Chemistry & Biology, Vol. 9, 185–193, February, 2002, ©2002 Elsevier Science Ltd. All rights reserved. PII S1074-5521(02)00106-0

Structure-Based Computational Database Screening, In Vitro Assay, and NMR Assessment of Compounds that Target TAR RNA

Kenneth E. Lind,[1] Zhihua Du,[1] Koh Fujinaga,[2] B. Matija Peterlin,[2,3,4] and Thomas L. James[1,5]
[1] Department of Pharmaceutical Chemistry
[2] Department of Medicine
[3] Department of Microbiology and Immunology
[4] Howard Hughes Medical Institute
University of California
San Francisco, California 94143

Summary

There has been little prior effort to discover new drugs on the basis of a unique RNA structure. Binding of the viral transactivator Tat to the 5′ bulge of the transactivation response (TAR) element is necessary for HIV-1 replication, so TAR RNA is a superb target. A computational approach was developed to screen a large chemical library for binding to a three-dimensional RNA structure. Scoring function development, flexible ligand docking, and limited target flexibility were essential. From the ranked list of compounds predicted to bind TAR, 43 were assayed for inhibition of the Tat-TAR interaction via electrophoretic mobility shift assays. Eleven compounds (between 0.1 and 1 μM) inhibited the Tat-TAR interaction, and some inhibited Tat transactivation in cells. NMR spectra verified specific binding to the 5′ bulge and no interaction with other regions of TAR.

Introduction

Ribonucleic acids are a promising, yet relatively untapped, target for drug design. Currently, most drugs on the market target proteins [1]. Yet, in many ways RNA may be a better target than a protein because it is upstream in the translation pathway. Inhibiting a single RNA molecule could prevent the production of thousands of proteins. As more three-dimensional (3D) RNA structures become available, unique binding sites will be defined for targeting.

There have been some attempts to discover drugs that interact with RNA [2, 3]. These have focused primarily on antibacterial agents because the target of some clinically important antibacterial drugs, originally discovered by soil sample screening, was found to be bacterial RNA. In general, these aminoglycosides are valuable [1], but they have undesirable features: they interact with other sites on nucleic acids in human cells, and bacteria develop resistance. Consequently, it is desirable to identify new classes of compounds for drug development and RNA targets that cannot be altered because of function. In humans, where RNA is expressed at even higher levels, drug candidates must be specific for a single RNA site.

In recent years, several drugs designed by the use of 3D protein structures were approved for clinical use [4], yet there has been little reported research on drug discovery using unique 3D RNA structures [5, 6]. Undoubtedly, this neglect has been due to the relative paucity of RNA structures. However, the number of RNA structures is increasing, and some of them represent potential targets for the design of new drugs.

Our goal is to develop a method to discover novel non-peptide, non-nucleotide compounds that interact with high affinity and specificity with a single RNA site. These compounds should also represent promising scaffolds for subsequent chemical modification that would enhance their pharmaceutical properties. One way to find such promising lead compounds is to screen the Available Chemicals Directory (ACD) or a corporate library of compounds. These typically comprise 10^5–10^6 compounds. Currently, high-throughput screening assays are set up at enormous expense to monitor experimentally the binding of compounds. However, several recent improvements in computational docking methodologies permit effective screening of compounds in silico. In particular, our inclusion of solvation effects for predicting accurate binding free energies served as a basis for developing a more accurate screen for ligand binding to RNA.

Acquired immunodeficiency syndrome (AIDS) is caused by the human immunodeficiency virus, type 1 (HIV-1). Unique structures on the HIV-1 RNA genome play essential roles in viral replication and are good targets for drug design. One such site is the transactivation response (TAR) element. It binds the viral transactivation protein, Tat. Both form a ternary complex with human cyclin T1 (CycT1), resulting in increased rates of elongation of transcription on the HIV-1 genome [7]. TAR RNA forms a short, bulged stem-loop structure (Figure 1) at the 5′ end of all viral transcripts [8]. Essential features of Tat and TAR have been delineated. In particular, the 5′ bulge and adjacent stem region are critical for Tat binding (critical residues are highlighted in red in Figure 1) [9, 10]. Interrupting the interaction between Tat and TAR blocks HIV-1 replication in infected cells [11].

There have been efforts to find small organic inhibitors of the HIV-1 Tat-TAR interaction [12]. Analogs of amino acids or nucleotides [13, 14] and aminoglycosides [15] can inhibit the binding of Tat to TAR. Aminoglycoside-arginine conjugates were effective inhibitors of the Tat-TAR interaction [16]. Whereas the best of the other compounds bind with micromolar affinity, some aminoglycosides that were conjugated with multiple arginines achieved binding in the 20–400 nM range.

In this report, we first describe the generation of an improved scoring function, which is essential for computational ligand screening of RNA targets. This virtual screening methodology should lead to faster and less expensive drug discovery. Next, we describe how we found new classes of ligands that block the Tat-TAR interaction. Virtual screening of the ACD identified possible candidates, in vitro electrophoretic mobility shift assays (EMSA) and NMR spectra demonstrated that

[5] Correspondence: james@picasso.ucsf.edu

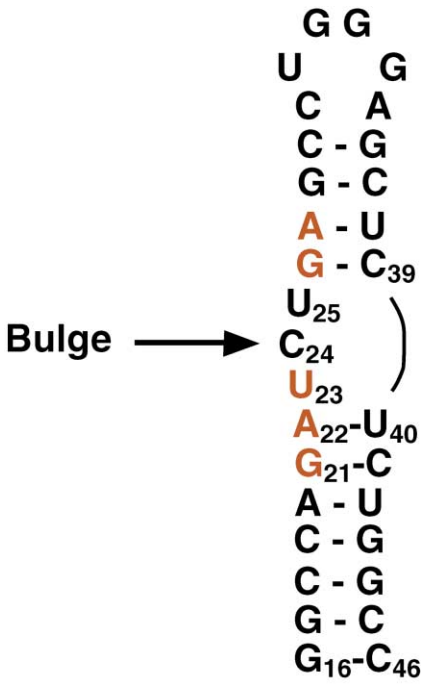

Figure 1. TAR RNA Secondary Structure

Residues shown in red are critical for binding of the Tat protein.

they bind specifically to the Tat binding site on TAR, and a functional assay showed that they inhibit Tat transactivation in cells.

Results

Basic Screening Rationale

We have previously described the basic rationale of our approach to discovering novel ligands capable of binding to unique 3D RNA structures [5, 6]. After selecting a promising target with available 3D structure, we perform virtual screening of the ACD, which contains 181,000 compounds, to identify candidates with good predicted affinity for this RNA target.

