

# SCIENTIFIC REPORTS



OPEN

## Mechanism of mRNA-STAR domain interaction: Molecular dynamics simulations of Mammalian Quaking STAR protein

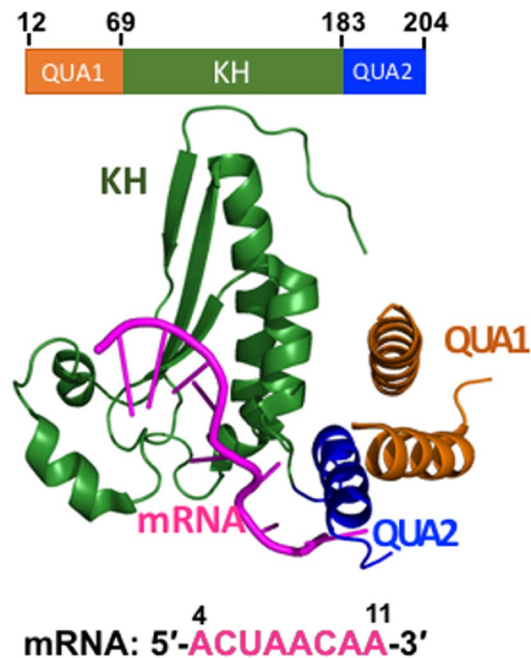
Monika Sharma  & C. R. Anirudh

STAR proteins are evolutionary conserved mRNA-binding proteins that post-transcriptionally regulate gene expression at all stages of RNA metabolism. These proteins possess conserved STAR domain that recognizes identical RNA regulatory elements as YUAAAY. Recently reported crystal structures show that STAR domain is composed of N-terminal QUA1, K-homology domain (KH) and C-terminal QUA2, and mRNA binding is mediated by KH-QUA2 domain. Here, we present simulation studies done to investigate binding of mRNA to STAR protein, mammalian Quaking protein (QKI). We carried out conventional MD simulations of STAR domain in presence and absence of mRNA, and studied the impact of mRNA on the stability, dynamics and underlying allosteric mechanism of STAR domain. Our unbiased simulations results show that presence of mRNA stabilizes the overall STAR domain by reducing the structural deviations, correlating the 'within-domain' motions, and maintaining the native contacts information. Absence of mRNA not only influenced the essential modes of motion of STAR domain, but also affected the connectivity of networks within STAR domain. We further explored the dissociation of mRNA from STAR domain using umbrella sampling simulations, and the results suggest that mRNA binding to STAR domain occurs in multi-step: first conformational selection of mRNA backbone conformations, followed by induced fit mechanism as nucleobases interact with STAR domain.

STAR (Signal Transduction and Activation of Ribonucleic acid) family of proteins are mRNA-binding proteins linking signal transduction with post-transcriptional gene regulation at various levels, such as mRNA shuttling, pre-mRNA splicing, and all stages of RNA metabolism<sup>1</sup>. Well studied examples include Splicing Factor 1 (SF1) mediating the intron recognition during spliceosome assembly<sup>2-4</sup>, Src-associated during mitosis 68-kDa protein (SAM 68) regulating alternative splicing in response to extracellular cues<sup>4-8</sup>, and Quaking related (QR) proteins<sup>9-12</sup>, critical in modulating development processes such as mammalian spermatogenesis (mice QKI), metazoan central nervous system development (GLD-1 in worms)<sup>13-16</sup>, sperm-to-oogenesis in hermaphrodites (GLD-1)<sup>17-19</sup>, or *Drosophila* wing development (HOW in flies)<sup>20-22</sup>. In humans, QKI proteins have been reported to be associated with numerous human pathologies like cancers<sup>12,23</sup> and neurological disorders such as human inherited ataxia, multiple sclerosis, or schizophrenia<sup>24,25</sup>. These recent findings warrant the need to understand the molecular basis of RNA recognition, binding, and regulation by STAR proteins.

Interestingly, all these proteins share a hallmark STAR domain of ~200 amino acids that recognizes and binds mRNA, and is conserved from yeast to humans<sup>1,24</sup>. STAR domain consists of N-terminal dimerization QUA1 domain, followed by mRNA binding extended single heterogeneous nuclear ribonucleoprotein K- Homology (KH) domain<sup>3,7,26-28</sup>. The latter is a conventional type I KH domain of  $\beta_1\alpha_1\alpha_2\beta_2\beta_3\alpha_3$  topology which recognizes four nucleotides of RNA by van der Waals forces, hydrophobic and electrostatic interactions<sup>29</sup>. Further downstream 2-3 nucleotides are recognized by fourth helix of C-terminal QUA2 domain<sup>28</sup>. Due to high sequence similarity, STAR family of proteins bind identical RNA regulatory elements (RREs) YUAAAY irrespective of their regulatory consequences<sup>28,30-32</sup>. Structure of the intact QKI STAR domains in complex with *in vivo* high affinity RNA targets has recently been solved<sup>28</sup>, revealing relative arrangement of QUA1, KH, and QUA2 domains within

Department of Chemical Sciences, Indian Institute of Science Education and Research (IISER), Sector 81, Knowledge City, SAS Nagar, Punjab, India. Correspondence and requests for materials should be addressed to M.S. (email: [msharma@iisermohali.ac.in](mailto:msharma@iisermohali.ac.in))



**Figure 1.** Crystal structure of monomer of QKI STAR-RNA complex (PDBid: 4JVH). Domain architecture of QKI domain is shown, consisting of N-terminal QUA1 domain (in orange), mRNA-binding KH-domain (in green), and C-terminal QUA2 domain (in blue). RNA target is shown in magenta.

each subunit of STAR homodimer. Both KH and QUA2 form an extended interface with bound mRNA (Fig. 1), and homodimerization is mediated exclusively by QUA1. Structure of STAR domain of mammalian Quaking protein in absence of RNA ligand is not available. However, structure of mRNA free KH-QUA2 region of *Xenopus* Quaking homologue, pXqua has been reported<sup>33</sup>. pXqua shows remarkable 94% sequence identity with human and mouse QKI proteins<sup>34</sup>. Based on <sup>15</sup>N relaxation parameters of QUA2 region of pXqua2, Maguire *et al.*<sup>33</sup> concluded that while KH domain is well defined, C-terminal QUA2 region in absence of mRNA ligand is more dynamic and partially structured in absence of mRNA (Figure S11A). Authors<sup>33</sup> speculated that QUA2 forms rigid helical structure once cognate mRNA is docked onto KH-QUA2 domain. These structures have thus, provided valuable insights into the molecular recognition of mRNA by STAR domain. However, they ignore information regarding the dynamic behavior of these domains that play critical role in their structural stability, and are relevant to the recognition and binding of mRNA. In this study, we employed atomistic molecular dynamics (MD) simulations to derive this information by studying the dynamics of intermolecular interactions in mRNA bound complexes of STAR domains of QKI. MD studies have been reliably applied previously to complexes of RNA with U1A protein containing the most abundantly found RNA recognition motif<sup>35–37</sup>. Therefore, the current study also aims to expand our understanding of the recognition of RNA elements by the second most occurring recognition fold in protein, K homology fold in STAR proteins.

Here, MD simulations have been carried out for STAR domain of Quaking protein in presence and absence of bound mRNA. Since our aim is to investigate influence of mRNA binding onto STAR domain, we have ignored the information of N-terminal dimerization QUA1 domain, and considered only truncated KH-QUA2 domain, as the latter is solely responsible for mRNA binding. We then analyzed the modes of dynamics using principal component analysis within the simulated systems and identified allosteric communications within the networks created using contact information and correlation data. Later on, we employed umbrella sampling technique to investigate how mRNA unbinds from pocket of STAR domain.

## Materials and Methods

**Simulated systems and Simulation protocol.** Molecular dynamics (MD) simulations of monomers of QKI with bound mRNA were carried out with explicit solvent. As our aim is to understand mRNA recognition and binding, we considered only mRNA binding KH-QUA2 domain for this study. Coordinates for mRNA bound KH-QUA2 domain of QKI were extracted from PDBid<sup>28</sup>: 4JVH (residues 69–204). In addition, RNA free conformations of QKI proteins were also simulated by removing mRNA coordinates from these mRNA-bound complexes. Herein, we refer to mRNA bound KH-QUA2 domain as HOLO state, and artificially generated mRNA free KH-QUA2 structure as APO state.

The CHARMM36 force field with the CMAP correction<sup>38</sup> was used to describe molecular interactions. All simulations were run with NAMD2.11<sup>39</sup>. The complexes were solvated in a rectangular TIP3P solvent box<sup>40</sup> and neutralized with Na<sup>+</sup> ions<sup>41</sup>, resulting in simulation box size of 84Åx83Åx84Å consisting ~60,000 atoms. Additional Na<sup>+</sup> and Cl<sup>-</sup> ions were added at a concentration of 0.15 M NaCl salt. Long-range electrostatic interactions were treated using the Particle Mesh Ewald (PME) technique<sup>42</sup> as implemented in NAMD<sup>43</sup>. A non-bonded

cut-off of 12 Å was applied to the Lennard-Jones potential and a cut-off for the direct-space part of the Coulomb forces with a switching function starting at distance 10 Å and reaching zero at 12 Å. A 2 fs time step was used for all MD simulations in combination with SHAKE<sup>44</sup> to holonomically constrain bonds involving hydrogens. Rigid water molecules were constrained using the SETTLE algorithm<sup>45</sup>. The systems were minimized under harmonic restraints of 2 kcal/mol/Å<sup>2</sup> in three steps. Initially, the backbone atoms of the protein and RNA were restrained while allowing only the water molecules to reorient for 50,000 steps. Then, restraints on only C $\alpha$  atoms of protein were applied to allow the protein and RNA to relax during an additional 50,000 steps of minimization. Finally, the restraints were reduced to 1 kcal/mol/Å<sup>2</sup> and the structures were minimized further for another 50,000 steps. The minimized structures were heated gradually from 0 K to 300 K at a rate of 30 K/60 ps while still maintaining the restraints. The simulation at 300 K was then continued for 100 ps before gradually releasing the restraints from 1 kcal/mol/Å<sup>2</sup> to 0 during another 200 ps. The final structures were then equilibrated for 10 ns under NPT ensemble. The simulation temperature and pressure was maintained at 300 K and 1 atm using Langevin dynamics<sup>46</sup> and Nose-Hoover Langevin piston method<sup>47,48</sup>, respectively. After equilibration, structures were subjected to NPT simulations. This method was repeated four times with different initial velocity assignments using random seed method to generate four independent production runs. The four runs span different simulation times of 75 ns, 100 ns, 200 ns, and 250 ns; reaching a total simulation time of 0.625  $\mu$ s for mRNA bound complexes. Similarly, mRNA free APO proteins were simulated three times generating three independent production runs spanning simulation times of 100 ns, 150 ns and 250 ns; resulting in total simulation time of 0.5  $\mu$ s.

