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

Chun Hua Li*, Zhi Cheng Zuo, Ji Guo Su, Xian Jin Xu and Cun Xin Wang*

aCollege of Life Science and Bioengineering, Beijing University of Technology, Pingleyuan 100, Chaoyang District, Beijing 100124, China; bCollege of Science, Yanshan University, Hebei Avenue 438, Qinhuangdao 066004, Hebei, China

Communicated by Ramaswamy H. Sarma

(Received 8 February 2012; final version received 11 April 2012)

The interaction of HIV-1 trans-activator protein Tat with its cognate trans-activation response element (TAR) RNA is critical for viral transcription and replication. Therefore, it has long been considered as an attractive target for the development of antiviral compounds. Recently, the conformationally constrained cyclic peptide mimetics of Tat have been tested to be a promising family of lead peptides. Here, we focused on two representative cyclic peptides termed as L-22 and KP-Z-41, both of which exhibit excellent inhibitory potency against Tat and TAR interaction. By means of molecular dynamics simulations, we obtained a detailed picture of the interactions between them and HIV-1 TAR RNA. In results, it is found that the binding modes of the two cyclic peptides to TAR RNA are almost identical at or near the bulge regions, whereas the binding interfaces at the apical loop exhibit large conformational heterogeneity. In addition, it is revealed that electrostatic interaction energy contributes much more to KP-Z-41 complex formation than to L-22 complex, which is the main source of energy that results in a higher binding affinity of KP-Z-41 over-22 for TAR RNA. Furthermore, we identified a conserved motif RRK (Arg-Arg-Lys) that is shown to be essential for specific binding of this class of cyclic peptides to TAR RNA. This work can provide a useful insight into the design and modification of cyclic peptide inhibitors targeting the association of HIV-1 Tat and TAR RNA.

Keywords: HIV-1 TAR RNA; cyclic peptide; molecular dynamics simulation; binding affinity; binding specificity

Introduction

The interaction of HIV-1 trans-activator Tat protein with its cognate trans-activation response element (TAR) RNA plays a central role for viral transcription and replication (Biswas et al., 1995; Gatignol et al., 1991). HIV-1 TAR RNA, located in the 5′-untranslated region (UTR) of nascent viral transcripts, is characterized by two helical stems separated by a tri-nucleotide bulge and a hexa-nucleotide apical loop capping the terminal stem (Aboulela, Karn, & Varani, 1995; Ratnasabapathy et al., 1990) (Figure 1(A)). The association of Tat with TAR RNA triggers the recruitment of the essential cellular cofactor transcription elongation factor b (P-TEFb) to RNA polymerase II (Karn & Stoltzfus, 2012; Peterlin & Price, 2006). Consequently, the progressive transcription elongation is promoted, substantially facilitating viral gene expression and virion production.

Interference with the binding of Tat to TAR RNA has been proved to be an especially attractive strategy for combating HIV infection. A variety of small molecule compounds (Cupelli & Hsu, 1995; Davis et al., 2004; Hwang et al., 2003; Mayer et al., 2006; Murchie et al., 2004; Puglisi et al., 1992; Raghunathan et al., 2006), antisense nucleotides (Darfeuille et al., 2004; Hamma et al., 2003; Kiviniemi and Virta, 2011; Toulme, Di Primo, & Moreau, 2001), and linear polypeptide analogs (Burns et al., 2008; Hamy et al., 1997, 1998; Huq, Wang & Rana, 1997; Lee et al., 2005; Tamilarasu, Huq & Rana, 2001) have been evaluated for Tat/TAR-interaction-blocking activity during the last two decades, but the vast majority of them have no sufficient potency and/or selectivity to warrant pharmaceutical development (Davidson et al., 2009). Recently, conformationally constrained peptidomimetics have caused widespread concern among researchers (Robinson, 2008). In particular, those cyclic beta-hairpin peptidomimetics of Tat protein have been successfully designed to prevent Tat–TAR association toward bovine immunodeficiency virus (BIV) (Athanassiou et al., 2004; Leeper et al., 2005) and HIV (Davidson et al., 2009, 2011; Lalonde et al., 2011;
Moehle et al., 2007). Additionally, this family of beta-hairpin peptidomimetics bears the virtue of resistance to proteolytic degradation (Moehle et al., 2007). Therefore, the cyclic peptide mimetics of Tat have become a class of promising inhibitors against HIV infection.

A number of studies have focused on TAR RNA and Tat molecule systems through biochemical techniques. At the same time, the theoretical approaches provided significant information as well. Molecular dynamics (MD) simulations have been successfully applied to the interpretation and prediction of structure–function relationships of biological macromolecules (Chen and Xiao, 2006; Cao and Wang, 2010; Su et al., 2011; Xu et al., 2011; Zhang and Liu, 2011). To our knowledge, Nifosi, Reyes, and Kollman (2000) first employed MD simulations to study the dynamic behavior of HIV-1 TAR RNA and its complex with arginamide. Subsequently, Reyes et al. (2001) investigated the differences of the structural dynamics and binding specificity between the two HIV TAR RNA complexes with Tat fragments of different lengths. Kulinski, Kulinska, Sarzynska, and Adamiak (2005) studied HIV-1 TAR RNA with the extensive MD simulation and confirmed the coexistence of alternative functional conformations of the apical loop. Mu and Stock (2006) presented a unique “minor groove” binding mode of a tripeptide to HIV-1 TAR RNA by combining docking and MD simulations. Musselman et al. (2007) elaborated the dynamical correlations within HIV-1 TAR RNA utilizing an isotropic reorientational eigenmode dynamics (iRED) analysis.

