Understanding Protein-RNA Interactions for Alternative Splicing During Gene Expression Using Molecular Dynamic Simulations

Joseph Daou
Department of Chemistry & Chemical Engineering/B.S. Chemistry
Dequan Xiao Ph.D.

Abstract
The alternative splicing during gene expression is regulated by a type of protein called *splicing factors*. Recently, a splicing factor, *SRSF2* protein was formed to recognize guanines and cytosines at the binding site of the RNA segment. A series of mutations were performed in experiments on SRSF2 to understand the binding interactions of the protein and RNA. However, the changes of binding interactions at the atomistic level due to the mutations have not been revealed. Here, we used molecular dynamics simulations to investigate the conformational changes and protein-RNA binding free energy changes due to a series of mutations. Through the extrapolation of intermediate states of molecular interactions during a fixed length of time throughout point mutations, we used free energy perturbation methods to compute the binding affinity between the mutated proteins and RNA segments. In addition, we also computed the binding free energy by comparing the free energy difference between the folded and unfolded states of the proteins. Our computed free energy changes due to mutations showed consistency with experimental findings. In addition, by molecular dynamic simulations for the solvated protein-RNA complexes, we revealed here the detailed interactions between the protein residues and the RNA for different mutants. Our results provide new insights on understanding the interaction between the splicing factor proteins and the RNA, which can be linked to understanding the pathological mechanisms of Leukemia and cancers.

Introduction

Biophysical researchers have shown, experimentally, that binding affinity deviations between the SRSF2 RRM, or RNA recognition motif protein, involved in alternative splicing, and its' respective RNA chain are directly linked to gene transcription errors in eukaryotic cells and may be the culprit for the onset of leukemia and even AIDS. SRSF2 (SC35) is able to bind degenerated RNA sequences with similar affinity in the nanomolar ranges. However, when a certain translation factor of this process, specifically the 2LEC protein, undergoes point-mutations at various residues, a change in protein-ligand dissociation occurs and thus the protein-RNA complex will either have a dramatic decrease in binding affinity or the protein may unfold. The main residue of interest for the 2LEC protein is the 95th proline amino acid. The point- mutation of this proline to 18 other amino acids, coupled with free energy perturbation to define free energy change, or ΔG, describes the atomistic properties of how binding dynamics affect RRM functions in splicing.

This paper will comment on how 2LEC's 95 residue (proline and other mutants) plays a special role in binding dynamics. In addition, 2LEB, the same protein bound to a different RNA RRM, will undergo similar mutations at 95. Lastly, six alanine mutations of 2LEC in which a *free energy perturbation* technique leads to the modeling of free energy increase due to mutation, and thus a weaker, more loosely-bound protein-RNA complex will be discussed. These new methodologies involve simulative molecular dynamics through computational calculations using biophysical modeling software. The concept utilizes the thermodynamic cycle of intermolecular forces. The advantages of free energy perturbation include precise and infinitely small sampling times, the vast array of intermediate states within amino acid introduction and even the option to solvate the RRM system for an in-vitro model comparable the environment of the protein-RNA complex.

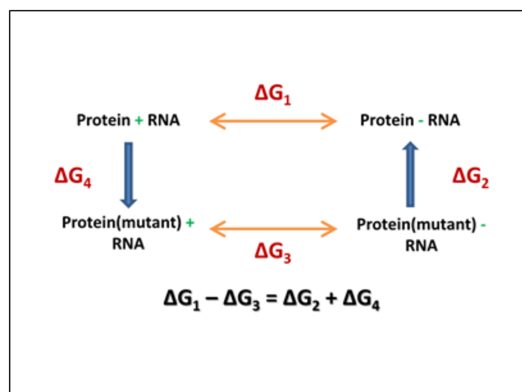

Figure 1: Thermodynamic Cycle of Protein-RNA System

The increase in free energy of a protein-RNA system caused by some mutation should follow the thermodynamic cycle where the mutated system's Gibbs' energy difference with and without RNA subtracted from the wild-type system's Gibbs' energy difference with and without RNA is equal to the combination of the free energy differences' of the protein undergoing mutation both with and without the presence of RNA.

Methods

I) *Equilibration of a Solvated Protein-RNA Complex*

Before explaining exactly how the free energy perturbation of a protein during six alanine mutations provides data that coincides with experimental results, one must first show that the mutation of a specific residue leads to a relaxed conformation of the protein-RNA system in a 10x10 Å "water-box" where the total energy of the system equilibrates over a period of hundreds of thousands

picoseconds (2 nanosecond simulation) with frequent sampling (1.0 femtosecond/timestep for >1,000,000 timesteps) time.

