Identification of ligands for RNA targets via structure-based virtual screening: HIV-1 TAR

Anton V. Filikov[a,*], Venkatraman Mohan[b], Timothy A. Vickers[b], Richard H. Griffey[b], P. Dan Cook[b], Ruben A. Abagyan[c] & Thomas L. James[a,**]
[a]*Department of Pharmaceutical Chemistry, University of California, San Francisco, CA 94143-0446, U.S.A.;* [b]*Isis Pharmaceuticals, Inc., Carlsbad, CA 92008-7208, U.S.A.;* [c]*Molsoft, L.L.C., P.O. Box 113, Metuchen, NJ 08840, U.S.A.*

Summary

Binding of the Tat protein to TAR RNA is necessary for viral replication of HIV-1. We screened the Available Chemicals Directory (ACD) to identify ligands to bind to a TAR RNA structure using a four-step docking procedure: rigid docking first, followed by three steps of flexible docking using a pseudobrownian Monte Carlo minimization in torsion angle space with progressively more detailed conformational sampling on a progressively smaller list of top-ranking compounds. To validate the procedure, we successfully docked ligands for five RNA complexes of known structure. For ranking ligands according to binding avidity, an empirical binding free energy function was developed which accounts, in particular, for solvation, isomerization free energy, and changes in conformational entropy. System-specific parameters for the function were derived on a training set of RNA/ligand complexes with known structure and affinity. To validate the free energy function, we screened the entire ACD for ligands for an RNA aptamer which binds L-arginine tightly. The native ligand ranked 17 out of ca. 153,000 compounds screened, i.e., the procedure is able to filter out >99.98% of the database and still retain the native ligand. Screening of the ACD for TAR ligands yielded a high rank for all known TAR ligands contained in the ACD and suggested several other potential TAR ligands. Eight of the highest ranking compounds not previously known to be ligands were assayed for inhibition of the Tat-TAR interaction, and two exhibited a CD_{50} of ca. 1 μM.

Abbreviations: HIV-1, human immunodeficiency virus – type 1; ACD, Available Chemicals Directory; Tat, trans-activating regulatory protein; TAR, RNA transactivation response element; CD_{50}, competitive dose – concentration of compound required to reduce to 50% the binding of protein to RNA; RMSD, atomic root-mean-square deviation.

Introduction

In recent years there has been much effort to design diagnostic and therapeutic agents which bind to protein receptors based on their three-dimensional structure. This effort has been fueled by the increasing availability of protein structures via X-ray crystallog-

raphy and NMR. However, there has been little or no effort to design drugs rationally on the basis of the sequence-dependent three-dimensional structure of DNA or RNA. The reason for this is undoubtedly due to the paucity of three-dimensional structures of possible targets available. However, structure determination methods have advanced to the state that structures of potential DNA and RNA targets are becoming available. Consequently, we can now entertain the concept of designing agents to bind to gene targets

*Present address: Xencor, 2585 Nina Street, Pasadena, CA 91107, U.S.A.
**To whom correspondence should be addressed. E-mail: james@picasso.ucsf.edu

based on the detailed three-dimensional structure of the target.

While the idea of targeting three-dimensional structure has been successfully applied to devise chemotherapeutic ligands to bind proteins, design of nucleic acid-binding ligands has been generally limited to utilizing primary structure, e.g., development of an antisense oligonucleotide to bind a specific stretch of mRNA. At best, the spatial consequences of the primary structure have been indirectly acknowledged in ligand design, e.g., narrowing of the minor groove in AT-rich moieties of a DNA duplex. Very recently, aminoglycosides capable of binding to the standard A-RNA duplex while discriminating against standard B-DNA have been described [1]. These various strategies have had some success, but with the numerous potential RNA and DNA tertiary structure targets, it would be good to develop alternative means to design binding agents. An initial foray in this direction has been recently described: identification of a carbocyanine dye strongly favoring binding to a guanine dimeric hairpin quadruplex over nonquadruplex DNA structures [2].

In principle, combinatorial chemistry provides a route to create and assay a large number of compounds. However, such an approach needs a scaffold upon which to build the many structural variations possible with combinatorial chemistry. We will indeed be very limited if we simply choose to build upon the few nonnucleotide, nonprotein ligands (e.g., aminoglycosides) known to bind RNA or DNA. Instead, we should be able to use computer-aided design methods to provide a list of novel lead compounds for assay and subsequent combinatorial chemistry.

Here we will describe the discovery of agents with the objective of disrupting the essential HIV-1 TAR-Tat interaction necessary for viral replication. This work serves the dual purpose of demonstrating the feasibility of our strategy to target RNA tertiary structure and yields lead compounds capable of binding to an extremely important RNA target with micromolar competitive dose.

Human immunodeficiency virus, type 1, or HIV-1 encodes a transactivating regulatory protein, called Tat, which regulates expression of all viral genes by increasing production of mature, full-length viral RNA [3, 4]. Tat acts by binding to a specific RNA target termed the transactivation response element (TAR). TAR consists of a short RNA stem-loop structure found at the 5′-ends of all nascent lentiviral transcripts. Binding of Tat to TAR is thought to lead to the recruit-

ment of cellular proteins and induce viral transcription [5]. Critical features of both the HIV-1 Tat protein and HIV-1 TAR have been delineated [6–12]. A direct correlation has been found between Tat binding to TAR RNA and the up-regulation of HIV-1 mRNA transcription [13, 14]. Interruption of the Tat-TAR interaction blocks HIV-1 replication in infected cells [15].

There has been significant interest in finding small molecules which target the HIV-1 Tat-TAR interaction. Amino acids or nucleotide-based analogs derived from Tat or TAR [16, 17] and aminoglycoside antibiotics [18] interfere with the Tat-TAR interaction. High toxicity or low bioavailability of these molecules prevent them from becoming viable anti-HIV-1 agents. Benzodiazepines [19] and epoxy steroids [20] can inhibit the cellular function of Tat. However, their antiviral activities can be explained by mechanisms other than inhibiting the Tat-TAR interaction [21]. High-throughput in vitro screening of fairly large compound libraries for inhibition of Tat-TAR interactions followed by assessment of activity in cell studies with Tat-activated reporter gene assays enables screening for new potent anti-HIV-1 agents [18]. Combinatorial chemistry methods combined with in vitro and in vivo assays resulted in discovery of a new hybrid peptoid/peptide oligomer of 9 residues that specifically inhibits the Tat/TAR interaction, both in vitro and in vivo [22]. Experimental determination of the three-dimensional structure of a fragment of the Tat protein [23] enabled use of computational methods for design of small ligands for Tat protein [24]. Although structures of HIV-1 [25] and BIV TAR [26] RNAs are available, there are no published examples of computational approaches for design of ligands for this molecule.

Computational screening of a very large database of compounds via their docking onto a receptor structure is a logical step in drug discovery programs where the receptor structure is available. There have been several successes in using computational database docking methods for discovery of ligands for proteins and RNA [1].

Theoretical prediction of association of a ligand with a receptor requires efficient sampling of the conformational space of the flexible ligand and a sufficiently accurate energy function to score the conformations obtained. In the past few years, several groups have developed docking methods which allow for ligand flexibility. Four different strategies are currently in use for docking of flexible ligands [31]: (a) Monte Carlo or molecular dynamics docking of

complete molecules [32–39], (b) in-site combinatorial search [40–45], (c) site mapping and fragment assembly [41, 42, 46, 47], and (d) evolutionary algorithms [48–51].

With increasing degrees of freedom, searching with either a molecular dynamics or Monte Carlo method in Cartesian coordinates becomes more unreliable due to the necessity of restricted sampling of conformational space [31, 52]. On the other hand, methods which work in internal coordinate space, such as ICM, are orders of magnitude faster [38, 39], since they cut down the number of variables by a factor of 7 on average. They can provide reasonable sampling of conformational space on a time-scale suitable for screening large databases. ICM uses pseudobrownian Monte Carlo minimization in torsion angle space coupled with local minimization in order to globally optimize an energy function. The function can include standard force-field terms (van der Waals, hydrogen bond, electrostatic and torsion angle terms) as well as entropy terms and solvation terms (solvent accessible surface based term, or Poisson, or MIMEL electrostatic terms). This continuous flexible docking method does not restrict the sampling but nevertheless is very fast because it works with torsional coordinates. Both grid and all-atom energy evaluation are available in ICM2.6.

An in-site combinatorial search explores the conformational space of small molecules in discrete steps. The problem is finding the combination of discrete states of constitutive residues that minimizes the energy function. The methods work fast and are successfully used for rough flexible-ligand docking.

Fragment assembly algorithms break a ligand into rigid fragments, align all of the fragments onto the protein target, and merge those fragments whose linkers are close enough to one another. To avoid a combinatorial explosion of the size of the conformational space for very flexible ligands, the search is narrowed down to a few promising head-fragment alignments, then focusing successive fragment alignments onto the linkers to which they must merge. A major drawback of the algorithms is the assumption that aligning isolated fragments on the binding site can provide a reasonable basis for modeling of the ligand conformation.

