
Publisher's PDF, also known as Version of record

License (if available):
CC BY

Link to published version (if available):
10.1021/jacs.6b09012

Link to publication record in Explore Bristol Research
PDF-document

This is the final published version of the article (version of record). It first appeared online via ACS at https://doi.org/10.1021/jacs.6b09012 . Please refer to any applicable terms of use of the publisher.

University of Bristol - Explore Bristol Research

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available: http://www.bristol.ac.uk/pure/about/ebr-terms
Structural Characterization of Arginine Fingers: Identification of an Arginine Finger for the Pyrophosphatase dUTPases

Gergely N. Nagy,∗†‡ Reynier Suárdíaz,§ Anna Lopata,‡∨ Olivér Ozohanics,∥ Károly Vékey,⊥∥ Beata G. Vértesesy,†⊥† Edina Rosta,∗†§∥

†Department of Biotechnology and Food Sciences, Budapest University of Technology and Economics, Budapest 1111, Hungary
§Institute of Enzymology, Research Centre for Natural Sciences, Hungarian Academy of Sciences, Budapest 1117, Hungary
∥MS Proteomics Research Group, Institute of Organic Chemistry, Research Centre for Natural Sciences, Hungarian Academy of Sciences, Budapest 1117, Hungary
‡Core Technologies Centre, Research Centre for Natural Sciences, Hungarian Academy of Sciences, Budapest 1117, Hungary
§Laboratory of Computational Biology, National Heart, Lung and Blood Institute, National Institutes of Health, Rockville, Maryland 10892-9314, United States

Abstract: Arginine finger is a highly conserved and essential residue in many GTPase and AAA+ ATPase enzymes that completes the active site from a distinct protomer, forming contacts with the γ-phosphate of the nucleotide. To date, no pyrophosphatase has been identified that employs an arginine finger fulfilling all of the above properties; all essential arginine fingers are used to catalyze the cleavage of the γ-phosphate. Here, we identify and unveil the role of a conserved arginine residue in trimeric dUTPases that meets all the criteria established for arginine fingers. We found that the conserved arginine adjacent to the P-loop-like motif enables structural organization of the active site for efficient catalysis via its nucleotide coordination, while its direct electrostatic role in transition state stabilization is secondary. An exhaustive structure-based comparison of analogous, conserved arginines from nucleotide hydrolases and transferases revealed a consensus amino acid location and orientation for contacting the γ-phosphate of the substrate nucleotide. Despite the structurally equivalent position, functional differences between arginine fingers of dUTPases and NTPases are explained on the basis of the unique chemistry performed by the pyrophosphatase dUTPases.

Introduction

Protein–protein interactions play crucial roles in ensuring optimal catalytic activity. Notably, interprotomer interactions often modulate the nucleotide triphosphate (NTP) hydrolysis and transfer reaction rates in NTP processing enzymes, e.g., AAA+ enzymes (ATPases associated with various cellular activities), GTPases, polymerases, kinases and dUTPases (DUT) assembled as homo- or heteropolymers. In many of these enzymes, an essential conserved residue, the arginine (Arg) finger, has been identified as a key entity in catalytic rate acceleration. The Arg finger constitutes a catalytically active complex with a neighboring enzyme monomer through interactions with the γ-phosphate of the nucleotide substrate. The Arg finger can originate either from a neighboring enzyme protomer (e.g., AAA+ enzymes, large GTP binding proteins) or from an effector protein (e.g., small GTPases). This residue was proposed to possess a catalytic role via nucleotide polarization and/or transition state (TS) stabilization in GTPases, ATPase motors, and helicases, and in signal transduction ATPases. Experimental data established that even the conservative replacement of the Arg finger diminished the catalytic rates while maintaining local structural integrity including P-loop coordination.

All proteins with catalytically essential Arg fingers reported in the literature to date catalyze the hydrolysis at the γ-phosphate group of the NTP. However, highly conserved arginines also assist the mechanism of NTP pyrophosphatases or transferases. Among them, dUTPases catalyze the pyrophosphorolysis of dUTP, a noncanonical DNA nucleotide. With this catalytic activity, dUTPases participate in preventive DNA repair and were recently shown to also exhibit regulatory functions related to, e.g., viral gene expression. Trimeric dUTPases host three active sites formed with the essential contribution of each protomer (Figure 1a). The nucleotide interactions at each active site are provided by five conserved motifs. The C-terminal arm of trimeric dUTPases encompasses the dynamic fifth conserved motif, which contains a conserved...
Arg directly preceding a glycine-rich P-loop-like motif (Figure 1). This Arg residue forms multiple hydrogen-bonding contacts with the γ-phosphate of the substrate and contributes to the functionally critical dUTP/dUDP substrate discrimination of dUTPases.\textsuperscript{20,21} Despite the wealth of biochemical, structural, and computational studies concerning dUTPases,\textsuperscript{20,22−27} structural and computational evidence assessing the role of this conserved Arg was missing.

Here, we determined crystallographic structures of Mycobacterium tuberculosis dUTPase (mtDUT), to establish the role of R140 as an arginine finger, performing kinetic and ligand binding measurements as well as molecular dynamics (MD) and QM/MM calculations. Our results together reveal the essential role of R140 in catalysis by primarily promoting active site organization. We further provide an exhaustive structure-based comparison of analogous conserved arginine residues present in a large number of nucleotide hydrolase and transferase enzyme classes, identifying a conserved spatial position for these positively charged Arg residues in one metal catalytic\textsuperscript{28} nucleotide hydrolase and transferase enzymes.

