

Computational Insights into the Binding Mechanism of OxyS sRNA with Chaperone Protein Hfq

Mengxin Li^{1#}, Yalong Cong^{1#}, Yifei Qi^{1,2} and John Z.H., Zhang^{1,2,3,4*}

¹Shanghai Engineering Research Center of Molecular Therapeutics and New Drug Development, Shanghai Key Laboratory of Green Chemistry & Chemical Process, School of Chemistry and Molecular Engineering, East China Normal University at Shanghai, 200062, China

²Shenzhen Institute of Synthetic Biology, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, People's Republic of China

³NYU-ECNU Center for Computational Chemistry at NYU Shanghai, Shanghai 200062, China

⁴Department of Chemistry, New York University, NY, NY10003, USA

* Correspondence: to: John.zhang@nyu.edu

#These authors contributed equally to this work.

Abstract: Under the oxidative stress condition, the small RNA (sRNA) OxyS that acts as essential post-transcriptional regulators of gene expression is produced and plays a regulatory function with the assistance of the RNA chaperone Hfq protein. Interestingly, experimental studies found that the N48A mutation of Hfq protein could enhance the binding affinity with OxyS while resulting in defection of gene regulation. But, how the Hfq protein interacts with sRNA OxyS and the origin of the stronger affinity of N48A mutation are both unclear. In this paper, molecular dynamics (MD) simulations were performed on the complex structure of Hfq and OxyS to explore their binding mechanism. The molecular mechanics generalized Born surface area (MM/GBSA) and interaction entropy (IE) method were combined to calculate the binding free energy between Hfq and OxyS sRNA, and the computational result is in excellent correlation with the experimental result. Per-residue decomposition of the binding free energy revealed that the enhanced binding ability of the N48A mutation mainly comes from the increased van der Waals interactions (vdW). This research explores the binding mechanism between OxyS and chaperone protein Hfq, and revealed the origin of the strong binding affinity of N48A mutation. The results provided important insights on the mechanism of gene expression regulation affected by protein mutations.

Keywords: small RNA OxyS; RNA chaperone Hfq protein; gene expression regulator; molecular dynamics simulations; binding free energy; interaction entropy

1. Introduction

In bacteria, the small RNA (sRNA) which cannot be encoded into protein owing to the absence of open reading frames plays a critical role in regulation of gene expression ^[1-4]. Under environmental stress such as oxidative stress, low temperature, iron ion concentration and so on, different types of sRNA are produced to regulate the expression of genes by targeting different messenger RNA (mRNA) and affecting their structures as well as translation efficiency ^[5-7]. In this process, the target mRNA can encode proteins in response to external changes and hostile environments. Based on the gene location, sRNA can be divided into two classes ^[8]: cis-encoded sRNA and trans-encoded sRNA. The cis-encoded sRNA is transcribed from the DNA strand complemented with target mRNA, promoting translation ^[9, 10]. The trans-encoded regulatory sRNA is transcribed from other genomic region, regulating the stability of the target mRNA by antisense mediation mechanism ^[11]. The latter referred as trans-encoded sRNA orchestrate dynamics response most strongly depend on the RNA chaperone Hfq that facilitates the formation of base pair between sRNA and target mRNA ^[12-14], and may change the stability and structure of the interacted RNA.

Hfq protein is a ring-shaped RNA-binding protein of the Sm/LSm family that consists of six identical subunits ^[15-18], in which each subunit includes a $\beta 5$ - $\alpha 1$ - β - $\beta 2$ - $\beta 3$ - $\beta 4$ conformation. The N-terminal structural motifs of the six subunits form a ring-shape core region. Structural studies have revealed that Hfq protein has three different binding sites: the proximal face, the distal face, and the lateral rim ^[19, 20]. The proximal binding site, located on one side of the ring, prefers to bind U-rich single-stranded RNA (ssRNA) while the distal binding site prefers A-rich ssRNA. The lateral rim can bind to A-rich and U-rich RNA substrates. The mutation of Hfq gene can reduce the adaption of bacteria in dynamic environment and cause defects in growth rate and tolerance of stress ^[21-25]. The ability of Hfq to influence the living process for bacteria is likely attributed to the role as sRNA chaperone protein. *In vivo*, the mutation of Hfq can affect the accumulation of the interacted sRNA, resulting the change of translation efficiency for mRNA, which in turn affects expression of gene.

Under the oxidative stress condition, the expression of OxyS sRNA is increased. As a regulator of pleiotropy, OxyS can regulate the gene expression by up-regulating and down-regulating a variety of genes with the help of chaperone protein Hfq. A-rich sequence between stem-loops b and c of OxyS sRNA was discovered as the binding region with Hfq protein ^[26]. To understand the mechanism how sRNA regulate mRNA, it is important to understand how Hfq binds to sRNA. However, the interaction mechanism between Hfq and sRNA at atomic levels is unclear. Wartell *et al.* ^[27] studied the interaction of Hfq with sRNA mRNAs, and concluded that the mechanism of the antisense regulation of OxyS for mRNA was not by the binding to rpoS mRNA but by sequestering Hfq. Callaghan *et al.* ^[28] discovered that the Hfq protein didn't change its own conformations obviously in the process of interacting with RNA and resulted in the alternation of sRNA structure, showing the importance of exploring the structure change of sRNA induced by Hfq protein. Barabas *et al.* ^[14] presented two Hfq-A18 complexes, and showed that base stacking can facilitate the interaction of target mRNA with Hfq-bound sRNA with

(ARN)_x motifs, putting forward to the interaction details. Shi et al. found that the N48A mutation of Hfq protein did not weaken the interaction of Hfq and ssRNA while dramatically enhanced the binding ability, causing sRNA–mRNA regulatory defects in experiment [29]. This phenomenon raises the question of whether N48 plays an important role in balancing Hfq–OxyS and the how the binding capacity of Hfq and OxyS influences the regulation function of gene. However, only few researchers explore the detailed interaction of Hfq and OxyS linker region at atomic level to understand the regulatory mechanism, especially with molecular dynamics (MD) simulations.

