Chapter 12
Virtual Screening for RNA-Interacting Small Molecules

Hyun-Ju Park and So-Jung Park

12.1 Virtual Screening and Docking Tools

Computational virtual screening strategy is useful in the very early stage of the drug discovery pipeline and provides a powerful tool for rapid discovery of small biologically active molecules. Such strategy can decrease the number of candidate compounds providing a good starting point for chemical synthesis and biological screening. Therefore, in many cases, virtual screening can be used prior to expensive experimental highthroughput screening. For this reason, since the terms of computational virtual screening came out in the late 1990s, it has been considered as a novel and essential technology in drug discovery. Ligand- and structure-based virtual screenings have been successfully applied to drug discovery programs in various disease areas. After prospective results for various protein targets were obtained, the utility of virtual screening to identify compounds for nucleic acid-based receptors has been the focus of much attention.

Many computational docking programs that can automatically dock small molecules into a binding site of a target receptor with minimum input from an operator have been developed and their applicability has been proven (McInnes 2007). In addition, the development of various free or commercially available databases allows for easy use of virtual screening. Each year, many successful cases of virtual screening against various targets including protein and nucleotides have been reported in several major review papers (Whitty and Kumaravel 2006; Seifert and Lang 2008; Villoutreix et al. 2009).

H.-J. Park (✉) • S.-J. Park
School of Pharmacy, Sungkyunkwan University, Suwon 440-746, Korea
e-mail: hyunju85@skku.edu

J.D. Dinman (ed.), *Biophysical Approaches to Translational Control of Gene Expression*, Biophysics for the Life Sciences 1, DOI 10.1007/978-1-4614-3991-2_12,

Table 12.1 Widely used docking algorithms

Name	URL	Short summary
AutoDock	http://autodock.scripps.edu/	Free software of automated small-molecules docking tools (rigid receptor, flexible ligand)
DOCK	http://dock.compbio.ucsf.edu/	Free software, small molecules-various receptor docking (protein, DNA, and RNA), protein–protein interaction
FlexX	http://www.biosolveit.de/flexx/	Fast computer program for predicting protein–ligand interactions
GOLD	http://www.ccdc.cam.ac.uk/products/life_sciences/gold/	Calculating the docking modes of small molecules in protein binding sites
Glide	http://www.schrodinger.com/	Fast flexible ligand docking program (small molecule-protein)
ICM	http://www.molsoft.com/	Automatic incorporation of flexibility into the ligand and receptor docking (protein, DNA, and RNA), protein–protein interaction
MORDOR	http://mondale.ucsf.edu/index_main_frame.html	Docking program using algorithm considering flexibility of both nucleic acid receptor and ligand
Surflex-Dock	http://www.tripos.com/	"Protomol"-guided flexible molecular docking program

The docking process concerns the prediction of ligand conformation and orientation within a targeted binding site (active site) (Kitchen et al. 2004). In order to carry out docking calculations, it is necessary to know the 3-dimensional (3D) structure of a target and the nature of the binding site. 3D structures are identified by X-ray crystallography or NMR experiments and can be downloaded from Protein Data Bank (PDB) (Berman et al. 2000) or predicted by homology modeling using various programs. The next step is to define the binding site by known information or prediction. When input is prepared, chemical compounds present in the database are docked into the defined binding site of the selected target receptor. There are two purposes of docking studies. One is to predict accurate ligand binding orientation referred as "molecular modeling" and the other is to predict activity or binding affinity referred as "scoring." A docking result is evaluated by ligand binding orientation through visual inspection and by binding affinity using the scoring function. Scoring function is designed to predict the biological activity or binding affinity through the evaluation of interactions between ligands and receptor (Halperin et al. 2002). Among various docking programs (Table 12.1), the most widely used programs are AutoDock (Morris et al. 1998; Huey et al. 2007), DOCK (Ewing et al. 2001), FlexX (Kramer et al. 1999; Stahl 2000), Gold (Jones et al. 1995), Glide (Zhou et al. 2001), Internal Coordinate Mechanics (ICM) (Abagyan and Totrov 1994), and Surflex-Dock (Jain 2003; Kellenberger et al. 2004). AutoDock is an automated flexible docking program designed to predict how small molecules (ligands) bind into the receptor structure. AutoDock is performed with an empirical

free energy force field based on a Lamarckian genetic algorithm, to bring about speedy prediction of conformation with calculated free energies of association (Morris et al. 1998). This program has application in X-ray crystallography, structure-based drug design, virtual screening, and protein–protein interaction studies. DOCK was introduced by the Kuntz group at UCSF and uses a rigid body docking algorithm and flexible ligand docking algorithm to dock the ligand into a negative image of the binding pocket (Ewing et al. 2001). FlexX is a fully automatic computer program for predicting protein–ligand interaction. FlexX can predict not only the lowest energy geometry of the complex of ligand with protein but also the binding affinities using an empirical scoring function (Böhm 1994). The descriptions of other widely used docking programs including commercial and free softwares are listed in Table 12.1.

