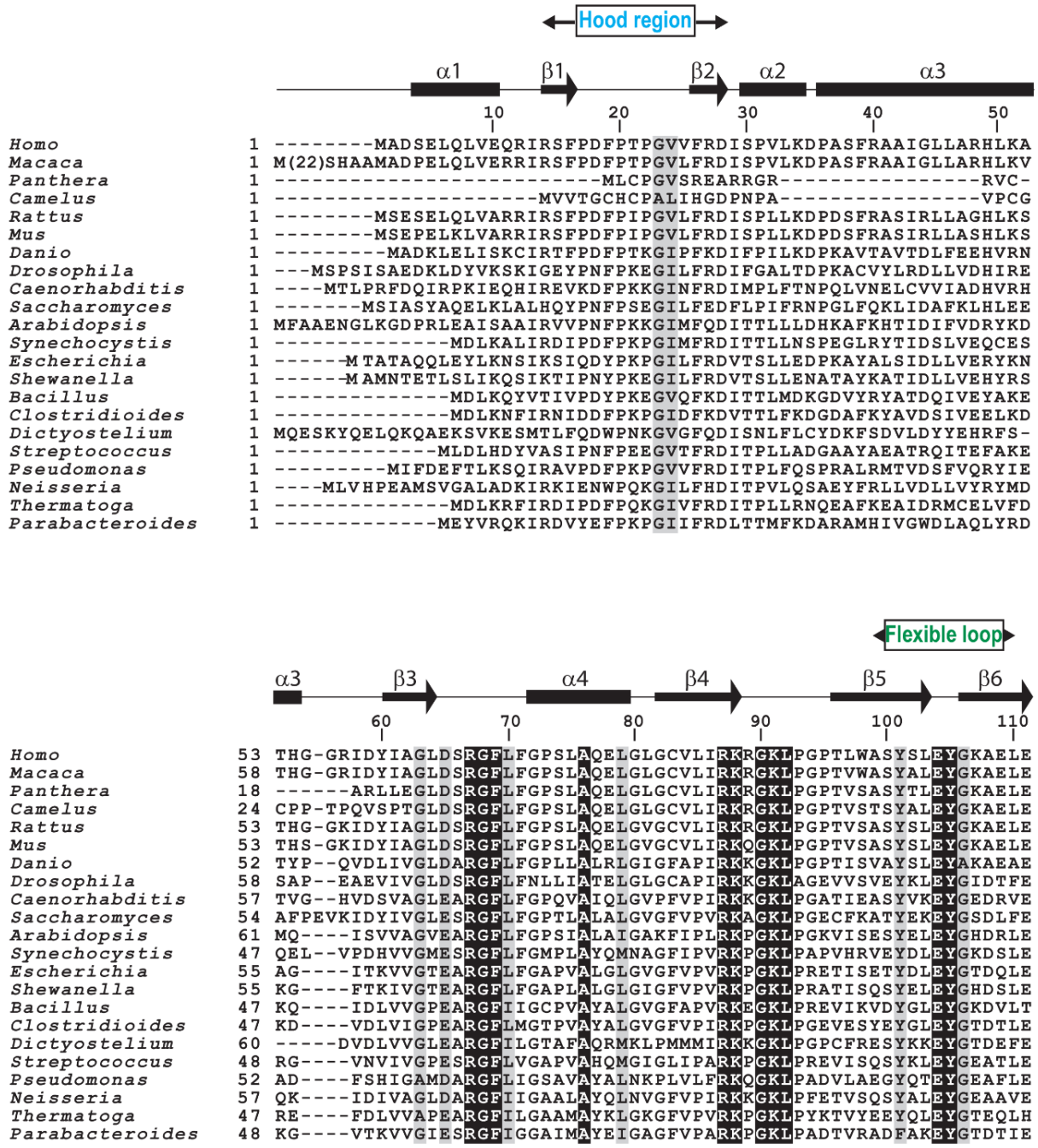
Supporting Information

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 : 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.