Compared with experiments, MD [30,31] simulations can provide more detailed dynamic information and have been widely used in the study of structure and function of biomolecules. For biological macromolecule systems, binding free energy is crucial to study binding mode and interaction strength. In free calculations, the accurate calculation of entropic contribution is difficult [32]. Recently, an interaction entropy (IE) method [33–35] was proposed for efficient calculation of entropy in protein-ligand and protein-protein bindings [36–39]. In the present study, the Molecular Mechanics/Generalized Born Surface Area (MM/GBSA) and IE methods were combined to explore detailed binding mode and gene expression regulation mechanism of two A-rich fragments 5'-AACUAAA-3' and 5'-AUAACUA-3' of the OxyS linker region with the Hfq protein (hereafter referred to as Ads system and Aus system). Moreover, in order to obtain more accurate and rigorous assessment of the impact of the N48A mutation, the free energy perturbation (FEP) method [40] was used to calculate the relative binding free energy.

2. Method

2.1 Molecular Dynamics Simulations

The initial crystal structures for the Ads (PDBID: 4QVD) and Aus systems (PDBID: 4QVC) [29] were obtained from the Protein Data Bank (PDB) [41]. The structure of the N48A mutant was obtained by performing mutation in the Leap module of AMBER18. The missing nucleotides were filled by PDBfixer (<https://github.com/openmm/pdbfixer>). The AMBER ff14SB and YIL force field were used to parameterize the protein and RNA [42]. Then the complex was solvated in a TIP3P [43] water box with a buffer distance of 12 Å. The counter ions were added to keep the system neutral. In order to relax the system, two steps of energy minimization were performed. First, the system was minimized by 5000 steps steepest descent followed by 5000 steps conjugate gradient minimization. In this process, solute was restrained with a force constant of 500 kcal/(mol * Å²). Second, the whole system was relaxed without any constraints. Then the whole system was heated from 0 to 300 K in 300 ps with the 5 kcal/(mol * Å²) harmonic constraint on solute. Finally, two-stage MD simulations in NPT ensemble were carried out for 100 ns in total. The first 50 ns simulation was performed for system equilibration and the second 50 ns was for production runs. During the simulation process, 1 kcal/(mol * Å²) harmonic constraints was included on the backbone of RNA. The trajectories were saved every 1 ps. The long-range electrostatic interactions were calculated by the particle mesh Ewald (PME) [44] and SHAKE

[45] algorithm was adopted to deal with all bonds involving hydrogen atom. The cut-off distance of 10 Å was used to deal with the non-bonded interactions. The Langevin thermostat and Berendsen barostat were used to control temperature and pressure.

2.2 MM/GBSA Method

The MM/GBSA approach was used to calculate the enthalpy of the binding free energy. The formula can be expressed by the following terms:

$$\Delta G_{bind} = G_{complex} - G_{protein} - G_{RNA} \quad (1)$$

Where the $G_{complex}$, $G_{protein}$ and G_{RNA} represent the free energy of complex, protein and RNA, respectively. The binding free energy also can be calculated by the following equation:

$$\Delta G_{bind} = \Delta H - T\Delta S \quad (2)$$

In which ΔH and $-T\Delta S$ represent enthalpy change and entropy change, respectively. Among them:

$$\Delta H = \Delta E_{vdw} + \Delta E_{ele} + \Delta G_{gb} + \Delta G_{np} \quad (3)$$

Where ΔE_{vdw} and ΔE_{ele} represent van der Waals interaction and electrostatic interaction between the protein and the RNA. ΔG_{gb} and ΔG_{np} represent polar solvation free energy and non-polar solvation free energy, respectively.

The ΔG_{gb} term belongs to electrostatic solvation free energy and can be calculated by the Generalized Born (GB) model. The non-polar solvation free energy can be obtained by the empirical solvent accessible surface (SASA) formula:

$$\Delta G_{np} = \gamma SASA + \beta \quad (4)$$

The values of γ and β were 0.005 kcal/(mol * Å²) and 0.000 kcal/mol.

2.3 Interaction Entropy

The interaction entropy method [33] was used to calculate the entropy contribution and the formula is given by

$$-T\Delta S = KT \ln \langle e^{\beta \Delta E_{int}} \rangle \quad (5)$$

where $\Delta E_{int} = E_{int} - \langle E_{int} \rangle$ represents the fluctuation of the interaction energy and $E_{int} = E_{vdw} + E_{ele}$. The exponential average can be calculated by simple time averaging (if energy has reached convergence),

$$\langle e^{\beta \Delta E_{int}} \rangle = \frac{1}{N} \sum_{i=1}^N e^{\beta \Delta E_{int}(t_i)} \quad (6)$$

In the process of numerical integration, to eliminate the energy “noises”, energy points that are three standard deviations away from the average were discarded. Finally, the binding free energy was obtained by the summation of enthalpy calculated by MM/GBSA and the entropy calculated by IE.

2.4 Free Energy Perturbation Calculations

To more accurately compute the relative binding free energy between the wild type and the mutant species, the free energy perturbation (FEP) was employed in the calculation. In the FEP method, the relative binding free energy $\Delta\Delta G$ between the mutant and the wild type is given by

$$\Delta\Delta G = \sum_{\lambda} \Delta G_{\lambda} \quad (7)$$

where ΔG_{λ} is computed as

$$\Delta G_{\lambda} = -KT \ln \left\langle e^{-\frac{V_{\lambda} + \Delta\lambda - V_{\lambda}}{KT}} \right\rangle_{\lambda} \quad (8)$$

where K and T are Boltzmann constant and absolute temperature. V_{λ} is defined as $V_{\lambda} = (1 - \lambda)V_{wt} + \lambda V_{mut}$ and λ is the coupling parameter that varies from 0 and 1.

The mutated residue was treated as softcore region and a total of 11 λ -windows were used. In each λ -window, the initial structure was optimized by 25000 steps of steepest descent followed by 25000 steps of conjugate gradient minimization. Then the system was heated from 0 K to 300 K. For each λ -window, a 3-ns MD was performed and 2000 snapshots were used for energy calculation.

