# Structure-Guided Optimization of Small Molecules Inhibiting Human Immunodeficiency Virus 1 Tat Association with the Human Coactivator p300/CREB Binding Protein-Associated Factor

Chongfeng Pan,[†] Mihaly Mezei,[‡] Shiraz Mujtaba,[‡] Michaela Muller,[‡] Lei Zeng,[‡] Jiaming Li,[†] Zhiyong Wang,*[,†] and Ming-Ming Zhou*[,‡]

*Hefei National Laboratory for Physical Science at Microscale and Department of Chemistry, University of Science and Technology of China, Hefei, Anhui 230026, People's Republic of China, and Department of Structural and Chemical Biology, Mount Sinai School of Medicine, New York University, One Gustave L. Levy Place, New York, New York 10029-6574*

*Received January 5, 2007*

**Abstract:** Human immunodeficiency virus 1 (HIV-1) *trans*-activator Tat recruits the human transcriptional coactivator PCAF (p300/CREB binding protein-associated factor) to facilitate transcription of the integrated HIV-1 provirus. We report here structure-based lead optimization of small-molecule inhibitors that block selectively Tat and PCAF association in cells. Our lead optimization was guided by grand-canonical ensemble simulation of the receptor/lead complex that leads to definition of chemical modifications with improved lead affinity through displacing weakly bound water molecules at the ligand−receptor interface.

The fundamental importance of the human immunodeficiency virus (HIV[a]) life cycle in the progression of the AIDS disease in infected subjects is well recognized.[1,2] Because of the development of resistance to the available anti-HIV drugs due to mutations in the targeted viral proteins,[3] continued viral production by chronically infected cells leads to HIV-based immune dysfunction and other disorders, including cancer. Recent studies suggest that new therapeutic strategies that target different steps in the viral life cycle could complement the existing anti-HIV therapies.[4] It has been shown that transcriptional activation of the integrated HIV provirus requires a molecular interaction between lysine 50-acetylated HIV-1 *trans*-activator Tat and the bromodomain (BRD) of the human transcriptional co-activator PCAF (p300/CREB binding protein-associated factor),[5,6] and blocking this interaction leads to a major reduction in Tat-mediated viral transcription.[6] These findings suggest that the Tat/PCAF recruitment could serve as a new therapeutic target for intervening HIV-1 replication. To validate this potential novel anti-HIV therapy target, we seek to develop small-molecule chemical inhibitors that selectively block this molecular interaction, which is essential for viral production. Toward this goal, we have identified from structure-based NMR screen N1-aryl-propane-1,3-diamine compounds (**1** and **2**, Figure 1) for the PCAF BRD binding to HIV-1 Tat.[7] However, these initial lead compounds require optimization to improve their potency for cell-based functional validation of this novel anti-HIV therapy target.

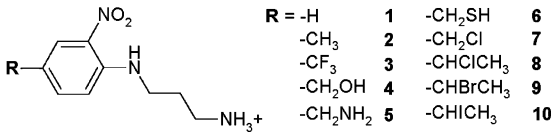


**Figure 1.** Structures of chemical ligands for the PCAF BRD.

Lead optimization often requires not only design and synthesis of appropriate chemical analogs but also estimation of the resulting effects of chemical modifications on target binding. Successful execution of the latter could reduce the burden of synthesis of a large number of chemical analogs for lead improvement. Toward this end, we report in this study a protocol for identifying sites for chemical modifications in a lead that would lead to an enhanced affinity to a receptor by defining water molecules weakly bound at the lead-receptor interface. This strategy involves three steps: (1) identifying water molecules trapped around a receptor-bound lead; (2) defining solvation sites near the ligand with low desolvation penalty; and (3) estimating effects of adding a new chemical group occupying solvation sites on receptor binding. Such effects are of the net change in free energy of binding of a ligand between energy gain in receptor binding upon chemical modifications and energy penalty for ligand desolvation at the point of the attachment.