Two different programs were used for docking the ligands to the target RNA: DOCK [17] and ICM [18]. DOCK is fast for matching compounds to a specific site on a target molecule. It accommodates rigid and flexible ligands and utilizes a relatively simple scoring function that evaluates van der Waals (VDW) and electrostatic interactions. The number of compounds in the ACD requires rapid docking and simple scoring. After DOCK runs, $<$20% of the ACD compounds ranked well enough for subsequent ICM screening. ICM is a more diverse scripting program that allows rapid exploration of conformational space via Monte Carlo searches using internal coordinates. ICM also permits flexible receptors and includes terms for hydrogen bonding and solvation.

Scoring Function for Ligand Binding to RNA

It is crucial to assess the relative binding strength of screened compounds. Potential energy functions (force fields) used during the docking process are necessarily simplified. Consequently, we developed an empirical scoring function for use in ranking compounds after docking created all of the ligand-RNA complexes [19].

The determination of several RNA-ligand complex structures has advanced understanding of interactions between RNA and ligand [20–33]. Binding affinities for these complexes were also reported [26, 27, 34–42]. We used these structures and binding energies first to develop a scoring function and subsequently to validate our procedure. These 13 RNA-ligand complexes with known structures and binding affinities include three different neomycin-RNA complexes and two separate tobramycin-RNA aptamers and are listed in Table 1. In addition, data for binding and nonbinding analogs of several RNA complexes were included in the calculations. For our computational methods, it is important that structures and binding energies of known test cases be reproduced. After several exploratory calculations to reproduce the structures of the known RNA-ligand complexes, we utilized the procedure described in the Experimental Procedures section.

Rigid redocking (RMSD $<$ 1 Å) accurately reproduced all 13 complexes (Table 1). The most accurate prediction of binding orientation was achieved when the electrostatic interactions were reduced by two; default values overemphasize the electronegative phosphate backbone. By reducing charged interactions (equivalent to salt condensation around the backbone [43]), we were able to recreate the original binding orientations. With flexible "anchor-and-grow" docking, arginine, citrullene, and the aminoglycosides did not bind in their original positions; other conformations had good DOCK scores. In part, this is a sampling issue because not all possible binding conformations were tried. With a large database, exhaustive sampling would be too time consuming. Also, this may reflect lack of a solvation term in the DOCK scoring function. Alternate conformations scored well in DOCK but could become disfavored if solvation were considered.

When the ligands were reminimized with ICM, they maintained their original docked conformation. Imposing light constraints sufficed to keep the original DOCK position. Because DOCK does not accurately predict true binding affinity, this is essential for determining an accurate scoring value for the complex resulting from DOCK.

After ICM reminimization, interaction energies for the ligand-RNA complex were decomposed into electrostatic, van der Waals, hydrogen bond, solvation, and entropic contributions from rotatable bonds [5]. The binding free energy is approximated as the sum of these terms:

$$\Delta G_{bind} \approx A \cdot \Delta E_{VDW} + B \cdot \Delta E_{El} + C \cdot \Delta E_{Hbond} + D \cdot \Delta E_{solv} + 0.3 \cdot \Delta E_{tors} \quad (1)$$

For each ΔE term, the energies of the free RNA and free ligand were subtracted from the energy of the complex.

Linear regression analysis using data for the thirteen known ligand-RNA complexes yielded values for coefficients A, B, C, and D for VDW, electrostatic, H bond, and solvation terms, respectively. Known binding energies were reproduced, but manual adjustment was carried

Table 1. RNA-Ligand Complexes with 3D Structure and Binding Constant Available

PDB Entry	Name	K_d (μM)	ΔG (kcal/mol)
1RAW [22]	AMP-RNA aptamer	2–14 [34]	−6.95
1KOC [23]	Arginine-RNA aptamer	60 [35]	−5.76
1F27 [24]	Biotin-RNA aptamer	6 [36]	−7.12
1KOD [23]	Citrullene-RNA aptamer	68 [35]	−5.68
1FMN [25]	FMN-RNA aptamer	0.5–200 [37]	−6
1BYJ [26]	Gentamicin-16S rRNA	0.01 [26]	−10.98
1EI2 [27]	Neomycin-TAU exon RNA	~1 [27]	−8
1NEM [33]	Neomycin-RNA aptamer	0.1 [38]	−9.6
1QD3 [29]	Neomycin-TAR	~1 [39]	−8
1PBR [30]	Paromomicin-16S rRNA	0.2 [40]	−9.19
1EHT [31]	Theophylline-RNA aptamer	0.4 [41]	−8.72
1TOB [28]	Tobramicin-RNA aptamer	0.0007 [42]	−12.48
2TOB [32]	Tobramicin-RNA aptamer	0.0012 [42]	−12.24

out to improve the separation of known binders from "non-binders." Values that were ultimately used were A = 0.15, B = 0.1, C = 0.5, and D = 0.025 for the full force field. These values reflect both binding and non-binding ligands scoring properly. Any ligand that experimentally does not bind its RNA should have a "poor" binding free energy (>0). Thus, while improving scoring of known non-binders, the final scoring function yields significant error in the predicted binding energy of known ligands (Table 2). Individual energy terms are well balanced. Ligands that bind through electrostatic interactions (e.g., aminoglycosides) show this preference, whereas those in stacking interactions (e.g., theophylline) predominantly have VDW contacts.

As a final test of the scoring function, each RNA structure was screened against a small database (175 compounds) that included all native ligands and ~150 other randomly chosen compounds. The scoring procedure worked well for selecting the native ligand for each RNA target. This validation ensured that true binders ranked high in our final list. For arginine and citrullene aptamers, the binding conformation of the ligand was not predicted very well, but predicted and experimental binding energies were relatively close. Residual differences were due to sampling with ICM. (When developing the scoring function, we did not allow the receptor to move so that we could approximate the binding energy for the reference experimental structure.) Importantly, the ranking of each ligand improved compared to the initial DOCK ranking. Complete results of the fitness test are accessible at http://picasso.ucsf.edu.

Database Screening against TAR

We used our database screening procedure, as described in the Experimental Procedures section and shown as a schematic in Figure 2, to find among the 181,000 compounds in the ACD several lead compounds that could bind to the 5′ bulge on TAR. A list of 500 compounds, all predicted to bind with high affinity to TAR, was generated. These compounds are diverse and share no common structural features. They possess aromatic moieties that interact with the bases, charged groups that bind to the phosphates, and H bond donors and acceptors that form other specific interactions. The only common "feature" appears to be size and shape; all compounds fit tightly into the 5′ bulge.