Evolutionary or genetic algorithms substantially speed up flexible docking by considerably narrowing the sampling of conformational space. However, this often results in a bias towards good hits found initially.

Combinatorial search, fragment assembly and evolutionary algorithms all speed up docking considerably by narrowing down sampling of conformational space due to their underlying simplifications. The methods are sometimes very effective but, because of these simplifications, fail in many cases.

For the current work, we use a four-step docking procedure. We combine a very fast rigid docking (DOCK, version 3.5 [53, 54]), used for a preliminary filtering of the database, and three steps of continuous flexible docking using a Monte Carlo search in torsion angle space (ICM, version 2.6) [38, 39]. Each flexible docking step is performed with progressively more detailed conformational sampling on a progressively smaller list of top-ranking compounds. This combination provides a very fine docking procedure while keeping run times still reasonable. Both grid and all-atom energy evaluation are available in ICM2.6. For the work described here, we used the more precise but slower all-atom energy evaluation. Our method is more precise, but computationally more expensive, than flexible docking procedures which use precalculated energy grids, combinatorial search, fragment assembly or other simplifications. Nevertheless, the method is still fast enough for database screening on common hardware.

To identify potential inhibitors of the Tat-TAR interaction, we have performed virtual screening of *ca.* 153,000 compounds contained in the Available Chemicals Directory (ACD-3D, version 95.1, Molecular Design Limited Information Systems, San Leandro, CA). Validation of our approach to docking flexible RNA ligands was carried out using structures of five known RNA-ligand complexes available in the Brookhaven Protein Data Bank [73]. Subsequent application of the method to HIV-1 TAR RNA led to a list of several potential ligands. Some of the best-ranked compounds were tested for inhibition of the Tat-TAR interaction in a scintillation proximity assay. Two of them exhibited a competitive dose CD_{50} of ca. 1 μM.

Methods

Simulation of known structures

To prepare the structures of the complexes for simulations using the pseudobrownian Monte Carlo search in torsion angle space (ICM, version 2.6), every complex was subjected to the following regularization procedure: the RNA molecule was assembled from ICM library residues and the ligand molecule was built with ideal covalent geometry using ICM tools; every

atom of these two molecules was tethered to the corresponding atom of the experimental structure; the tethers were minimized, then torsion angles for the residues far away from the binding site (typically, more than 3 residues away) were fixed, and a Monte Carlo search of ca. 30,000 steps was performed, with tether energy and restraint energy terms added to the usual force field. The distances between atoms involved in ligand-RNA hydrogen bonds were restrained during the regularization. There are five structures of RNA/ligand complexes, where the ligand is not a protein or oligonucleotide, available in the Brookhaven Data Bank: ATP-binding RNA aptamer in complex with AMP (1raw) [27], RNA aptamer complexed with arginine (1koc) [28], flavin mononucleotide-RNA aptamer complex (1fmn) [29], RNA aptamer complexed with L-citrulline (1kod) [28], and tobramycin-RNA aptamer complex (1tob) [30]. The starting structures for the ligands were obtained by randomization of translational, rotational and torsional coordinates of the ligand for 1raw, 1koc and 1kod complexes, so that the atomic root-mean-square deviation (RMSD) with respect to the experimental ligand conformation was >15 Å. For 1fmn and 1tob complexes, randomization could not be performed because these binding sites are partially closed. So, the ligands in these cases were partially displaced from the experimental positions by means of restrained minimization as far as the binding site permits, and the torsion angles were randomized. A Monte Carlo search was performed with three loose distance restraints, which keep the ligand within 15 Å of the binding site. The energy expression for the Monte Carlo procedure included van der Waals, hydrogen bond, distance-dependent electrostatic and torsion angle terms as well as the restraint energy term, or solvation term based on solvent accessible surface area with distance-independent electrostatic term [38, 39]. The energy expression for local minimization (150 steps performed after every Monte Carlo step) included van der Waals, hydrogen bond, distance-dependent electrostatic and torsion angle terms. Geometrically different (as evaluated by the root- mean-square displacement of the ligand atoms) and low energy conformations were accumulated in a conformational stack as described in Reference 74.

Threading the HIV sequence onto the BIV TAR structure

The 29-nucleotide fragment of HIV TAR RNA examined in one NMR structural study [25] and the
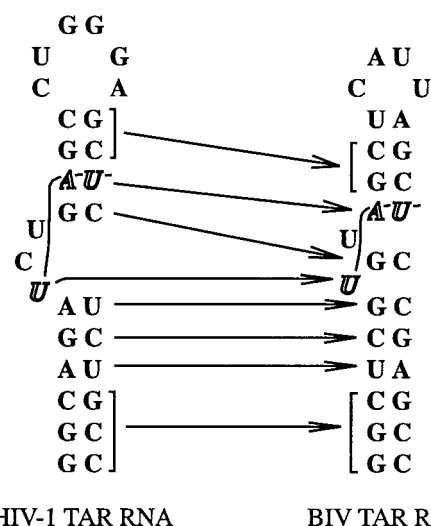

Figure 1. The sequences of the HIV-1 and BIV TAR RNA hairpin loops. The residues involved in base triples are in italics. The alignment used for threading is shown by arrows.

28-nucleotide fragment of BIV TAR RNA used in another NMR structural study [26] are shown in Figure 1. For reasons cited subsequently, we decided to use the BIV structure but with the HIV sequence for our target RNA. The structural model of the bulge moiety of TAR was built by threading the HIV sequence onto the BIV TAR structure using a regularization procedure very similar to that described above, with the alignment shown in Figure 1. For mismatched residues, only N1 and N9 atoms were tethered. After minimization of the tethers, the torsion angles in the first three stem base pairs and in the loop were fixed, and a Monte Carlo search of ca. 30,000 steps was performed, with tether energy and restraint energy terms added to the usual force field. Hydrogen bond distance restraints were imposed for base pairs and for the triple base.

Scintillation proximity assay

The peptide ISIS RP-350 consists of 39 amino acid residues from position 48 to 86 of the HIV-1 Tat protein. The peptide is labeled to high specific activity (100 µCi ml^{-1}) with ^{125}I at Amersham Life Sciences. Streptavidin-coated SPA beads are incubated for 20 min at room temperature in a PRB buffer (50 mM Tris, pH 8.0, 1.5 mM MgCl$_2$, 50 mM KCl, 10% glycerol and 0.01% NP-40) with 0.1 µCi of the labeled peptide and 100 nM ISIS 5832, an RNA oligonucleotide with a 3′-biotin. Incubation can be done in the presence or absence of compounds un-
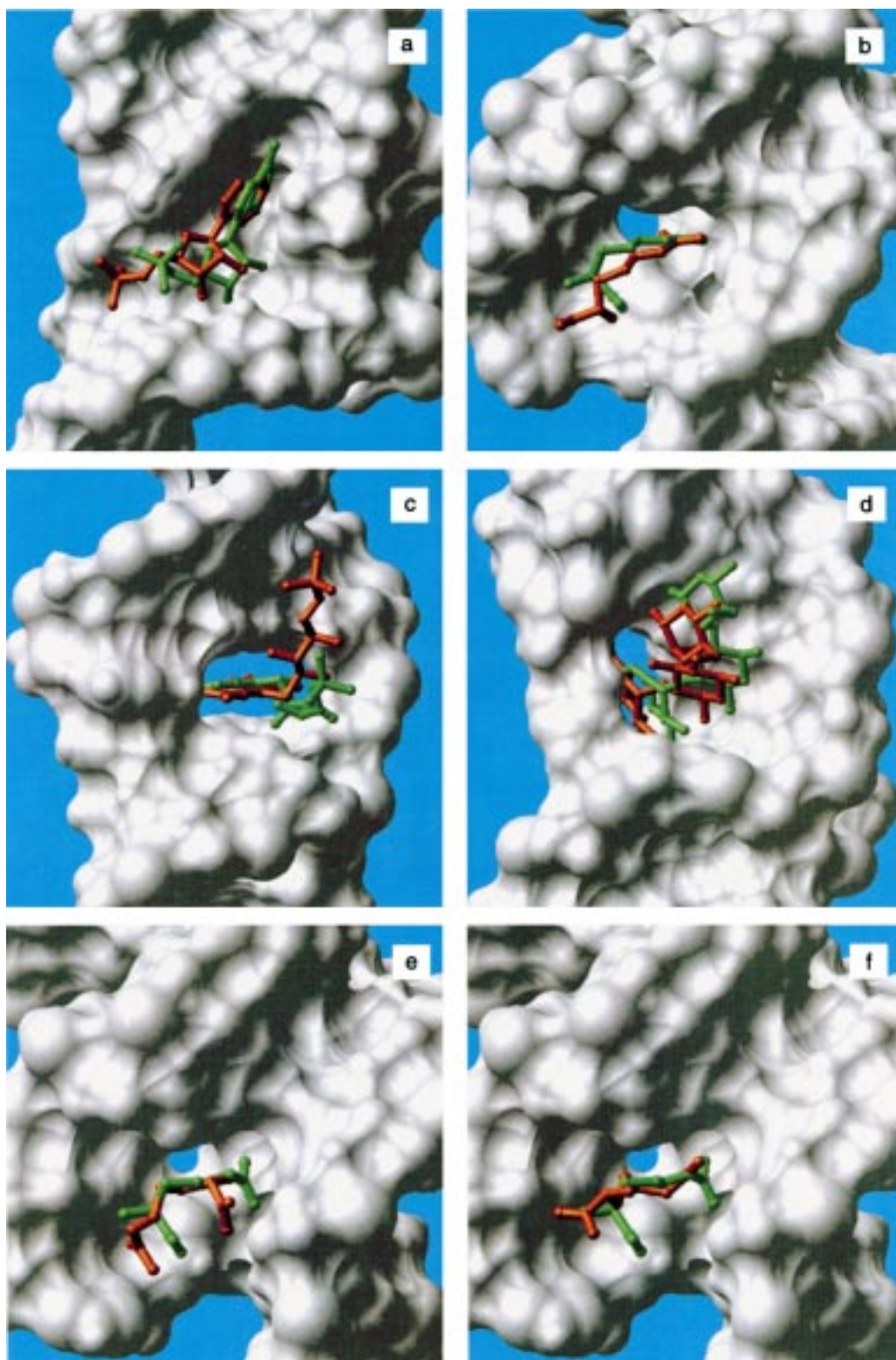
Figure 2. The results of flexible docking of ligands onto rigid RNA receptors for complexes of known structure. (a) 1raw, (b) 1koc, (c) 1fmn, (d) 1tob, (e) 1kod: these simulations were carried out with the solvent modeled via distance-dependent electrostatics; (f) 1kod: the solvent is modeled via atomic solvation parameters, electrostatics is attenuated ($\epsilon = 20$). Simulated conformations are colored red; experimental NMR conformations are green.

