Supplementary Information

Structure and the mechanism of eukaryotic FMN adenylyltransferase

Carlos Huerta¹, Dominika Borek¹, Mischa Machius¹, Nick V. Grishin², and Hong Zhang^{1,*}

Supplementary Methods

Supplementary References

Supplementary Figures S1-3

Supplementary Movie

Supplementary Methods

<u>Materials</u>

ATP, AMPCPP (α , β -methyleneadenosine 5'-triphosphate), FMN, FAD and all other reagents were purchased from Sigma-Aldrich, unless stated otherwise. High-throughput (HT) crystal screens, 24-well crystallization plates, Micro-seeding beads, sodium acetate buffer and hydrochloric acid solution for the crystallization experiments were purchased from Hampton Research. EnzChek[®] pyrophosphate assay kit was purchased from Molecular Probes. Genomic DNA from *Candida glabrata* (strain NCYC 388, ATCC 36909D) was purchased from ATCC.

<u>Cloning of CgFMNAT</u>

The following primers were used to amplify *Cg*FMNAT gene from *C. glabrata* genomic DNA: forward primer 5'-GGGCCATGGTGATGCGTTTGGGTGACGCTGC-3', encoding a *NcoI* restriction site (bold) prior to the start codon, and the reverse primer 5'-GGGGTCGACTCATTTCTTTTAATTCTTCCTGCTCTTTC-3', encoding a *SalI* restriction site (bold). Two nucleotides were added after the *NcoI* site in order to have in-frame translation of the gene, resulting in the addition of four residues (Gly-Ala-Met-Val) before the starting Met. The sequence of insert was confirmed by DNA sequencing facility at the McDermott Center at UT Southwestern Medical Center.

Data Collection, SAD phasing, model building and refinement

Two data sets for native apo- C_g FMNAT were collected: one 1.50 Å resolution set was collected at beam-line 19-BM and the 1.20 Å resolution set was collected at beam-line 19-ID. The crystals belong to either P3₁21 or P3₂21 space group for the native and SeMet apo- C_g FMNAT data sets. Aside from the N-terminal methionine, C_g FMNAT has only one non-terminal methionine. Heavy atom search and phasing were performed using SHELXC/D and E

as incorporated in the HKL3000 package.¹ Two Se sites were found and used for calculating phases. Refinement of heavy atom positions and phases was also performed with MLPHARE² followed by density modification using DM³ in the CCP4 package.⁴ The figure-of-merit after MLPHARE was 0.15 and significantly improved to 0.70 after DM for the 2.18 Å SAD data set. During phase determination, the space group was determined to be P3₂21 with one *Cg*FMNAT molecule per asymmetric unit. A model composed of 169 residues out of 308 was built using Resolve⁵. These phases were extended to 1.50 Å of the native data set using DM, which resulted in a much improved electron density map. A model of 272 residues out of 308 was built using ARP/wARP⁶. Refinement against the 1.20 Å native apo-*Cg*FMNAT data set proceeded in two stages. The first stage consisted of restrained refinements of isotropic B-factors with tight restraints. In the second stage, anisotropic B-factors were included in the refinements and stereochemical restraints were gradually loosened with each refinement step. The similar refinement procedure was performed for product ternary complex. Water molecules were added by ARP/wARP or PHENIX⁷ and manually inspected.

Molecular Modeling

All images of structures, ligands, electron density maps, electrostatic potential surface and movie were generated using *i*PyMol (http://www.pymol.org) and *e*Movie⁸. The pdb2pqr server⁹ was used to generate a pqr file using PARSE atomic radii and charges.¹⁰ Structural alignment was performed using Secondary Structure Matching.¹¹

