This journal is © The Royal Society of Chemistry 2011

Supplementary Figures

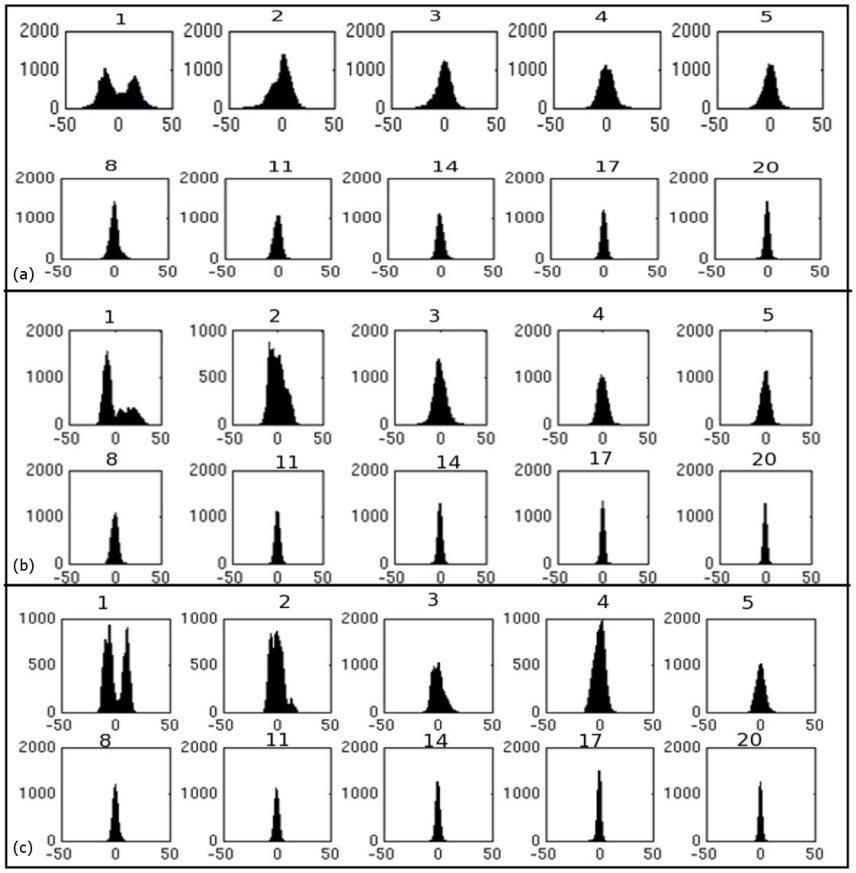


Figure S1: Histograms resulting from the projection of structures from (a) Sys1, (b) Sys2, and (c) Sys3 onto the top 20 modes (i.e the eigenvectors corresponding to the top 20 eigenvalues). Each figure is labelled by its corresponding eigenvalue rank (i.e 1, 2 ,... ,20). Each histogram has the same x-scale (-50 to 50) for easy comparison. The x-axis represents the projection of structures binned together on the corresponding essential mode. The number of bins used in the histogram is 50. It is evident from the plot that for each system, the top/top two modes exhibit a bimodal distribution whereas the remaining components show progressively sharpening unimodal distribution which reflects their decreasing contribution to variance. These modes mainly represent minor Gaussian fluctuations about a mean position and do not contribute majorly to conformational variations.

This journal is © The Royal Society of Chemistry 2011

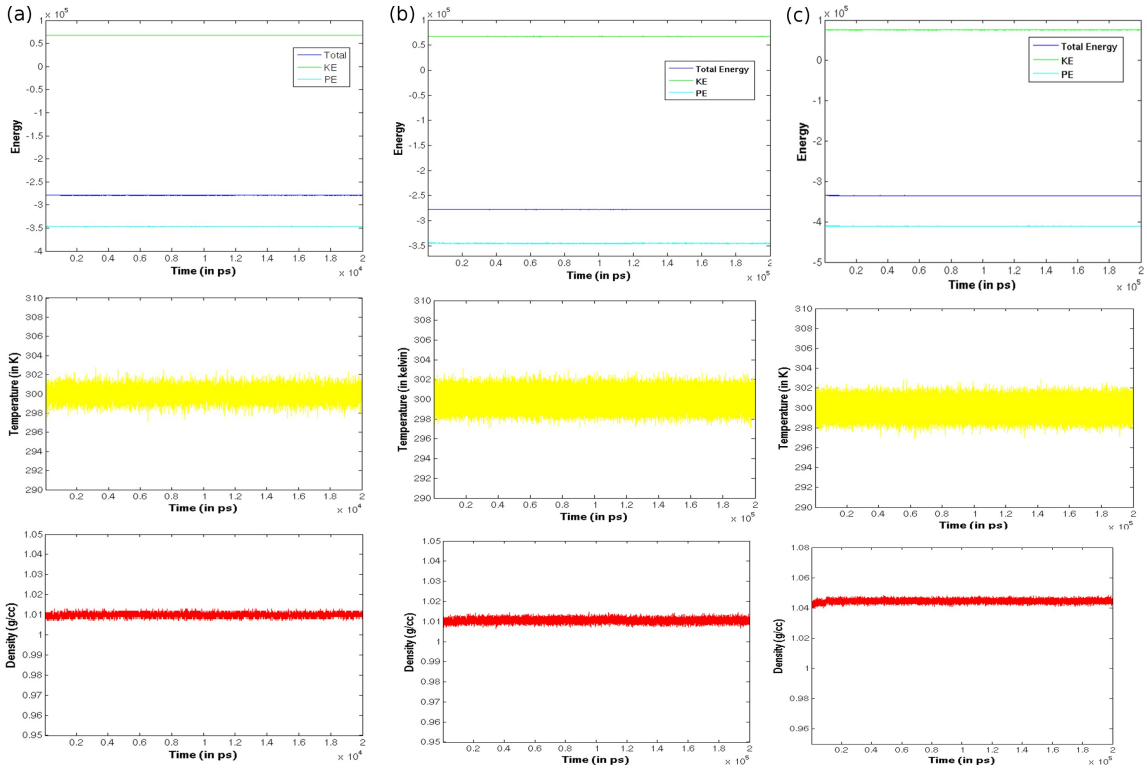
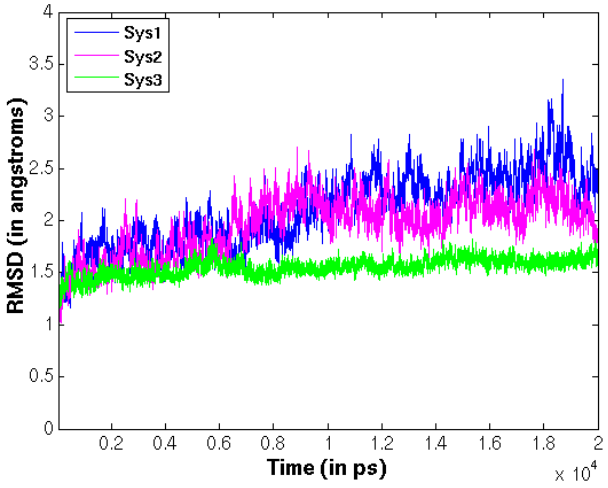


