JOÃO VICTOR BADARÓ DE MORAES

STRUCTURAL MODELING OF E-NTPDase-SUBSTRATE COMPLEXES PRESERVING CATALYTIC EXPERIMENTAL FEATURES

Dissertation submitted to the Cell and Structural Biology Graduate Program of the Universidade Federal de Viçosa in partial fulfillment of the requirements for the degree of *Magister Scientiae*.

Adviser: Juliana Lopes Rangel Fietto

Co-adviser: Marcelo Depólo Polêto

VIÇOSA - MINAS GERAIS
2024

Ficha catalográfica elaborada pela Biblioteca Central da Universidade Federal de Viçosa - Campus Viçosa

T
M827n
2024

Moraes, João Victor Badaró de, 2024-
Structural modelling of E-NTPDase-substrate complexes preserving catalytic experimental features / João Victor Badaró de Moraes. – Viçosa, MG, 2024.
1 dissertação eletrônica (50 f.): il. (algumas color.).

Texto em inglês.
Inclui apêndices.
Orientador: Juliana Lopes Rangel Fietto.
Dissertação (mestrado) - Universidade Federal de Viçosa, Departamento de Biologia Geral, 2024.
Referências bibliográficas: f. 40-44.
DOI: https://doi.org/10.47328/ufvbbt.2024.066
Modo de acesso: World Wide Web.

1. Enzimas. 2. Reações químicas. 3. Fosfatos. I. Fietto, Juliana Lopes Rangel, 1971-. II. Universidade Federal de Viçosa. Departamento de Biologia Geral. Programa de Pós-Graduação em Biologia Celular e Estrutural. III. Título.

CDD 22. ed. 572.78

Bibliotecário(a) responsável: Euzébio Luiz Pinto CRB-6/3317

JOÃO VICTOR BADARÓ DE MORAES

STRUCTURAL MODELING OF E-NTPDase-SUBSTRATE COMPLEXES PRESERVING CATALYTIC EXPERIMENTAL FEATURES

Dissertation submitted to the Cell and Structural Biology Graduate Program of the Universidade Federal de Viçosa in partial fulfillment of the requirements for the degree of *Magister Scientiae*.

APPROVED: February 29, 2024.

Assent:

Documento assinado digitalmente
gov.br
JOAO VICTOR BADARO DE MORAES
Data: 19/07/2024 11:10:17-0300
Verifique em https://validar.iti.gov.br

João Victor Badaró de Moraes
Author

Documento assinado digitalmente
gov.br
JULIANA LOPES RANGEL FIETTO
Data: 19/07/2024 11:32:01-0300
Verifique em https://validar.iti.gov.br

Juliana Lopes Rangel Fietto

Adviser

I dedicate this work to my mother and loved ones.

ACKNOWLEDGEMENTS

To my mother, Sra. Marilza Ricardina Badaró, who supported and encouraged me to pursue the formal education that she never had the opportunity to have. As a single mother, you sacrificed many things in your life to provide for me. For that, I will never be able to thank you enough. I also would like to thank my family for always being there for me.

To my beloved partner, Raissa Barbosa de Castro, for all the support. This journey was much easier with you by my side. Thank you for your patience, understanding, and unwavering belief in me. I hope to be as much of a positive force in your life as you are in mine.

To my adviser, Dra. Juliana Lopes Rangel Fietto, for believing in me and providing me with the opportunity to learn and grow as a researcher in her laboratory over the past six years.

To my co-adviser, Dr. Marcelo Depólo Polêto, for introducing me to the world of Molecular Dynamics simulations and for all the support and attention during this period.

To my friends and colleagues from the Labaratório de Infectologia Molecular Animal and the Laboratório de Biologia Celular Bioprodudos I and II, thank you for the wonderful moments we have shared over these years. A special thanks to Dra. Nancy Pavione, working with you was an enriching experience.

Thank you to the Americal Kennel Club for granting the scholarship. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001.

To the Funtadação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG), for the financial suport.

Thank you to the Instituto de Biotecnologia Aplicada à Agropecuária (BIOAGRO) for providing a proper space to develop this research.

ABSTRACT

BADARÓ DE MORAES, João Victor, M.Sc., Universidade Federal de Viçosa, February, 2024. **Structural modeling of E-NTPDase-substrate complexes preserving catalytic experimental features**. Adviser: Juliana Lopes Rangel Fietto. Co-advisers: Marcelo Depólo Polêto.

The Ecto-Nucleoside Triphosphate Diphosphohydrolase (E-NTPDase or NTPDase) family, essential for converting tri- and diphosphate nucleotides to monophosphates, significantly influences purinergic/pyrimidinergic signaling. However, detailed structural insights, particularly for NTPDase-nucleotide complexes, are scarce. The limited structural information leads to challenges in rationally optimizing and exploring the therapeutic potential of these enzymes. Addressing this limitation, this study introduces a computational strategy for assembling NTPDase-substrate complexes. While Molecular Docking may serve this purpose, limitations such as the inability to precisely reproduce experimentally characterized ligand conformations can undermine the complexes' reliability and further computational studies relying on them. Leveraging the active site's high conservation across NTPDases, we hypothesized a consistent substrate binding conformation. Pursuing that, we structurally aligned the available experimental structures in substrate-bound productive states and identified a canonical substrate conformation prevalent in the majority of experimentally resolved complexes. We also observed other general features of the complexes, which were conserved in all the characterized experimental structures. In light of these conserved features, we proposed a method for modeling NTPDase-nucleotide complexes that are compliant with the available experimental data. These complexes can be modeled by transferring crystallized substrate structures to well-modeled NTPDase structures, also carrying their cofactors and relevant water molecules. Subsequent energy minimization and equilibration through Molecular Dynamics simulations yield a final structure that closely resembles the conserved features characterized in experimental structures. We demonstrated this approach's utility by modeling complexes for each human NTPDase with ATP, ADP, GTP, GDP, UTP, and UDP, presenting a novel methodological avenue for future research.

Keywords: E-NTPDase; Computational Modeling; Enzyme-Substrate Complex.

RESUMO

BADARÓ DE MORAES, João Victor, M.Sc., Universidade Federal de Viçosa, fevereiro, 2024. **Modelagem estrutural de complexos E-NTPDase-substrato preservando características experimentais catalíticas**. Orientadora: Juliana Lopes Rangel Fietto. Co-orientador: Marcelo Depólo Polêto.

A família Ecto-Nucleosídeo Trifosfato Difosfohidrolase (E-NTPDase ou NTPDase), essencial para converter nucleotídeos tri e difosfato em monofosfatos, influencia significativamente a sinalização purinérgica/pirimidinérgica. No entanto, informações estruturais detalhadas, particularmente para complexos NTPDase-nucleotídeo, são escassas. A limitada informação estrutural leva a desafios na otimização racional e na exploração do potencial terapêutico dessas enzimas. Abordando essa limitação, este estudo introduz uma estratégia computacional para montar complexos substrato-NTPDase. Embora o Docking Molecular possa servir a esse propósito, sua incapacidade de reproduzir precisamente conformações de ligantes caracterizadas experimentalmente limitam a confiabilidade dos complexos formados. Aproveitando a alta conservação do sítio ativo entre as NTPDases, hipotetizamos uma conformação de ligação de substrato consistente. Com isso, alinhamos estruturalmente as estruturas experimentais disponíveis em estados produtivos com substrato ligado e identificamos uma conformação de substrato canônica, prevalente na maioria dos complexos experimentalmente resolvidos. Também observamos outras características gerais dos complexos, que foram conservadas em todas as estruturas experimentais caracterizadas. À luz dessas características conservadas, propusemos um método para a modelagem de complexos NTPDase-nucleotídeo que estejam em conformidade com os dados experimentais disponíveis. Esses complexos podem ser modelados transferindo estruturas de substrato cristalizadas para estruturas de NTPDase bem modeladas, carregando também seus cofatores e moléculas de água relevantes. A minimização de energia subsequente e a equilibração por meio de simulações de Dinâmica Molecular produzem uma estrutura final que se assemelha estreitamente às características conservadas caracterizadas em estruturas experimentais. Demonstramos a utilidade dessa abordagem modelando complexos para cada NTPDase humana com ATP, ADP, GTP, GDP, UTP e UDP, apresentando uma nova via metodológica para pesquisas futuras.

Palavras-chave: E-NTPDase; Modelagem Molecular; Complexo Enzima-Substrato.

SUMMARY

Structural Modeling of E-NTPDase-Substrate Complexes Preserving Catalytic Experimental Features

João Victor Badaró de Moraes[1], Marcelo Depólo Polêto[2], Raissa Barbosa de Castro[3], Gustavo Costa Bressan[3], Raphael de Souza Vasconcellos[3], Juliana Rangel Fietto[3,4*]

[1] General Biology Department, Universidade Federal de Viçosa, Viçosa, Minas Gerais, Brazil.

[2] Department of Biochemistry, Virginia Tech, Blacksburg, Virginia, USA.

[3] Biochemistry and Molecular Biology Department, Universidade Federal de Viçosa, Viçosa, Minas Gerais, Brazil.

[4] Instituto de Biotecnologia Aplicada à Agropecuária (Bioagro), Universidade Federal de Viçosa, Viçosa, Minas Gerais, Brazil.

***Corresponding author:** Juliana Lopes Rangel Fietto

Phone: +55 (31) 3612-2464

E-mail: jufietto@ufv.br

Abstract

The Ecto-Nucleoside Triphosphate Diphosphohydrolase (E-NTPDase or NTPDase) family, essential for converting tri- and diphosphate nucleotides to monophosphates, significantly influences purinergic/pyrimidinergic signaling. However, detailed structural insights, particularly for NTPDase-nucleotide complexes, are scarce. The limited structural information leads to challenges in rationally optimizing and exploring the therapeutic potential of these enzymes. Addressing this limitation, this study introduces a computational strategy for assembling NTPDase-substrate complexes. While Molecular Docking may serve this purpose, the inability to precisely reproduce experimentally characterized ligand conformations limits its use and reliability. Leveraging the active site's high conservation across NTPDases, we hypothesized a consistent substrate binding conformation. Pursuing that, we structurally aligned the available experimental structures in substrate-bound productive states and identified a canonical substrate conformation prevalent in the majority of the complexes. We also observed other general features conserved in all the characterized experimental structures. Thus, we proposed a method for modeling NTPDase-nucleotide complexes compliant with the available experimental data that comprises the transferring of crystallized substrate structures to well-modeled NTPDase structures, also carrying their cofactors and relevant water molecules. Subsequent energy minimization and equilibration through Molecular Dynamics simulations yield a final structure that closely resembles the conserved features characterized in experimental structures. We demonstrated this approach's utility by modeling complexes for each human NTPDase with ATP, ADP, GTP, GDP, UTP, and UDP, presenting a novel methodological avenue for future research.

Keywords: E-NTPDase, Computational Modeling, Enzyme-Substrate Complex.

1. Introduction

The Ecto Nucleoside triphosphate diphosphohydrolase (E-NTPDases or NTPDases) enzyme family plays a crucial role in hydrolyzing triphosphate (NTPs) and diphosphate (NDPs) nucleotides into their monophosphate forms (NMPs). Together with the 5'-nucleotidase, the E-NTPDases modulate the purinergic/pyrimidinergic signaling pathway by influencing the balance between nucleotide forms and free nucleosides in the cellular environment [1–3]. This family has been described in multiple organisms with a wide range of functions, from virulence factors essential for trypanosomatids nutrition to a critical factor for immune response and tumor regulation [4–8]. Despite their importance, there is still much to learn about the family's structure and function.

More than 15 years ago, the first E-NTPDase structure was experimentally resolved, providing insight into the function of active site residues, substrate binding mode, and catalytic mechanism [9]. Since then, a total of 48 structures from seven different organisms have been reported, consolidating the knowledge of the general structural features of the family. However, only 15 of these structures were resolved in a productive bound state with substrate analogs [9–12]. Among the eight human E-NTPDase isoforms (HsNTPDase1-8), HsNTPDase4 is the only one with a resolved structure [13]. Unfortunately, only the apo form of the enzyme could be captured.