Supplementary references

- 1. Minor, W., Cymborowski, M., Otwinowski, Z. & Chruszcz, M. (2006). HKL-3000: the integration of data reduction and structure solution--from diffraction images to an initial model in minutes. *Acta Crystallogr D Biol Crystallogr* **62**, 859-66.
- Otwinowski, Z. (1991). Maximum likelihood refinement of heavy atom parameters. In Isomorphous Replacement and Anomalous Scattering (Wolf, W., Evans, P. R. & Leslie, A. G. W., eds.), pp. 80-86. Daresbury Laboratory, Warrington, UK.
- 3. Cowtan, K. (1994). 'DM': an automated procedure for phase improvement by density modification. *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography* **31**, 34-38.
- 4. Potterton, E., Briggs, P., Turkenburg, M. & Dodson, E. (2003). A graphical user interface to the CCP4 program suite. *Acta Crystallogr D Biol Crystallogr* **59**, 1131-7.
- 5. Terwilliger, T. C. (2003). SOLVE and RESOLVE: automated structure solution and density modification. *Methods Enzymol* **374**, 22-37.
- 6. Morris, R. J., Perrakis, A. & Lamzin, V. S. (2003). ARP/wARP and automatic interpretation of protein electron density maps. *Methods Enzymol* **374**, 229-44.
- Adams, P. D., Grosse-Kunstleve, R. W., Hung, L. W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K. & Terwilliger, T. C. (2002). PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr D Biol Crystallogr* 58, 1948-54.
- 8. Hodis, E., Schreiber, G., Rother, K. & Sussman, J. L. (2007). eMovie: a storyboard-based tool for making molecular movies. *Trends Biochem Sci* **32**, 199-204.
- 9. Dolinsky, T. J., Nielsen, J. E., McCammon, J. A. & Baker, N. A. (2004). PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res* **32**, W665-7.
- 10. Sitkoff, D., Sharp, K. A. & Honig, B. (1994). Accurate Calculation of Hydration Free Energies Using Macroscopic Solvent Models. *The Journal of Physical Chemistry* **98**, 1978-1988.
- 11. Terwilliger, T. C. (2000). Maximum-likelihood density modification. *Acta Crystallogr*. **D56**, 965-972.
- 12. Pei, J. & Grishin, N. V. (2007). PROMALS: towards accurate multiple sequence alignments of distantly related proteins. *Bioinformatics* **23**, 802-8.

Supplementary Figures



Figure S1. CgFMNAT elution profile from gel filtration column

The single peak maximum occurs at 68.035 ml, corresponding to a molecular weight of ~38 kDa. The theoretical molecular mass for CgFMNAT is 35,568 Da.