Figure S2: Plots of thermodynamic parameters (like total energy, kinetic energy, potential energy, temperature and density) Vs. Time for (a) Sys1, (b) Sys2, and (c) Sys3. All three simulations exhibit overall thermodynamic stability. For ease of comparison, each plot has the same scales along x and y-axes for a given parameter. Time along the x-axis starts from 30ps in each plot.



This journal is © The Royal Society of Chemistry 2011

Figure S3: Root-mean-square deviation (RMSD) profile for Sys1 (native DhPylRS), Sys2 (DhPylRS+2yly), and Sys3 (DhPylRS+2yly+2tRNA) with reference to the minimized crystal structures. Time along the x-axis starts from 30ps.

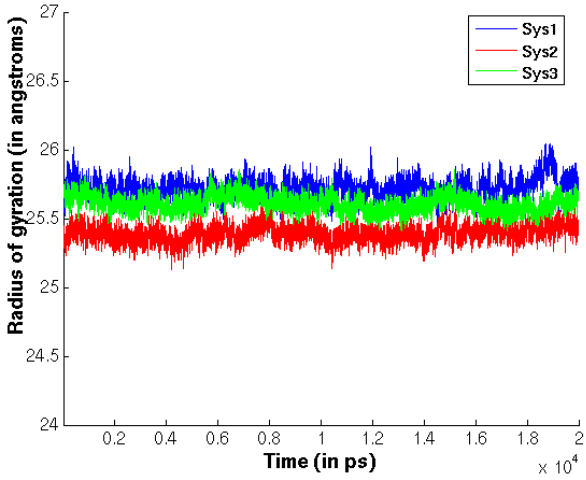


Figure S4: Plot of radius of gyration Vs. Time for Sys1 (native DhPylRS), Sys2 (DhPylRS+2yly), and Sys3 (DhPylRS+2yly+2tRNA). The radius of gyration time-series seems to have reached a plateau indicating convergence of simulations. Time along the x-axis starts from 30ps.

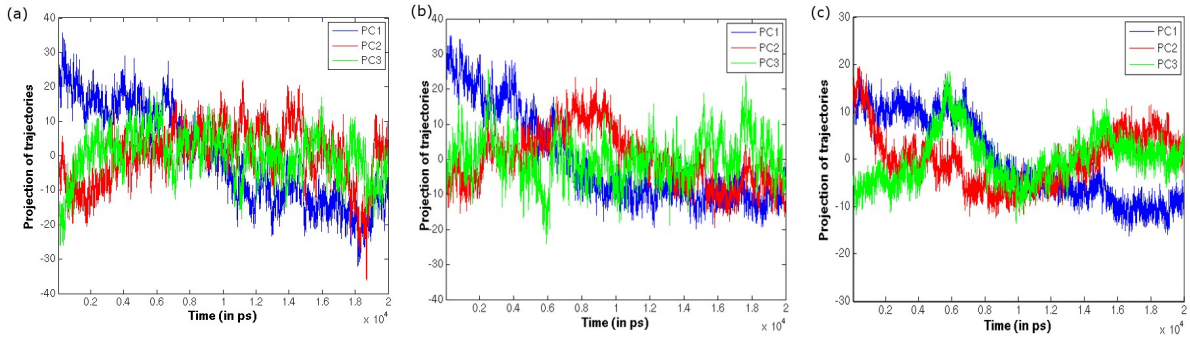


Figure S5: Projection of trajectories on top three principal components for (a) Sys1 (native DhPylRS), (b) Sys2 (DhPylRS+2yly), and (c) Sys3 (DhPylRS+2yly+2tRNA).The scales are same along the two axes for all the plots for ease of comparison.

This journal is © The Royal Society of Chemistry 2011

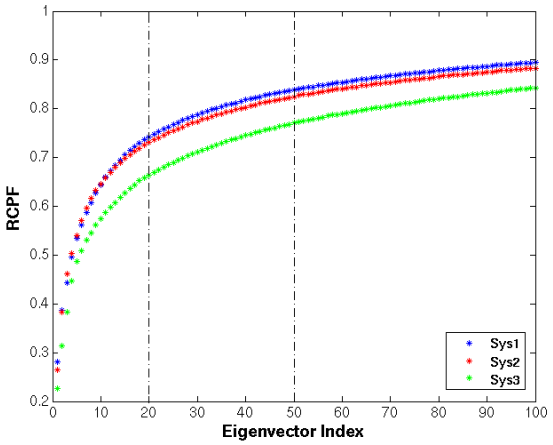


Figure S6: Relative Cumulative Positional fluctuation for the three systems (Sys1-3) are presented, which shows about 65-75% of the overall fluctuation is captured by the top 20 modes (~80% of total fluctuation is captured by top 50 modes).

