Comparison of Coarse-Grained and Atomistic-Level Simulations for Aminoacyl-tRNA Synthetases

Alexander M. Strom, Stephanie N. Tadayon, Yer Yang, Sudeep Bhattacharyya, and Sanchita Hati
Department of Chemistry, University of Wisconsin-Eau Claire, WI 54702

Abstract

Aminoacyl-tRNA Synthetases are a group of multi-domain enzymes responsible for catalyzing the covalent attachment of an amino acid to its corresponding tRNA forming an aminoacyl-tRNA. A characteristic of AARSs is the large scale conformational changes they undergo during enzymatic activity. These large scale motions alternate the protein back and forth between its unbound inactive state, and its bound active state.

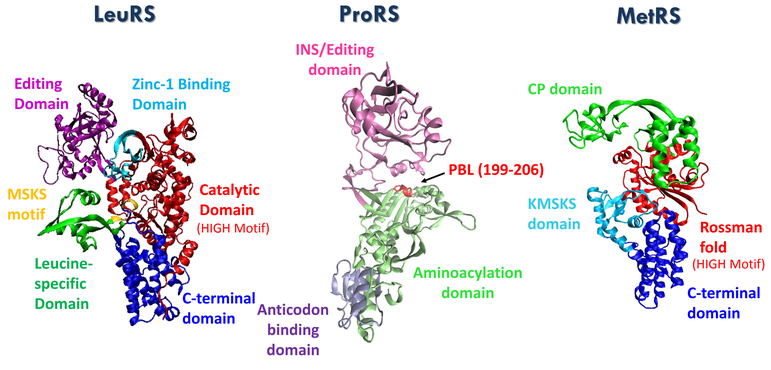
In this study, we observed the differences between three different simulation types. Normal mode analysis (NMA), atomistic-level molecular dynamics (MD) and coarse-grained molecular dynamics simulations (CMD). These three procedures were conducted on *Escherichia coli methionyl-tRNA synthetase (Ec* MetRS), *Enterococcus faecalis* prolyl-tRNA synthetase (Ef ProRS), and *Thermus thermophilus* leucyl-tRNA synthetase (Tt LeuRS). In an atomistic-level MD simulation, each atom's motion is calculated using Newton's second law (F = ma) to create a highly detailed representation of a molecular movements and fluctuations. On the other hand, coarse-grained simulations like CMD and NMA treat a molecule as an elastic mass-spring network of grouped atoms. Coarse-grained simulations have been observed to sacrifice some detail but require dramatically less computational power saving time and financial resources.

By studying these three methods, we hope to determining a more economical yet accurate approach for studying large, multi-domain proteins like aminoacyl-tRNA synthetases using todays computational tools.

Background

Aminoacyl-tRNA Synthetases (AARS)

- Are multi-domain enzymes [*1*]
- Catalyze the esterification reaction to covalently attach an amino acid to its cognate tRNA to form aminoacyl-tRNA
- Contain additional editing domains that catalyze the hydrolysis of non-cognate aminoacyl-tRNA [*2*]
- Coupled-domain motion plays a key role in catalysis [*3,4*]
- In many cases, substrate-induced conformational change is observed that can encompasses large protein domains.



Objectives

Experimentally, it is difficult to study protein motions. In contrast, Newtonian mechanics, when applied under either atomistic or coarse-grain approximations, produces reliable representations of molecular movements. The objectives are as follows:

- ✓ To produce simulated biologically relevant motions of AARSs that typically occur in nano to millisecond time-scales
- ✓ To compare the effectiveness and reliability of normal mode analysis and coarse-grained MD with principle component analysis from a MD simulation trajectories
- ✓ To be able to extend our studies to examine pre-existing pathways of site-to-site communications in aminoacyl-tRNA synthetases

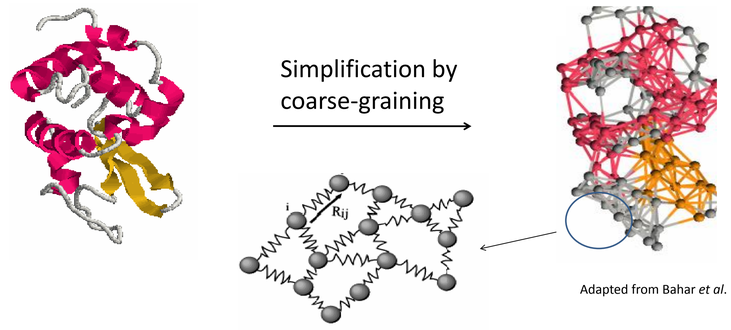
Methods

CMD Simulating a protein molecule and its motion with coarse-grained approximation [*5*].
MD Simulating a protein molecule and its motion with all-atom representation.
NMA Calculation and analysis of normal modes of a protein molecules with coarse-grained approximation [*6*].