der study in a volume of 50 μl in an opaque 96-well plate. Following incubation, the plates are spun at 1000 rpm for 5 min to settle the SPA beads. The biotinylated TAR oligonucleotide binds the streptavidin-coated SPA beads. ^{125}I-Tat peptide associated with the biotinylated TAR oligonucleotide excites the scintillant in the SPA bead, resulting in a quantifiable signal which can be read in the TopCount 96-well scintillation counter. Compounds that interfere with the Tat/TAR interaction result in ^{125}I-Tat floating free in buffer where excited electrons are quenched before transferring energy to scintillant in the SPA bead. This is observed as a decrease in signal. The data were expressed as CD_{50} values (compound concentration required to achieve a 50% decrease in intensity of the signal).

Results and discussion

Validation via simulations of known structures

One can use a variety of possible energy functions with ICM2.6. In order to choose the most appropriate energy function and run-times and to validate the docking technique, we performed simulations of ligand conformations for RNA/ligand complexes of known structure starting from a random ligand conformation. The RNA molecules were kept rigid during the simulation. We used five structures available in the Brookhaven Data Bank. The details on generating the starting conformations can be found in the Methods section.

The results of the simulations are shown in Table 1 and Figure 2. The ligand conformations for 1raw, 1koc, 1fmn and 1tob are reproduced reasonably well with the standard energy function, which includes torsion angle energy term, van der Waals term, hydrogen bond term and distance-dependent electrostatics with $\epsilon = 4r$. The ligand conformation for 1kod is not reproduced well (RMSD = 4.3 Å for all heavy atoms) when using this energy function but approaches the experimental structure (RMSD = 2.5 Å for all heavy atoms) only with the solvent modeled via atomic solvation parameters [55, 56] and with attenuated electrostatics ($\epsilon = 20$). The simulation reaches convergence in only 3 hrs on an SGI R5000/180, because there are many local minima separated by only $\cong 0.3\%$ energy from the global minimum, which corresponds to the conformation consistent with the experimental one. We ran simulations with higher ($\epsilon = 30, 40$) and lower

($\epsilon = 1, 4, 10$) dielectric constants: all of them find the experimental conformation, but only with $\epsilon = 20$ does it correspond to the global energy minimum.

In the case of 1raw, 1koc and 1fmn, the atomic RMSD relative to the experimental conformation for the recognition moiety is more meaningful than total RMSD, because the rest of the molecule is positioned much more approximately in the NMR structures. The sugar-phosphate moiety in 1raw hardly forms specific contacts with the receptor, since chemical modifications in this fragment do not alter the binding much [27]. The different models deposited for the 1fmn structure in the Brookhaven Data Bank reveal that the sugar-phosphate tail of the flavin mononucleotide does not have a determined conformation, i.e., it is inappropriate to compare this part of the molecule with the simulation. In 1koc, the COO^- and NH_3^+ groups switched places relative to the NMR structure (Figure 2b). Again, the groups are positioned fairly arbitrarily in the NMR structure, too, because there are no nonexchangeable protons in these groups, i.e., there are no experimental restraints.

It is informative to examine the relationship between energy and RMSD from the experimental structure. Such a dependence for one complex, 1raw, is shown in Figure 3. The points represent best-energy conformations for fifty conformational families obtained in the Monte Carlo simulation. The algorithm found conformations closer to the experimental structure than the energy minimum at RMSD≈2.3 Å, but these conformations have considerably higher energy.

Because four of five experimental structures were reproduced successfully with the solvent modeled via distance-dependent electrostatics, we decided to use this simpler energy function for automated database docking. In order to rank the ligands after docking is completed, we need a function to estimate their binding free energies.

Binding free energy function for ranking ligands

The energy function used in the flexible docking procedure cannot be used to estimate binding affinity, because it represents the potential energy – not the free energy – of the molecular system. It can only be used for very crude scoring of the ligands. For a more reliable estimation of affinities, we had to develop an empirical method which gives adequate estimates for the molecular systems used in this work, i.e., for complexes of small molecules with RNA. However, even this empirical method will be highly imperfect.

Table 1. Results of flexible docking of ligands on rigid RNA receptors for complexes of known structure taken from the Brookhaven Protein Databank. The ligands are listed according to the code assigned by Brookhaven. The atomic RMSD is calculated with respect to the experimental conformation

PDB code[a]	Starting RMSD (Å)	Energy function	Final RMSD (Å) all heavy atoms	Final RMSD (Å) recognition moiety atoms[b]	Run-time, SGI R5000/180 (min)
1raw	randomized, >15	A[c]	2.3	0.9	50
1koc	randomized, >15	A[c]	3.4	0.55	6
1kod	randomized, >15	A[c]	4.3	3.7	60
1kod	randomized, >15	B[d]	2.5	1.9	180
1fmn	6.4	A[c]	3.2	0.8	20
1tob	5.1	A[c]	2.4	–	20

[a]1raw: ATP-binding RNA aptamer in complex with AMP [27]; 1koc: arginine-RNA aptamer [28]; 1fmn: FMN-RNA aptamer complex [29]; 1kod: RNA aptamer complexed with citrulline [28]; and 1tob: tobramycin-RNA aptamer complex [30].
[b]RMSD values for the recognition moiety atoms are calculated for the adenine in 1raw, guanidino group in 1koc, flavin moiety in 1fmn and side-chain heavy atoms in 1kod.
[e]Torsion angle energy term, van der Waals term, hydrogen bond term and distance-dependent electrostatics with $\epsilon = 4r$.
[d]Torsion angle energy term, van der Waals term, hydrogen bond term and Coulomb electrostatics with $\epsilon = 20$ and solvation term, calculated via atomic solvation parameters [55].
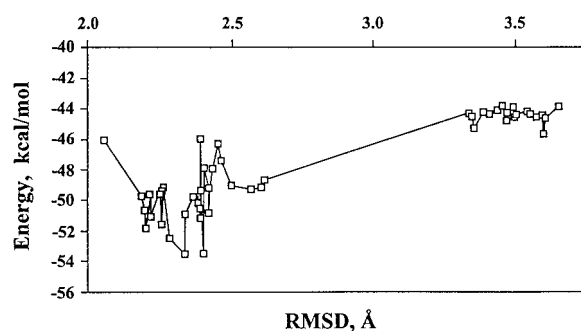

Figure 3. Relationship between energy and atomic RMSD from the experimental structure for docking of AMP on its RNA aptamer (PDB code 1raw). The points represent best energy conformations for fifty conformational families obtained during Monte Carlo simulation.

Figure 4. Thermodynamic cycle used to derive the empirical function for the free energy of binding.

Its disadvantage is that when we use it for screening a large database of potential ligands, there will be several high-ranking compounds which will turn out to be poor ligands when we do experimental assays and we will undoubtedly miss some good ligands. Its advantage is that it may be the best thing available for scoring and, as we will see, it can provide a reasonably effective screening tool such that some good ligands will achieve high scores and will subsequently be tested experimentally. Furthermore, with future knowledge about RNA-ligand interactions, the method can be improved.
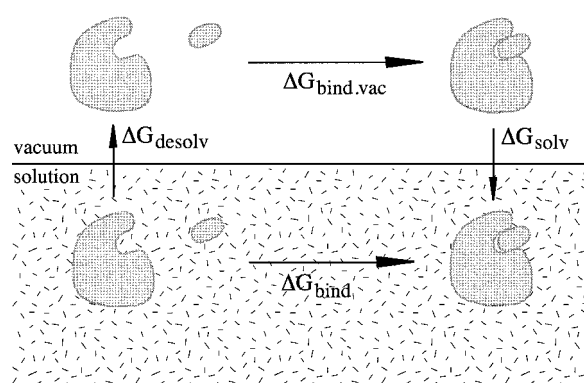
There are a number of empirical methods for estimation of the free energy of binding [24, 38, 45, 57–61, 75, 76]. Most of the methods calculate binding free energy on the basis of additive empirical free energy terms evaluating ΔG, the difference of Gibbs energies between the noncovalent complex and its parts. An alternative approach is based on knowledge-based potentials derived from statistical analysis of crystal structures of protein-ligand complexes [61].

