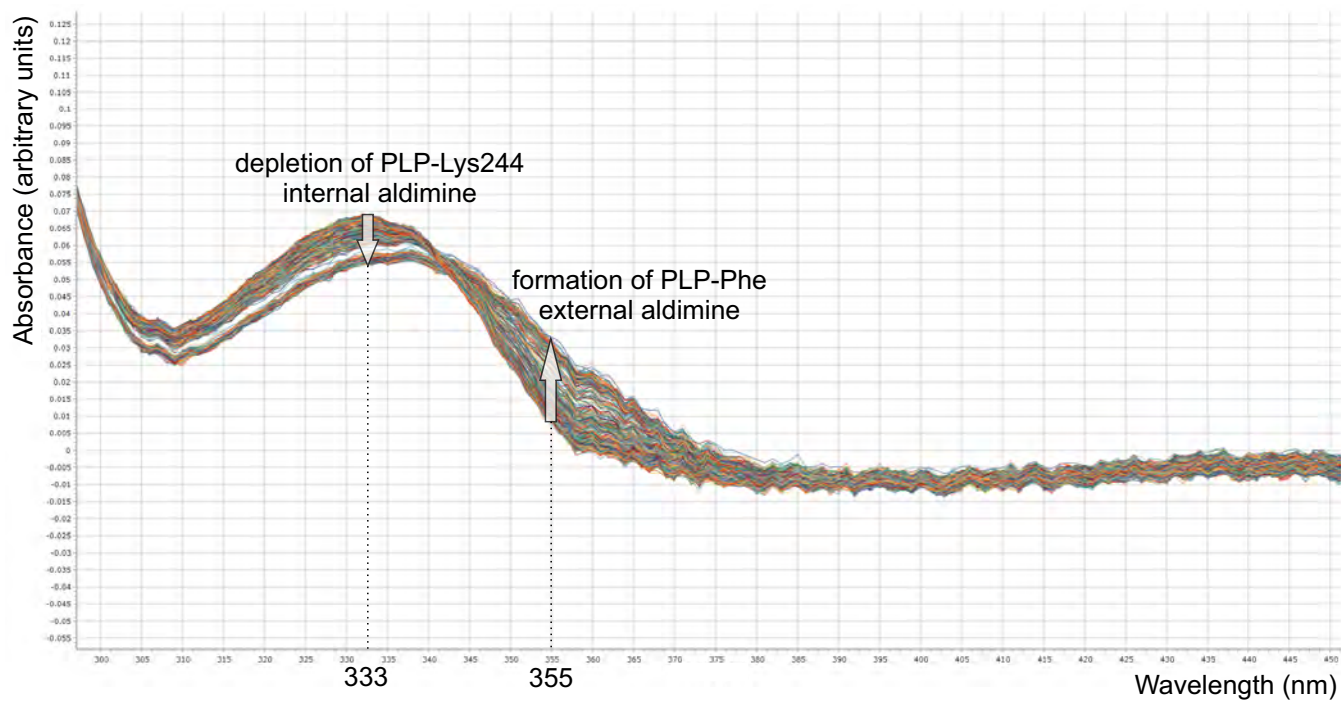
Supplementary Figure S1. Kinetic analysis of phenylalanine binding by MtHISN6. Formation of the external aldimine, was monitored by absorbance increase at around 355 nm for 4 min. The assay was performed in the mixture containing 25 mM HEPES-NaOH pH 7.5, 100 mM KCl, 50 mM NaCl, 20 µM PLP, and 20 µM MtHISN6. Either Phe was added at 1 mM final concentration.