12.2 Computational Programs to Predict 2D and 3D RNA Structures

In recent years, while the number of identified RNA sequences has rapidly increased, the number of known 3D structures has not kept pace with it. For this reason, there is a large gap between the number of known RNA sequences and 3D structures. For example, tRNA is one of the most structurally well characterized RNAs and its 1,101,833 characterized sequences are reported in the Rfam (Gardner et al. 2009), a database of sequence families of structural RNAs; however, only 170 structures are reported. To apply structure-based drug design approaches to the identification of RNA binding ligands, computational programs are required for prediction of RNA structures. Several computer programs have been developed for folding of RNA secondary structures, and modeling of RNA 2D and 3D structures. Those computational tools are summarized in Tables 12.2 and 12.3.

12.3 RNA-Targeted Virtual Screening

Many clinical antibiotics including macrolides, aminoglycosides, and others targeting bacterial ribosomal RNA (rRNA) reveal that RNA is the important target for drug development (Knowles et al. 2002; Hermann 2005). The appearance of drug resistance is the most critical problem in treating bacterial (Neu 1992) and viral infections (Perrin and Telenti 1998). RNAs contain highly conserved structural and functional motifs that may serve as drug targets, so the development of resistance to drugs targeting RNA can be slower than that to drugs targeting protein (Gallego and Varani 2001). In contrast to DNA, which mostly has a double-stranded helix structure, RNA is generally single-stranded and folds into complex 3D structures that provide unique pockets for small molecules (Foloppe et al. 2006), thus making RNA an attractive drug target.

Table 12.2 RNA secondary structure viewers/editors programs

Name	URL	Description
PseudoViewer (Han et al. 1999; Byun and Han 2009)	http://wilab.inha.ac.kr/pseudoviewer/	visualize RNA pseudoknot 2D structure automatically
RNAdraw (Matzura and Wennborg 1996)	http://www.rnadraw.com/	RNA 2D structure prediction, analysis, and visualization
RNA Movies (Evers and Giegerich 1999)	http://bibiserv.techfak.uni-bielefeld.de/rnamovies/	System for the visualization of RNA secondary structure spaces
RNAView/RnamlView (Yang et al. 2003)	http://ndbserver.rutgers.edu	Automatically generate 2D displays of RNA/DNA secondary structures with tertiary interactions
VARNA (Darty et al. 2009)	http://varna.lri.fr	Automated drawing, visualization, and annotation of the secondary structure of RNA
RNA Designer (Andronescu et al. 2004)	http://www.rnasoft.ca/cgi-bin/RNAsoft/RNAdesigner	Design de novo RNA structures with certain structural properties
RnaViz 2 (De Rijk et al. 2003)	http://rnaviz.sourceforge.net/	Drawings of RNA secondary structure with portability and structure annotation
Vienna RNA (Hofacker 2003)	http://www.tbi.univie.ac.at/~ivo/RNA/	Program for the prediction and comparison of RNA secondary structures

In spite of these advantages, RNA has not been focus of structure-based drug design, not only due to lack of information of RNA 3D structures, but also due to the sequence-specific unique features of the binding pockets in RNA. The binding pocket of protein usually lies deep in an internal region, separated from solvent. In RNA targets, the binding pockets are large and flat, located along the surface, and relatively exposed to solvent. Therefore, in using docking algorithms to discover RNA-binding drugs, the physicochemical properties of RNA, such as conformational flexibility, high negative charge, and solvation, should be taken into account more accurately than those for proteins. Despite such differences between protein and RNA targets, classical protein-ligand docking programs have sometimes successfully performed in RNA-targeted virtual screenings. For example, Kuntz first reported a successful virtual screening using DOCK 3.5 program to identify small molecules in the Available Chemicals Directory (ACD) that targeted an RNA double helix (Chen et al. 1997). Following the first study, many research groups have reported successful studies of virtual screening targeting RNA through protein-based docking methods (Filikov et al. 2000; Lind et al. 2002; Kang et al. 2004; Park et al. 2008, 2011). Meanwhile, since most of the available docking methods were developed for protein targets, their compliance with RNA targets has been evaluated extensively, and recently Li et al. demonstrated that two widely-used protein docking programs, GOLD 4.0 and Glide 5.0, are appropriate for structure-based drug design and virtual screening for RNA targets (Detering and Varani 2004; Li et al. 2010).