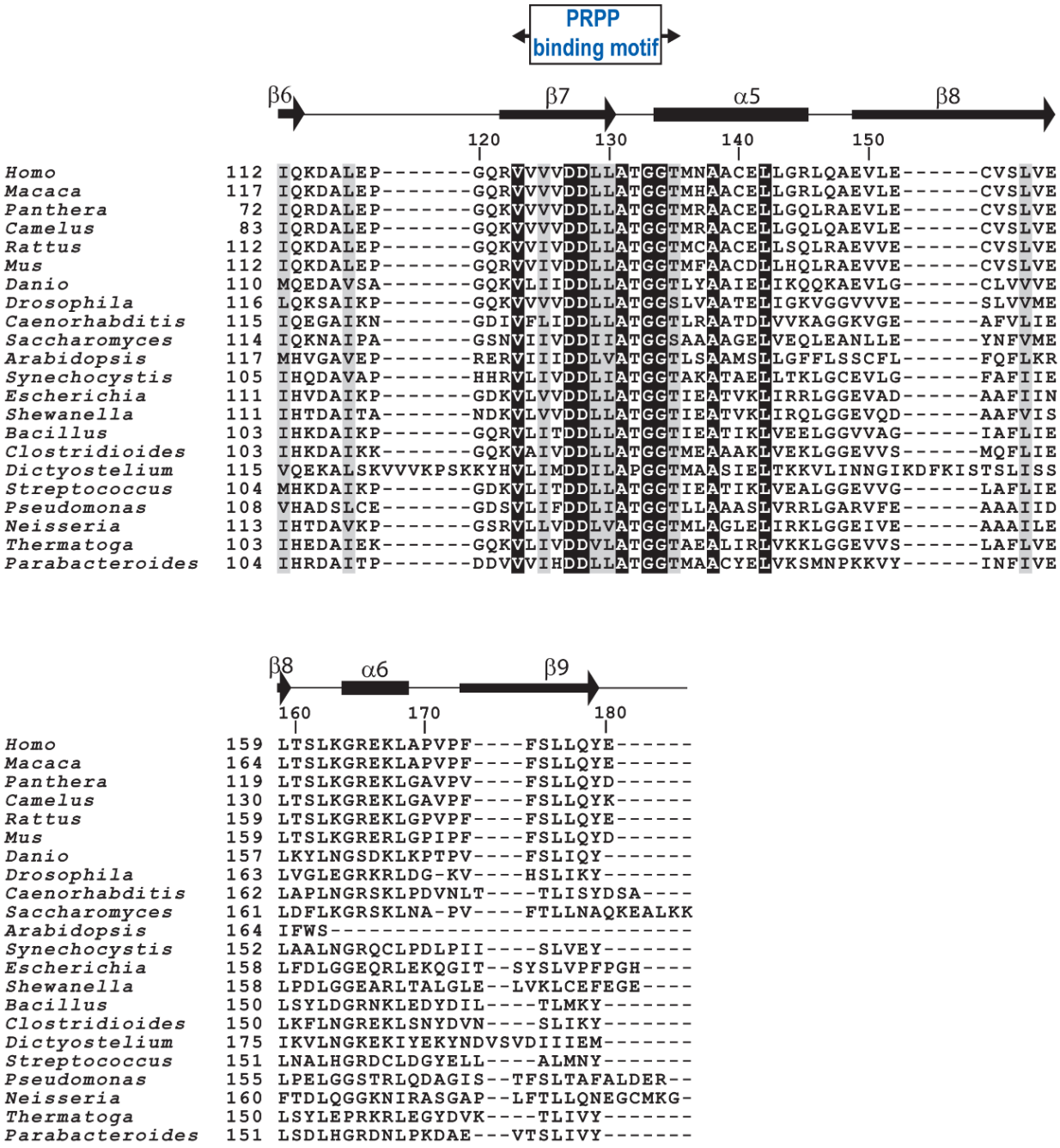
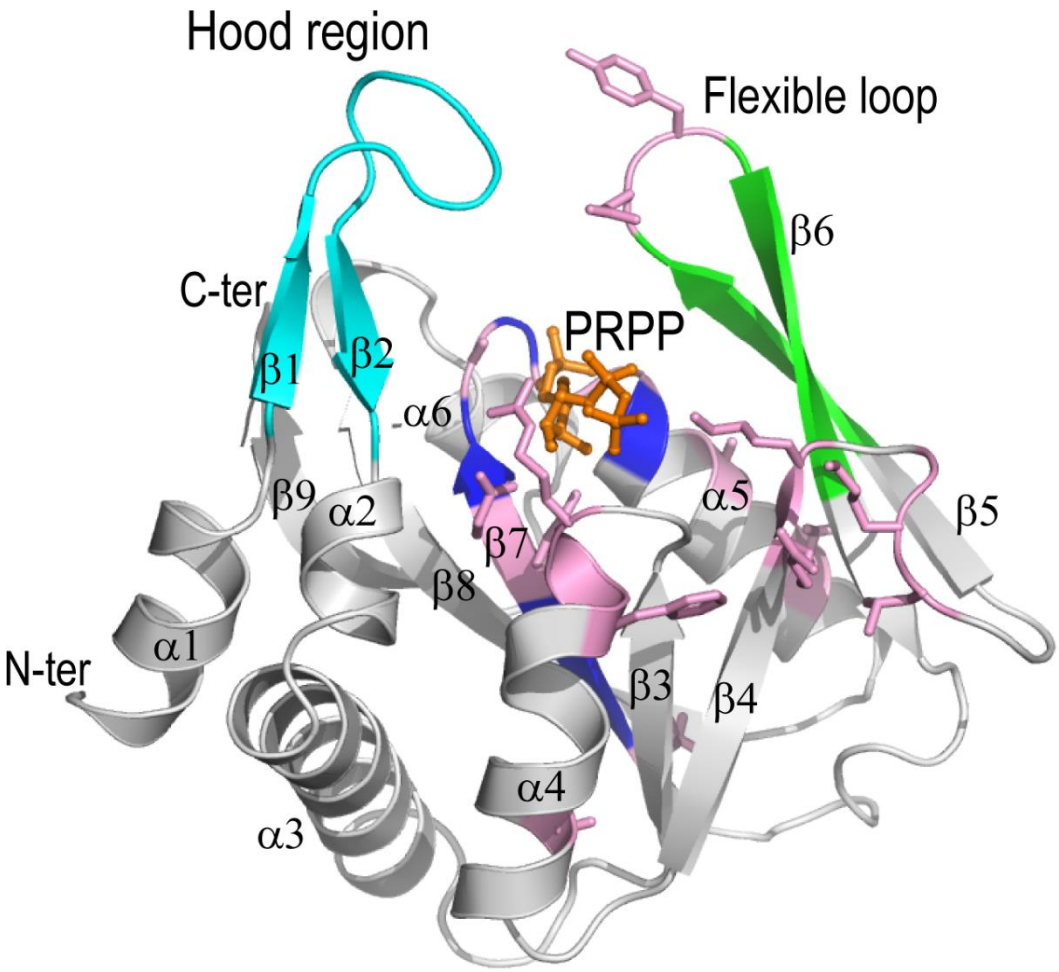
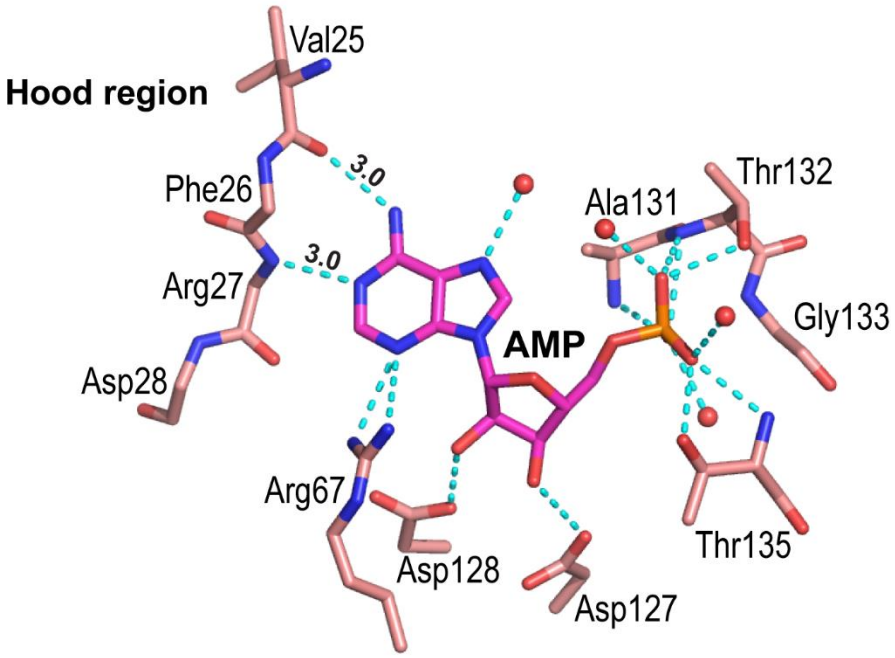

PRPP
binding motif
β6
β7
α5
β8
120
130
140
150
Homo 112 IQKDALEP-------GQRVVVVDDLLATGGTMNAACELLGRLQAEVLE------CVSLVE
Macaca 117 IQKDALEP-------GQRVVVVDDLLATGGTMHAACELLGRLQAEVLE------CVSLVE
Panthera 72 IQRDALEP-------GQKVVVVDDLLATGGTMRAACELLGQLRAEVLE------CVSLVE
Camelus 83 IQRDALEP-------GQKVVVVDDLLATGGTMRAACELLGQLQAEVLE------CVSLVE
Rattus 112 IQKDALEP-------GQKVVIVDDLLATGGTMCAACELLSQLRAEVVE------CVSLVE
Mus 112 IQKDALEP-------GQRVVIVDDLLATGGTMFAACDLLHQLRAEVVE------CVSLVE
Danio 110 MQEDAVSA-------GQKVLIIDDLLATGGTLYAAIELIKQQKAEVLG------CLVVVE
Drosophila 116 LQKSAIKP-------GQKVVVVDDLLATGGSLVAATELIGKVGGVVVE------SLVVME