Supplementary Figure S2 (Page 2-4). Chemical formulas and estimated binding energies of top 95 results from Virtual Screening.
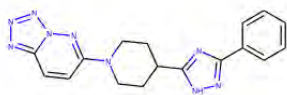
ZINC000095393759
-10.769 kcal/mol
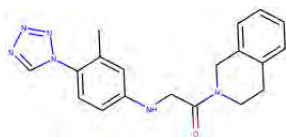
ZINC000065566055
-10.734 kcal/mol
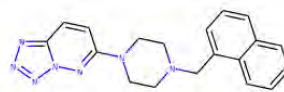
ZINC000014139069
-10.727 kcal/mol
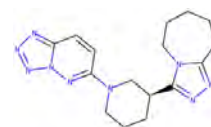
ZINC000011085217
-10.661 kcal/mol
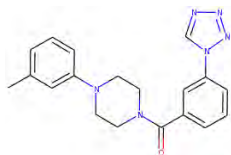
ZINC000033238693
-10.606 kcal/mol
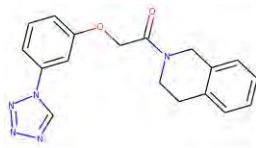
ZINC000005673503
-10.605 kcal/mol
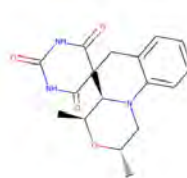
ZINC000009009723
-10.605 kcal/mol
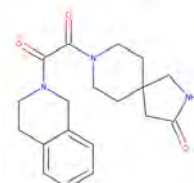
ZINC000067674282
-10.571 kcal/mol
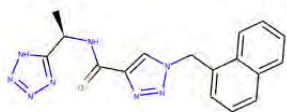
ZINC000072140159
-10.557 kcal/mol
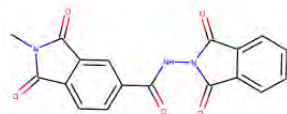
ZINC000005013760
-10.537 kcal/mol
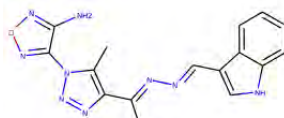
ZINC000000061971
-10.479 kcal/mol
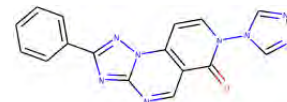
ZINC000012323188
-10.477 kcal/mol
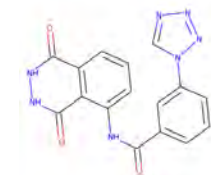
ZINC000072275309
-10.471 kcal/mol
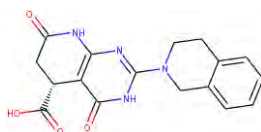
ZINC000019961288
-10.465 kcal/mol
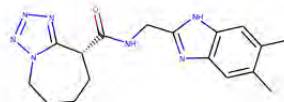
ZINC000584897329
-10.459 kcal/mol
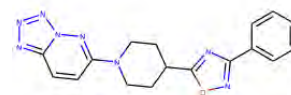
ZINC000095350149
-10.444 kcal/mol
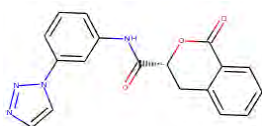
ZINC000120437127
-10.422 kcal/mol
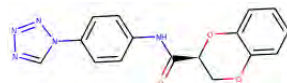
ZINC000000412886
-10.421 kcal/mol
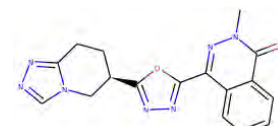
ZINC000089741012
-10.385 kcal/mol
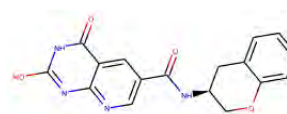
ZINC000053047092
-10.38 kcal/mol
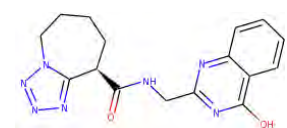
ZINC000299793701
-10.374 kcal/mol
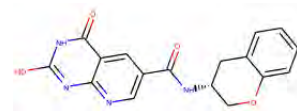
ZINC000053047094
-10.374 kcal/mol
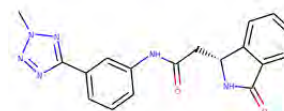
ZINC000079486951
-10.368 kcal/mol
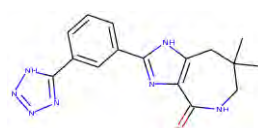
ZINC000067975156
-10.364 kcal/mol
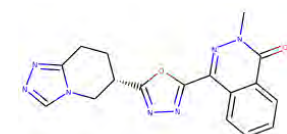
ZINC000089741013
-10.353 kcal/mol
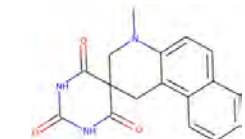
ZINC000009015982
-10.35 kcal/mol
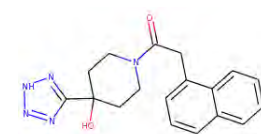
ZINC000585139781
-10.338 kcal/mol
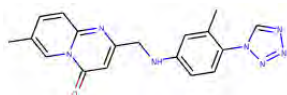
ZINC000012521901
-10.336 kcal/mol
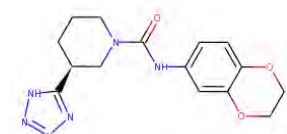
ZINC000257192291
-10.334 kcal/mol
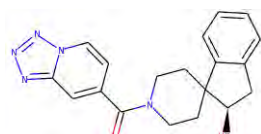
ZINC000072152603
-10.333 kcal/mol
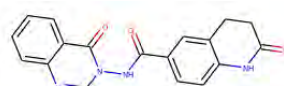
ZINC000032841367
-10.328 kcal/mol
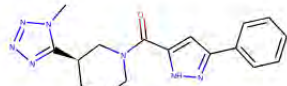
ZINC000257184561
-10.323 kcal/mol