\section*{Methods}

\textbf{Chemicals.} DpnI restriction enzyme as well as Phusion polymerase were obtained from New England Biolabs (Ipswich, MA). IPTG was obtained from Fisher Scientific GmbH, Germany. Ni-NTA was from Qiagen. Materials for electrophoresis were from Bio-Rad. Phenol red was obtained from Merck (Kenilworth, NJ); dUTP, α,β-imido dUTP (dUPNPP), and dUDP were from Jena Biosciences. All other chemicals were of analytical grade and were purchased from Sigma-Aldrich (St. Louis, MO).

\textbf{Mutagenesis and Protein Expression.} Site-directed mutagenesis was performed by the Quikchange method (Agilent) and was verified by sequencing (MWG Eurofins GmbH, Martinsried, Germany). We interrogated the effect of a lysine surrogate to the R140 residue by performing the arginine/lysine exchange in mtDUT\textsuperscript{H145W} construct which displays similar kinetic parameters to the wild type mtDUT.\textsuperscript{29} The following primers were used to create mtDUT\textsuperscript{H145W} from the mtDUTH\textsuperscript{H145W} construct: 5′−3′ GCCCTCGACATCCCAAGGGCGACGAGGTCGGCTGGGG and 3′−5′ CCCAGGCCACCGTGGCAATCCCATGGGCCAGGGCGACGAGGTCGGCTGGGG, respectively. For mtDUT\textsuperscript{G143STOP}, the primers 5′−3′ CCCCGCCGCGACTGAGCGCCACGGTCTCCTC and 3′−5′ GAGGAACCGTGCGCCAGTGCCGGGG enabled the introduction of a stop codon in place of G143 residue of mtDUT\textsuperscript{H145W}, thus disrupting the first β-turn of the P-loop-like motif. Proteins were expressed and purified as described previously.\textsuperscript{29} Briefly, recombinant dUTPase with an N-terminal 6xHis-tag was expressed in E. coli strain BL21(DE3)-pLysS. The final supernatant after cell extraction was loaded onto a Ni-NTA column and purified according to the Novagen protocol. Proteins were dialyzed against a buffer containing 20 mM HEPES/NaOH pH 7.5, 100 mM NaCl, 2 mM MgCl\textsubscript{2}, and 1 mM DTT. Protein concentration was measured by UV absorbance using extinction coefficients $ε_{280} = 8480 \text{ M}^{-1} \text{ cm}^{-1}$ for mtDUT\textsuperscript{H145W} and $ε_{280} = 2980 \text{ M}^{-1} \text{ cm}^{-1}$ for mtDUT\textsuperscript{G143STOP}.

\textbf{ESI-MS.} In the mass spectrometric study of protein complexes, a commercial Waters QTOF Premier instrument (Waters, Milford, MA) equipped with an electrospray ionization source (Waters, Milford, MA) was used in the positive ion mode. Mass spectra were obtained under native conditions; the ions were generated from aqueous 10 mM NH\textsubscript{4}HCO\textsubscript{3} pH 7.5 buffer solution containing M. tuberculosis dUTPase protein constructs at 0.4 μM monomer concentration. These conditions allowed transfer of the native protein complex present in the solution into the gas phase. The capillary voltage was 3600 V; the sampling cone voltage was 125 V, and the temperature of the source was kept at 80 °C. Collision cell pressure was 3.38 × 10\textsuperscript{−3} mbar, and ion guide gas flow was 35.00 mL/min. Mass spectra were recorded using the software MassLynx 4.1 (Waters, Milford, MA) in the mass range 1400−5000 m/z.

\textbf{Enzyme Activity Assay.} dUTPase activity was followed by a continuous spectrophotometric phenol red assay as described in ref 30.
Briefly, the assay that measures proton release upon dUTP hydrolysis was performed in 10 mm path length cuvettes thermostated to 20 °C in a Spectro 200 spectrophotometer with a reaction volume of 1000 μL. Due to low enzyme activity of the investigated enzyme point mutants, mtDUTR140K,H145W and mtDUTG143STOP enzyme concentrations were set to 1 and 5 μM, respectively, in 1 mM HEPES pH 7.5 buffer containing 100 mM KCl, 40 μM phenol red, and 5 mM MgCl2. The reaction was started by a final quick addition of 5–80 μM dUTP to the reaction mixture; then, absorbance decrease was recorded for 1–5 min. The measured absorbance decline was corrected for the change of the absorbance for the same time interval in absence of dUTP substrate. The average of three measurements was calculated at each data point. Kinetic data were fitted with Michaelis–Menten equation.