Arginine binds weakly to TAR (K_d ~1 mM), and three arginine analogs are in the final list of 500 compounds. The highest ranking (β-napthyl-arg) is at position 87, with a score of −4.38 kcal/mol. Several arginine analogs with aromatic groups are present in the ACD. Aromatic groups add favorable stacking interactions that are not present with arginine alone. Aminoglycosides also bind TAR but do not appear in the final 500; most are large

Table 2. Binding Affinities (kcal/mole) of Native Ligands Used for Parameterization of Scoring Function

PDB Entry	$\Delta G \approx -RT \ln K_d$	Predicted Score[a]	Difference
1RAW (AMP)	−6.95	−3.8	+3.2
1KOC (Arg)	−5.76	−3.0	+2.8
1F27 (Biotin)	−7.12	−4.5	+2.6
1KOD (Cit)	−5.68	−3.2	+2.4
1FMN (FMN)	−6	−3.9	+2.1
1BYJ (Gent)	−10.98	−8.9	+2.1
1EI2 (Neo 1)	−8	−6.0	+2.0
1NEM (Neo 2)	−9.6	−13.0	−3.4
1QD3 (Neo 3)	−8	−11.2	−3.2
1PBR (Par)	−9.19	−8.7	+0.5
1EHT (Theo)	−8.72	−6.5	+2.2
1TOB (Tob 1)	−12.48	−11.3	+1.2
2TOB (Tob 2)	−12.24	−11.7	+0.5

[a] Predicted binding free energy (kcal/mol) for native ligands based on scoring function in Equation 1.

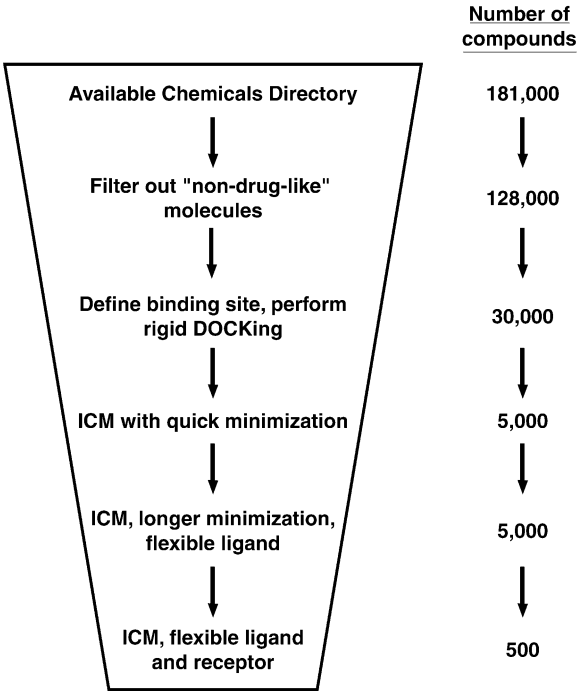

Figure 2. Flow for Computational Screening of a Chemical Database, e.g., the Available Chemicals Directory

The number of compounds, after scoring and ranking, that are passed along to the subsequent stage is shown to the right.

and thus are filtered out by the Lipinski rules [44]. Also, the 5′ bulge structure does not permit entry of aminoglycosides; large structural rearrangements would be necessary. Our method permits some flexibility of RNA in the final stages of calculation, but it is still rather limited. Previous studies showed a rearrangement of the 5′ bulge upon neomycin binding [29].

For the screening in vitro, we purchased 50 of the top 200 compounds on the final list. Some of the leads on the list were eliminated based on chemical intuition that they would not be good drugs, high cost (at least for our lab), and apparent lack of availability (in spite of the ACD listing). After assaying these, we purchased four more phenothiazines.

RNA Binding Assays In Vitro (Electrophoretic Mobility Shift Assay)

EMSA is an effective method for determining whether TAR forms a complex with Tat and human cyclin T1 [7]. In the presence of TAR, containing the 5′ bulge and central loop, and these proteins, a lower mobility complex is observed on polyacrylamide gels (Figure 2, lanes 1 and 2). The lack of higher-order complex formation signifies that a compound can block this RNA-protein interaction. Indeed, of the 50 compounds, eight were found to completely prevent the binding between Tat and TAR at concentrations between 0.1 and 1 μM, with hybrid CycT1-Tat protein and TAR concentrations each at 0.1 μM. Three others partially inhibited at 1 μM (data not shown). Prochlorperazine was one of the best inhibitors, so four analogs were also purchased and assayed.

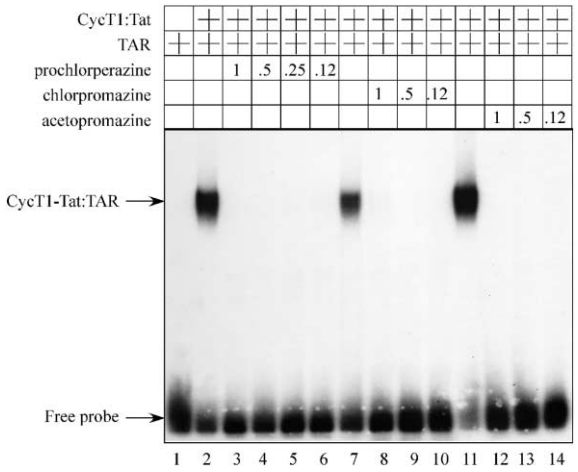

Figure 3. Electrophoretic Mobility Shift Assay for Monitoring Inhibition of the Interaction of Tat with TAR in the Presence of CycT1

Binding among CycT1, Tat, and TAR can be disrupted by some phenothiazines. Equimolar concentrations (0.1 μM) of radiolabeled TAR and the hybrid CycT1-Tat protein were incubated alone or with decreasing concentrations of different compounds and separated by nondenaturing gel electrophoresis. Data with prochlorperazine, chlorpromazine, and acetylpromazine are presented. Numbers refer to the concentration of these compounds in μM. CycT1-Tat fusion protein was present in lanes 2–14. Free probe is presented in lane 1. Arrows point to the RNA-protein complex between CycT1-Tat and TAR and to free TAR.

Three phenothiazines (prochlorperazine, chlorpromazine, and acetopromazine) each inhibited the Tat-TAR interaction at all concentrations examined (Figure 3, lanes 3–6, 8–10, and 12 and 13, respectively), but trifluoperazine and thioethylperazine did not (our unpublished data). These phenothiazines are used clinically as antipsychotic, sedative, and antiemetic agents. Recent studies also suggest that they have antibiotic properties [45].

Figure 4 presents the compounds that inhibit at concentrations of 0.1–1 μM. Two are fluorescein analogs, another appears to be an intercalator, and the others have pharmaceutical potential. The EMSA data clearly support the computational screening results. Although some compounds bind, modifying these ligands could yield even better inhibitors.