**Analysis of the trajectories.** Trajectory analysis or in particular, structure analyses involving the deviations (RMSD, RMSF), intramolecular contacts, native contacts within RNA binding KH domain, protein-mRNA contacts, were carried out using VMD. RMSD and RMSF calculations have been done with reference to crystal structure, and for conformations sampled at every 2 ps during runs. Structural convergence of different runs was assessed based on computation of root mean square average correlation (RAC) of protein KH-QUA2 domain at different time intervals using the *rmscorr* command<sup>49</sup> in CPPTRAJ<sup>50</sup>. For RAC calculation, the fit was done over C $\alpha$  atoms of running average structures of residues 69–204 over the full trajectories calculated at each time interval from 0 ps to 250 ns referenced in the RMS fit to the average structure over the entire trajectory run. For evaluation of native contacts, we calculated for two measures: Q, the fraction of native contacts and Qs, similarity of natively contacting residue pairs to their native distances using carma software<sup>51</sup>. For Q and Qs calculation, a native contact is defined when C $\alpha$  atom of one residue interacts with C $\alpha$  of another residue (excluding the adjacent two neighboring residues) within cutoff distance of 8.0 Å. Qs is similar to Q except that a Gaussian penalty is introduced if the native contact is far from its native distance<sup>52</sup>. Native contacts between KH-QUA2 domain were calculated for KH-QUA2 domain after pooling in all the conformations sampled at every 2 ps during different trajectory runs. Variations in interactions between mRNA and protein residues within conformation sampled at every 2 ps were analyzed using VMD for different trajectory runs of HOLO state.

Principal component analysis has been done with Bio3D package written in R<sup>53,54</sup>. A variance-covariance matrix was constructed using C $\alpha$  atoms after pooling in all the conformations (sampled every ps) from the trajectories, and time-dependent projections of the first three principle components onto the trajectories have been studied. Contact map was constructed for C $\alpha$  atom of crystal structure using Bio3D package, considering contacts between cutoff of 8 Å, excluding neighboring residue.

**Determination of allosteric communications.** Allosteric networks within the proteins were identified using the NetworkView plugin<sup>55,56</sup> of VMD. The dynamical networks were constructed using data from all the trajectories of the protein-RNA complexes sampled every 2 ps. For each molecular system, a network graph was constructed with two nodes for each nucleotide molecule, N1 and N9 atoms were used to define the nodes for the base atoms for pyrimidines and purines, respectively; and phosphorus, P atom was used to define the nodes for the sugar and backbone phosphate atoms. To define nodes for each protein residue, C $\alpha$  atoms were used. All the conformations were pooled in to calculate the local-contact matrix, where a contact was defined between the two nodes (excluding the neighboring nodes) within 4.5 Å, and observed for at least 75% of the simulation time. The so-constructed contact matrix was then weighed by the correlation values of the two end nodes, where  $C_{ij}$  is the correlation matrix calculated as  $C_{ij} = \langle \Delta r_i \cdot \Delta r_j \rangle / \langle \Delta r_i^2 \rangle^{1/2} \langle \Delta r_j^2 \rangle^{1/2}$ . The correlation matrices were calculated using carma software<sup>51</sup>, and the shortest paths between two nodes in the network were found using the Floyd-Warshall algorithm<sup>57</sup>. Characteristic path lengths (CPL), average of all the shortest paths, were calculated as indicative of size of networks for both HOLO and APO states. The community structure was identified by using Girvan-Newman algorithm<sup>58</sup>, which uses edge betweenness to detect community peripheries. Betweenness of an edge is defined as the number of shortest paths between pairs of nodes that run along it, and is a measure of influence of a node over the flow of information between other nodes. Community detection is an iterative process, where each cycle consists of calculation of betweenness of all edges in the network, followed by removal of edge with the highest betweenness until no edge remain.

**Umbrella sampling simulations.** The limitation with the conventional MD is that the simulated systems get easily trapped in their local minima, and thus, cannot be used for sampling conformational transition processes such as the unbinding or binding of ligands. We employed umbrella sampling technique<sup>59–61</sup> to investigate the unbinding of mRNA from STAR domain. In order to simulate the unbinding process, distance between the centers of mass defined by group of atoms of STAR domain as well as for mRNA was used as reaction coordinate (RC). For mRNA, the P and O5' atoms of nucleic backbone of nucleotides 5 to 9 were used to calculate the center of mass. For STAR domain, center of mass was calculated for group of C $\alpha$  atoms present within 5.0 Å of mRNA in crystal structure. The RC was extended from initial distance of 10.5 Å to 20 Å (Figure S12). Umbrella sampling run consisted 15 simulation windows with 0.5 Å in length for RC between 10.5 Å to 15 Å; and 1 Å for RC between 15 Å

to 20 Å. Here an elastic constant of 5 kcal/mol Å<sup>2</sup> was used in all the US simulation windows. For each window, 50 ns were performed to converge the sampling, thus approximate 750 ns US simulations have been performed for the complex.

The aim of this study is to investigate the dissociation of mRNA from STAR domain, and not to calculate the binding free energy. Thus, we did not use the earlier proposed constrained scheme for ligand unbinding<sup>62–64</sup>. Nevertheless, to prevent the drifting of the systems, especially for STAR domain, two restraints were applied. First, orientation restraint of 100 kcal/mol Å<sup>2</sup> was added to the C $\alpha$  atoms of the protein to tackle the rotational drift. Secondly, restraint of 10 kcal/mol Å<sup>2</sup> was added to the C $\alpha$  atoms that are 15 Å away from the mRNA in the crystal structure to restrain them to the center of mass as calculated for similar atoms in the crystal structure. Potential of mean force (PMF) was constructed along the RC using Weighted Histogram Analysis Method (WHAM)<sup>65,66</sup>, after discarding the first 5 ns as equilibration period. Error bars were calculated with Monte Carlo bootstrap error analysis. RC was separated into 1000 bins for the WHAM calculation, with iteration tolerance value set to 0.0001. Temperature was set to 300 K to keep consistency with the simulation temperature.

It is possible to estimate multidimensional PMFs along additional degrees of freedom, whether biased or unbiased, with the assumption that all other degrees of freedom orthogonal to the biasing coordinates were thoroughly sampled<sup>67</sup>. To investigate the energy profile as mRNA explored different conformations within each US window, we calculated two-dimensional PMF along the rmsd of backbone atoms of mRNA as additional degree of freedom and biased RC. The rmsd values were calculated with respect to crystal structure for backbone atoms of nucleotides 5 to 10 (excluding O1P and O2P oxygen atoms of phosphate backbone moiety) for conformation sampled within each window, and these additional coordinates were added to the biasing coordinates with zero force constant. For obtaining two-dimensional PMF profile along RC and rmsd, the data was binned into 500 bins for 2D-WHAM calculation with tolerance value of 0.0001 and temperature set to 300 K.

Backbone conformations sampled by mRNA within each window during umbrella sampling was quantified in detail using two measures: sugar puckering configurations and pseudo-torsion angles. These calculations were performed for mRNA conformation sampled using 3DNA software<sup>68</sup>. The backbone pseudo-torsion angles<sup>69</sup> were defined as:  $\eta$ : C4'(i-1)-P(i)-C4'(i)-P(i+1) and  $\theta$ : P(i)-C4'(i)-P(i+1)-C4'(i+1).

## Results

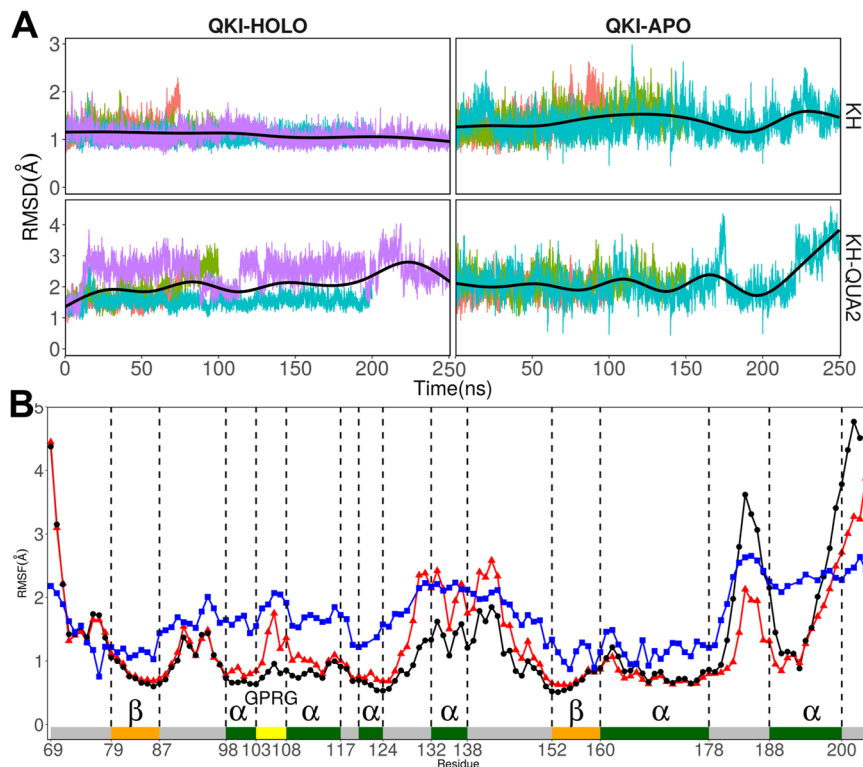
We have carried out simulations of mRNA bound and mRNA free extended KH (KH-QUA2) domains of QKI proteins. Throughout this manuscript, for convenience and clarity, we will refer these mRNA bound structures as HOLO states and mRNA free structures as APO states.

