Supplementary Information

Structure and the mechanism of eukaryotic FMN adenylyltransferase

Carlos Huerta[1], Dominika Borek[1], Mischa Machius[1], Nick V. Grishin[2], and Hong Zhang[1,*]

Supplementary Methods

Supplementary References

Supplementary Figures S1-3

Supplementary Movie

Supplementary Methods

Materials

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.

Cloning of CgFMNAT

The following primers were used to amplify *Cg*FMNAT gene from *C. glabrata* genomic DNA: forward primer 5'-GGG**CCATGG**TGATGCGTTTGGGTGACGCTGC-3', encoding a *Nco*I restriction site (bold) prior to the start codon, and the reverse primer 5'-GGG**GTCGAC**TCATTTCTTTTTAATTCTTCCTGCTCTTTC-3', encoding a *Sal*I restriction site (bold). Two nucleotides were added after the *Nco*I 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-*Cg*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 $P3_121$ or $P3_221$ space group for the native and SeMet apo-*Cg*FMNAT data sets. Aside from the N-terminal methionine, *Cg*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 MLPHARE[2] followed by density modification using DM[3] 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 $P3_221$ with one *Cg*FMNAT molecule per asymmetric unit. A model composed of 169 residues out of 308 was built using Resolve[5]. 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/wARP[6]. 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 PHENIX[7] 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*Movie[8]. The pdb2pqr server[9] 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.
2. 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.
7. 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

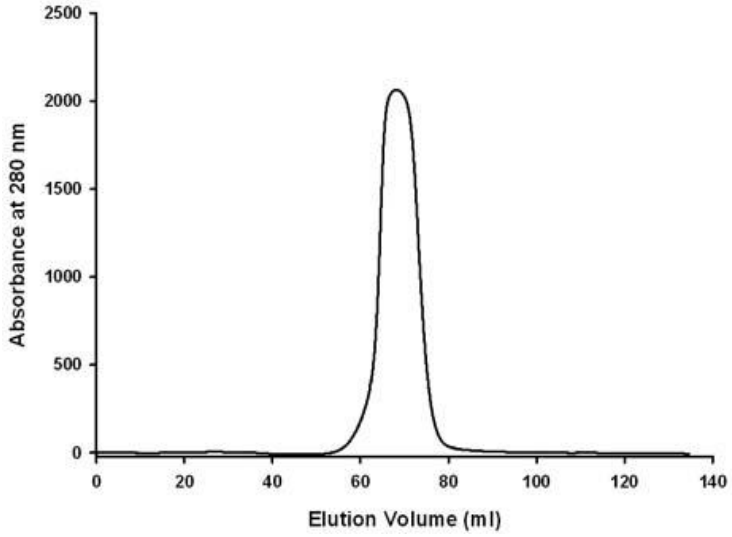

Figure S1. *Cg*FMNAT elution profile from gel filtration column
The single peak maximum occurs at 68.035 ml, corresponding to a molecular weight of ~38 kDa. The theoretical molecular mass for *Cg*FMNAT is 35,568 Da.