Functional Assays in Cells

We studied some compounds in cells. Prochlorperazine results are shown in Figure 5. We cotransfected target and effector plasmids in the presence of prochlorperazine in HeLa cells. Two different sets of targets and effectors were used. First, HIV-1 LTR and Tat were coexpressed. Second, as a control, we utilized the heterologous tethering system of the Regulator of Expression of Virion genes (Rev) and its Rev response element (RRE) RNA. If specific, our compounds should block Tat transactivation via TAR and not have any effects on the heterologous tethering of the RevTat fusion protein via RRE. Indeed, at concentrations between 5 and 20 μM, prochlorperazine inhibited Tat transactivation up to 6-fold (Figure 5, bars 2 and 5) in a dose-dependent fashion. Importantly, this compound had no effect on the hybrid RevTat protein on the RRE (Figure 5, bars 7 and 10). With

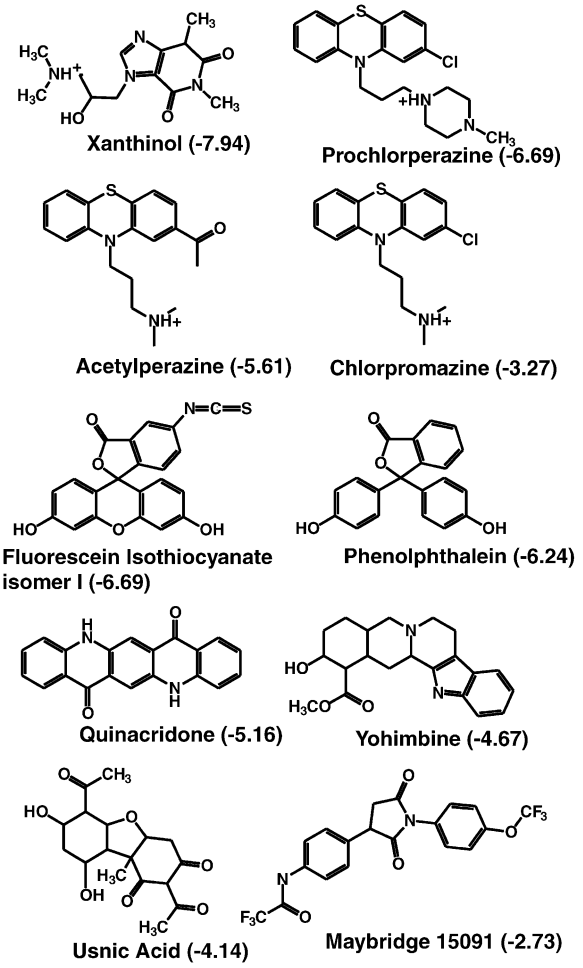

Figure 4. Ligands Found Experimentally to Bind to the TAR 5′ Bulge

Ligands found to bind ($\Delta G \approx -6$ kcal) to the TAR 5′ bulge at concentrations of 1 μM or less as determined by EMSA. Binding free energies (kcal/mole) predicted by Equation 1 are in parentheses.

prochlorperazine >20 μM, significant cellular toxicity resulted. In short, some compounds specifically inhibit the interaction among CycT1, Tat and TAR, and transactivation in cells.

NMR Screening of Selected Compounds

With structure-based drug discovery, lead compounds identified by virtual screening are almost always aimed at the active site of the protein or nucleic-acid target. Compounds identified by virtual screening bind a specific target. In vitro assays confirm inhibition but do not explicitly show that the ligand binds to the designated site, and they do not preclude its binding to other sites. However, NMR not only shows whether a ligand binds to the specific target but also reveals all sites on the target that interact with the ligand. This requires nuclear resonance assignments for the target. We saved substantial time by using resonance assignments that Prof. Jamie Williamson sent to us for HIV-1 TAR. His lab has revealed valuable insights into interactions between peptides derived from Tat and TAR [46, 47]. We primarily needed to reconcile our spectra with those assignments,

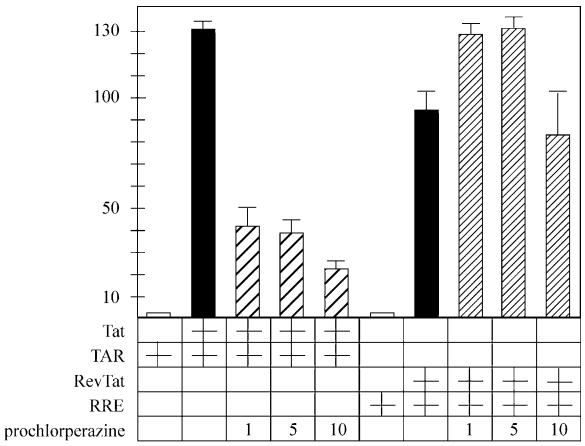

Figure 5. CAT Assay Results Designed to Test Whether or Not Prochlorperazine, Which Inhibits the Tat-TAR Interaction, Can Inhibit Tat Transactivation in Cells

Tat transactivation is clearly inhibited by prochlorperazine in cells. HeLa cells coexpressed either HIV-1 LTR (TAR) and Tat or pRRESCAT (RRE) and the RevTat fusion protein. Increasing concentrations of prochlorperazine diminished Tat transactivation via TAR but, as a control, had no effect on that between Rev and RRE. As observed in lanes 2 and 5, Tat transactivation was equivalent in native and heterologous RNS-tethering systems. The data represent three experiments with depicted standard errors. Expression levels from both target plasmids were set to one (lanes 1 and 6).

especially when differences were due to TAR-ligand complex formation.

NMR experiments with several lead compounds confirmed binding to the 5′ bulge of TAR. In particular, chlorpromazine and its relatives produced very distinct chemical shifts for resonances emanating from bulge residues. For example, the 5′ bulge contains three pyrimidine residues (U23, C24, and U25) that can be monitored by their H5-H6 crosspeaks in 2D TOCSY spectra. NMR experiments confirmed acetylpromazine binding to TAR. Most satisfying, chemical shifts changed upon the addition of ligand only for nuclei in the 5′ bulge and not in the stem or loop. Figure 6A illustrates this with pyrimidine H5–H6 crosspeaks. A comparison of chemical shifts with and without acetylpromazine revealed the following substantial (>0.1 ppm) changes (downfield and upfield shifts are denoted by positive and negative values, in ppm, respectively): A22H8 (+0.12), A22H4′ (+0.14), U23H6 (+0.16), U23H5 (+0.22), U23H1′ (+0.14), U23H2′ (+0.13), U23H4′ (−0.15), C24H1′ (+0.10), U25H2′ (−0.11), U25H3′ (−0.17), G26H1′ (−0.22), C39H1′ (+0.13), C39H3′ (+0.15), U40H6 (−0.11), U40H5 (−0.10), U40H1′ (−0.18), C41H6 (−0.12), and C41H5 (−0.13). All chemical-shift changes were observed for nuclei in the bulge residues U23, C24, U25 and adjacent base-pairs (A22, G26, C39, U40, and C41). Even with a 5-fold excess of acetylpromazine, no chemical-shift changes were detected from nuclei in other regions of TAR. The chemical-shift data show that the binding of acetylpromazine to TAR is specific to the 5′ bulge. Nonspecific ligand-RNA interactions are not involved in the binding.