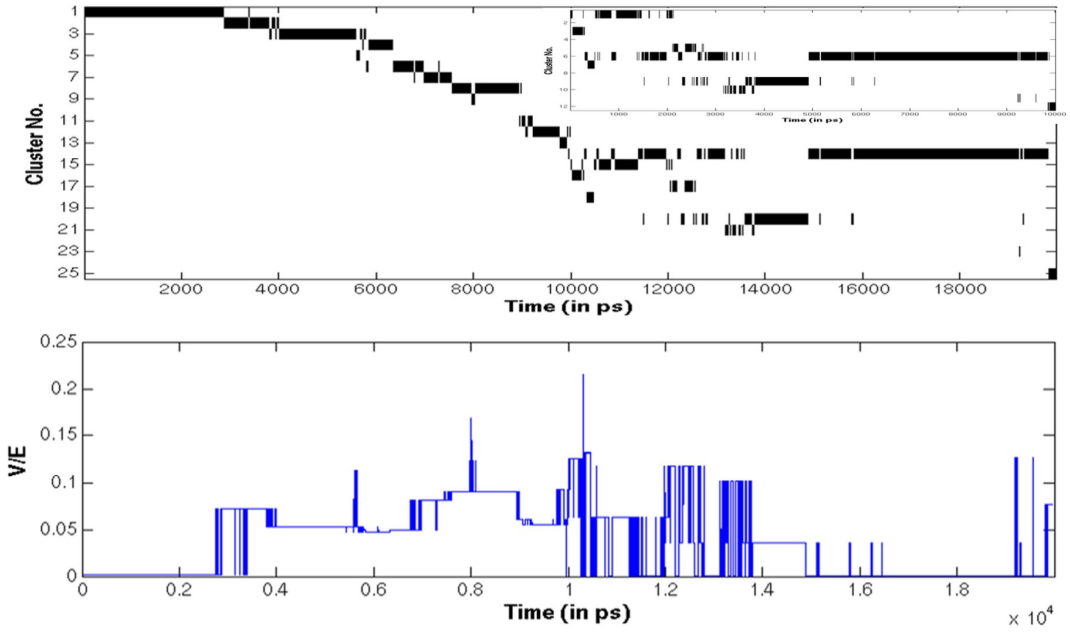


Figure S7(a)

This journal is © The Royal Society of Chemistry 2011

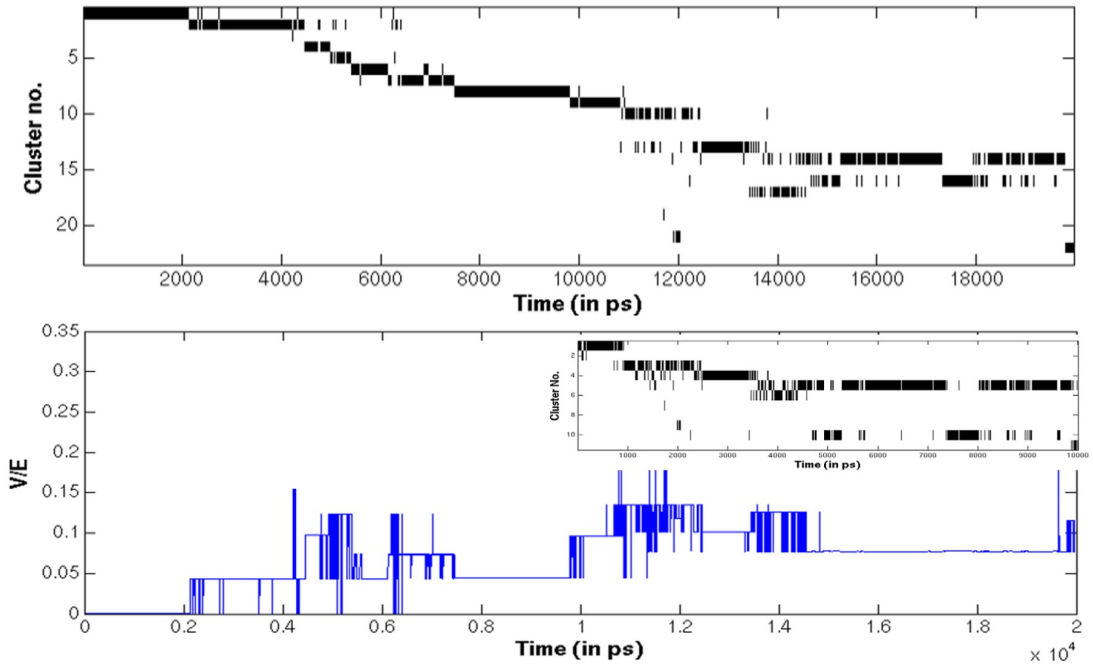


Figure S7(b)

Figure S7: Conformational clusters obtained from quantum clustering of the snapshots for (a) DhPylRS (Sys1) and (b) DhPylRS + Pyl-AMP (Sys2). 25 and 23 conformational clusters are identified for Sys1 and Sys2 respectively (upper panel) based on their V(x)/E values. The closer conformations are binned in a particular potential well by QC. No clusters are formed by an overlap of snapshots from pre and post equilibrium range. The lower panel depicts the corresponding V(x)/E values for the conformational clusters given in the upper panel along the trajectory. In the insets, the clusters obtained from independent QCs of the equilibrated trajectories (10-20ns) for Sys1-2 shows very similar clustering pattern in the time range of 10-20ns with the results obtained from QC on full 20 ns trajectory. Time along the x-axis starts from 30ps.

This journal is © The Royal Society of Chemistry 2011

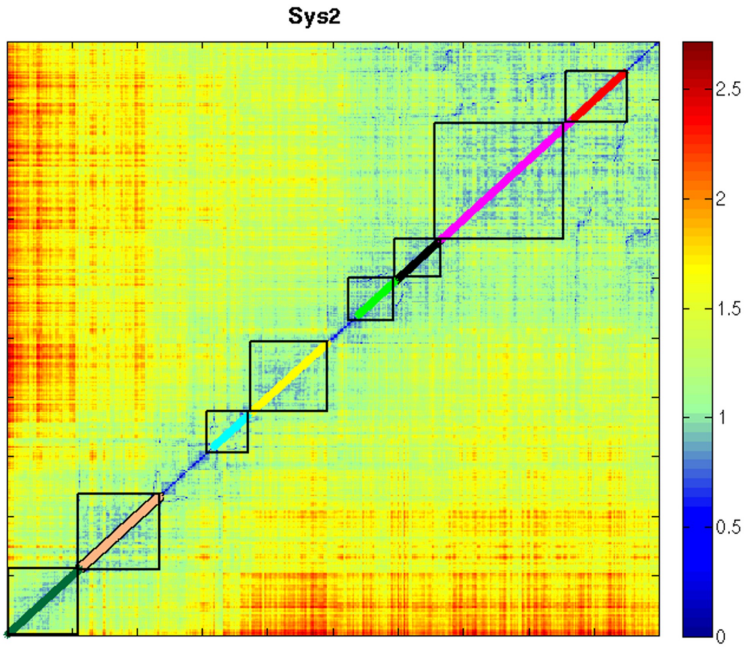


Figure S8: 2D-RMSD plot for Sys2 re-arranged based on the clusters obtained from QC. The eight highly populated clusters are highlighted and labelled and their corresponding time points in the MD trajectory are according to that in Figure S7(b). The 2D-RMS deviation is less variable within a cluster (approximately < 1 Å) in contrast to that across the clusters.