		PP-loop						ADE motif			
		40	50	60	70	80	90	100	110	120)
		1	la a			I	1	1		1	
CGFMNAT	37	TTKSILINETF-	-PKWSPLNGE	EISFSYNGGK	DCQVLLLLY	LSCLWEYYIV	KLSQSQFDGKF	HRFPLTKLF	TVFIDHDDT	FKTLENF	120
2goy	41	41 SPQDILK-AAF-EHFGDELWISFSGAED-VVLVDMAWKLNRNVKVFSLDTGRL 31 DAEGRVA-WAL-DNLPGEVVLSSSFGIQAAVSLHLVNQIRPDIPVILTDTGYLF							VF <mark>5L</mark> DTGRL	HPETYRF	97
Isur	31								FPETYRF	88	
1zun	14	ESTHITK-EVA-	-AEFDN	APVMLYSIGK	DSAVMLHLA	KKAFFPG		KLPFF	VMHVDIRWK	FQEMYRE	62
2054	24		-TSR(DST ACRIA						80
100m	216	DAVARTREOVG-		VTI CLSCCV	DSSVTAMLI	HRAT		GKNI T	CVEVDNGL	RINEAFO	272
rabu	LIU	DATATICEQTO						GRATE		NENE/ LEQ	
									ARG1 motif	8	
		13	30			140	150	160	170	1	
			1			1	1	t 💷			
CGFMNAT	121	IEETSLRYSL	SLYESDRD	(CETM/	AEAFETFLQVF	-PETKAIVI	GIRHTDPFG	EHLKPIQ	178
2goy	98	IDQVREHYG1	IAIDVLSPDF	PRLLEPLVKE	KGLFSFYRD	GHGECCGIRK	IEPLKRKLA	GVRAWAT	GQRRDQSPGTR	SQVAVLE	178
lsur	89	IDELTDKLKL	NLKVYRATE	SAAWQEARY	GKLWEQGVE	GIEKYNDINK	VEPMNRALKEL	NAQTWFA	GLRREQSG-SR	ANLPVLA	170
1zun	75	RDQMVEEMGL	DLITHIN			-SAKHTDIMK	TEGLKQALDKH	GFDAAFO	GARRDEEK-SR	AKERVYS	149
2007	04					GIEAQARQAR	YQAFARILL	PGEVLVI	AQHLDDQCEI-	-FLLALK	171
1 apm	273	VI DMEGDHEG-I		DELS		AKEKTTORVE				ESAA	346
rghii	215					ARREITOR	VEVIDEEACKE	LOVINILAC		LJAA	540
							v-Phos	nhate moti	f		
		180			190	200	210	220	230	240	
		1				1		and a state of the	1	1	
CGFMNAT	179	KTDAN		WPDF	YRLQPLLHW	/NLANIWSFLL	YSNEPICE <mark>L</mark> YR	YGFTSLGNV	EETLPNPHLRK	DKNSTPL	244
2goy	179	TDG		-AFSTPEKPL	YKENPLSSM	TSEEVWGYTR	MLELPYNSLHE	RGYISIGCE	PCTRPVLPN		238
	-						Contract of the second states of the second states of the	and and and an and a second se			
1sur	171	IQ		RGV	FKVLPIIDW	DNRTIYQYLQ	KHGLKYHP <mark>L</mark> WD	EGYLSVGDT	Н		215
lsur lzun	171 150	IQ	VQRPELWNVY	NGNVNKGES	FKVLPIIDW	/DNRTIYQYLQI /TELDIWQYIYI	KHGLKYHP <mark>L</mark> WD LEGIPIVP <mark>L</mark> Y-	E <mark>GY</mark> LSVGDT FAA	H		215 211
1sur 1zun 1ni5 2nsv	171 150 130	IQFRDSKHRWDPKN RGSGPAGLSAMA	NQRPELWNVN NEVSEF	/NGNVNKGES	FKVLPIIDW IRVFPLSNW RLIRPLLAR	/DNRTIYQYLQ /TELDIWQYIY	KHGLKYHP <mark>L</mark> WD LEGIPIVP <mark>LY-</mark> QYDLRWIED	EGYLSVGDT	Н		215 211 177 210
lsur lzun lni5 2nsy lanm	171 150 130 172 347	IQ FRDSKHRWDPKM RGSGPAGLSAM	NQRPELWNVY	/NGNVNKGES	FKVLPIIDW IRVFPLSNW RLIRPLLAR ADLLPLTGL	DNRTIYQYLQ TELDIWQYIY TRGELVQWAR TKRQGRTLLK	KHGLKYHP <mark>L</mark> WD LEGIPIVP <mark>LY-</mark> QYDLRWIED ELGAPER-LYL	EGYLSVGDT	H		215 211 177 219 404
lsur lzun lni5 2nsy lgpm	171 150 130 172 347	IQ FRDSKHRWDPKN RGSGPAGLSAMA	NQRPELWNVY	RGV /NGNVNKGES AGT GDGG	FKVLPIIDW IRVFPLSNW RLIRPLLAR ADLLPLTGL GLVEPLKEL	VDNRTIYQYLQ TELDIWQYIYI TRGELVQWAR(.TKRQGRTLLKI FKDEVRKIGLI	KHGLKYHP <mark>L</mark> WD LEGIPIVP <u>LY-</u> QYDLRWIED ELGAPER-LYL ELGLPYDMLY-	E <mark>GY</mark> LSVGDT FAA KEPTADLLC RHPFP	Н РЕКРQQS		215 211 177 219 404
lsur lzun lni5 2nsy lgpm	171 150 130 172 347	IQ FRDSKHRWDPKN RGSGPAGLSAM/	NQRPELWNVN AEVSEF	RGV YNGNVNKGES AGT GDGG. KM	FKVLPIIDW IRVFPLSNW RLIRPLLAR ADLLPLTGL GLVEPLKEL	/DNRTIYQYLQ /TELDIWQYIYI !TRGELVQWAR !TKRQGRTLLKI .FKDEVRKIGLI	KHGLKYHPLWD LEGIPIVPLY- QYDLRWIED ELGAPER-LYL ELGLPYDMLY-	E <mark>GYLSVG</mark> DT FAA KEPTADLLC RHPFP f	H		215 211 177 219 404
lsur lzun lni5 2nsy lgpm	171 150 130 172 347	250	NQRPELWNV AEVSEF 260	RGV YNGNVNKGES AGT GDGG KM 270	FKVLPIIDW IRVFPLSNW RLIRPLLAR ADLLPLTGL GLVEPLKEL 280	VDNRTIYQYLQ /TELDIWQYIYI TRGELVQWARI .TKRQGRTLLKI .FKDEVRKIGLI 290	KHGLKYHPLWD LEGIPIVPLY- QYDLRWIED ELGAPER-LYL ELGLPYDMLY- ARG2 moti 300	E <mark>GYLSVG</mark> PT FAA KEPTADLLC RHPFP f	H		215 211 177 219 404
