Determination of the formylglycinamide ribonucleotide amidotransferase ammonia pathway by combining 3D- RISM theory with experiment

Ajay S. Tanwar¹, Daniel J. Sindhikara², Fumio Hirata^{2*}, Ruchi Anand^{1*}

¹Department of Chemistry, Indian Institute of Technology, IIT-Bombay, Mumbai 400076, India

²College of Life Sciences, Ritsumeikan University and Molecular Design Frontier Co. Ltd. Japan

SUPPORTING INFORMATION

Table of contents

1. Material and Methods	S2
2. Supplementary Figures	S8
3. Supplementary Tables	S16
4. References	S18

1. Materials and methods

Background of 3D RISM. In short the 3D-RISM and 1D-RISM are based on the site-based interpretation of the molecular Ornstein-Zernike equation.¹

$$h(12) = c(12) + \frac{\rho}{\Omega} \int c(12)h(23)d(3)$$

Here h and c are the total and direct correlation functions, respectively. 1 and 2 represent the coordinates of molecules 1 and 2 , and 3 represents a third molecule integrated over all space. ρ represents the solvent density and Ω is the normalization constant for the angular space. Both 1D and 3D-RISM utilize classical atomic interaction potentials and closure approximations such as the KH closure used here.^{2,3} 1D-RISM often utilizes rigid solvent structure, orientational averaging and dielectric constraints for expedited convergence.^{4,5} To model a solvated biomolecular system with 3D-RISM, first 1D-RISM must be performed on the solvent mixture to obtain the solvent correlation functions. These solvent data, in combination, with the solute biomolecular structure and parameters are then input into the 3D-RISM calculation which is then iterated until convergence. The result of this calculation is the 3D equilibrium solvent distributions in the presence of a static solute. While the quick convergence of solvent distributions of 3D-RISM is advantageous over explicit solvent sampling methods,⁶ this advantage is increased orders of magnitude when dilute co-solvents are concerned, such as ammonia in the present study. It is this huge advantage of 3D-RISM/RISM that we seek to exploit with the pathway determination method presented here.

System parameters. Unless otherwise specified, force field parameters were taken from the ff99SB force field parameter set⁷ using the *tLeap* module of *ambertools* 12.⁸

TrpCage. The first structure from PDBID 1L2Y was used for the calculation.⁹ The PDB was typed using tLEaP from AMBERtools. The correlations from the salt/water mixture 1D-

RISM calculation were used for these 3D-RISM calculations (see 1D-RISM calculations below). For the *RismPath* calculation, the “water oxygen” pathway was determined using 10 nodes connecting endpoints located at the center of the cage (2.937,-2.635,-0.544) and a point near the N-terminal asparagine (-8.063, 6.365, -2.044).

CPS. PDBID 1JDB¹⁰ was used for the single point calculation of CPS. Only protein chains B and C were used for the calculation. Protonation states were determined using by the PROPKA¹¹⁻¹⁴ via the PDB2PQR server.^{15,16} The intermolecular correlations from the ammonia/water mixture 1D-RISM calculation were used for the CPS 3D-RISM calculations(see 1D-RISM calculations below). For the *RismPath* calculation, 20 nodes were used. The endpoints were within the small subunit active site (62.587, 70.337, -77.863) and the large subunit active site (59.310, 45.895, -48.929).

PurL. Solute structure was taken from PDBID: 1T3T. Protonation states were first determined using PROPKA¹¹⁻¹⁴ via the PDB2PQR server^{15,16} on the apo structure. Protonable side chains were then checked visually to see whether ligand presence might affect the protonation state. As a result Glu718 and Glu896 near the structural ADP were determined to be unprotonated. There was a missing loop spanning residues 448-466. This loop was modeled using the analogous loop in 2HS4(*ThermotogasmPurL*).¹⁷ All natural residues in the resulting protein were modeled using ff99SB force field parameters. A glutamylthioester was modeled using BCC charges¹⁸ and GAFF parameters¹⁹ leaving the carboxylate unprotonated. The structural ADP charge calculations were calculated using HF/6-31G* QM calculation using Gaussian 09.²⁰ Structural ion parameters for SO₄ and Mg⁺⁺ were taken from the literature.^{21,22} The correlations from the ammonia/water mixture 1D-RISM calculation were used for these 3D-RISM calculations (see 1D-RISM calculations below). For the

RismPath calculation, 15 nodes were used. The endpoints were at near the gate formed by Phe1094 and Phe1165 in the glutaminase domain (102.000, 89.000, -28.500) and the FGAR/ATP binding site (114.510, 76.403, -46.347).