Table 12.3 Computer program for predicting RNA 3D structure

Name	URL	Short summary
Mode RNA (Rother et al. 2011)	http://genesilico.pl/moderna/	Automatic program for comparative modeling of RNA 3D structures. Require sequence alignment and structural template
MMB (formerly RNA Builder) (Flores et al. 2010)	https://simtk.org/home/rnatoolbox	Automatic program for generating model RNA structures from 2D and template structure by simulating in parallel at multiple levels of details
PARADISE (Assemble and S2S) (Jossinet et al. 2010; Jossinet and Westhof 2005)	http://paradise-ibmc.u-strasbg.fr/	Assemble: automatic program for intuitive graphical interface to study and construct complex 3D RNA structures. S2S: graphical system to easily display, manipulate, and interconnect heterogeneous RNA data like multiple sequence alignments, 2D and 3D structures
RNA2D3D (Martinez et al. 2008)	http://www-lmmb.ncifcrf.gov/	Manual manipulation program for RNA 3D modeling with conversion of RNA 2D structures to 3D
ERNA-3D (Zwieb and Müller 1997)	http://www.erna-3d.de/	Molecular Modeling Expert System to develop for the generation of models of RNA and protein molecules. Need to manual manipulation
MC-Fold/MC-Sym (Parisien and Major 2008)	http://www.major.iric.ca/MC-Pipeline/	A web-hosted service for prediction of RNA 3D structure with input 2D secondary
NAST (Jonikas et al. 2009)	https://simtk.org/home/nast	Predicting RNA 3D model from 2D structure
DMD/iFoldRNA (Sharma et al. 2008)	http://danger.med.unc.edu/tools.php	Web portal for interactive RNA folding simulations. Enable to perform molecular dynamics simulations of RNA using coarse-grained structural models (two-beads/residue)
YUP (Tan et al. 2006)	http://rumour.biology.gatech.edu/YammpWeb/	RNA molecular modeling program based on PYTHON. Offer such methods as Monte Carlo, Molecular Mechanics and Energy Minimization

Several modified algorithms were established for RNA-targeted virtual screening. For example, Morley and Afshar established an empirical scoring function that appropriately describes steric, polar, and charged interactions in RNA-ligand complexes (Morley and Afshar 2004). This scoring function was implanted in RiboDock program and validated for docking screening of a large-size chemical database. Virtual screening with this program successfully identified novel ligands for the bacterial ribosomal A-site (Foloppe et al. 2004). Moitessier et al. applied a unique method to the AutoDock docking process, taking into account inherent RNA flexibility and key water molecules, and this modified docking tool was validated by

the docking of aminoglycosides to the ribosomal RNA A-site (Moitessier et al. 2006). In Kuntz et al.'s study, 70 experimental RNA-ligand complexes from PDB were re-docked using the DOCK 6 program, and the resulting docked conformations were rescored with AMBER generalized Born/solvent-accessible surface area (GB/SA) and Poisson–Blotzmann/SA (PB/SA) scoring functions in combination with explicit water molecules and sodium counterions. The success rate for reproducing experimental binding modes was significantly improved by using AMBER GB/SA or PB/SA (Lang et al. 2009).