Understanding the NTPDase's structure and interactions with different substrates can be crucial for rationalizing the enzyme's specificities and to explore their possible biotechnological applications. In this scenario, computational techniques can offer valuable insights into enzyme-substrate complexes and aid studies in understanding the structure/function relationship [14].

Although not replacing experimental structures, AlphaFold 2 enables the production of high-quality protein models [15, 16]. However, accurately describing an NTPDase-substrate complex depends not only on a reliable protein structure but also on the positioning of the substrate molecule in the active site. Even though Molecular Docking software may serve this purpose, these programs aim to balance accuracy and speed. Therefore, deviations in reproducing experimentally characterized complexes < 2 Å are accepted and used as an efficiency metric when applying such a method [17, 18]. As we showcase in the supplementary material with the Re-Docking of ATP and ADP analogs co-crystalized with the Rattus norvegicus NTPDase2 (PDB code: 3CJA and 4BR0, respectively), imprecisions within this range of variation can result in unusual substrate conformation that ultimately constrains the biological relevance of the generated complexes (Figure S1). To overcome this limitation, we may base the description of such complexes on the available experimental data, thus producing complexes that fit the conserved features observed in multiple crystal structures.

Based on a careful examination of multiple NTPDase experimentally resolved structures in a productive bound state, we propose transferring the experimentally resolved substrate-cofactor complex and the relevant water molecules to a modeled NTPDase structure as an alternative method for describing unknown NTPDase-substrate complexes by computational means. To illustrate our approach, we built and described complexes of each HsNTPDase with multiple substrates. Herein, we share the outcomes of our study and encourage other researchers to apply this strategy in their investigations.

2. Methods

2.1. Co-crystalized substrate structural alignment and analysis

Considering the Adenosine-5'-[(β, γ)-imido]triphosphate (AMP-PNP) ATP analog and the Adenosine-5'-[(α, β)-imido]diphosphate (AMP-NP) ADP analog, co-crystalized with the RnNTPDase2 (PDB code: 3CJA and 4BR0, respectively) as reference structures for triphosphate and diphosphate nucleotides upon binding to an NTPDase, respectively, we calculate the Root Mean Square Deviation (RMSD) between pairs of structures using the PyMOL Molecular Graphics System Version 4.6 Schrödinger, LLC. The best-fitted RMSD value was considered. As summarized in Supplementary Table 1, structures of the *Rattus norvegicus* NTPDase2 [9, 12], *Legionella pneumophila* NTPDase1 [12], and *Toxoplasma gondii* NTPDase1 and 3 [10, 11] retrieved from the Protein Data Bank (PDB) were considered in the analysis of substrate binding mode [19]. Distance calculations were also performed using PyMOL.

2.2. HsNTPDases Molecular Modeling

The reference sequence of each HsNTPDase used in the modeling process was obtained from the UniProt database [20]. To identify the transmembrane or signal peptide regions in the sequence, we used the DeepTMHMM web server [21]. Supplementary Table 1 compiles the UniProt entry code and the modeled sequence regions of each HsNTPDase1-8. The structures of the HsNTPDase1-8 ectodomains were built using the AlphaFold2 engine through the ColabFold AlphaFold2_mmseqs2 Jupyter notebook platform with default settings and using the RnNTPDase2 (PDB code: 3CJA) as a custom template [15, 22]. For each enzyme, five models were generated. Following the automatic quality score classification of the platform, each enzyme's best model was selected and used in the subsequent steps. To ensure that the models

were in a suitable conformation for describing its complex with substrates, we structurally aligned each of the best models with the custom template to evaluate the similarity of the structures using PyMOL software. The overall similarity was evaluated by measuring the RMSD based on the alpha-carbon coordinates. As shown in the supplementary material (Figure S2), out of the five Apyrase Conserved Regions (ACR1-5), ACR1 and ACR4 positions drastically change if the enzyme is in the open or closed state. For this reason, the alignments were also evaluated visually in regard to the positioning of the mentioned regions.

2.3. Substrate transferring from experimental structures to modeled HsNTPDases

For each modeled enzyme, we built complexes with the primary nucleotides hydrolyzed by these enzymes: ATP, ADP, GTP, GDP, UTP, and UDP. Using PyMOL, each HsNTPDase was structurally aligned with a reference crystal structure to transfer the substrate from the crystal to the human enzymes. The ATP, ADP, GTP, and UTP substrate structures were extracted from RnNTPDase2 crystals PDB code 3CJA, 4BR0, 4BQZ, and 4BR2, respectively. GDP and UDP were manually constructed by changing the nitrogenous base information in the ADP analog on the RnNTPDase2 crystal structure. All the mentioned crystal structures were crystallized with non-hydrolyzable analogs of each substrate. To reconstruct the substrate molecule, the nitrogen N3β or N3α in the triphosphates and diphosphates were manually substituted into oxygen atoms. In all complexes, we used calcium as a cofactor. We also included the four water molecules in the calcium first hydration shell and the two water molecules predicted to have a role in the catalytic process and present um multiple crystal structures [9, 10, 12].

2.4. Molecular Dynamics energy minimization and equilibration

Before the construction of each MD simulation system, the protonation state of the enzyme-substrate complex was evaluated and adjusted using the PlayMolecule ProteinPrepare web server [23]. Using the CHARMM-GUI web server, a simulation system was created for each HsNTPDase-substrate complex [24]. The complexes were solvated with the modified TIP3P water model [25–27] and neutralized with K^+ and Cl^- to a final concentration of 0.15 M. The CHARMM36m force field was used to describe the interactions of all atoms in the simulation box, and the WYF parameters for cation-π interaction were also added [28, 29]. The Chemistry at Harvard Molecular Mechanics software (CHARMM) was employed to perform the energy minimization and equilibration [30]. Energy minimization utilized the steepest descent and the Adopted Basis Newton-Raphson methods, each for 50 steps [31, 32]. Positional restraints with a force constant of 1.0 kcal/mol/$Å^2$ were applied to the backbone and side chain atoms during this stage. Subsequently, equilibration was performed over 10,000 steps with a 1 fs integration step, employing the NVT ensemble for dynamic equilibration. For the HsNTPDase5-ATP complex, vacuum energy minimization was applied before the complex. The resulting structure was then put in solution and treated like the other complexes.

3. Results and Discussion

3.1. Substrate structure upon binding from experimental structures

The NTPDase family members possess a remarkable capability to process diverse nucleotides within a shared active site [2, 33, 34]. Given the high degree of conservation of the NTPDase active site in both sequence and structure, it is reasonable to anticipate that the conformation of substrates upon binding would exhibit a degree of similarity. The structural overlay of

experimentally resolved substrate structures characterized across various NTPDase isoforms and from varying organisms available in the Protein Data Bank (PDB) supports this expectation, revealing a consistent substrate conformation upon binding to an NTPDase. Among the 15 experimentally resolved structures representing productive substrate-bound states available on the PDB, 12 exhibit a shared substrate conformation. This conformation is characterized by the nucleoside adopting an anti-configuration, with the phosphates and the nitrogenous base forming a linear-like structure, as illustrated by Figure 1. Figures 1A, B, and C visually illustrate this substrate conformation, utilizing ATP analogs co-crystalized with *R. novergicus* NTPDase2 (A), *L. pneumophila* NTPDase1 (B), and *T. gondii* NTPDase 3 (C).

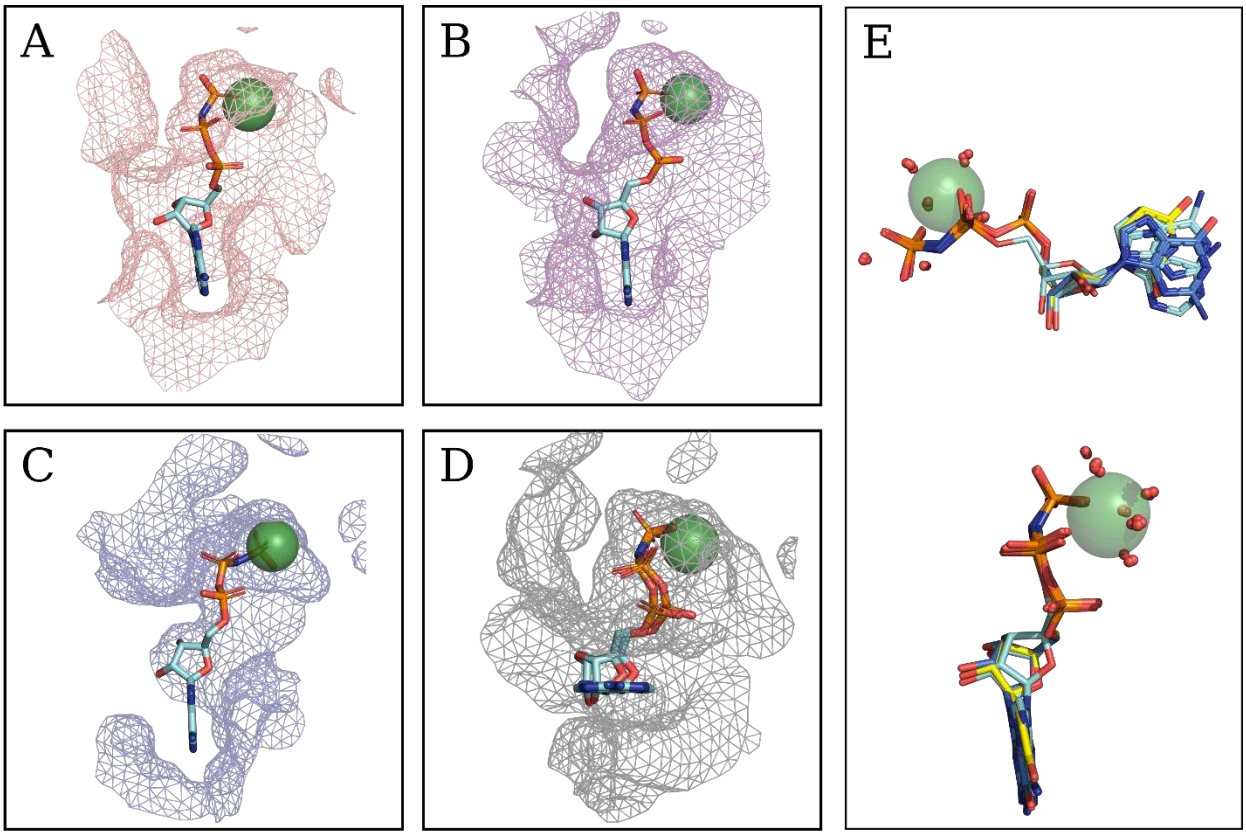

Figure 1 – Binding mode of ATP analogs on multiple NTPDases. The substrate accommodation within the NTPDases active site illustrated targeting resolved structures with ATP analog. The active site is portrayed as a mesh surface, the cofactor as a green sphere, and the substrates as sticks, with colors corresponding to the nitrogenous base. The figure illustrates the productive binding mode of ATP analogs on (A) RnNTPDase2 (PDB code: 3CJA), (B) LpNTPDase1 (PDB code: 4BRA), (C) TgNTPDase3 (PDB code: 4A5A), and (D) TgNTPDase1 (PDB code: 4KH4). Additionally, (E) presents a structural alignment of all substrates experimentally resolved in complex with the RnNTPDase2. The small red spheres represent water molecules of the cofactor's first hydration shell and the catalytic waters.

Utilizing ATP and ADP analogs co-crystallized with RnNTPDase2 (PDB code: 3CJA and 4BR0) as benchmark structures for triphosphate and diphosphate nucleotides, respectively. To evaluate the similarity among co-crystalized substrates, we calculate the RMSD of NTPs and NDPs compared to the reference structures using the PyMOL software. The resultant RMSD values are summarized in Table 1. Among the triphosphate substrates adopting the linear-like conformation, the highest RMSD value is observed when comparing the reference to the TgNTPDase3 ATP analog (PDB code: 4A5A), with a deviation of merely 1.22 Å (Figure 2A). In contrast, an RMSD of 0.32 Å is attained when aligning the diphosphate reference structure with the ADP analog bound to LpNTPDase1 (PDB code: 4BRC), representing the highest value among diphosphate nucleotides (Figure 2B). Figure 1E depicts the alignment of ATP, ADP, GTP, and UTP analogs co-crystallized with RnNTPDase2, showcasing a nearly identical conformation across different substrate types. Due to its frequent occurrence and similarity in multiple crystal structures, we designate this linear-like substrate conformation of NTPs and NDPs as the canonical substrate structure upon binding to an NTPDase.