The estimation of such effects are realized by Monte Carlo simulations of a lead−receptor complex in the grand-canonical (T, V, $\mu$) ensemble (GCE),[8,9] in which a system's temperature and volume but not the number of molecules are kept constant. Besides conventional rotations and translations, the GCE simulation is also performed with periodical insertions and deletions of water molecules. This ensures adequate sampling of isolated cavities that exist between the receptor and the ligand or inside the receptor, which are otherwise difficult to treat by simulations in canonical, microcanonical, or isothermal−isobaric ensembles.[10−13] The "cavity biased" variant attempts insertion of a molecule only if a cavity of an appropriate radius is found.[9,11] Insertion and deletion attempts are accepted with probabilities $P_i$ and $P_d$, respectively

$$P_i = \min\{1, P_N^{\mathrm{cav}} \exp[B + (E(r_{N+1}) - E(r_N))/kT]/(N+1)\}$$

$$P_d = \min\{1, N \exp[-B + (E(r_N) - E(r_{N-1}))/kT]/P_{N-1}^{\mathrm{cav}}\}$$

where $E(r_N)$ is the potential energy of the system of $N$ molecules at configuration $r_N$, $P_N^{\mathrm{cav}}$ is the probability of finding a cavity of specific size, $k$ is the Boltzmann constant, and $T$ is the absolute temperature. The parameter $B$ controls the final density of the system. To match experimental conditions, $B$ is tuned to a value that results in the experimental water density in the outer layers of the simulation cell, using a recently developed automatic tuning method.[14] The water sites around a ligand are calculated by extracting water within 4 Å of a ligand.[15] At each site, orientation of a water molecule is an average of those contributing to that site. Overall, this strategy provides a sensitive probe of local chemical potential in the vicinity of the ligand.

The simulation defines generic solvation sites,[15] mean positions of water molecules around the ligand, with fractional occupancy calculated from the number and position of water at the corresponding sites. Water molecules found trapped at the

* To whom correspondence should be addressed. Phone: 212-659-8652 (M.-M.Z.); +86-551-3603185 (Z.Y.W.) Fax: 212-849-2456 (M.-M.Z.); +86-551-3631760 (Z.Y.W.). E-mail: ming-ming.zhou@mssm.edu (M.-M.Z.); zwang3@ustc.edu.cn (Z.Y.W.).

[†] University of Science and Technology of China.

[‡] Mount Sinai School of Medicine.

[a] Abbreviations: BRD, bromodomain; CBP, CREB-binding protein; GCE, grand-canonical ensemble; HIV, human immunodeficiency; NMR, nuclear magnetic resonance; PCAF, p300/CBP associated factor.

10.1021/jm070014g CCC: $37.00 
Published on Web 04/20/2007

ligand/receptor interface are indicative of sites on the ligand for modifications. Because trapping a water molecule is entropically unfavorable, displacing it by a chemical group may improve lead affinity to a receptor. During simulation, surroundings of a ligand are examined for sites with high probability of successful water deletions. Water sitting at such sites may be displaced by a new functional group tethered to the ligand without incurring significant desolvation penalty. Such spots are thus deemed as good candidate sites for chemical modifications in lead optimization.

Change in binding affinity of a ligand, $\Delta\Delta G^{bind}$, can in principle be calculated as a thermocyle that involves free-energy simulations on the ligand in free and receptor-bound states.[16] However, such simulations require high precision because $\Delta\Delta G^{bind}$ is typically a small value between two simulated quantities. To circumvent this problem, we estimated solvation free energy of a ligand by defining the extent it destabilizes the surrounding water. Once a region of low desolvation penalty is identified in the GCE simulation, an atom X is selected on the ligand in that region, replacement groups are designed, and effects are evaluated.

We tested this approach on a chemical ligand *N*1-(4-methyl-2-nitro-phenyl)-propane-1,3-diamine (**2**) bound to the BRD of the co-activator PCAF.[5] We selected the most representative structure of the 20 final NMR structures[5] for this study, as judged by the lowest root-mean-square deviations. We used the program Simulaid[17] to orient and fit the complex in the smallest enclosing rectangular box to minimize the number of water molecules needed for the simulation. This box was enlarged in all three directions to ensure a water layer of a 15 Å thickness.