**Stability of STAR domain.** We analyzed changes in root mean square deviations (RMSD) values of the protein backbone atoms of QKI domain with respect to the crystal structure coordinates in response to the presence or absence of mRNA. Figure 2A shows the time evolution of RMSD values of backbone atoms of truncated KH-QUA2 protein by using crystal structure as reference for each simulation run. On an average, RMS deviations are lower for mRNA bound HOLO structures compared to mRNA free APO structures. To further investigate which part of STAR domain is contributing to large RMSD values, we carried out domain-wise RMSD calculations of KH and QUA2 domains. The time evolution plots of RMSDs of KH domain during each simulation run (Fig. 2A) indicate that mRNA binding stabilizes the KH domain and present low RMSD values in HOLO states. Inclusion of helical C-terminal QUA2 domain increases the RMSD values for both HOLO and APO states, indicating high flexibility of C-terminal domain irrespective of presence of mRNA.

The root mean square fluctuations (rmsf) is another indicator of the flexibility of certain residue during simulations. We calculated rmsf values of C $\alpha$  atoms of each protein residue, averaged over all the conformations sampled during different independent simulations runs for both HOLO and APO structures. For comparison with crystal structure, we converted B-factors or temperature factors to rmsf values by using Debye-Waller formula  $\left( RMSF = \left( \frac{3B}{8\pi^2} \right)^{1/2} \right)$  Figure 2B shows that in presence of mRNA, KH domain in HOLO state shows less fluctuations than observed in APO states. mRNA is bound onto hydrophobic surface created by the residues from helices  $\alpha$ 3,  $\alpha$ 4, strand  $\beta$ 2, conserved GPRG, and loops connecting  $\beta$ 2 and  $\alpha$ 5 of KH domain. Although these residues show high mobility in crystal, the rmsf values are lower in HOLO state than in APO state. The residues at the loops and at both N- and C-terminal regions are flexible in both mRNA bound and free states. It is noteworthy that residues in the C-terminal QUA2 region show elevated experimental RMSF, and exhibit large RMSF values in both APO and HOLO states, indicating that QUA2 domain is still relatively flexible in solution irrespective of presence of mRNA. The higher mobility of QUA2 domain is further confirmed by the comparison with its higher rms deviations observed for both APO and HOLO states.

We further assessed the convergence of each simulation run by computing pseudo-autocorrelation functions for RMSD values: root mean square (RMS) average correlation (RAC). A converged trajectory is when slope of RAC values approaches zero, indicating no further changes in the average structure as sampling time increases<sup>49</sup>. For QKI system, we fit each trajectory to the overall average structure and observed that for both HOLO and APO states: as simulation time increases, RAC slope smoothly decreases nearing zero at 250 ns (Figure S13). However, at this point we will not exclude the possibility that time scales may remain shorter to observe any drastic changes that would have occurred over longer timescales and reflected in the RAC values.

Comparing mRNA free APO state simulation results with experimentally available pXqua KH-QUA2 domain structure, we observe that though KH-QUA2 domain is well defined in both proteins, QUA2 domain is more flexible in pXqua protein. Solution structure of mRNA-free APO form of pXqua KH-QUA2<sup>33</sup> shows significant flexibility in C-terminal QUA2 domain with disordered helical structure and possesses different orientations with respect to  $\alpha$ 6 of KH domain (Figure S11A). It should be noted here that pXqua QUA2 domain is larger in size (26 residues: 188–214) compared to QKI QUA2 domain (16 residues: 188–204). In addition, overall secondary



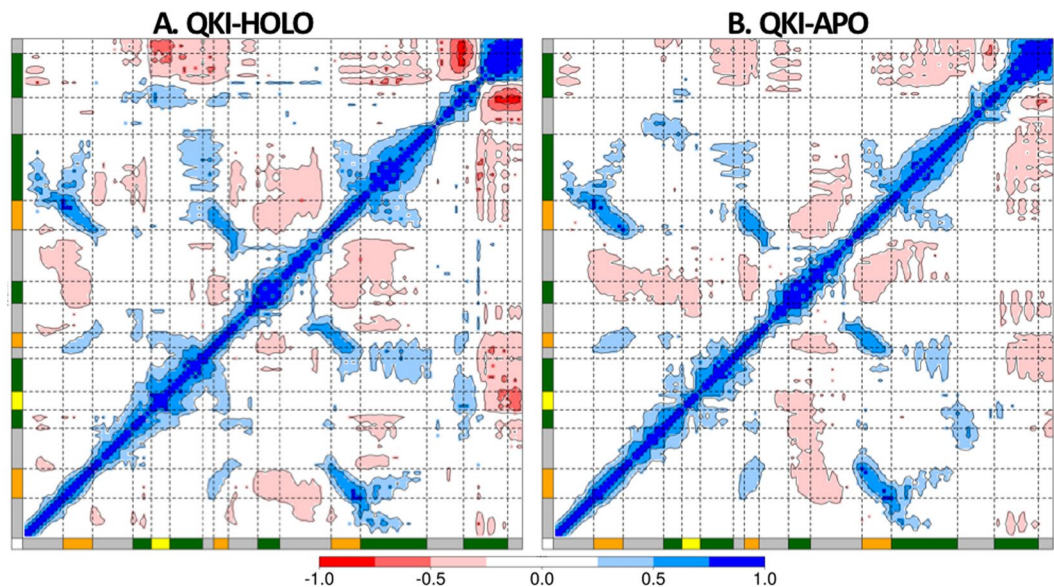
**Figure 2.** (A) Time-evolution of RMSD of STAR domain of QKI protein in presence of mRNA (QKI-HOLO state) and in absence of mRNA (QKI-APO state). RMSD values during independent runs are shown in varied colors, with averages shown in black thick line. (B) RMSF for C atoms of KH-QUA2 domain of QKI protein averaged over all the simulation runs. Data with black circles and black lines is for HOLO state, red colored triangles with red lines is for APO state, and blue colored squares with blue lines correspond to converted B-factors for crystal structure (4JVH). Secondary structure elements: alpha helices, beta sheets and loops are annotated with boxes of green color, orange color and grey color. 'GPRG' motif is annotated with yellow colored box.

structure assignment of solution structure of mRNA free pXqua KH-QUA2 is considerably different from X-ray determined structure of QKI KH-QUA2 (Figure S11B). This may be the reason that we do not see drastic changes in our QKI-APO state simulations, besides the limitations in timescales of conventional MD simulations.

The fact that Maguire *et al.* speculated<sup>33</sup> that QUA2 domain ought to be preformed even in absence of ligand in order to have effective mRNA binding; and since no experimental structural information is available for either mRNA bound pXqua KH-QUA2 or mRNA free QKI KH-QUA2 domain, it remains unclear about the exact conformational transition of KH-QUA2 domain going from ligand free APO state to mRNA bound HOLO state.

**Correlations within STAR domain.** We calculated the correlation matrix for C $\alpha$  position fluctuations of QKI for HOLO and APO states. Dynamic cross-correlation maps for the two states, representing correlation coefficients calculated as averaged over all the conformations sampled during the simulations, are shown in Fig. 3. The whole range of correlation from  $-1$  to  $+1$  is represented in three ranges: blue color corresponding to positive correlation values ranging from  $0.25$  to  $1$ ; red color corresponding to negative correlation values ranging from  $-0.25$  to  $-1$ ; and white color corresponding to weak or no-correlation values ranging from  $-0.25$  to  $+0.25$ . The extent of correlation or anti-correlation is indicated by variation in the intensity of respective blue or red color. A quick glance at the maps indicates that there is loss of correlation when mRNA is not bound as shown by reduction in colored islands in DCC maps of APO vs HOLO. Comparing with the contact map constructed from crystal structure coordinates of QKI STAR domain (pdbID: 4JVH) (Figure S14), we see that residues that form contact within crystal structure are positively correlated in both HOLO and APO states.

