Supporting Information

Structural basis for substrate selectivity and nucleophilic substitution mechanisms in human adenine phosphoribosyltransferase catalyzed reaction

Mohammad Ozeir, Jessica Huyet, Marie-Claude Burgevin, Benoît Pinson, Françoise Chesney, Jean-Marc Remy, Abdul Rauf Siddiqi, Roland Lupoli, Gregory Pinon, Christelle Saint-Marc, Jean-Francois Gibert, Renaud Morales, Irène Ceballos-Picot, Robert Barouki, Bertrand Daignan-Fornier, Anne Olivier-Bandini, Franck Augé, Pierre Nioche

Table of contents:

Figure S1:	2
Figure S2:	4
Note S1:	5
Figure S3:	6
Note S2:	7
Table S1:	8
Figure S4:	9
Figure S5:	10
Figure S6:	11
References:	12

Figure S1: Structure-based sequence alignment of APRT from *Homo sapiens* (human), *Macaca fascicularis*, *Panthera tigris*, *Camelus ferus*, *Rattus norvegicus*, *Mus musculus*, *Danio rerio*, *Drosophila melanogaster*, *Caenorhabditis elegans*, *Saccharomyces cerevisiae*, *Arabidopsis thaliana*, *Synechocystis sp. PCC6803*, *Escherichia coli*, *Shewanella oneidensis*, *Bacillus subtillis str.168*, *Clostridioides difficile*, *Dictyostelium discoideum*, *Streptococcus pneumonia*, *Pseudomonas aeruginosa*, *Neisseria meningitides*, *Thermatoga maritima*, *Parabacteroides sp. YL27*. Secondary structures were defined using the DSSP server (1). The hood region, the flexible loop and the PRPP binding motif are indicated. The strictly conserved residues are black shadowed.



								Fleit	xible loop
		α3	β3		α4	β4		β5	β6
		_	60	70 I	80 I		90 I	100 	110
Macaca Panthera Camelus Rattus Mus Danio Drosophila Caenorhabditis	58 18 24 53 52 52 57	THG-GRI A CPP-TPQ THG-GKI THS-GKI TYP-QV SAPEA TYG-HY	DYIAGLD NELLEGLD VSPTGLD DYIAGLD DYIAGLD VDLIVGLD LEVIVGLD	SRGFLFGPS SRGFLFGPS SRGFLFGPS SRGFLFGPS SRGFLFGPL SRGFLFGPL SRGFLFGPL	LAQELGI LAQELGI LAQELGU LAQELGV LAQELGV LALRLGI IATELGI VATOLGV	GCVLIRK GCVLIRK GCVLIRK GCVLIRK GCVLIRK GFAPIRK GCAPIRK	RGKLPGPT RGKLPGPT RGKLPGPT QGKLPGPT KGKLPGPT KGKLAGEV KGKLAGEV	VWASYA VSASYA VSTSYA VSASYS ISVAYS VSVEYA	LEYGKAELE LEYGKAELE LEYGKAELE LEYGKAELE LEYGKAELE LEYGKAELE LEYGIDTFE KEYGEDBVE
Saccharomyces Arabidopsis Synechocystis Escherichia Shewanella Bacillus Clostridioides Dictyostelium Streptococus	54 47 55 47 47 47 60	AFPEVKI MQI QELVF AGI KGF KQI KDV DV	DYIVGLE SVVAGVE DHVVGME TKVVGTE DLVVGPE DLVVGPE DLVVGPE	SRGFLFGPT ARGFLFGPS SRGFLFGAP ARGFLFGAP ARGFIIGCP ARGFIIGCP ARGFILGTA SRGFILCCAP	LALALGV IALAIGA VALGLGV VALGLGU VAYALGV FAQRMKL	GFVPVRK KFIPLRK GFIPVRK GFVPVRK GFVPVRK GFAPVRK GFVPIRK PMMMIRK	AGKLPGEC PGKLPGKV PGKLPAPV PGKLPREI PGKLPREV PGKLPGEV KGKLPGEV	FKATYE ISESYE ISETYI ISQSYE IKVDYG FRESYE TSOSYE	KEYGSDLFE LEYGHDRLE DLEYGTDQLE LEYGHDSLE LEYGHDSLE LEYGHDSLE LEYGHDFE LEYGTDTLE KEYGTDFFE
streptococcus Pseudomonas Neisseria Thermatoga Parabacteroides	48 52 57 47 48	ADF QKI REF KGV	SHIGAMD DIVAGLD DLVVAPE TKVVGIE	SRGFLVGAP ARGFLIGSA ARGFILGAA S <mark>RGF</mark> IGGAI	VAHQMGI VAYALNK LAYQLNV MAYKLGK MAYEIGA	GEIPARK PLVLFRK GFVPIRK GFVPVRK	PGKLPREV QGKLPADV KGKLPFEI PGKLPYKI P <mark>GKL</mark> PADI	USQSIF VSQSIA VYEEYQ VRADFA	LEYGEATLE DEYGEAFLE LEYGEAAVE LEYGTEQLH KEYGTDTIE