**Circular Dichroism Intensity Titrations.** Near-UV CD spectra were recorded at 20 °C on a JASCO 720 spectropolarimeter using a 10 mm path length quartz cuvette. Protein (50 μM) was titrated with stepwise addition of αβ-imido-dUTP (dUPNPP) in a buffer containing 10 mM potassium phosphate pH 7.5 and 1 mM MgCl2 as described in ref 20. A spectrum in the range λ = 250–290 nm was recorded at each nucleotide concentration. Differential curves were obtained by subtracting the CD spectra of nucleotide–buffer titrations from that of the corresponding nucleotide–protein titration point. Differential ellipticity at λmax = 260 nm was plotted against the nucleotide concentration by using Origin software (OriginLab Corp, Northampton, MA). Dissociation constants were determined by fitting the data with quadratic eq 1:

$$y = s + \frac{A(c + x + K)}{2x} \sqrt{(c + x + K)^2 - 4cx}$$

Here $s = y$ at $x = 0$; $A$ is the amplitude of the circular dichroism intensity change, $c$ is the concentration of the protein, and $K$ is the dissociation constant.

**Crystallization.** mtDUTR140K,H145W and mtDUTG143STOP proteins were crystallized using the hanging drop method and applying the same crystallization condition that was used for the wild type enzyme.29 Briefly, 4.5 mg/mL dUTPase together with 6.7 mM dUPNPP, 20 mM MgCl2, and 0.1% sodium azide were mixed with the reservoir solution containing 0.1 M TRIS/HCl pH 7.5, 1.5 M ammonium sulfate, and 12.5% glycerol in a 1:1 μL:1 μL ratio.

**Data Collection and Structure Determination.** Complete crystallographic data sets of the acquired crystals were recorded at a Supernova sealed tube home source diffractometer (Agilent) for mtDUTR140K,H145W, as well as at the ID32-2 beamline of ESRF (Grenoble), for mtDUTG143STOP. Data were processed using XDS and XScale31 programs. An initial electron density map was obtained for both mutant enzyme:dUPNPP:Mg2+ complexes by Fourier synthesis from the phases of the mtDUTR140K,H145W (structure: 3HZA) as the model. Model building and refinement was carried out using the Coot33 program and REFMAC36 from the CCP4 suite.34 To model anisotropic displacements of the protein complexes, TLS parametrization was implemented within the program REFMAC. Later, refinement of high resolution mtDUTG143STOP structure was pursued using anisotropic B factors. At the final cycles of this refinement, B factor restraints were reduced to 0.7. Final models contain the bound substrate analogue dUPNPP:Mg2+ complex. For the sake of clarity, residues are numbered according to the physiological *M. tuberculosis* dUTPase sequence, disregarding the recombinant tag in the atomic coordinate files as well as figures and throughout the text. Coordinates and structure factor data of mtDUTR140K,H145W and mtDUTG143STOP were deposited to the Protein Data Bank with accession codes 5EDD and 5ECT, respectively (cf. Table 2).

**Molecular Dynamics Simulations.** The coordinates of the wild type (WT) mycobacterial dUTPase enzyme in complex with the isosteric substrate analogue, αβ-imido-dUTP, and Mg2+ were taken from the PDB 2PY4 structure of 1.49 Å resolution. To convert the crystallized substrate analogue to dUTP, the imido NH moiety was replaced by O. The R140K and G143STOP mutants were generated using the Mutagenesis utility of Pymol43 starting from the wild type 2PY4 structure. The simulations were set-up using standard simulation protocols with CHARMM-GUI.44,45 Hydrogen coordinates were generated using standard protonation states for all ionizable residues using CHARMM.46 The system was solvated with a pre-equilibrated TIP3P cubic water box of 80 Å3. Water molecules were randomly replaced by ions to ensure the neutralization of the system, and an additional KCl salt concentration corresponding to 0.15 M. We run production trajectories of 300 ns after 10 ns of equilibration using NAMD.47 Temperature and pressure were held constant at 300 K and 1 atm, respectively. We used the CHARM3648 force field with periodic boundary conditions and the particle mesh Ewald method49 for the long-range electrostatics in combination with a 12 Å cutoff for the evaluation of the nonbonded interactions, and the same protocol was followed for the corresponding MD simulations.

**QM/MM Calculations.** Geometries of the reactant and transition states of the phosphate cleavage reaction in WT dUTPase were taken from our previous work.24 The system was trimmed to a sphere of 21 Å radius centered at the Pz atom of the dUTP. Atoms further than 15 Å from the Pz atom of dUTP were kept frozen during all QM/MM calculations. No cutoffs were introduced for the nonbonding MM and QM/MM interactions. We have used Q-Chem v4.150,51 and Gaussian09 D050 program packages to perform the QM calculations using the B3LYP functional in combination with 6-31+G(d,p) basis sets. The quantum mechanical system was coupled with the CHARMM program52 for the MM region. A full electrostatic embedding scheme53 has been adopted in all the calculations, and hydrogen link atoms have been used to treat the QM/MM boundaries. The energy minimized reaction pathway was obtained using a QM region consisting of the dUTP molecule; the backbone of T81 and I82; the side chains of D83, S65, H145, S147, and S148; the amide group of Q13; the guanidine group of R140; the Mg2+ ion with its full coordination shell; and nearby crystallographic waters to account for 132 atoms. We investigated the reaction barrier using different QM/MM model systems. To separate the geometrical factors from the electrostatic and charge transfer interactions, we used the fully geometry optimized WT reaction profile,54 and recalculated the barrier height using different models for the R64 and R140 side chains. We included them either (i) in the QM region, corresponding to the minimized profile, accounting fully for charge transfer and electrostatic effects; (ii) in the MM region, allowing only electrostatic effects; or (iii) with its atomic charges set to zero (neutralized) during the evaluations of the QM/MM wave function and energies, only preserving its geometrical effects.