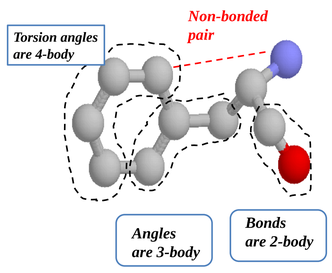
Coarse-Grained Modeling

Summary

- Coarse-graining is a simulation method where residues are treated as a series of beads and springs, usually the C_α atoms.
- Both CMD and NMA use coarse-graining to produce a bead and spring network.



Atomistic Modeling and Molecular Dynamics Simulations



$$U(\vec{R}) = \underbrace{\sum_{bonds} k_i^{bond}(r_i - r_0)^2}_{U_{bond}} + \underbrace{\sum_{angles} k_i^{angle}(\theta_i - \theta_0)^2}_{U_{angle}} + \underbrace{\sum_{dihedrals} k_i^{dihe}[1 + \cos(n_i\phi_i + \delta_i)]}_{U_{dihedral}} + \underbrace{\sum_i \sum_{j \neq i} 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + \sum_i \sum_{j \neq i} \frac{q_i q_j}{\epsilon r_{ij}}}_{U_{nonbond}}$$

- Potential energy, *U*, is a function of the conformation C of the macromolecule. The problem of "minimizing *U*" can be stated as finding C such that *U*(C) is minimum.
- In the present study, 25ns MD simulations were performed.

Principle Component Analysis

- MD and CMD simulations contain the dynamic information of a large number of atoms. To explore the collective motion of a group of atoms, a systematic study of these simulations data is carried out through the use of Principal Component Analysis (PCA) [*7, 8*].
- PCA is a quantitatively rigorous method that simplifies the problem of a multi-dimensional dataset replacing a group of variables with a single new variable, called a principal component.
- Principle components were compared with normal modes seeing as they produce comparable types of results.

Calculation of Harmonic Normal Modes

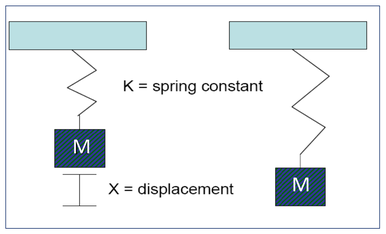
$$V = \frac{1}{2}k\left[\sum_{j,i\neq j}\Gamma_{ij}\left[(r_{ij} - r_{ij}^{o})^2\right]\right]$$

$$F = Ma = M\frac{\partial^2 X}{\partial t^2}$$

Displacement as a function of time:

$$X(t) = A\sin(2\pi\nu t)$$

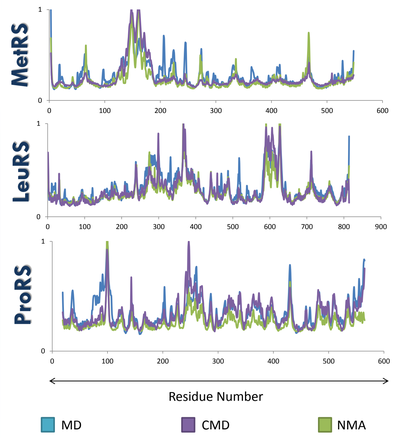
$$-4\pi^2\nu^2 M \cdot X = k \cdot X$$



- A set of algorithms allows us to derive the frequencies and displacements of each residue as the mass-spring network moves back and forth giving a simple yet useful representation of the flexible protein backbone.

Results

Thermal Fluctuations



- A comparison of normalized thermal fluctuations produced by all combined modes of each simulation method.