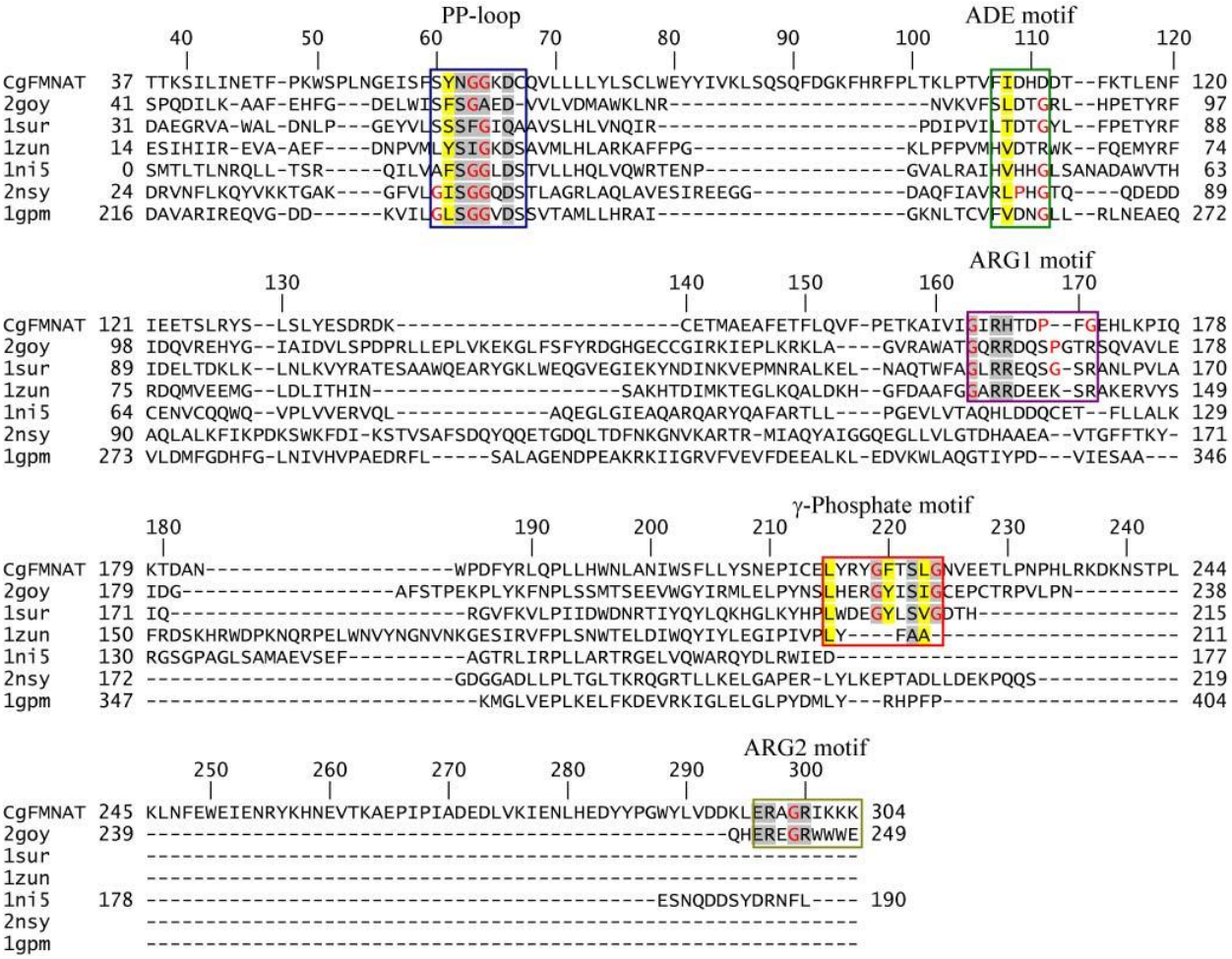

Figure S2. Structure-based multiple sequence alignment of representative "adenine nucleotide α hydrolase-like" superfamily proteins

Each sequence is labeled by the PDB code, except *Cg*FMNAT. The residue numbers of *Cg*FMNAT 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 (NH^{3+}-dependent NAD synthetase) and 1gpm (GMP synthetase).