Apparently, these researches based on MD simulations concentrated on either the free TAR RNA or the TAR RNA associated with linear peptides or small molecules. With respect to the HIV-1 TAR RNA complexed with the aforementioned cyclic beta-hairpin peptidomimetics of Tat (cyclic peptide for short in the following parts), the relationships between the properties of structure, dynamics, and energetics and the high affinity and specificity recognition have not yet been clearly understood. In this work, we chose two representative cyclic peptides (Figure 1B and Table 1), termed L-22 and KP-Z-41, respectively, as case studies (Davidson et al., 2009; Davidson, Patora-Komisarska, Robinson, & Varani, 2011). These two peptides were selected for their potent inhibitory activities and available complex structures with TAR RNA. Here, we delineated a detailed picture of the interactions between them and HIV-1 TAR RNA via all-atom MD simulations with explicit water solvent. Based on the two sets of simulation runs, we revealed the differences in binding mode and affinity of these two cyclic peptides to TAR RNA. Moreover, a conserved motif RRK (Arg-Arg-Lys) was identified which plays a central role in the high affinity and specificity recognition of TAR RNA by this class of cyclic peptide inhibitors.
Materials and methods

Molecular dynamics simulations

The two initial complex structures (Figure 1(C) and (D)) were derived from Protein Data Bank (PDB entries: 2KDQ for L-22/TAR RNA complex (Davidson et al., 2009) and 2KX5 for KP-Z-41/TAR RNA complex (Davidson, Patora-Komisarska, Robinson, & Varani, 2011). Each complex system contains an ensemble of 10 structure models resolved by nuclear magnetic resonance spectroscopy (NMR), in which the structure with the minimum cumulative root-mean-square displacement (RMSD) from the rest of the ensemble was used as the starting coordinate. In addition, the respective monomer coordinates were obtained by splitting the selected complex models. For each type of peptide–RNA complex, three separate MD simulations involving the complex, receptor TAR RNA, and ligand cyclic peptide were conducted independently. The two sets of simulations were setup based on the same procedures as follows.

The AMBER10 program package (Case et al., 2005) and the force field parm99SB (Perez et al., 2007) were chosen for simulations. Every structure was immersed in a truncated octahedral box filled with the TIP3P water molecules, keeping a minimum distance of 10 Å between the solute and each face of the box. The proper type and number of counterions (sodium or chloride ions) were added to neutralize the charged solute molecule(s). The solvated systems were subjected to a thorough energy minimization and then were gradually heated to 300 K with weak harmonic constraints over a period of 1 ns. The production run without constraints was extended up to 40 ns at constant temperature (300 K) and constant pressure (1 atm) with an integration time step of 2 fs. The trajectory of production phase was saved every 1 ps, thus yielding 40,000 snapshots for each system. The long-range electrostatic interactions were treated with particle mesh Ewald method (Darden, York, & Pedersen, 1993), while covalent bonds involving hydrogen atoms were constrained by the SHAKE algorithm (Ryckaert, Ciccotti, & Berendsen, 1977).

Binding free energy calculation and free energy decomposition

The binding free energy calculations were carried out based on molecular mechanics Poisson–Boltzmann surface area (MM-PBSA) and molecular mechanics generalized Born (MM-GBSA) approaches, using the pbsa program as implemented in AMBER10. In principle, the binding free energy can be evaluated on the basis of either the “single trajectory protocol” or the “separate trajectory protocol” (Reyes & Kollman, 2000). Compared with the former protocol, the latter is computationally more expensive, but it is more accurate due to the conformational energy of adaption taken into account explicitly (Reyes & Kollman, 2000). In this study, the separate trajectory protocol was adopted in view of the high flexibility of RNA. The binding free energies of cyclic peptides to TAR RNA were estimated according to the following equation:

$$\Delta G_{\text{binding}} = \bar{G}_{\text{complex}} - \bar{G}_{\text{peptide}} - \bar{G}_{\text{RNA}}$$

where the absolute free energies of the complex ($\bar{G}_{\text{complex}}$), peptide ($\bar{G}_{\text{peptide}}$), and RNA ($\bar{G}_{\text{RNA}}$) were averaged over 3000 snapshots taken at 10 ps intervals from the last 30 ns equilibration MD trajectories. The contributions of conformational entropy were also considered through the normal mode analysis. More details about the binding free energy calculation are included in Supplementary material. In addition, the free energy decomposition on a per-residue basis was performed with the MM-GBSA approach.

Figure 2. Time evolution of the mass-weighted heavy atom RMSDs for L-22 complex (panel A) and KP-Z-41 complex (panel B) in combination with their respective isolated monomers. The prefix apo refers to the ligand-free state.
Results

Assessment of structural stability and flexibility during simulations

Conformational changes of L-22 complex, KP-Z-41 complex, and their respective isolated monomers over the course of simulations were monitored by the mass-weighted heavy atom RMSDs in reference to the starting structures as a function of time. From Figure 2, the RMSD for each system has an updrift during the early MD simulation. It can be partly attributed to the sudden removal of constraints on the solute molecule when the simulation progressed into the production phase from the gradual heating stage (see Materials and methods). Noticeably, those free monomers show a larger fluctuation in RMSD as compared to the two complexes within the first 5 ns of the simulations, which is mainly because the monomer simulations were started from the conformations separated from the complexes. On the whole, all these simulation runs tend to reach a plateau since 10 ns. Hence, all the following analyses were performed on the last 30 ns equilibrated MD trajectories.

To track the extent of variation in flexibility of individual residues/nucleotides upon binding, the per-residue/nucleotide root-mean-square fluctuations (RMSFs) of heavy atoms were computed over respective simulation runs (see Figure 3). As shown in Figure 3(A) and (B), the overall flexibility of each peptide is reduced upon TAR RNA binding. Rather intriguingly, the fluctuations occurring at the P-end moiety of each bound peptide are much larger than those observed for the corresponding part in the free one. Upon examination, we found that the A-end freedom degrees of both peptides are largely constrained due to extensive contacts with the major groove binding pockets; by contrast, their P-ends deviating from the major groove are relatively more free to oscillate owe to the lack of structural constrains imposed by TAR (see Supplementary material Table SI). This can account for the exceptional fluctuations in the P-end of each bound peptide.