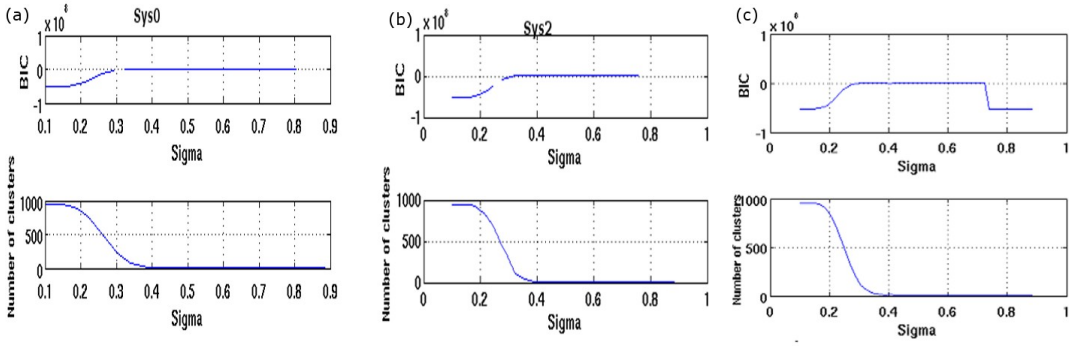


Figure S9: The upper panel shows the variation of BIC score with the σ values and the lower panel exhibits the variation of the number of clusters as a function of σ for (a) Sys1, (b) Sys2, and (c) Sys3 respectively. The values of σ corresponding to the maximum BIC score is chosen for optimal clustering of conformational data in each of the three systems [details of the σ values are presented in Table S1].

This journal is © The Royal Society of Chemistry 2011

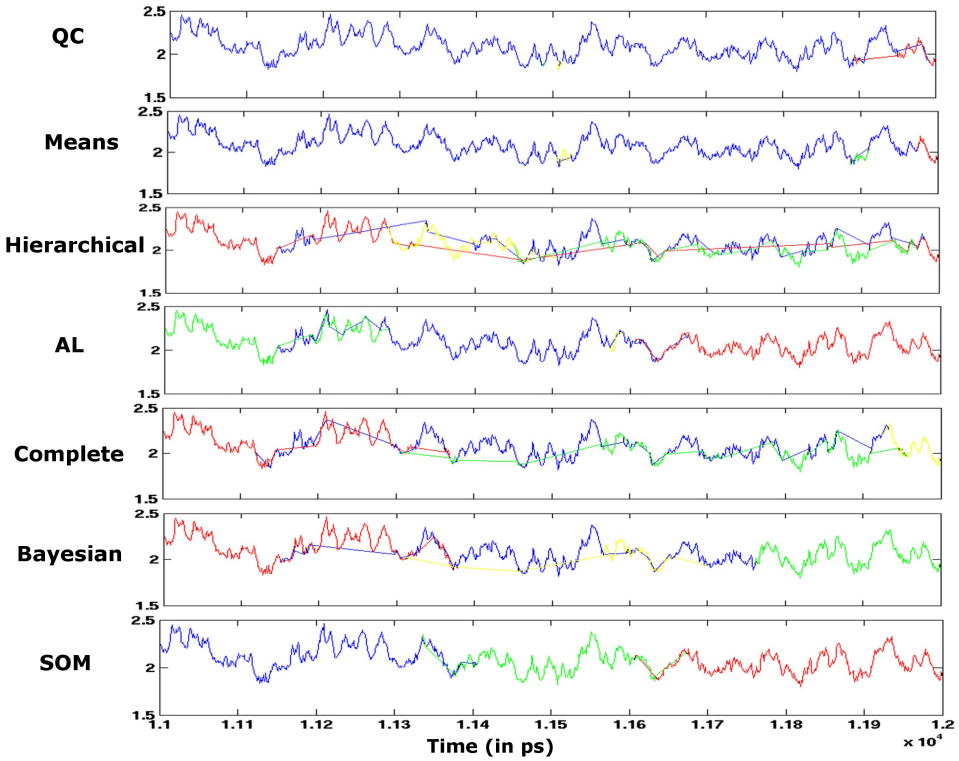


Figure S10: Clusters obtained from the application of seven different clustering algorithms for test data [Sys2 (11-12 ns)] shown in an RMSD versus time plot. The four clusters in each case (with the exception of the three clusters obtained from SOM) are coloured according to their size (blue > red > green > yellow). The clustering patterns are different in each case with exception of some similarity in pattern between QC and means for this smaller subset of data.

This journal is © The Royal Society of Chemistry 2011

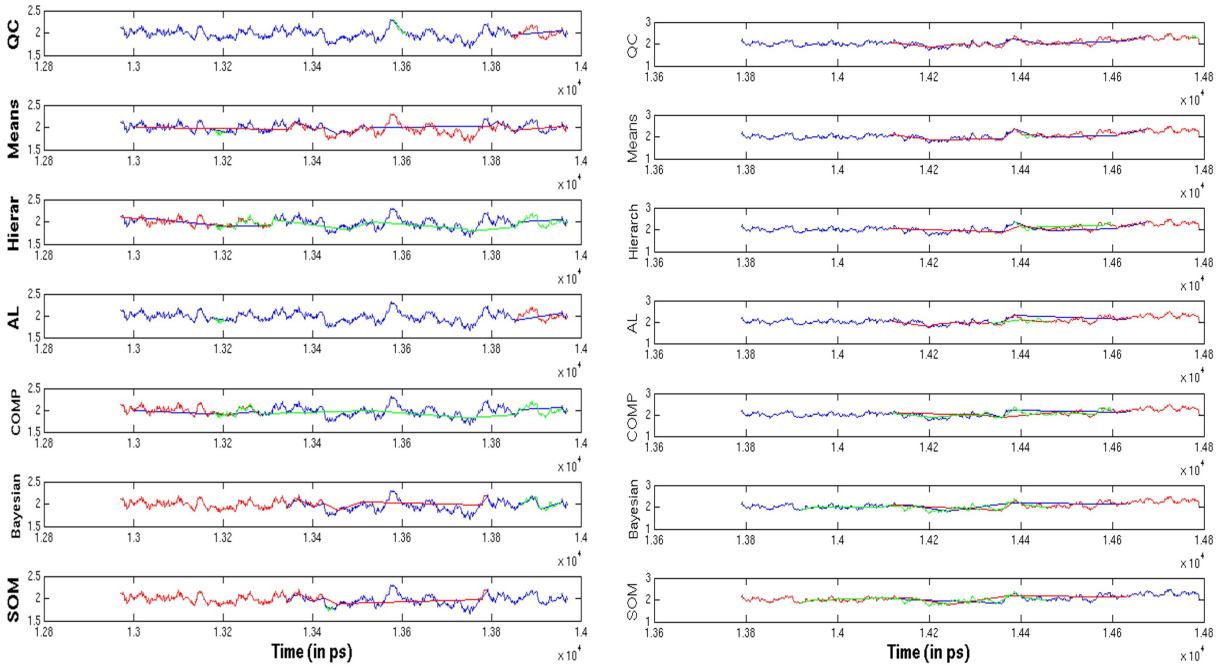


Figure S11: Clusters obtained from the application of seven different clustering algorithms for test data [Sys2 (13-14/14-15 ns)] shown in an RMSD versus time plot. The three clusters in each case are coloured according to their size (blue > red > green). The clustering patterns are different in each case with exception of some similarity in pattern between QC and means/average linkage for this smaller subset of data.