## Results

### Integriy of Both R140 and the P-Loop-like Element Is Essential for Catalytic Efficiency

We engineered two mutant proteins (mtDUTR140K,H145W and mtDUTG143STOP) to

---

Table 1. Kinetic and Ligand Binding Parameters of the dUTPase Variants

<table>
<thead>
<tr>
<th>enzyme</th>
<th>mtDUTWT</th>
<th>mtDUTR140K,H145W</th>
<th>mtDUTG143STOP</th>
<th>mtDUTR140K,H145W</th>
<th>mtDUTG143STOP</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k_{in}$ ($s^{-1}$)</td>
<td>3.0 ± 0.3</td>
<td>0.7 ± 0.04</td>
<td>0.0035 ± 0.0001</td>
<td>0.0045 ± 0.0003</td>
<td>0.00089 ± 0.00004</td>
</tr>
<tr>
<td>$K_d$ (μM)</td>
<td>1.4 ± 0.4</td>
<td>1.3 ± 0.4</td>
<td>6.7 ± 0.4</td>
<td>1.1 ± 0.3</td>
<td>8.2 ± 2.5</td>
</tr>
<tr>
<td>$K_{d,protein}$ (μM)</td>
<td>n.d.</td>
<td>0.9 ± 0.57</td>
<td>3.9 ± 1.3</td>
<td>5.7 ± 1.5</td>
<td>5.2 ± 1.0</td>
</tr>
</tbody>
</table>

*From ref 24. $K_d$ measurements were performed by near-UV CD titration, as described in the Methods section; for titration curves, see also Figure S2.*
elucidate the role of key constituents of the C-terminal arm including the R140 in substrate coordination and catalysis. Notably, the conserved arginine is replaced by a lysine in the first construct, whereas the C-terminal truncation of the second construct at the 143rd residue position (mtDUTG143STOP) disrupts only the P-loop-like element, while the R140 is

Table 2. Crystallographic Data Collection and Refinement Statistics

<table>
<thead>
<tr>
<th></th>
<th>mtDUTR140K</th>
<th>mtDUTG143STOP</th>
</tr>
</thead>
<tbody>
<tr>
<td>PDB</td>
<td>5EDD</td>
<td>5ECT</td>
</tr>
<tr>
<td>Data Collection</td>
<td></td>
<td></td>
</tr>
<tr>
<td>space group</td>
<td>P6_3</td>
<td>P6_3</td>
</tr>
<tr>
<td>cell dimensions a, c (Å)</td>
<td>55.05, 83.83</td>
<td>54.78, 84.08</td>
</tr>
<tr>
<td>resolution (Å)</td>
<td>41.9–1.97 (2.02–1.97)</td>
<td>26.0–1.30 (1.33–1.30)</td>
</tr>
<tr>
<td>wavelength (Å)</td>
<td>1.5406</td>
<td>0.8729</td>
</tr>
<tr>
<td>Rmerge</td>
<td>0.130 (0.497)</td>
<td>0.028 (0.561)</td>
</tr>
<tr>
<td>⟨I/σ(I)⟩</td>
<td>6.1 (1.6)</td>
<td>21.23 (1.95)</td>
</tr>
<tr>
<td>completeness (%)</td>
<td>99.6 (99.6)</td>
<td>99.1 (99.3)</td>
</tr>
<tr>
<td>redundancy</td>
<td>3.5 (2.5)</td>
<td>3.3 (3.3)</td>
</tr>
<tr>
<td>Reﬁnement Statistics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>resolution (Å)</td>
<td>19.19–1.97</td>
<td></td>
</tr>
<tr>
<td>no. reﬂns</td>
<td>35829</td>
<td>34820</td>
</tr>
<tr>
<td>Rwork/Rfree</td>
<td>0.160/0.211</td>
<td>0.123/0.157</td>
</tr>
<tr>
<td>no. atoms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>protein</td>
<td>1026</td>
<td>1070</td>
</tr>
<tr>
<td>ligand/ion</td>
<td>37</td>
<td>39</td>
</tr>
<tr>
<td>water</td>
<td>159</td>
<td>127</td>
</tr>
<tr>
<td>B factors</td>
<td></td>
<td></td>
</tr>
<tr>
<td>protein</td>
<td>21.0</td>
<td>22.3</td>
</tr>
<tr>
<td>ligand/ion</td>
<td>20.2</td>
<td>19.1</td>
</tr>
<tr>
<td>water</td>
<td>32.0</td>
<td>31.6</td>
</tr>
<tr>
<td>rms deviations</td>
<td></td>
<td></td>
</tr>
<tr>
<td>bond length (Å)</td>
<td>0.019</td>
<td>0.019</td>
</tr>
<tr>
<td>bond angle (deg)</td>
<td>1.95</td>
<td>1.76</td>
</tr>
</tbody>
</table>

Values in parentheses are for the highest resolution shell.