Here we employ an empirical free energy function derived from the thermodynamic cycle shown in Figure 4:

$$\Delta G_{bind} = \Delta G_{desolv} + \Delta G_{bind.vac} + \Delta G_{solv} \quad (1)$$

where ΔG_{bind} is the Gibbs free energy of binding in solution, ΔG_{desolv} is the free energy of desolvation, i.e., of transfering the unbound molecular system to vacuum, $\Delta G_{bind.vac}$ is the free energy of binding in vacuum and ΔG_{solv} is the free energy of solvation, i.e., transfer of the complex back to solvent. $\Delta G_{desolv} + \Delta G_{solv}$ can be calculated via the approach of Wesson and Eisenberg [56]. This method uses atomic solvation parameters derived from vapor-solution transfer experiments performed on small molecules. When a ligand binds to a macromolecular receptor, it displaces partially frozen water molecules from the volume of the binding pocket, i.e., these water molecules gain entropy; we call this a cavity entropy effect. This entropy gain is probably not fully accounted for when $\Delta G_{desolv} + \Delta G_{solv}$ is calculated via atomic parameters because they are derived from experiments with small molecules: they do not have cavities. The entropy gain can be very high: measurements of sorption isotherms of water vapor on solid ovalbumin at several temperatures give values up to 6 kcal/mol [62]. We calculate $\Delta G_{desolv} + \Delta G_{solv}$ as a sum of a solvation term, calculated via the atomic parameters of Abagyan [55], and a term $\Delta G_{cav\ entropy}$, which accounts for this cavity entropy effect:

$$\Delta G_{desolv} + \Delta G_{solv} = \Sigma_i \Delta S_i \sigma_i + \Delta G_{cav\ entropy}. \quad (2)$$

where Σ_i means summation over all nonhydrogen atoms in the molecular system, ΔS_i is the difference of the solvent accessible surface area of atom i in the bound and in the unbound state, and σ_i is a solvation parameter for atom i. For the purpose of calculating this solvation term, we distinguish eleven atom types and use the solvation parameters of Abagyan [55] shown in Table 2. Hydrogens are ignored in this term.

The free energy of binding in vacuo can be partitioned into the following terms:

$$\Delta G_{bind.vac} = \Delta G_{el} + \Delta G_{vdW} + \Delta G_{hb}$$
$$+ \Delta G_{isomer} + \Delta G_{tr/rot} + \Delta G_{tors} \quad (3)$$

where ΔG_{el}, ΔG_{vdW} and ΔG_{hb} account for intermolecular interactions: they are respectively electrostatic interaction, van der Waals interaction and hydrogen bond terms; ΔG_{isomer} is a conformational isomerization term which accounts for the change of the

Table 2. Atomic solvation parameters [55] used in the empirical free energy function

σ (cal/(mol Å^2))	Radius (Å)	Atom type
10	1.95	C aliphatic
−9	1.8	C aromatic
−163	1.7	N uncharged
−280	1.7	N^+, N_ξ in Lys$^+$
−220	1.7	$N_{\eta 1}$, $N_{\eta 2}$ in Arg$^+$
−114	1.6	O hydroxyl
−64	1.4	O carbonyl
−280	1.4	O^- in Glu, Asp
−174	1.4	O in COOH
−22	2.0	S in SH
−92	1.85	S in Met or S-S

molecular internal energies on binding; it includes both enthalpic and entropic components: this term can be called intramolecular 'strain' (B. Honig, oral presentation at 'Molecular Recognition in Drug Design: Docking and Scoring', February 6–7, 1998, San Francisco). $\Delta G_{tr/rot}$ is a term accounting for three translational and three rotational degrees of freedom lost by the ligand on binding, and ΔG_{tors} is a term which accounts for the loss of bond configurational entropy due to freezing of some torsion angles upon complex formation. Note that we do not isolate into a separate term the change in torsional enthalpy on binding: it is included in ΔG_{isomer}.

The terms ΔG_{el}, ΔG_{vdW} and ΔG_{hb} are macroscopic values, i.e., averages over molecular ensembles. To estimate them, we assume that these terms are proportional to corresponding potential energy terms ΔE_{el}, ΔE_{vdW} and ΔE_{hb} calculated for the minimized conformation of a single molecular complex, i.e., to microscopic values. The proportionality factor we obtain by fitting to experimental binding data. To estimate the potential energy terms, we use a force field [63] with RNA charges taken from Veal and Wilson [64], Gasteiger and Marsili charges for the ligands calculated with Sybyl 6.3, and a dielectric constant $\epsilon = 1$ since we are calculating binding in vacuo. The values calculated for these terms are very approximate, since the force field was developed so that some physical properties are reproduced but not the binding energies. The least accurate term is ΔE_{el}, since we only roughly know the charges and have a very crude figure for the dielectric constant. The inaccuracy of these terms provides another reason why they
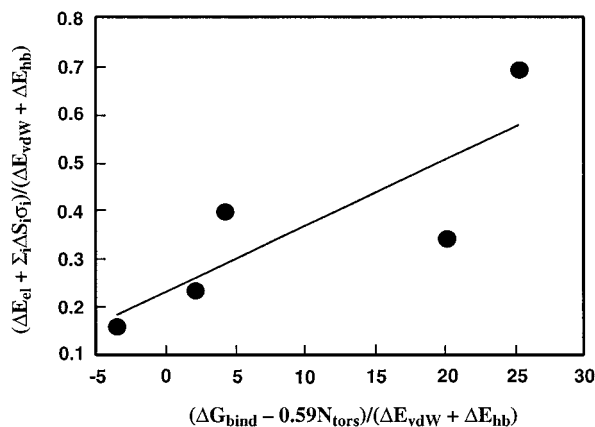
Figure 5. Linear regression of experimental binding data with Equation 6 results in the following values for the empirical parameters: $a = 0.014 \pm 0.005$, $b = 0.22 \pm 0.08$.

should be scaled by fitting to experimental data. The sum of these three terms represents an estimate of the major part of the enthalpy of interaction; another part of the enthalpy is in ΔG_{isomer}, but this is most probably a smaller part except when we deal with a major rearrangement of the molecules upon binding. These terms are microscopic values. Using them we try to estimate the enthalpy, which is a macroscopic function, i.e., an average over a molecular ensemble. This provides another reason why these terms should be scaled by fitting to experimental data.

The estimation of $\Delta G_{desolv} + \Delta G_{solv}$ through atomic solvation coefficients is very approximate, too; we simply do not have a better method at present. So, we come to the following equation, which uses the parameter a to be determined by fitting to experimental binding data:

$$\Delta G_{bind} = (\Delta E_{el} + \Delta E_{vdW} + \Delta E_{hb} + \Sigma_i \Delta S_i \sigma_i) * a$$
$$+ \Delta G_{cav\ entropy} + \Delta G_{tr/rot}$$
$$+ \Delta G_{isomer} + \Delta G_{tors} \qquad (4)$$

where ΔE_{el}, ΔE_{vdW} and ΔE_{hb} are the differences in corresponding potential energies between the unbound and bound states calculated via the force field.

The calculation of ΔG_{isomer} and $\Delta G_{cav\ entropy}$ presents a major problem; at present, there are no methods which can reliably cope with this challenge. Here we use the following approximation. We assume that the sum $\Delta G_{cav\ entropy} + \Delta G_{tr/rot} + \Delta G_{isomer}$ is proportional to the core interaction energy, represented by the following term: $(\Delta E_{el} + \Delta E_{vdW} + \Delta E_{hb} + \Sigma_i \Delta$

$S_i\ \sigma_i) * a$. Thus we can rewrite Equation 4 as:

$$\Delta G_{bind} = (\Delta E_{el} + \Delta E_{vdW} + \Delta E_{hb} + \Sigma_i \Delta S_i \sigma_i) * a$$
$$+ \Delta G_{tors}. \qquad (5)$$

Parameter a now incorporates $\Delta G_{cav\ entropy} + \Delta G_{tr/rot} + \Delta G_{isomer}$ as well. Although a larger isomerization energy is not always compensated by a larger interaction energy, and a larger interaction energy does not always mean more water molecules are pushed out, we utilize this assumption as a crude approximation.

We calculate ΔG_{tors} by multiplying the number of rotatable bonds fixed on complex formation by a coefficient 0.59 kcal/mol (1 cal = 4.19 J), which we calculated by averaging the data on conformational entropy of amino acid side chains [65].