3. Results and Discussion

3.1 MD Simulation and Binding Free Energies

To reduce the statistical errors, three independent 100-ns trajectories with random initial velocities were generated for each complex system. The root means square deviation (RMSD) of the backbone atoms in Hfq and Oxys sRNA relative to the initial structure for all systems were calculated to ensure that the systems had reached equilibrium during the simulation (Figure S1-S4). In the 100-ns simulation, the first 50 ns was used for the convergence of the structure, and the last 50 ns was used for analysis. In order to obtain better convergence of the IE method, the RMSD of the gas phase interaction energy (E_{int}) of 10-ns window during the last 50 ns simulation was calculated (Figure S5, S6). Based on the RMSD calculation, the 10-ns window with the smallest fluctuation was used for further energy calculation and analysis. Besides, IE of a 10-ns window with the smallest RMSD for all systems was also depicted in Figure S7 and S8, showing that the entropy changes over time had a good convergence.

The detailed energies of each trajectory were summarized in Table 1. The binding free energy of three simulations were basically consistent within the acceptable error range. The average binding free energy of Ads, Ads (N48A), Aus and Aus (N48A) were -27.7, -29.9, -29.7 and -31.3 kcal/mol, which were in line with the trend of experimental values of -8.0, -9.6, -10.5 and -10.6 kcal/mol. The calculated affinity of Aus was stronger than that of Ads, and the affinities of Ads (N48A) and Aus (N48A) mutants were also stronger than the respective wild types, which were in agreement with experimental trend and demonstrated that the simulation and calculated method were reliable.

Table 1. The experimental and calculated binding free energy from three MD trajectories. (All energy values are given in kcal/mol)

System	Simulation	ΔH	$-T\Delta S$	ΔG_{cat}	ΔG_{exp}
Ads	No. 1	-62.4±0.4	33.8±0.0	-28.6±0.4	
	No. 2	-60.3±0.5	35.1±0.0	-25.2±0.5	
	No. 3	-62.4±0.5	33.3±0.0	-29.2±0.5	
	Ave	-61.7	34.0	-27.7	
Ads (N48A)	No. 1	-65.4±0.4	36.5±0.0	-28.9±0.4	
	No. 2	-66.5±0.4	38.2±0.0	-28.3±0.4	
	No. 3	-64.1±0.3	31.5±0.0	-32.6±0.3	
	Ave	-65.3	35.4	-29.9	
Aus	No. 1	-61.3±0.4	30.7±0.0	-30.6±0.4	-10.5
	No. 2	-64.2±0.4	32.7±0.0	-31.5±0.4	
	No. 3	-67.5±0.4	40.4±0.0	-27.2±0.4	
	Ave	-64.3	34.6	-29.7	
Aus (N48A)	No. 1	-68.7±0.4	37.8±0.0	-30.9±0.4	-10.6
	No. 2	-67.7±0.3	33.3±0.0	-34.4±0.3	
	No. 3	-65.8±0.4	37.1±0.0	-28.7±0.4	
	Ave	-67.4	36.1	-31.3	

3.2 Comparison of Binding Mode of Ads and Aus

Experimental studies found that the binding affinities of Ads and Aus differs by 2.5 kcal/mol despite the high similarity of the RNA sequences. In order to explore the reason of stronger binding ability in the Aus system, residue decomposition was performed to measure the per-residue contribution. The key residues in Ads and Aus that contributed more than -2.5 kcal/mol were plotted in **Figure 1**. The stronger binding affinity in Aus mainly came from residues TYR203 (-10.2 kcal/mol in Aus vs -7.5 kcal/mol in Ads) and LYS209 (-16.3 kcal/mol vs -9.5 kcal/mol). It was worth noting that residues ILE268 and LYS269 had stronger binding affinity in Ads compared with Aus.

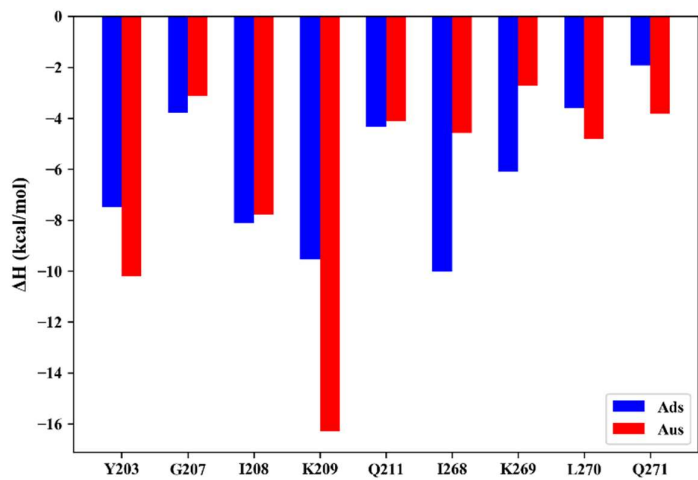


Figure 1. The key residues that contribute more than -2.5 kcal/mol in Ads and Aus.

To further understand the origin of the binding energy difference, the contribution of key residues in Ads and Aus were decomposed into van der Waals (vdW), electrostatic, polar solvation energy, and non-polar solvation energy (**Figure 2**). Compared with Ads, the favorable contribution for TYR203 and LYS209 in Aus mainly came from the electrostatic interaction and polar solvation energy. The main source for the weaker interaction of ILE268 in Aus mainly originated from the vdW. For LYS269, the weaker energy mainly came from vdW and electrostatic interaction.

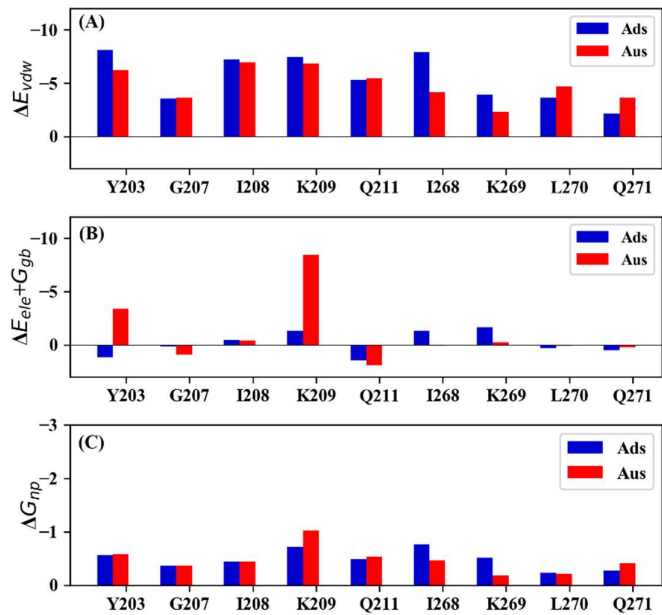


Figure 2. Detailed energy decomposition of Ads and Aus. (A) van der Waals (vdW). (B) sum of electrostatic and polar solvation energy. (C) nonpolar solvation energy.