Table 1 - Substrate similarity

Enzyme / PDB code	Corresponding Nucleotide	RMSD (Å)
Triphosphates		
RnNTPDase2 / 4BQZ	GTP	0.28
RnNTPDase2 / 4BR2	UTP	0.1
LpNTPDase1 / 4BRA	ATP	1.13
LpNTPDase1 / 4BRD	ATP	1.07
LpNTPDase1 / 4BRG	GTP	0.6
LpNTPDase1 / 4BRK	UTP	1.04
TgNTPDase3 / 4A5A	ATP	1.22
Diphosphate		
LpNTPDase1 / 4BRC	ADP	0.32
LpNTPDase1 / 4BRL	GDP	0.25
LpNTPDase1 / 4BRE	ADP	0.14

While most NTPDases accommodate the canonical substrate structure, TgNTPDase1 presents an alternative binding mode attributed to a distinct arrangement of residues within its binding site [10]. Notably, TgNTPDase1 is the sole resolved NTPDase structure that physically obstructs the space necessary to establish the canonical substrate structure, as evidenced by the enzyme surface in Figure 1D. The nitrogenous base of the ATP analog observed upon binding to TgNTPDase1 (PDB code: 4A5A) forms an approximate 90° angle with the nitrogenous base of the reference NTP structure (Figure 2C). A similar pattern, albeit with a less pronounced angle, is noted when

comparing the ADP analog co-crystalized with the same enzyme and its reference structure (Figure 2D). Nevertheless, despite this divergent orientation of the nitrogenous base observed in the TgNTPDase1 complexes, the phosphate and ribose moieties of ATP and ADP analogs still can be aligned with RMSD values of 1.64 and 0.64 Å, respectively, when compared to their respective reference structures.

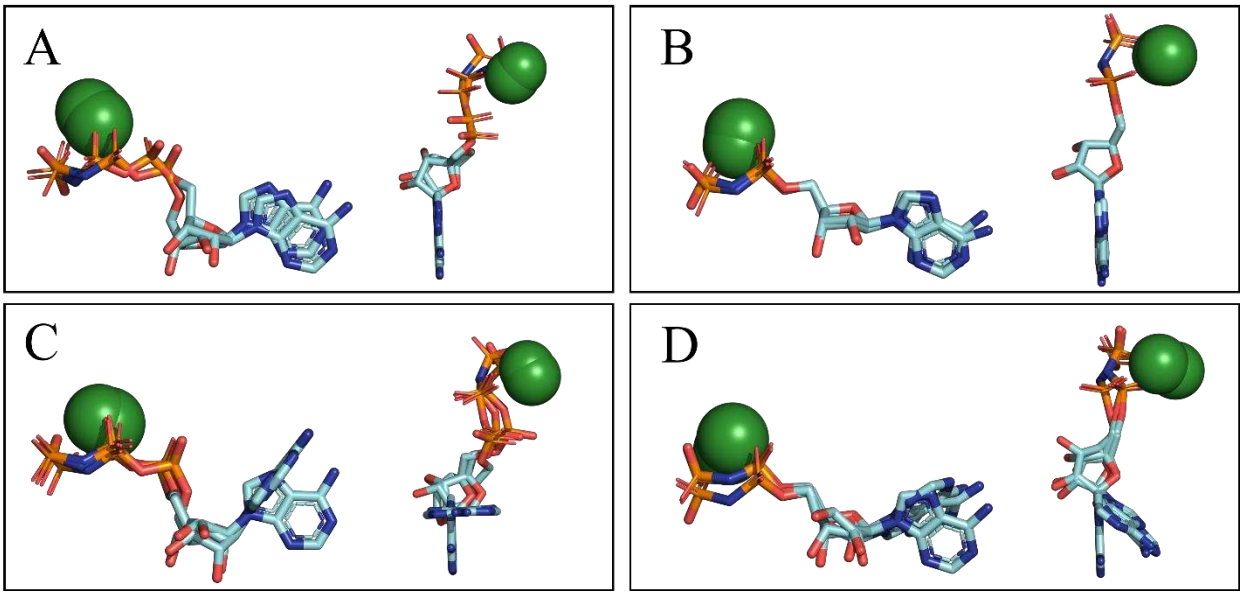

Figure 2 - Structural alignment of substrate analogs. The figures display the structural alignment of the triphosphate (A) and diphosphate (B) nucleotides in the canonical conformation with the highest RMSD values compared to their respective references. Figures C and D show the structural alignment of the alternative binding mode of ATP and ADP analogs, respectively, observed in TgNTPDase1. To enhance visualization of the similarities and differences, Figure 2A-D presents two perspectives of the same alignment per section: one lateral (on the left) and one from above (on the right).

In the case of the LpNTPDase1-UTP complex, two available PDB structures report different binding modes. One presents the canonical substrate structure (PDB code: 4BRK), and the other

in an alternative-like binding mode (PDB code: 4BRI) (Figure S3). As presented in Figure 1B, the binding site of LpNTPDase1 does not physically obstruct the substrate from adopting a linear-like conformation. Additionally, the positioning of the ribose in the alternative-like structure differs from the alternative binding mode characterized in TgNTPDase1. Consequently, the significance of this ambiguous UTP binding conformation in LpNTPDase1 remains to be determined.

Despite variations in substrate conformation or binding mode, certain general features persist in all crystal structures of productive NTPDase-substrate complexes. These include the bidentate interaction of the cofactor with the nucleotides and a symmetric interaction between the aspartic acid of the DXG motifs on ACRs 1 and ACR4 with the cofactor. Excluding the abnormal TgNTPDase1 measurements, when calculated from the DXG aspartic acid's Cγ, the average distances are 4.80 Å and 4.84 Å, respectively. Individual measurements are presented in Supplementary Table 2. The presence of four water molecules in the cofactor's first hydration shell, a catalytic water, and an additional water molecule close to the last phosphate group of NTPs and NDPs, is also conserved in all crystal structures where the water molecules' positions were determined (which excludes the TgNTPDase1 and 3 complexes). Regarding the ribose configuration, substrates co-crystallized with RnNTPDase2 consistently exhibit a C2' endo configuration, while the C3' endo conformation predominates in other structures.

The notion of what is a productive NTPDase-nucleotide complex is based on the set of conserved hydrogen bonds formed between the enzyme and the phosphate moiety of nucleotides, the position of the cofactor, and the presence of six conserved water molecules in the active site for the experimental structures with resolution for the water molecules [12]. Given the activity of the RnNTPDase2 in crystal form [9], this productive binding mode is expected to be a good

representation of the pre-catalytic structure of these complexes. Thus, the aforementioned features can be used as quantitative and qualitative metrics for characterizing an unknown productive NTPDase-substrate complex.

The structural analysis presented here underscores the high degree of conservation in substrate structure upon binding across various NTPDases. Based on these observations, we can imply that unknown NTPDase-substrate complexes probably exhibit these common characteristics. The experimental structures currently available also indicate that barring any physical obstruction similar to what is seen in TgNTPDase1, different nucleotides are likely to adopt the canonical conformation when productively bound to an NTPDase. In light of this context, computational studies intending to characterize or investigate unknown NTPDase-substrate complexes should base their investigations on initial structures that emulate the previously described characteristics.

The exploration of computational methods for studying the structure and functions of NTPDases remains relatively unexplored. To date, only two papers have ventured into applying Molecular Docking and Molecular Dynamics Simulations to investigate human NTPDases in complexes with substrates [35, 36]. In both of these works, Molecular Docking was the method of choice for constructing the NTPDase-substrate complex. Although a valid method for this purpose, even when attempting to reproduce an experimentally known complex through a Re-Docking process, successful predicted poses often fall within a < 2 Å RMSD range of structural difference [17, 18]. However, as illustrated in Figure S1, Docking poses within this deviation range can fall short in capturing the previously discussed conserved features in the experimental structures. Consequently, the Docking inaccuracies can constrain the biological relevance of an NTPDase-substrate complex described using such methodology and bias more complex computational

studies like MD simulation, which is highly dependent on the initial coordinates state of the simulated system [37].

In this context, the necessity for an alternative method to construct reliable NTPDase-substrate complexes that accurately incorporate the conserved structure and features previously described is apparent. To address this need, we propose transferring a crystallized substrate-cofactor complex and relevant water molecules to a well-modeled NTPDase structure. By maintaining the conserved characteristics of substrate structure and interaction, this approach offers an alternative option for characterizing and studying unresolved NTPDase-substrate complexes.

Co-crystalized substrate analogs can be readily converted into the desired nucleotide, and adjustments can be made to the cofactor type, such as switching from Ca^{2+} to Mg^{2+}, if necessary. Once assembled, the system can undergo optimization through molecular dynamics (MD) energy minimization and equilibration, ensuring proper alignment between the protein and transferred molecules. The resulting complex is anticipated to encompass all conserved features observed in crystal structures, serving as a robust starting point for more sophisticated computational methods or a reliable model for describing the NTPDase-substrate complex.

3.2. Application of the proposed substrate transferring protocol

To exemplify the efficacy of the proposed transferring method, we utilized PyMOL and selected substrate structures experimentally resolved to construct complexes of HsNTPDase1-8 with ATP, ADP, GTP, GDP, UTP, and UDP. The enzymes are discussed in pairs following their phylogenetic relationship previously characterized [33].

The HsNTPDase1-8 were modeled using the ColabFold engine [22], with RnNTPDase2 (PDB code: 3CJA) serving as a custom reference structure [9]. The rationale behind utilizing a reference structure was to bias the modeling process toward a specific protein state, particularly the closed enzyme state associated with a productive binding mode [12]. To ensure that a closed state was achieved for each of the HsNTPDase1-8, we structurally aligned the enzymes with the template structure, evaluating the RMSD and the position of ACR1 and 4. As depicted in Figure S4, the modeled enzyme structures nicely match the template structure, confirming a closed state for all of them. These models were then used to build each complex described next. Further characterization will be discussed elsewhere regarding the conformational changes associated with each HsNTPDase (manuscript in preparation).

Table 2 – Measurements and interaction of the transferred substrate to HsNTPDases.