Caenorhabditis 115 IQEGAIKN-------GDIVFLIDDLLATGGTLRAATDLVVKAGGKVGE------AFVLIE
Saccharomyces 114 IQKNAIPA-------GSNVIIVDDIIATGGSAAAAGELVEQLEANLLE------YNFVME
Arabidopsis 117 MHVGAVEP-------RERVIIIDDLVATGGTLSAAMSLLGFFLSSCFL------FQFLKR
Synechocystis 105 IHQDAVAP-------HHRVLIVDDLIATGGTAKATAELLTKLGCEVLG------FAFIIE
Escherichia 111 IHVDAIKP-------GDKVLVVDDLLATGGTIEATVKLIRRLGGEVAD------AAFIIN
Shewanella 111 IHTDAITA-------NDKVLVVDDLLATGGTIEATVKLIRQLGGEVQD------AAFVIS
Bacillus 103 IHKDAIKP-------GQRVLITDDLLATGGTIEATIKLVEELGGVVAG------IAFLIE
Clostridioides 103 IHKDAIKK-------GQKVAIVDDLLATGGTMEAAAKLVEKLGGEVVS------MQFLIE
Dictyostelium 115 VQEKALSKVVVKPSKKYHVLIMDDILAPGGTMAASIELTKKVLINNGIKDFKISTSLISS
Streptococcus 104 MHKDAIKP-------GDKVLITDDLLATGGTIEATIKLVEALGGEVVG------LAFLIE