The bulge region contains several uridine residues, so a sample with only U residues uniformly labeled in ^{13}C and ^{15}N was prepared. Figure 6B shows the ribose

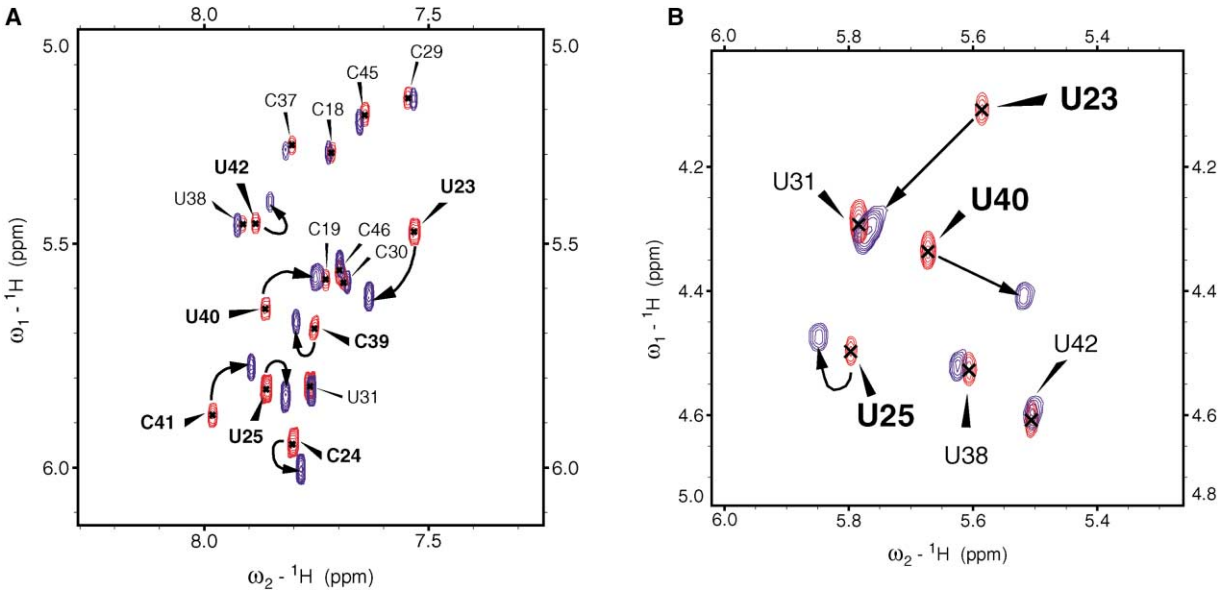

Figure 6. NMR Experiments Show that Acetylpromazine Binds Specifically to the 5′ Bulge of TAR and Not to the Stem or Loop Regions

(A) 2D TOCSY spectrum showing pyrimidine H5-H6 crosspeaks for TAR. Red: free RNA; blue: complex with acetylpromazine. Conditions: 1.5 mM RNA, 20 mM phosphate, 20 mM NaCl (pH 6.5), 35°C. Green arrows indicate chemical-shift changes.

(B) Ribose H1′-H2′ region of the HCCH-COSY spectrum for {^{13}C,^{15}N-labeled U}TAR in the absence (red) and presence (blue) of acetylpromazine. Arrows identify crosspeaks that shift with added ligand.

H1′-H2′ region of HCCH-COSY spectra for the {^{13}C,^{15}N-labeled U} TAR in the absence and presence of acetylpromazine. The HCCH-COSY spectral data confirm the conclusions from TOCSY spectra, and both confirm the computational as well as in vitro and in vivo assay results, showing that the compounds bind in a specific manner to this RNA. In particular, NMR chemical shifts are observed only for residues A22, U23, C24, U25, G26, C39, U40, and C41 (see Figure 1 to observe how these residues define the binding site).

Discussion

The computational screening of compounds for binding to 3D protein structures has led to clinically useful drugs [4]. Typically, initial screens found weak (~10–100 μM) binders that could be modified to yield nanomolar binding compounds—the improvement in ligand affinity needed is equivalent to the energy involved in forming a couple of additional hydrogen bonds. This combination of computational and chemical optimization approaches has been quite successful in moving compounds into clinical trials.

We have developed a computational method that predicts the binding of small molecules to RNA targets. Critical to our success, the scoring function was optimized to produce approximate free-energy scores as well as relative ranks. The fact that ten of our approximately 50 selected compounds, at concentrations of 0.1–1 μM, inhibited the binding between TAR and Tat shows satisfactorily that the scoring function worked well. It is even more encouraging that the compounds are specific for the unique 5′ bulge in TAR. However, the ~40 false positives show that improvements are needed. Future changes to the scoring function, in particular better treatment of solvation effects with the generalized Born or Poisson-Boltzmann approaches in DOCK and ICM, should improve selection. Developments permitting greater flexibility in the RNA target would also help.

We identified acetylpromazine as an exciting lead compound that binds specifically to the same TAR site that is recognized by the Tat protein. Determination of the structure of this complex is progressing. This could allow us to suggest modifications designed to enhance affinity and specificity for the 5′ bulge. The structure can also be compared to computational predictions to improve our scoring function. With few RNA-ligand structures currently available, each additional one could greatly improve our ability to make predictions for drug discovery.