According to the per-nucleotide RMSF plots for apo-TAR RNA (free state) and holo-TAR RNA (bound state) in Figure 3(C) and (D), it can be seen that TAR RNA overall becomes much less flexible upon binding of peptide. For the two regions apical loop and bulge, they exhibit large fluctuations in both free and bound states, which can be expected due to lack of hydrogen bonding constrains within them. Comparing the fluctuations of the two holo-TAR RNAs, three nucleotide sites should be specially pointed out: G33 and A35 in the apical loop and U23 in the bulge. Both G33 and A35 show the opposite fluctuating tendencies, while U23 shares the lowest fluctuations. In the following two sections concerning the analyses of hydrogen bonding and binding interfaces of complexes, we will give an account for the origins of these unusual fluctuations at length.
Comparative analyses of hydrogen bonding between complexes

To reveal similarities and differences in hydrogen bonding pattern, we carried out comparative analyses of the hydrogen bonds (H-bonds) with an occupancy > 45% over both complex trajectories. Table 2 presents all these intermolecular (upper part) and partial intra-RNA H-bonds (lower part). From Table 2, it is found that the guanidinium groups of Arg3 and Arg5 in L-22 make stable H-bonds with the bases of G26 and G28, respectively. Similarly, the stable H-bond pairs also form between the side chains of Arg5 and Arg7 in KP-Z-41 and the bases of G26 and G28, respectively. The H-bonding of G28 to arginine observed here is in agreement with the NMR data (Davidson et al., 2009, 2011). The H-bonds between G26 and arginine were not immediately established in the NMR experiments, but this nucleobase is in proximity to the residue side chain in both NMR ensembles, and analogous H-bonding interactions (G11 and Arg3) are present in the BIV TAR RNA complex with a peptidomimetic BIV-2 (just one residue replaced compared to L-22) (Leeper, Athanassiou, Dias, Robinson, & Varani, 2005). Thus, it is likely that this type of bis-arginine (bis-Arg) among these cyclic peptides is associated with specific nucleotide recognition. Coincidently, both side chains of Thr4 in L-22 and the corresponding Gln6 in KP-Z-41 form H-bonds with the same phosphate oxygen of bulge nucleotide C24. Arg1 in L-22 and Arg3 in KP-Z-41 make contacts with the phosphate of A22 in the lower helix, and the latter is additionally H-bonded to the phosphate of G21. These H-bondings with the nucleotides in the bulge and lower

| Table 2. Comparison of hydrogen bonds for the two complexes from the MD trajectories. |

| L-22 complex trajectory | KP-Z-41 complex trajectory |
|------------------------|-----------------------------|
| **Donor Acceptor**     | **Donor Acceptor**           | **% Occupied**     | **% Occupied**     |
| **Intermolecular H-bonds**a | **Intermolecular H-bonds**a | **Intermolecular H-bonds**a | **Intermolecular H-bonds**a |
| ARG3-NH1-HH12          | RG26-O6                     | ARG5-NH1-HH12      | RG26-O6                     | 95.49 | 95.81 |
| RH2-HH22               | RG26-N7                     | RH2-HH22           | RG26-N7                     | 95.49 | 95.81 |
| NE-HE                  | RU23-O2P                    | RH2-HH22           | RG26-N7                     | 95.49 | 95.81 |
| RH2-HH21               | RU23-O5'                    | RH2-HH22           | RG26-N7                     | 95.49 | 95.81 |
| ARG5-NH2-HH22          | RG28-N7                     | ARG5-NH1-HH12      | RG28-N7                     | 95.49 | 95.81 |
| RH1-HH12               | RG28-N7                     | RH2-HH22           | RG28-N7                     | 95.49 | 95.81 |
| THR4-O2-HH21           | RC24-O2P                    | RH2-HH22           | RG28-N7                     | 95.49 | 95.81 |
| ARG1-NH1-HH12          | RA22-O2P                    | RH2-HH22           | RG28-N7                     | 95.49 | 95.81 |
| RH2-HH22               | RA22-O2P                    | RH2-HH22           | RG28-N7                     | 95.49 | 95.81 |
| ARG9-NH1-HH12          | RG33-O2'                    | ARG5-NH1-HH12      | RG28-N7                     | 95.49 | 95.81 |
| RH8-NH2-HH21           | RC30-O2                     | RH2-HH22           | RG28-N7                     | 95.49 | 95.81 |
| RH1-HH11               | RG34-O6                     | RH2-HH22           | RG28-N7                     | 95.49 | 95.81 |
| NE-HE                  | RC30-O2                     | RH2-HH22           | RG28-N7                     | 95.49 | 95.81 |
| RH1-HH12               | RG33-O5'                    | RH2-HH22           | RG28-N7                     | 95.49 | 95.81 |
| RH2-HH22               | RG33-O5'                    | RH2-HH22           | RG28-N7                     | 95.49 | 95.81 |

| **Intra-RNA H-bonds**b | **Intra-RNA H-bonds**b | **% Occupied**     | **% Occupied**     |
|------------------------|------------------------|-------------------|-------------------|
| RU23-N3-H3             | RA27-N7                 | 99.99             | 99.99             |
| RA27-N6-H62            | RU23-O4                 | 94.72             | 96.32             |
| RA27-N6-H61            | RU23-O4                 | 94.72             | 96.32             |
| RU38-N3-H3             | RA27-N1                 | 90.25             | 93.10             |
| RU40-N3-H3             | RA27-N1                 | 90.25             | 93.10             |
| RA22-N6-H61            | RU40-O4                 | 97.34             | 98.46             |
| RA22-N6-H61            | RU40-O4                 | 97.34             | 98.46             |
| RG26-N2-H21            | RC39-O6                 | 99.99             | 99.99             |
| N1-H1                  | RC39-N3                 | 99.99             | 99.99             |
| RG33-N2-H21            | RC92-O2                 | 74.01             | 100.00            |
| RH1-H1                 | RC39-N3                 | 69.67             | 99.99             |
| RG29-N4-H41            | RG33-O6                 | 69.89             | 99.99             |

Only the H-bonds with an occupancy >45% are considered.

*aUpper panel shows the H-bonds between peptide and RNA.

*bLower lists H-bonds formed by base pairing at or near the bulge.