As small molecules induce RNA conformational changes by binding to structures from preexisting dynamic ensembles (Puglisi et al. 1992; Zhang et al. 2007; Frank et al. 2009; Cruz and Westhof 2009; Fulle and Gohlke 2010), large conformational changes in RNA receptors after binding with small molecules during virtual screening must be accounted for. However, current protocols do not consider this point, limiting the range of target structures for the discovery of small molecules. Guilbert and James therefore developed a flexible docking program called MORDOR, which supports flexibility in the ligand and limited flexibility in the RNA for induced-fit binding. MORDOR performed well not only on 57 test sets of RNA-ligand complexes by retrieving experimental poses within 2.5 Å with a 74% success rate, but also in discovering ligands for novel targets such as human telomerase RNA (Guilbert and James 2008; Pinto et al. 2008). When comparing the practicality of DOCK 6 and MORDOR, DOCK 6 screened about 3–10 times faster than MORDOR, while MORDOR performed better on ligands with a large number of rotatable bonds (Lang et al. 2009). Most recently, Al-Hashimi's group reported a new strategy for virtual screening targeting an RNA dynamic ensemble constructed by combining NMR spectroscopy and computational molecular dynamics (MD) (Stelzer et al. 2011). This strategy takes into account large degrees of RNA conformational adaptation during virtual screening. This approach was applied to a search for small molecules for HIV-1 TAR (transactivation response element) RNA (*vide infra*).

12.4 Successful Application of Virtual Screening to RNA Receptors

12.4.1 Case 1: HIV-1 TAR RNA Hairpin

As one of the best characterized RNA-based regulatory machineries, the interaction of TAR with Tat protein has been focused upon as a target to inhibit HIV-1 replication (Yang 2005). Ligands that inhibit the TAR-Tat interaction can be developed as anti-AIDS drugs. Various molecules for TAR RNA that inhibit Tat binding have been identified, and many of those molecules were discovered by virtual screening of chemical libraries. In earlier studies, James used automatic docking methods (DOCK and ICM), employing flexible docking with Monte Carlo simulation and optimized scoring function, identifying phenothiazine compounds from ACD as

TAR RNA ligands (Filikov et al. 2000; Lind et al. 2002). To circumvent the limitation of incorporating RNA flexibility for structure-based virtual screening, and to find novel scaffolds for TAR-Tat inhibitors, ligand-based virtual screening was conducted. The SQUID fuzzy pharmacophore search method successfully identified a novel heterocyclic compound with an order of magnitude improved activity compared to known phenothiazine compounds (Renner et al. 2005).

In 2011, Al-Hashimi's group developed the most outstanding technology for RNA-targeted virtual screening through intensive generation of an ensemble of TAR RNA conformers and a robust re-docking validation test. Their strategy was effectively applied to virtual screening of a relatively small-sized chemical library containing 51,000 compounds, and they identified netilmicin, a selective HIV-1 TAR RNA binder, that inhibited HIV-1 replication in vivo (Stelzer et al. 2011). Details of their study are described below.

12.4.1.1 Validation of Docking Program

To test the accuracy of docking, a total of 96 small molecule-bound RNA structures downloaded from the PDB were assessed for docking performance. All docking performances were carried out using the ICM docking program (Abagyan and Totrov 1994) and results were evaluated by ICM Score and RMSD between native ligand (extracted from RNA structures) and predicted orientation after docking. The binding energies based on the ICM Score were predicted with high accuracy ($R = 0.71$). In more than half of cases (53%), the predicted conformations matched the X-ray or NMR structures within 2.5 Å RMSD.

12.4.1.2 Preparation of HIV-1 TAR RNA Ensemble

To decide on a suitable RNA ensemble, the accuracies of docking with two sets of RNA ensembles of HIV-1 TAR were compared. One ensemble, named $TAR^{NMR\text{-}MD}$, consisted of 20 conformers generated by SAS selection (select-and-sample strategy, Frank et al. 2009). To construct $TAR^{NMR\text{-}MD}$, HIV-1 TAR structure (PDB id: 1ANR, Aboul-ela et al. 1996) was downloaded and molecular dynamics simulations were performed by measuring NMR residual dipolar couplings (RDC) in elongated RNA. The other RNA ensemble, named TAR^{MD}, consisted of 20 randomly selected snapshots from an 80-ns MD simulation of apo-TAR with backbone RMSDs ranging from 3 to 80 Å. The X-ray structure (PDB id: 397D, Ippolito and Steitz 1998) and 20 NMR structures (PDB id: 1ANR, Aboul-ela et al. 1996) of apo-TAR were downloaded from the PDB. A test set ligand library containing 38 known ligands for TAR RNA was obtained from the published literatures. To test the accuracy of the RNA ensemble, virtual screening of the test set ligands targeting the RNA ensembles, $TAR^{NMR\text{-}MD}$ and TAR^{MD}, was conducted, respectively. Virtual screening was conducted using ICM after binding pockets were predicted by the ICM PocketFinder module based on the calculated surface area and volume of cavities on the receptor surface.