In QKI, mRNA binds onto KH domain within the cleft formed by helices  $\alpha_3$ ,  $\alpha_4$ ,  $\alpha_6$ , and strands  $\beta_1$ ,  $\beta_2$ ,  $\beta_3$  on one side and  $\alpha_5$  on other side. The binding is further extended by helix  $\alpha_7$  of QUA2 domain. mRNA specifically interacts with the residues from helices  $\alpha_3$ ,  $\alpha_4$ , strand  $\beta_2$ , conserved GPRG, and loops connecting  $\beta_2$  and  $\alpha_5$  of KH domain. These residues are observed to be highly self-correlated as indicated by the blue-colored islands in DCC map of HOLO state. This self-correlation decreases in absence of bound mRNA as observed from DCC map of APO state. Interestingly, looking at the correlation between two sides of cleft of mRNA binding region of KH domain, we observe that in presence of mRNA, correlation of  $\alpha_5$  on one side of the cleft with mRNA-binding helices  $\alpha_3$  and  $\alpha_4$  on other side of cleft is lost.  $\alpha_5$  is anti-correlated with helix  $\alpha_6$  and strands  $\beta_1$ ,  $\beta_2$ , and  $\beta_3$  in HOLO state. In APO state, as mRNA is absent, though  $\alpha_5$  remains anti-correlated with the strands  $\beta_1$ ,  $\beta_2$ , and



**Figure 3.** Dynamic cross correlation maps calculated as time-averaged for C $\alpha$  atoms of KH-QUA2 domain of QKI protein (**A**) in presence of bound mRNA, QKI-HOLO and (**B**) in absence of mRNA, QKI-APO. Secondary structure elements are shown as in Fig. 2.

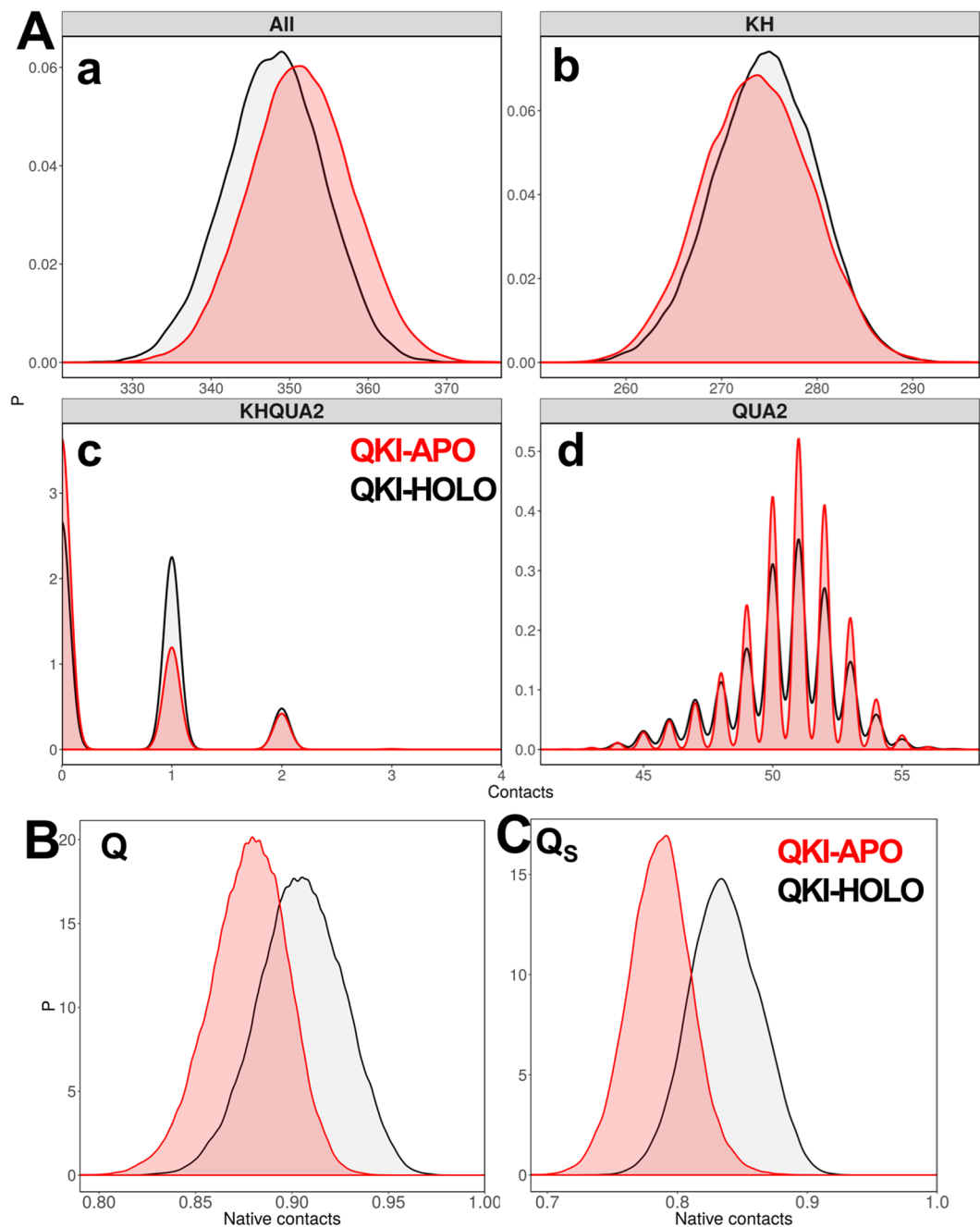
$\beta$ 3 on the other side of cleft, it loses the correlation with helix  $\alpha$ 6 and gets anti-correlated with mRNA-binding helices  $\alpha$ 3 and  $\alpha$ 4 on other side of cleft.

In presence of mRNA, motion of helix  $\alpha$ 7 of C-terminal QUA2 domain is anti-correlated with all the structural elements of KH domain, except  $\beta$ 1 and  $\alpha$ 5. In absence of mRNA, helix  $\alpha$ 7 gets anti-correlated with  $\alpha$ 5, and for the other SSEs, the extent of anti-correlation reduces as observed by reduction in the size as well as intensity of red colored islands. The generated DCC maps for both HOLO and APO states thus, disclosed that mRNA binding influences the nature of correlated motions of STAR domain.

**Impact of mRNA on the interactions within STAR domain.** To assess the impact of mRNA on the intra-residual contacts within STAR domain, we calculated total contacts for conformations sampled during each simulation run. Whenever one C $\alpha$  atom sees another C $\alpha$  atom within 6.0 Å, we define it as a contact. The distribution plots for the contacts (Fig. 4A) for the truncated KH-QUA2 domain suggest that in absence of mRNA, mRNA binding maxi-KH domain has more intramolecular contacts, which reduces in presence of mRNA. We further breakdown these contacts as intramolecular contacts within KH domain, within QUA2 domain, and inter-domain contacts between KH and QUA2. Interestingly, we observed that while for KH domain, the intramolecular domain contacts are slightly more in HOLO state than in APO state; the increase in total number of contacts in STAR domain is in fact consequence of increase in (a) intramolecular contacts within QUA2 domain and (b) interdomain contacts of QUA2 with KH domain. This suggests that in absence of mRNA, protein residues are free to form another set of interactions and inter-domain interactions between KH and QUA2 increases. Now, the question arises about the fate of native contacts that are present in experimental crystal structures; and how important is the presence of mRNA for intramolecular contacts within STAR domain.

To answer this query and further evaluate the structural differences in STAR domain associated with mRNA binding with respect to the experimental crystal structure, we calculated the fraction of native contacts (Q) and the similarity of natively contacting residue pairs to their native distances ( $Q_s$ ) for each conformation sampled during all the simulation runs for HOLO and APO states. Plotting the probability distribution functions for Q (Fig. 4B) and  $Q_s$  (Fig. 4C) for both HOLO and APO states, we observe that absence of mRNA shifts the peak plot to the left indicating loss of native contacts in APO state. The similarity measure,  $Q_s$  not only takes into consideration the existence of contact, but also quantify how close the contact is to the native distance<sup>52</sup>. Thus, the differences in HOLO and APO state peaks for  $Q_s$  in comparison to Q are modest yet subtle. The contact analyses indicate that though in absence of mRNA, STAR domain forms more number of intramolecular contacts; presence of mRNA is essential for maintaining the native contacts within STAR domain.

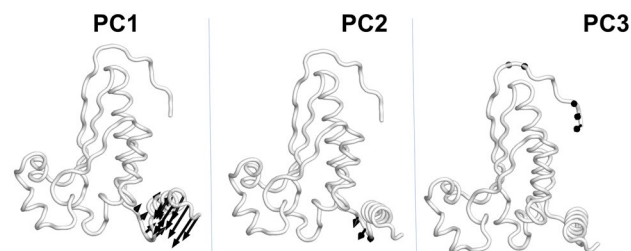
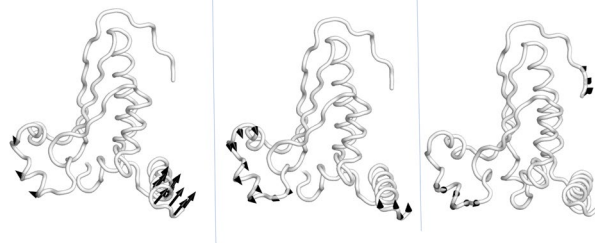
**Relative motions of STAR domain in response to mRNA.** Principal component analysis (PCA) separates conformational space of a system into essential subspaces by filtering global slow motions from the fast motions<sup>70</sup>. Thus, majority of the dynamics can be described by a surprisingly low number of collective degrees of freedom. Figure S15 shows that eigenvalues for both HOLO and APO states drop steeply in the first ten eigenmodes; and the detailed information related of the top three principal movements (PC1, PC2, and PC3) is presented in Fig. 5. The pointing of the arrows shows the direction of the motion vector and, the length of arrows represents the motion magnitude. We observe that STAR domain conducts differential motion modes in presence or absence of bound mRNA. In mRNA bound HOLO state, first principal mode is dominated by hinge movements in only QUA2 domain with stable KH domain. In APO state, when mRNA is not bound movements are



**Figure 4.** (A) Distribution of native contacts within STAR domain of QKI protein during the simulations (a) within maxi KH (KH-QUA2) domain, (b) within KH domain only, (c) between KH domain and QUA2 domain, and (d) within QUA2 domain. Probability distribution functions for (B) fraction of native contacts,  $Q$  and (C) similarity of natively contacting residue pairs to their native distances,  $Q_s$  for both HOLO (in black) and APO (in red) states.

observed in helix  $\alpha_5$  and QUA2 domain. For QUA2 domain, the direction of the movements in APO seems to be opposite to those observed in HOLO state. In second and third modes of HOLO state, movements are observed for flexible KH-QUA2 linker, and N-terminal of KH domain, respectively. For APO state, both second mode and third mode is dominated with correlated movements of helix  $\alpha_5$  with C-terminal of QUA2 domain and N-terminal of KH domain, respectively. The results of PCA analysis are in consistent with the DCCM analysis that motion of  $\alpha_5$  of KH domain is negatively correlated with motion of  $\alpha_7$  of QUA2 domain as suggested by their movements in opposite directions in the first mode of APO state.