2
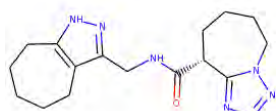
ZINC000584902046
-10.323 kcal/mol
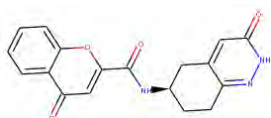
ZINC000238863626
-10.311 kcal/mol
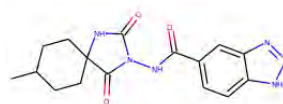
ZINC000012530714
-10.296 kcal/mol
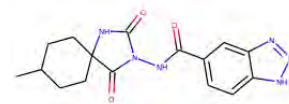
ZINC000079484287
-10.29 kcal/mol
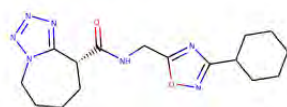
ZINC000299777609
-10.289 kcal/mol
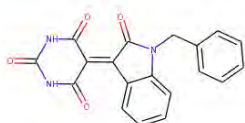
ZINC000005603208
-10.275 kcal/mol
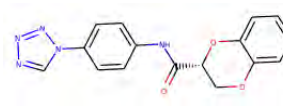
ZINC000000412887
-10.275 kcal/mol
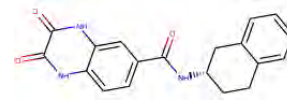
ZINC000070753421
-10.271 kcal/mol
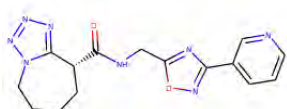
ZINC000299780538
-10.271 kcal/mol
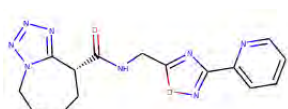
ZINC000299803011
-10.271 kcal/mol
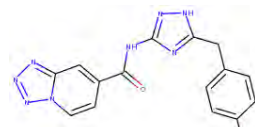
ZINC000095475137
-10.269 kcal/mol
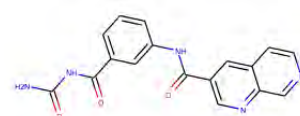
ZINC000338752544
-10.248 kcal/mol
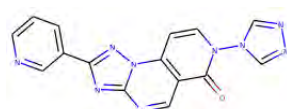
ZINC000012321580
-10.248 kcal/mol
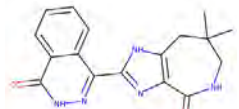
ZINC000101991666
-10.247 kcal/mol
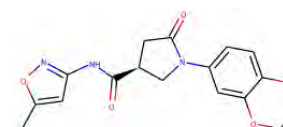
ZINC000040467596
-10.246 kcal/mol
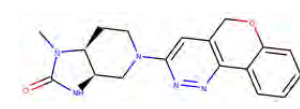
ZINC000863949062
-10.245 kcal/mol
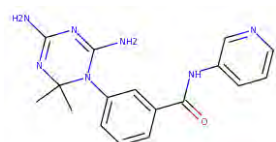
ZINC000001716762
-10.24 kcal/mol
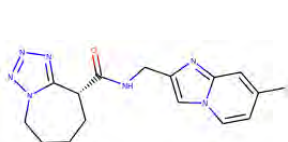
ZINC000299761345
-10.238 kcal/mol
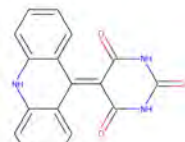
ZINC000223812065
-10.237 kcal/mol
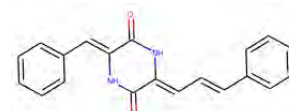
ZINC000104203224
-10.237 kcal/mol
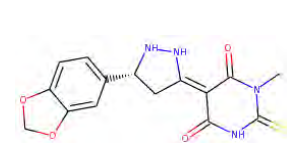
ZINC000013552790
-10.234 kcal/mol
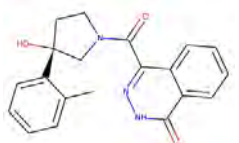
ZINC000219921959
-10.23 kcal/mol
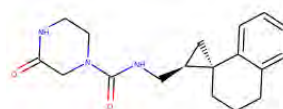
ZINC000095414421
-10.229 kcal/mol
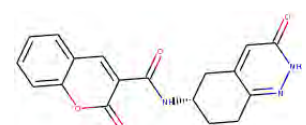
ZINC000238818218
-10.227 kcal/mol
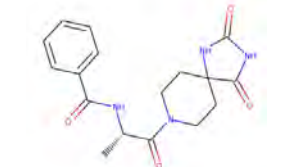
ZINC000040485334
-10.225 kcal/mol
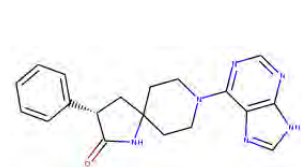
ZINC000426383995
-10.221 kcal/mol
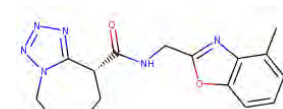
ZINC000299781159
-10.22 kcal/mol
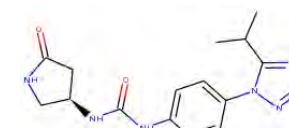
ZINC000408514726
-10.219 kcal/mol
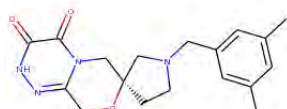
ZINC000306131692
-10.218 kcal/mol
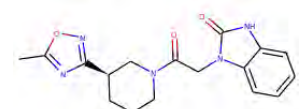
ZINC000072232293
-10.214 kcal/mol
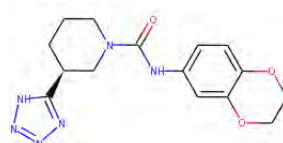
ZINC000257192291
-10.214 kcal/mol
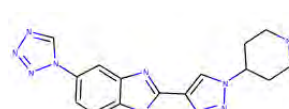
ZINC000091847748
-10.214 kcal/mol