To further explore the origin of the favorable and weak residues, the hydrogen bond analysis based on MD trajectory was performed (**Table 2**). The favorable electrostatic interaction for TYR203 in Aus mainly arisen from the hydrogen bond between TYR203 and nucleotide adenine, with the occupancy of 95.5%, whereas the same hydrogen bond didn't form in Ads. Similarly, residue LYS269 formed a hydrogen bond in Ads with the occupancy of 62.1%, but didn't form any hydrogen bond in Aus, resulting the stronger electrostatic interaction in Ads. The average distance between the backbone N atom of LYS269 and OP1 atom of uracil 368 in Ads was 11.9 Å longer than that of Aus, which may result in weaker vdW interaction. For residues LYS209 and ILE268, the conformations of these residues were analyzed (**Figure 3**). The charged group of LYS209 formed a salt bridge with phosphate group of RNA in Aus with average distance of 3.7 Å (**Figure 3A, B**). However, in Ads the phosphate group of nucleotide moved away from LYS209, making it impossible to form the same salt bridge (**Figure 4, 5A**). As for ILE268, the vdW hydrophobic interaction in Ads was significantly stronger than that of Aus. According to the structural analysis in Figure 3C and D, the alkyl of ILE268 formed CH- π interaction with the base of adenine, in which the distance between alkyl with the center of mass of nucleobases was 5.0 and 5.6 Å in Ads and Aus, respectively. The shorten distance in Ads was the reason for the higher vdW interaction. Besides, the distribution of the distance in Figure 5B also showed the shorten distance in Ads. Although ILE268 and LYS269 had stronger enthalpy in Ads, but the increased values did not exceed the stronger contribution of residues TYR203 and LYS209 in Aus.

Table 2. Occupancy of hydrogen bonds for Ads and Aus.

	Acceptor	Donor	Occupancy (%)	Distance (Å)	Angle (°)
Ads	A369@OP2 ^a	TYR203@OH	0.0	5.3	111.0
	U368@OP1	LYS269@N	62.1	3.8	154.8
Aus	A367@OP2 ^a	TYR203@OH	95.5	2.7	163.4
	U366@OP1	LYS269@N	0.0	11.9	132.9

^aA367 in Aus corresponds to A369 in Ads

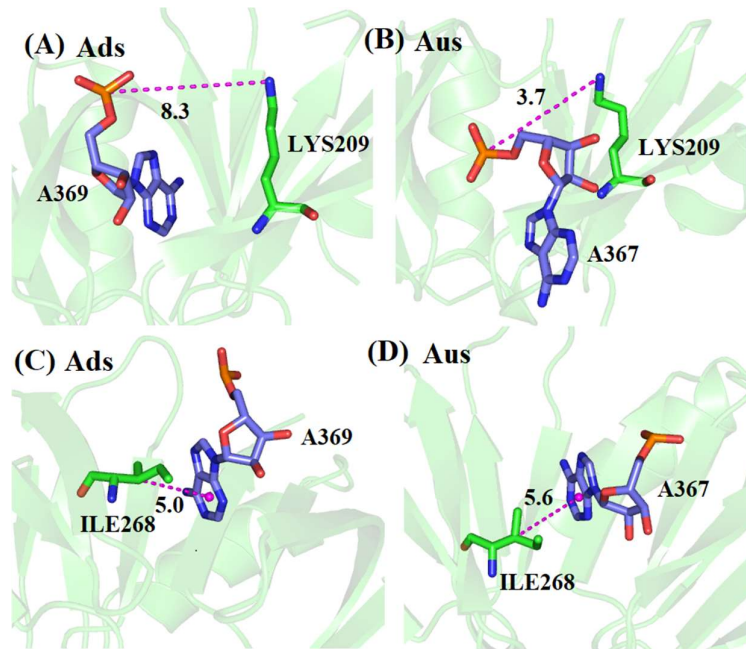


Figure 3. The conformations in the lowest-energy structures of MD trajectory for key residues LYS209 and ILE268. (A) LYS209 in Ads. (B) LYS209 in Aus. (C) ILE268 in Ads. (D) ILE268 in Aus.

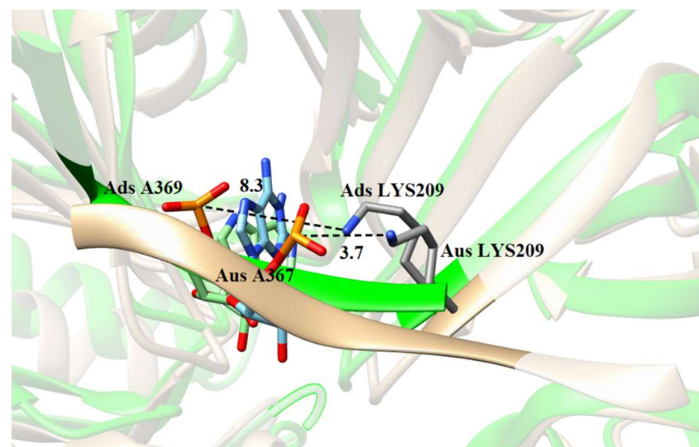


Figure 4. The superposition of Ads and Aus structure in the lowest-energy structures of MD trajectory (Aus in gold, Ads in green). The distance (Å) between the charged group of LYS209 and the phosphate group of adenine in Ads and Aus systems were labeled.

