Application of Statistical-Thermal Coupling Analysis to Identify Residue-Residue Interaction Networks that Facilitate Coupled-Domain Dynamics in Aminoacyl-tRNA Synthetases

Brent P. Lehman, Stephanie N. Tadayon, Arrianna M. Zirbes, James M. Johnson, Sudeep Bhattacharyya, and Sanchita Hati
Department of Chemistry, University of Wisconsin-Eau Claire, WI 54702

Abstract

Aminoacyl-tRNA synthetases (ARSs) are multi-domain proteins that catalyze covalent attachment of amino acids to the 3′-end of the cognate tRNA molecules. Long-range domain-domain communications are known to play an important role in the function of ARSs. Earlier studies have demonstrated that the coupling of domain motions is important in mediating long-range inter-domain communication in modular proteins. However, the molecular mechanism of coupled-domain motions and long-range communication in ARSs has remained poorly understood. In the present study, the molecular mechanism underlying the coupled-domain motions in ARSs has been probed. Specifically, a bioinformatics-based analysis has been performed to trace "pre-existing" interaction networks that facilitate coupled-domain dynamics. Herein, we report the Statistical Thermal Coupling Analysis (STCA) method, which integrates the information of dynamic coupling of residues with their evolutionary features (conserved and coevolved). The STCA method is applied to three ARS systems: methionyl-, leucyl- and prolyl-tRNA synthetases. The present study demonstrated that the dynamic coupling among distant domains of these ARSs is maintained through networks of evolutionarily constrained residues that are engaged in correlated motion. Moreover, multiple residue-residue interaction networks between distant functional domains are revealed in this study. Existence of these interaction networks is consistent with mutational data and is supported by computational studies.

Aminoacyl-tRNA Synthetases

Aminoacyl-tRNA synthetases are multi-domain proteins, which are responsible for catalyzing the aminoacylation of tRNA in a two-step reaction:

$$AA + ATP + ARS \longrightarrow AA\text{-}AMP.ARS + PP_i \quad (1)$$

$$AA\text{-}AMP.ARS + tRNA \longrightarrow AA\text{-}tRNA + AMP + ARS \quad (2)$$

- ARSs can be considered **allosteric** because tRNA binding at the anticodon binding domain induces conformational change in the remote domains and facilitates catalysis[1]
- **Two proposed models for long-range allosteric communications**[2]
  - **The "induced-fit" model:** substrate-induced conformational change propagated through a single residue-residue interaction pathway.
  - **The "population-shift" model:** a perturbation at a distant site that alters the conformational equilibrium through "pre-existing" multiple pathways of residue-residue interactions.
  - **Pre-Existing pathways:** Communication pathways are inherent to the protein and new pathways cannot be generated. There are multiple allosteric pathways within a protein and information can only be sent through these pathways.[2]

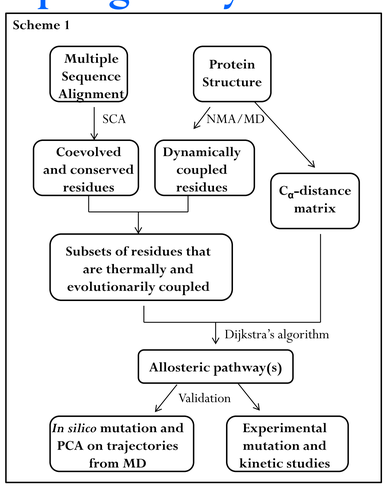
Hypothesis

- Long-range site-to-site communications propagate through networks of conserved and coevolved residues that are thermally coupled.[3]
- The **pre-existing pathways** in a multi-domain protein can be traced by identifying the evolutionarily as well as thermally coupled residues that form contiguous network of interactions.