The results suggested that $TAR^{NMR\text{-}MD}$ improved the accuracy of docking compared with TAR^{MD}. From this result $TAR^{NMR\text{-}MD}$ was chosen for further virtual screening to identify small molecule TAR-Tat inhibitors.

12.4.1.3 Virtual Screening Against TAR Dynamic Ensemble

A chemical database consisting of 49,166 compounds was obtained from the Center for Chemical Genomics at the University of Michigan and 2,060 compounds from the author's in-house library. Based on the ICM Score, the top 57 commercially available hit compounds were selected and their binding activities were tested by fluorescence-based assays. This identified six small molecules that bound to TAR with high affinity (K_d = 55 nM to 122 μM) and inhibited TAR interaction with Tat (K_i = 710 nM to 169 μM) in vitro. Among them, netilmicin (K_d = ~1.4 μM) bound to HIV-1 TAR with the highest selectivity over HIV-2 TAR. Netilmicin repressed Tat-mediated transactivation of the HIV-1 promoter through its interaction with TAR in live human T-cells and inhibited HIV-1 replication in the HIV-1 indicator cell line TZM-bl.

12.4.2 Case 2: RNA Pseudoknots

Ribosomal −1 frameshifting (−1 FS) is an essential event during translation for the synthesis of two or more proteins encoded by overlapping reading frames on a single mRNA (Dinman and Berry 2007) in many RNA viruses such as retroviruses, coronaviruses, yeast, plant virus, and even bacteria (Jacks and Varmus 1985; Brierley et al. 1991; Chamorro et al. 1992; Tzeng et al. 1992; ten Dam et al. 1994; Kang and Tinoco 1997; Jacobs et al. 2007). Two *cis*-acting elements are required to regulate −1 FS. One is a slippery sequence where ribosome-associated tRNAs slip, and the other is RNA secondary structure such as a hairpin or pseudoknot that promotes ribosome pausing. Thermodynamic or kinetic control of RNA secondary structure folding may be important in regulating the efficiency of −1 FS. Human immunodeficiency virus type 1 (HIV-1) utilizes −1 FS to regulate the expression ratio of Gag to Gag-Pol, which is critical for the production of infectious virion particles (Paulus et al. 1999). The RNA stem-loop sequence that is involved in −1 FS of HIV-1 is highly conserved in the main subtypes of HIV-1 (Gareiss and Miller 2009). Mutations of this sequence reduce −1 FS and decrease viral infectivity and replication (Baril et al. 2003; Dulude et al. 2006). In severe acute respiratory syndrome coronavirus (SARS-CoV), replicase genes mainly encode two large replicative polyproteins (pp1a and pp1ab) which are expressed by two partially overlapped open reading frames ORF 1a and ORF 1b. As ORF 1b has no independent translation initiation site, polyprotein pp1ab encoded by ORF 1b is only translated as a fused protein form with ORF 1a through −1 FS. As pp1ab includes RNA-dependent RNA polymerase (RdRp) and other replication components which are important proteins for viral replication,

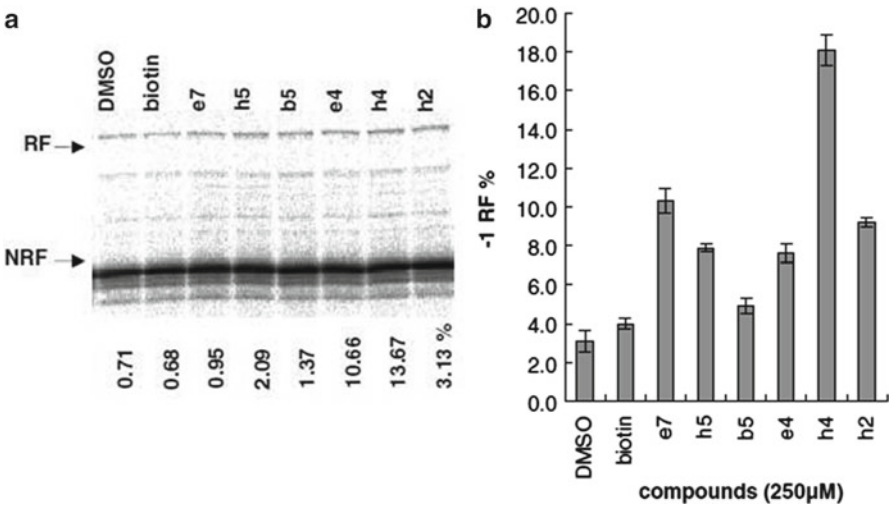

Fig. 12.1 −1 frameshifting efficiencies induced by biotin aptamer RNA pseudoknot in the presence of compounds (250 μM) determined by SDS–PAGE (**a**) and dual luciferase assay (**b**). −1 FS % values are shown at the *bottom* of the autoradiogram of SDS–PAGE. Each −1 FS % value from dual luciferase assay is the average of triplicate experiments