lsur lzun lni5 2nsy lgpm	171 150 130 172 347	IQ FRDSKHRWDPKN RGSGPAGLSAM/ 	NQRPELWNV AEVSEF 260	RGV YNGNVNKGES GDGG KM 270 	FKVLPIIDW IRVFPLSNW RLIRPLLAR ADLLPLTGL GLVEPLKEL 280	NDNRTIYQYLQ TELDIWQYIYI TRGELVQWAR TKRQGRTLLK FKDEVRKIGLI 290 	KHGLKYHPLWD LEGIPIVPLY- QYDLRWIED ELGAPER-LYL ELGLPYDMLY- ARG2 moti 300	E <mark>GYLSVG</mark> PT FAA KEPTADLLC RHPFP f	H		215 211 177 219 404
1sur 1zun 1ni5 2nsy 1gpm CgFMNAT	171 150 130 172 347 245	IQ FRDSKHRWDPKN RGSGPAGLSAM/ 	VQRPELWNV AEVSEF 260 /KHNEVTKAE	RGV YNGNVNKGES GDGG. KM 270 EPIPIADEDL	FKVLPIIDW IRVFPLSNW RLIRPLLAR ADLLPLTGL GLVEPLKEL 280 VKIENLHED	NDNRTIYQYLQ TELDIWQYIY TRGELVQWAR TKRQGRTLLKI FKDEVRKIGL 290 YYPGWYLVDDI	KHGLKYHPLWD LEGIPIVPLY- QYDLRWIED ELGAPER-LYL ELGLPYDMLY- ARG2 moti 300 L KLERAGRIKKK	EGYLSVGPT FAA KEPTADLLC RHPFP f 304	H		215 211 177 219 404
1sur 1zun 1ni5 2nsy 1gpm CgFMNAT 2goy	171 150 130 172 347 245 239	250 KLNFEWEIENR	NQRPELWNVY AEVSEF 260 /KHNEVTKAE	RGV YNGNVNKGES GDGG. KM 270 EPIPIADEDL	FKVLPIIDW IRVFPLSNW RLIRPLLAR ADLLPLTGL GLVEPLKEL 280 VKIENLHED	NDNRTIYQYLQ ITELDIWQYIY TRGELVQWARI TKRQGRTLLKI FKDEVRKIGLI 290 IYYPGWYLVDDI	KHGLKYHPLWD LEGIPIVPLY- QYDLRWIED ELGAPER-LYL ELGLPYDMLY- ARG2 moti 300 L KLERAGRIKKK QHEREGRWWWE	EGYLSVGDT FAA KEPTADLLC RHPFP f 304 249	H		215 211 177 219 404
1sur 1zun 1ni5 2nsy 1gpm CgFMNAT 2goy 1sur	171 150 130 172 347 245 239	250 KLNFEWEIENR	NQRPELWNVY AEVSEF 260 I (KHNEVTKAE	RGV YNGNVNKGES GDGG. KM 270 EPIPIADEDL	FKVLPIIDW IRVFPLSNW RLIRPLLAR ADLLPLTGL GLVEPLKEL 280 VKIENLHED	NDNRTIYQYLQ TELDIWQYIY TRGELVQWAR(TKRQGRTLLKI FKDEVRKIGLI 290 NYYPGWYLVDDI	KHGLKYHPLWD LEGIPIVPLY- QYDLRWIED ELGAPER-LYL ELGLPYDMLY- ARG2 moti 300 J KLERAGRIKKK QHEREGRWWWE	EGYLSVGDT FAA KEPTADLLC RHPFP f 304 249	H		215 211 177 219 404
lsur lzun lni5 2nsy lgpm CgFMNAT 2goy lsur lzun laif	171 150 130 172 347 245 239	IQ FRDSKHRWDPKN RGSGPAGLSAM/ 250 KLNFEWEIENRY	NQRPELWNVY AEVSEF 260 I YKHNEVTKAE	RGV YNGNVNKGES AGT GDGG. KM 270 PIPIADEDL	FKVLPIIDW IRVFPLSNW RLIRPLLAR ADLLPLTGL GLVEPLKEL 280 VKIENLHEC	VDNRTIYQYLQ VTELDIWQYIY VTRGELVQWAR TKRQGRTLLKI FKDEVRKIGLI 290 I VYYPGWYLVDDI	KHGLKYHPLWD LEGIPIVPLY- QYDLRWIED ELGAPER-LYL ELGLPYDMLY- ARG2 moti 300 J KLERACRIKKK QHEREGRWWWE	EGYLSVGDT FAA KEPTADLLC RHPFP f 304 249	H		215 211 177 219 404
lsur lzun lni5 2nsy lgpm CgFMNAT 2goy lsur lzun lni5 2nsy	171 150 130 172 347 245 239 178	IQ FRDSKHRWDPKN RGSGPAGLSAM/ 250 KLNFEWEIENRY	NQRPELWNVY AEVSEF 260 I YKHNEVTKAE	270 PIPIADEDL	FKVLPIIDW IRVFPLSNW RLIRPLLAR ADLLPLTGL GLVEPLKEL 280 I VKIENLHEC	VDNRTIYQYLQ VTELDIWQYIY TRGELVQWAR TKRQGRTLLKI FKDEVRKIGLI VYPGWYLVDDI ESNQDD	KHGLKYHPLWD LEGIPIVPLY- QYDLRWIED ELGAPER-LYL ELGLPYDMLY- ARG2 moti 300 L KLERAGRIKKK QHEREGRWWWE 	EGYLSVGDT FAA RHPFP f 304 249 190	H		215 211 177 219 404
lsur lzun lni5 2nsy lgpm CgFMNAT 2goy lsur lzun lni5 2nsy lgpm	171 150 130 172 347 245 239 178	250 KLNFEWEIENR	NQRPELWNV AEVSEF 260 YKHNEVTKAE	RGV YNGNVNKGES AGT GDGG KM 270 PIPIADEDL 	FKVLPIIDW IRVFPLSNW RLIRPLLAR ADLLPLTGL GLVEPLKEL VKIENLHEC	VDNRTIYQYLQ VTELDIWQYIY TRGELVQWAR TKRQGRTLLKI FKDEVRKIGLI VYYPGWYLVDDI	KHGLKYHPLWD LEGIPIVPLY- QYDLRWIED ELGAPER-LYL ELGLPYDMLY- ARG2 moti 300 I KLERAGRIKKK QHEREGRWWWE SYDRNFL	EGYLSVGDT FAA RHPFP f 304 249 190	H		215 211 177 219 404