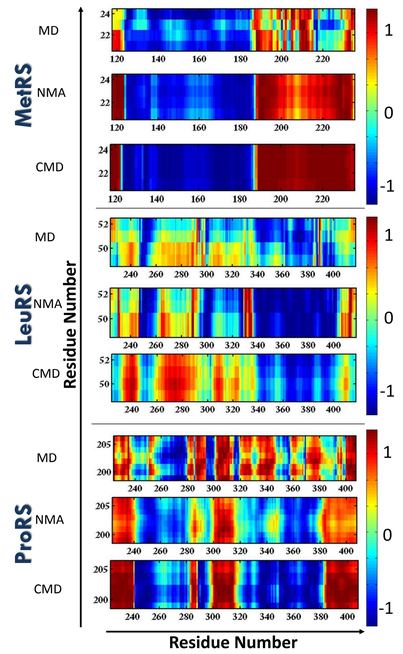
Overlap Calculations

Protein System	Mode	MD vs NMA	MD vs CMD
Ec MetRS	1	0.92	0.87
	2	0.86	0.78
	3	0.76	0.72
Tt LeuRS	1	0.79	0.61
	2	0.88	0.88
	3	0.87	0.88
Ef ProRS	1	0.80	0.80
	2	0.77	0.78
	3	0.73	0.71

- Overlap of initial and final conformation states comparing x, y, z coordinates of the first three modes generated by these simulation methods.

Catalytically Relevant Loop & Domain Analysis

Cross Correlation Maps



- MetRS: CP domain vs. HIGH loop
- LeuRS: CP1 domain vs. HIGH loop
- ProRS: INS domain vs. PBL

Domain & Loop Overlap Values

MD vs NMA

Domain	Overlap Value	Loop	Overlap Value
Ec MetRS		Ec MetRS	
CP (118-235)	0.97	HIGH (21-24)	0.99
		MSKS (332-336)	0.97
Ef ProRS		Ef ProRS	
INS (224-407)	0.79	PBL (199-206)	0.84
Tt LeuRS		Tt LeuRS	
CP1 (224-417)	0.86	HIGH (49-52)	0.96
ZBD (154-189)	0.85	MSKS (638-641)	0.89
LSD (577-634)	0.84		

MD vs CMD

Domain	Overlap Value	Loop	Overlap Value
Ec MetRS		Ec MetRS	
CP (118-235)	0.97	HIGH (21-24)	0.99
		MSKS (332-336)	0.98
Ef ProRS		Ef ProRS	
INS (224-407)	0.82	PBL (199-206)	0.88
Tt LeuRS		Tt LeuRS	
CP1 (224-417)	0.89	HIGH (49-52)	0.98
ZBD (154-189)	0.89	MSKS (638-641)	0.96
LSD (577-634)	0.92		

- A comparison of overlap values for catalytically relevant domains and loops for mode 1.

Conclusions

- ❖ We believe our data, both quantitatively (through overlap calculations) and qualitatively (through cross-correlation matrices and thermal fluctuations) indicates that low-cost normal mode analysis as well as coarse-grained molecular dynamics can serve as a reliable substitute for high-cost MD simulations when simulating the large-scale, low-frequency motions that multi-domain protein like AARSs exhibit.
- ❖ In the future, we plan to expand our study by investigating substrate-bound forms of these enzymes.

References

(1) Crepin et al. (2006) *Structure* 14, 1511-1525.
(2) Beuning et al. (2001) *J. Biol. Chem.* 276, 30779-30785.
(3) Weimer et al. (2009) *J. Biol. Chem. 284,* 10088-99.*
(4) Sanford et al. (2012) Biochemistry 51, 2146-2156.*
(5) Gorecki et al. (2009) *J. Comput. Chem.* 30, 2364-2373.
(6) Eyal et al. (2006*) Bioinformatics*,22, 2619-2627.
(7) Glykos et al. (2006) *J. Comput. Chem.*, 27, 1765-1768.
(8) Mueller et al. (2011) *J. Phys. Chem. B 115*, 3632-3641.*
* *Research conducted by undergraduate students at UWEC*

Acknowledgements:
-National Institute of Health (Grant number GM085779)
-National Science Foundation through TeraGrid Resources
-Office of Research & Sponsored Programs at UWEC
-Learning and Technology Services at UWEC