Methods

Statistical-Thermal-Coupling Analysis

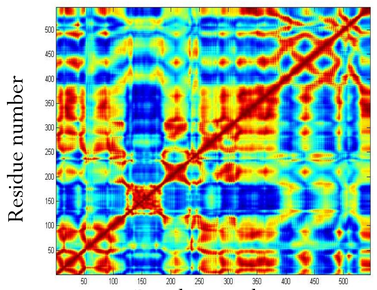
- In this method, the information of protein dynamics (from the Normal Mode Analysis[4] or Molecular Dynamics Simulations[5]) has been computationally integrated to the sequence conservation and coevolutionary data (from the Statistical Coupling Analysis[6]). The shortest path has been determined using the Dijkstra's algorithm.[7]
- Predicted pathways are being verified using *in silico* or experimental mutational studies.



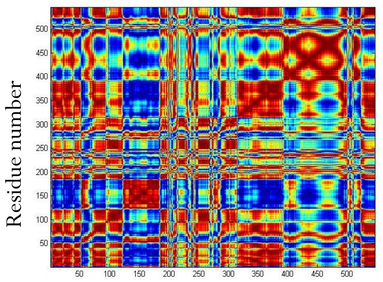
Results

Predicted Pathways - *E. coli* MetRS

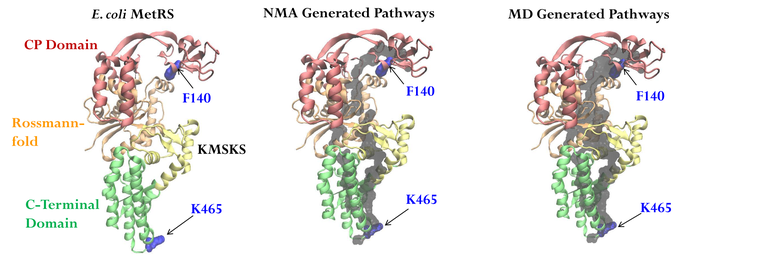
NMA Cross-Correlation Matrix



MD Cross-Correlation Matrix

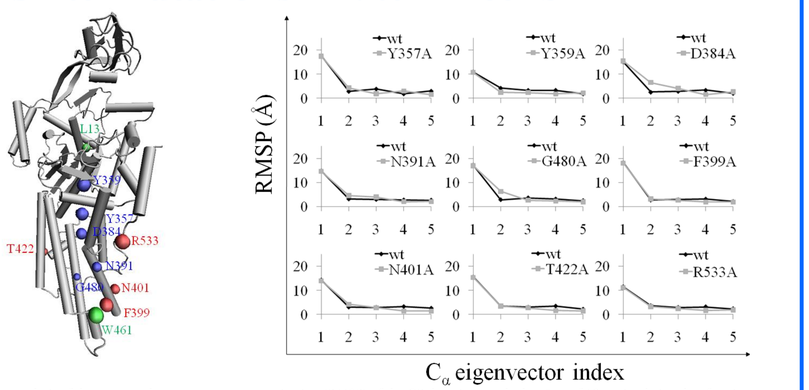


Pathway	Parameters ($C_{ij} > 0.8$)	Residue-Residue Interaction Networks from NMA/MD	% Residue Identity
I	SCANMA: 0.4; Consv: 0.55; Distance ≤ 8.5	**NMA**: A467 → K465 → W461 → P460 → C477 → G480 → F484 → L487 → L491 → P493 → V35 → F47 → C49 → A50 → D52 → H54 →Y237 → P236 → M134 → D129 → S175 → K142	100
		MD: A467 → K465 → W461 → P460 → C477 → G480 → F484 → L487 → L491 → P493 → V35 → F47 → C49 → A50 → D52 → H54 →Y237 → P236 → M134 → D129 → S175 → K142	
II	SCANMA: 0.4; Consv: 0.5; Distance ≤ 8.75	**NMA**: A467 → K465 → P460 → C477 → G480 → F484 → L487 → L491 → P493 →V35 → F47 → C49 → A50 → D52 → H54 →Y237 → M134 → D129 → S175 → K142	91
		MD: A467 → K465 → W461 → P460 → C477 → G480 → F484 → L487 → L491 → P493 →V35 → F47 → C49 → A50 → D52 → H54 →Y237 → P236 → M134 → D129 → S175 → K142	
III	SCANMA: 0.4; Consv: 0.6; Distance ≤ 8.0	**NMA**: A467 → K465 → P460 → R395 → N391 → K388 → D384 → R380 →Y358 →Y359 → L26 → H24 → P14 → D52 → H54 →Y237 → M134 → D129 → S175 → K142	90
		MD: A467 → K465 → P460 →V455 → N452 → K388 → D384 → R380 →Y358 →Y359 → L26 → H24 → P14 → D52 → H54 →Y237 → M134 → D129 → S175 → K142	