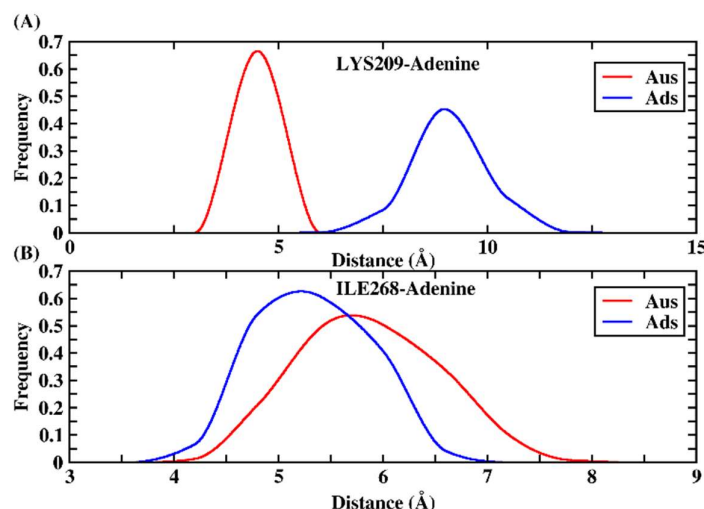


Figure 5. The distribution of distance for key residue LYS209 and ILE268 in Ads and Aus. (A) the distance between the charged group of LYS209 and the phosphate group of adenine. (B) distance between the alkyl of residue ILE268 and the center of mass of nucleobases.

3.3 Mutational Effect of Ads and Aus

Experimental studies showed that the N48A mutation improved the binding affinity for both Ads and Aus [29]. According to the calculated values based the combination of MM/GBSA and IE (**Table 1**), the average binding free energy of Ads (N48A) and Aus (N48A) increased by 2.2 and 1.6 kcal/mol compared with that of the wild type. In order to further verify the reliability of our results, the binding free energy difference between wild and mutated systems were calculated used FEP methods (**Table 3**). The mutation enhanced the binding in Ads and Aus, with the values of 2.9 and 4.1 kcal/mol, which was in line with the experiment trends and basically consistent with the results calculated by MM/GBSA and IE, indicating that our results are reliable. Because the FEP method simulated a series of nonphysical intermediate states and could not provide detail energy information to deeply explore the interaction mechanism, the following analyses of detailed atomic interactions were based on the results calculated by MM/GBSA.

Table 3. The free energy change caused by the N48A mutation calculated with MM/GBSA/IE and FEP. All values are in kcal/mol.

System	$\Delta\Delta G_{GB}$	$\Delta\Delta G_{FEP}$	$\Delta\Delta G_{exp}$
Ads (N48A)	-2.2	-2.9	-1.6
Aus (N48A)	-1.6	-4.1	-0.2

In order to obtain more detailed interaction information and reveal the molecular mechanism of the mutational effect, the residue-based free energy decomposition method was applied to the wild type and mutant systems (**Figure 6**). The stronger binding affinity in Ads (N48A) mainly came from the residues

LYS269, LEU270 and GLN271. For Aus (N48A), TYR203, ILE268 and LEU270 were main beneficial residues for the stronger binding ability. It was worth noting that GLN271 contributed more than 1 kcal/mol in Aus than that in Aus (N48A).

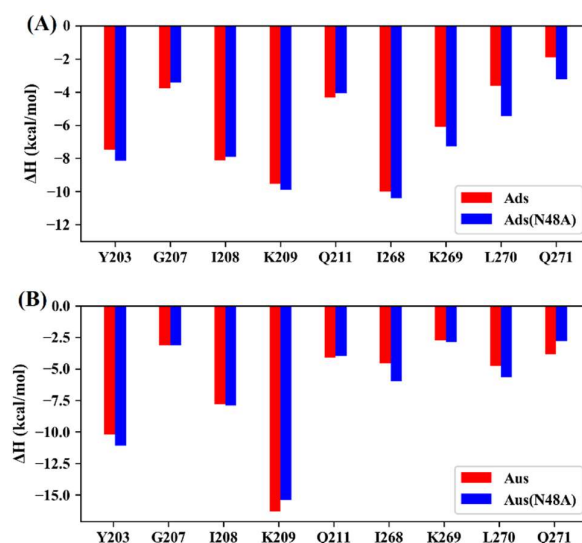


Figure 6. Key residues in the wild type and mutated systems. (A) wild type and mutant for Ads systems. (B) wild type and mutant for Aus systems.

To further explore the mechanism of the affinity change and obtain a more intuitive comparison, the conformations of these residues were plotted in the Figure 7 and Figure 8 based on the lowest-energy structures from MD simulation. The interaction of LYS269 with RNA mainly came from the alkyl of the residue and base of the cytosine, with the contribution of -6.0 and -7.2 kcal/mol in Ads and Ads (N48A) (Figure 6A and Figure 7A). From the energy decomposition analyses, the increased energy mainly depended on the vdW interaction. The average distance of alkyl with nucleotide was 7.0 Å in Ads while the value was 6.0 Å in Ads (N48A). Similarly, the interaction of residue LEU270 with RNA mainly came from the interaction of alkyl and the base of adenine, which provided average contribution of -3.6 and -5.4 kcal/mol in Ads and Ads (N48A). The average distance of alkyl and the center of mass of the base was 5.7 Å in Ads and 4.7 Å in Ads (N48A), causing the increasing of the vdW interaction. The interaction of residue GLN271 with RNA was mostly contributed by CH- π interaction. After the mutation, the enthalpy changed from -1.9 to -3.2 kcal/mol. The increased energy was mainly attributed to the vdW interaction in which the distance of alkyl with bases of uracil was shortened from 6.9 to 5.9 Å. Moreover, the peaks of distance distribution for LYS269, LEU270 and GLN271 with the interacted nucleotide were all smaller in Ads (N48A) compared with Ads (Figure 7C-E).