We performed GCE simulation of the PCAF BRD/**2** complex using the program MMC[17] and the cavity-biased technique,[9] with insertions only made into voids. We extended preferential sampling to insertions and deletions,[18] thereby ensuring a thorough sampling of regions in the vicinity of the ligand. The simulation on the PCAF BRD/**2** complex led to the identification of five water molecules, three of which are trapped between the protein and the ligand with significant propensity for deletion replacement (W3−W5, Figure 2A,B). Notably, one such water molecule (W3) was found in the acetyl group binding pocket that is filled by the acetylated lysine 50 (K50ac) of Tat in the PCAF BRD/Tat-K50ac peptide complex but unoccupied when bound to the ligand **2**. Linking of an acetyl moiety to **2** with an aliphatic carbon chain has been shown to displace that water molecule in the acetyl binding pocket, which results in a major enhancement in binding affinity to the protein (unpublished results, Z.Y. Wang and M.-M. Zhou).

Besides the protein/ligand interface, the *para*-methyl group on the aniline ring of **2** was shown to be associated with water prone for deletion replacement (Figure 2B). Due to its solvent exposure nature, we reasoned that derivatization at this methyl group would likely replace the weakly bound water without incurring major desolvation penalty. To identify appropriate functional groups for new chemical analogs, we then performed the GCE simulation for several hypothetical derivatives that contain the following functional groups at the *para*-methyl position in **2**: $-CF_3$, $-CH_2OH$, $-CH_2NH_2$, $-CH_2SH$, $-CH_2Cl$, $-CHClCH_3$, $-CHBrCH_3$, and $-CHICH_3$ (compounds **3**−**10**, as listed in Figure 1 and Table 1). The simulations were carried out using corresponding structure models that were built on the template of the PCAF BRD/**2** structure. During the calculations, the torsion angles in the different derivatives and the orientation of the methyl group were sampled, while the rest of the structure was held rigid.

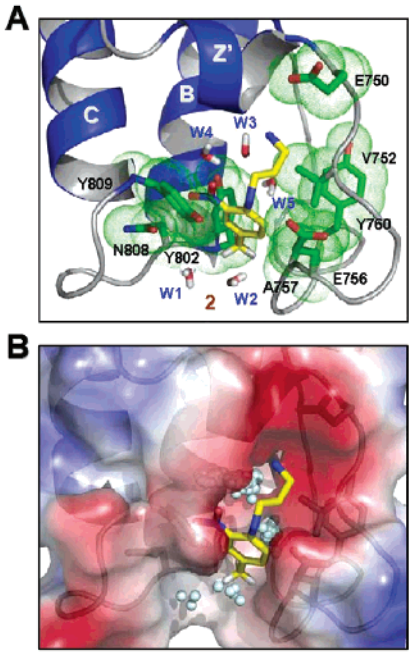


**Figure 2.** (A), (B) Ribbon and surface representations of the PCAF BRD/**2** complex structure showing mean positions of trapped water molecules (W1−W5) identified in the GCE simulation. The deletion sites are indicated as spheres in B.

**Table 1.** Number of Water Deletions in the Vicinity of the Ligand

| Rc | region | | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 3.5 | NP1 | in | 216 | 149 | 575 | 538 | 418 | 551 | 397 | 400 | 626 |
| 3.5 | NP1 | out | 32 | 20 | 131 | 43 | 272 | 32 | 26 | 37 | 212 |
| 3.5 | NP1 | total | 248 | 169 | 706 | 581 | 690 | 583 | 423 | 437 | 838 |
| 3.5 | CZ | in | 109 | 87 | 139 | 158 | 225 | 298 | 139 | 242 | 221 |
| 3.5 | CZ | out | 31 | 15 | 129 | 11 | 266 | 14 | 25 | 33 | 196 |
| 3.5 | CZ | total | 140 | 102 | 368 | 169 | 491 | 312 | 164 | 275 | 417 |
| 4.0 | NP1 | in | 290 | 198 | 771 | 634 | 586 | 675 | 574 | 508 | 608 |
| 4.0 | NP1 | out | 36 | 32 | 159 | 65 | 347 | 73 | 52 | 44 | 275 |
| 4.0 | NP1 | total | 326 | 230 | 930 | 701 | 933 | 848 | 626 | 552 | 973 |
| 4.0 | CZ | in | 90 | 99 | 365 | 367 | 326 | 414 | 219 | 324 | 403 |
| 4.0 | CZ | out | 31 | 35 | 143 | 18 | 335 | 17 | 37 | 37 | 258 |
| 4.0 | CZ | total | 221 | 134 | 508 | 385 | 661 | 431 | 256 | 361 | 661 |