For proper perturbation of interaction energy between various residues and a specific RNA sequence, equilibration of a protein *in-aqua* is used as the preliminary step for calculating free energy changes computationally. The protein-RNA complex was solvated in a periodic 3-D box with a 10 Å distance between the edge of the complex and the boundary of the box in each dimension. Equilibration, or energy minimization, of the whole system was conducted at 310 K and room pressure. Langevin dynamics was included in MD simulations and all investigated molecular systems equilibrated for 2 nanoseconds with a time-step of 1.0 femtosecond. These equilibration steps usually showed dramatic decrease in total energy for each mutant starting from the initial geometry. At the end of equilibration, the energy for the solvated complexes reached a steady fluctuation.

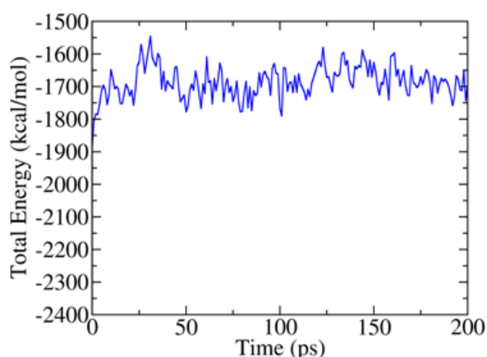

Figure 2: 2LEC Wild-Type Equilibration

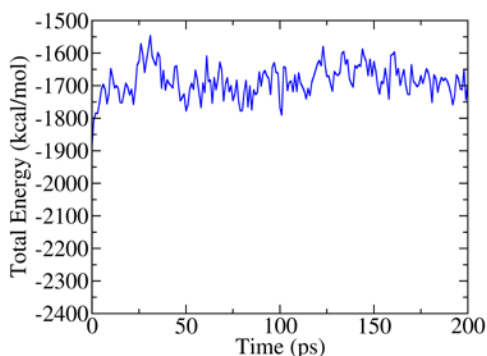

Figure 3: 2LEC P95A Mutant

II) *Free Energy Calculations Using Perturbation Methods*

The free energy perturbation (FEP) method is based on the statistical mechanics of free energy under the alchemical change due to protein mutations. The alchemical change of residues is implemented by a continuous variable λ with the range of λ=0.0 to λ=1.0. Throughout the course of a mutation, intermediate states, where hybrid amino acids are exhibited, are represented by λ=0-1. The first state (λ=0.0), includes the atomic conditions for the wild type protein-RNA complex, while the final state (λ=1.0), includes the atomic conditions for the mutated complex. The FEP calculations by NAMD2 include a short equilibration step, followed by forward steps (λ=0→1) with decoupled intramolecular interactions either on or off, and lastly backward steps (λ=1→0) with the same decoupling options. After each perturbation for several previously-experimented mutations, ΔG was found to change with a similar trend as the experimental values for the same point mutation. The conditions used for the FEP calculation for 2LEC (i.e., SRSF2 + 5'-UGGAGU-3') and 2LEB (i.e., SRSF2 + 5'-UCCAGU-3') mutations included a 3-D water-box for the in-aqua environment at 310 K.

III) *Binding Free Energy Calculation Using FoldX*

After solvation and equilibration of 2LEC and 2LEB variants, an alternative technique was used to calculate the binding free energy before and after mutations of a specific residue. The 95th residue of both 2LEC and 2LEB is proline, which is a special instance of amino acid due to its' amine-containing ring. Normally on these proteins, the proline's ring nitrogen acts as a hydrogen bond acceptor and therefore partakes in hydrogen bonding with binding RNA. If the 95th residue is mutated, hydrogen bonding towards the end of the peptide chain is diminished and other hydrophobic interactions may occur between amino acids and individual bases. We used FoldX to compute the free energy before and after the mutations. Instead of using an alchemical perturbation method, FoldX computed the free energy based on the free energy difference between the folded and unfolded protein states. Following the binding energy calculations, we used YASARA to examine hydrogen bonding, hydrophobic interactions, and π-π bonding interactions between proteins and respective RNA chains.