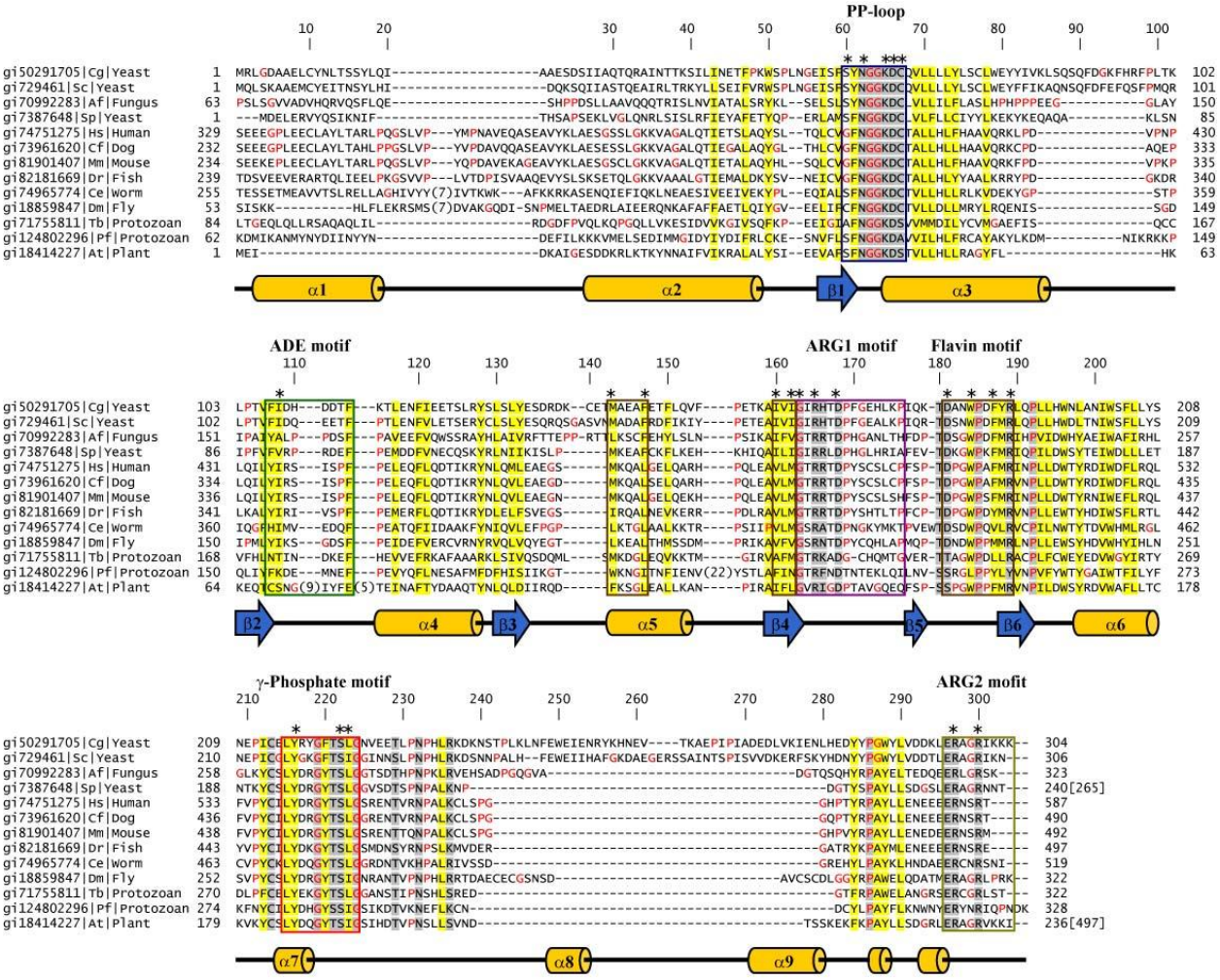

Figure S3. Multiple sequence alignment of eukaryotic FMNATs

The residue numbers and secondary structure elements of *Cg*FMNAT 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 *Cg*FMNAT 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 PROMALS[12] and manually modified.

Supplementary Movie

*Cg*FMNAT active site in action

Using the four structures of *Cg*FMNAT, the movie illustrates the intermediate movements of *Cg*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 Mg^{2+} ion is shown as a green sphere. The movie starts with apo-*Cg*FMNAT transitioning to the ATP complex with ATP fading-in. Next, the ATP complex is shown transitioning to the Mg^{2+}ATP complex. This segment illustrates the conformational changes of ATP when Mg^{2+} ion binds and how the protein accommodates Mg^{2+}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.