The number of successful water deletions in the vicinity of each ligand derivative is shown in Table 1. For each compound, we calculated corresponding deletion sites within 3.5 and 4.0 Å of the ligand and for sites between the ligand and the protein (labeled as "in"), as well as for sites on the solvent exposed site of the protein (labeled as "out"). Furthermore, sites were counted in the proximity of the entire ligand (labeled as "NP1") or only near the chemical groups involved in the derivatization, up to and including the anchoring carbon in the aniline ring (labeled as "CZ").

Because attachment of heavy atoms to a chemical ligand likely results in an increase of its surface area, comparison of deletion counts was done in individual groups. The salient features that we concluded from GCE calculations with these hypothetical chemical derivatives are as follows: (i) the $-CF_3$ derivative **3** has better solvation than ligand **1**. The latter lacks the *para*-methyl group; (ii) of the derivatives that contain $-CH_2OH$, $-CH_2NH_2$, $-CH_2SH$, and $-CH_2Cl$ (**4**−**7**) at the *para*-methyl group, **5** shows consistently lower water deletion counts at both 3.5 and 4.0 Å distances for the entire molecule as well as near the derivation site; and (iii) the compounds with an extra methyl display solvation strength in the order of $-CHClCH_3$ (**8**) > $-CHBrCH_3$ (**9**) ≫ $-CHICH_3$ (**10**), of which the iodine derivative **10** is markedly less well solvated than the former

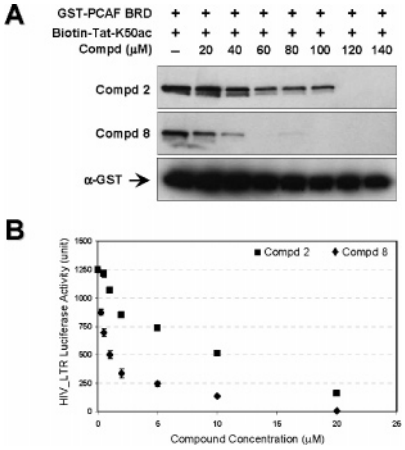


**Figure 3.** Blocking of PCAF BRD/Tat-K50ac binding by **2** or **8**. (A) In vitro inhibition of **2** or **8** for GST-PCAF BRD binding to a biotinylated Tat-K50ac peptide, as assessed by anti-GST Western blot. (B) HIV LTR luciferase assay assessing efficacy of **2** or **8** in inhibiting Tat-mediated HIV transcription in the cell.

two. Notably, the GCE calculated deletion counts between the entire ligand and the proximity of the *para*-methyl group are in accord for most of the compounds, with the exception of **6** (Table 1). In the latter case, the $-CH_2SH$ group is better solvated at the protein/ligand interface than at the entire ligand, but this gain was countered by its worsened solvation on the solvent-exposed side.

To validate the predications of chemical derivations from the GCE simulations, we chemically synthesized several **2** analogs, that is, **3**, **4**, **5**, and **8**, which were predicted in the aforementioned GCE simulations to have improved binding affinity for the PCAF BRD. The corresponding synthesis routes and conditions for these chemical derivatives are described in detail in Support Information. We then assessed the ability of these newly synthesized chemical analogs in blocking PCAF BRD binding to a Tat-K50ac peptide in a competition assay. In this assay, inhibition of a ligand against PCAF BRD binding to the biotinylated Tat-K50ac peptide that was bound to streptavidin agarose beads was measured in a concentration-dependent manner. Of these **2** analogs, **3**, **4**, and **5** showed a modest increase in the inhibition against PCAF BRD/Tat binding as compared to **2** (data not shown), whereas the new ligand **8**, *N*-(3-amino-propyl)-4-(1-chloroethyl)-2-nitrobenzene amine, exhibited an estimated 15-fold improvement of the inhibitory activity in vitro (Figure 3A).