Figure 2. Flexibility and disorder of mtDUT and its mutants affecting the Arg finger and the P-loop-like elements from crystallographic structures (top) and MD simulations (bottom). Cartoon representation is shown of the (a) wild type mtDUT (PDB: 2PY4), (b) mtDUTR140K,H145W, and (c) mtDUTG143STOP structures. The local thermal motions defined by B factors in crystal structures (upper row, a–c) are compared with the observed flexibility in MD simulations (lower row, d–f). The superposition of MD simulation snapshots collected at uniform intervals is shown. The disordered C-terminal region and the loop S18–D24 are highlighted; the dUTP and the R/K140 are shown as licorice. Color code: blue, more rigid; red, more flexible. Only one dUTPase protomer is highlighted with the other two protomers shown as transparent.
retained in the sequence. The H145W mutation introduces an intrinsic Trp fluorescence sensor that we have extensively used for the characterization of the enzymatic mechanism of various dUTPases. The mtDUTH145W enzyme was reported to behave as wild type. Both mtDUTR140K,H145W and mtDUTG143STOP form homotrimers, as confirmed by ESI-MS (Figure S1, Table S1). Steady-state kinetic measurements revealed severely deteriorated dUTPase activities for these constructs with corresponding kcat values comparable to that of mtDUTT138STOP and devoid of the P-loop-like element as well as of R140. (Table 1, Figure S2). Near-UV circular dichroism equilibrium ligand binding measurements showed that both mtDUTR140K,H145W and mtDUTG143STOP have a moderate 5-fold decreased affinity toward the slowly hydrolyzable substrate analogue α,β-imido dUTP (dUPNPP) (Table 1, Figure S2).

High Resolution mtDUT Crystal Structures Provide Evidence That R140 Is Critical for Proper Active Site Organization. We determined the mtDUTR140K,H145W and mtDUTG143STOP crystal structures to 1.97 and 1.30 Å resolution, respectively (Table 2). These structures display similar overall fold as the mtDUTWT (PDB: 2PY4) with all atom rms deviations of 0.172 and 0.174 Å, respectively. In both mutant structures, the dUPNPP:Mg2+ complex is clearly visible in the enzyme active site in a conformation similar to that in the wild type enzyme (Figure S3). In addition, an electron density is present in both mutant structures at the approximate position of the catalytic water molecule. Strikingly, the complete C-terminal arm, encompassing the K140 and the P-loop-like element, becomes disordered in the R140K mutant (Figure 2, Figure S3). These alterations render the highly conserved R64 to be the sole residue that still directly interacts with the dNTP γ-phosphate. The structure of the mtDUTG143STOP also lacks the P-loop-like element, simply due to the engineered truncation. Nevertheless, the electron density corresponding to R140 together with vicinal residues (138–141) is resolved, despite the C-terminal arm segment 134–137 not being resolved in this structure. Here, the bidentate interaction of R140 with the nucleotide γ-phosphate persists. However, R64, the other interaction partner of the nucleotide γ-phosphate, is shifted by 0.9 Å away from its wild type location, contributing to a twisted conformation of the triphosphate chain (Figure S3). In both structures, the abolish conserved interactions of the nucleotide with the P-loop-like element are substituted by ordered water molecules (Figure S3). In accordance with the loss of important interactions, the triphosphate groups of the nucleotides are characterized with an increased flexibility as deduced from B factor analysis (Figure S4).

Alteration of R140 Affects Ligand Traffic. Crystallographic B factor analysis of mtDUTR140K,H145W revealed that the loop (S18–D24) between the first and the second β-sheets directly adjacent to the first conserved motif appears with increased flexibility compared to mtDUTWT (Figure 2). A comparison with other deposited mtDUT structures revealed that increased flexibility of the loop (S18–D24) is induced by exchange or deletion of R140, which engages in an interprotomer interaction with D24 (Table S2). This contact together with the H21–S139 interaction is of key importance to the ordering of the C-terminal arm to the active site as the preceding part of the arm (residues 130–138) is not involved in polar interactions with other protein segments. This anchoring of R140 further facilitates interactions of the P-loop-like element (141–148) with the triphosphate chain of the nucleotide. Molecular dynamics simulations revealed that the dynamic movements of this loop facilitate the entry and departure of the nucleotide to the active site.

MD Simulations and QM/MM Calculations Reveal the Role of Arginine Residues in the Catalytic Mechanism of dUTPase. MD simulations were performed for the mtDUTR140K, mtDUTG143STOP as well as for the mtDUTWT trimer structures with dUTP substrate present at the active sites. Each simulation was initiated using the wild type structure as template, and was run for 300 ns. In the case of the mtDUTR140K mutant, the C-terminal arm became flexible in two of the three active sites of the mtDUTR140K trimer within the first 10 ns, whereas the third active site became disordered only after about 150 ns, reflected by the distance between K140 and the γ-phosphate (Figure 2e, Figure S5). The disorder at the C-terminal region is accompanied by the increased disorder of the S18–D24 loop of the interacting protomer, indicating extensive local structural perturbations. In contrast, MD simulations of mtDUTG143STOP indicated low flexibility for the available part of the C-terminal arm together with unaltered γ-phosphate coordination by R140 (Figure 2f, Figure S5). Nevertheless, the interactions with the other highly conserved arginine, R64, and the nucleotide γ-phosphate became looser, in accordance with the structural results (Figures S3 and S5). To probe the significance of R64 and R140 coordination on the catalytic mechanism of dUTPase, we calculated the reaction barrier heights of the pyrophosphate cleavage reaction at the QM/MM level using several model systems. We found that the electrostatic influence of R140 in the catalytic step is only minor (Figure 3 and Figure S6). The change in the activation energy was within 2 kcal/mol, regardless of whether the R140 side chain was present in the QM region or in the MM region, or if it was neutralized. In contrast, the absence of R64 from the MM region rendered the enzyme virtually inactive with a reaction barrier increase by more than 10 kcal/mol (Figure 3 and Figure S6). Including R64 in the QM region did not change the reaction barrier considerably. Therefore, we suggest that the key effect of this latter residue on the catalysis is mainly electrostatic, whereas R140 does not have a major electrostatic stabilization effect for the transition state.