			•	PRPP binding mo	tif			
		β6		β7	α5	5	β 8	
		7	120	130	140	150		
Ното	112	IQKDALEP	ĠQR <mark>V</mark> N	VVVDDLLAT	GGIMNAACE	LGRLQAEVLE	cv	SLVE
Macaca	117	IQKDALEP	GQR <mark>V</mark> V	7VVDDLLAT	GGTMHAACE	LGRLQAEVLE	CV	SLVE
Panthera	72	IQRDALEP	GQKV	/VVDDLLAT	GGTMRAACE	LLGQLRAEVLE	CV	SLVE
Camelus	83	IQRDALEP	GQKV	/VVDDLLAT	GGTMRA ACE	LGQLQAEVLE	CV	SLVE
Rattus	112	IQKDALEP	GQKV	/IVDDLLAT	GGTMCAACE	LLSQLRAEVVE	CV	SLVE
Mus	112	IQKDALEP	GQRV	/IVDDLLAT	GGTMFA ACD	LLHQLRAEVVE	CV	SLVE
Danio	110	MQEDAVSA	GQKVI	LIIDDLLAT	GGTLYA AIE	LIKQQKAEVLG	CI	VVVE
Drosophila	116	LQKSAIKP	GQKV	7VVDDLLAT	GGSLVA ATE	LIGKVGGVVVE	SI	VVME
Caenorhabditis	115	IQEGAIKN	GDIVI	FLIDDLLAT	GGTLRA ATD	LVVKAGGKVGE	AF	'VLIE
Saccharomyces	114	IQKNAIPA	GSNV1	IVDDIIAT	GGSAAA <mark>a</mark> age	LVEQLEANLLE	YN	FVME
Arabidopsis	117	MHVGAVEP	RERVI	IIIDDLVAT	GGTLS<mark>A</mark>AMS	LLGFFLSSCFL	FC	FLKR
Synechocystis	105	IHQDAVAP	HHRVI	LIVDDLIAT	GGTAK<mark>A</mark>TAE	LLTKLGCEVLG	FA	FIIE
Eschericĥia	111	IHVDAIKP	GDKVI	LVVDDLLAT	GGTIEATVK	LIRRLGGEVAD	AA	FIIN
Shewanella	111	IHTDAITA	NDKVI	LVVDDLLAT	GGTIEATVK	LIRQLGGEVQD	AA	FVIS
Bacillus	103	IHKDAIKP	GQRVI	LITDDLLAT	GGTIEATIK	LVEELGGVVÃG	IA	FLIE
Clostridioides	103	IHKDAIKK	GOKVA	AIVDDLLAT	GGTMEA AAK	LVEKLGGEVVS	MC	FLIE
Dictyostelium	115	VQEKALSKVVVK	PSKKYHVI	LIMDDILAP	GGTMAASIE	LTKKVLINNGI	KDFKISTS	LISS
Streptococcus	104	MHKDAIKP	GDKVI	LITDDLLAT	GGTIEATIK	LVEALGGEVVG	LA	FLIE
Pseudomonas	108	VHADSLCE	GDSVI	LIFDDLIAT	GGTLLAAAS	LVRRLGARVFE	AA	AIID
Neisseria	113	IHTDAVKP	GSRVI	LVDDLVAT	GGTMLAGLE	LIRKLGGEIVE	AA	AILE
Thermatoga	103	IHEDAIEK	gokvi	LIVDDVLAT	GGTAEALIR	LVKKLGGEVVS	LA	FLVE
Parabacteroides	104	IHRDAITP	DÕVVV	/IHDDLLAT	GGTMA <mark>A</mark> CYE	LVKSMNPKKVY	in	FIVE