Enzyme	Substrate	RMSD (Å)	Asp DXG ACR1-Cofactor Distance	Asp DXG ACR4 - Cofactor Distance	ACR residues forming hydrogen bonds with the substrate phosphate moiety		
					ACR1	ACR2	ACR4
NTPDase1	ATP	0.32	5.1	5.1	HIS59, SER58, SER57	GLY133, ALA132, THR131	GLY215, GLY216, ALA217,

							SER218
NTPDase1	ADP	0.26	4.8	6.0	HIS59, SER58, SER57	GLY133, ALA132, THR131	GLY215, GLY216, ALA217, SER218
NTPDase1	GTP	0.35	5.1	5.4	HIS59, SER58, SER57	GLY133, ALA132, THR131	GLY215, GLY216, ALA217, SER218
NTPDase1	GDP	0.32	4.9	5.2	HIS59, SER58, SER57	GLY133, ALA132, THR131	GLY215, GLY216, ALA217, SER218
NTPDase1	UTP	0.34	5.3	5.0	HIS59, SER58, SER57	GLY133, ALA132, THR131	GLY216, ALA217, SER218
NTPDase1	UDP	0.94	4.9	5.1	HIS59, SER58, SER57	GLY133, THR131	GLY216, ALA217, SER218
NTPDase2	ATP	0.26	5.2	5.1	HIS50, SER49, SER48	ALA123, THR122	GLY204, ALA205, SER206
NTPDase2	ADP	0.25	5.7	4.9	HIS50, SER49, SER48	GLY124, ALA123, THR122	GLY203, GLY204, ALA205, SER206
NTPDase2	GTP	0.30	5.0	5.4	HIS50, SER49, SER48	GLY124, ALA123, THR122	GLY204, ALA205, SER206
NTPDase2	GDP	0.13	5.8	5.0	HIS50, SER49, SER48	GLY124, ALA123, THR122	GLY204, ALA205, SER206
NTPDase2	UTP	0.28	5.3	5.0	HIS50, SER49, SER48	GLY124, ALA123, THR122	GLY204, ALA205, SER206
NTPDase2	UDP	0.87	5.6	4.9	HIS50, SER49,	GLY124, THR122	GLY203, GLY204,

					SER48		ALA205, SER206
NTPDase3	ATP	0.32	4.9	5.0	ARG67, SER66, SER65	GLY141, ALA140, THR139	GLY222, ALA223, SER224
NTPDase3	ADP	0.38	5.1	5.1	ARG67, SER66, SER65	ALA140, THR139	GLY221, GLY222, ALA223, SER224
NTPDase3	GTP	0.27	5.1	5.1	ARG67, SER66, SER65	GLY141, ALA140, THR139	GLY221, GLY222, ALA223, SER224
NTPDase3	GDP	0.38	5.5	4.6	ARG67, SER66, SER65	GLY141, ALA140, THR139	GLY221, GLY222, ALA223, SER224
NTPDase3	UTP	0.50	5.1	4.8	ARG67, SER66, SER65	GLY141, ALA140, THR139	GLY221, GLY222, ALA223, SER224
NTPDase3	UDP	0.92	5.7	4.5	ARG67, SER66, SER65	GLY141, ALA140, THR139	GLY221, GLY222, ALA223, SER224
NTPDase4	ATP	0.39	5.1	5.1	SER98, SER97	ALA180, THR179	GLY273, GLY274, ALA275, SER276
NTPDase4	ADP	0.63	5.4	4.9	GLY99, SER98, SER97	GLY181, ALA180, THR179	GLY273, GLY274, ALA275, SER276
NTPDase4	GTP	0.31	5.0	5.3	GLY99, SER98, SER97	GLY181, ALA180, THR179	GLY274, ALA275, SER276
NTPDase4	GDP	071	5.0	5.4	SER98,	ALA180,	GLY274,

					SER97	THR179	ALA275, SER276
NTPDase4	UTP	0.33	5.1	5.4	GLY99, SER98, SER97	GLY181, ALA180, THR179	GLY273, GLY274, ALA275, SER276
NTPDase4	UDP	0.95	5.5	5.0	GLY99, SER98, SER97	GLY181, ALA180, THR179	GLY273, GLY274, ALA275, SER276
NTPDase5	ATP	-	4.6	5.1	-	-	-
NTPDase5	ADP	0.35	4.7	4.8	GLY39, THR38, SER37	GLY112, ALA111, THR110	GLY182, ALA183, SER184
NTPDase5	GTP	0.54	4.7	4.9	GLY39, THR38, SER37, GLY36	GLY112, ALA111, THR110	GLY182, ALA183, SER184
NTPDase5	GDP	0.33	4.9	4.9	GLY39, THR38, SER37	GLY112, ALA111, THR110	GLY181, GLY182, ALA183, SER184
NTPDase5	UTP	0.32	4.9	5.1	GLY39, THR38, SER37	GLY112, ALA111, THR110	GLY182, ALA183, SER184
NTPDase5	UDP	0.81	5.0	5.0	GLY39, THR38, SER37	GLY112, ALA111, THR110	GLY182, ALA183, SER184
NTPDase6	ATP	0.41	5.1	4.9	GLY52, THR51, SER50	GLY124, ALA123, THR122	GLY194, GLY195, SER196
NTPDase6	ADP	0.30	5.1	5.3	GLY52, THR51, SER50	GLY124, ALA123, THR122	GLY193, GLY194, GLY195, SER196
NTPDase6	GTP	0.36	4.9	4.9	GLY52, THR51,	GLY124, ALA123,	GLY194, GLY195,

					SER50	THR122	SER196
NTPDase6	GDP	0.37	4.9	4.7	GLY52, THR51, SER50	GLY124, ALA123, THR122	GLY193, GLY194, GLY195, SER196
NTPDase6	UTP	0.44	4.9	4.8	GLY52, THR51, SER50	GLY124, ALA123, THR122	GLY193, GLY194, GLY195, SER196
NTPDase6	UDP	0.91	4.7	5.6	GLY52, THR51, SER50	GLY124, ALA123, THR122	GLY193, GLY194, GLY195, SER196
NTPDase7	ATP	0.41	5.1	5.4	GLY45, SER44, SER43	ALA126, THR127	GLY211, GLY112, ALA213, SER214
NTPDase7	ADP	0.64	5.1	5.0	GLY45, SER44, SER43	ALA126, THR127	GLY112, ALA213, SER214
NTPDase7	GTP	0.27	4.8	5.7	GLY45, SER44, SER43	GLY127, ALA126, THR127	GLY112, ALA213, SER214
NTPDase7	GDP	0.62	5.0	5.4	GLY45, SER44, SER43	ALA126, THR127	GLY112, ALA213, SER214
NTPDase7	UTP	0.32	4.6	5.9	GLY45, SER44, SER43	GLY127, ALA126, THR127	GLY112, ALA213, SER214
NTPDase7	UDP	0.89	4.9	5.7	GLY45, SER44, SER43	ALA126, THR127	GLY211, GLY112, ALA213, SER214
NTPDase8	ATP	0.44	5.0	5.6	HIS53, SER52, SER51	GLY217, ALA126, THR125	GLY207, GLY208, ALA209, SER210

NTPDase8	ADP	0.38	5.4	4.8	HIS53, SER52, SER51	GLY217, ALA126, THR125	GLY208, ALA209, SER210
NTPDase8	GTP	0.46	5.6	4.7	HIS53, SER52, SER51	GLY217, THR125	GLY208, ALA209, SER210
NTPDase8	GDP	0.19	4.6	5.1	HIS53, SER52, SER51	GLY217, ALA126, THR125	GLY207, GLY208, ALA209, SER210
NTPDase8	UTP	0.77	5.0	5.3	HIS53, SER52, SER51	GLY217, ALA126, THR125	GLY208, ALA209, SER210
NTPDase8	UDP	0.79	5.4	5.0	HIS53, SER52, SER51	GLY217, ALA126, THR125	GLY207, GLY208, ALA209, SER210

The following figures, presented in a comparable perspective, highlight the interactions of each enzyme with the ribose and nitrogenous base of the studied substrates. Notably, as previously characterized, the phosphate moiety of all nucleotides engages in highly conserved hydrogen bonds with the catalytic site formed by the ACR1-5 [12]. To enhance clarity and focus on more distinguishing features of each complex, these conserved hydrogen bonds are not explicitly depicted in the figures. Nevertheless, the residues engaging in these interactions are summarized in Table 2 alongside with the RMSD values of each substrate compared to the NTP and NDP complex reference structure and the distance values between the aspartic on each DXG motif and the cofactor.

3.3. HsNTPDase1 and HsNTPDase8

The HsNTPDase1 and 8 exhibit a general preference for NTPs over NDPs, a feature shared with the other cell surface-located NTPDases in different organisms. Nevertheless, each isoform displays its unique characteristics. The HsNTPDase1 has a higher activity towards adenine nucleotides and a similar hydrolytic capability when processing NTPs and NDPs, while the HsNTPDase8 displays the characteristic preference for NTPs over NDPs regardless of the nucleobase type [33, 38–41].

In HsNTPDase1, the stabilization of purinergic nucleotides in the bound state is mediated by a π-π stacking interaction between Phe365, Tyr408, and the substrate's nitrogenous base (Figure 3A, B, D, and E). Conversely, pyrimidinergic nucleotides such as UTP and UDP can only engage in one π interaction with Phe365 (Figure 3C and F). In comparison, HsNTPDase8 interacts with all substrate bases through a single π interaction via the Tyr357 residue. Notably, Trp398 replaces Tyr408 in HsNTPDase1 and does not interact with any of the substrates (Figure 3G-L). However, in solution, conformational changes may place Trp398 in an optimal position to also form π interactions with substrates.

Multiple residues in both enzymes form hydrogen bonds with ribose substrates. Asp259 and Lys258 residues interact with both ATP and ADP in HsNTPDase1. Additionally, ATP forms a hydrogen bond with Tyr408, and ADP forms one with Tyr412 through the N1 atom of the adenine base (Figure 3A and D). Both guanine nucleotides form a similar hydrogen bond with Tyr412. Upon binding to an NTPDase, diphosphate nucleotides such as ADP adopt an elongated structure, and the hydrogen bond with the nitrogenous base and others shared with ATP may help stabilize this conformation. GTP and UTP form a hydrogen bond with Ser58 of the ACR1

residue, which is predicted to interact with the phosphate moiety (Figure 3 C and D). The competition between the ribose and the phosphates for the same interaction partner could cause instability in these complexes. In contrast, in HsNTPDase8, the enzyme-ribose interactions remain highly conserved, with Arg249, Asp250, and Asn354 forming similar hydrogen bonds with all nucleotides (Figure 3 G-L).

The Arg85 residue in HsNTPDase1, although not predicted to interact with any tested nucleotide, is consistently located close to the substrate in all complexes, a unique feature of this enzyme (Figure 3AF). Arg85 is a positively charged amino acid that can interact with NTPs and NDPs' phosphate moiety. This residue may be essential for HsNTPDase1's distinctive properties, such as directly converting ATP to AMP [42].

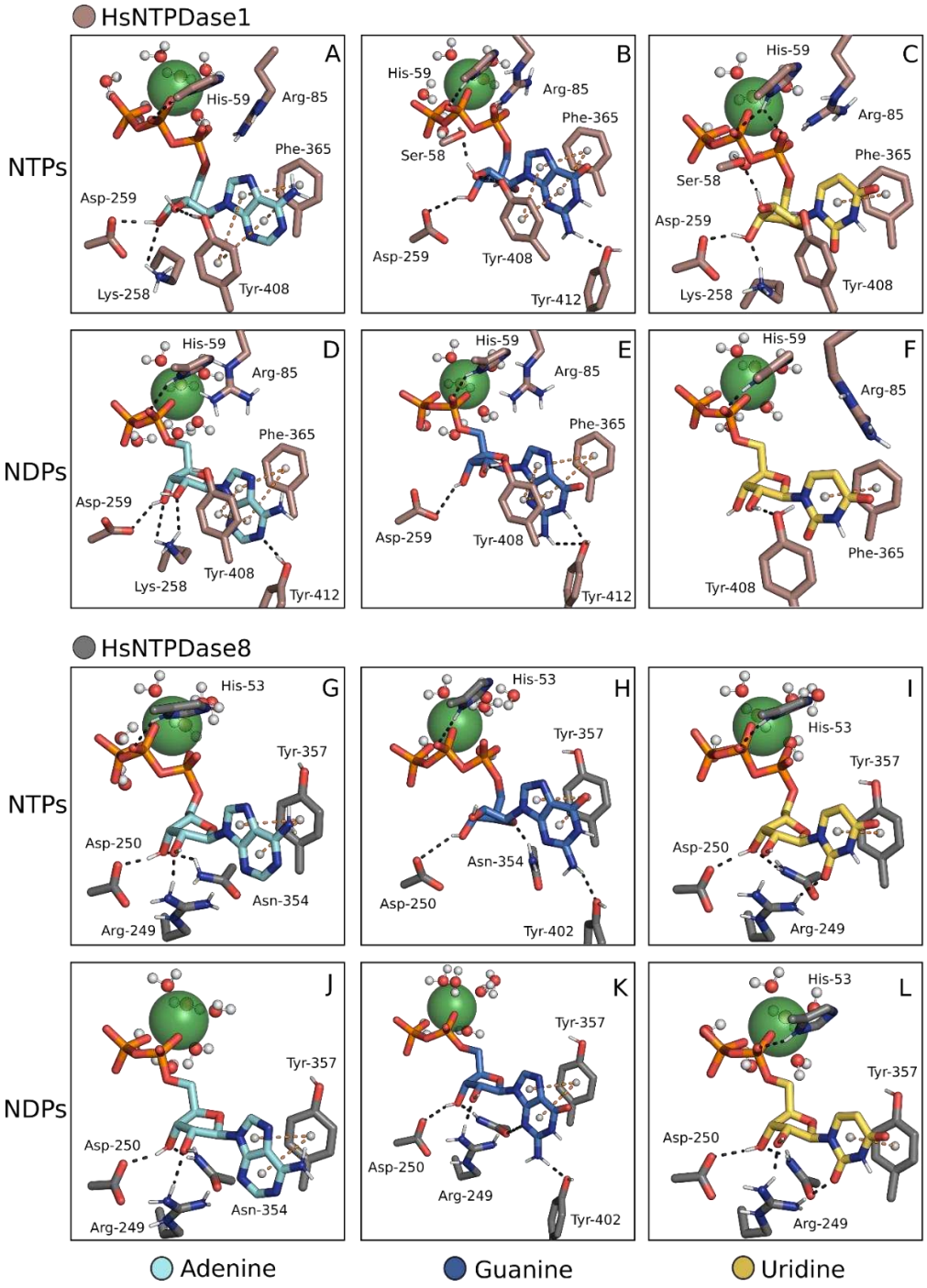

Figure 3 - HsNTPDase1 and 8 in complex with multiple substrates. NTPDase-substrate complexes assembled by the proposed method. HsNTPDase1 is in complex with ATP, GTP, UTP, ADP, GDP, UDP (A-F), and HsNTPDase8 with the same substrates (G-L). Ca^{2+} represented as a green sphere, was the cofactor of choice.