*cThe unique base pairing of C29 and G33 in L-22 complex (see the text for more details).
stem of TAR RNA accord well with the intermolecular NOE measurements (Davidson et al., 2009, 2011). Arg15 in KP-Z-41, located opposite Arg3, interacts with the phosphate of C37 in the upper helix. It should be noted that the equivalent residue to Arg15 in KP-Z-41 does not exist in L-22 (see Table 1). At the apical loop, Arg9 of L-22 makes an H-bond with the ribose hydroxyl group of G33, while Arg11 of KP-Z-41 forms contacts with phosphate oxygens of G34, which are consistent with the NMR data (Davidson et al., 2009, 2011). In contrast to the lack of any H-bonds mediated by Arg10 in KP-Z-41 complex, the corresponding Arg8 in L-22 complex is able to form five H-bonds with nucleotides C30, G33, and G34 in the apical loop (see below Figure 4(B)), which confirms the observed low flexibility of Arg8 upon binding (Figure 3(A)). As to RNA, the phosphate group of loop nucleotide A35 makes the stable H-bond with the backbone N atom of Cys13 in KP-Z-41 (see below Figure 4(C)), leading into the diminished mobility of A35 observed in KP-Z-41 complex run rather than in L-22 complex run (Figure 3(C) and (D)).

Finally, summing up the H-bonds at interface, we can see for KP-Z-41 complex, five residues Arg5, Gln6, Gly9, Arg11, and Cys13 are involved in H-bonding with RNA apical loop or bulge and for L-22 complex, the residues are Arg3, Thr4, Arg8, and Arg9. It is these extensive H-bonding interactions that collectively drive TAR RNA to wrap around the cyclic peptides.

Additionally, we investigated the H-bonds of base pairs at or around the bulge (see lower part of Table 2). As indicated by the high occupancy, the Hoogsteen H-bond alignment between U23 and A27 (see below Figure 4(C)), leading into the diminished mobility of A35 observed in KP-Z-41 complex run rather than in L-22 complex run (Figure 3(C) and (D)). Finally, summing up the H-bonds at interface, we can see for KP-Z-41 complex, five residues Arg5, Gln6, Gly9, Arg11, and Cys13 are involved in H-bonding with RNA apical loop or bulge and for L-22 complex, the residues are Arg3, Thr4, Arg8, and Arg9. It is these extensive H-bonding interactions that collectively drive TAR RNA to wrap around the cyclic peptides.

Comparison of peptide-RNA binding interfaces

To further dissect the differences between the two peptide-RNA binding interfaces, we performed a structural comparison between the two lowest energy
conformations obtained from the respective complex trajectories. Figure 4(A) illustrates the structure superimposition between TAR RNAs bound to L-22 and KP-Z-41, respectively, with a pairwise backbone RMSD of 3.5 Å (as compared to 4.6 Å computed for the two initial NMR structures). From Figure 4(A), both the RNA structures globally converge well toward each other, including the region at the interhelical junction (marked with rectangle). The most distinct difference between them lies in the apical loop moiety (marked with ellipse). Hence, we focused on the structural organization around the loop (see Figure 4(B) and (C)). As shown in Figure 4(B), Arg8 in L-22 is sandwiched between G32 and G33 to form cation–π interactions while the Arg9 is stacked below G34. Interestingly, the corresponding Arg10 and Arg11 in KP-Z-41 are also involved in analogous stacking interactions with G32 and G34, respectively (Figure 4(C)); however, the orientations of both bases against the stacked residues are distinct from those observed in L-22 complex (Figure 4(B)). In addition, it can be seen in Figure 4(C) that the G33 at the apex of loop is away from the upper helix major groove and most solvent-exposed. This observation shows structural evidence for the large fluctuation of G33 in KP-Z-41/TAR complex (Figure 3(D)). Similarly, the nucleotide A35 in L-22/TAR complex is also exposed to the solution (Figure 4(B)) and thereby exhibiting a high flexibility (Figure 3(C)). With regard to the holo-TAR RNA with L-22, it is the very juxtaposition of Arg8 guanidinium with bases of G34 and C30 via H-bonds as well as the base pairing between G33 and C29 that leads into the apparent distortion of apical loop in a “zigzag” course (Figure 4(B)), which is in contrast with the “bow” shape observed in KP-Z-41 complex (Figure 4(C)). Together with the H-bonding analyses above, it is suggested that the structural arrangements between L-22 complex and KP-Z-41 complex are nearly identical at the interhelical junction, whereas the local conformations at the apical loop exhibit large heterogeneity.

### Binding free energies and residual energy decomposition

To obtain insights into the contributions of different energy terms to the affinity of cyclic peptides for HIV-1 TAR RNA, the binding free energies for L-22/TAR RNA and KP-Z-41/TAR RNA were computed with both MM-PBSA and MM-GBSA methods (for comparison). Table 3 contains various contributions to the binding free energy as calculated using the “separate trajectory protocol” mentioned in Materials and methods. Overall, both GB and PB models yielded qualitatively very similar results. The calculated total binding free energies are remarkably favorable for the formation of both L-22 complex (−33.59 and −46.23 kcal/mol for PB and GB models, respectively) and KP-Z-41 complex (−60.09 and −66.55 kcal/mol, respectively). As a result, KP-Z-41 is predicted to bind to target TAR RNA with a higher affinity than L-22, which is in qualitative agreement with the experimental data (Davidson et al., 2009, 2011).

### Table 3. Binding free energy components for L-22 complex and KP-Z-41 complex.

| Term       | L-22/TAR RNA | Mean | SE  | KP-Z-41/TAR RNA | Mean | SE  |
|------------|--------------|------|-----|-----------------|------|-----|
| ΔELE       | −4672.38     | (3.47) |     | −3311.32        | (3.01) |     |
| ΔVDW       | −69.80       | (0.42) |     | −65.49          | (0.41) |     |
| ΔINT       | 8.01         | (0.72) |     | 9.27            | (0.70) |     |
| ΔSURNP     | −11.63       | (0.04) |     | −12.14          | (0.04) |     |
| ΔPB_{CAL}  | 4655.50      | (3.36) |     | 5259.61         | (2.63) |     |
| ΔPB_{SOL}  | 4643.87      | (3.37) |     | 5247.47         | (2.71) |     |
| ΔPB_{ELE}  | −16.88       | (0.73) |     | −51.71          | (0.73) |     |
| ΔPB_{TOT}  | −90.29       | (0.79) |     | −120.07         | (0.75) |     |
| ΔGB_{CAL}  | 4642.86      | (3.49) |     | 5253.16         | (2.74) |     |
| ΔGB_{SOL}  | 4631.23      | (3.49) |     | 5241.02         | (2.77) |     |
| ΔGB_{ELE}  | −29.52       | (0.55) |     | −58.16          | (0.50) |     |
| ΔGB_{TOT}  | −102.94      | (0.72) |     | −126.52         | (0.68) |     |
| TAS_{TOT}  | −56.71       | (0.26) |     | −59.97          | (0.25) |     |
| ΔG^{GBd}   | −33.59       |       |     | −60.09          |       |     |
| ΔG^{PBd}   | −46.23       |       |     | −66.55          |       |     |