Significance

The computational screening of compounds that bind to three-dimensional protein structures has led to several novel, clinically useful drugs [4]. Typically, these initial screens identified weak (~10–100 μM) binders that could be modified to yield nanomolar binding compounds—the needed improvement in ligand affinity is equivalent to the energy involved in forming a couple of additional hydrogen bonds. The combination of computational lead generation and chemical optimization has been quite successful in moving compounds into clinical trials. Other than initial work from the University of California-San Francisco, there has been no report of the computational targeting of RNA. Here we have demonstrated the concept with a protocol that predicts the binding of small molecules to RNA targets. Developing an effective scoring function

to assess the binding of these compounds to RNA has been of paramount importance. We have optimized this scoring function to produce both approximate binding free energies and relative ranks. The resulting candidates are non-peptide, non-nucleotide organic compounds whose molecular weight is generally less than 600 Da. Although the method is far from perfect and there is room for improvement, our initial results with HIV-1 TAR as a target are very encouraging. We have identified promising drug candidate leads that inhibit the crucial Tat-TAR interaction, according to an in vitro assay, at a concentration of ~0.1–1 μM. NMR spectra of TAR complexed with one promising lead, acetylpromazine, demonstrate that the compounds bind to the targeted bulge site on TAR and do not bind to the stem or loop regions. Furthermore, we have shown that some of them also inhibit Tat transactivation in cells. Optimization of these compounds to enhance their affinity and decrease their side effects will lead to a new class of chemical agents for inhibiting the HIV virus.

Experimental Procedures

Computational Screening with DOCK

A flow diagram showing our virtual screening procedure that works well for targeting RNA structures is shown in Figure 2. The DOCK program was utilized for the initial screening of ligands [17]. First, the known binding sites were filled with sphere sets from the SPHEREGEN module of DOCK. Rigid redocking of the ligands with DOCK was then employed to reproduce the bound conformation in 500 orientation steps with energy (electrostatic and VDW) scoring. Different scaling factors for both the electrostatic and VDW terms were tested for their ability to predict the correct structure. For comparison, flexible docking via the "anchor-and-grow" method was also used for the production of binding conformations [48].

Development of the Scoring Function

The ICM program [18] was used to explore ligand binding in more detail for a subset of structures that scored well in binding to RNA via DOCK. Once binding orientations were determined with DOCK, ICM was used to determine a binding score. The docked conformation was tethered to its original position with light (0.1 kcal/mol/Å^2) constraints and subjected to 5000 Monte Carlo search steps with 500 minimization steps. This maintains the binding conformation and adjusts for slight variations between the DOCK and ICM force fields. In particular, small differences in VDW calculations can lead to large differences in scores if the ligands are not allowed to reposition.

To test the scoring function, each RNA target was screened against a small database of compounds (all native ligands plus ~150 random compounds). With ICM rescoring, a two-step approach was used. First, the ligand was tethered with 0.1 kcal/mol/Å^2 constraints for 500 Monte Carlo steps with 50 minimization steps. Next, both the ligand and RNA residues within 5 Å were free to move for 5000 Monte Carlo steps (with 500 minimization steps). The same procedure was used subsequently for large database screening.

Database Screening against TAR RNA

The 181,000 compounds in the Available Chemicals Directory, 1998 version, (MDL Information Systems) were scanned to detect lead compounds.

The HIV-1 TAR target site (the bulge and adjacent base pairs in Figure 1) was built from a model of BIV TAR [5]. The 181,000 compounds in the ACD were scanned to detect lead compounds as follows. Initially, a simple filter eliminated "non-drug-like" molecules, based on rules outlined by Lipinski [44]. For our situation, requisite ligand characteristics were a maximum molecular weight of 600, a maximum total of ten hydrogen bond donors or acceptors, a formal charge between −5 and +5, and a maximum of 20 rotatable bonds. The resulting 128,000 drug-like compounds were screened with DOCK. Each compound was permitted 500 rigid orientations. Energy scoring was used, with a distance-dependent dielectric of 4 r. In addition, electrostatic interactions were reduced by a factor of two based on results we obtained with the small database screening. The top 30,000 compounds were kept.

Subsequently, ICM was used in three sequential steps to refine the initial DOCK results. First, each of the 30,000 compounds resulting from the DOCK search was reminimized and scored with 500 Monte Carlo and 50 minimization steps. The scoring function (Equation 1) was used to reorder the original DOCK list. Next, the top-scoring 5000 compounds were permitted flexibility in a longer calculation with 5000 Monte Carlo and 500 minimization steps, with rescoring. Finally, RNA residues within 5 Å of the binding pocket, as well as the ligand, were allowed to move during the Monte Carlo search. The 500 highest ranking compounds that result are all predicted to bind with high affinity to the 5′ bulge site.

Preparation of TAR RNA

Samples of the 31 nucleotide TAR (Figure 1) were prepared by in vitro transcription with T7 RNA polymerase and a synthetic DNA template [49]. Most preparations yielded unlabeled TAR, but TAR ^{13}C,^{15}N-labeled only at U residues was prepared by the use of labeled UTP (Isotec).

In vitro Electrophoretic Mobility Shift Assay (EMSA)

To assess whether compounds selected by computational methods actually bind to the 5′ bulge of TAR, an in vitro EMSA assay was developed [7]. Fifty compounds were selected from the final list of 500 (vide supra) and ordered. Each compound was dissolved to 10 μM in DMSO and incubated with ^{32}P-radiolabeled TAR (~1 μM) for 20 minutes at 30°C in EMSA binding buffer (30 mM Tris-HCl [pH 8.0]/70 mM KCl/0.01% Nonidet P-40/5.5 mM $MgCl_2$/1 mM DTT/13% glycerol), and poly-dIdC and poly-rIrC blockers. A hybrid protein composed of Tat and CycT1 was added with incubation for 20 min at 37°C. The final concentrations of hybrid Tat-CycT1 and TAR were 0.104 mM and 0.10 mM, respectively. The mixture was loaded onto a 6% nondenaturing SDS gel; electrophoretic power was 4 W for 3 hr at 4°C. The gel was dried, and autoradiography exposed positions of the TAR and Tat-CycT1:TAR complex on the gel.

Chloramphenicol Acetyl Transferase (CAT) Assay

HeLa cells were preincubated with drug in concentrations ranging from 0.01 to 10 μM. DNA constructs containing the engineered CAT gene preceded by TAR or RRE promoters were transfected into the HeLa cells with Lipofectin. Cells were incubated at 37°C, 5% CO_2 for 5 hr. Cells were rinsed, fresh drug and media (3 ml 10% Fetal Calf Serum, DMEM) were added, and then the cells were incubated for 3 days at 37°C, 5% CO_2. Cells were collected by rinsing with ~1 μl phosphate buffer and centrifugation for 4 min at 4000 rpm. Lysis buffer was added to the pellet, and centrifugation for 10 min at 14,000 rpm followed. The supernatant was heated to 65°C for 5 min (to remove background protein expression) and centrifuged for 10 min at 14,000 rpm. 100 μl supernatant, 1 mg chloramphenicol, 1 μg ^3H-acetyl-CoA, and EconoFluor solution were mixed and immediately placed into the scintillation counter. We measured CAT enzyme activity by detecting the amount of ^3H-acetyl-chloramphenicol.