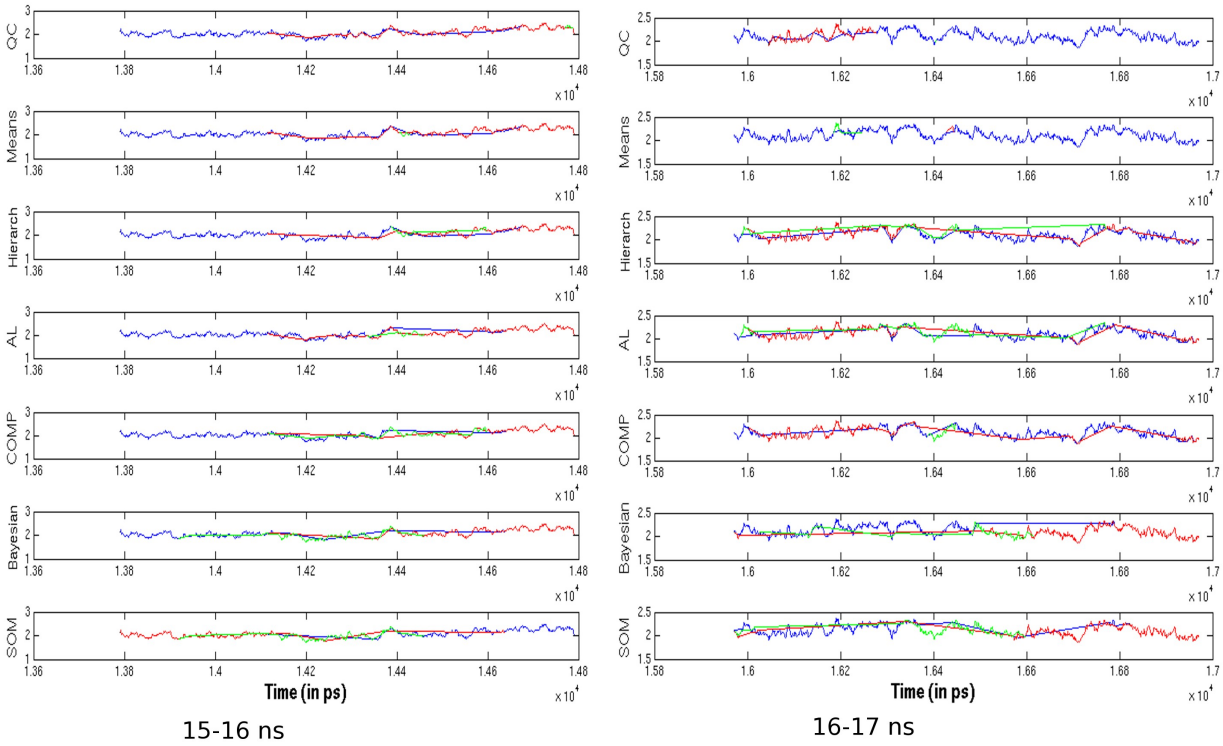


Figure S12: Clusters obtained from the application of seven different clustering algorithms for test data [Sys2 (15-16/16-17 ns)] shown in an RMSD versus time plot. The three clusters in each case are coloured according to their size (blue > red > green). The clustering patterns are different in each case with exception of some similarity in pattern between QC and means/average linkage for this smaller subset of data.

This journal is © The Royal Society of Chemistry 2011

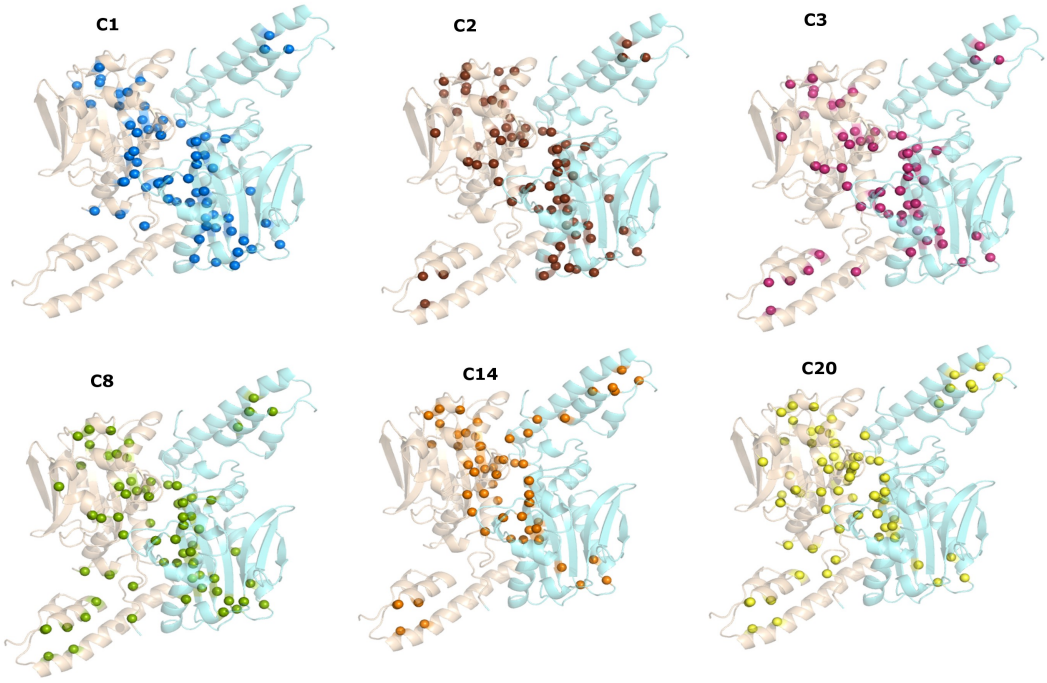


Figure S13(a)