IV	SCANMA: 0.4; Consv: 0.55; Distance ≤ 8.75	**NMA**: A467 → K465 → P460 → C477 → R395 → N391 → K388 → D384 → R380 →Y358 →Y359 → L26 → H24 → P14 → D52 → H54 →Y237 → M134 → D129 → S175 → K142	82
		MD: A467 → K465 → P460 →V455 → N452 → K388 → D384 → R380 →Y358 →Y359 → I29 → H28 → C11 → P14 → D52 → H54 →Y237 → M134 → D129 → S175 → K142	
V	SCANMA: 0.4; Consv: 0.5; Distance ≤ 9	**NMA**:A467 → K465 → P460 → R395 → N391 →V386 →V381 →Y358 →Y359 → L26 → H24 → P14 → D52 → H54 →Y237 → M134 → D129 → S175 → K142	77
		MD: A467 → K465 → P460 → R395 → N391 →V386 →V381 →Y358 →Y359 → L26 → H24 → P14 → D52 → G55 →T56 → P137 → F140 → K142	
VI	SCANMA: 0.4; Consv: 0.55; Distance ≤ 8.5	**NMA**: A467 → K465 → W461 → P460 → C477 → R395 → N391 →K388 → D384 → R380 →Y358 → Y359 → L26 → H24 → P14 → D52 → H54 →Y237 → P236 → M134 → D129 → S175 → K142	74
		MD: A467 → K465 → W461 → P460 →V455 → N452 → K388 → D384 → R380 →Y358 → R356 → F87 → F84 → H80 → D51 → A53 → H54 →Y237 → P236 → M134 → D129 → S175 → K142	



Predicted pathways between K465 and F140 of *Ec* MetRS (pdb code: 1QQT) obtained from NMA and MD STCA analysis. Underlined residues in the table above have been shown experimentally to alter *Ec* MetRS function.

In Silico Mutations – *E. coli* MetRS

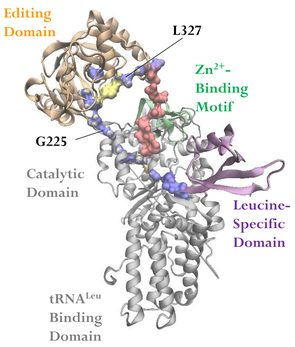


Pair-wise analysis of the computed root-mean-square-projection (RMSP) of the CP domain over the last 5 ns of 10 ns simulation data for eigenvectors 1 to 5. The positive controls are Y357A, Y359A, D384A, N391A, and G480A, whereas negative controls are F399A, N401A, T422A, and R533A.

Predicted Pathways – *E. coli* LeuRS

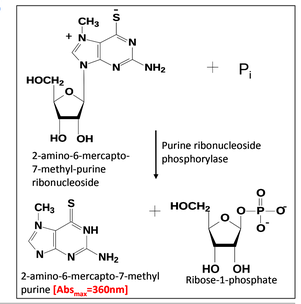
Pathways	Domains	Residues
I	ED to LSD	**E292** → A293 → L327 → V335 → T247 → T248 → S227 → G225 → Q220 → A534 → G530 → M568 → **V569**
II	ED to CD	**E292** → A293 → L327 →Y330 → G331 → T247 → T248 → S227 → G225 → Q220 → A534 → **H537**
III	ED to CD	**E292** → A291 → R185 → N157 → N168 → E169 → M536 → **H537**