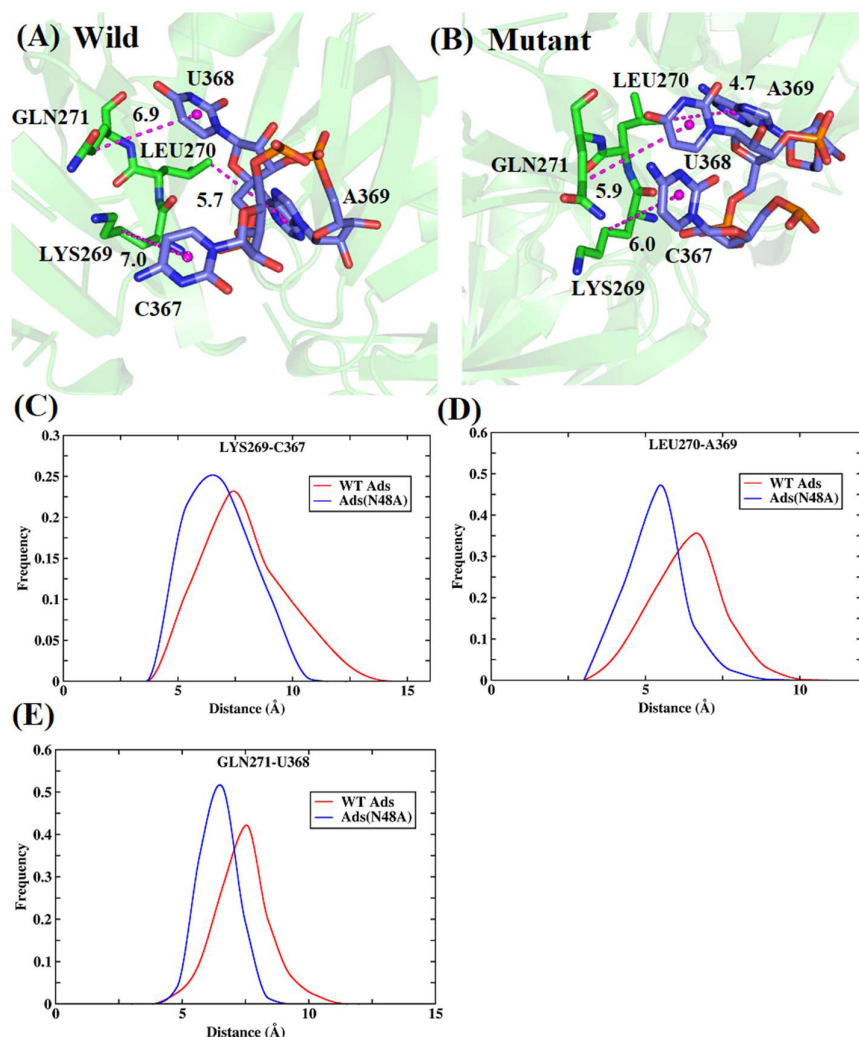


Figure 7. The conformations of the lowest-energy structures and distance distribution of key residues in the Ads systems. (A) Ads. (B) Ads (N48A). (C-E) distance distribution between alkyl of LYS269 and base of cytosine (C), alkyl of LEU270 and base of adenine (D), and alkyl of GLN271 and base of uracil (E).

The binding modes between key residues and RNA in Aus were similar to those in Ads. The benzene ring of TYR203 formed π - π hydrophobic interaction with the base of adenine (**Figure 8A** and **8B**). After the N48A mutation, as the distance of the benzene ring with the base was shortened from 4.7 to 4.0 Å, the enthalpy increases from -10.1 to -11.1 kcal/mol. Besides, the alkyl of ILE268 and LEU270 formed CH- π hydrophobic interaction with the nitrogenous base of RNA. The contribution of ILE268 and LEU270 in Aus was -4.5 and -4.7 kcal/mol while the contribution in Aus (N48A) changed to -5.9 and -5.6 kcal/mol, with the increased energy -1.4 and -0.9 kcal/mol. The average distance between center of mass of nitrogenous base and alkyl of ILE268 was 5.6 and 4.9 Å in Aus and Aus (N48A). For LEU270, the distance was 4.4 and 4.1 Å, respectively. The shorten distance in Aus (N48A) led to

stronger vdW interaction. There was a noticeable residue GLN271 with lower energy in mutant compared with wild type. Because of the conformational change caused by the mutation, the distance of alkyl group with nitrogenous base was increased from 4.7 to 5.6 Å, causing weaker vdW interactions. Besides, from the distance distribution in Figure 8 C-F, the distance peaks for above mentioned stronger residues with their own interacted adenine in mutant were all smaller than wild type. The above analyses suggested that the N48A mutation can affect the conformations of the surrounding residues which led to stronger interaction with sRNA.

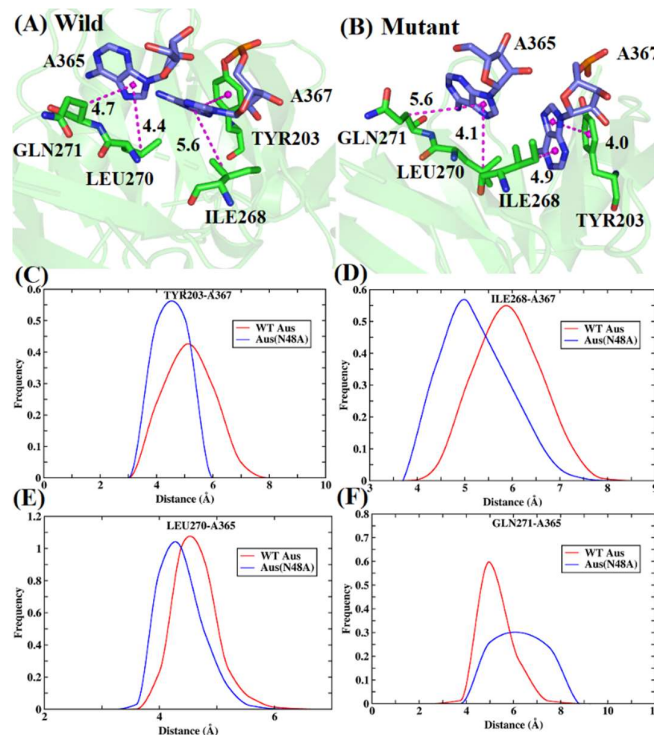


Figure 8. The conformations of the lowest-energy structures and distance distribution of key residues in the Aus systems. (A) Aus. (B) Aus (N48A). (C-F) distance distribution between benzene ring of TYR203 and base of adenine (C), alkyl of ILE268 and base of adenine (D), alkyl of LEU270 and base of adenine (E), and alkyl of GLN271 and base of adenine (F).