All values are given in kcal/mol.  
*ΔE = E(complex) – E(RNA) – E(peptide).*  
ELE, molecular mechanics electrostatic energy;  
VDW, molecular mechanics van der Waals energy;  
INT, internal strain energy;  
SURNP, nonpolar solvation energy;  
PB_{CAL}, electrostatic solvation energy approximated by PB model;  
PB_{SOL}, electrostatic solvation energy approximated by GB model;  
PB_{ELE}, electrostatic solvation energy approximated by GB model;  
GB_{CAL}, electrostatic solvation energy approximated by GB model;  
GB_{SOL}, electrostatic solvation energy approximated by GB model;  
GB_{ELE}, electrostatic solvation energy approximated by GB model;  
TB_{TOT}, solute entropy.  
Average over 3000 (300 in the case of entropy calculations) snapshots.  
Standard error of mean value in parenthesis.

G^{PBd} = ΔPB_{TOT} – TΔS_{TOT},  
G^{GBd} = ΔGB_{TOT} – TΔS_{TOT}.
As reported in Table 3, the molecular mechanics electrostatic energy ($\Delta ELE$) strongly favors the association of cyclic peptides with TAR RNA, while the electrostatic solvation terms ($\Delta PB_{CAL}$ and $\Delta GB_{CAL}$) significantly oppose complex formation. Both are rather substantial in magnitude but compensate each other, overall resulting in favorable total electrostatic contributions. From PB model, the absolute values of total electrostatic energies ($\Delta PB_{ELE}$) reach to $\sim$17 kcal/mol for L-22/TAR RNA and $\sim$52 kcal/mol for KP-Z-41/TAR RNA; from GB model, the corresponding values ($\Delta GB_{ELE}$) are $\sim$30 and $\sim$58 kcal/mol for these two complexes. The van der Waals interaction ($\Delta VDW$) energy, the major contributor, is favorable to the binding free energy by 65–70 kcal/mol. Also, the nonpolar solvation term ($\Delta SUR_{NP}$) contributes favorably to the binding affinity by 11–12 kcal/mol. The contributions from internal energy ($\Delta INT$), however, are unfavorable by 8–9 kcal/mol, indicating that the conformational changes of the binding partners lead to internal strain upon complex formation. In addition, the changes in the conformation entropy ($-T\Delta S$) disfavor both peptides binding to TAR RNA by up to 57–60 kcal/mol. Apparently, taking the entropic contribution into account (more reasonable) shifts the binding free energy to much more positive values. This is as expected given that the complex formation is accompanied with a considerable loss of conformational flexibility especially for TAR RNA.

Adding together the above terms (marked in italic in Table 3), $\Delta VDW + \Delta INT + \Delta PB_{ELE} + \Delta GB_{ELE} + \Delta SUR_{NP} - T\Delta S_{TOT}$ yield the total gains in free energy of cyclic peptides binding to TAR RNA. Comparing the magnitudes of all these energy terms between the two complexes, we found that $\Delta VDW$, $\Delta INT$, $\Delta SUR_{NP}$, and $T\Delta S_{TOT}$ make comparable contributions to the binding free energies of the two complexes. By contrast, the total electrostatic terms ($\Delta PB_{ELE}$ and $\Delta GB_{ELE}$) differ markedly between complexes, according to both PB and GB models. It is evident that the total electrostatic contribution favors the association of HIV-1 TAR RNA with KP-Z-41 much more than that with L-22. This is the main source of energy that results in a higher affinity of KP-Z-41 over L-22 for target TAR RNA. Figure S2 illustrates the electrostatic potentials mapped on the surfaces of the peptides from their respective lowest energy complex conformations. In comparison with L-22 (Figure S2A), KP-Z-41 (Figure S2B) apparently forms a wider area of positive potential (blue) at the RNA binding interface and this area was examined to match well with the highly electronegative RNA via electrostatic complementarity. The extra positive potential regions at KP-Z-41 are mostly owing to the two distal residues Arg1 and Arg15 at P-end (Figure S2B). These can account for the more favorable electrostatic contribution to KP-Z-41 complex formation as compared to L-22 complex.

In order to identify the key residues important for association, the free energy decomposition on residual basis was performed through MM-GBSA method. As shown in Figure 5, the basic residues of each peptide substantially contribute to the binding to TAR RNA, with an energy value $<-7$ kcal/mol. Overall, the charged residues of KP-Z-41 make almost equivalent contributions (around $-9$ kcal/mol) to the binding free energy, whereas the binding energy per residue differs markedly between charged residues in L-22. In particular, Arg3 and Arg8 in L-22 make the most energetic contributions, almost twice as much as those of other basic residues. It is unsurprising since these two residues participate in the H-bonding network (Table 2) and stacking interactions with adjacent nucleotides (Figure 4(B)). With respect to KP-Z-41, it is a point worth noting that its two P-end residues Arg1 and Arg15 are considerable contributors although relatively away from the RNA phosphate backbone, which is consistent with the observation that the electrostatic force can extend up to 11 Å from the RNA phosphate group (Bahadur, Kannan & Zacharias, 2009; Garcia-Garcia & Draper, 2003; Law et al., 2006). Given the lack of the corresponding residues in L-22, it is indicated that the enhanced electrostatic contribution within KP-Z-41 complex (as mentioned above) can be partly ascribed to the two residues Arg1 and Arg15 in the extended P-end.