3
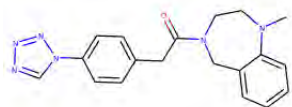
ZINC000068859472
-10.213 kcal/mol
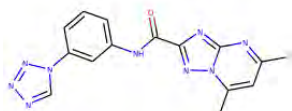
ZINC000006879180
-10.212 kcal/mol
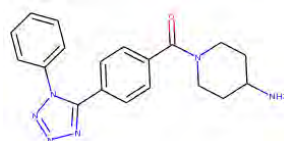
ZINC000426382918
-10.211 kcal/mol
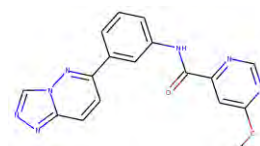
ZINC000225530229
-10.206 kcal/mol
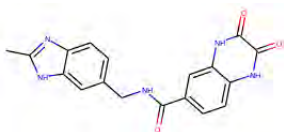
ZINC000215029558
-10.204 kcal/mol
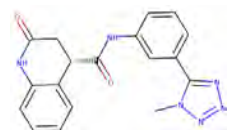
ZINC000010092434
-10.204 kcal/mol
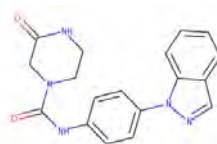
ZINC000096010636
-10.203 kcal/mol
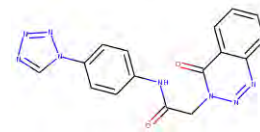
ZINC000015192914
-10.201 kcal/mol
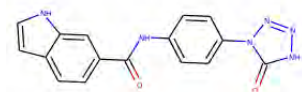
ZINC000194302053
-10.198 kcal/mol
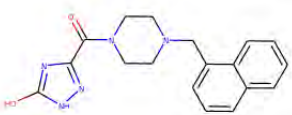
ZINC000080210896
-10.196 kcal/mol
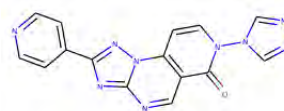
ZINC000020744776
-10.195 kcal/mol
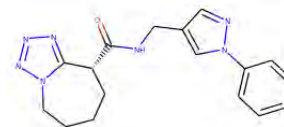
ZINC000299782575
-10.195 kcal/mol
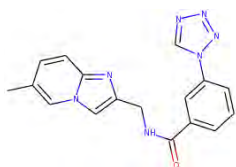
ZINC000057267836
-10.193 kcal/mol
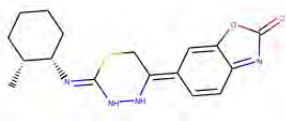
ZINC000606348184
-10.189 kcal/mol
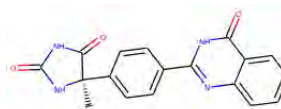
ZINC000053845962
-10.183 kcal/mol
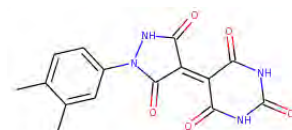
ZINC000004502149
-10.182 kcal/mol
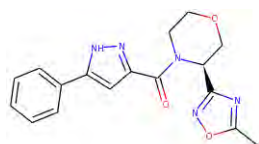
ZINC000095468258
-10.179 kcal/mol
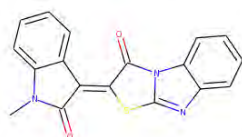
ZINC000000133083
-10.179 kcal/mol
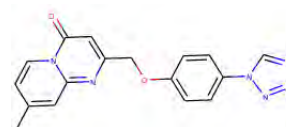
ZINC000014352345
-10.178 kcal/mol
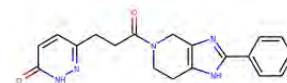
ZINC000096141842
-10.176 kcal/mol
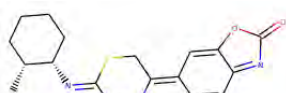
ZINC000606348184
-10.175 kcal/mol
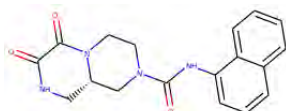
ZINC000079046209
-10.174 kcal/mol
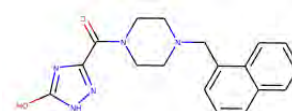
ZINC000080210896
-10.172 kcal/mol
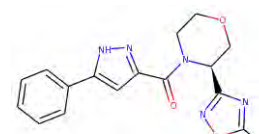
ZINC000095468258
-10.171 kcal/mol
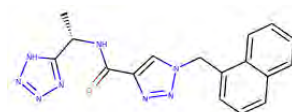
ZINC000072140158
-10.17 kcal/mol
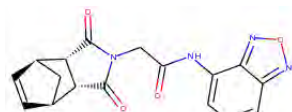
ZINC000012390796
-10.168 kcal/mol
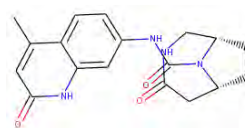
ZINC000095506432
-10.167 kcal/mol
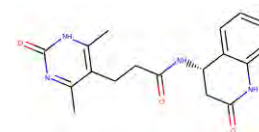
ZINC000216164490
-10.167 kcal/mol
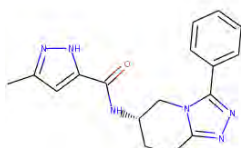
ZINC000095998152
-10.166 kcal/mol
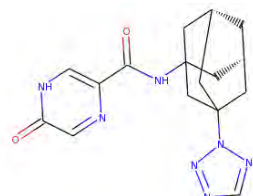
ZINC000670464124
-10.165 kcal/mol
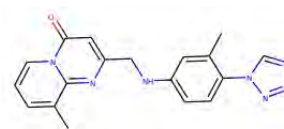
ZINC000044908085
-10.165 kcal/mol