Calculation parameters

1D-RISM calculations. All solvent mixture susceptibilities were solved using the *rism1d* module of ambertools using D-RISM at 298.13K with a grid spacing of 0.025Å and 16384 grid points with the KH-closure. Two solvent mixtures were used for calculations, either a salt/water mixture or ammonia/water mixture. For the salt water mixture the components were 55.5M SPC/e water and 0.01M Na⁺Cl⁻ using Aqvist parameters.²³ For the ammonia and water mixture, ammonia parameters were taken the Rizzo and Jorgensen's work.²⁴ For 1D-RISM calculations, though, all non-bonded terms should be non-zero, so the formula by Luchko et al was used to determine the terms of the “embedded” atoms.²⁵

Equation 1

$$\sigma_e/2 = \sigma_{host}/2 - r_{eh}$$

Equation 2

$$\epsilon_e = \epsilon_h * 0.1$$

Here, *e* and *h* represent the embedded atom (hydrogens), and *h* represents the host atom (nitrogen). *r_{eh}* is the N-H bond distance. The final parameters for ammonia are listed in the following table (Table S1). This model ammonia was mixed at 0.001M with TIP3P water at 55.5M.

3D-RISM calculations. Subsequent 3D-RISM calculations were performed using the *rism3d.snglpnt* module of ambertools using a grid spacing of 0.5 Å and a minimum buffer of 10 Å or greater using the KH-closure.

Molecular dynamics cascade. For the case of PurL where a pathway could be determined using the crystal structure, an MD cascade was performed to introduce small structural fluctuations. All calculations were restrained using 5 kcal/mol/Å² restraints on carbon-alpha and sulphate atoms. All MD calculations were performed in using GB solvent (igb=1). The cascade began with a minimization, using sander for 500 steepest-descent cycles. The minimized structures were then heated from 0 to 300K over 30 ps using *pmemd.cuda*. Subsequently production calculations at 300K were run for 5 ns using *pmemd.cuda*. Snapshots were extracted from the production calculations every 50 ps using *cpptraj*. Finally 3D-RISM and *RismPath* calculations were performed on the extracted snapshots.

Rismpath calculations. *RismPath* calculations were run using $k_{\text{band}} = 80 \text{ kcal/mol/Å}^2$, a maximum of 1000 minimization cycles, a maximum path-length cutoff of 1.5x the linear end-end distance, a maximum gap cutoff of 1.5x the average node-node distance, and 500 total path attempts (trials). The number of nodes was tuned to the system as described above in system parameters. Otherwise, the parameters were first tuned until they both give reasonable results visually (e.g. paths don't pass through solute atoms) and so that they consistently yield identical pass/fail results, and then finally applied uniformly to all systems.

Cloning, expression and purification. Point mutants of the enzyme were made by site directed mutagenesis (SDM) using the strategy employed by the Quikchange (Stratagene) kits. In cases where this strategy failed, overlap and extension PCR followed by cloning using KpnI and BamHI enzymes was employed. Mutation V333I was constructed by SDM method whereas S312F, A384F and A384I were made by overlap PCR method. Phusion high

fidelity DNA polymerase enzyme was used for all the PCR reaction. All the protein were expressed and purified by previously described procedure.²⁶

X-ray crystallography. Diffraction data for V333I was collected on Rigaku FR-E+ Super bright microfocus rotating anode at National Institute of immunology (Delhi, India). The V333I crystals were grown by hanging drop vapour diffusion as described previously.²⁷ Data was collected using anR AXIS IV++ detector, 1.0° oscillation, 3 min exposure time and 240 mm crystal to detector distance at a wavelength of 1.54 Å. V333I crystals were in the space group P6₅ with unit cell dimensions a=b=146.57 Å, c= 141.32 Å. The data was processed using HKL 2000.²⁸ The structure was determined by performing molecular replacement using the native StPurL structure (PDB ID IT3T) and refined using CNS.²⁹ All manual model building was performed using the program Coot.³⁰ The crystallographic data statistics is shown in Table S2.