Discussion

**Essential role of RRK motif in cyclic peptide/TAR RNA recognition**

With respect to this class of lead peptides, it is interesting to explore whether there exists a certain consensus motif that is important for potent and specific binding to TAR RNA. This is also useful for the design of cyclic peptide inhibitors. On the ground of our MD studies and available experimental data, here, we carefully identified...
a conserved motif Arg-Arg-Lys (RRK, see Figure 6) that is shown to be critical for the high affinity and specificity association with TAR RNA. The residue Lys (Lys6 in L-22 and Lys8 in KP-Z-41, see Figure 6(C) and (F)) at A-end serves to guide the correct location of peptide relative to TAR RNA, while the bis-Arg sandwich (Arg3-Arg5 in L-22 and Arg5-Arg7 in KP-Z-41, see Figure 6(A) and (D)) adjacent to that mentioned Lys chiefly contributes to the specific recognition. This motif is proposed based on the following analyses.

On the one hand, the bis-Arg sandwich is involved in specific base recognition, which can be explained in terms of hydrogen bonding and electrostatic features at the binding interface. As revealed by the H-bond analysis (Table 2), for each complex the bis-Arg side chains of A-end exclusively make stable H-bonds with the nucleobases of G26 and G28, respectively. Other basic residues, however, either make contacts with RNA phosphate backbones or form inconsistent H-bonds with the nucleobases. Moreover, we carried out electrostatic potential calculations on around the bis-Arg sandwich and observed two negative potential pockets above and below the base plane of U23 (Figure 6(B) and (E)), respectively. It is clear that the bis-Arg sandwich helps to establish the base triple by buttressing the conserved nucleotide U23 in the flexible bulge. Additionally, it is noteworthy that the two electrostatic potential pockets are dominantly formed by electronegative groups not from the backbone phosphates, but from the bases in the major groove near the bulge (Figure 6(B) and (E)). In previous studies, the analogous bis-Arg signature is also found important for the binding of Tat-derived nonclosed peptides (Calabro, Daugherty & Frankel, 2005; Puglisi et al., 1995; Ye, Kumar, & Patel, 1995) and other cyclic peptides (Athanassiou et al., 2007; Leeper et al., 2005) to BIV TAR RNA. Taken together, these results demonstrate that the bis-Arg sandwich is of essence in conferring recognition specificity.

On the other hand, the residue Lys at A-end is related to the proper positioning of the cyclic peptide relative to TAR RNA. To our knowledge, all known cyclic peptides (Athanassiou et al., 2007; Davidson et al., 2009, 2011; Leeper et al., 2005) adopt the opposite orientation toward TAR RNA compared to nonclosed peptides (Aboulela et al., 1995; Puglisi, Chen, Blanchard, & Frankel, 1995; Ye et al., 1995). On this point, we explain that the residue Lys at A-end plays a primary role in orienting the correct face of cyclic peptides toward TAR RNA. As illustrated in Figure 6(C) and (F), for each peptide the positively charged terminus of Lys at A-end penetrates into an electrostatic potential pocket composed of the phosphate groups of nucleotides C24 through G28 at the interhelical junction. Thus, the Lys is regarded as a key residue for the initial electrostatic attraction (before binding) and neutralization (upon binding) with TAR RNA, and this kind of role regarding Lys apparently

Figure 6. Arg-Arg-Lys (RRK) motif identified in L-22 complex (upper panel) and in KP-Z-41 complex (lower panel), respectively. (A) bis-Arg sandwich in L-22/TAR RNA. The two Arg residues are shown as blue ball,-stick model, G26 and G28 as green stick/ribbon style, and other nucleotides A22, U23, and A27 in cyan, yellow, and orange, respectively. The black dotted lines indicate specific H-bonds from Table 2. (B) Surface potential map around the two electrostatic pockets as outlined in (A). One is formed between nucleotides A22 and U23 and the electronegative groups of U23 and A22 bases and the phosphate backbone between them constitute the charge cluster. The other is above U23, composed of electronegative groups on U23 (O2, N3, O4), A27 (N6, N7), and G28 (O6, N7). (C) The positively charged side chain of Lys6 at L-22 A-end is oriented toward a negative potential pocket formed by the phosphate groups of U24 through C28. (D), (E), and (F) The Arg5-Arg7-Lys8 combination found in KP-Z-41/TAR RNA complex with the same drawing and coloring schemes as in (A), (B), and (C), respectively. Electrostatic potentials were color-coded on the molecular surfaces from red (−20 kT) to blue (+20 kT). All the figures were prepared using the two lowest energy structures from the respective complex trajectories.
cannot be conferred by those hydrophobic residues at P-end. Moreover, the experiments found that the substitutions of A-end amino acid lysine with some other residues affect the binding affinity to TAR RNA to a different extent (Athanassiou et al., 2007; Davidson et al., 2009, 2011).

**Conclusion and perspective**

In this work, we focused on the two representative cyclic peptide inhibitors namely L-22 and KP-Z-41 and delineated detailed characteristics of the interactions and recognition between them and HIV-1 TAR RNA via MD simulations. It is found that the TAR RNA base triple (U23: A27–U38) near the bulge is well defined over the course of both complex simulation runs. The binding modes of these two peptides to TAR RNA are almost identical at the interhelical junction, whereas the binding interfaces at the apical loop exhibit large conformational heterogeneity. In addition, we revealed that the electrostatic interaction energy contributes much more to KP-Z-41 complex formation than to L-22 complex, which is partly attributed to the two P-end residues Arg1 and Arg15 in KP-Z-41. This is the main source of energy that results in a higher affinity of KP-Z-41 than L-22 for TAR RNA. Furthermore, we identified a conserved motif RRK that is shown to be essential for the high affinity and selectivity recognition of TAR RNA by this class of lead peptides.

The results of this work could provide some implications for the improvement in activity of Tat peptidomimetic inhibitors. The basic RRK combination identified here is crucial for the maintenance of conserved contacts of peptides with TAR RNA; hence, this motif should be exploited for future development of cyclic peptides. Meanwhile, its location on the beta-hairpin scaffold needs to be optimized for desired potency. Another point worth noting is that, compared with the intimate contacts of RRK motif with the area near the bulge, the interactions around the apical loop are relatively weak, as high fluctuations still observed in this region. This implies that the additional binding affinity for TAR could be obtained by introducing the residues with larger side chains (e.g. Trp, Phe, and His) in peptides to form more extensive interactions with RNA loop.