4

September 2, 2024

Wojciech Witek, MSc
Department of Structural Biology of Eukaryotes
Institute of Bioorganic Chemistry
Polish Academy of Sciences
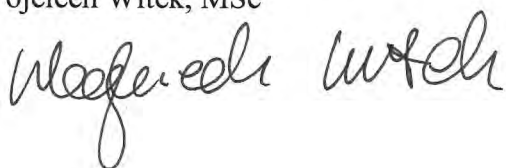wwitek@ibch.poznan.pl
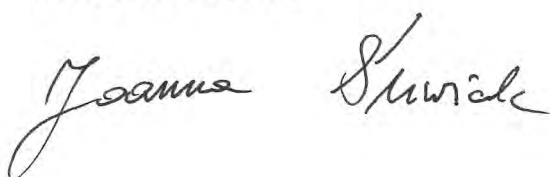
To whom it may concern:
Regarding my doctoral defense

This letter certifies my contribution to the following publication, for which I was the first author:

Witek Wojciech, Śliwiak Joanna, Ruszkowski Miłosz (2021) Structural and mechanistic insights into the bifunctional HISN2 enzyme catalyzing the second and third steps of histidine biosynthesis in plants, Scientific Reports 11(1):9647, https://doi.org/10.1038/s41598-021-88920-2

My contributions were to: refine and analyze crystal structures, design mutants and conduct kinetic measurements, conduct *in silico* analyses, prepare figures and write the manuscript

Wojciech Witek, MSc

September 2, 2024

Joanna Śliwiak, PhD
Laboratory of Protein Engineering
Institute of Bioorganic Chemistry
Polish Academy of Sciences
sliwiak@ibch.poznan.pl

To whom it may concern:
Regarding the doctoral defense of Wojciech Witek, MSc

This letter certifies my contribution to the following publication, for which I was a co-author:

Witek Wojciech, Śliwiak Joanna, Ruszkowski Miłosz (2021) Structural and mechanistic insights into the bifunctional HISN2 enzyme catalyzing the second and third steps of histidine biosynthesis in plants, Scientific Reports 11(1):9647, https://doi.org/10.1038/s41598-021-88920-2

My contributions were to: perform and analyze ITC experiments

Joanna Śliwiak, PhD

September 2, 2024

Miłosz Ruszkowski, dr. habil.
Head of Department of Structural Biology of Eukaryotes
Institute of Bioorganic Chemistry
Polish Academy of Sciences
mruszkowski@ibch.poznan.pl

To whom it may concern:
<u>Regarding the doctoral defense of Wojciech Witek, MSc</u>