The most approximate terms in Equation 5 are ΔE_{el} and $\Sigma_i \Delta S_i \sigma_i$; there are large uncertainties in the parameters determining these terms. Because of this, we choose to weigh these terms separately from ΔE_{vdW} and ΔE_{hb} by using a second parameter:

$$\Delta G_{bind} = (\Delta E_{el} + \Sigma_i \Delta S_i \sigma_i) * a$$
$$+ (\Delta E_{vdW} + \Delta E_{hb}) * b$$
$$+ 0.59 * N_{tors}, \qquad (6)$$

where N_{tors} is the number of rotatable bonds in the ligand molecule which become fixed upon complex formation.

The binding constants are available for all five NMR structures of small ligands complexed with RNA that are deposited in the Brookhaven Data Bank. Consequently, we derived the parameters a and b by linear regression of Equation 6 for these structures (Figure 5). It results in the following values for the parameters: $a = 0.014 \pm 0.005$, $b = 0.22 \pm 0.08$. The correlation of the experimental points and the predicted energies is reasonable ($R = 0.83$). The standard error for the calculated binding free energy is ± 2.7 kcal/mol. As more structures of RNAs complexed with small ligands become available, the parameters can be further refined.

The fact that ΔE_{el} is multiplied by a small factor in Equation 6 may not be unexpected, but it may seem surprising that $\Sigma_i \Delta S_i \sigma_i$ is also multiplied by the same factor: this term is based on atomic solvation parameters that are fitted to transfer free energy, a directly measured thermodynamic quantity. There are at least two reasons for scaling this term. The atomic solvation parameters are derived for derivatives of amino acids. We do not know parameters for RNA: we must use protein parameters as an approximation. The second reason is simple: we would like to introduce a

Table 3. Comparison of experimental free energies of binding for RNA/ligand complexes of known structure with those calculated via Equation 6 with coefficients $a = 0.014$ and $b = 0.22$

PDB code	ΔG_{bind} experiment (kcal/mol)	ΔG_{bind} calculated by Equation 6[f] (kcal/mol)	ΔG_{bind} calculated by Equation 7[g] (kcal/mol)
1raw	−6.8[a]	−7.2	−7.4
1koc	−6.8[b]	−10.8	−10.1
1kod	−6.8[c]	−3.6	−4.2
1fmn	−7.2[d]	−7.8	−7.6
1tob	−12.4[e]	−9.4	−8.8

[a]Reference 71.
[b]Reference 28.
[c]Reference 28.
[d]Reference 29.
[e]Reference 72.
[f]The standard error for the calculated binding free energy is ±2.7 kcal/mol.
[g]The standard error for the calculated binding free energy is ±2.5 kcal/mol.

third parameter and scale ΔE_{el} and $\Sigma_i \Delta S_i \sigma_i$ differently, but with only five experimental points available at this time we cannot use more than two parameters. As more structures and binding constants become available, a third parameter can be introduced which should significantly improve calculated affinities. For proteins many more experimental points are available. So, when applying Equation 6 for protein complexes, a third parameter should certainly be used.

The calculated and experimental free energies of binding for these five complexes are compared in Table 3. For these complexes, we counted the number of ligand rotatable bonds fixed on binding (N_{tors}) by inspecting the complexes visually. For purposes of automatic database docking, this step must be excluded, i.e., we must use the number of ligand rotatable bonds $N_{rot.b.}$ instead of N_{tors}. Correspondingly, the coefficient 0.59 in the conformational entropy term should be decreased because (a) symmetrical groups should not be counted: switching between their rotamers produces indistinguishable states; (b) rotatable bonds for groups like -OH or -SH should not be counted either since the hydrogen atom is small, and these groups usually have enough room to rotate; (c) some of the rotatable bonds in the ligand are already partially fixed in the unbound state, and some of them are not completely fixed in the bound state: the actual loss of entropy is much less than in the transition from a completely free to a completely fixed state. Consequently, we use an empirical value of 0.3 for the coefficient.

Therefore, the actual equation used for scoring ligands from the database is:

$$\Delta G_{bind} = (\Delta E_{el} + \Sigma_i \Delta S_i \sigma_i) * 0.014 + (\Delta E_{vdW} + \Delta E_{hb}) * 0.22 + 0.3 * N_{rot.b.} \quad (7)$$

where $N_{rot.b.}$ is the number of rotatable bonds in the ligand. A drawback of this energy function is the unavailability of solvation coefficients for atoms other than C, O, N, and S. So ligands which contain other atoms have to be treated separately; at present, we ignore compounds which contain other atoms. Inclusion in the future will require that we have some complexes with other atoms for our training set. The energy function must be parameterized on a training set of complexes. The more similar the complexes in the training set are to the target complex, the better the prediction will be. If no data on similar complexes is available, we must use some general parameters; the prediction will be more approximate and it is expected that the ranking of ligand binding will be less accurate.

Validation of the empirical free energy function (Equation 7) via screening for a known ligand

As a validation test for the empirical free energy function (Equation 7) and for the docking protocol, we performed screening of the ACD for ligands for the L-arginine RNA aptamer. This complex was chosen from the five RNA complexes with small molecules available in the Brookhaven database, because this is the only one which has a rather tight ligand-binding pocket fit.

The program DOCK [53, 54], version 3.5, was utilized for rigid docking of compounds from the Available Chemicals Directory (ca. 153,000 compounds) to provide preliminary filtering of the database. Fifty overlapping spheres were generated in order to describe the solvent-accessible surface of the binding site. The cluster of spheres was manually edited to achieve uniform coverage of the binding site. Spheres outside the binding site were eliminated. Contact scoring without normalization for number of atoms was used. No chemical matching was used. Five steric clashes per compound were allowed. L-arginine showed up with the rank of 8728 out of 153,000. The highest scoring compounds (ca. 20% of the database) were subsequently subjected to flexible docking using ICM, version 2.6 [38, 39] with evaluation via the empirical free energy function we developed (Equation 7). Run times at this step averaged ca. 1.5 min per molecule on an SGI R5000/180.
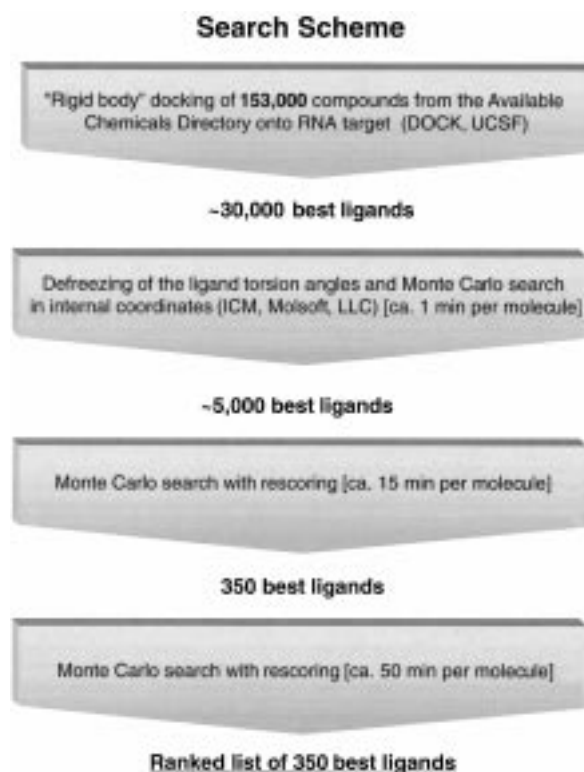
Search Scheme

"Rigid body" docking of **153,000** compounds from the Available Chemicals Directory onto RNA target (DOCK, UCSF)

~30,000 best ligands

Defreezing of the ligand torsion angles and Monte Carlo search in internal coordinates (ICM, Molsoft, LLC) [ca. 1 min per molecule]

~5,000 best ligands

Monte Carlo search with rescoring [ca. 15 min per molecule]

350 best ligands

Monte Carlo search with rescoring [ca. 50 min per molecule]

Ranked list of 350 best ligands

Figure 6. Flowchart of the docking procedure. Numerical values correspond specifically to targeting the arginine aptamer.

The best scoring 20% of these ligands were redocked with ten times longer sampling per molecule. Following this, 350 compounds were redocked again with sampling of ca. 50 min per molecule, which corresponds to ca. 50,000 conformations (energy function calls) per molecule or ca. 3300 Monte Carlo steps. This three-step calculation took about a week on ten SGI R5000/180 machines running in parallel. This sampling typically ensures convergence for compounds with 7–8 rotatable bonds. For every run of flexible docking, the initial conformations of the ligands were obtained by randomizing torsion angles and rotational coordinates and by randomizing the translational coordinates within a 10 Å-diameter sphere around the center of the binding site. This three-step flexible docking procedure improved the rank of L-arginine to 195. Figure 6 shows the flowchart for the docking procedure. The algorithm is not able to recognize that the guanidino group is planar, i.e., its bonds are not rotatable. With rotatable bonds, arginine finds a different energy minimum with significantly worse binding energy. If the bonds in the guanidino group are fixed, which requires intervention of the operator,

L-arginine ends up in a conformation virtually identical to the one obtained in the section 'Validation via simulation of known structures' with an RMSD of 3.4/0.55 Å (all heavy atoms/recognition moiety heavy atoms). On the scoring list it moves to rank 17 with $\Delta G_{calculated} = -11.08$. (The experimental value is -6.8 kcal/mol [28]). Note that this value differs from that in Table 3 (-10.8 kcal/mol), because it was calculated by Equation 7 for automated database docking rather than by Equation 6 used for calculating values in Table 3. The ten best ranking compounds are shown in Figure 7. D-arginine showed up at the rank of 3672, and L-citrulline at the rank of 13260. Some of the compounds in the database can have multiple charge states and protoisomers. For purposes of automatic database docking, we only consider the protonation states present in the database.

