Supplemental Information for the paper:

The interactions and recognition of cyclic peptide mimetics of Tat with HIV-1 TAR RNA: a molecular dynamics simulation study

Chun Hua Li ^{a,*}, Zhi Cheng Zuo ^a, Ji Guo Su ^b, Xian Jin Xu ^a, Cun Xin Wang ^{a,*}

^a College of Life Science and Bioengineering, Beijing University of Technology, Beijing 100124, China

^b College of Science, Yanshan University, Qinhuangdao 066004, China

Details of binding free energy of calculation

In this paper, we applied the “separate trajectory protocol” (Kollman et al., 2000; Reyes & Kollman, 2000) to the estimation of the bind free energy. This computational strategy entails three simulation trajectories involving the complex and its isolated partners. Here, a set of 3000 snapshots extracted at 10 ps intervals from the last 30 ns of each MD simulations were used to compute the binding free energy (ΔG) between the cyclic peptide and TAR RNA according to the following equation (Srinivasan et al., 1998; Massova & Kollman, 1999):

$$\Delta G_{binding} = \overline{G}_{complex} - \overline{G}_{peptide} - \overline{G}_{RNA} \quad [1]$$

where $\overline{G}$ is the (absolute) average free energy calculated as the sum of the molecular mechanics (MM) energy (i.e. gas phase energy), solvation energy and solute entropy.

$\overline{G}$ can be formulized as follows:

$$\overline{G} = \overline{E}_{MM} + \overline{G}_{sol} - T\overline{S} \quad [2]$$

where $\overline{E}_{MM}$ is the sum of the van der Waals energy, electrostatic energy and internal bonded energy (bond, angle, dihedral energies). $\overline{E}_{MM}$ is expressed by the following equation:

$$\overline{E}_{MM} = \overline{E}_{vdW} + \overline{E}_{ELE} + \overline{E}_{INT} \quad [3]$$

The solvation energy, $\overline{G}_{sol}$, is approximated as the sum of the polar and nonpolar solvation contributions using a continuum solvation model:

$$\overline{G}_{sol} = \overline{G}_{polar} + \overline{G}_{nonpolar} \quad [4]$$

The above polar electrostatic contribution to the solvation energy was calculated through the finite-difference Poisson-Boltzmann (FDPB) and Generalized Born (GB) solvation models. The solute and the surrounding solvent were assigned a dielectric constant of 1 and 80, respectively. Here, we did not taken the ionic effects into account for the calculations of both PB and GB. As to PB model, the grid resolution of 0.5 Å per grid point was applied and 1000 linear iterations were required for energy convergence. Concerning GB model, the IGB=2 option was passed to *sander* program of AMBER. The nonpolar hydrophobic contribution was approximated to be proportional to the solvent-accessible surface area (SASA):

$$\overline{G}_{nonpolar} = \gamma * \overline{SASA} + \beta \quad [5]$$

The *SASA* was calculated using the linear combination of pairwise overlays (LCOP) method (Weiser, Shenkin & Still, 1999). The coefficients were set as follows: $\gamma = 0.00542 \text{ kcal} \cdot \text{\AA}^{-2} \cdot \text{mol}^{-1}$ and $\beta = 0.92 \text{ kcal/mol}$. The probe radius was set to 1.4 Å to define the molecular surface.

The normal-mode analysis was performed to evaluate the conformational entropy change using the *nmode* program in AMBER10. Due to the high computational demand, only 300 snapshots were used for entropy estimation. Snapshots were collected at equally spaced intervals (100 ps) of the last 30 ns equilibration trajectories from each simulation system. Prior to the normal analysis, thorough minimizations were carried out with a distance-dependent dielectric constant ($\epsilon = 4R_{ij}$, R_{ij} is the distance between two atoms) including all non-bonded interactions until the root-mean-square of the elements of gradient vector was less than $5 \times 10^{-4} \text{ kcal} \cdot \text{mol}^{-1} \cdot \text{\AA}^{-1}$.