Pseudomonas 108 VHADSLCE-------GDSVLIFDDLIATGGTLLAAASLVRRLGARVFE------AAAIID
Neisseria 113 IHTDAVKP-------GSRVLLVDDLVATGGTMLAGLELIRKLGGEIVE------AAAILE
Thermatoga 103 IHEDAIEK-------GQKVLIVDDVLATGGTAEALIRLVKKLGGEVVS------LAFLVE
Parabacteroides 104 IHRDAITP-------DDVVVIHDDLLATGGTMAACYELVKSMNPKKVY------INFIVE
β8
α6
β9
160
170
180
Homo 159 LTSLKGREKLAPVPF----FSLLQYE------
Macaca 164 LTSLKGREKLAPVPF----FSLLQYE------
Panthera 119 LTSLKGREKLGAVPV----FSLLQYD------
Camelus 130 LTSLKGREKLGAVPF----FSLLQYK------
Rattus 159 LTSLKGREKLGPVPF----FSLLQYE------
Mus 159 LTSLKGRERLGPIPF----FSLLQYD------
Danio 157 LKYLNGSDKLKPTPV----FSLIQY-------
Drosophila 163 LVGLEGRKRLDG-KV----HSLIKY-------
Caenorhabditis 162 LAPLNGRSKLPDVNLT----TLISYDSA----
Saccharomyces 161 LDFLKGRSKLNA-PV----FTLLNAQKEALKK