−1 FS is essential for the synthesis of enzymatic proteins. The stability of the RNA pseudoknot that induces −1 FS in the SARS-CoV (SARS-pseudoknot) also has a dramatic effect on −1 FS efficiency. Therefore, the RNA secondary structure in the −1 FS site has emerged as an attractive target for drug development (Baranov et al 2005; Plant et al 2005; Su et al. 2005).

Park first conducted virtual screening against the RNA pseudoknot in the −1 FS site to discover ligands that change −1 FS efficiency (Park et al. 2008). In this pilot study, they used the −1 FS system containing biotin aptamer RNA pseudoknot as the RNA secondary structure element. Biotin RNA aptamer was the only ligand-bound RNA pseudoknot structure (PDB id: 1F27) determined by X-crystallography (Nix et al. 2000). Park et al. used the conventional 2D and 3D pharmacophore search program Unity (Martin 1992) for primary database filtering and the FlexX docking program for final docking screening. RNA flexibility was not considered and ligand flexibility was given during FlexX run. Out of about 80,000 compounds in the chemical DB, they obtained 37 hits which increased −1 FS. Compound **h4** showed the highest activity in the in vitro transcription and translation coupled assay (Fig. 12.1). The FlexX-docked pose of **h4** is shown in Fig. 12.2. The docking mode of **h4** is similar to that of biotin; however, **h4** forms a stronger interaction with the receptor RNA (Fig. 12.2a). Compound **h4** forms hydrogen bonds with O4′ and the 2-carbonyl oxygen of uracil ring of U7 which is one of the critical residues for interaction with biotin in the X-ray structure, and an additional hydrogen bond with ribose O2′ atom of A16 (Fig. 12.2b). These interactions may alter the stability of the RNA pseudoknot and increase the stalled time of the ribosome on the slippery site, thus increasing the rates of −1 FS.

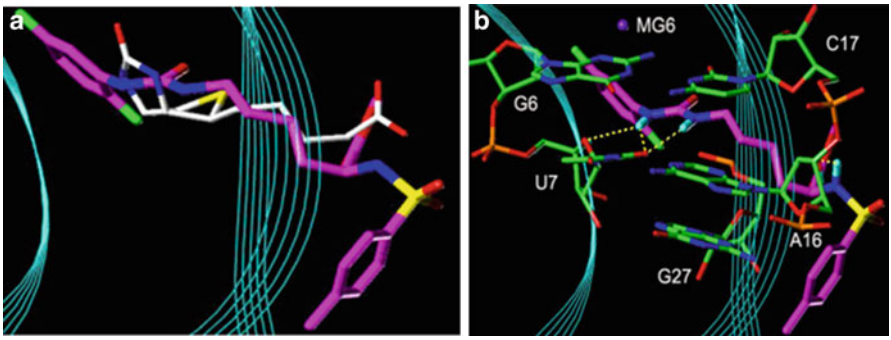

Fig. 12.2 (**a**) Overlay of FlexX-docked pose of **h4** and X-ray pose of biotin in the aptamer RNA pseudoknot. The ligand is rendered in capped stick. Carbon atoms of **h4** are *magenta* and those of biotin are *white*. *Cyan* lined-ribbon represents the backbone of RNA. (**b**) Docked model of **h4** in complex with biotin-pseudoknot complex. The residues in the active site are rendered in stick. Carbon atoms of pseudoknot are *green*, *oxygen red*, *nitrogen blue*, and *phosphorus orange*. *Yellow dashed lines* are hydrogen bonds