Circular dichroism (CD) spectroscopy. CD spectra were recorded on a Jasco J-815 CD spectrometer using 0.5 mg mL⁻¹ protein in phosphate buffer (50 mM sodium phosphate, 100 mM NaCl). Scans were performed at 20 °C using 0.1 cm path length quartz cuvettes with 8 sec differential integration time at a scan rate of 50 nm min⁻¹. Thermal denaturation studies were performed by measuring signal at 222 nm while raising the temperature from 20 °C to 95 °C at 1 °C min⁻¹ ramp rate.²⁶

FGAM synthetase assay. The activity was measured using a previously published protocol.^{26,31} In short, a coupled reaction was used where FGAR was converted into FGAM by PurL, which was then converted to AIR (aminoimidazole ribonucleotide) by PurM protein. The reactions were carried out in 150 µL volume with 50 mM phosphate buffer pH 7.5, 20

mM MgCl₂, 80 mM KCl, 50 mM L-glutamine, 10 mM ATP, various concentration (0-3.2mM) of FGAR, 5μl PurL and 5μl PurM. The reaction was incubated at 37°C for 10 min and quenched by the addition of 25μl of 20% trichloroacetic acid (pH 1.4). The concentration of AIR was estimated by using a modified Bratton-Marshall assay.³¹ The intensity of the pink color was recorded at 500 nm on Spectramax M5 instrument.

Glutamate dehydrogenase (GDH) assay. The amount of leaky ammonia and glutamate production were measured using GDH assay. The produced ammonia reacts with α-ketoglutarate and NADPH (reduced form of nicotinamide adenine dinucleotide phosphate) in presence of glutamate dehydrogenase (GDH) enzyme and form glutamate and oxidized NADP⁺. The decrease in absorbance of NADPH due to oxidation is measured at 340nm.^{32,33}

2. Supplementary Figures

Figure S1. *RismPath* algorithm workflow

Figure S2. Demonstration of *RismPath* on TrpCage. TrpCage is shown in cartoon format colored by secondary structure. Endpoints for the calculation are shown as yellow spheres. (a) 3D-RISM distribution of water oxygen ($g(r) > 3.5$) is shown as green glass. (b) Resulting path of *RismPath* calculation using 11 nodes. Nodes are spheres colored from red to blue based on the node index.

Figure S3. *RismPath* demonstration on ammonia pathway of CPS. CPS is shown in cartoon format in white glass. Endpoints are shown in yellow. The resulting *RismPath* path is shown as spheres colored from red to blue. The pathway is somewhat convoluted as it has to bend around secondary structures. (a) Side view. (b) Axial (path axis) view.

Figure S4. Potential “Blocking Mutations” for ammonia pathway in PurL. Wild type paths shown as small green spheres. Mutant residues shown in ball-and-stick format.

Figure S5. Resulting paths from all Path2 blocking mutations. Path 1 and 2 are shown as red and blue tubes respectively. *RismPath* path nodes from the WT and all mutant calculations are shown as green and orange spheres respectively. An apparent 3rd path appears near but 9 Å distant from path 1.

Figure S6. 10% SDS PAGE gel of purified StPurL (wild type) and mutants proteins.

Figure S7. (a) CD spectra and (b) thermal denaturation profile of the wild type StPurL protein and various mutants.

Figure S8. Glutamate production of native and mutants protein.

3. Supplementary Tables

Table S1. Ammonia parameters

	Charge (e)	σ (Å)	ϵ (kcal/mol)
N	-1.02	3.42	0.170
H	0.34	1.4	0.017

Table S2. Crystallographic data statistics

<i>Data collection^a</i>	
Protein	V333I mutant
Space group	P6 ₅
Cell parameters	
a, b, c (Å)	146.57,146.57,141.32
α, β, γ (°)	90,90,120
Resolution (Å)	2.90
Multiplicity	5.2 (5.1)
Completeness (%)	98.1 (99.1)
I/ σ	10.2 (4.5)
R _{sym} (%)	16.7 (40.0)
Total no. of reflections	241004
No. of Unique reflections	46347
<i>Refinement</i>	
Resolution range (Å)	30.0-2.90
No. of reflections total	35235

No. of reflections test set	3537
R _{work} /R _{free} (%)	17.9/23.7
No. of atoms	
Total	10182
Protein	9902
Ligands	30
Ion	30
Water	220
rmsd	
Bond lengths (Å)	0.006
Bond angles (°)	1.31
Ramachandran plot	
Most favoured region (%)	93.4
Additionally allowed region (%)	5.9
Outliers (%)	0.7