NMR Screening of Selected Compounds

All NMR spectra were acquired on Varian Inova 600 MHz spectrometers, processed with NMRpipe [50], and analyzed with SPARKY [51]. Typical sample conditions were 1–1.5 mM RNA in 20 mM sodium phosphate buffer and 20 mM NaCl (pH 6.5). Each compound was in 2-fold excess over RNA. 2D TOCSY spectra were collected in 2.5 hr at 35°C in D_2O with 60 ms mixing time, MLEV-17 mixing sequence, 512 FIDs of 4096 complex points, and a 6000 Hz spectral width in both dimensions. The phase-sensitive mode in the indirectly detected dimension was achieved by the TPPI method [52]. The 2D (^1H, ^1H) HCCH-COSY spectrum was collected for TAR ^{13}C, ^{15}N-labeled only at U residues. Coupling constants were optimized for ribose with J_{HC} = 165 Hz and J_{CC} = 40 Hz. Spectral widths were 2200

Hz in both ^{1}H dimensions. 512 FIDs of 2048 complex points were collected.

Supplemental Material

A supplemental table is available at http://picasso.ucsf.edu. This table shows the predicted binding affinities and relative rank of native ligands with their RNA targets. The data were obtained with the computational screening procedure described in the Experimental procedures.

Acknowledgments

This work was supported principally by grant AI46967 and partially by AI44515 from the National Institutes of Health. We are grateful to Prof. I.D. Kuntz for helpful discussions and access to DOCK, to Prof. Ruben Abagyan for access to ICM, and to Prof. James R. Williamson for generously providing resonance assignments for TAR.

Received: August 21, 2001
Accepted: October 30, 2001

References

1. Schroeder, R., Waldsich, C., and Wank, H. (2000). Modulation of RNA function by aminoglycoside antibiotics. EMBO J. *19*, 1–9.
2. Chen, Q., Shafer, R.H., and Kuntz, I.D. (1997). Structure-based discovery of ligands targeted to the RNA double helix. Biochemistry *36*, 11402–11407.
3. Hermann, T., and Westhof, E. (1999). Docking of cationic antibiotics to negatively charged pockets in RNA folds. J. Med. Chem. *42*, 1250–1261.
4. Charifson, P.S., and Kuntz, I.D. (1997). Recent successes and continuing limitations in computer-aided drug design. In Practical Application of Computer-Aided Drug Design, P.S. Charifson, ed., (New York: Marcel Dekker), pp. 1–37.
5. Filikov, A.V., Mohan, V., Vickers, T.A., Griffey, R.H., Cook, P.D., Abagyan, R.A., and James, T.L. (2000). Identification of ligands for RNA targets via structure-based virtual screening: HIV-1 TAR. J. Comput. Aided Mol. Des. *14*, 593–610.
6. James, T.L., Lind, K.E., Filikov, A.V., and Mujeeb, A. (2000). Three-dimensional RNA structure-based drug discovery. In Structure, Motion, Interaction and Expression of Biological Macromolecules, R.H. Sarma, ed. (New York: Adenine Press), pp. 201–205.
7. Wimmer, J., Fujinaga, K., Taube, R., Cujec, T.P., Zhu, Y., Peng, J., Price, D.H., and Peterlin, B.M. (1999). Interactions between Tat and TAR and human immunodeficiency virus replication are facilitated by human cyclin T1 but not cyclins T2a or T2b. Virology *255*, 182–189.
8. Kao, S.-Y., Calman, A.F., Luciw, P.A., and Peterlin, B.M. (1987). Anti-termination of transcription within the long terminal repeat of HIV-1 by *tat* gene product. Nature *330*, 489–493.
9. Hauber, J., Malim, M.H., and Cullen, B.R. (1989). Mutational analysis of the conserved basic domain of human immunodeficiency virus *tat* protein. J. Virol. *63*, 1181–1187.
10. Delling, U., Reid, L.S., Barnett, R.W., Ma, M.Y., Climie, S., Sumner-Smith, M., and Sonenberg, N. (1992). Conserved nucleotides in the TAR RNA sStem of human immunodeficiency virus type 1 are critical for Tat binding and trans activation: model for TAR RNA tertiary structure. J. Virol. *66*, 3018–3025.
11. Hsu, M.-C., Schutt, A.D., Holly, M., Slice, L.W., Sherman, M.I., Richman, D.D., Potash, M.J., and Volsky, D.J. (1991). Inhibition Of HIV replication in acute and chronic infections in vitro by a Tat antagonist. Science *254*, 1799–1802.
12. Mei, H.-Y., Cui, M., Heldsinger, A., Lemrow, S.M., Loo, J.A., Sannes-Lowery, K.A., Sharmeen, L., and Czarnik, A.W. (1998). Inhibitors of protein-RNA complexation that target the RNA: specific recognition of human immunodeficiency virus type 1 TAR RNA by small organic molecules. Biochemistry *37*, 14204–14212.
13. Sullenger, B.A., Gallardo, H.F., Ungers, G.E., and Gilboa, E. (1991). Analysis of trans-acting response decoy RNA-mediated inhibition of human immunodeficiency virus type-1 transactivation. J. Virol. *65*, 6811–6816.
14. Hamy, F., Brondani, V., Flörsheimer, A., Stark, W., Blommers, M.J.J., and Klimkait, T. (1998). A new class of HIV-1 Tat antagonist acting through Tat-TAR inhibition. Biochemistry *37*, 5086–5095.
15. Zapp, M.L., Stern, S., and Green, M.R. (1993). Small molecules that selectively block RNA binding of HIV-1 Rev protein inhibit Rev function and viral production. Cell *74*, 969–978.
16. Litovchick, A., Evdokimov, A.G., and Lapidot, A. (2000). Aminoglycoside-arginine conjugates that bind TAR RNA: synthesis, characterization, and antiviral activity. Biochemistry *39*, 2838–2852.
17. Ewing, T.J.A., and Kuntz, I.D. (1997). Critical evaluation of search algorithms for automated molecular docking and database screening. J. Comput. Chem. *18*, 1175–1189.
18. Abagyan, R.A., Totrov, M.M., and Kuznetsov, D.N. (1994). ICM—a new method for protein modeling and design—applications to docking and structure prediction from the distorted native conformation. J. Comp. Chem. *15*, 488–506.
19. Filikov, A.V., and James, T.L. (1998). Structure-based design of ligands for protein basic domains: application to the HIV-1 Tat protein. J. Comp. Aided Molec. Design *12*, 1–12.
20. Hermann, T., and Patel, D.J. (2000). Biochemistry—adaptive recognition by nucleic acid aptamers. Science *287*, 820–825.
21. Patel, D.J., Suri, A.K., Jiang, F., Jiang, L.C., Fan, P., Kumar, R.A., and Nonin, S. (1997). Structure, recognition and adaptive binding in RNA aptamer complexes. J. Mol. Biol. *272*, 645–664.
22. Dieckmann, T., Suzuki, E., Nakamura, G.K., and Feigon, J. (1996). Solution structure of an ATP-binding RNA aptamer reveals a novel fold. RNA *2*, 628–640.
23. Yang, Y., Kochoyan, M., Burgstaller, P., Westhof, E., and Famulok, M. (1996). Structural basis of ligand discrimination by two related RNA aptamers resolved by NMR spectroscopy. Science *272*, 1343–1347.
24. Nix, J., Sussman, D., and Wilson, C. (2000). The 1.3 angstrom crystal structure of a biotin-binding pseudoknot and the basis for RNA molecular recognition. J. Mol. Biol. *296*, 1235–1244.
25. Fan, P., Suri, A.K., and Fiala, R. (1996). Molecular recognition in the FMN-RNA aptamer complex. J. Mol. Biol. *258*, 480–500.
26. Yoshizawa, S., Fourmy, D., and Puglisi, J.D. (1998). Structural origins of gentamicin antibiotic action. EMBO J. *17*, 6437–6448.
27. Varani, L., Spillantini, M.G., Goedert, M., and Varani, G. (2000). Structural basis for recognition of the RNA major groove in the tau exon 10 splicing regulatory element by aminoglycoside antibiotics. Nucleic Acids Res *28*, 710–719.
28. Jiang, L., Suri, A.K., Fiala, R., and Patel, D.J. (1997). Saccharide-RNA recognition in an aminoglycoside antibiotic-RNA aptamer complex. Chem. Biol. *4*, 35–50.
29. Faber, C., Sticht, H., Schweimer, K., and Rosch, P. (2000). Structural rearrangements of HIV-1 Tat-responsive RNA upon binding of neomycin B. J. Biol. Chem. *275*, 20660–20666.
30. Fourmy, D., Recht, M.I., Blanchard, S.C., and Puglisi, J.D. (1996). Structure of the A site of *Escherichia coli* 16S ribosomal RNA complexed with an aminoglycoside antibiotic. Science *274*, 1367–1371.
31. Zimmermann, G.R., Jenison, R.D., Wick, C.L., Simorre, J.P., and Pardi, A. (1997). Interlocking structural motifs mediate molecular discrimination by a theophylline-binding RNA. Nat. Struct. Biol. *4*, 644–649.
32. Jiang, L.C., and Patel, D.J. (1998). Solution structure of the tobramycin-RNA aptamer complex. Nat. Struct. Biol. *5*, 769–774.
33. Jiang, L.C., Majumdar, A., Hu, W.D., Jaishree, T.J., Xu, W.K., and Patel, D.J. (1999). Saccharide-RNA recognition in a complex formed between neomycin B and an RNA aptamer. Struct. Fold. Des. *7*, 817–827.
34. Sassanfar, M., and Szostak, J.W. (1993). An RNA motif that binds ATP. Nature *3364*, 550–553.
35. Famulok, M. (1994). Molecular recognition of amino acids by RNA-aptamers—an L-citrulline binding RNA motif and its evolution into an L-arginine binder. J. Am. Chem. Soc. *116*, 1698–1706.
36. Wilson, C., Nix, J., and Szostak, J. (1998). Functional require-