Structural Comparison of Arg Finger–NTP Interactions for All Enzyme Classes. To assess the general structural arrangement of Arg–NTP interactions, we analyzed the positions of Arg residues from a large number of enzyme classes (Figure 4) that form hydrogen bonds with a nucleoside triphosphate substrate from high resolution crystallographic structures (Figure S7). We selected enzyme structures with a single bound Mg2+ cofactor in the active site, which is a key common feature among dUTPases, GTPases, or AAA+ enzymes, and for all of the known Arg finger examples (Figure 5), to the best of our knowledge. The structural alignment identifies a large Arg cluster near the γ-phosphate, with ATPases and GTPases as primary examples with an Arg finger residue (Figure 4, Figure S7). A second, more scattered smaller cluster corresponds to various kinases and the R64 of dUTPase (Figure 4a). These results suggest that an arginine residue positioned to contact the γ-phosphate at the specific orientation as shown by the main cluster is generally advantageous for the γ-phosphate catalytic cleavage reactions. The fifth motif R140 in mtDUT protomers also occupies this equivalent position. Nevertheless, an important functional difference arises from the fact that dUTPases catalyze a chemically different catalytic reaction as compared with GTPases and ATPases (Figure 4). For trimeric and monomeric dUTPases, the nucleophilic attack
dUTPase showed only 30–50% decrease in activity. R64 contacts both the β- and γ-phosphates (Figure 4b), and at the same time completes the active site, at the interface between two protomers. R64 is part of a different protomer than the one coordinating the nucleoside group. In contrast to R140, our QM/MM results predicted that R64 electrostatically stabilizes the TS of the pyrophosphorolysis reaction. This is also in accordance with a previous mutational study in *Saccharomyces cerevisiae* that reported very low activity for the corresponding Ala mutant R68A. Furthermore, a positively charged residue (R/K) occupies this position almost invariably, and the highly conserved R64 is substituted by Lys in only about 3% of fully functional trimeric dUTPases (e.g., ref 22) (Figure 1c and Table S2).

R140 is a highly conserved residue among trimeric and monomeric dUTPases (Figure 1b). Its naturally occurring homologues are very rare (less than 0.3%, see Table S3), and do not favor the conserved Lys exchange. This is in good agreement with our results that conservative R140K mutation renders the enzyme inactive. Accordingly, Arg/Lys, Arg/Ala point mutants of dUTPases were also seriously compromised in their catalytic efficiency on the basis of prior studies including herpes viral, lentiviral, retroviral, yeast, and human sources.

Figure 3. Role of arginine residues in mtDUT catalysis as assessed by QM/MM calculations. (a) Transition state configuration of the *M. tuberculosis* dUTPase active site. Residues always included in the QM region (licorice, sphere for the Mg2+) and the R64 and R140 moieties (transparent thin sticks) are shown. The reaction coordinate was defined as a linear combination of distances involving the breaking and forming of bonds (yellow dashed lines): r(\(O−Pr\)) − r(\(Ow−Pr\)) + 0.5\(r(\Ow−Hw)−0.5r(\Hw−Ow−C\O\)). (b) QM/MM energy barriers for the catalytic step when the side chains of R64 and/or R140 are (i) included in the QM region (QM), (ii) included in the MM region (MM), or (iii) neutralized in the MM region (MM0) during QM/MM calculations.

occur at the α-phosphate; therefore, their cognate Arg corresponding to R140 is not involved in the direct electronic stabilization of the transition state. In contrast, the γ-phosphate serves as the reaction center upon the nucleophilic attack in GTPases and ATPases, in which the Arg fingers can indeed stabilize the transition state.

**DISCUSSION**

**Functional Roles of Conserved Arginines at the Active Site of dUTPases.** Three conserved arginine residues are present in the active site of dUTPases: R64 (second motif), R110 (fourth motif), and R140 (fifth motif). Both R64 and R140 directly contact the triphosphate chain (Figure 1c). Arg140 coordinates the nucleotide γ-phosphate, analogously to the Arg fingers in AAA+ enzymes. The fourth motif R110 does not directly coordinate the substrate nucleotide, and the corresponding R111A mutant of *Saccharomyces cerevisiae*
electrostatic and structural contributions to catalysis. More
detailed analysis could also be carried out to evaluate the
electrostatic preorganization toward the transition state, as
done, e.g., by Florian and Warshel.53 To compare with our
biochemical data, calculations of accurate free energy barriers
are needed, taking into account geometrical relaxation that
requires more extensive free energy calculations.