Results and Discussion

After geometry relaxation using molecular dynamics, we found that the trend of calculated $\Delta\Delta G$ by the FEP method is consistent with the experimental $\Delta\Delta G$ for a series of mutations, e.g., D42A, R61A, R86A and Y44A. From the relaxed conformations, the specific interactions (i.e., hydrogen-binding, π-π interaction, and hydrophobic interaction) were revealed before and after the mutations for the protein-RNA complexes.

Table 1: Binding energy between protein and RNA for a series of mutations at site 95 of SRSF2

		ΔG (kcal/mol)	
Amino Acid Classification	Proline 95 Mutation	2LEC	2LEB
	Wild Type	-5.26	-2.10
Positively Charged Side Chains	Arginine (R)	-3.17	-6.87
	Histidine (H)	-10.72	-7.45
	Lysine (K)	-6.20	-10.78
Negatively Charged Side Chains	Aspartic Acid (D)	-9.68	-7.27
	Glutamic Acid (E)	-7.02	-8.39
Polar, Neutral Side Chains	Serine (S)	-13.43	-8.20
	Thereonine (T)	-9.17	-12.17
	Asparagine (N)	-4.36	-11.36
	Glutamine (Q)	-9.95	-7.05
Special Cases	Cysteine (C)	-9.59	-12.49
	Glycine (G)	-4.69	-8.56
Hydrophobic Side Chains	Alanine (A)	-7.70	-1.97
	Valine (V)	-13.14	-11.44
	Tyrosine (Y)	-13.87	-4.73
	Isoleucine (I)	-8.99	-7.51
	Tryptophan (W)	-10.44	-8.12
	Leucine (L)	-5.04	-16.01
	Methionine (M)	-11.32	-12.37
	Phenylalanine (F)	-4.86	-4.01

Table 2: Comparison for ΔΔG from the experimental measurement and the FEP calculations

	Experimental			FEP
2LEC	K_d (μM)	ΔG (kcal/mol)	ΔΔG (kcal/mol)	ΔΔG (kcal/mol)
Wild Type	0.22	-9.316		
D42A	3.7	-7.704	+1.612	+100
R61A	5	-7.518	+1.788	+70
R86A	0.46	-8.988	+0.328	+48
Y44A	5	-7.518	+1.788	+22