Arabidopsis 164 IFWS----------
Synechocystis 152 LAALNGRQCLPDLPII----SLVEY-------
Escherichia 158 LFDLGGEQRLEKQGIT--SYSLVPFPGH----
Shewanella 158 LPDLGGEARLTALGLE--LVKLCEFEGE----
Bacillus 150 LSYLDGRNKLEDYDIL----TLMKY-------
Clostridioides 150 LKFLNGREKLSNYDVN----SLIKY-------
Dictyostelium 175 IKVLNGKEKIYEKYNDVSVDIIIEM-------
Streptococcus 151 LNALHGRDCLDGYELL----ALMNY-------
Pseudomonas 155 LPELGGSTRLQDAGIS--TFSLTAFALDER--
Neisseria 160 FTDLQGGKNIRASGAP--LFTLLQNEGCMKG-
Thermatoga 150 LSYLEPRKRLEGYDVK----TLIVY-------
Parabacteroides 151 LSDLHGRDNLPKDAE---VTSLIVY-------

Figure S2 : Structure of a PRPP-Mg^{2+}-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.

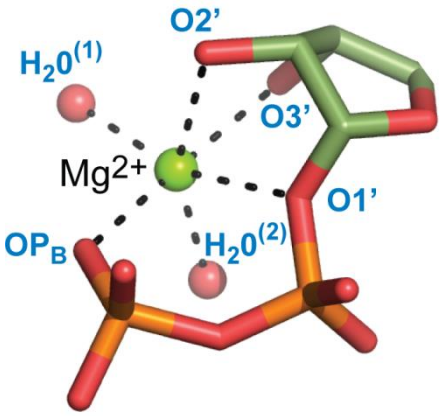
Figure S3 : Interaction of AMP in the hAPRT active site (PDB_ID: 6FCL).