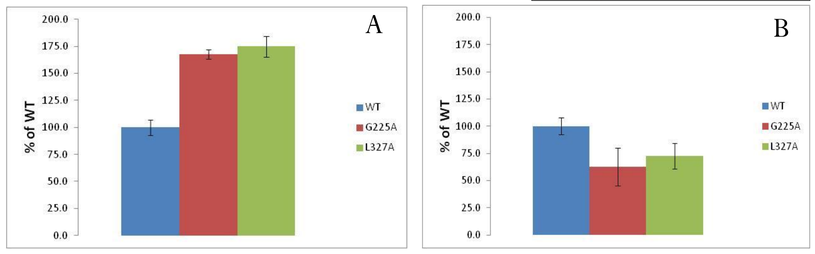
Predicted pathways obtained from STCA-NMA analysis for Ec LeuRS [homology model structure was generated using *T. Thermophilus* LeuRS structure (pdb code: 1H3N)]. The underlined residues represent residues that are known to have significant impact on catalysis upon mutation (ref. 3 and references therein). Pathway I (Blue), pathway II (yellow), and pathways III (red) are shown in the figure in right.



Spectrophotometric Analysis

- **Adenylate formation**
  $E + Leu + ATP \longrightarrow E.Leu\text{-}AMP + PP_i$
- **Aminoacylation Reaction**
  $E + Leu + ATP + tRNA^{Leu} \longrightarrow Leu\text{-}tRNA^{Leu} + AMP + PP_i + E$
- **PP_i + PP_iase → 2P_i**
- **Quantification of P_i using PNP**
  [Lloyd et al, Nucleic Acids Research (1995) 23, 2886]





Comparison of catalytic efficiencies for leucine activation (A) and Leu-tRNA^{Leu} formation (B) between wild-type (WT) LeuRS and the two mutants, G225A and L327A. Error bars are ± standard error. Absorbance values were measured at 360 nm.

Conclusions

- The STCA offers an alternative method for predicting pathways of allosteric communication in multi-domain proteins.
- The pathways identified by the STCA method using NMA and MD simulations bear significant similarities (>70% identify).
- The pathways identified by the STCA method bear significant similarities to previously reported pathways obtained from the MD simulations and protein structure networks analysis.[8]
- Coarse-grained simulations can be used for large proteins to analyze correlated motions.

Future Directions

- Compare and contrast the MD and NMA based pathways using computational and experimental mutagenesis as well as kinetic studies for *E. coli* MetRS.
- To probe potential long-range communication pathways in *E. coli* LeuRS by conducting kinetic studies, specifically the amino acid activation, aminoacylation, and post-transfer editing assays, to determine the catalytic efficiencies of the wild-type and mutant enzymes. Targets for mutation have been identified as A534, N157, and E169.

References

1. Budiman, M. E., Knaggs, M. H., Fetrow, J. S., and Alexander, R. W. (2007) *Proteins* 68, 670-689.
2. del Sol, A., Tsai, C. J., Ma, B., and Nussinov, R. (2009) *Structure* 17, 1042-1050.
3. Weimer, K. M., Shane, B. L., Brunetto, M., Bhattacharyya, S., and Hati, S. (2009) *J. Biol. Chem.* 284, 10088-10099.
4. Eyal, E., Yang, L. W., and Bahar, I. (2006) *Bioinformatics* 22, 2619-2627.
5. Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kale, L., and Schulten, K. (2005) *J. Comput. Chem.* 26, 1781-1802.
6. Lockless, S. W., and Ranganathan, R. (1999) *Science* 286, 295-299.
7. Dijkstra, E. W. (1959) *Numerische Mathematik*;1:269-271.
8. Ghosh, A., and Vishveshwara, S. (2007) *Proc. Natl. Acad. Sci. U. S. A.* 104, 15711-15716.

Acknowledgements:
-National Institute of Health (grant number: GM085779)
-TeraGrid Resources
-Office of Research & Sponsored Programs at UWEC