The empirical free energy function (Equation 7) could be compared with functions available in the literature. Here we limit ourselves to comparing results using Equation 7 with the simplest and most popular functions. Figure 8 shows the correlation of the energy calculated by Equation 7 for the 50 best ligands (TAR RNA run) with: (a) van der Waals interaction energy (R = 0.50), (b) the sum of van der Waals, hydrogen bonding and Coulomb electrostatic ($\epsilon = 4$) terms (R = 0.38), and (c) the sum of van der Waals, hydrogen bonding and Coulomb electrostatic ($\epsilon = 1$) terms (R = 0.27). As we see the correlation is poor. The correlation is also poor with the Coulomb electrostatic interaction term only (R = 0.23) and with the hydrogen bonding term (R = 0.10). The poor correlation, of course, demonstrates that results from function (7) are distinct from the simpler functions; it does not prove that function (7) is better though. We believe that taking into account such major effects as solvation, isomerization free energy and conformational entropy with scaling factors weighed using experimental data, not only dramatically rescores the list of ligands, but dramatically improves the 'hit rate'. The 'hit rate' would appear to be rather good from the limited sample of compounds experimentally tested (vide infra), but the extensive calculations entailed in direct studies on comparison of 'hit rates' achieved with different scoring functions should be the subject of a separate work.

To understand how much flexible docking improves the result, we need to rank conformations obtained by DOCK with the same formula we use for ranking ligands after flexible docking. The rank of L-arginine improves from 8728 to 6347 when calculated

ACD Code	Structure	Predicted ΔG_{bind} (kcal/mol)
00047650		-13.3
00154115		-13.0
00154110		-12.8
00027918		-12.8
00083503		-12.8
00151463		-12.6
00187126		-12.2
00044544		-11.9
00085579		-11.6
00037067		-11.5

Figure 7. The ten highest ranking compounds resulting from screening the ACD for L-arginine RNA aptamer ligands. L-Arginine itself (with guanidino group torsion angles being fixed) has a rank of 17 out of the 153,000 compounds in the ACD, with $\Delta G_{calculated} = -11.08$ kcal/mol.

by Equation 7. Rigorous comparison of rigid and flexible docking is beyond the scope of the current work: we only used rigid docking for crude and quick preliminary filtering of the database. Most probably the rank of L-arginine after rigid docking can be improved by using more extensive sampling and tuning parameters with the DOCK algorithm. The rigid docking can, however, rank a known ligand very low. Some good inhibitors may be filtered out in the early stage of docking. The severity of the problem is, of course, system-dependent.

As we see, the method is quite effective for database screening for RNA ligands: with reasonable runtimes the final list of compounds can be thousands of times smaller than the original database and still contain the native ligand.

Structural model for HIV-1 TAR RNA

Figure 1 shows the 29-nucleotide fragment of HIV TAR RNA used in an NMR structural study [25] and the 28-nucleotide fragment of BIV TAR RNA studied [26]. The HIV TAR RNA contains a hairpin loop with
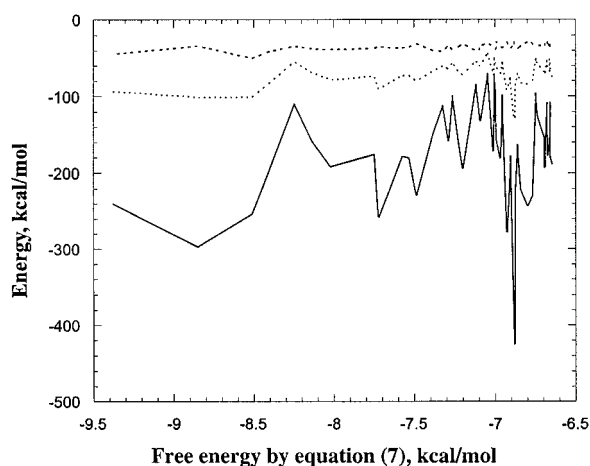
Figure 8. Correlation of the energy calculated by Equation 7 for the 50 best ligands, from the TAR RNA run, with: dashed line – van der Waals interaction energy (R = 0.50), dotted line – the sum of van der Waals, hydrogen bonding and Coulomb electrostatic ($\epsilon = 4$) terms (R = 0.38), and solid line – the sum of van der Waals, hydrogen bonding and Coulomb electrostatic ($\epsilon = 1$) terms (R = 0.27).

a three-base bulge in the stem. Biochemical studies of HIV Tat demonstrated that one arginine residue is very important for specific binding [6]. NMR studies have shown that the RNA changes conformation upon peptide binding and that an argininamide induces a similar change, although a higher concentration is required than for the basic peptide [66]. A structural model was developed in which the guanidinium group of the arginine is bound at the three-base bulge, contacting two phosphates and the edge of a base. In addition, there was evidence for the formation of a base triple UAU adjacent to the bulge site [66]. Recently, a more detailed study of TAR RNA complexed with a Tat-derived peptide was carried out, which supports the site and mode of binding of the arginine but disputes the formation of the base triple [25]. Unfortunately, the peptide resonances were not assigned in the study. As a result, the overall geometry of the RNA structure is not well-defined: the size of the peptide binding site (the major groove at the bulge region) varies by almost a factor of two among the twenty models deposited in the Brookhaven Data Bank. While it is commendable that the authors are conservative in defining the conformation of their subject RNA, this level of precision makes it very difficult to do structure-based ligand design; the major groove at the bulge region is obviously the most interesting region for the design.

The BIV Tat-TAR complex is quite similar to that of HIV. Biochemical studies revealed that the RNA binding site for the Tat protein is again a bulge-

containing hairpin, but with a different type of bulge (Figure 3). Complexes of 14- and 17-residue peptide fragments of BIV Tat bound to a 28-nucleotide TAR RNA have been characterized in detail by NMR [26, 67]. Extensive assignments of intermolecular contacts have been carried out, resulting in a well-defined structure of the complex [26]. As with the HIV Tat-TAR interaction, there is a crucial arginine residue – the R73 sidechain is positioned near the G11-C25 base pair, and this residue hydrogen bonds both to the edge of G11 and to the phosphate of the 9–10 step, reminiscent of the arginine fork described in the HIV system [6]. BIV TAR forms the base triple U10-A13-U24. Although the bulge of the BIV TAR differs from that of HIV, the geometries of the bulge regions are very similar for both structures when bound to the peptide.

NMR studies of both free and complexed HIV TAR RNA came to different conclusions regarding formation of a base triple [25, 66]. If the base triple forms, this creates a unique structure, so that if a ligand for this site is designed, it most probably would be very specific for this RNA. Consequently, for the purposes of drug design, the structure with the base triple is clearly preferable. Even if the base triple is not formed in the peptide- or protein-bound structure but constitutes a low-energy structure, i.e., it exists at least part of the time in solution or can be easily induced by the appropriate ligand, it is still preferable to use it for ligand design for reasons of specificity.

So, we decided to use a base triple-containing structure for the ligand design. The structural model of the bulge fraction of TAR was built by threading the HIV sequence on to the BIV TAR structure using a regularization procedure [68] with the alignment shown in Figure 1.