		β8	α6		β9
		160		170	180
Ното	159	LTSLK	GREKL	APVP	FFSLLQYE
Macaca	164	LTSLK	GREKL	APVP	FFSLLQYE
Panthera	119	LTSLK	GREKL	GAVP	VFSLLQYD
Camelus	130	LTSLK	GREKL	GAVP	FFSLLQYK
Rattus	159	LTSLK	GREKL	GPVP	FFSLLQYE
Mus	159	LTSLK	GRERL	GPIP	FFSLLQYD
Danio	157	LKYLN	GSDKL	KPTP	VFSLIQY
Drosophila	163	LVGLE	GRKRL	DG-K	VHSLIKY
Caenorhabditis	162	LAPLN	GRSKL	PDVN	LTTLISYDSA
Saccharomyces	161	LDFLK	GRSKL	NA-P	VFTLLNAQKEALKK
Arabidopsīs	164	IFWS-			
Synechocystis	152	LAALN	GRQCL	PDLP	IISLVEY
Escherichia	158	LFDLG	GEQRL	EKQG	ITSYSLVPFPGH
Shewanella	158	LPDLG	GEARL	TALG	LELVKLCEFEGE
Bacillus	150	LSYLD	GRNKL	EDYD	LTLMKY
Clostridioides	150	LKFLN	GREKL	SNYD	VNSLIKY
Dictyostelium	175	IKVLN	GKEKI	YEKYI	NDVSVDIIIEM
Streptococcus	151	LNALH	GRDCL	DGYE	LLALMNY
Pseudomonas	155	LPELG	GSTRL	QDAG:	ISTFSLTAFALDER
Neisseria	160	FTDLQ	GGKNI	RASG	APLFTLLQNEGCMKG-
Thermatoga	150	LSYLE	PRKRL	EGYD	VKTLIVY
Parabacteroides	151	LSDLH	GRDNL	PKDAI	EVTSLIVY
Shewanella Bacillus Clostridioides Dictyostelium Streptococcus Pseudomonas Neisseria Thermatoga Parabacteroides	158 150 150 175 151 155 160 150 151	LPDLG LSYLD LKFLN IKVLN LNALH LPELG FTDLQ LSYLE LSDLH	GEARL GRNKL GREKL GKEKI GRDCL GSTRL GGKNI PRKRL GRDNL	TALG EDYD SNYD YEKYI DGYE QDAG RASG EGYD PKDA	LE LVKLCEFEGE IL TLMKY NDVSVDIIIEM LLALMNY IS TFSLTAFALDER AP LFTLLQNEGCMK VK TLIVY E VTSLIVY

Figure S2 : Structure of a PRPP-Mg²⁺-hAPRT monomer (PDB_ID: 6FCH) showing the hood region (cyan), the flexible loop (green) and the PRPP binding motif (blue). The strictly conserved residues shown in Figure S1 are illustrated in pink. The PRPP molecule, in orange, point to the position of the active site.



Supporting Note S1 : In all of the structures reported here, the ligand occupancies were estimated by 1) inspecting the OMIT map Fo-Fc electron density; 2) comparing the ligand B-factors to the surrounding amino acids B-factors; 3) calculating the RSCC (Real Space Correlation Coefficient), RSR (Real Space R-value) and LLDF (Local Ligand Density Fit) values (2). Then, the estimated occupancy values were based on a continuous electron density map calculated at 3 σ , comparable B-factors for both ligands and surrounding amino acids, low RSR, high RSCC (> 0.90) values and LLDF values near or lower than 2.

Figure S3 : Interaction of AMP in the hAPRT active site (PDB_ID: 6FCL).



Supporting Note S2: Details on the possible $S_N 2$ pathways. a) A two-step anionic $S_N 2$ pathway: The first step is a shift of the hydrogen on the purine/N7 position to Glu104. The resulting negative charge on the purine is transferred from position N7 to N9, which favors the nucleophilic attack toward the C1' position of PRPP in a second step, which leads to the formation of the nucleotide and a pyrophosphate that inherits the negative charge initially carried by Glu104; b) A one-step neutral pathway: A concerted pathway where the proton shift and the nucleophilic attack occur simultaneously, the negative charge of Glu104 is transferred directly to the pyrophosphate without the formation of any intermediate; c) A two-step cationic $S_N 2$ pathway: The nucleophilic attack of the purine/N9 nitrogen towards the C1' position of PRPP is the first event, which leads to the formation of a pyrophosphate and a nucleotide protonated on its N7 position. This positively charged nucleotide intermediate transfers its proton to Glu104 in a second step.