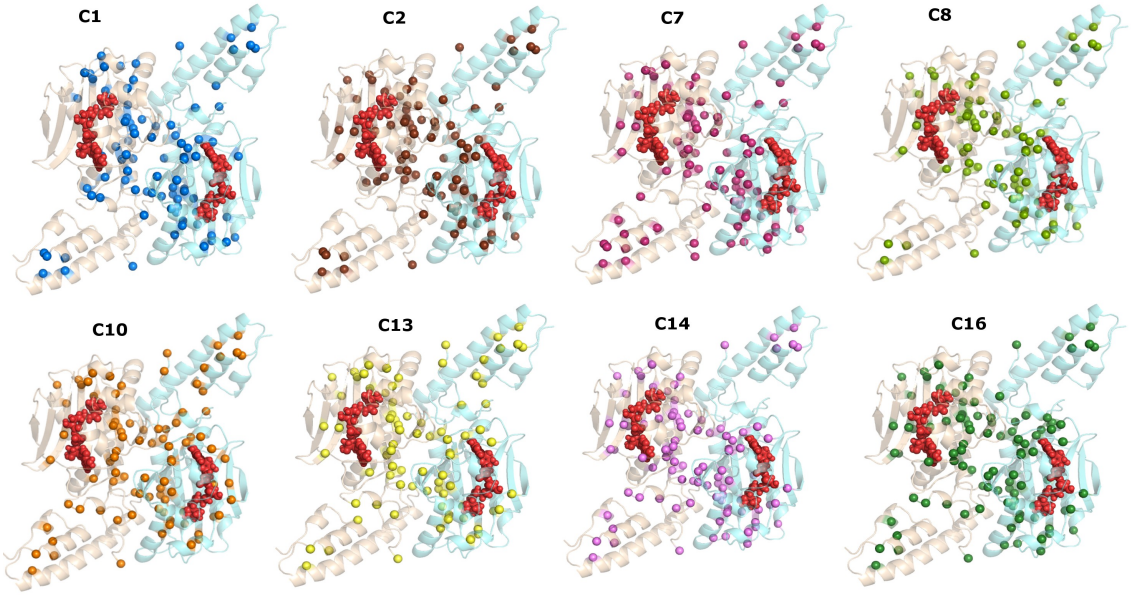


Figure S13(b)

This journal is © The Royal Society of Chemistry 2011

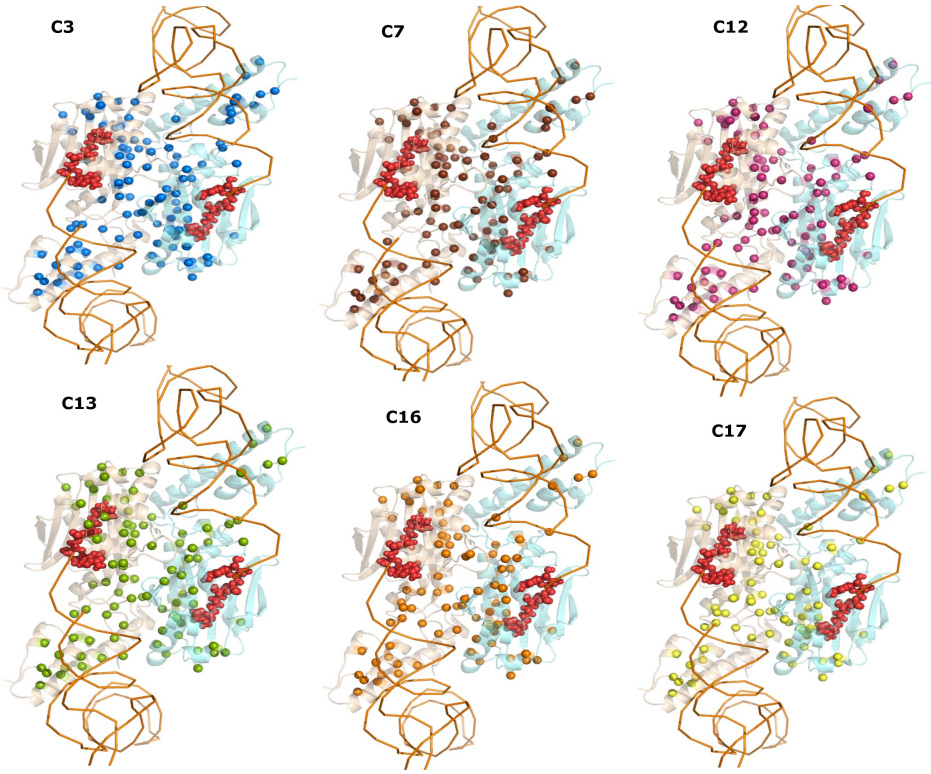


Figure S13(c)

Figure S13: Pictorial representation of the dynamically stable cliques for the conformations corresponding to a particular cluster (cluster numbers marked in the figure) in (a) Sys1, (b) Sys2, and (c) Sys3 respectively as van der Waals' spheres (only Cα atoms are shown for clarity). The major low energy conformational clusters (conformational clusters with > 5% contribution to total population) in each system are represented here. The protein is depicted as new cartoon with the two subunits being coloured differently and the ligand Pyl-AMP is represented as red van der Waals' spheres. The $tRNA^{Pyl}$ is depicted as blue ribbons.

This journal is © The Royal Society of Chemistry 2011

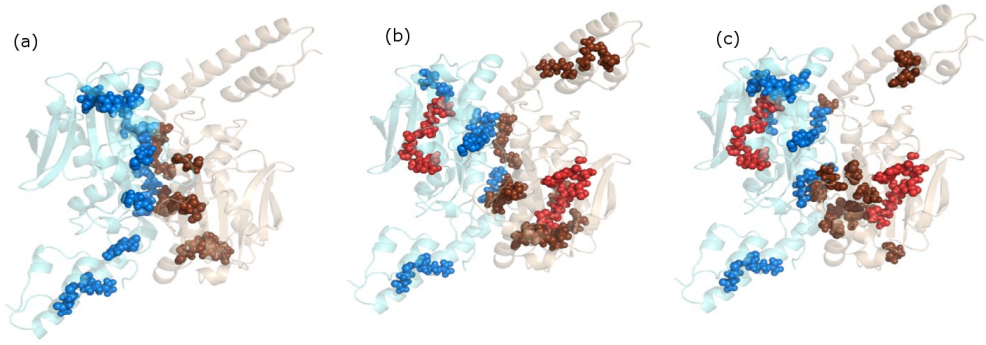


Figure S14: Hubs that are part of cliques are pictorially represented for (a) Sys1, (b) Sys2, and (c) Sys3 respectively as van der Waals' spheres (blue for one subunit, brown for another). It can be noted that there is considerable difference between such residues for the protein DhPylRS under different states of ligation. An exhaustive list for the three systems is given in Table S2. The protein is depicted as new cartoon with the two subunits being coloured differently and the ligand Pyl-AMP is represented as red van der Waals' spheres.