Figure S2. Structure-based multiple sequence alignment of representative "adenine nucleotide α hydrolase-like" superfamily proteins

Each sequence is labeled by the PDB code, except CgFMNAT. The residue numbers of CgFMNAT are marked at the top of the alignment. The sequence numbers at the beginning and end of the alignment for each sequence are provided. Conserved structural motifs are boxed and conserved residues within the motifs highlighted. Glycine and proline residues are in red; conserved glycine and charged/polar residues are highlighted in gray and conserved hydrophobic/aromatic residues in yellow. Members of PAPS reductase-like family are 2goy (APS reductase), 1sur (PAPS reductase) and 1zun (ATP sulfurylase); member of PP-loop ATPase family is 1ni5 (tRNA-Ile-lysidine synthetase); and members of N-type ATP pyrophosphatases are 2nsy (NH³⁺-dependent NAD synthetase) and 1gpm (GMP synthetase).



Figure S3. Multiple sequence alignment of eukaryotic FMNATs

The residue numbers and secondary structure elements of CgFMNAT are marked at the top and bottom of the alignment, respectively. Each sequence is labeled by the gi number, species name abbreviation, and common species name. The first and last residues of each sequence are numbered at the beginning and end, with numbers in parentheses indicating residues not shown and total number of residues in brackets. Structural motifs of CgFMNAT are boxed. Residues involved in substrate/product interaction are indicated by asterisks (*). Glycine and proline residues are in red; conserved glycine, proline and charged/polar residues are highlighted in gray and conserved hydrophobic/aromatic residues in yellow. Abbreviation of species name is as follow: Cg, *Candida glabrata*; Sc, *Saccharomyces cerevisiae*; Af, *Aspergillus fumigates*; Sp, *Schizosaccharomyces pombe*; Hs, *Homo sapiens*; Cf, *Canis familiaris*; Mm, *Mus musculus*; Dr, *Danio rerio*; Ce, *Caenorhabditis elegans*; Dm, *Drosophila melanogaster*; Tb, *Trypanosoma brucei*; Pf, *Plasmodium falciparum* and At, *Arabidopsis thaliana*. Sequence alignment was constructed by PROMALS¹² and manually modified.

Supplementary Movie

CgFMNAT active site in action

Using the four structures of C_g FMNAT, the movie illustrates the intermediate movements of C_g FMNAT and the substrates during catalysis through linear interpolation using *i*Pymol and *e*Movie. The main chain of the protein is displayed as loops and colored brown. Side chain of residues known to interact with substrates and products are shown as sticks and colored by atom type with carbon atoms in teal. The ATP, FMN and FAD molecules are shown as sticks and are colored by atom type with carbon atoms white for ATP and yellow for the flavins. The Mg²⁺ ion is shown as a green sphere. The movie starts with apo- C_g FMNAT transitioning to the ATP complex with ATP fading-in. Next, the ATP complex is shown transitioning to the Mg²⁺ATP complex. This segment illustrates the conformational changes of ATP when Mg²⁺ ion binds and how the protein accommodates Mg²⁺ATP. The FMN binding is accompanied by the subtle conformational changes of the flavin motif. In the last scene, the substrates are shown to be converted to products.