Identification of an Arg Finger in dUTPases. On the
basis of the following properties, we propose that R140 fulfills
the requirements of being an Arg finger: (i) It is a highly
conserved residue. (ii) Its mutation (even conservative)
seriously compromises the catalytic activity. (iii) It emerges
from a separate protein subunit that forms the substrate binding
site and establishes interactions with the γ-phosphate of the
substrate NTP. The conserved arginine residue of dUTPases
thus displays functional and structural similarities to the Arg
finger motifs of AAA+ enzymes, GTPase activating proteins
(GAPs) for small G proteins, and other nucleoside triphosphate
hydrolyzing enzymes that form interactions with the γ-
phosphate of the substrate. For a more in-depth structural
analysis, we provided a thorough structure-based comparison of
arginines and, within, Arg fingers employed by NTP hydrolase
enzymes (Figure 4). Importantly, while the P-loop-like element
was identified in other pyrophosphatase enzymes as well,54−56
they do not employ an Arg finger. Notably, conserved Arg
residues are also present in two metal-ion catalyst enzymes57
(e.g., dimeric dUTPases, polymerases); however, these residues
originated from the same protein subunit as the protomer
forming the active site. Therefore, these are not considered Arg
fingers. Moreover, the positioning of the arginines in these
enzymes does not seem to follow an obvious cluster around the
triphosphate chain (Figure S8).

Interestingly, R64 also fulfills the requirements for an arginine
finger, thus presenting dUTPases with two arginine fingers.

Figure 4. Arg fingers from various NTP hydrolase/transferase enzymes. (a) Arg residues occupy clusters in specific positions around the γ-phosphate in NTP cleaving enzymes. The nucleoside group (gray) is shown for dUTPase. The triphosphate group and its coordinating Arg residues are shown as sticks (blue for the same protomer as the one that coordinates the substrate, green or pink for additional different protomers). Mg\(^{2+}\) ions are shown as green spheres. (b) Analogous positions of the Arg fingers around the nucleotide for dUTPase (PDB: 2PY4,58 note the presence of αβ-
imido dUTP substrate analogue instead of dUTP) and for the MoeB protein (transparent, 1JWA59). (c, d) Analogous positions of the Arg fingers from a GTPase (PDB: 2QTV, c) and an ATPase (PDB: 2JIZ, d). Note that the nucleophilic attack (red arrows) of the respective nucleophile water molecule (W_{cat}) takes place on different phosphate moieties for the pyrophosphatase dUTPases and MoeB vs ATPases and GTPases.
Dual essential arginine fingers were reported for several GTPase and AAA+ enzymes, including Myo9b-RhoGAP,\textsuperscript{58} ClpB/Hsp104,\textsuperscript{59} ClpA, ClpC, p97/VCP/Cdc48,\textsuperscript{60} and NSF.\textsuperscript{61} In these enzymes the available structural data suggests that the second arginine contacts the nucleotide sugar group, not the triphosphate chain. Our structural analysis over many NTP hydrolysis enzymes has identified that there is a well-defined structural region occupied by arginines (Figure 4a) around the γ-phosphate. This supports a consensus position of arginine fingers (either intrinsic\textsuperscript{50} or “classical” from a distinct subunit) corresponding to R140, which is located at the same region with respect to the phosphate chain and the magnesium ion in a large number of nucleotide triphosphate converting enzymes.

Besides dUTPases, we identified only one other example among NTP pyrophosphatase or pyrophosphorylase enzymes that employs an Arg finger (Figure 5). This enzyme family performs various essential signaling and regulatory functions, and belongs to the eukaryotic ubiquitin activating (E1) enzymes\textsuperscript{62} or to the analogous prokaryotic protein families, e.g., MoeB/MoaD\textsuperscript{63} and Thif/ThiS\textsuperscript{64} playing roles in the biosynthesis of molybdopterin, and thiamine, respectively. These Arg fingers, however, are not required for catalytic activity with the respective Ala mutations representing an active enzyme.\textsuperscript{65} Note also that the currently available crystallographic structures do not have a Mg\textsuperscript{2+} ion cofactor bound, which may affect the positioning of the Arg residues (in particular for Thif, see Figure S7).

While the E1-like pyrophosphatase enzymes were reported to possess an Arg finger, they do not strictly fulfill all the required criteria, as the R/K or even R/A mutations do not affect the catalytic activity\textsuperscript{66} (cf., criteria for Arg fingers above). This fact implies that E1-like pyrophosphatases in addition to dUTPases do not use their structurally analogous Arg fingers for the electrostatic stabilization of the TS. This contrasts the functional role of Arg fingers from ATPases and GTPases where these catalytically essential residues directly stabilize the TS. Nevertheless, as the R140 Arg finger structurally positions the P-loop-like motif of dUTPases, its presence is critical for ensuring optimal catalytic activity.