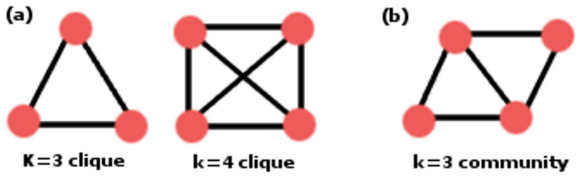


Figure S15: (a) Schematic description of k=3/k=4 cliques with 3/4 nodes completely connected to each other and (b) k=3 community formed by two k=3 cliques sharing k-1 edges respectively.

This journal is © The Royal Society of Chemistry 2011

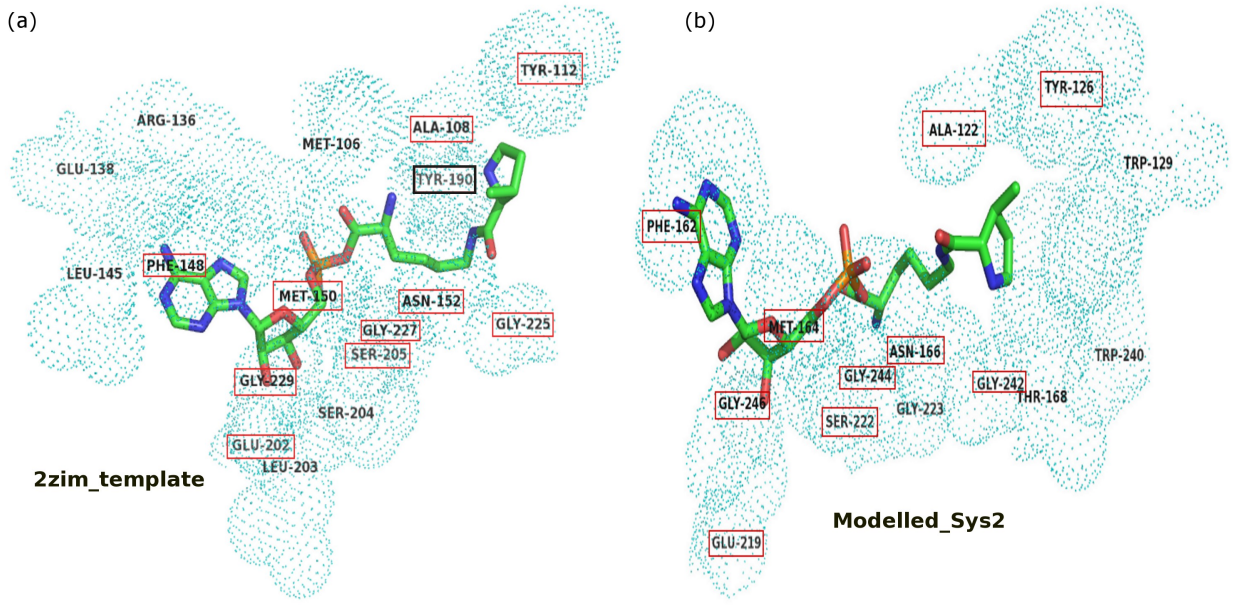


Figure S16: Residues within 3.5 Å of YLY in the (a) template and (b) modelled structure. Strikingly, majority of the intercating residues (highlighted as red rectangles) are retained in the vicinity of YLY in almost identical positions in the equilibrated modelled structure. Tyr-190 is also included in the modelled structure by a slight increase in the cut-off radius to 3.8Å around the ligand. The residue numbering in this figure is done by renumbering the residues in the corresponding PDB structures (2zim and 2zni:chain C) from 1.

Electronic Supplementary Material (ESI) for Molecular BioSystems
This journal is © The Royal Society of Chemistry 2011

Supplementary Methods

Modeling of the ligand bound states: Sys2 and Sys3

The crystal structure of native DhPylRS and DhPylRS in complex with $tRNA^{Pyl}$ are available (PDB id: 2znj and 2zni respectively)[1] but the DhPylRS structure bound to both $tRNA^{Pyl}$ and Pyl-AMP is not available. However, the structure of *M.mazei* pyrrolysyl-tRNA synthetase bound to Pyl-AMP has been solved by X-ray crystallography (PDB id: 2zim)[2]. So the DhPylRS+Pyl-AMP (Sys2) complex is modelled based on the crystal structure with PDB id: 2zim using the closed form asymmetric subunit from PDB id: 2znj as our template. All-atom RMSD between the two structures (2zim and the closed form of 2znj) is ~1.3Å with very little variations in the catalytic domain. Also the all-atom RMSD between 2zni and 2znj closed subunit is only ~0.8Å. We further analyzed the ligand binding pocket (residues within 3.5Å of YLY) for the template and the equilibrated modelled structures. Strikingly, there is a significant overlap between the residues around the ligand [Figure S16], validating our modelling exercise. DhPylRS+Pyl-AMP+tRNA (Sys3) is constructed by adding tRNA from PDB id: 2zni to Sys2. The parameters for Pyl-AMP are generated using the antechamber module in AMBER 9[3]. The electrostatic charges of the units are evaluated using quantum mechanical calculation [using *Gaussian* 03[4]] at the Hartree-Fork level of theory with 6-31G* basis set and using the RESP fit procedure.

Optimization of σ using BIC score

The method of quantum clustering (QC) models the clustering problem in an efficient manner by associating each data point with a Gaussian wave function and employing the Schrödinger equation to obtain potential weighted minima in the data space. The data points are further binned into each of these minima to obtain the clusters. The method has been described in detail elsewhere[5, 6].

The only variable parameter in the QC method is σ, the width of the Gaussian at half minimum. The higher the σ values, the fewer the clusters detected by the algorithm. In order to maximize objectivity, Varshavsky *et al*. proposed the use of an internal criterion known as Bayesian Information (BIC)[7]. It is a model based analysis assuming the data to be generated by a mixture of underlying normal probability distributions. BIC has been used for the purpose of parameter

This journal is © The Royal Society of Chemistry 2011

(σ) tuning to optimize the quality and objectivity of clustering using QC method. BIC is defined as follows:

$$BIC \equiv 2l_M(x,\hat{\Theta}) - m_M \log(N) \approx 2\log p(x \mid M) + const$$

Where $l_M(x,\Theta)$ is the mixture log likelihood (of the data x and the predicted model Θ) which is maximized under the condition that m_M (a function of the number of independent parameters) is minimized. To obtain the best clustering parameter, Varshavsky *et al.* suggested the use of this BIC score[7]. The parameter σ is varied from 0.1 to 0.9 and the optimal value of σ for quantum clustering is computed.