This letter certifies my contribution to the following publication, for which I was the corresponding author:

Witek Wojciech, Śliwiak Joanna, Ruszkowski Miłosz (2021) Structural and mechanistic insights into the bifunctional HISN2 enzyme catalyzing the second and third steps of histidine biosynthesis in plants, Scientific Reports 11(1):9647, https://doi.org/10.1038/s41598-021-88920-2

My contributions were to: design experiments, obtain crystals, conduct kinetic measurements, solve structures, and edit the text

Miłosz Ruszkowski, dr. habil

August 30, 2024

Wojciech Witek, MSc
Department of Structural Biology of Eukaryotes
Institute of Bioorganic Chemistry
Polish Academy of Sciences
wwitek@ibch.poznan.pl

To whom it may concern:
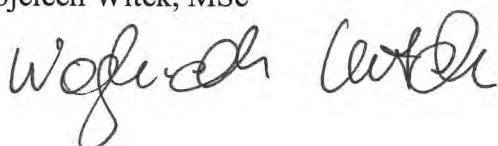Regarding my doctoral defense

This letter certifies my contribution to the following publication, for which I was the first author:

Witek W., Imiołczyk B., Ruszkowski M. (2024) *Structural, kinetic, and evolutionary peculiarities of HISN3, a plant 5'-ProFAR isomerase,* Plant Physiology and Biochemistry 215 109065

https://doi.org/10.1016/j.plaphy.2024.109065

My contributions were to: produce the proteins, measure enzymatic activities, prepare ProFAR, participate in crystallization, solve and refine crystal structures, prepare figures and write the manuscript (first and revised versions)

Wojciech Witek, MSc

August 30, 2024

Barbara Imiołczyk, PhD
Department of Structural Biology of Eukaryotes
Institute of Bioorganic Chemistry
Polish Academy of Sciences
barim@ibch.poznan.pl

To whom it may concern:
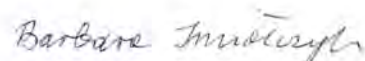Regarding the doctoral defense of Wojciech Witek, MSc

This letter certifies my contribution to the following publication, for which I was a co-author:

Witek W., Imiołczyk B., Ruszkowski M. (2024) *Structural, kinetic, and evolutionary peculiarities of HISN3, a plant 5'-ProFAR isomerase,* Plant Physiology and Biochemistry 215 109065

https://doi.org/10.1016/j.plaphy.2024.109065

My contribution :

x) optimization of expression and purification of HISN3 resulting in pure, stable protein and best diffracting crystals,

x) production of Mt histidine biosynthetic pathway gene products: HISN1, HISN2, HISN3 and E.coli pyrophosphatase with and without Histag used in multistep enzymatic reaction generating substrate for HISN3,

x) proposing simple and efficient method of HISN3 substrate isolation using in multistep enzymatic reaction proteins with Histag and after the reaction retaining them on NI-NTA column. The substrate is present in the flow through,

x) Production of HISN3 protein for enzymatic reaction kinetics.

Barbara Imiołczyk

Barbara Imiołczyk, PhD

August 30, 2024

Miłosz Ruszkowski, dr. habil.
Head of Department of Structural Biology of Eukaryotes
Institute of Bioorganic Chemistry
Polish Academy of Sciences
mruszkowski@ibch.poznan.pl

To whom it may concern:
<u>Regarding the doctoral defense of Wojciech Witek, MSc</u>

This letter certifies my contribution to the following publication, for which I was the corresponding author:

Witek W., Imiołczyk B., Ruszkowski M. (2024) *Structural, kinetic, and evolutionary peculiarities of HISN3, a plant 5'-ProFAR isomerase,* Plant Physiology and Biochemistry 215 109065

https://doi.org/10.1016/j.plaphy.2024.109065

My contributions were to: obtain funding, conceptualize and supervise all the work, edit the manuscript (the first and revised versions)

(signature)

Miłosz Ruszkowski, dr. habil

July 1, 2024

Wojciech Witek, MSc
Department of Structural Biology of Eukaryotes
Institute of Bioorganic Chemistry
Polish Academy of Sciences
wwitek@ibch.poznan.pl

To whom it may concern:
<u>Regarding my doctoral defense</u>

This letter certifies my contribution to the following publication, for which I was the first author:

Witek W., Śliwiak J., Rawski M., Ruszkowski M. (2024) *Targeting imidazole-glycerol phosphate dehydratase in plants: novel approach for structural and functional studies, and inhibitor blueprinting,* Frontiers in Plant Science 15;15:1343980,
https://doi.org/10.3389/fpls.2024.1343980

My contributions were to: produce and purify the enzyme, obtain crystals, prepare samples for cryoEM, solve and refine crystal structures and HISN5- IG2 cryoEM structure, analyze all structures, produce IGP, conduct phylogenetic studies, prepare figures and write the manuscript