Table S1: Comparison of the magnesium ion geometry in human APRT wild type and variant structures.

The atom configuration is indicated in the figure below. The values discussed are indicated in red.



	Hx-PRPP- hAPRT ^{wt}	ADE- PRPP- hAPRT ^{wt}	PRPP- hAPRT ^{wt}	ADE-PRPP- Mg ²⁺ - hAPRT ^{Y105F-} ^{14days} (14 days)	ADE-PRPP- Mg ²⁺ - hAPRT ^{Y105F-} ^{14days} (14 days) (mol B. with	ADE-PRPP- Mg ²⁺ - hAPRT ^{Y105F-} ^{30days} (30 days) (mol B. with
				AMP)	PRPP)	AMP)
PDB_ID	6HGQ	6FCI	6FCH	6FD5	6FD5	6FD6
References	This work	(3)	(3)	(3)	(3)	(3)
			•			
Bond distance (Å)						
01'-Mg ²⁺	2.2	2.2	2.4	1.8	2.3	1.7
O2'-Mg ²⁺	2.1	2.1	2.4	2.7	2.4	2.8
O3'-Mg ²⁺	2.2	2.3	2.3	2.2	2.4	2.2
OP _B -Mg ²⁺	2.0	2.0	2.2	2.2	2.2	2.2
H ₂ O(1)-Mg ²⁺	2.1	2.1	2.3	2.3	2.4	2.2
H ₂ O(2)-Mg ²⁺	2.1	2.1	2.1	2.3	2.1	2.5
Angle between ato	oms (°)			1	1	
01'-Mg ²⁺ -02'	75	74	69	79	72	80
01'-Mg ²⁺ -03'	84	84	78	88	82	86
02'-Mg ²⁺ -03'	78	75	71	63	70	64
O1'-Mg ²⁺ -H ₂ O(2)	89	87	91	92	85	80
O1'-Mg ²⁺ -OP _B	95	94	95	95	95	91
H ₂ O(1)-Mg ²⁺ -O2'	97	106	96	110	93	110
H ₂ O(1)-Mg ²⁺ -O3'	87	92	89	93	88	95
H ₂ O(1)-Mg ²⁺ - H ₂ O(2)	97	92	102	78	107	90
$H_2O(1)-Mg^{2+}-OP_B$	93	89	96	84	96	87

Figure S4: Octahedral coordination of the magnesium ion in the ADE-PRPP-hAPRT^{Y105F-14days} complex structure 14 days post-crystallization (first subunit) (3).



Figure S5 : Superimposition of both molecules of the asymmetric unit in the ADE-PRPP-Mg²⁺-hAPRT^{Y105F-14days} variant structure showing the displacement of the magnesium ion. The magnesium ion in the ADE-PRPP-Mg²⁺-hAPRT^{Y105F-30days} second subunit is not shown for clarity but superimposes to the ion colored in magenta (3).



Figure S6: Superimposition of AMP-hAPRT^{wt} (magenta) onto the first subunit of ADE-PRPP-Mg²⁺hAPRT^{Y105F-14days} structure (cyan), which showed the formation of the products AMP and PPi *in crystallo* (3).



REFERENCES:

- 1. Kabsch, W., and Sander, C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*. **22**, 2577–2637
- 2. Deller, M. C., and Rupp, B. (2015) Models of protein–ligand crystal structures: trust, but verify. *J. Comput. Aided Mol. Des.* **29**, 817–836
- 3. Huyet, J., Ozeir, M., Burgevin, M.-C., Pinson, B., Chesney, F., Remy, J.-M., Siddiqi, A. R., Lupoli, R., Pinon, G., Saint-Marc, C., Gibert, J.-F., Morales, R., Ceballos-Picot, I., Barouki, R., Daignan-Fornier, B., Olivier-Bandini, A., Augé, F., and Nioche, P. (2018) Structural Insights into the Forward and Reverse Enzymatic Reactions in Human Adenine Phosphoribosyltransferase. *Cell Chem. Biol.* **25**, 666-676.e4