**Acknowledgments**

This work was supported by the Chinese Natural Science Foundation (Grant Nos. 31171267, 10974008 and 109747014), the Beijing Natural Science Foundation (Grant No. 4102006), the Beijing Municipal Education Commission (Grant No. KM201310005030), and the Fundamental Research Fund for Beijing Municipal Education Commission Science and Technology Innovation Platform and International S&T Cooperation Program of China (Grant No. 2010DFA31710).

**Supplementary material**

The supplementary material for this paper is available online at http://dx.doi.10.1080/07391102.2012.698248.

**References**

Aboulela, F., Karn, J., & Varani, G. (1995). The structure of the human-immunodeficiency-virus type-1 TAR RNA reveals principles of RNA recognition by Tat protein. *Journal of Molecular Biology*, 253, 313–332.

Athanassiou, Z., Dias, R.L.A., Moehle, K., Dobson, N., Varani, G., & Robinson, J.A. (2004). Structural mimicry of retroviral Tat proteins by constrained, beta-hairpin peptidomimetics: Ligands with high affinity and selectivity for viral TAR RNA regulatory elements. *Journal of the American Chemical Society*, 126, 6906–6913.

Athanassiou, Z., Patora, K., Dias, R.L.A., Moehle, K., Robinson, J.A., & Varani, G. (2007). Structure-guided peptidomimetic design leads to nanomolar beta-hairpin inhibitors of the Tat-TAR interaction of bovine immunodeficiency Virus. *Biochemistry*, 46, 741–751.

Bahadur, R.P., Kann, S., & Zacharias, M. (2009). Binding of the bacteriophage P22 N-peptide to the boxB RNA motif studied by molecular dynamics simulations. *Biophysical Journal*, 97, 3139–3149.

Biswa, D.K., Salas, T.R., Wang, F.L., Ahlers, C.M., Dezube, B.J., & Pardee, A.B. (1995). A Tat-induced auto-up-regulatory loop for superactivation of the human-immunodeficiency-virus type-1 promoter. *Journal of Virology*, 69, 7437–7444.

Burns, V.A., Bobay, B.G., Basso, A., Cavanagh, J., & Melander, C. (2008). Targeting RNA with cysteine-constrained peptides. *Bioorganic & Medicinal Chemistry Letters*, 18, 565–567.

Calabro, V., Daugherty, M.D., & Frankel, A.D. (2005). JDV tat-BIV TAR-A single intermolecular contact mediates intramolecular stabilization of both RNA and protein. *Proceedings of the National Academy of Sciences of the United States of America*, 102, 6849–6854.

Cao, Z., & Wang, J. (2010). A comparative study of two different force fields on structural and thermodynamics character of H1 peptide via molecular dynamics simulations. *Journal of Biomolecular Structure & Dynamics*, 27, 651–661.

Case, D.A., Cheatham, T.E. 3rd, Darden, T., Gohlke, H., Luo, R., Merz, K.M.Jr., … Woods, R.J. (2005). The Amber biomolecular simulation programs. *Journal of Computational Chemistry*, 26, 1668–1688.

Chen, C., & Xiao, Y. (2006). Molecular dynamics simulations of folding processes of a beta-hairpin in an implicit solvent. *Physical Biology*, 3, 161–171.

Cupelli, L.A., & Hsu, M.C. (1995). The human immunodeficiency virus type 1 Tat antagonist, Ro 5–3335, predominantly inhibits transcription initiation from the viral promoter. *Journal of Virology*, 69, 2640–2643.

Darden, T., York, D., & Pedersen, L. (1993). Particle Mesh Ewald – an N·Log(N) method for Ewald sums in large systems. *Journal of Chemical Physics*, 98, 10089–10092.

Darfeuille, F., Hansen, J.B., Orum, H., Di Primo, C., & Toulmou, J.J. (2004). LNA/DNA chimeric oligomers mimic RNA aptamers targeted to the TAR RNA element of HIV-1. *Nucleic Acids Research*, 32, 3101–3107.

Davidson, A., Leeper, T.C., Athanassiou, Z., Patora-Komisarska, K., Karn, J., Robinson, J.A., & Varani, G. (2009). Simultaneous recognition of HIV-1 TAR RNA bulge and
loop sequences by cyclic peptide mimics of Tat protein. 

Proceedings of the National Academy of Sciences of the United States of America, 106, 11931–11936.

Davidson, A., Patora-Komisarska, K., Robinson, J.A., & Varani, G. (2011). Essential structural requirements for specific recognition of HIV TAR RNA by peptide mimetics of Tat protein. Nucleic Acids Research, 39, 248–256.

Davis, B., Afshar, M., Varani, G., Murchie, A.I.H., Karn, J., Lentzen, G., ... Bou-ellal, F. (2004). Rational design of inhibitors of HIV-1 TAR RNA through the stabilisation of electrostatic “hot spots”. Journal of Molecular Biology, 336, 343–356.

Garcia-Garcia, C., & Draper, D.E. (2003). Electrostatic interactions in a peptide–RNA complex. Journal of Molecular Biology, 331, 75–88.

Hamy, F., Brondani, V., Florsheimer, A., Stark, W., Blommers, M.J.J., & Klimkait, T. (1998). A new class of HIV-1 Tat antagonist acting through Tat-TAR interaction. Biochemistry, 37, 5086–5095.

Hamy, F., Felder, E.R., Heizmann, G., Lazdins, J., AboulEla, F., Varani, G., Karn, J., & Klimkait, T. (1997). An inhibitor of the Tat/TAR RNA interaction that effectively suppresses HIV-1 replication. Proceedings of the National Academy of Sciences of the United States of America, 94, 3548–3553.

Huq, I., Wang, X., & Rana, T.M. (1997). Specific recognition of HIV-1 TAR RNA by a D-Tat peptide. Natural Structural Biology, 4, 881–882.