Wojciech Witek, MSc

July 1, 2024

Joanna Śliwiak, PhD
Laboratory of Protein Engineering
Institute of Bioorganic Chemistry
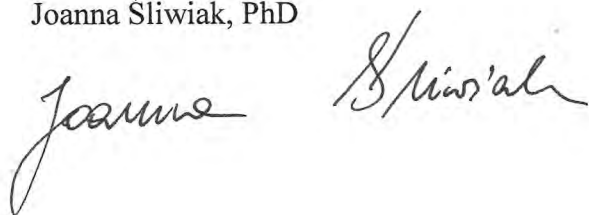Polish Academy of Sciences
sliwiak@ibch.poznan.pl

To whom it may concern:
Regarding the doctoral defense of Wojciech Witek, MSc

This letter certifies my contribution to the following publication, for which I was a co-author:

Witek W., Śliwiak J., Rawski M., Ruszkowski M. (2024) *Targeting imidazole-glycerol phosphate dehydratase in plants: novel approach for structural and functional studies, and inhibitor blueprinting,* Frontiers in Plant Science 15;15:1343980, https://doi.org/10.3389/fpls.2024.1343980

My contributions were to: perform and analyze ITC experiments

Joanna Śliwiak, PhD

July 1, 2024

Michał Rawski, PhD
Cryo-EM Facility
SOLARIS National Synchrotron Radiation Centre
michal.rawski@uj.edu.pl

To whom it may concern:
<u>Regarding the doctoral defense of Wojciech Witek, MSc</u>

This letter certifies my contribution to the following publication, for which I was a co-author:

Witek W., Śliwiak J., Rawski M., Ruszkowski M. (2024) *Targeting imidazole-glycerol phosphate dehydratase in plants: novel approach for structural and functional studies, and inhibitor blueprinting,* Frontiers in Plant Science 15;15:1343980,
https://doi.org/10.3389/fpls.2024.1343980

My contributions were to: vitrify sample grids and collect cryoEM data

Michał Rawski, PhD

July 1, 2024

Miłosz Ruszkowski, dr. habil.
Head of Department of Structural Biology of Eukaryotes
Institute of Bioorganic Chemistry
Polish Academy of Sciences
mruszkowski@ibch.poznan.pl

To whom it may concern:
Regarding the doctoral defense of Wojciech Witek, MSc

This letter certifies my contribution to the following publication, for which I was the corresponding author:

Witek W., Śliwiak J., Rawski M., Ruszkowski M. (2024) *Targeting imidazole-glycerol phosphate dehydratase in plants: novel approach for structural and functional studies, and inhibitor blueprinting,* Frontiers in Plant Science 15;15:1343980,
https://doi.org/10.3389/fpls.2024.1343980

My contributions were to: obtain funding, solve and refine cryoEM structures, supervise all the work, edit the manuscript

Miłosz Ruszkowski, dr. habil

July 1, 2024

Maria Rutkiewicz, PhD
Department of Structural Biology of Eukaryotes
Institute of Bioorganic Chemistry
Polish Academy of Sciences
mrutkiewicz@ibch.poznan.pl

To whom it may concern:
Regarding the doctoral defense of Wojciech Witek, MSc

This letter certifies my contribution to the following publication, for which I was the first author:

Rutkiewicz Maria, Nogues Isabel, Witek Wojciech, Angelaccio Sebastiana, Contestabile Roberto, Ruszkowski Miłosz (**2023**) Insights into the substrate specificity, structure, and dynamics of plant histidinol-phosphate aminotransferase (HISN6), *Plant Physiology and Biochemistry* 196:759-773, https://doi.org/10.1016/j.plaphy.2023.02.017

My contributions were to: crystallize *Mt*HISN6 in the open and closed states, crystallize the *Mt*HISN6_HOLP complex, refine and analyze all structures, prepare relevant figures and write the manuscript

Maria Rutkiewicz

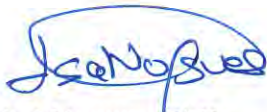Maria Rutkiewicz, PhD

July 1, 2024

Isabel Nogues, PhD
Research Institute on Terrestrial Ecosystems
National Research Council
isabel.nogues@ibaf.cnr.it

To whom it may concern:
Regarding the doctoral defense of Wojciech Witek, MSc

This letter certifies my contribution to the following publication, for which I was a co-author:

Rutkiewicz Maria, Nogues Isabel, Witek Wojciech, Angelaccio Sebastiana, Contestabile Roberto, Ruszkowski Miłosz (**2023**) Insights into the substrate specificity, structure, and dynamics of plant histidinol-phosphate aminotransferase (HISN6), *Plant Physiology and Biochemistry* 196:759-773, https://doi.org/10.1016/j.plaphy.2023.02.017