**Network Analysis of STAR domain in response to mRNA.** We combined the correlation data obtained from MD simulations with the structure-based network analysis to determine evolution of residue interaction networks in presence or absence of mRNA. Protein structure networks were constructed by incorporating the

**A. QKI-HOLO****B. QKI-APO**

**Figure 5.** Motions along the first three principal components are shown for (A) mRNA bound HOLO and (B) mRNA free APO state.

topology-based residue connectivity in combination with the contact maps of residue cross-correlations obtained from MD simulations. Each residue and nucleotide represents a node in the network, and two nodes are connected by an edge if (a) nodes are non-neighbors, and (b) they are in contact within 4.5 Å of each other for more than 75% of simulation time. The significance of node is determined further by the dynamic information flow through that edge as measured by the correlation between respective nodes. The extent of correlation between the two nodes is depicted by the width of the edge, with wider edges indicating strongly correlated nodes. The weighted dynamical network representations for HOLO and APO states in Fig. 6A depicted the changes in interactions among residues in response to mRNA. The differences in networks are significantly visible in the mRNA binding cleft region of STAR domain as nodes are observed to have wider edges for helices  $\alpha 5$ ,  $\alpha 6$ ,  $\alpha 7$  (QUA2 domain) and strand  $\beta 3$ , suggesting enhanced stability and correlation within the STAR domain in presence of mRNA. Similar trend is observed in DCCM analysis discussed above.

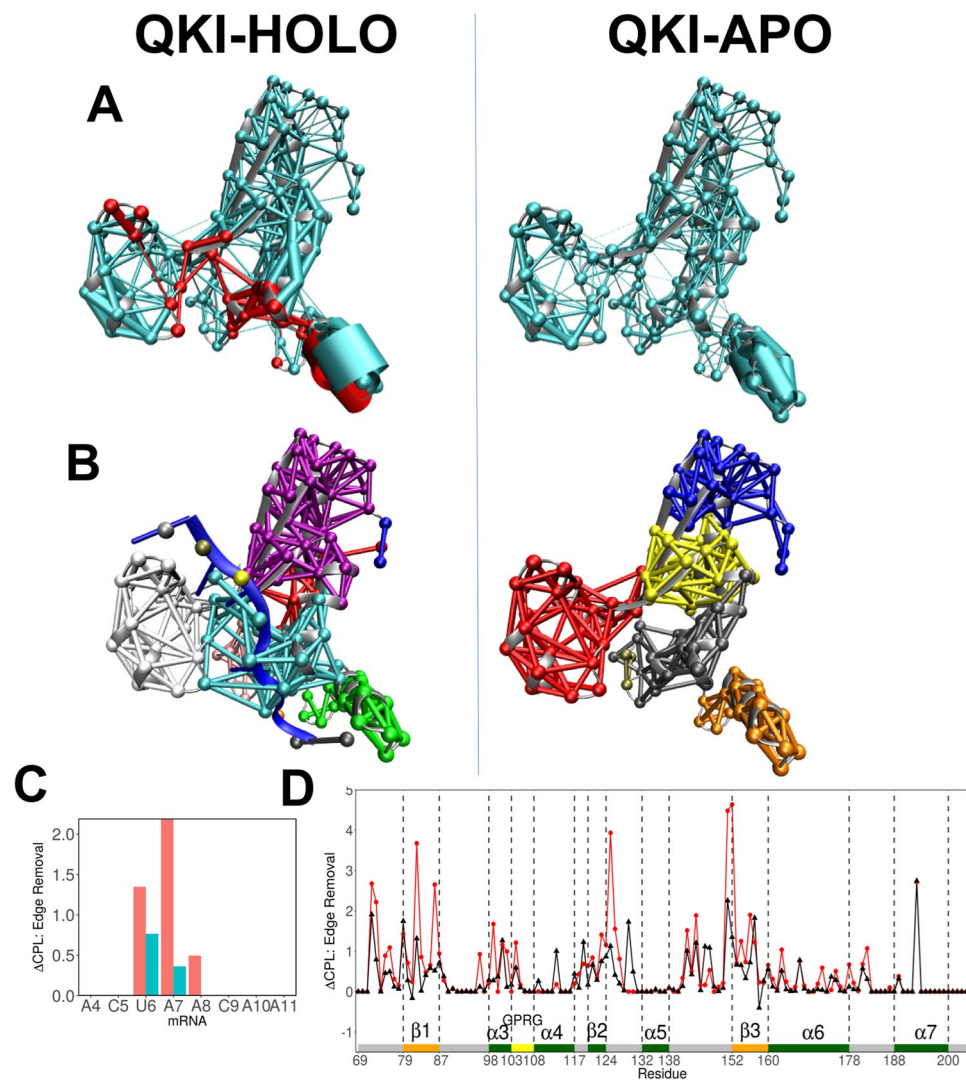
We analyzed the variations in the connectivity of networks in response to mRNA by viewing the changes in formation of local communities (Fig. 6B). Nodes within one community are more strongly interconnected to each other, compared to nodes within other communities. The way Girvan-Newman algorithm splits the HOLO and APO states is suggestive of the influence of mRNA binding to the STAR domain. In APO state, distinct communities are formed around mRNA binding cleft region: (a) N-terminal upper right part of cleft ( $\beta 1$ ,  $\beta 2$ ,  $\alpha 4$ ), (b) middle right part of cleft ( $\beta 3$ ,  $\alpha 3$ ), (c) left part of cleft ( $\alpha 5$ ), (d) down right of cleft ( $\alpha 6$ ), and (e) QUA2 domain ( $\alpha 7$ ). For HOLO state, the arrangement of communities changes and major communities observed are: (a) upper right and middle of cleft region merges as one community, (b) interface mRNA nucleotides and their protein interaction partners, (c) left part of cleft ( $\alpha 5$ ), and (d) QUA2 domain ( $\alpha 7$ ). Interestingly, residues from  $\alpha 3$  (on right side of cleft) and  $\alpha 5$  (on left side of cleft) form one community along with bound mRNA in HOLO state. In APO state, these form distinct communities.

To further identify nucleotides involved in communication across the interface, we looked at the differences in characteristic path length (CPL) upon edge removal for each nucleotide (Fig. 6C). Sethi *et al.*<sup>54</sup> observed that for allosteric signal transmission in the aaRS:tRNA complex, nucleotides involved in communication significantly increased the edge CPL. Similarly, for STAR-mRNA complex, nucleotides that significantly increased the edge CPL are the ones forming direct interactions with STAR domain: U6, A7, and A8. These residues form part of invariant core sequence observed via genome-wide search of STAR binding element of QKI: NA(A > C)U(A >> C)A (underlined nucleotides correspond to U6, A7 and A8)<sup>31,32</sup>. The stronger correlation of these nucleotides with STAR domain is also reflected in their interaction analyses. These three nucleotides form consistent native interactions with their interaction partners in STAR domain during all simulation runs compared to other nucleotides in the mRNA element (Figure S16).

We also analyzed variations in CPL upon complete removal of nodes (Figure S17) and edges (Fig. 6D) for STAR domain for HOLO and APO states. Though the results of node removal are a bit noisy, edge removal results seem to be more informative. The nucleotide binding residues in STAR domain are observed to increase CPL not only in HOLO state, but also in APO state. A noteworthy comparison is the alone increase in CPL upon removal of Q193 for QUA2 domain irrespective of presence of mRNA. Q193 is involved in interaction with U6 of mRNA, and this interaction remains consistent throughout the simulations of HOLO state.

**Mapping simulation data with mutational data.** Deleterious effects have been observed for N-ethyl-N-nitrosourea (ENU)-induced hydrophobic-to-polar point mutation, V157E in KH domain of QKI<sup>24,71</sup>.