Hwang, S., Tamlarasu, N., Kibler, K., Cao, H., Ali, A., Ping, Y. H., Jeang, K.T., & Rana, T.M. (2003). Discovery of a small molecule Tat-trans-activation-responsive RNA antagonist that potently inhibits human immunodeficiency virus-1 replication. Journal of Biological Chemistry, 278, 39092–39103.

Karn, J., & Stoltzfus, C.M. (2012). Transcriptional and post-transcriptional regulation of HIV-1 gene expression. Cold Spring Harbor Perspectives in Medicine, 4, a036150.

Kiviniemi, A., & Virta, P. (2011). Characterization of a human TAR RNA-binding protein that activates the HIV-1 LTR. Science, 251, 1597–1600.

Lee, C.W., Cao, H., Ichiyama, K., & Rana, T.M. (2005). Design and synthesis of a novel peptidomimetic inhibitor of HIV-1 Tat-TAR interactions: Squaryli diamide as a new potential bioisostere of unsubstituted guanidine. Bioorganic & Medicinal Chemistry Letters, 15, 4243–4246.

Leeper, T.C., Athanassiou, Z., Dias, R.L.A., Robinson, J.A., & Varani, G. (2005). Cyclic peptide-TAR RNA recognition by a cyclic peptidomimetic of Tat protein. Biochemistry, 44, 12362–12372.

Mayer, M., Lang, P.T., Gerber, S., Madrid, P.B., Pinto, I.G., Guy, R.K., & James, T.L. (2006). Synthesis and testing of a focused phenothiazine library for binding to HIV-1 TAR RNA. Chemistry & Biology, 13, 993–1000.

Mochile, K., Athanassiou, Z., Patora, K., Davidson, A., Varani, G., & Robinson, J.A. (2007). Design of beta-hairpin peptidomimetics that inhibit binding of alpha-helical HIV-1 Rev protein to the rev response element RNA. Angewandte Chemie (International ed. in English), 46, 9101–9104.

Mu, Y.G., & Stock, G. (2006). Conformational dynamics of RNA-peptide binding: A molecular dynamics simulation study. Biophysical Journal, 90, 391–399.

Murchie, A.I.H., Davis, B., Isel, C., Afshar, M., Drysdale, M. J., Bower, J., ... Karn, J. (2004). Structure-based drug design targeting an inactive RNA conformation: Exploiting the flexibility of HIV-1 TAR RNA. Journal of Molecular Biology, 336, 625–638.

Musselman, C., Al-Hashimi, H.M., & Andricioaei, I. (2007). IRED analysis of TAR RNA reveals motional coupling, long-range correlations, and a dynamical hinge. Biophysical Journal, 93, 411–422.

Nifosi, R., Reyes, C.M., & Kollman, P.A. (2000). Molecular dynamics studies of the HIV-1 TAR and its complex with argininamide. Nucleic Acids Research, 28, 4944–4955.

Perez, A., Marchan, I., Svozil, D., Sponer, J., Cheatham, T. E.3rd, Laughton, C.A., & Orozco, M. (2007). Refinement of the AMBER force field for nucleic acids: Improving the description of alpha/gamma conformers. Biophysical Journal, 92, 3817–3829.

Peterlin, B.M., & Price, D.H. (2006). Controlling the elongation phase of transcription with P-TEFb. Molecular Cell, 23, 297–305.

Puglisi, J.D., Chen, L., Blanchard, S., & Frankel, A.D. (1995). Solution structure of a bovine immunodeficiency virus Tat-TAR peptide-RNA complex. Science, 270, 1200–1203.

Raghunathan, D., Sanchez-Pedregal, V.M., Junker, J., Schwieck, C., Kalese, M., Kirschning, A., & Carlonagru, T. (2006). TAR-RNA recognition by a novel cyclic aminoglycoside analogue. Nucleic Acids Research, 34, 3599–3608.

Ratnasabapathy, R., Sheldon, M., Johal, L., & Hernandez, N. (1990). The HIV-1 long terminal repeat contains an unusual element that induces the synthesis of short RNAs from various messenger-Rna and Snr promoters. Genes & Development, 4, 2061–2074.

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.
Reyes, C.M., Nifosi, R., Frankel, A.D., & Kollman, P.A. (2001). Molecular dynamics and binding specificity analysis of the bovine immunodeficiency virus BIV Tat-TAR complex. Biophysical Journal, 80, 2833–2842.
Robinson, J.A. (2008). Beta-hairpin peptidomimetics: Design, structures and biological activities. Accounts of Chemical Research, 41, 1278–1288.
Ryckaert, J.P., Ciccotti, G., & Berendsen, H.J.C. (1977). Numerical-integration of cartesian equations of motion of a system with constraints – molecular-dynamics of N-alkanes. Journal of Computational Physics, 23, 327–341.
Su, J.G., Xu, X.J., Li, C.H., Chen, W.Z., & Wang, C.X. (2011). An analysis of the influence of protein intrinsic dynamical properties on its thermal unfolding behavior. Journal of Biomolecular Structure & Dynamics, 29, 105–121.
Tamilarasu, N., Huq, I., & Rana, T.M. (2001). Targeting RNA with peptidomimetic oligomers in human cells. Bioorganic & Medicinal Chemistry Letters, 11, 505–507.
Toulme, J.J., Di Primo, C., & Moreau, S. (2001). Modulation of RNA function by oligonucleotides recognizing RNA structure. Progress in Nucleic Acid Research and Molecular Biology, 69, 1–46.
Xu, X., Su, J., Chen, W., & Wang, C. (2011). Thermal stability and unfolding pathways of Sso7d and its mutant F31A: Insight from molecular dynamics simulation. Journal of Biomolecular Structure & Dynamics, 28, 717–727.
Ye, X., Kumar, R.A., & Patel, D.J. (1995). Molecular recognition in the bovine immunodeficiency virus Tat peptide-TAR RNA complex. Chemistry & Biology, 2, 827–840.
Zhang, J., & Liu, D.D. (2011). Molecular dynamics studies on the structural stability of wild-type dog prion protein. Journal of Biomolecular Structure & Dynamics, 28, 861–869.