3.4. HsNTPDase2 and HsNTPDase3

The other two cell surface-located NTPDases found in humans are the HsNTPDase2 and the HsNTPdase3. While closely related to the other members of this group, isoforms 2 and 3 exhibit unique characteristics and are identified in their respective branches on the phylogenetic tree [33]. Regarding their activity, the HsNTPDase3 has a standard preference for NTPs over NDPs. Compared to other cell surface NTPDases, HsNTPDase2 stands out as the isoform with the highest preference for triphosphate nucleotides [33, 38].

In HsNTPDase2, Arg394 is functionally similar to HsNTPDase1 Tyr408 in stabilizing substrate bases. This is accomplished through a cation-π and a π-π interaction mediated by the Tyr350 residue (Figure 4A-F). Conversely, HsNTPDase3 employs Tyr417 and Tyr375 residues to establish the standard π-π stacking interaction with the substrate bases (Figure 4G-L). In the context of hydrogen bonds involving the ribose of the nucleotides, HsNTPDase2 forms four hydrogen bonds with ATP via residues Arg245 and Asp246 (Figure 4A). These identical residues form fewer hydrogen bonds with other substrates (Figure 4B-F). In contrast, HsNTPDase3 engages in fewer interactions with the ribose moiety, with each adenine nucleotide forming two bonds (Figure 4G and J).

Notably, an arginine on ACR1 distinguishes HsNTPDase3 from other cell surface HsNTPDases that possess a histidine residue in the sixth position of ACR1. The histidine residue in other enzymes forms a hydrogen bond between its Nδ and an oxygen atom of the Pβ or Pα in NTPs or NDPs complexes, respectively (Figure 3A-L). In HsNTPDase3, the Arg67 backbone nitrogen forms a hydrogen bond with the substrates, similar to the interaction mediated by the histidine residue in other HsNTPDases (Figure 4G-L).

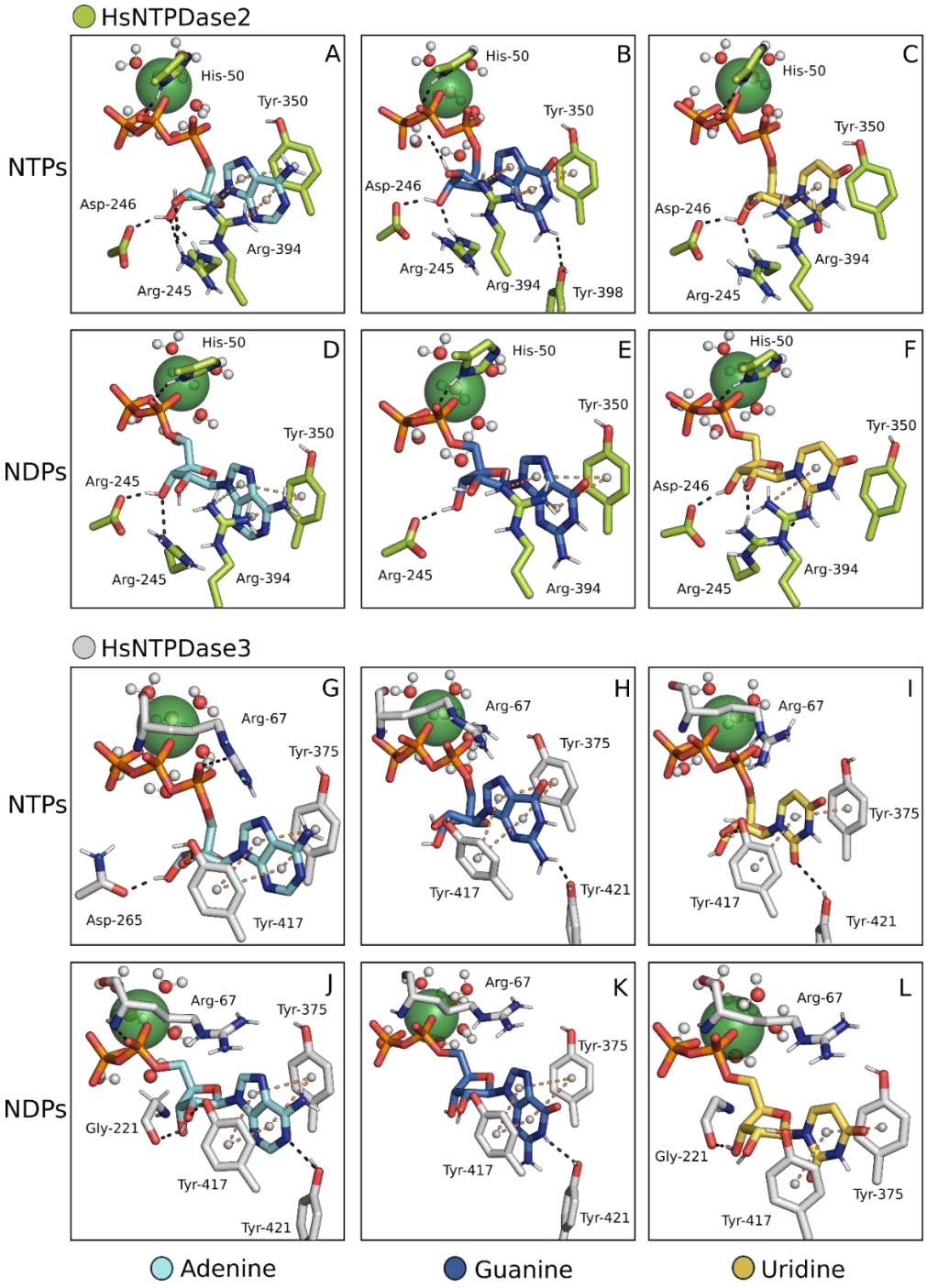

Figure 4 - HsNTPDase2 and 3 in complex with multiple substrates. NTPDase-substrate complexes assembled by the proposed method. HsNTPDase2 is in complex with ATP, GTP, UTP, ADP, GDP, UDP (A-F), and HsNTPDase3 with the same substrates (G-L). Ca^{2+}, represented as a green sphere, was the cofactor of choice.

3.5. HsNTPDase4 and HsNTPDase7

Notably, the NTPDase family exhibits a highly conserved structure, but sequence conservation is typically low, even among isoforms of the same species. For instance, when comparing human NTPDase isoforms, HsNTPDase4 and HsNTPDase7 demonstrate the highest sequence similarity, sharing up to 67% of their ectodomain sequence. Notably, both enzymes favor uridine substrates, but HsNTPDase4 is more active towards UDP [43], whereas HsNTPDase7 is more active towards UTP [44]. Furthermore, both HsNTPDase4 and 7 stabilize the nitrogenous base of their nucleotides by employing a single π interaction with the nitrogenous base, which is mediated by Tyr436 (Figure 5A-F) and Tyr374 Figure (5G-L) residues, respectively.

Compared to other HsNTPDase residues, these tyrosine residues move closer to the catalytic site, facilitating the formation of π-π interaction with pyrimidinergic substrates. Unlike other isoforms, HsNTPDase4 and 7 promote multiple hydrogen bonds with all substrates' nitrogenous bases (Figure 5A-L). The interactions between these residues may play a crucial role in their specific characteristics.

The nucleotide-binding sites of these two enzymes contain conserved residues that interact with the ribose part of the nucleotides. In HsNTPDase4, the residues Asn331, Glu433, and Lys492, mediate this interaction (Figure 5A-F). On HsNTPDase7, the corresponding residues Asn269, Glu370, and Lys430 are responsible for forming the mentioned interaction (Figure 5G-L). His484 also plays a role in substrate interaction in the HsNTPDase4-ATP complex.

All intracellular HsNTPDases share a common characteristic: the presence of an arginine residue immediately after ACR1. This residue interacts with the phosphate moiety of all nucleotides (Figures 5 and 6 A-L). Based on the information gathered from enzymes related to the NTPDase

family, an arginine residue in this position might be a crucial aspect of NDPase activity displayed by these enzymes. Supporting this idea, it has been found that adding an arginine in the corresponding spot of HsNTPDase3 enhances its diphosphatase activity [45].

The reduced activity of hsNTPDase4 towards adenine nucleotides is another notable feature of these enzymes. It has been suggested that this characteristic may be linked to steric impediments imposed by His480 [13]. The same can be implied for HsNTPDase7 via the corresponding reside His418. In our models, these residues' positions do not obstruct any substrate positioning. The His480 and His418 form hydrogen bonds with multiple substrates, including ATP and ADP (Figure 5A, B, D, E). However, the histidine side chain can be highly flexible. Moving toward an adenine substrate might elicit a repulsive response between the imidazolium ring of both residues and the adenine amino group. Conversely, guanine and uridine nucleotides might have an opposite response due to the carboxyl groups on both bases. Figure 5H shows a hydrogen bond between the His418 side chain and the ATP carboxyl group.

