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

Joseph Daou and Dequan Xiao*

Department of Chemistry and Chemical Engineering, University of New Haven, West Haven, CT 06516

Abstract

The alternative splicing during gene expression is regulated by a type of proteins 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.

Methods

Equilibration of a Solvated Protein-RNA Complex

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.

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.

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.

2LEC Binding Interactions Before & After Mutation

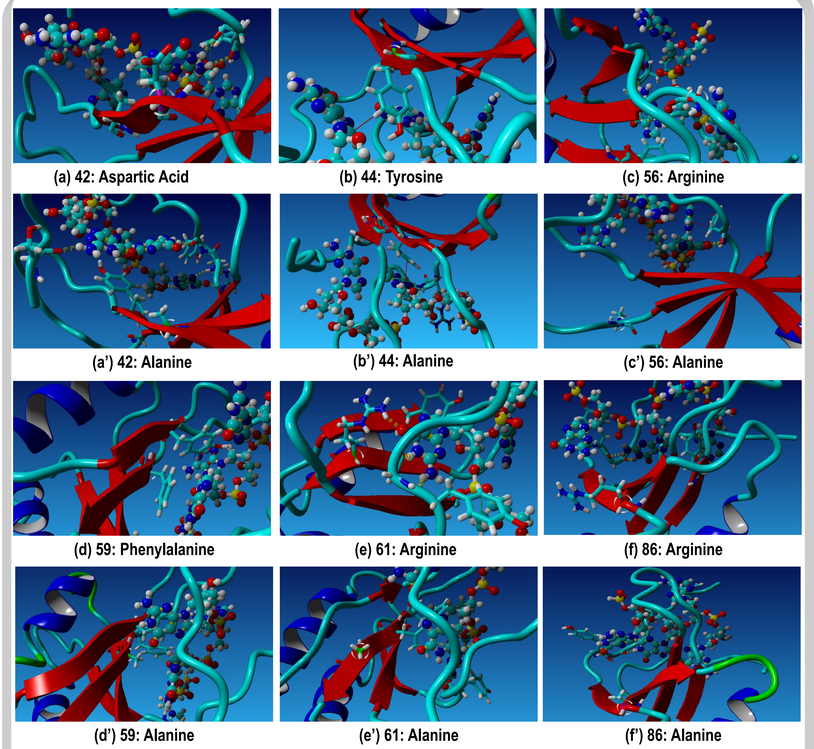

Figure 1. Illustrated interactions between the SRSF2 protein and the RNA sequences at particular mutation sites before (a-f) and after (a'-f') the mutations. The hydrogen bonding interactions are denoted by yellow dashes, the π-π interactions are denoted by red lines, and the hydrophobic interactions are denoted by green lines.

FEP Binding Free Energy Calculations

Table 1. 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

Binding Energy Calculations by FoldX

Table 2. Binding energy between protein and RNA for a series of mutations at site 95 of SRSF2. The mutated protein-RNA complexes were first relaxed in a solvated water box, then the binding energy was computed using FoldX by accounting the conformational relaxation of the protein and RNA backbones. The wild-type residue at site 95 (proline) was mutated to all the other possible residues to analyze the influence of different types of mutations to the binding energy between protein and RNA. The calculations were performed for both the 2LEC and 2LEB complexes.

		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

Conclusions

After geometry relaxation using molecular dynamics, we found that the trend of calculated ΔΔG by the FEP method is consistent with the experimental ΔΔG 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.

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.