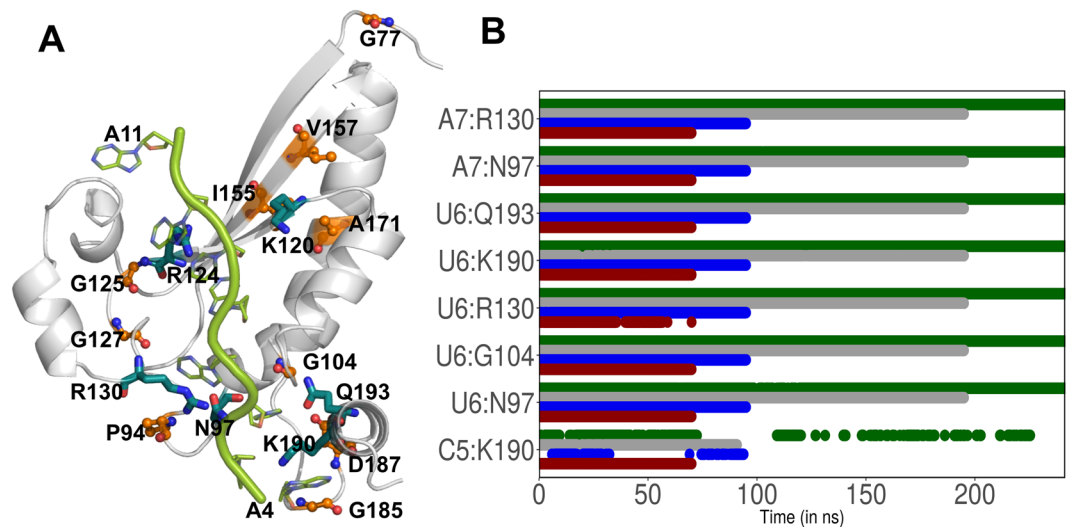


**Figure 6.** (A) Networks, weighed by correlation data, within KH-QUA2 domain of QKI protein is shown for mRNA-bound HOLO state and mRNA free APO state. Red colored nodes and edges correspond to interactions within 8 Å of mRNA. (B) Splitting of the networks into different communities are shown. (C) Change in CPL upon edge removal of nucleotide is shown for mRNA. Red and blue colored bars correspond to two nodes representing nucleotide base and sugar-phosphate backbone, respectively. (D) Change in CPL upon edge removal shown for KH-QUA2 domain for mRNA bound HOLO state (black colored) and mRNA free APO state (red colored). Structural elements are depicted as in Fig. 2B.

Another hydrophobic-to-polar mutation I155N showed severe developmental defects in zebrafish<sup>72</sup>. We analyzed solvent accessible surface area (SASA) values for these residues and observed average values as: I155 ( $0.06 \pm 0.29$ ), V157 ( $5.93 \pm 3.97$ ) for HOLO state, and I155 ( $0.06 \pm 0.25$ ), V157 ( $5.06 \pm 4.05$ ) for APO state simulations. Lower SASA values observed for these residues irrespective of mRNA presence indicate that these residues are buried within the hydrophobic pocket, and suggest that their mutation to polar residues will affect mRNA binding by strongly impairing the protein stability.

In homologous GLD-1 protein, eight missense mutations abolishing essential GLD-1 function in the germline have been identified: G200E, P217L, G227D, G248R, G250R, A294T, G308E, and D310N<sup>19</sup>. In QKI, these residues are equivalent to: G77, P94, G104, G125, G127, A171, G185, and D187, respectively. Teplova *et al.*<sup>28</sup> investigated *in vivo* and *in vitro* RNA-binding affinities with the mutations of key amino acids within the KH and QUA2 domains of QKI contacting specific bases of mRNA. Three double-mutants were identified: K190A/Q193A, N97A/R130A, and K120A/R124A<sup>28</sup>. As measured by ITC experiments K190A/Q193A mutation resulted in 2-fold reduction in *in vitro* RRE binding affinities, N97A/R130A mutation drops affinity by 10-fold, and K120A/R124A by 20-fold. Similar trend is observed for *in vivo* RRE binding quantified with PAR-CLIP. The signal for K190A/Q193A was reduced 5-fold, the signal for N97A/R130A was reduced 10-fold, and the signal for K120A/R124A was reduced 20-fold compared with wild-type protein.

Spatial organization of these residues in the bound mRNA-STAR domain complex indicates (Fig. 7A) that probable effect of mutations of residues: N97, G104, K120, R124, G125, R130, K190, and Q193 is because of their



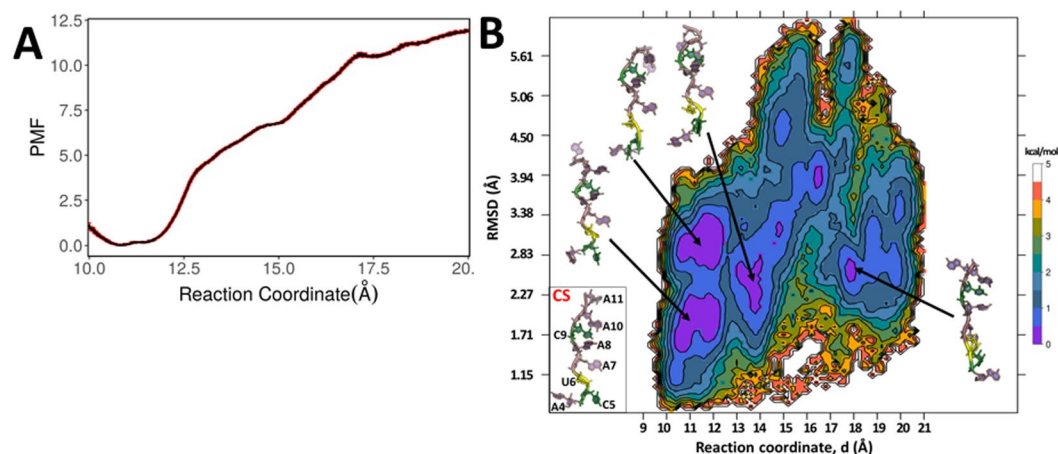
**Figure 7.** (A) Spatial organization of various mutations (reported for STAR proteins) in QKI protein. (B) Time evolution plots depicting interactions between mRNA nucleotide and its amino acid partner in STAR domain. Here interactions present in crystal structure are considered. Different colors indicate four different runs for HOLO state.

role in interacting mRNA nucleotides. Looking at the time evolution plots of native mRNA-protein interaction analysis (Fig. 7B), we observed that K190 interacts with C5 and U6 for more than 75% of simulation time. The interactions of U6 with N97, G104, R130 and Q193 are consistent throughout the simulations. U6 forms hydrogen bonds with G104 and Q193. Similarly, interactions of A7 with N97 and R130 are consistent throughout the simulations. As per deviation analysis, elevated rmsf values are observed for N97, G104, K120, R124, G125, R130 and Q193 residues in mRNA free APO states in comparison with bound HOLO states. Interestingly, some of these residues also show increase in characteristic path length for both APO and HOLO states, indicating their importance in functioning of STAR domain. Eight missense equivalent residues in QKI show comparative large increase in  $\Delta$ CPL upon node removal in APO state than in comparison with HOLO state, corroborating their role in structural stability of STAR domain (Table S11). Residues that interact with mRNA show increase in  $\Delta$ CPL upon node removal in HOLO state than in APO state indicating their role in recognition and binding of mRNA (Table S11).

**Dissociation of mRNA from bound STAR complex.** Conventional MD simulations have their limitations when it comes to simulate transition processes involving higher energetic barriers. For instance, simulation of dissociation of mRNA from bound STAR domain may take even more than millisecond of level of conventional simulations. To overcome this problem, we carried out umbrella sampling simulations to investigate unbinding process of mRNA from KH-QUA2 domain of QKI protein. We considered reaction coordinate as separation distance of mRNA from QKI domain, and from initial value of 10.5 Å, the distance was extended to 20 Å. We calculated PMF as a function of this reaction coordinate to determine the energetics of dissociation of mRNA from bound QKI-STAR complex (Fig. 8A). The comparative analysis of PMF profile indicates that dissociation of mRNA is energetically unfavorable process, and PMF values are lower for separation distance between 10.6–11.5 Å. This corresponds to the range observed for RC during conventional MD of bound HOLO state. As distance is decreased beyond this range, the energies rise resulting in dissociation of mRNA from the complex.

As we investigate further the conformational sampling of mRNA along the RC, we calculated two-dimensional PMF along biased reaction coordinate and unbiased coordinate as rmsd of mRNA with respect to crystal structure (Fig. 8B). We observe that as the mRNA dissociates from the complex, it has larger rmsd values from the initial structure. Interestingly, for the favorable RC range of 10.6–11.5 Å, we observed two minima at rmsd ranges of 1.7–2.2 Å and 2.8–3.3 Å as indicated by violet-colored islands in two-dimensional plot. mRNA within this RC range show lesser rmsd values than 4 Å as it is still interacting with protein. However, moving along RC beyond, we do observe violet colored minima at RC range of 13–14 Å with rmsd values of 1.8–2.6 Å. Few scattered lower energy conformations with rmsd less than 3 Å are observed at higher RC values. This suggests that even in absence of STAR domain, backbone of single stranded mRNA is able to adopt conformations similar to crystal structure. Looking at the representative structures for minima, we find that backbone conformations are similar to that observed in crystal structure but nucleobases adopt different conformations.

For such a small mRNA, RMSD values of more than 2 Å may include diverse backbone conformations. Thus, we further assessed the nucleic backbone conformations by calculating sugar pucker angles (Figure S18) and pseudo-torsions (Figure S19) for the nucleotides within each simulation window. In crystal structure, ribose sugars of nucleotides 4–7 are C2'-endo and 8–11 are C3'-endo. Looking at the pucker behavior of these residues, we observe that as mRNA dissociates from complex, the nucleotides 4–7 show transitions from C2'-endo to C3'-endo and occasionally to O4'-endo as well. However, nucleotides 8–11 remain primarily as C3'-endo. This similar behavior



**Figure 8.** (A) 1-D Potential of mean force (PMF) profile along the biased reaction coordinate as obtained from 15 umbrella sampling simulations within range of 10.5–20 Å. Errors in the PMF profile are shown in red. (B) Two-dimensional free energy profile along the biased reaction coordinate,  $d$  and rmsd values determined from umbrella sampling simulations. The representative structures are shown for each minima observed by violet-colored islands. Crystal structure of bound mRNA is shown for comparison.

is reflected in behavior of pseudo-torsions, where  $\eta$ - $\theta$  plots are analyzed and compared with crystal structure bound mRNA (Figure S19). We analyzed the time-dependent variations of pseudo-torsions as dial plots (Figure S110) where variations of  $\eta$  and  $\theta$  for each nucleotide of mRNA conformation sampled within each simulation window is shown. Mapping the pseudo-torsion plots with 2D PMF plot, we observe that mRNA backbone comprising nucleotides U6, A7, A8, C9, and A10 show similar patterns at low and higher RC values and these may contribute to violet colored minima at higher RC values. For instance, conformations sampled at RC 10.5 Å for backbone variations of U6, A7, A8, C9, and A10 are visited at RC value of 14–14.5 Å, or even at 17–19 Å.