Table SI Statistics of TAR RNA nucleotides interacting with the peptide A-ends and P-ends with an occupancy > 99 % computed from the complex trajectories

		^a TAR RNA				^d Total
		Bulge	Apical loop	Upper stem	Lower stem	
^b L-22	A-end	U23, C24, U25	C30, U31, G32, G33, G34, A35	G26, A27, G28, C29, G36, C37	/	15
	P-end	/	A35	G36, C37	G21, A22	5
^c KP-Z-41	A-end	U23, C24	G32, G33, G34, A35	G26, A27, G28, C29, G36, C37, C39	G21, A22, U40	16
	P-end	/	/	G36, C37	G18, C19, A20, G21, A22	7

The contact nucleotide is the one that has at least one atom within 4 Å of any atom of the peptide.

^a TAR RNA can be grouped into four secondary structure elements. See details in Figure 1.

^b For L-22, A-end is comprised of T4, R5, K6, G7, R8, R9; P-end: R1, V2, R11, I12, ^DP13, P14.

^c For KP-Z-41, A-end: R5, Q6, R7, K8, G9, R10, R11, I12; P-end: R1, V2, R3, I14, R15, I16, ^DP17, P18.

^d Total number of TAR RNA nucleotides contacting with the A-end or P-end of cyclic peptide.

Figure S1: The representative TAR local conformations around the bulge respectively from the simulations for L-22/TAR complex and its apo-TAR (A) and for KP-Z-41/TAR complex and its apo-TAR (B). The black dotted lines indicate the base pairing hydrogen bonds. The bulge (U23 to U25) backbone is rendered in red. The establishment of base tripe (blue) leads to a noticeable twist in the bulge backbone, facilitating the base pairing of A22-U40 (orange) and G26-C39 (green).

Figure S2: Electrostatic potential surface for L-22 (A) and KP-Z-41 (B) viewed from the binding interface. Basic residues are labeled by corresponding texts. The negative electrostatic potential is represented in red (-5 kT), the neutral in white and the positive in blue (+5 kT). The Adaptive Poisson-Boltzmann Solver (APBS) plug-in within VMD software (Humphrey, Dalke & Schulten, 1996; Baker et al., 2001) was utilized to perform the electrostatic potential calculations. The temperature and ionic strength of the simulated solution were set to 298 K and 150 mM, respectively. Other parameters were set as their default values in APBS. These two figures are prepared using the two lowest-energy complex structures from their respective MD trajectories.

Supplementary References

- Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. (2001). Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci U S A*, 98, 10037-10041.
- Humphrey, W., Dalke, A. & Schulten, K. (1996). VMD: visual molecular dynamics. *J Mol Graph*, 14, 33-38, 27-38.
- Kollman, P. A., Massova, I., Reyes, C., Kuhn, B., Huo, S., Chong, L., Lee, M., Lee, T., Duan, Y., Wang, W., Donini, O., Cieplak, P., Srinivasan, J., Case, D. A. & Cheatham, T. E., 3rd (2000). Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc Chem Res*, 33, 889-897.

- Massova, I. & Kollman, P. A. (1999). Computational alanine scanning to probe protein-protein interactions: A novel approach to evaluate binding free energies. *Journal of the American Chemical Society*, 121, 8133-8143.
- Reyes, C. M. & Kollman, P. A. (2000). Structure and thermodynamics of RNA-protein binding: using molecular dynamics and free energy analyses to calculate the free energies of binding and conformational change. *Journal of Molecular Biology*, 297, 1145-1158.
- Srinivasan, J., Cheatham, T. E., Cieplak, P., Kollman, P. A. & Case, D. A. (1998). Continuum solvent studies of the stability of DNA, RNA, and phosphoramidate - DNA helices. *Journal of the American Chemical Society*, 120, 9401-9409.
- Weiser, J., Shenkin, P. S. & Still, W. C. (1999). Approximate atomic surfaces from linear combinations of pairwise overlaps (LCPO). *Journal of Computational Chemistry*, 20, 217-230.