4. Conclusion

In this paper, MD simulation and binding free energy calculation were performed to explore the binding mechanism of sRNA (Oxys) and chaperone protein (Hfq). The calculated binding free energy from MM/GBSA and IE was consistent with experimental results. The decomposition of per-residue energy showed that TYR203 and LYS209 played key roles in the favorable interaction of Aus compared with Ads, in which the electrostatic interaction was the major driving force. Moreover, the stronger binding affinity of the N48A mutants relative to the wild types in both Ads and Aus mainly came from

vdW hydrophobic interactions. Structural analyses indicated that the mutation changes the conformation of surrounding residues.

This study provided atomic-level interaction mechanism of sRNA and Hfq based on dynamics information and it deepened our understanding of the binding affinity change caused by the mutation. It is expected that the results could provide useful information for modifying the protein by analyzing the interaction between Hfq protein and RNA.

Acknowledgment

This work was supported by the National Key R&D Program of China (Grant No. 2019YFA0905200), the National Natural Science Foundation of China (Grant No. 21933010), NYU-ECNU Center for Computational Chemistry at NYU Shanghai. We also thank the Supercomputer Center of East China Normal University (ECNU Public Platform for Innovation 001) for providing us computational time.

REFERENCES

1. Brosius, J.; Tiedge, H., RNomenclature. *RNA Biol.* **2004**, *1* (2), 81-83.
2. Hoe, C.-H.; Raabe, C. A.; Rozhdestvensky, T. S.; Tang, T.-H., Bacterial sRNAs: Regulation in stress. *Int. J. Med. Microbiol.* **2013**, *303* (5), 217-229.
3. Altuvia, S.; Weinstein-Fischer, D.; Zhang, A.; Postow, L.; Storz, G., A Small, Stable RNA Induced by Oxidative Stress: Role as a Pleiotropic Regulator and Antimutator. *Cell* **1997**, *90* (1), 43-53.
4. Chambers, J. R.; Sauer, K., Small RNAs and their role in biofilm formation. *Trends Microbiol.* **2013**, *21* (1), 39-49.
5. Gottesman, S.; McCullen, C. A.; Guillier, M.; Vanderpool, C. K.; Majdalani, N.; Benhammou, J.; Thompson, K. M.; FitzGerald, P. C.; Sowa, N. A.; FitzGerald, D. J., Small RNA regulators and the bacterial response to stress. *Cold Spring Harbor Symp. Quant. Biol.* **2006**, *71*, 1-11.
6. De Lay, N.; Schu, D. J.; Gottesman, S., Bacterial Small RNA-based Negative Regulation: Hfq and Its Accomplices. *J. Biol. Chem.* **2013**, *288* (12), 7996-8003.
7. Bossi, L.; Figueroa-Bossi, N., Competing endogenous RNAs: a target-centric view of small RNA regulation in bacteria. *Nat. Rev. Microbiol.* **2016**, *14* (12), 775-784.
8. Lalaouna, D.; Simoneau-Roy, M.; Lafontaine, D.; Massé, E., Regulatory RNAs and target mRNA decay in prokaryotes. *Biochim. Biophys. Acta, Gene Regul. Mech.* **2013**, *1829* (6), 742-747.
9. Chinni, S. V.; Raabe, C. A.; Zakaria, R.; Randau, G.; Hoe, C. H.; Zemann, A.; Brosius, J.; Tang, T.-H.; Rozhdestvensky, T. S., Experimental identification and characterization of 97 novel npcRNA candidates in Salmonella enterica serovar Typhi. *Nucleic Acids Res.* **2010**, *38* (17), 5893-5908.
10. Waters, L. S.; Storz, G., Regulatory RNAs in Bacteria. *Cell* **2009**, *136* (4), 615-628.
11. Saramago, M.; Bárria, C.; dos Santos, R. F.; Silva, I. J.; Pobre, V.; Domingues, S.; Andrade, J. M.; Viegas, S. C.; Arraiano, C. M., The role of RNases in the regulation of small RNAs. *Curr. Opin. Microbiol.* **2014**, *18*, 105-115.
12. Quendera, A. P.; Seixas, A. F.; dos Santos, R. F.; Santos, I.; Silva, J. P. N.; Arraiano, C. M.; Andrade, J. M., RNA-Binding Proteins Driving the Regulatory Activity of Small Non-coding RNAs in Bacteria. *Front. Mol. Biosci.* **2020**, *7*, 9.
13. Shimoni, Y.; Friedlander, G.; Hetzroni, G.; Niv, G.; Altuvia, S.; Biham, O.; Margalit, H., Regulation of gene expression by small non-coding RNAs: a quantitative view. *Mol. Syst. Biol.* **2007**, *3* (1), 138.
14. Schulz, E. C.; Seiler, M.; Zuliani, C.; Voigt, F.; Rybin, V.; Pogenberg, V.; Mücke, N.; Wilmanns, M.; Gibson, T. J.; Barabas, O., Intermolecular base stacking mediates RNA-RNA interaction in a crystal structure of the RNA chaperone Hfq. *Sci. Rep.* **2017**, *7* (1), 9903.
15. Sauter, C.; Basquin, J. r. m.; Suck, D., Sm-like proteins in Eubacteria: the crystal structure of the Hfq protein from Escherichia coli. *Nucleic Acids Res.* **2003**, *31* (14), 4091-4098.
16. Weichenrieder, O., RNA binding by Hfq and ring-forming (L)Sm proteins A trade-off between optimal sequence readout and RNA backbone conformation. *RNA Biol.* **2014**, *11* (5), 537-549.