Comparison of Quantum clustering with other clustering algorithms

A rigorous comparison of quantum clustering method against six other clustering algorithms (means, hierarchical, average-linkage, complete, Bayesian, and SOM) is shown in Figure S9-Figure S11 for small subsets of MD data for Sys2 (five 1 ns windows between 10 and 20 ns). The σ values for quantum clustering of each small dataset ($Sys2^{test}$) are optimized using the BIC score. The clustering at σ_{opt} yielded 4/3 clusters for the windows 11-12ns/13-14, 14-15, 15-16, 16-17 ns respectively. The cluster count was fixed at four and three while running the other six algorithms using the PTRAJ module in AmberTools1.2 for the corresponding test windows from Sys2. It is evident from the figures that each clustering algorithm gave a different clustering pattern for the data ($Sys2^{test}$) with the exception of QC and means and/or average linkage where the clustering pattern was similar to each other for a given window. As a consequence of this similarity over majority of test windows, we further choose means algorithm for an extensive comparison with the quantum clustering method for the trajectories obtained from Sys1-3.

Network parameters associated with high connection: Cliques/communities and Hubs

In general network terminology, the parameters cliques/communities represent tightly connected regions of the network [8] and hubs represent highly connected points. In the context of PSN, these parameters are used to identify the rigid regions in the protein structures and to recognize the changes that take place due to the binding of ligands[9].

This journal is © The Royal Society of Chemistry 2011

k-clique: A k-clique is defined as a set of k nodes (points represented by amino acids) in which each node is connected to all the other nodes.[8] Figure S14(a) schematically shows a k=3/k=4 clique in which all the three/four nodes are connected to each other.

k-clique community: According to mathematical literature, a k-clique community has been defined as the assemblage of k-cliques that can be percolated through a series of adjacent k-cliques.[8] In the present study, a k-clique community is one in which two k-cliques share k-1 nodes. Figure S14(b) schematically shows a k=3 community where two k=3 cliques share an edge (i.e., 3-1=2nodes).

k-Clique community finding algorithm: The community search approach employed by us is based on the algorithm proposed by Palla *et al* [8]. We have used Cfinder[10] to obtain the k-clique community from PSNs. In majority of the cases we obtain k=3 cliques (with a few exception of k=4 cliques) at the chosen I_{min} =3.9%. k-clique communities with an overlap of k-1 nodes are obtained using Cfinder. The communities with k-2 node overlaps are obtained by manual inspection of the cliques and communities.

Hubs: Hubs are defined as nodes connected by four or more edges to its neighbours in the PSN at the chosen value of I_{min}.

Dynamically Stable Cliques, Communities, and Hubs: The network parameters are considered to be dynamically stable if they are present in more than 70% of the simulation snapshots within a cluster. The two-dimensional representation of these dynamically stable cliques and communities are drawn using Cfinder[10].

References

1. K. Nozawa, P. O'Donoghue, S. Gundllapalli, Y. Araiso, R. Ishitani, T. Umehara, D. Soll and O. Nureki, *Nature,* 2009, **457,** 1163-1167.
2. J. M. Kavran, S. Gundllapalli, P. O'Donoghue, M. Englert, D. Soll and T. A. Steitz, *Proc. Natl. Acad. Sci. USA,* 2007, **104,** 11268-11273.
3. D. A. Case, T. A. Darden, T. E. Cheatham III, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, K. M. Merz, D. A. Pearlman, M. Crowley, R. C. Walker, W. Zhang, B. Wang, S. Hayik, A. Roitberg, G. Seabra, K. F. Wong, F. Paesani, X. Wu, S. Brozell, V. Tsui, H. Gohlke, L. Yang, C. Tan, J. Mongan, V. Hornak, G. Cui, P. Beroza, D. H. Mathews, C. Schafmeister, W. S. Ross and P. A. Kollman *AMBER9*, University of California, San Fransisco: 2006.

This journal is © The Royal Society of Chemistry 2011

4. M. J. T. Frisch, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, Jr., J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; and Pople, J. A. *Gaussian 03, Revision C.02*, Gaussian, Inc., Wallingford CT: 2004.
5. D. Horn and A. Gottlieb, *Phys Rev Lett,* 2002, **88,** 18702.
6. D. Horn and I. Axel, *Bioinformatics,* 2003, **19,** 1110-1115.
7. R. Varshavsky, D. Horn and M. Linial, Clustering Algorithms Optimizer: A Framework for Large Datasets. In *Bioinformatics Research and Applications*, 2007; pp 85-96.
8. G. Palla, I. Derenyi, I. Farkas and T. Vicsek, *Nature,* 2005, **435,** 814-818.
9. A. Ghosh and S. Vishveshwara, *Biochemistry,* 2008, **47,** 11398-11407.
10. B. Adamcsek, G. Palla, I. J. Farkas, I. Derenyi and T. Vicsek, *Bioinformatics,* 2006, **22,** 1021-1023.

This journal is © The Royal Society of Chemistry 2011

Supplementary Tables

Table S1: Optimized σ values for Sys1-3

System	No. of Iterations	Optimized σ value
Sys1	50	0.3390
Sys2	50	0.3220
Sys3	50	0.3050

Table S2: Comparison of hub residues participating in cliques for Sys1-3

Sys1	***Sys2***	***Sys3***
	14W*	
		17V*
18Q	*18Q 18Q**	*18Q*
		*20Q**
21R	*21R 21R**	*21R*
		69H*
70R 70R*	70R 70R*	70R
78E 78E*	78E*	
85H 85H*	85H	
		90V*
91Q 91Q*	91Q 91Q*	91Q
		93V*
	94T*	
		127L*
128R 128R*	128R 128R*	128R
131L*		
		130M*
		135L*
136Y	136Y*	136Y
		138L*

This journal is © The Royal Society of Chemistry 2011

129W 129W*	129W*	129W
	154E 154E*	
		158C*
171E 171E*		
		178T*
179E 179E*	179E 179E*	179E
182T*	182T	
	190R*	
	197W*	197W
		249P*
250W	250W*	250W
		275R
279Y		

The hub residues that are part of cliques are depicted for Sys1-3. The ones that are completely conserved/ partially conserved/exhibits semi-conservative mutation are highlighted as green/blue/pink respectively. The residues in this table which directly interact with tRNA are given in italics.