These analyses suggest that conformation selection should be primary binding mechanism for mRNA to be recognized by STAR domain, and nucleobases rearrange later in vicinity of protein to interact with amino acid partners, probably via induced fit mechanism.

## Discussion

In this study, we have carried out simulations of mRNA bound and mRNA free truncated maxi-KH (KH-QUA2) domain of STAR mammalian QKI protein. mRNA binds to KH domain by interacting with residues from helices  $\alpha 3$ ,  $\alpha 4$ , strand  $\beta 2$ , conserved GPRG, and loops connecting  $\beta 2$  and  $\alpha 5$  of KH domain. The binding is further extended by helix  $\alpha 7$  of QUA2 domain. Presence of mRNA lowers the deviations and fluctuations for the interacting regions within KH domain. Single helical QUA2 domain on the other hand is quite flexible irrespective of mRNA presence. Presence of mRNA influences the correlated motions within STAR domain as self-correlation within these binding regions decreases or changes in absence of bound mRNA. The underline reason behind this is attributed to variations in intra-residual contacts in KH-QUA2 domain in presence or absence of mRNA. Total number of intramolecular contacts within KH-QUA2 domain increases in absence of mRNA, though number of contacts within KH domain are less in APO state than in HOLO state. In APO state, flexible C-terminal QUA2 domain is free to have more interactions within itself and with mRNA-free KH domain. On the contrary, native contacts are maintained in presence of bound mRNA throughout the simulations of HOLO state. The first principal mode detected with PCA of APO state showed opening motions for mRNA binding cleft in absence of mRNA, indicating flexibility of KH-QUA2 domain where mRNA bind onto.

We further analyzed the KH-QUA2 domain in absence and presence of mRNA as network of interactions between the residues, with residues defined as nodes and interactions within them as edges. These edges were weighed with correlation data, and stronger correlations are observed in presence of mRNA. We further divided the network into local communities, and observed distinct communities in HOLO and APO states. C-terminal QUA2 domain in both HOLO and APO state forms one distinct community. More communities are observed in mRNA binding region of KH domain in APO state than in HOLO state. We computed differences in characteristic path length upon removal of each node or edge in the network. Noteworthy, we found that if we remove nodes or edges consisting nucleotides or residues that are experimentally reported significant for binding and regulation of mRNA by STAR domain in QKI or related STAR proteins, considerable differences in characteristic path length are observed.

We further investigated dissociation of mRNA from KH-QUA2 domain using enhanced umbrella sampling simulations. PMF profile calculated along biased reaction coordinate suggested that mRNA dissociation is an energetic unfavorable process, with minima near the bound values of complex. We further explored PMF along rmsd of mRNA with respect to the crystal structure as additional degree of freedom, and observed that even at higher RC values (indicating dissociated mRNA from STAR domain), few minima at lower rmsd values are observed. Mapping mRNA conformations to such regions, we observed that mRNA is able to adopt backbone conformations in absence of bound protein but the nucleobases adopt different conformations. This was further substantiated with calculation of sugar pucker angles and pseudotorsions to define backbone conformations of mRNA sampled within each umbrella sampling simulation window.

Based on our analysis on conventional and umbrella sampling simulations, we propose that mRNA binding is multi-step process. In APO state, STAR domain is flexible around mRNA binding cleft region formed by helix  $\alpha 5$  on left side of cleft, and rest of KH and C-terminal QUA2 domain on right side of cleft. mRNA in free state is able to adopt multiple conformations, out of which extended backbone conformations similar to crystal structure are binding competent forms. Such conformations successfully interact onto the cleft region, and the nucleotides rearrange to form specific interactions with STAR domain via induced fit mechanism. A multi-step binding mode has very well observed in other biological complex systems<sup>73–77</sup> as well. In order to delineate exact mechanism of mRNA binding, more extensive simulation studies are required which are beyond the scope of this research and is a subject of further interest.

## Conclusion

By using conventional simulations, we have investigated the influence of mRNA binding onto the KH-QUA2 domain of STAR Quaking (QKI) protein. The effect of mRNA binding has been analyzed on the basis of deviations, correlations, intramolecular contacts, and interaction network. Our results mapped with the experimental mutational data reported. We further explored the dissociation of mRNA from STAR domain by employing enhanced umbrella sampling techniques and we observed that binding of mRNA onto STAR domain is an interplay of conformational selection of extended mRNA backbone conformations followed by induced fit of nucleobases to their amino acid partners in STAR domain.

## References

- Vernet, C. & Artzt, K. STAR, a gene family involved in signal transduction and activation of RNA. *Trends Genet.* **13**, 479–484 (1997).
- Arning, S., Grüter, P., Bilbe, G. & Krämer, A. Mammalian splicing factor SF1 is encoded by variant cDNAs and binds to RNA. *RNA* **2**, 794–810 (1996).
- Liu, Z. *et al.* Structural basis for recognition of the intron branch site RNA by splicing factor 1. *Science* **294**, 1098–1102 (2001).
- Matter, N., Herrlich, P. & König, H. Signal-dependent regulation of splicing via phosphorylation of Sam68. *Nature* **420**, 691–695 (2002).
- Reddy, T. R. *et al.* Inhibition of HIV replication by dominant negative mutants of Sam68, a functional homolog of HIV-1 Rev. *Nat. Med.* **5**, 635–642 (1999).
- Coyle, J. H. *et al.* Sam68 Enhances the Cytoplasmic Utilization of Intron-Containing RNA and Is Functionally Regulated by the Nuclear Kinase Sik/BRK. *Mol. Cell. Biol.* **23**, 92–103 (2003).
- Lukong, K. E. & Richard, S. Sam68, the KH domain-containing superSTAR. *Biochim. Biophys. Acta* **1653**, 73–86 (2003).
- Paronetto, M. P., Achsel, T., Massiello, A., Chalfant, C. E. & Sette, C. The RNA-binding protein Sam68 modulates the alternative splicing of Bcl-x. *J. Cell Biol.* **176**, 929–939 (2007).
- Larocque, D. & Richard, S. QUAKING KH domain proteins as regulators of glial cell fate and myelination. *RNA Biol.* **2**, 37–40 (2005).
- Larocque, D. *et al.* Nuclear Retention of MBP mRNAs in the Quaking Viable Mice. *Neuron* **36**, 815–829 (2002).
- Larocque, D. *et al.* Protection of p27(Kip1) mRNA by quaking RNA binding proteins promotes oligodendrocyte differentiation. *Nat. Neurosci.* **8**, 27–33 (2005).
- Yang, G. *et al.* RNA-binding protein quaking, a critical regulator of colon epithelial differentiation and a suppressor of colon cancer. *Gastroenterology* **138**, 231–240.e1–5 (2010).
- Francis, R., Barton, M. K., Kimble, J. & Schedl, T. *gld-1*, a tumor suppressor gene required for oocyte development in *Caenorhabditis elegans*. *Genetics* **139**, 579–606 (1995).
- Lee, M.-H. & Schedl, T. Translation repression by GLD-1 protects its mRNA targets from nonsense-mediated mRNA decay in *C. elegans*. *Genes Dev.* **18**, 1047–1059 (2004).
- Schumacher, B. *et al.* Translational repression of *C. elegans* p53 by GLD-1 regulates DNA damage-induced apoptosis. *Cell* **120**, 357–368 (2005).
- Crittenden, S. L. *et al.* A conserved RNA-binding protein controls germline stem cells in *Caenorhabditis elegans*. *Nature* **417**, 660–663 (2002).
- Francis, R., Maine, E. & Schedl, T. Analysis of the multiple roles of *gld-1* in germline development: interactions with the sex determination cascade and the *glp-1* signaling pathway. *Genetics* **139**, 607–630 (1995).
- Jan, E., Motzny, C. K., Graves, L. E. & Goodwin, E. B. The STAR protein, GLD-1, is a translational regulator of sexual identity in *Caenorhabditis elegans*. *EMBO J.* **18**, 258–269 (1999).
- Jones, A. R. & Schedl, T. Mutations in *gld-1*, a female germ cell-specific tumor suppressor gene in *Caenorhabditis elegans*, affect a conserved domain also found in Src-associated protein Sam68. *Genes Dev.* **9**, 1491–1504 (1995).
- Nabel-Rosen, H., Dorevitch, N., Reuveny, A. & Volk, T. The balance between two isoforms of the *Drosophila* RNA-binding protein how controls tendon cell differentiation. *Mol. Cell* **4**, 573–584 (1999).
- Volk, T., Israeli, D., Nir, R. & Toledano-Katchalski, H. Tissue development and RNA control: 'HOW' is it coordinated? *Trends Genet. TIG* **24**, 94–101 (2008).
- Zaffran, S., Astier, M., Gratecos, D. & Sémériva, M. The held out wings (how) *Drosophila* gene encodes a putative RNA-binding protein involved in the control of muscular and cardiac activity. *Dev. Camb. Engl.* **124**, 2087–2098 (1997).
- Novikov, L. *et al.* QKI-mediated alternative splicing of the histone variant MacroH2A1 regulates cancer cell proliferation. *Mol. Cell. Biol.* **31**, 4244–4255 (2011).
- Chénard, C. A. & Richard, S. New implications for the QUAKING RNA binding protein in human disease. *J. Neurosci. Res.* **86**, 233–242 (2008).
- Bockbrader, K. & Feng, Y. Essential function, sophisticated regulation and pathological impact of the selective RNA-binding protein QKI in CNS myelin development. *Future Neurol.* **3**, 655–668 (2008).
- Ryder, S. P. & Williamson, J. R. Specificity of the STAR/GSG domain protein Qk1: implications for the regulation of myelination. *RNA NYN* **10**, 1449–1458 (2004).
- Beuck, C. *et al.* Structure of the GLD-1 homodimerization domain: Insights into STAR protein-mediated translational regulation. *Struct. Lond. Engl.* **1993**(18), 377–389 (2010).
- Teplova, M. *et al.* Structure-function studies of STAR family Quaking proteins bound to their *in vivo* RNA target sites. *Genes Dev.* **27**, 928–940 (2013).
- Valverde, R., Edwards, L. & Regan, L. Structure and function of KH domains. *FEBS J.* **275**, 2712–2726 (2008).
- Galarneau, A. & Richard, S. Target RNA motif and target mRNAs of the Quaking STAR protein. *Nat. Struct. Mol. Biol.* **12**, 691–698 (2005).
- Hafner, M. *et al.* Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* **141**, 129–141 (2010).
- Jungkamp, A.-C. *et al.* *In vivo* and transcriptome-wide identification of RNA binding protein target sites. *Mol. Cell* **44**, 828–840 (2011).