Screening for TAR RNA ligands

Screening of the Available Chemicals Directory for TAR RNA ligands was performed using the same protocol as used for docking of the arginine aptamer (vide supra). The final list of potential ligands generated by the procedure contains all known ligands for TAR RNA. L-arginine [experimental $K_i = 4$ mM [69]] appears at a relative rank of 1.0%, and spermine (a generic ligand for nucleic acids) is at 0.22% [relative rank $= 100 *$ rank/(number of compounds in database)]. All aminoglycoside antibiotics, which are known to be ligands for TAR RNA, appear at reasonable ranks: neomycin, 1.1% [experimental IC$_{50}$ = 0.92 μM]; gentamycin, 1.0% [experi-

ACD Code	Structure	Predicted ΔG_{bind} (kcal/mol)
00102170		-9.4
00195005		-8.8
00192509		-8.5
00135881		-8.4
00046964		-8.2
00032135		-8.1
00102404		-8.0
00190628		-7.7

Figure 9. The highest ranking compounds, with their scores, from screening of the Available Chemicals Directory for HIV-1 TAR RNA ligands.
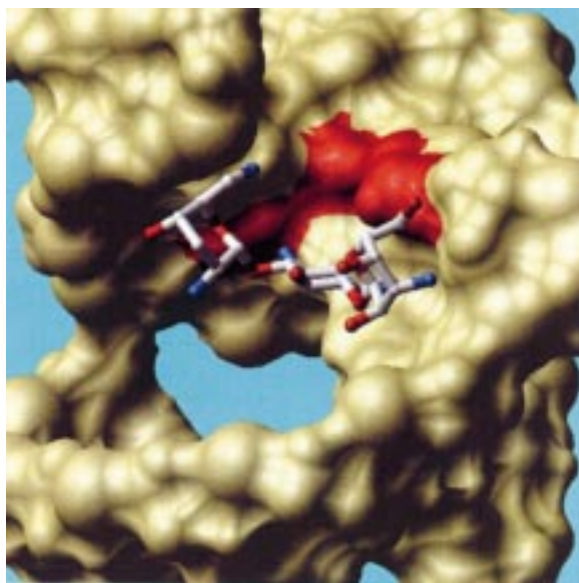
Figure 10. The conformation of tobramycin docked onto the TAR RNA structural model resulting from the docking procedure. It is interesting to compare the results obtained at every stage of the docking procedure. Figure 11 compares the ranks obtained by the best 35 compounds (TAR RNA run), following the final run of 50 min Monte Carlo per molecule, with ranking obtained at earlier stages in the docking procedure. Correlation increases with each stage of docking: R = 0.025 for rigid docking, R = 0.39 for 1.5 min of Monte Carlo per compound, R = 0.46 for 15 min of Monte Carlo per compound. If a large number of compounds is to be screened experimentally, the final CPU-intensive Monte Carlo run we used here is not necessary; much shorter times suffice.

Figure 11. Ranks obtained at various stages in the docking procedure against TAR RNA for the 35 compounds achieving the highest ranking following the final run of 50 min Monte Carlo simulation per molecule. The thin solid line is for rigid docking (R = 0.025), the sparsely dotted line is for 1.5 min Monte Carlo simulation per compound (R = 0.39), and the heavily dotted line is for 15 min Monte Carlo simulation per compound (R = 0.46). This set of calculations was carried out on a fraction of the compounds in the Available Chemicals Directory (12,000 compounds out of 153,000); the results should be typical of the entire ACD.

mental IC_{50} = 45 μM]; streptomycin, 0.9% [experimental IC_{50} = 9.5 μM]; paromomycin, 1.7%; kanamycin 1.3%; neamine, 0.29%; tobramycin, 0.04%, lividomycin 2.7% and sisomicin 0.14%. The latter six antibiotics inhibit Tat peptide binding to antibiotics inhibit Tat peptide binding to TAR RNA at a concentration of 3 mM [70]; Benkamycin [3 mM inhibitor [70]] was not considered, because it was absent from the version of the ACD we used. Figure 9 shows the highest ranking compounds with their scores. Figure 9 shows that many compounds predicted to be good ligands for TAR RNA are likely to be false positives as drug candidates because of their chemical nature: too many net charges or too polar to be orally active. In other words they are not 'drug-like'. In the current study, we only search for compounds that disrupt TAR-tat binding and ignore bioavailability, specificity, toxicity and other desirable properties for drug leads. Naturally, a ranked list would need to be subsequently examined for these properties as well before subsequent studies of lead compounds were
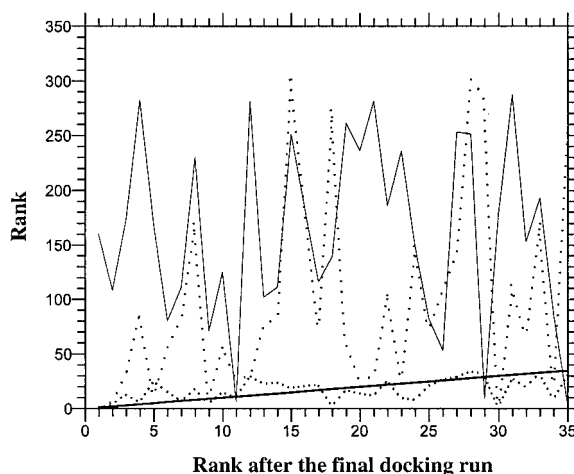
performed. Figure 10 shows the conformation of tobramycin docked on the TAR RNA structural model resulting from the procedure.

It is interesting to compare the results obtained at every stage of the docking procedure. Figure 11 compares the ranks obtained by the best 35 compounds (TAR RNA run), following the final run of 50 min Monte Carlo per molecule, with ranking obtained at earlier stages in the docking procedure. Correlation increases with each stage of docking: R = 0.025 for rigid docking, R = 0.39 for 1.5 min of Monte Carlo per compound, R = 0.46 for 15 min of Monte Carlo per compound. If a large number of compounds is to be screened experimentally, the final CPU-intensive Monte Carlo run we used here is not necessary; much shorter times suffice.

A large number of compounds in the ACD are insoluble. Nevertheless, we used all of them for computational screening. Only compounds soluble in water (or at least water-miscible solvents) can be readily tested experimentally, but promising compounds insoluble in water can serve as precursors for chemical modifications which may make them soluble. These modifications can be designed computationally, so they do not adversely affect binding.
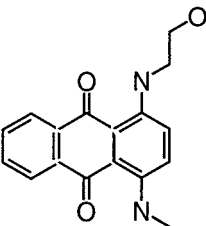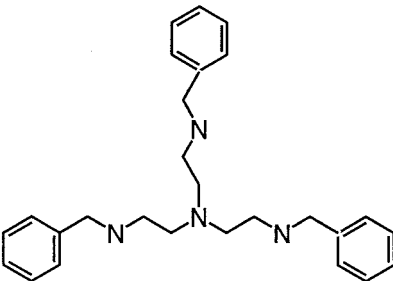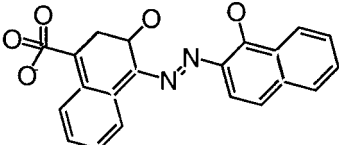
ACD Code	Structure	Predicted ΔG_{bind} (kcal/mol)	Exptl CD_{50} (μM)
00001199		-5.1	ca. 1
00192509		-8.5	ca. 1
00003934		-5.1	ca. 50

Figure 12. Inhibition of Tat-TAR interaction, assayed experimentally, by several compounds which had a high ranking in the virtual screening procedure.

Experimental test for inhibition of Tat-TAR interaction

A small number of the higher ranking compounds were selected to provide an experimental test of our procedure. The compounds were selected for assay before the final docking run was performed. We selected readily available, soluble and nonpoisonous high-ranking compounds from the 15-min-of-Monte Carlo-per-compound run. A scintillation proximity assay (see Methods) was used to investigate inhibition of the Tat-TAR interaction by eight compounds which achieved high ranks in the virtual screening procedure. Additional compounds will be tested experimentally. The results of the inhibition study are shown in Figure 12. Two of the compounds tested exhibited a CD_{50} (concentration of compound required to reduce to 50% Tat binding to TAR) of about 1 μM. We conclude that the docking method we have devised is quite effective for database screening for RNA ligands. It is interesting to examine the rank order of these three

compounds at each stage in the virtual screening procedure. The ligands, listed by ACD code, received the following ranks at the four stages (DOCK, MC1.5, MC15, MC50) of docking:

> 00001199 : 295 (MC50), 28 (MC15),
> 3417 (MC1.5), 6171 (DOCK),
> 00192509 : 3 (MC50), 3 (MC15),
> 2720 (MC1.5), 6919 (DOCK),
> 00003934 : 286 (MC50), 34 (MC15),
> 1724 (MC1.5), 1938 (DOCK).

It is doubtful that any of these three would have been included in an experimental screen following the initial two docking steps unless a rather large screen was performed. In an industrial setting, screening several hundred of the highest ranking compounds which were readily available and soluble would certainly be feasible.

Those ligands which exhibit micromolar CD_{50} can serve as leads for drug design. The assay we use tests the ability of compounds to inhibit binding of the tat

peptide to TAR RNA rather then their ability to bind to TAR RNA. Strictly speaking therefore, we cannot claim that the compounds actually bind to TAR in the manner anticipated. The issue of specificity of binding to TAR RNA versus other compounds and other nucleic acids will be the subject of a separate study.