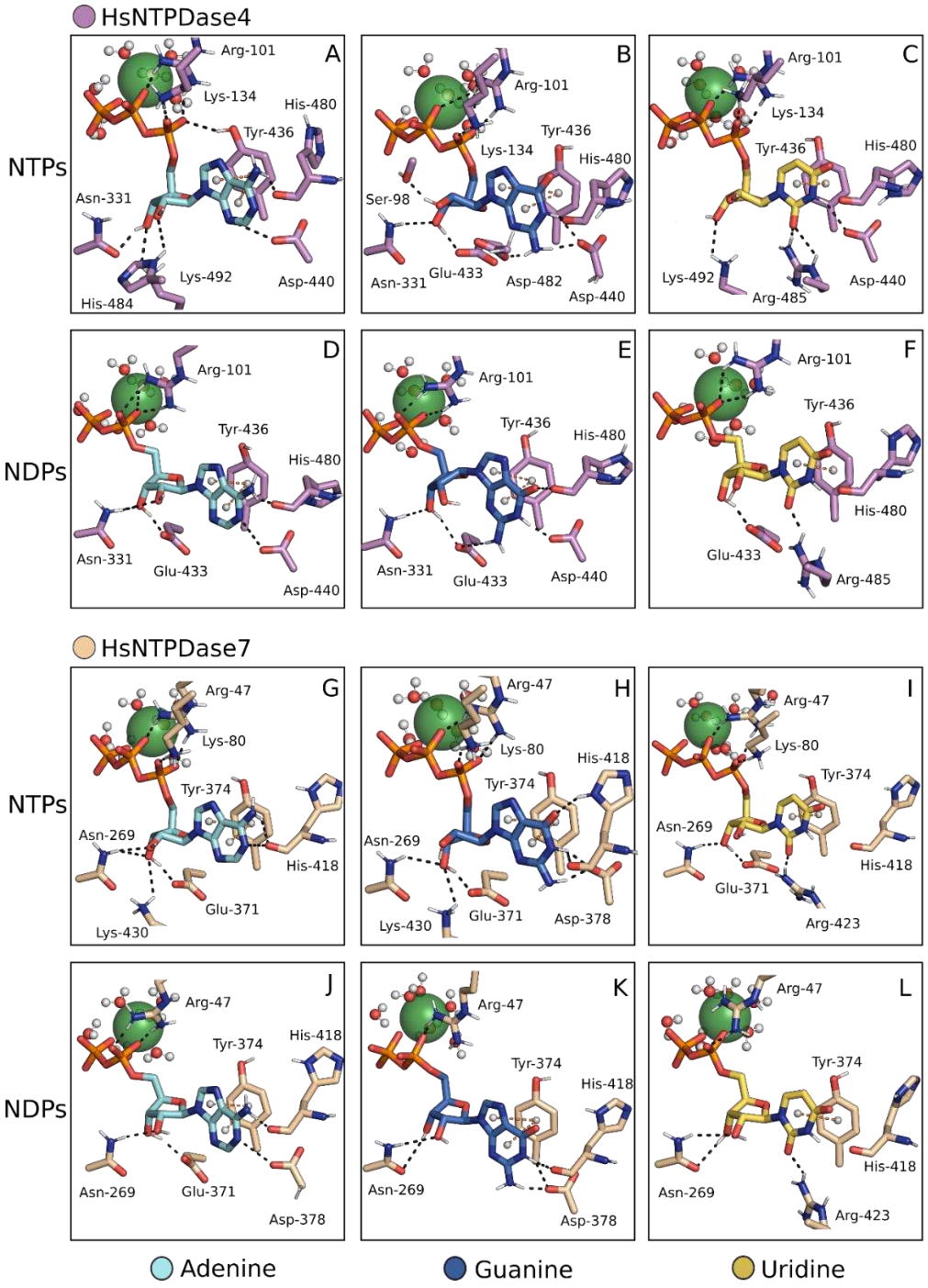

Figure 5 - HsNTPDase4 and 7 in complex with multiple substrates. NTPDase-substrate complexes assembled by the proposed method. HsNTPDase4 is in complex with ATP, GTP, UTP, ADP, GDP, UDP (A-F), and HsNTPDase7 with the same substrates (G-L). Ca^{2+}, represented as a green sphere, was the cofactor of choice.

3.6. HsNTPDase5 and HsNTPDase6

The energy minimization and equilibration processes were applied to each complex to achieve the most energetically favorable complex. However, the standard energy minimization protocol used for all other complexes was ineffective in accommodating the ATP molecule in the HsNTpdase5 active site. Interestingly, previous studies on HsNTPDase5 have indicated little to no activity toward ATP, suggesting that forming a productive complex of ATP with HsNTPDase5 is not a favorable process [46, 47]. To overcome this limitation, the enzyme-substrate complex was first minimized in a vacuum and then in solution.

The complexes with HsNTPDase 5 and 6, despite having a substrate structure resembling the crystallographic information, deviate from those formed with other HsNTPDases. Isoform 5 forms a cation -π interaction mediated through Asp134 that stabilizes the nucleotide base (Figure 6A-F). Furthermore, Tyr311 engages in a T-shaped π-π interaction with the nucleotide bases in ADP and UDP complexes (Figure 6D and F). Asp327 forms a cation-π interaction with purinergic bases and a hydrogen bond with uridine nucleotides (Figure 6G-L), while Tyr324 forms T-shaped π-π interactions with ADP and GDP (Figure 6G and K). Unlike other isoforms, no hydrogen bonds are formed between the ribose of substrates and the enzymes.

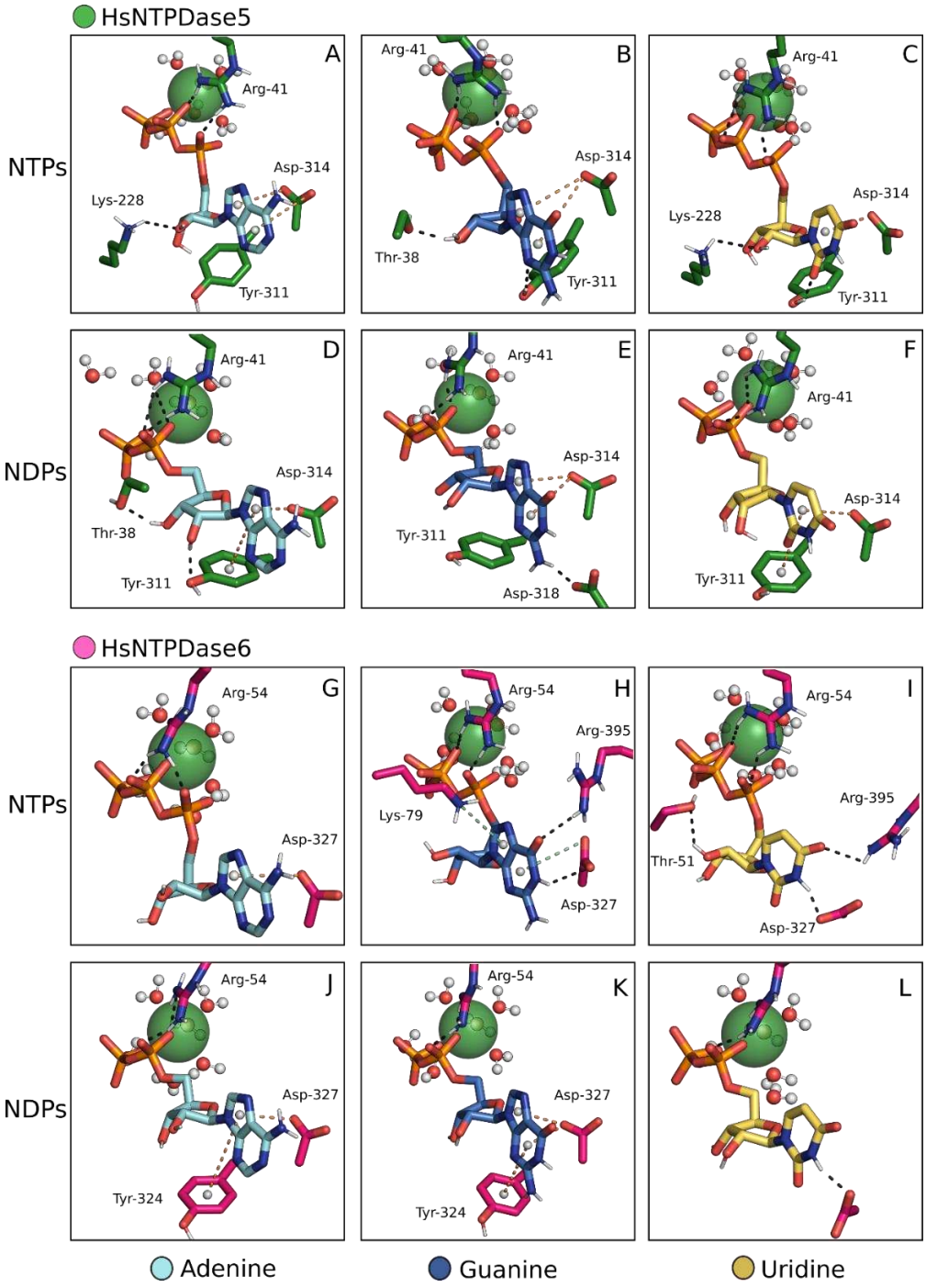

Figure 6 - HsNTPDase5 and 6 in complex with multiple substrates. NTPDase-substrate complexes assembled by the proposed method. HsNTPDase5 is in complex with ATP, GTP, UTP, ADP, GDP, UDP (A-F), and HsNTPDase6 with the same substrates (G-L). Ca^{2+}, represented as a green sphere, was the cofactor of choice.

4. Conclusion

The analysis of NTPDase-substrate complexes across various isoforms reveals a highly conserved substrate conformation upon binding to several NTPDases, with most structures exhibiting a canonical anti-configuration of the nucleoside in a linear-like arrangement of phosphates and nitrogenous base. Variations in substrate structure while maintaining fundamental enzyme-substrate interactions, as the alternative binding mode seen in TgNTPDase1, underscores that the adaptability of the NTPDases to process multiple substrates is likely linked to the stabilization of a conserved nucleotide structure upon binding. This structural conservation suggests that unknown NTPDase-substrate complexes probably share these features, providing a foundation for computational modeling efforts.

In light of these observations, we propose that transferring co-crystallized substrate-cofactor complexes and important water molecules to well-modeled NTPDase structures offers a reliable method for assembling unknown NTPDase-substrate complexes. Refining these complexes through molecular dynamics energy minimization and equilibration allows for the adaptation of the enzyme and transferred molecules to one another. This method ensures that the conserved features observed in experimental structures are accurately reproduced, enhancing the reliability of computational predictions regarding NTPDase-nucleotide complexes.

We have successfully modeled and described the HsNTPDase1-8 with multiple substrates, demonstrating the utility of our approach. Although a static model limits the depth of information accessible, these structures serve as literature-consistent starting points for more comprehensive characterization through MD simulation. Further investigation utilizing these models can lead to

a better understanding of the HsNTPDases specificities, potentially providing the necessary information for its optimization, unleashing the therapeutic potential of these enzymes.

Overall, in the absence of experimental structures, we encourage researchers to apply this method on their own structural investigation of the NTPDase family.

References

1. Knowles AF (2011) The GDA1_CD39 superfamily: NTPDases with diverse functions. Purinergic Signal 7:21–45. https://doi.org/10.1007/s11302-010-9214-7

2. Sévigny J (2018) E-NTPDase Family. In: Encyclopedia of Signaling Molecules. Springer, Cham, pp 1544–1553

3. Burnstock G, Satchell DG, Smythe A (1972) A comparison of the excitatory and inhibitory effects of non-adrenergic, non-cholinergic nerve stimulation and exogenously applied ATP on a variety of smooth muscle preparations from different vertebrate species. Br J Pharmac 46:234–242. https://doi.org/10.1111/j.1476-5381.1972.tb06868.x

4. Paes-Vieira L, Gomes-Vieira AL, Meyer-Fernandes JR (2021) E-NTPDases: Possible Roles on Host-Parasite Interactions and Therapeutic Opportunities. Front Cell Infect Microbiol 11

5. da Silva W, Ribeiro IC, Agripino J de M, et al (2023) Leishmania infantum NTPDase1 and NTPDase2 play an important role in infection and nitric oxide production in macrophages. Acta Trop 237:106732. https://doi.org/10.1016/J.ACTATROPICA.2022.106732

6. Dwyer KM, Deaglio S, Gao W, et al (2007) CD39 and control of cellular immune responses. Purinergic Signal 3:171–180. https://doi.org/10.1007/S11302-006-9050-Y/FIGURES/3

7. Liu Y, Li Z, Zhao X, et al (2023) Review immune response of targeting CD39 in cancer. Biomark Res 11:1–12. https://doi.org/10.1186/S40364-023-00500-W/TABLES/1

8. Baghbani E, Noorolyai S, Shanehbandi D, et al (2021) Regulation of immune responses through CD39 and CD73 in cancer: Novel checkpoints. https://doi.org/10.1016/j.lfs.2021.119826

9. Zebisch M, Strä N (2008) Structural insight into signal conversion and inactivation by NTPDase2 in purinergic signaling. PNAS 105:6882–6887

10. Krug U, Totzauer R, Zebisch M, Sträter N (2013) The ATP/ADP substrate specificity switch between Toxoplasma gondii NTPDase1 and NTPDase3 is caused by an altered mode of binding of the substrate base. ChemBioChem 14:2292–2300. https://doi.org/10.1002/cbic.201300441