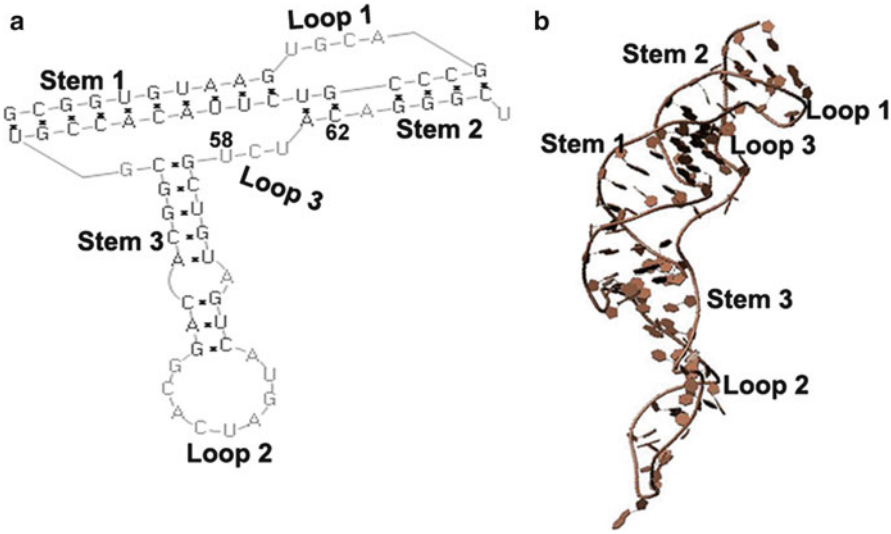

Fig. 12.3 (**a**) Two-dimensional model of SARS-pseudoknot generated by the PSEUDOVIEWER program. (**b**) Three-dimensional structural model of the SARS-pseudoknot used in this study. It was optimized by molecular dynamics simulation using the Amber 8.0 program. *Brown ribbon* renders the phosphate backbone of the RNA pseudoknot

Based on these successful results, Park applied the virtual screening strategy to discover ligands for the SARS-pseudoknot, even though its 3D structure was not completely determined. It was known that the SARS-pseudoknot has a unique 3 stem-3 loop structure. They built a 3D model using the RNA pseudoknot predicting program (PSEUDOVIEWER) and Sybyl molecular modeling software, and then

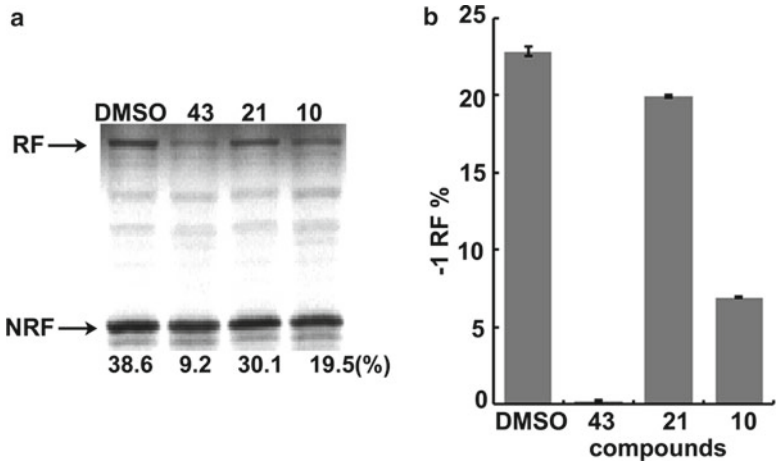

Fig. 12.4 Measurements of −1 FS efficiencies by in vitro TNT assay. (**a**) The −1 FS efficiencies (%) in the presence of **43**, **21**, and **10** obtained from SDS-PAGE analysis. The nonframeshifting product (NRF) is the renilla luciferase protein, and the frameshifting product (RF) is a firefly luciferase-renilla luciferase fusion protein. (**b**) The −1 FS efficiencies obtained from dual luciferase assays

optimized the RNA structure by AMBER molecular dynamics simulation (Case et al. 2005) (Fig. 12.3). Using the DOCK 4.0 program, flexible docking of a commercially-available chemical DB (Leadquest) was conducted. The chemical DB was the same as that previously used for screening against the biotin aptamer RNA pseudoknot. A total of 35 compounds inhibited −1 FS, and three structurally analogous compounds (**43**, **21**, and **10** in Fig. 12.4) reduced −1 FS selectively. Compound **43** was the most active, decreasing −1 FS efficiency by 80%. In HEK 293 cells, **43** inhibited −1 FS in a concentration-dependent manner with an IC_{50} value of approximately 0.45 μM (Park et al. 2011). The docking model of **43** in complex with the SARS-pseudoknot is shown in Fig. 12.5, and reveals that **43** interacts with various residues including key residues to maintain −1 frameshifting efficiency by hydrogen bonds. Compound **43** interacts with the carbonyl oxygen atom (O2) of the U58 uracil base by hydrogen bond, and the nitrogen atom in the thiazole ring of **43** forms a hydrogen bond (2.8 Å) with the 2′-OH group of ribose of C62. The hydrogen bond between the thiazole moiety and the receptor pseudoknot was identified as one of the key intermolecular interactions.