We further showed that **8** exhibited $EC_{50}$ of $\sim 1$ $\mu$M (as compared to an estimated $EC_{50}$ of 10 $\mu$M for **2**) in its cellular efficacy on inhibiting Tat-mediated transcription of the viral promoter in a HIV-1 LTR-luciferase reporter gene assay (Figure 3B). Moreover, our study showed that the PCAF BRD ligands are selective over other members of the structurally conserved BRD family,[19] including the BRD from the human transcriptional co-activator CBP (CREB binding protein). It has recently been shown that CBP via its BRD can bind specifically to the human tumor suppressor p53 at the C-terminal acetylated lysine 382 (K382ac)[20] and that this CBP BRD/p53-K382ac interaction is responsible for p53 acetylation-dependent co-activator recruitment in response to DNA damage or external stress signals, a step essential for p53-induced transcriptional activation of cyclin-dependent kinase inhibitor p21 in G1 cell cycle arrest.[20,21] While transcriptional activity of p53 can be modulated by CBP BRD-specific small molecules,[22,23] neither **8** nor **2** showed any effect at 5 $\mu$M on p53-induced p21 transcription in response to DNA damage upon UV treatment (data not shown), thereby demonstrating the high selectivity of these PCAF BRD-specific ligands over other structurally similar proteins. Collectively, these data confirm our water deletion replacement predictions by the GCE simulations and support the notion that appropriate modifications of a lead that displaces its surrounding weakly trapped water when bound to a receptor can result in an enhancement of lead binding to the receptor. Furthermore, these studies highlight the potential of the markedly improved lead **8** as a tool in facilitating our further validation of PCAF BRD/HIV Tat association as a target for new anti-HIV therapy.

In summary, our study reported here shows that identification of displaceable bound water molecules near the receptor-bound ligand that are determined based on the open-ensemble simulation may offer a valuable strategy in guiding chemical modifications of lead optimization in small-molecule drug design.

**Acknowledgment.** We thank the National Institutes of Health (to M.M.Z. and S.M.), the GlaxoSmithKline and the Campbell Foundations (to M.M.Z.), and National Science Foundation of China (#30572234 to Z.Y.W.) for financial support of this work.

**Supporting Information Available:** Experimental details and procedures. This material is available free of charge via the Internet at http://pubs.acs.org.