My contributions were to: perform the kinetic measurements and analyze the data

Isabel Nogues, PhD

July 1, 2024

Wojciech Witek, MSc
Department of Structural Biology of Eukaryotes
Institute of Bioorganic Chemistry
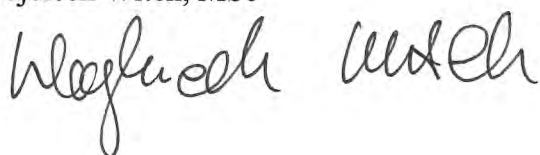Polish Academy of Sciences
wwitek@ibch.poznan.pl

To whom it may concern:
Regarding doctoral defense of Wojciech Witek, MSc

This letter is to certify my contribution to the following publication of which I was a co-author.

Rutkiewicz Maria, Nogues Isabel, Witek Wojciech, Angelaccio Sebastiana, Contestabile Roberto, Ruszkowski Miłosz (**2023**) Insights into the substrate specificity, structure, and dynamics of plant histidinol-phosphate aminotransferase (HISN6), *Plant Physiology and Biochemistry* 196:759-773, https://doi.org/10.1016/j.plaphy.2023.02.017

My contributions were to: obtain the apo*Mt*HISN6 structure and analyze the HISN6 evolution, write the introduction and evolution sections
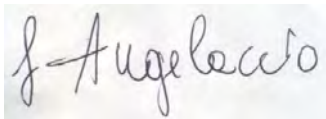
Wojciech Witek, MSc

July1, 2024

Sebastiana Angelaccio, PhD
Dipartimento di Scienze Biochimiche "A. Rossi Fanelli"
Istituto Pasteur Italia-Fondazione Cenci Bolognetti
Sapienza Università di Roma
sebastiana.angelaccio@uniroma1.it

To whom it may concern:
<u>Regarding the doctoral defense of Wojciech Witek, MSc</u>

This letter certifies my contribution to the following publication, for which I was a co-author:

Rutkiewicz Maria, Nogues Isabel, Witek Wojciech, Angelaccio Sebastiana, Contestabile Roberto, Ruszkowski Miłosz (**2023**) Insights into the substrate specificity, structure, and dynamics of plant histidinol-phosphate aminotransferase (HISN6), *Plant Physiology and Biochemistry* 196:759-773, https://doi.org/10.1016/j.plaphy.2023.02.017

My contributions were to: perform the kinetic measurements

Sebastiana Angelaccio, PhD

July 1, 2024

Roberto Contestabile, PhD
Dipartimento di Scienze Biochimiche "A. Rossi Fanelli"
Istituto Pasteur Italia-Fondazione Cenci Bolognetti
Sapienza Universita di Roma
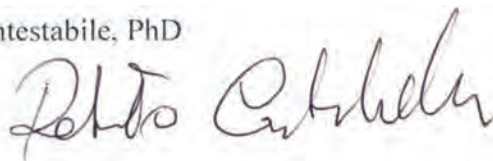roberto.contestabile@uniromal.it

To whom it may concern:
Regarding the doctoral defense of Wojciech Witek, MSc

This letter certifies my contribution to the following publication, for which I was a co-author:

Rutkiewicz Maria, Nogues Isabel, Witek Wojciech, Angelaccio Sebastiana, Contestabile Roberto, Ruszkowski Miłosz (2023) Insights into the substrate specificity, structure, and dynamics of plant histidinol-phosphate aminotransferase (HISN6), *Plant Physiology and Biochemistry* 196:759-773, https://doi.org/10.1016/j.plaphy.2023.02.017

My contributions were to: analyze the kinetic data

Roberto Contestabile, PhD

July 1, 2024

Miłosz Ruszkowski, dr. habil.
Head of Department of Structural Biology of Eukaryotes
Institute of Bioorganic Chemistry
Polish Academy of Sciences
mruszkowski@ibch.poznan.pl

To whom it may concern:
Regarding the doctoral defense of Wojciech Witek, MSc

This letter certifies my contribution to the following publication, for which I was the corresponding author:

Rutkiewicz Maria, Nogues Isabel, Witek Wojciech, Angelaccio Sebastiana, Contestabile Roberto, Ruszkowski Miłosz (**2023**) Insights into the substrate specificity, structure, and dynamics of plant histidinol-phosphate aminotransferase (HISN6), *Plant Physiology and Biochemistry* 196:759-773, https://doi.org/10.1016/j.plaphy.2023.02.017

My contributions were to: obtain other HISN6 structures, write parts of the manuscript, edit the text

Miłosz Ruszkowski, dr. habil