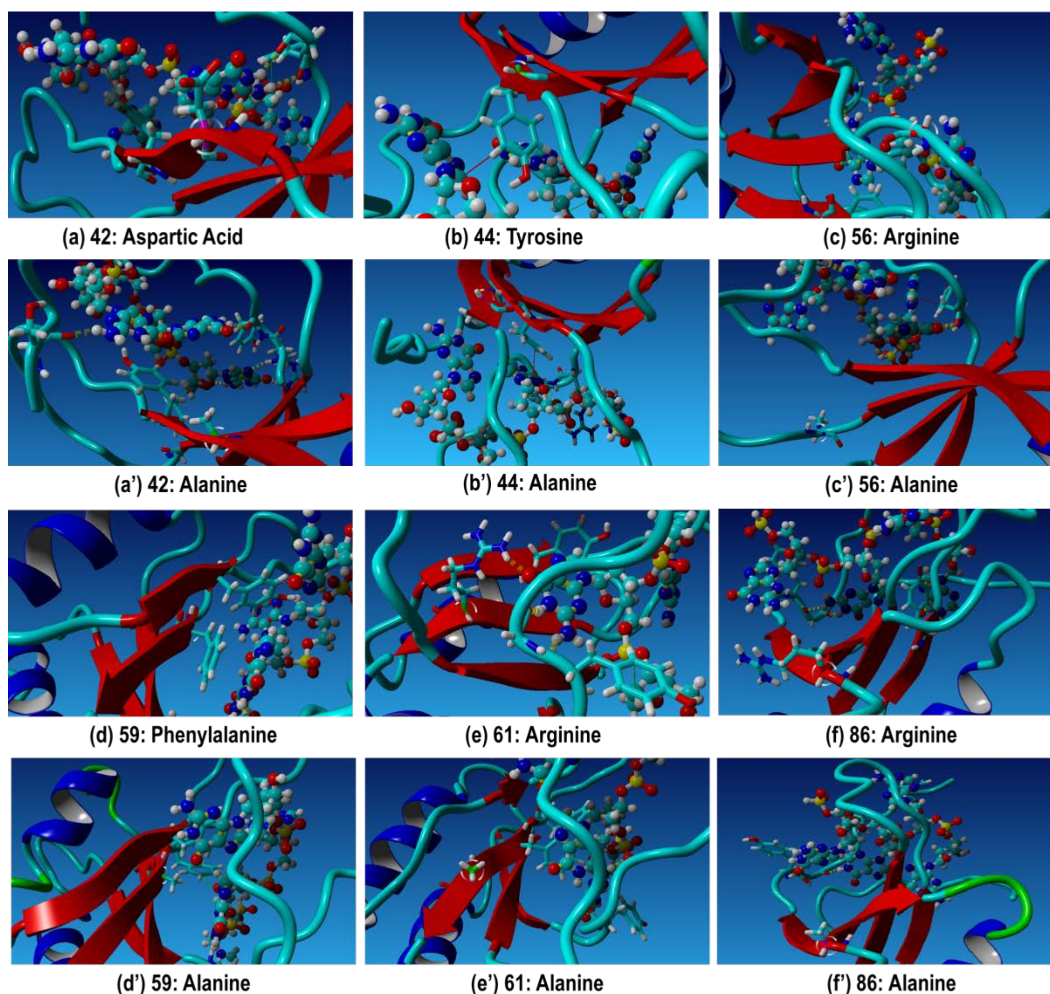

Figure 4: Illustrated interactions between the SRSF2 protein and the RNA sequences at particular mutation sites before (a-f) and after (a'-f') the mutations. The hydrogen bonding interactions are denoted by yellow dashes, the π-π interactions are denoted by red lines, and the hydrophobic interactions are denoted by green lines.

Conclusion

For 2LEC, after mutating residue 95 from proline, the binding affinity increases for nearly all the mutants with the exception of the mutations to arginine, asparagine, glycine, leucine, and phenylalanine. For 2LEB, the binding affinity increases after a mutation to other amino acids with the exception of the mutation to alanine. For 2LEC, the mutations from proline to serine, valine, and tyrosine resulted in the greatest increase in binding affinity out of all possible point-mutations. For 2LEB, the mutations to cysteine, leucine, and methionine resulted in the greatest increase in binding affinity.

This is the first finding for understanding the interactions between SRSF2 and RNA sequences by combining the molecular dynamics simulations and free energy calculations methods. We found that the mutations to SRSF2 at the aforementioned key sites (e.g. 42, 61, 86, 44, and 95) actually induced the change of the prevalent interactions between the protein and RNA, leading to the change of binding free energies. Thus, our findings provide new insights on understanding the pathological mechanisms of leukemia and cancer associated with SRSF2 mutations.

References:

1) Daubner, Gerrit M., Antoine Cléry, Sandrine Jayne, James Stevenin, and Frédéric H-T Allain. "A Syn–anti Conformational Difference Allows SRSF2 to Recognize Guanines and Cytosines Equally Well." *The EMBO Journal* 31.1 (2011): 162-74. Web.

Acknowledgements:

I would like to extend sincere gratitude to my research advisor, Dr. Xiao, of the University of New Haven's Chemistry and Chemical Engineering department, the directors and committee-members of the SURF program at the University of New Haven, and also all members of the Xiao research group, especially Armand Sebastian who contributed to the expansion and illustration of my results. Also, thanks to the Chemistry department for securing and forwarding the grants necessary for the funding of the department's workstation and CPU server, the two machines in which, without their use, none of my energy calculations would have been possible.

Biography:

Joseph Daou is a Class of 2015 Chemistry and Forensic Science Double Major at the University of New Haven. He enthusiastically conducts computational and theoretical chemistry research in the Xiao research group under the advisement of Dr. Dequan Xiao, a physical chemistry professor in the Chemistry department. He is also a quality control/compound handling intern in the Chemistry department of Arvinas Inc., a start-up pharmaceutical company located in New Haven, CT that specializes in protein degradation to inhibit tumor growth. Joseph just recently returned from San Francisco, CA where he held a poster session presentation within the biological division of the 248th American Chemical Society National Meeting and Exposition. He looks forward to continue research projects in the future and has aspirations of becoming a research scientist one day. He hopes to begin graduate research and teacher assistantship within either an Environmental Chemistry or Chemistry Ph.D. program in 2015. He wants to concentrate his graduate work in the fields of analytical, theoretical, and environmental chemistry. For leisure, Joseph enjoys playing on the UNH Men's Volleyball Club team, freshwater fishing, swimming, and cooking.