## References

(1) Wei, X.; Ghosh, S. K.; Taylor, M. E.; Johnson, V. A.; Emini, E. A.; Deutsch, P.; Lifson, J. D.; Bonhoeffer, S.; Nowak, M. A.; Hahn, B. H. Viral dynamics in human immunodeficiency virus type 1 infection. *Nature* **1995**, *373*, 117−122.
(2) Ho, D. D.; Neumann, A. U.; Perelson, A. S.; Chen, W.; Leonard, J. M.; Markowitz, M. Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection. *Nature* **1995**, *373*, 123−126.
(3) Ho, D. D.; Zhang, L. HIV-1 rebound after anti-retroviral therapy. *Nat. Med.* **2000**, *6* (7), 736−737.
(4) Jeang, K.-T.; Xiao, H.; Rich, E. A. Multifaceted activities of the HIV-1 transactivator of transcription, Tat. *J. Biol. Chem.* **1999**, *274*, 28837−28840.
(5) Mujtaba, S.; He, Y.; Zeng, L.; Farooq, A.; Carlson, J. E.; Ott, M.; Verdin, E.; Zhou, M. M. Structural basis of lysine-acetylated HIV-1 Tat recognition by PCAF bromodomain. *Mol. Cell* **2002**, *9* (3), 575−586.
(6) Dorr, A.; Kiermer, V.; Pedal, A.; Rackwitz, H. R.; Henklein, P.; Schubert, U.; Zhou, M. M.; Verdin, E.; Ott, M. Transcriptional synergy between Tat and PCAF is dependent on the binding of acetylated Tat to the PCAF bromodomain. *EMBO J.* **2002**, *21* (11), 2715−2723.
(7) Zeng, L.; Li, J.; Muller, M.; Yan, S.; Mujtaba, S.; Pan, C.; Wang, Z.; Zhou, M. M. Selective small molecules blocking HIV-1 Tat and coactivator PCAF association. *J. Am. Chem. Soc.* **2005**, *127* (8), 2376−2377.
(8) Adams, D. J. Grand canonical ensemble Monte Carlo for a Lennard−Jones fluid. *Mol. Phys.* **1975**, *29*, 307−311.
(9) Mezei, M. Grand-canonical ensemble Monte Carlo study of dense liquid Lennard-Jones, soft spheres and water. *Mol. Phys.* **1987**, *61* (3), 565−582.
(10) Tieleman, D. P.; Berendsen, H. J. A molecular dynamics study of the pores formed by *Escherichia coli* OmpF porin in a fully hydrated palmitoyloleoylphosphatidylcholine bilayer. *Biophys. J.* **1998**, *74* (6), 2786−2801.
(11) Resat, H.; Mezei, M. Grand canonical Monte Carlo simulation of water positions in crystal hydrates. *J. Am. Chem. Soc.* **1994**, *116*, 7451−7452.
(12) Verdonk, M. L.; Chessari, G.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Nissink, J. W.; Taylor, R. D.; Taylor, R. Modeling water molecules in protein−ligand docking using GOLD. *J. Med. Chem.* **2005**, *48* (20), 6504−6515.
(13) Li, Z.; Lazaridis, T. The effect of water displacement on binding thermodynamics: concanavalin A. *J. Phys. Chem. B* **2005**, *109* (1), 662−670.
(14) Speidel, J. A.; Banfelder, J. R.; Mezei, M. Automatic control of solvent density in grand canonical ensemble Monte Carlo simulations. *J. Chem. Theory Comput.* **2006**, *2*, 1429−1434.
(15) Mezei, M.; Beveridge, D. L. Generic solvation sites in a crystal. *J. Comput. Chem.* **1984**, *6*, 523−527.

(16) Tempe, B. L.; McCammon, J. A. Ligand−receptor interactions. *J. Comput. Chem.* **1984**, *8*, 281−283.
(17) Mezei, M. URL: http://inka.mssm.edu/∼mezei.
(18) Owicki, J. C. *Computer Modeling of Matter*; ACS Symposium Series; American Chemical Society: Washington, DC, 1978; Vol. 86.
(19) Zeng, L.; Zhou, M. M. Bromodomain: An acetyl-lysine binding domain. *FEBS Lett.* **2002**, *513* (1), 124−128.
(20) Mujtaba, S.; He, Y.; Zeng, L.; Yan, S.; Plotnikova, O.; Sachchidanand; Sanchez, R.; Zeleznik-Le, N. J.; Ronai, Z.; Zhou, M. M. Structural mechanism of the bromodomain of the coactivator CBP in p53 transcriptional activation. *Mol. Cell* **2004**, *13* (2), 251−263.
(21) Barlev, N. A.; Liu, L.; Chehab, N. H.; Mansfield, K.; Harris, K. G.; Halazonetis, T. D.; Berger, S. L. Acetylation of p53 activates transcription through recruitment of coactivators/histone acetyltransferases. *Mol. Cell* **2001**, *8*, 1243−1254.
(22) Sachchidanand; Resnick-Silverman, L.; Yan, S.; Mutjaba, S.; Liu, W. J.; Zeng, L.; Manfredi, J. J.; Zhou, M. M. Target structure-based discovery of small molecules that block human p53 and CREB binding protein association. *Chem. Biol.* **2006**, *13* (1), 81−90.
(23) Mujtaba, S.; Zeng, L.; Zhou, M. M. Modulating molecular functions of p53 with small molecules. *Cell Cycle* **2006**, *5* (22), 2575−2578.

JM070014G