17. Brennan, R. G.; Link, T. M., Hfq structure, function and ligand binding. *Curr. Opin. Microbiol.* **2007**, *10* (2), 125-133.
18. Møller, T.; Franch, T.; Højrup, P.; Keene, D. R.; Bächinger, H. P.; Brennan, R. G.; Valentin-Hansen, P., Hfq: A Bacterial Sm-like Protein that Mediates RNA-RNA Interaction. *Mol. Cell* **2002**, *9* (1), 23-30.
19. Otaka, H.; Ishikawa, H.; Morita, T.; Aiba, H., PolyU tail of rho-independent terminator of bacterial small RNAs is essential for Hfq action. *Proc. Natl. Acad. Sci.* **2011**, *108* (32), 13059.
20. Kovach, A. R.; Hoff, K. E.; Canty, J. T.; Orans, J.; Brennan, R. G., Recognition of U-rich RNA by Hfq from the Gram-positive pathogen *Listeria monocytogenes*. *Rna* **2014**, *20* (10), 1548-1559.
21. Sauer, E.; Schmidt, S.; Weichenrieder, O., Small RNA binding to the lateral surface of Hfq hexamers and structural rearrangements upon mRNA target recognition. *Proc. Natl. Acad. Sci.* **2012**, *109* (24), 9396.
22. Papenfort, K.; Vanderpool, C. K., Target activation by regulatory RNAs in bacteria. *Fems Microbiol. Rev.* **2015**, *39* (3), 362-378.
23. Chao, Y.; Vogel, J., The role of Hfq in bacterial pathogens. *Curr. Opin. Microbiol.* **2010**, *13* (1), 24-33.
24. Papenfort, K.; Vogel, J., Regulatory RNA in Bacterial Pathogens. *Cell Host & Microbe* **2010**, *8* (1), 116-127.
25. Chambers, J. R.; Bender, K. S., The RNA Chaperone Hfq Is Important for Growth and Stress Tolerance in *Francisella novicida*. *PLoS One* **2011**, *6* (5).
26. Zhang, A.; Wassarman, K. M.; Ortega, J.; Steven, A. C.; Storz, G., The Sm-like Hfq Protein Increases OxyS RNA Interaction with Target mRNAs. *Mol. Cell* **2002**, *9* (1), 11-22.
27. Updegrove, T. B.; Wartell, R. M., The influence of *Escherichia coli* Hfq mutations on RNA binding and sRNA•mRNA duplex formation in rpoS riboregulation. *Biochim. Biophys. Acta, Gene Regul. Mech.* **2011**, *1809* (10), 532-540.
28. Henderson, C. A.; Vincent, H. A.; Casamento, A.; Stone, C. M.; Phillips, J. O.; Cary, P. D.; Sobott, F.; Gowers, D. M.; Taylor, J. E.; Callaghan, A. J., Hfq binding changes the structure of *Escherichia coli* small noncoding RNAs OxyS and RprA, which are involved in the riboregulation of rpoS. *Rna* **2013**, *19* (8), 1089-1104.
29. Wang, L. J.; Wang, W. W.; Li, F. D.; Zhang, J. H.; Wu, J. H.; Gong, Q. G.; Shi, Y. Y., Structural insights into the recognition of the internal A-rich linker from OxyS sRNA by *Escherichia coli* Hfq. *Nucleic Acids Res.* **2015**, *43* (4), 2400-2411.
30. McCammon, J. A.; Gelin, B. R.; Karplus, M., Dynamics of folded proteins. *Nature* **1977**, *267* (5612), 585-590.
31. Elcock, A. H., Molecular simulations of diffusion and association in multimacromolecular systems. In *Numerical Computer Methods, Pt D*, Brand, L.; Johnson, M. L., Eds. Elsevier Academic Press Inc: San Diego, 2004; Vol. 383, pp 166-198.

32. Winkler, D. A., Ligand Entropy Is Hard but Should Not Be Ignored. *J. Chem. Inf. Model.* **2020**, *60* (10), 4421-4423.
33. Duan, L.; Liu, X.; Zhang, J. Z. H., Interaction Entropy: A New Paradigm for Highly Efficient and Reliable Computation of Protein–Ligand Binding Free Energy. *J. Am. Chem. Soc.* **2016**, *138* (17), 5722-5728.
34. Yan, Y.; Yang, M.; Ji, C. G.; Zhang, J. Z. H., Interaction Entropy for Computational Alanine Scanning. *J. Chem. Inf. Model.* **2017**, *57* (5), 1112-1122.
35. Liu, X.; Peng, L.; Zhou, Y.; Zhang, Y.; Zhang, J. Z. H., Computational Alanine Scanning with Interaction Entropy for Protein–Ligand Binding Free Energies. *J. Chem. Theory Comput.* **2018**, *14* (3), 1772-1780.
36. Song, J.; Qiu, L.; Zhang, J. Z. H., Erratum on “An efficient method for computing excess free energy of liquid”. *Sci. China. Chem.* **2018**, *61* (12), 1638-1638.
37. Qiu, L.; Yan, Y.; Sun, Z.; Song, J.; Zhang, J. Z. H., Interaction entropy for computational alanine scanning in protein–protein binding. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2018**, *8* (2), e1342.
38. Zou, Y.; Qian, Z.; Sun, Y.; Wei, G.; Zhang, Q., Orcein-Related Small Molecule O4 Destabilizes hIAPP Protofibrils by Interacting Mostly with the Amyloidogenic Core Region. *J. Phys. Chem. B* **2017**, *121* (39), 9203-9212.
39. Thai, N. Q.; Nguyen, H. L.; Linh, H. Q.; Li, M. S., Protocol for fast screening of multi-target drug candidates: Application to Alzheimer’s disease. *J. Mol. Graphics Modell.* **2017**, *77*, 121-129.
40. Shirts, M. R.; Pitner, J. W.; Swope, W. C.; Pande, V. S., Extremely precise free energy calculations of amino acid side chain analogs: Comparison of common molecular mechanics force fields for proteins. *J. Chem. Phys.* **2003**, *119* (11), 5740-5761.
41. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E., The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28* (1), 235-242.
42. Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C., ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11* (8), 3696-3713.
43. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79* (2), 926-935.
44. Darden, T.; York, D.; Pedersen, L., Particle mesh Ewald: an N.log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98* (12), 10089-10092.
45. Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C., Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **1977**, *23* (3), 327-341.