**Evolutionary Development of dUTPases: Recruiting the Arg Finger and P-Loop-like Motif, and Emergence of Novel Functions.** The β-fold dUTPase superfamily includes dCTP deaminases (DCD), bifunctional dCTP deaminase/dUTPase (DCD-DUT), and dUTPases. These enzymes share similar overall fold and subunit assembly, yet the active site of DCD is assembled from only two and not from three subunits.\textsuperscript{64} Among these enzymes, only monofunctional dUTPases employ the P-loop-like motif, even though DCD-DUT is also acting as a pyrophosphatase. It has been proposed that the Last Universal Common Ancestor (LUCA) of dUTPases could have possessed bifunctional DCD/DUT activity.\textsuperscript{66} The monofunctional DUT enzymes were evolved later from this ancient promiscuous enzyme, according to the common proposed model of enzyme catalysis evolution.\textsuperscript{66} An Arg or Lys, not coupled to a P-loop-like motif, is present in some DCD-DUTs at a sequential position corresponding to the fifth motif Arg (R140) of trimeric DUTs.\textsuperscript{67} On the basis of the currently available structural information, however, these residues do not coordinate the cognate dCTP or dUTP nucleotides (e.g., PDB 2HXD). Instead, another conserved Lys coordinates the γ-phosphate of the nucleotide in DCD and DCD-DUT enzymes.\textsuperscript{68} On the basis of these observations, we suggest that the acquisition of the P-loop-like motif might have enabled the antecedent arginine to adopt the function of an Arg finger. The novel Arg finger function together with the P-loop-like fifth motif probably contributed to their increased substrate specificity and dUTPase catalytic rate acceleration.\textsuperscript{20}

Intriguingly, it has recently been revealed that some trimeric dUTPases possess additional functionality beyond their role in the nucleotide metabolism.\textsuperscript{17,23,47} Moonlighting activity generally associated with viral transfection and expression regulation was reported for many viral dUTPases.\textsuperscript{18,68–70} It has been previously suggested that several viruses acquired their dut gene from their host organism via horizontal gene transfer.\textsuperscript{71} Although the underlying molecular mechanism is not yet fully deciphered, the acquired fifth motif together with other elements was suggested to play a role in the observed moonlighting function,\textsuperscript{72} and within, the involvement of the Arg finger may also be envisaged. Nevertheless, in some cases the novel immunoregulatory function of viral dUTPases was presented in the absence of cognate dUTPase activity\textsuperscript{73} and of the P-loop-like motif.\textsuperscript{74,75}
CONCLUSION

We showed that the Arg from the fifth motif (R140) fulfills the criteria used to identify Arg fingers, therefore associating this conserved arginine of dUTPases with Arg fingers for the first time. We demonstrated that this highly conserved residue is required for full catalytic activity, promoting catalysis while being located on distinct subunit and exerting its effects on interprotomer interactions around the substrate. Functional differences between Arg fingers of dUTPases and of other Arg finger-employing P-loop NTPases may be explained on the basis of the different role of the cognate Arg finger in the catalytic mechanism. Notably, currently known enzymes with Arg fingers catalyze the cleavage between the β−γ phosphate groups of the respective nucleotides, with the single exception of E1-like enzymes, for which, however, the conserved Arg fingers are not essential for catalysis. As the cleavage takes place between the α−β phosphates for pyrophosphatases, R140 being further from the cleavage site has a more pronounced structural role, whereas R64, another conserved Arg residue from a different subunit, can be directly involved in TS stabilization. Together, these findings support our hypothesis that the role of the Arg finger in dUTPases is to correctly position the P-loop residues.

The distinct structural composition of dUTPases renders them highly analogous to NTPases, and to GTPases in particular, albeit with a unique chemistry. It remains to be explored how these structural features are related to the recently identified moonlighting function that dUTPase enzymes play in viral signaling and how their evolutionary development took place.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.6b09012.

Additional figures, including ESI-MS spectra, steady-state kinetic measurements and ligand binding isotherms, active site structural observations, QM/MM energy profiles, structural comparison, and position of conserved Arg residues in enzymes with two metal-ion catalytic mechanism; additional tables, including molecular masses, crystallographic B factor analysis, and conservation of active site Arg positions in dUTPases (PDF).

AUTHOR INFORMATION

Corresponding Authors

*nyag.gergely@ttk.mta.hu
*vertessy@mail.bme.hu
*edina.rosta@kcl.ac.uk

Present Addresses

1. Laboratory of Molecular Physiology, National Heart, Lung and Blood Institute, National Institutes of Health, Bethesda, MD 20892, United States, and Astbury Centre for Structural Molecular Biology, Faculty of Biological Sciences, University of Leeds, Leeds LS2 9JT, United Kingdom.
2. Szent István University, Ybl Miklós Faculty of Architecture and Civil Engineering, Budapest 1442, Hungary.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

E.R. gratefully acknowledges Dr. Kei-ichi Okazaki for helpful discussions, and P.G. Jambrina for help with structural alignments. We acknowledge the European Synchrotron Radiation Facility for provision of synchrotron radiation facilities, and we would like to thank Alexander Popov for assistance in using beamline ID23-2. We gratefully thank Veronika Harmat for home source and synchrotron data collection, respectively. We acknowledge computer time on ARCHER granted via the UK High-End Computing Consortium for Biomolecular Simulation, HECBioSim (http://www.hecbiosim.ac.uk), supported by EPSRC (Grant EP/L000253/1) and the computational resources of the NIH biowulf cluster. R.S. acknowledges the EC for a Marie Curie fellowship (Project 622711). E.R. acknowledges the BB/N007001/1 BBSRC grant. This work was supported by the National Research, Development and Innovation Office Hungarian Scientific Research Fund [OTKA K115993, K119493, K104986], the MedinProt program of the Hungarian Academy of Sciences, the International Centre for Genetic Engineering and Biotechnology [ICGEB CRP/HUN14-01], and the European Commission FP7 Biostruct-X project [Contract 283570]. J.T. is supported by the Bolyai János Research Scholarship of the Hungarian Academy of Sciences.

REFERENCES