33. Maguire, M. L. *et al.* Solution structure and backbone dynamics of the KH-QUA2 region of the Xenopus STAR/GSG quaking protein. *J. Mol. Biol.* **348**, 265–279 (2005).
34. Zorn, A. M. & Krieg, P. A. The KH domain protein encoded by quaking functions as a dimer and is essential for notochord development in Xenopus embryos. *Genes Dev.* **11**, 2176–2190 (1997).
35. Kormos, B. L., Baranger, A. M. & Beveridge, D. L. Do collective atomic fluctuations account for cooperative effects? Molecular dynamics studies of the U1A-RNA complex. *J. Am. Chem. Soc.* **128**, 8992–8993 (2006).
36. Kormos, B. L., Baranger, A. M. & Beveridge, D. L. A study of collective atomic fluctuations and cooperativity in the U1A-RNA complex based on molecular dynamics simulations. *J. Struct. Biol.* **157**, 500–513 (2007).
37. Showalter, S. A. & Hall, K. B. Correlated motions in the U1 snRNA stem/loop 2:U1A RBD1 complex. *Biophys. J.* **89**, 2046–2058 (2005).
38. Huang, J. & MacKerell, A. D. CHARMM36 all-atom additive protein force field: validation based on comparison to NMR data. *J. Comput. Chem.* **34**, 2135–2145 (2013).
39. Phillips, J. C. *et al.* Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26**, 1781–1802 (2005).
40. 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.* **79**, 926–935 (1983).
41. Beglov, D. & Roux, B. Finite representation of an infinite bulk system: Solvent boundary potential for computer simulations. *J. Chem. Phys.* **100**, 9050–9063 (1994).
42. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: An N-log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089–10092 (1993).
43. Essmann, U. *et al.* A smooth particle mesh Ewald method. *J. Chem. Phys.* **103**, 8577–8593 (1995).
44. 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.* **23**, 327–341 (1977).
45. Miyamoto, S. & Kollman, P. A. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J. Comput. Chem.* **13**, 952–962 (1992).
46. Brünger, A., Brooks, C. L. & Karplus, M. Stochastic boundary conditions for molecular dynamics simulations of ST2 water. *Chem. Phys. Lett.* **105**, 495–500 (1984).
47. Martyna, G. J., Tobias, D. J. & Klein, M. L. Constant pressure molecular dynamics algorithms. *J. Chem. Phys.* **101**, 4177–4189 (1994).
48. Feller, S. E., Zhang, Y., Pastor, R. W. & Brooks, B. R. Constant pressure molecular dynamics simulation: The Langevin piston method. *J. Chem. Phys.* **103**, 4613–4621 (1995).
49. Galindo-Murillo, R., Roe, D. R. & Cheatham, T. E. Convergence and reproducibility in molecular dynamics simulations of the DNA duplex d(GCACGAACGAACGAACGC). *Biochim. Biophys. Acta* **1850**, 1041–1058 (2015).
50. Roe, D. R. & Cheatham, T. E. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **9**, 3084–3095 (2013).
51. Glykos, N. M. Software news and updates. *Carma*: a molecular dynamics analysis program. *J. Comput. Chem.* **27**, 1765–1768 (2006).
52. Cho, S. S., Levy, Y. & Wolynes, P. G. P versus Q: Structural reaction coordinates capture protein folding on smooth landscapes. *Proc. Natl. Acad. Sci.* **103**, 586–591 (2006).
53. Grant, B. J., Rodrigues, A. P. C., ElSawy, K. M., McCammon, J. A. & Cavas, L. S. D. Bio3d: an R package for the comparative analysis of protein structures. *Bioinforma. Oxf. Engl.* **22**, 2695–2696 (2006).
54. Skjærven, L., Yao, X.-Q., Scarabelli, G. & Grant, B. J. Integrating protein structural dynamics and evolutionary analysis with Bio3D. *BMC Bioinformatics* **15**, 399 (2014).
55. Sethi, A., Eargle, J., Black, A. A. & Luthey-Schulten, Z. Dynamical networks in tRNA:protein complexes. *Proc. Natl. Acad. Sci. USA* **106**, 6620–6625 (2009).
56. Eargle, J. & Luthey-Schulten, Z. NetworkView: 3D display and analysis of protein-RNA interaction networks. *Bioinformatics* **28**, 3000–3001 (2012).
57. Floyd, R. W. Algorithm 97: Shortest Path. *Commun ACM* **5**, 345– (1962).
58. Girvan, M. & Newman, M. E. J. Community structure in social and biological networks. *Proc. Natl. Acad. Sci. USA* **99**, 7821–7826 (2002).
59. Torrie, G. M. & Valleau, J. P. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J. Comput. Phys.* **23**, 187–199 (1977).
60. Adcock, S. A. & McCammon, J. A. Molecular dynamics: survey of methods for simulating the activity of proteins. *Chem. Rev.* **106**, 1589–1615 (2006).
61. Kästner, J. Umbrella sampling. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **1**, 932–942 (2011).
62. Woo, H.-J. & Roux, B. Calculation of absolute protein-ligand binding free energy from computer simulations. *Proc. Natl. Acad. Sci. USA* **102**, 6825–6830 (2005).
63. Doudou, S., Burton, N. A. & Henchman, R. H. Standard Free Energy of Binding from a One-Dimensional Potential of Mean Force. *J. Chem. Theory Comput.* **5**, 909–918 (2009).
64. Gumbart, J. C., Roux, B. & Chipot, C. Standard Binding Free Energies from Computer Simulations: What Is the Best Strategy? *J. Chem. Theory Comput.* **9**, 794–802 (2013).
65. Kumar, S., Rosenberg, J. M., Bouzida, D., Swendsen, R. H. & Kollman, P. A. THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.* **13**, 1011–1021 (1992).
66. Souaille, M. & Roux, B. Extension to the weighted histogram analysis method: combining umbrella sampling with free energy calculations. *Comput. Phys. Commun.* **135**, 40–57 (2001).
67. Banavali, N. K. & Roux, B. Free Energy Landscape of A-DNA to B-DNA Conversion in Aqueous Solution. *J. Am. Chem. Soc.* **127**, 6866–6876 (2005).
68. Lu, X.-J. & Olson, W. K. 3DNA: a versatile, integrated software system for the analysis, rebuilding and visualization of three-dimensional nucleic-acid structures. *Nat. Protoc.* **3**, 1213–1227 (2008).
69. Keating, K. S., Humphris, E. L. & Pyle, A. M. A new way to see RNA. *Q. Rev. Biophys.* **44**, 433–466 (2011).
70. Amadei, A., Linssen, A. B., de Groot, B. L., van Aalten, D. M. & Berendsen, H. J. An efficient method for sampling the essential subspace of proteins. *J. Biomol. Struct. Dyn.* **13**, 615–625 (1996).
71. Justice, M. J. & Bode, V. C. Three ENU-induced alleles of the murine quaking locus are recessive embryonic lethal mutations. *Genet. Res.* **51**, 95–102 (1988).
72. Lobbardi, R. *et al.* Fine-tuning of Hh signaling by the RNA-binding protein Quaking to control muscle development. *Dev. Camb. Engl.* **138**, 1783–1794 (2011).
73. Qin, F. *et al.* Induced fit or conformational selection for RNA/U1A folding. *RNA NYN* **16**, 1053–1061 (2010).
74. Changeux, J.-P. & Edelstein, S. Conformational selection or induced fit? 50 years of debate resolved. *FI000 Biol. Rep.* **3** (2011).
75. Bucher, D., Grant, B. J. & McCammon, J. A. Induced fit or conformational selection? The role of the semi-closed state in the maltose binding protein. *Biochemistry (Mosc.)* **50**, 10530–10539 (2011).
76. Vogt, A. D. & Di Cera, E. Conformational Selection or Induced Fit? A Critical Appraisal of the Kinetic Mechanism. *Biochemistry (Mosc.)* **51**, 5894–5902 (2012).
77. Morando, M. A. *et al.* Conformational Selection and Induced Fit Mechanisms in the Binding of an Anticancer Drug to the c-Src Kinase. *Sci. Rep.* **6**, 24439 (2016).

## Acknowledgements

MS thanks Department of Science and Technology (DST), India for INSPIRE Award and research grant (IFA14-CH-165).

## Author Contributions

M.S. conceived, designed, and performed the experiments. M.S. analyzed the data and wrote the paper. M.S. and A.C.R. contributed the literature materials, and reviewed the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-017-12930-2>.

**Competing Interests:** The authors declare that they have no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017