Acknowledgements

References

1. Chen, Q., Shafer, R.H. and Kuntz, I.D., Biochemistry, 36 (1997) 11402.
2. Chen, Q., Kuntz, I.D. and Shafer, R.H., Proc. Natl. Acad. Sci. USA, 93 (1996) 2635.
3. Kao, S.-Y., Calman, A.F., Luciw, P.A. and Peterlin, B.M., Nature, 330 (1987) 489.
4. Marciniak, R.A. and Sharp, P.A., EMBO J., 10 (1991) 4189.
5. Carroll, R., Peterlin, B.M. and Derse, D., J. Virol., 66 (1992) 2000.
6. Calnan, B.J., Tidor, B., Biancalana, S., Hudson, D. and Frankel, A.D., Science, 252 (1991) 1167.
7. Karn, J., Dingwall, C., Finch, J.T., Heaphy, S. and Gait, M.J., Biochimie, 73 (1991) 9.
8. Loret, E.P., Georgel, P., Johnson, W.C. Jr. and Ho, P.S., Proc. Natl. Acad. Sci. USA, 89 (1992) 9734.
9. Hauber, J., Malim, M.H. and Cullen, B.R., J. Virol., 63 (1989) 1181.
10. Weeks, K.M., Ampe, C., Schultz, S.C., Steitz, T.A. and Crothers, D.M., Science, 249 (1990) 1281.
11. Aboul-ela, F., Karn, J. and Varani, G., Nucleic Acids Res., 24 (1996) 3974.
12. Puglisi, J.D., Chen, L., Frankel, A.D. and Williamson, J.R., Proc. Natl. Acad. Sci. USA, 90 (1993) 3680.
13. Delling, U., Reid, L.S., Barnett, R.W., Ma, M.Y., Climie, S., Sumner-Smith, M. and Sonenberg, N., J. Virol., 66 (1992) 3018.
14. Dingwall, C. et al., EMBO J., 9 (1990) 4145.
15. Johnston, M.I. and Hoth, D.F., Science, 260 (1993) 1286.
16. Sullenger, B.A., Gallardo, H.F., Ungers, G.E. and Gilboa, E., J. Virol., 65 (1991) 6811.
17. Ecker, D.J., Vickers, T.A., Bruice, T.W., Freier, S.M., Jenison, R.D., Manoharan, M. and Zounes, M., Science, 14 (1992) 958.
18. Mei, H.-Y., Galan, A.A., Halim, N.S., Mack, D.P., Moreland, D.W., Sanders, K.B., Truong, H.N. and Czarnik, A.W., Bioorg. Med. Chem. Lett., 5 (1995) 2755.
19. Hsu, M.-C., Schutt, A.D., Holly, M., Slice, L.W., Sherman, M.I., Richman, D.D., Potash, M.J. and Volsky, D.J., Science, 254 (1991) 1799.
20. Michne, W.F., Schroeder, J.D., Bailey, T.R., Young, D.C., Hughes, J.V. and Dutko, F.J., J. Med. Chem., 36 (1993) 2701.
21. Cupelli, L.A. and Hsu, M.C., J. Virol., 69 (1995) 2640.
22. Hamy, F., Felder, E.R., Heizmann, G., Lazdins, J., Aboul-ela, F., Varani, G., Karn, J. and Klimkait, T., Proc. Natl. Acad. Sci. USA, 94 (1997) 3548.
23. Mujeeb, A., Parslow, T.G., Yuan, Y.-C. and James, T.L., J. Biomol. Struct. Dyn., 13 (1996) 649.
24. Filikov, A.V. and James, T.L., J. Comput.-Aided Mol. Design, 12 (1998) 1.
25. Aboul-ela, F., Karn, J. and Varani, G., J. Mol. Biol., 253 (1995) 313.
26. Ye, X., Kumar, R.A. and Patel, D.J., Chem. Biol., 2 (1995) 827.
27. Dieckmann, T., Suzuki, E., Nakamura, G.K. and Feigon, J., RNA, 2 (1996) 628.
28. Yang, Y., Kochoyan, M., Burgstaller, P., Westhof, E. and Famulok, M., Science, 272 (1996) 1343.
29. Fan, P., Suri, A.K. and Fiala, R., J. Mol. Biol., 258 (1996) 480.
30. Jiang, L., Suri, A.K., Fiala, R. and Patel, D.J., Chem. Biol., 4 (1997) 35.
31. Rosenfeld, R., Vajda, S. and DeLisi, C., Ann. Rev. Biophys. Biomol. Struct., 24 (1995) 677.
32. Goodsell, D.S. and Olson, A.J., Proteins, 8 (1990) 195.
33. Yue, S.Y., Protein Eng., 4 (1990) 177.
34. Di Nola, A., Roccantano, D. and Berendsen, H.J., Proteins, 19 (1994) 174.
35. Luty, B.A., Wasserman, Z.R., Stouten, P.F.W. and Hodge, C.N., J. Comput. Chem., 16 (1995) 454.
36. Caflisch, A., Niederer, P. and Anliker, M., Proteins, 14 (1992) 102.
37. Hart, T.N. and Read, R.J., Proteins, 13 (1992) 206.
38. Totrov, M.M. and Abagyan, R.A., Proteins, (1997) Suppl. 1, 215.
39. Totrov, M.M. and Abagyan, R.A., J. Comput. Chem., 15 (1994) 1105.
40. Leach, A.R. and Kuntz, I.D., J. Comput. Chem., 13 (1992) 730.
41. Mizutani, M.Y., Tomioka, N. and Itai, A., J. Mol. Biol., 243 (1994) 3310.
42. Rarey, M., Kramer, B., Lengauer, T. and Klebe, G., J. Mol. Biol., 261 (1996) 470.
43. Sandak, B., Nussinov, R. and Wolfson, H.J., Comput. Appl. Biosci., 11 (1995) 87.
44. Desmet, J., Demaeyer, M., Hazes, B. and Laster, I., Nature, 356 (1992) 539.
45. Vajda, S., Weng, Z., Rosenfeld, R. and DeLisi, C., Biochemistry, 33 (1994) 13977.
46. DesJarlais, R.L., Sheridan, R.P., Seibel, G.L., Dixon, J.S., Kuntz, I.D. and Venkatarghavan, R., J. Med. Chem., 31 (1988) 722.
47. Welch, W., Ruppert, J. and Jain, A.N., Chem. Biol., 3 (1996) 449.
48. Clark, K.P. and Ajay, J. Comput. Chem., 16 (1995) 1210.
49. Jones, G., Willett, P. and Glen, R.C., J. Comput.-Aided Mol. Design, 9 (1995) 532.
50. Gehlhaar, D.K., Moerder, K.E., Zichi, D., Sherman, C.J., Ogden, R.C. and Freer, S.T., J. Med. Chem., 38 (1995) 466.
51. Oshiro, C.M., Kuntz, I.D. and Dixon, J.S., J. Comput.-Aided Mol. Design, 9 (1995) 113.
52. Verlinde, C.L. and Hol, W.G., Structure, 2 (1994) 577.

610

53. Meng, E.C., Shoichet, B.K. and Kuntz, I.D., J. Comput. Chem., 13 (1992) 505.
54. Shoichet, B.K., Bodian, D.L. and Kuntz, I.D., J. Comput. Chem., 113 (1992) 380.
55. Abagyan, R.A., In van Gunsteren, W.F., Weiner, P.K. and Wilkinson, A.J. (Eds), Computer Simulation of Biomolecular Systems: Theoretical and Experimental Applications, vol. 3, Kluwer Academic Publishers, Dordrecht, 1997, pp. 363–394.
56. Wesson, L. and Eisenberg, D., Protein Sci., 1 (1992) 227.
57. Novotny, J., Bruccoleri, R.E. and Saul, F.A., Biochemistry, 28 (1989) 4735.
58. Horton, N. and Lewis, M., Protein Sci., 1 (1992) 169.
59. Murphy, K.P., Xie, D., Garcia, K.C., Amzel, L.M. and Freire, E., Proteins, 15 (1993) 113.
60. Wilson, C., Mace, J.E. and Agard, D.A., J. Mol. Biol., 220 (1991) 495.
61. Wallquist, A., Jernigan, R.L. and Covell, D.G., Protein Sci., 4 (1995) 1881.
62. Bryan, W.P., Science, 266 (1994) 1726.
63. Momany, F.A., McGuire, R.F., Burgess, A.W. and Scheraga, H.A., J. Phys. Chem., 79 (1975) 2361.
64. Veal, J.H. and Wilson, W.D., J. Biomol. Struct. Dyn., 8 (1991) 1119.
65. Pickett, S.D. and Sternberg, M.J., J. Mol. Biol., 231 (1993) 825.
66. Puglisi, J.D., Tan, R.Y., Calnan, B.J., Frankel, A.D. et al., Science, 257 (1992) 76.
67. Puglisi, J.D., Chen, L., Blanchard, S. and Frankel, A.D., Science, 270 (1995) 1200.
68. Abagyan, R.A., Totrov, M.M. and Kuznetsov, D.N., J. Comput. Chem., 15 (1994) 488.
69. Tao, J. and Frankel, A.D., Proc. Natl. Acad. Sci. USA, 90 (1992) 2723.
70. Mei, H.-Y. et al., Bioorg. Med. Chem., 5 (1997) 1173.
71. Sassanfar, M. and Szostak, J.W., Nature, 3364 (1993) 550.
72. Wang, Y., Killian, J., Hamasaki, K. and Rando, R.R., Biochemistry, 35 (1996) 12338.
73. Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N. and Bourne, P.E., Nucleic Acids Res., 28 (2000) 235.
74. Abagyan, R.A. and Argos, P., J. Mol. Biol., 225 (1992) 519.
75. Bohm, H.J., J. Comput.-Aided Mol. Design, 12 (1998) 309.
76. Bohm, H.J., J. Comput.-Aided Mol. Design, 8 (1994) 243.