11. Krug U, Zebisch M, Krauss M, Sträter N (2012) Structural insight into activation mechanism of Toxoplasma gondii nucleoside triphosphate diphosphohydrolases by

disulfide reduction. Journal of Biological Chemistry 287:3051–3066. https://doi.org/10.1074/JBC.M111.294348

12. Zebisch M, Krauss M, Schäfer P, et al (2013) Crystallographic Snapshots along the Reaction Pathway of Nucleoside Triphosphate Diphosphohydrolases. Structure 21:1460–1475. https://doi.org/10.1016/J.STR.2013.05.016

13. Gorelik A, Labriola JM, Illes K, Nagar B (2020) Crystal structure of the nucleotide-metabolizing enzyme NTPDase4. Protein Science 29:2054–2061. https://doi.org/10.1002/pro.3926

14. Wu L, Qin L, Nie Y, et al (2022) Computer-aided understanding and engineering of enzymatic selectivity. Biotechnol Adv 54:107793. https://doi.org/10.1016/J.BIOTECHADV.2021.107793

15. Jumper J, Evans R, Pritzel A, et al (2021) Highly accurate protein structure prediction with AlphaFold. Nature 596:583–589. https://doi.org/10.1038/s41586-021-03819-2

16. Terwilliger TC, Liebschner D, Croll TI, et al (2023) AlphaFold predictions are valuable hypotheses and accelerate but do not replace experimental structure determination. Nature Methods 2023 1–7. https://doi.org/10.1038/s41592-023-02087-4

17. Jakhar R, Dangi M, Khichi A, Chhillar AK (2019) Relevance of Molecular Docking Studies in Drug Designing. Curr Bioinform 15:270–278. https://doi.org/10.2174/1574893615666191219094216

18. Saikia S, Bordoloi M (2018) Molecular Docking: Challenges, Advances and its Use in Drug Discovery Perspective. Curr Drug Targets 20:501–521. https://doi.org/10.2174/1389450119666181022153016

19. Berman HM, Westbrook J, Feng Z, et al (2000) The Protein Data Bank. Nucleic Acids Res 28:235–242. https://doi.org/10.1093/NAR/28.1.235

20. Apweiler R, Bairoch A, Wu CH, et al (2004) UniProt: the Universal Protein knowledgebase. Nucleic Acids Res 32:D115. https://doi.org/10.1093/NAR/GKH131

21. Hallgren J, Tsirigos KD, Damgaard Pedersen M, et al (2022) DeepTMHMM predicts alpha and beta transmembrane proteins using deep neural networks. bioRxiv. https://doi.org/10.1101/2022.04.08.487609

22. Mirdita M, Schütze K, Moriwaki Y, et al (2022) ColabFold: making protein folding accessible to all. Nature Methods 2022 19:6 19:679–682. https://doi.org/10.1038/s41592-022-01488-1

23. Martínez-Rosell G, Giorgino T, De Fabritiis G (2017) PlayMolecule ProteinPrepare: A Web Application for Protein Preparation for Molecular Dynamics Simulations. J Chem Inf Model 57:1511–1516. https://doi.org/10.1021/ACS.JCIM.7B00190

24. Lee J, Cheng X, Swails JM, et al (2016) CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field. J Chem Theory Comput 12:405–413. https://doi.org/10.1021/ACS.JCTC.5B00935

25. Neria E, Fischer S, Karplus M (1996) Simulation of activation free energies in molecular systems. Journal of Chemical Physics 105:1902–1921. https://doi.org/10.1063/1.472061

26. Jorgensen WL, Chandrasekhar J, Madura JD, et al (1983) Comparison of simple potential functions for simulating liquid water. J Chem Phys 79:926–935. https://doi.org/10.1063/1.445869

27. Durell SR, Brooks BR, Ben-Naim A (1994) Solvent-Induced Forces between Two Hydrophilic Groups. J Phys Chem 98:2198–2202. https://doi.org/10.1021/j100059a038

28. Huang J, Rauscher S, Nawrocki G, et al (2016) CHARMM36m: an improved force field for folded and intrinsically disordered proteins. Nature Methods 2016 14:1 14:71–73. https://doi.org/10.1038/nmeth.4067

29. Khan HM, MacKerell AD, Reuter N (2019) Cation-π Interactions between Methylated Ammonium Groups and Tryptophan in the CHARMM36 Additive Force Field. J Chem Theory Comput 15:7–12. https://doi.org/10.1021/ACS.JCTC.8B00839

30. Brooks BR, Brooks CL, Mackerell AD, et al (2009) CHARMM: The Biomolecular Simulation Program. J Comput Chem 30:1545. https://doi.org/10.1002/JCC.21287

31. E. J. Haug, J. S. Arora, K. Matsui (1976) A Steepest-Descent Method for Optimization of Mechanical Systems. J Optim Theory Appl 19:401–424. https://doi.org/10.1007/BF00941484

32. Ypmat TJ (1995) Historical development of the newton-raphson method. Society for Industrial and Applied Mathematics 37:3. https://doi.org/10.1137/1037125

33. Robson SC, Sévigny J, Zimmermann H (2006) The E-NTPDase family of ectonucleotidases: Structure function relationships and pathophysiological significance. Purinergic Signal 2:409. https://doi.org/10.1007/S11302-006-9003-5

34. Laliberte JF, Beaudoin AR (1983) Sequential hydrolysis of the γ- and β-phosphate groups of ATP by the ATP diphosphohydrolase from pig pancreas. Biochim Biophys Acta 742:9–15. https://doi.org/10.1016/0167-4838(83)90352-7

35. Vadlamani VMK, Gunasinghe KKJ, Chee XW, et al (2023) Human soluble CD39 displays substrate inhibition in a substrate-specific manner. Scientific Reports 2023 13:1 13:1–12. https://doi.org/10.1038/s41598-023-36257-3

36. Iqbal J, Shah SJA (2018) Molecular dynamic simulations reveal structural insights into substrate and inhibitor binding modes and functionality of Ecto-Nucleoside Triphosphate Diphosphohydrolases. Scientific Reports 2018 8:1 8:1–11. https://doi.org/10.1038/s41598-018-20971-4

37. Hollingsworth SA, Dror RO (2018) Molecular Dynamics Simulation for All. Neuron 99:1129–1143. https://doi.org/10.1016/j.neuron.2018.08.011

38. Kukulski F, Lévesque SA, Lavoie ÉG, et al (2005) Comparative hydrolysis of P2 receptor agonists by NTPDases 1, 2, 3 and 8. Purinergic Signal 1:193–204. https://doi.org/10.1007/S11302-005-6217-X/TABLES/3

39. Christoforidis S, Papamarcaki T, Galaris D, et al (1995) Purification and Properties of Human Placental ATP Diphosphohydrolase. Eur J Biochem 234:66–74. https://doi.org/10.1111/j.1432-1033.1995.066_c.x

40. Fausther M, Lecka J, Kukulski F, et al (2007) Cloning, purification, and identification of the liver canalicular ecto-ATPase as NTPDase8. Am J Physiol Gastrointest Liver Physiol 292:785–795. https://doi.org/10.1152/ajpgi.00293.2006

41. Knowles AF, Li C (2006) Molecular cloning and characterization of expressed human ecto-nucleoside triphosphate diphosphohydrolase 8 (E-NTPDase 8) and its soluble extracellular domain. Biochemistry 45:7323–7333. https://doi.org/10.1021/bi052268e

42. Heine P, Braun N, Heilbronn A, Zimmermann H (1999) Functional characterization of rat ecto-ATPase and ecto-ATP diphosphohydrolase after heterologous expression in CHO cells. Eur J Biochem 262:102–107. https://doi.org/10.1046/J.1432-1327.1999.00347.X

43. Wang TF, Guidotti G (1998) Golgi localization and functional expression of human uridine diphosphatase. J Biol Chem 273:11392–11399. https://doi.org/10.1074/JBC.273.18.11392

44. Shi J Da, Kukar T, Wang CY, et al (2001) Molecular Cloning and Characterization of a Novel Mammalian Endo-apyrase (LALP1). Journal of Biological Chemistry 276:17474–17478. https://doi.org/10.1074/JBC.M011569200

45. Moeckel D, Jeong SS, Sun X, et al (2014) Optimizing human apyrase to treat arterial thrombosis and limit reperfusion injury without increasing bleeding risk. Sci Transl Med 6:. https://doi.org/10.1126/SCITRANSLMED.3009246

46. Mulero JJ, Yeung G, Nelken ST, Ford JE (1999) CD39-L4 Is a Secreted Human Apyrase, Specific for the Hydrolysis of Nucleoside Diphosphates. Journal of Biological Chemistry 274:20064–20067. https://doi.org/10.1074/JBC.274.29.20064

47. Murphy-Piedmonte DM, Crawford PA, Kirley TL (2005) Bacterial expression, folding, purification and characterization of soluble NTPDase5 (CD39L4) ecto-nucleotidase. Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics 1747:251–259. https://doi.org/10.1016/J.BBAPAP.2004.11.017

Attachments

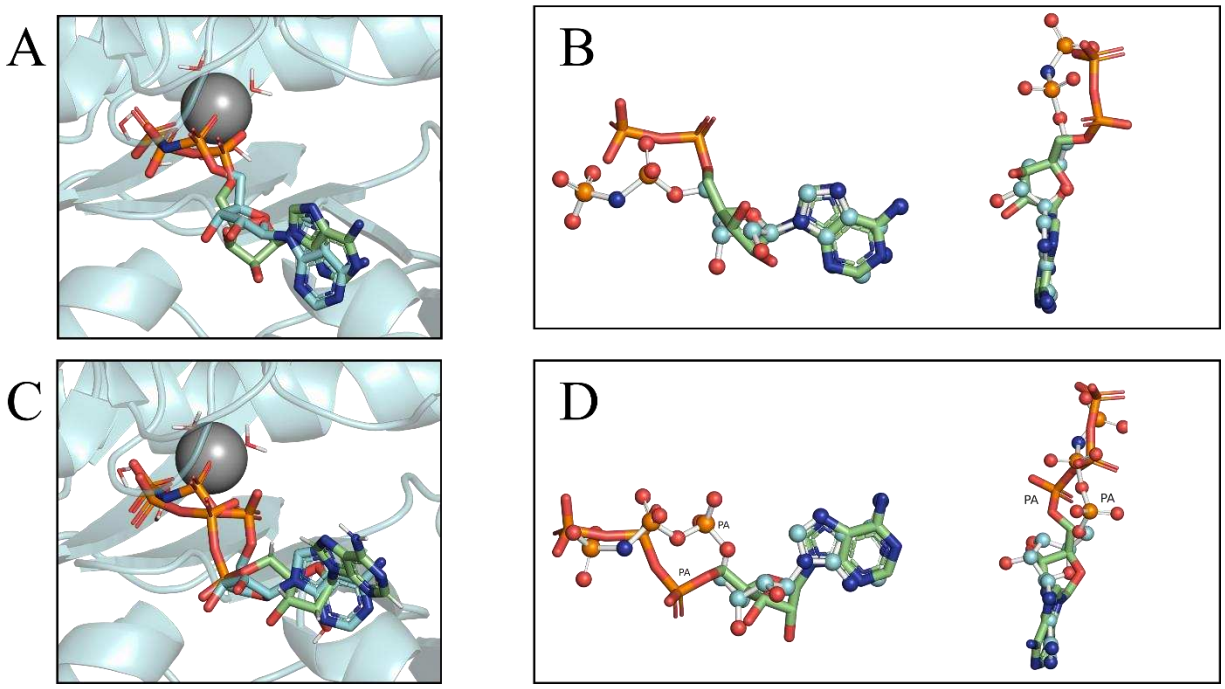

Figure S1 – Binding mode as predicted by Molecular Docking. The RnNTPDase2 crystal structures 4BR0 and 3CJA (cartoon) and the co-crystalized ADP and ATP analogs (sticks) are presented in cyan. A and C show the best-predicted pose of ADP and ATP (green sticks in both), respectively, in the RnNTDase2. B and D show the structural alignment of each best-predicted pose (green sticks) with its corresponding nucleotide analog experimentally resolved with the RnNTPDase2 (represented as balls and sticks).