Supporting Note S2: Details on the possible S_N2 pathways. a) A two-step anionic S_N2 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_N2 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) (mol A, with AMP)	ADE-PRPP-Mg^{2+}-hAPRT$^{Y105F-14days}$ (14 days) (mol B, with PRPP)	ADE-PRPP-Mg^{2+}-hAPRT$^{Y105F-30days}$ (30 days) (mol B, with AMP)
PDB_ID	6HGQ	6FCI	6FCH	6FD5	6FD5	6FD6
References	This work	(3)	(3)	(3)	(3)	(3)
Bond distance (Å)						
O1'-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_2O(1)-Mg^{2+}	2.1	2.1	2.3	2.3	2.4	2.2
H_2O(2)-Mg^{2+}	2.1	2.1	2.1	2.3	2.1	2.5
Angle between atoms (°)						
O1'-Mg^{2+}-O2'	75	74	69	79	72	80
O1'-Mg^{2+}-O3'	84	84	78	88	82	86
O2'-Mg^{2+}-O3'	78	75	71	63	70	64
O1'-Mg^{2+}-H_2O(2)	89	87	91	92	85	80
O1'-Mg^{2+}-OP_B	95	94	95	95	95	91
H_2O(1)-Mg^{2+}-O2'	97	106	96	110	93	110
H_2O(1)-Mg^{2+}-O3'	87	92	89	93	88	95
H_2O(1)-Mg^{2+}-H_2O(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$^{Y105F\text{-}14days}$ complex structure 14 days post-crystallization (first subunit) (3).

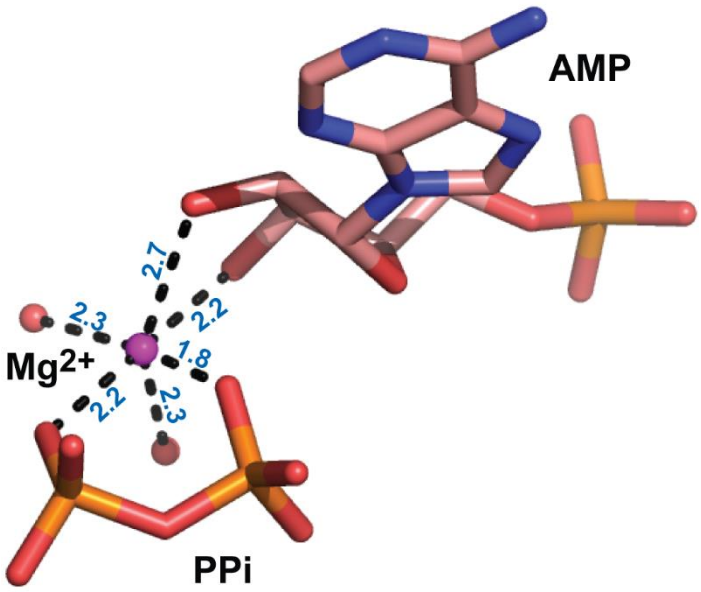


Figure S5 : Superimposition of both molecules of the asymmetric unit in the ADE-PRPP-Mg^{2+}-$hAPRT^{Y105F\text{-}14days}$ variant structure showing the displacement of the magnesium ion. The magnesium ion in the ADE-PRPP-Mg^{2+}-$hAPRT^{Y105F\text{-}30days}$ second subunit is not shown for clarity but superimposes to the ion colored in magenta (3).