ments for specific ligand recognition by a biotin-binding RNA pseudoknot. Biochemistry *37*, 14410–14419.

37. Burgstaller, P., and Famulok, M. (1994). Isolation of RNA aptamers for biological cofactors by in vitro selection. Angew Chem. Int. Ed. *33*, 1084–1087.
38. Wallis, M.G., Vonahsen, U., Schroeder, R., and Famulok, M. (1995). A novel RNA motif for neomycin recognition. Chem. Biol. *2*, 543–552.
39. Wang, S., Huber, P.W., Cui, M., Czarnik, A.W., and Mei, H.-Y. (1998). Binding of neomycin to the TAR element of HIV-1 RNA induces dissociation of Tat protein by an allosteric mechanism. Biochemistry *37*, 5549–5557.
40. Recht, M.I., Fourmy, D., Blanchard, S.C., Dahlquist, K.D., and Puglisi, J.D. (1996). RNA sequence determinants for aminoglycoside binding to an a-site rRNA model oligonucleotide. J. Mol. Biol. *262*, 421–436.
41. Jenison, R.D., Gill, S.C., Pardi, A., and Polisky, B. (1994). High-resolution molecular discrimination by RNA. Science *263*, 1425–1428.
42. Wang, Y., and Rando, R.R. (1995). Specific binding of aminoglycoside antibiotics to RNA. Chem. Biol. *2*, 281–290.
43. Wagner, K., Keyes, E., Kephart, T.W., and Edwards, G. (1997). Analytical Debye-Hückel model for electrostatic potentials around dissolved DNA. Biophys. J. *73*, 21–30.
44. Lipinski, C.A. (2000). Drug-like properties and the causes of poor solubility and poor permeability. J. Pharmacol. Toxicol. Method *44*, 235–249.
45. Amaral, L., and Kristiansen, J.E. (2001). Activity of phenothiazines against antibiotic resistant Mycobacterium tuberculosis: a review supporting further studies that may elucidate the potential use of thioridazine as anti-tuberculosis therapy. J. Antimicrob. Chemother. *47*, 505–511.
46. Puglisi, J.D., Tan, R.Y., Calnan, B.J., Frankel, A.D., and Williamson, J.R. (1992). Conformation of the TAR RNA-arginine complex by NMR spectroscopy. Science *257*, 76–80.
47. Brodsky, A.S., and Williamson, J.R. (1997). Solution structure of the HIV-2 TAR-argininamide complex. J. Mol. Biol. *267*, 624–639.
48. Leach, A.R., and Kuntz, I.D. (1992). Conformational analysis of flexible ligands in macromolecular receptor sites. J. Comp. Chem. *13*, 730–748.
49. Milligan, J.F., and Uhlenbeck, O.C. (1989). Synthesis of small RNAs using T7 RNA polymerase. In Methods in Enzymology, J.E. Dahlberg and J.N. Abelson, eds. (New York: Academic Press), pp. 51–62.
50. Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J., and Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J. Biomolec. NMR *6*, 277–293.
51. Goddard, T.D., and Kneller, D.G. (1998). SPARKY. 3.0. University of California, San Francisco.
52. Redfield, A.G., and Kunz, S.D. (1985). Quadrature fourier NMR detection: simplex multiplex for dual detection and discussion. J. Magn. Reson. *19*, 250–254.