To illustrate our point, the DockThor web server was employed to dock ADP and ATP on the RnNTPDase2. Figures S1B and S1D show the discrepancies between predicted and experimental substrate conformations. Despite the nitid difference, the docked and experimental NDP and NTP structures can be aligned with an RMSD of 0.543 and 1.33, respectively.

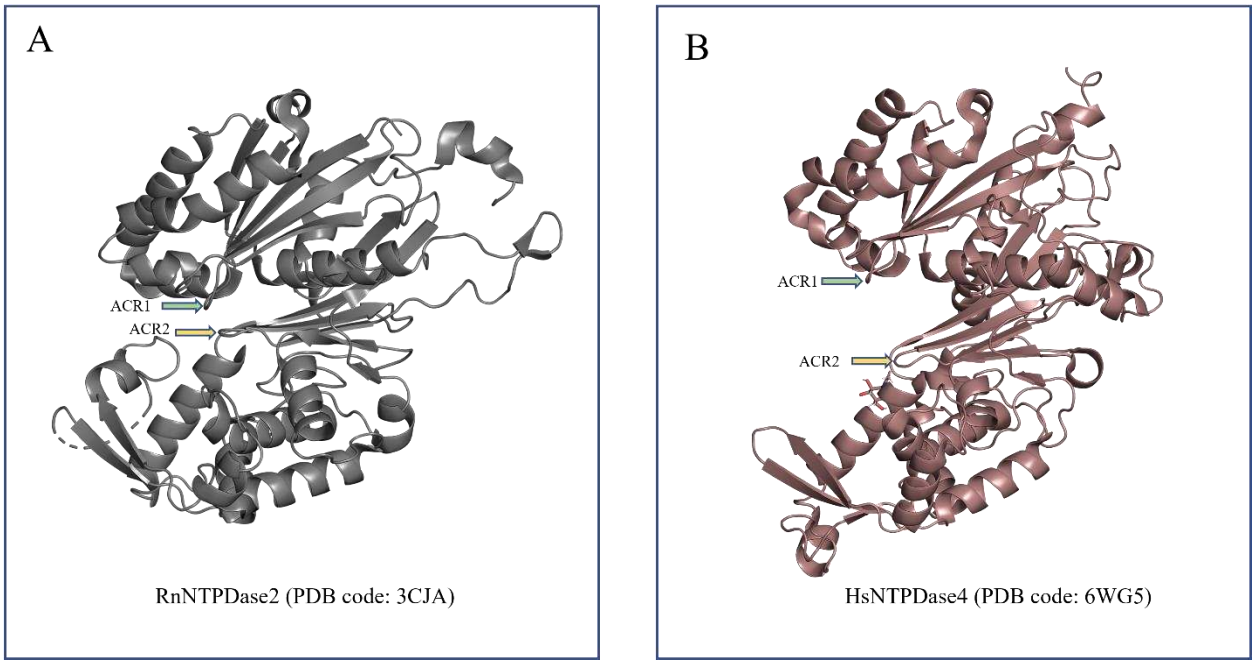

Figure S2 – Experimentally characterized closed and open states. The distance difference between ACR1 and ACR4 in the RnNTPDase2 closed conformation (A) and the HsNTPDase4 open conformation(B). The structures correspond do the PDB code 3CJA and 6WG5, respectively.

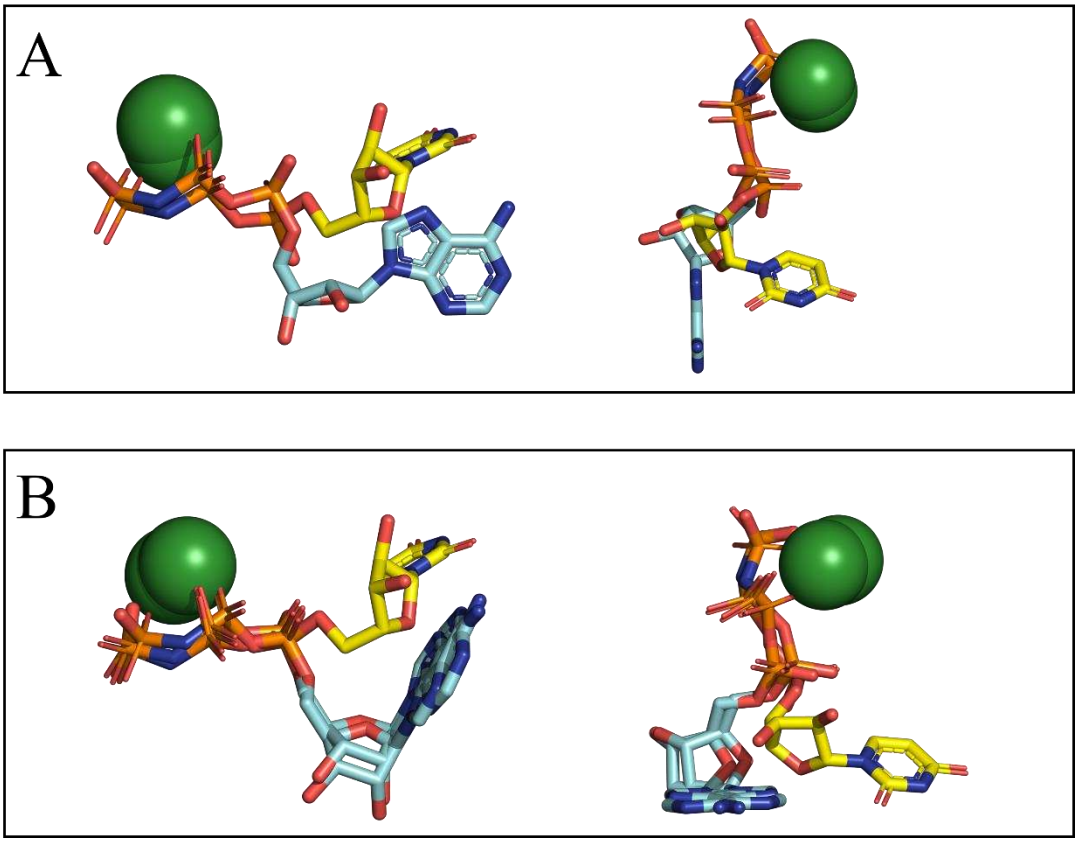

Figure S3 – Alternative-like UTP structure upon binding to LpNTPDase1. The alignment of the UTP alternative-like structure with the reference NTP structure (A) and with the ATP molecule in the alternative binding mode (B) shows the unique features of this unambiguous binding mode. To enhance visualization of the similarities and differences, two perspectives of the same alignment per section are presented: one lateral (on the left) and one from above (on the right).

The alternative-like UTP structure observed with the LpNTPDase1 (PDB code: 4BRI) is a unique conformation that remains to be rationalized. Although the phosphate moiety and cofactor reproduce the general characteristics observed in other complexes, the ribose is positioned in a unique conformation among all crystal structures.

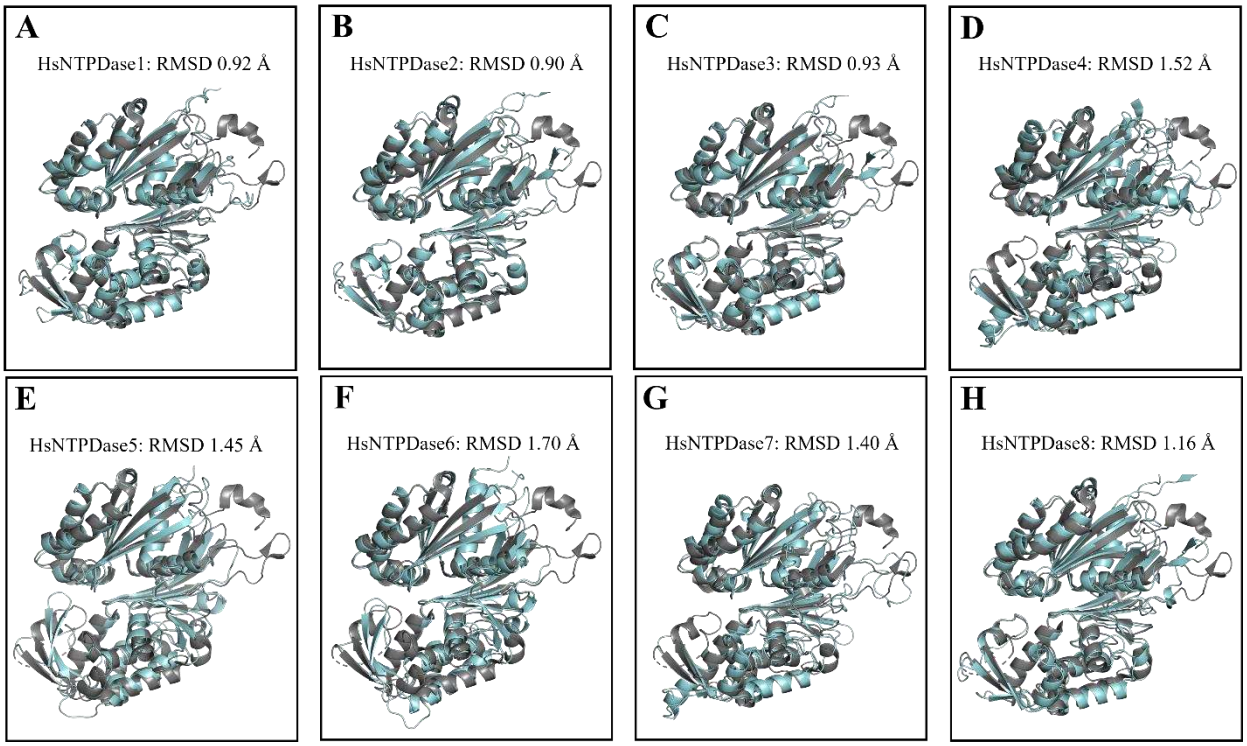

Figure S4 – HsNTPDase closed state. A-H shows the structural alignment of the HsNTPDase1-8 best model with the RnNTPDase2 (PDB code: 3CJA), respectively. By the similarity of the structures, we can ensure that all of them are in a closed state suitable for describing their complex with substrates.

Supplementary Table 1 – Modeled enzymes

Enzyme	UniProt Entry Code	Modeled residues
HsNTPDase1	P49961	46-467
HsNTPDase2	Q9Y5L3	37-452
HsNTPDase3	O75355	54-476
HsNTPDase4	Q9Y227	86-543
HsNTPDase5	O7536	26-408
HsNTPDase6	O75354	39-421
HsNTPDase7	Q9NQZ7	32-481
HsNTPDase8	Q5MY95	40-457

UniProt entry and regions of the sequence modeled of each HsNTPDase.

Supplementary Table 2 - Experimental structures with a productive substrate binding mode

Enzyme / PDB code	**Substrate Analog**	**Asp DXG ACR1 - Cofactor Distance (Å)**	**Asp DXG ACR2 - Cofactor Distance (Å)**
RnNTPDase2 / 3CJA		5.0	5.1
RnNTPDase2 / 4BQZ		4.9	4.8
RnNTPDase2 / 4BR2		5.0	5.0
LpNTPDase1 / 4BRA		4.7	4.8
LpNTPDase1 / 4BRD		4.8	4.7

LpNTPDase1 / 4BRG	4.8	4.8
LpNTPDase1 / 4BRK	4.7	4.8
TgNTPDase1 / 4KH4	3.9	5.7
TgNTPDase3 / 4A5A	4.8	4.9
RnNTPDase2 / 4BR0	4.9	5.1
LpNTPDase1 / 4BRC	4.8	4.7
LpNTPDase1 / 4BRI	4.7	4.7
LpNTPDase1 / 4BRL	4.7	4.7
LpNTPDase1 / 4BRE	4.7	4.8
TgNTPDase1 / 4KH5	4.0	6.5