---

^avalues for the highest-resolution shell are given in parentheses

4. References

- (1) Hirata, F. *Kluwer Academic Pub.*, **2003**, 24.
- (2) Kovalenko, A.; Hirata, F. *J. Chem. Phys.* **1999**, *110*, 10095.
- (3) Kovalenko, A.; Hirata, F. *J Phys Chem B* **1999**, *103*, 7942.
- (4) Perkyns, J.; Pettitt, B. M. *J. Chem. Phys.* **1992**, *97*, 7656.
- (5) Perkyns, J. S.; Montgomery Pettitt, B. *Chem. Phys. Lett.* **1992**, *190*, 626.
- (6) Stumpe, M. C.; Blinov, N.; Wishart, D.; Kovalenko, A.; Pande, V. S. *J Phys Chem B* **2010**, *115*, 319.
- (7) Hornak, V. e. a. *J Comput Chem* **2006**, *65*, 712.
- (8) Case, D. A. e. a. <http://ambermd.org> **2012**.
- (9) Neidigh, J. W.; Fesinmeyer, R. M.; Andersen, N. H. *Nat Struct Mol Biol* **2002**, *9*, 425.
- (10) Thoden, J. B.; Raushel, F. M.; Benning, M. M.; Rayment, I.; Holden, H. M. *Acta Crystallogr. D Biol. Crystallogr.* **1999**, *55*, 8.
- (11) Søndergaard, C. R.; Olsson, M. H. M.; Rostkowski, M.; Jensen, J. H. *J. Am. Chem. Soc.* **2011**, *7*, 2284.
- (12) Li, H.; Robertson, A. D.; Jensen, J. H. *Proteins: Structure, Function, and Bioinformatics* **2005**, *61*, 704.
- (13) Olsson, M. H. M.; Søndergaard, C. R.; Rostkowski, M.; Jensen, J. H. *J. Am. Chem. Soc.* **2011**, *7*, 525.
- (14) Bas, D. C.; Rogers, D. M.; Jensen, J. H. *Proteins: Structure, Function, and Bioinformatics* **2008**, *73*, 765.
- (15) Dolinsky, T. J.; Nielsen, J. E.; McCammon, J. A.; Baker, N. A. *Nucleic Acids Res* **2004**, *32*, W665.
- (16) Dolinsky, T. J.; Czodrowski, P.; Li, H.; Nielsen, J. E.; Jensen, J. H.; Klebe, G.; Baker, N. A. *Nucleic Acids Res.* **2007**, *35*, W522.

- (17) Morar, M.; Anand, R.; Hoskins, A. A.; Stubbe, J.; Ealick, S. E. *Biochemistry* **2006**, *45*, 14880.
- (18) Jakalian, A.; Jack, D. B.; Bayly, C. I. *J Comput Chem* **2002**, *23*, 1623.
- (19) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J Comput Chem* **2004**, *25*, 1157.
- (20) Frisch, M. J. e. a. *Wallingford, CT 2* **2009**, *4*.
- (21) Cannon, W. R.; Pettitt, B. M.; McCammon, J. A. *J. Phys. Chem.* **1994**, *98*, 6225.
- (22) Allnér, O.; Nilsson, L.; Villa, A. *J. Am. Chem. Soc* **2012**, *8*, 1493.
- (23) Aqvist, J. *J. Phys. Chem.* **1990**, *94*, 8021.
- (24) Rizzo, R. C.; Jorgensen, W. L. *J. Am. Chem. Soc* **1999**, *121*, 4827.
- (25) Luchko, T.; Gusarov, S.; Roe, D. R.; Simmerling, C.; Case, D. A.; Tuszynski, J.; Kovalenko, A. *J Chem Theory Comput* **2010**, *9*, 607.
- (26) Tanwar, A. S.; Goyal, V. D.; Choudhary, D.; Panjekar, S.; Anand, R. *PLoS ONE* **2013**, *11*, e77781.
- (27) Anand, R.; Hoskins, A. A.; Stubbe, J.; Ealick, S. E. *Biochemistry* **2004**, *43*, 10328.
- (28) Otwinowski, Z. O.; Minor, W. *Methods in Enzymology* **1997**, *276*, 307.
- (29) Brunger, A. T.; Adams, P. D.; Clore, G. M.; DeLano, W. L.; Gros, P.; Grosse-Kunstleve; R. W., J.; J. S., K., J., ; Nilges, M.; Pannu, N. S.; Read, R. J.; Rice, L. M.; Simonson, T., and ; Warren, G. L. *Acta Crystallogr D Biol Crystallogr* **1998**, *54*, 904.
- (30) Emsley P, L. B., Scott WG, Cowtan K *Acta Cryst D* **2010**, *66*, 486.
- (31) Anand, R.; Hoskins, A. A.; Bennett, E. M.; Sintchak, M. D.; Stubbe, J.; Ealick, S. E. *Biochemistry* **2004**, *43*, 10343.
- (32) Bera, A. K.; Smith, J. L.; Zalkin, H. *J. Biol. Chem.* **2000**, *275*, 7975.
- (33) List, F.; Vega, M. C.; Razeto, A.; Hager, M. C.; Sterner, R.; Wilmanns, M. *Chem. Biol.* **2012**, *19*, 1589.