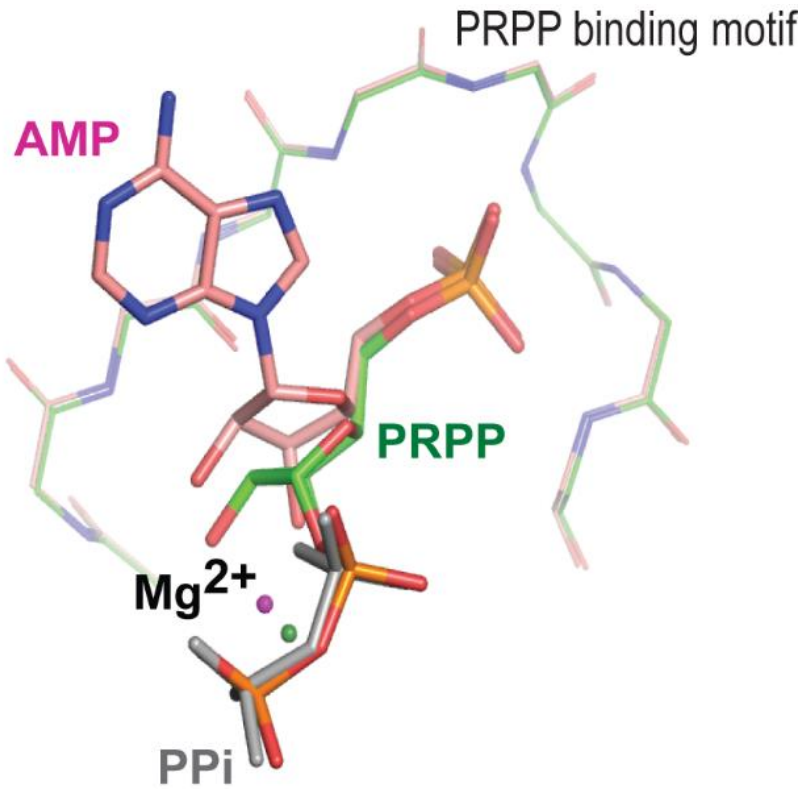
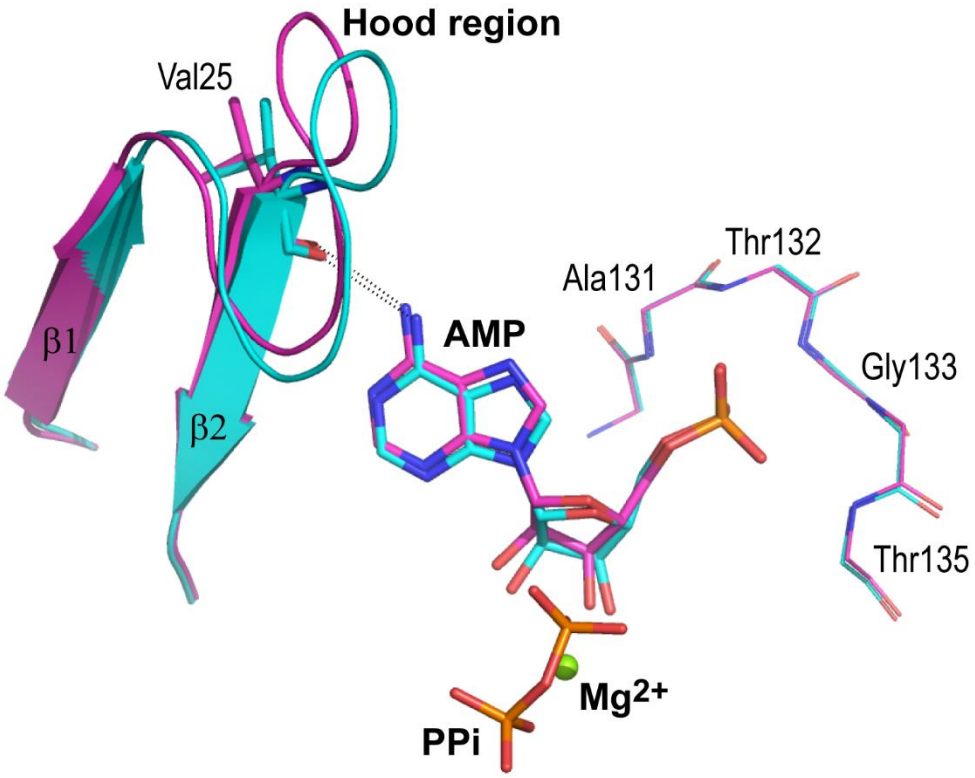


Figure S6 : Superimposition of AMP-hAPRT^{wt} (magenta) onto the first subunit of ADE-PRPP-Mg^{2+}-hAPRT$^{Y105F\text{-}14days}$ structure (cyan), which showed the formation of the products AMP and PPi *in crystallo* (3).



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