12.4.3 Case 3: Riboswitches (Metabolite-Sensing mRNAs)

Riboswitches are highly organized domains within 5′-UTRs of mRNAs and undergo alternate conformational switches. Riboswitches consist of two domains, an aptamer domain that is a binding site for an effector metabolite, and an expression platform that prompts changes in gene expression. Upon metabolite binding, one of the riboswitch conformers is stabilized, and is capable of controlling expression of a

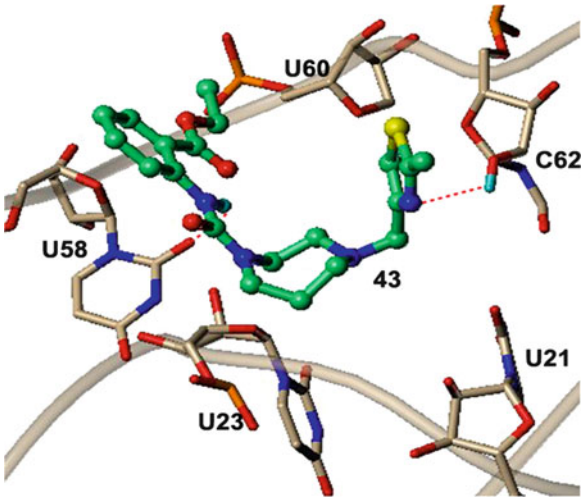

Fig. 12.5 Docked model of 43: SARS-psuedoknot complex generated by DOCK 4.0. Several residues in the binding site are rendered in capped stick (*brown carbon*), **43** in ball and stick (*green carbon*), and *red dashed lines* indicate hydrogen bonds

downstream gene either at the transcriptional or translational level (Miranda-Ríos et al. 2001; Mironov et al. 2002; Gilbert and Batey 2005) in bacteria (Winkler et al. 2002; Mandal et al. 2003) as well as in some plants and fungi (Sudarsan et al. 2003; Bocobza and Aharoni 2008). Therefore, the aptamer domain is an essential element of gene regulation. Riboswitches were originally discovered as an antibiotic target (Blount and Breaker 2006; Lee et al. 2009; Kim et al. 2009) and also have the potential to be developed into designer riboswitches for genetics studies (Suess and Weigand 2008). About 20 classes of riboswitches have been reported and a large number (probably up to 100) of new classes may await discovery. Among them, guanine riboswitches are one of the validated targets for development of new antibacterial drugs. X-ray crystal structures of more than 10 riboswitches have been determined, and thus structure-based drug design approaches can be applied to this target (Serganov 2010). Daldrop et al. tried virtual screening to discover novel ligands targeting the purine riboswitch using the program DOCK3.5.54 with minor modification of the scoring function (Daldrop et al. 2011). The *Bacillus subtilis xpt-pbuX* guanine riboswitch carrying a C74U mutation (called GRA, PDB id: 2G9C) (Gilbert et al. 2006) in complex with pyrimidine-2,4,6-triamine was used as a target receptor for docking.

12.4.3.1 Validation of Docking Program

All docking and virtual screening studies were carried out using a slightly modified version of the DOCK3.5.54 program that incorporated RNA-specific parameters to calculate van der Waals and electrostatic energies. This group first tested whether native ligand was correctly reproduced in terms of its binding geometry after self-docking. The RMSD between native ligand and the docking result was measured.

RMSD was less than 0.34 Å, which showed that the docked model was close to native. Secondly, the accuracy of the prediction was tested by docking with known ligands and decoys. The test set consisted of eight known ligands, with binding affinities ranged from 0.01 to 100 μM. For all 15 decoys, no binding was detected at up to 300 μM; except for guanine that was tested up to its solubility limit. All compounds in the test set docked into the active site and the results were analyzed by sorting the docking scores. Seven out of the eight top-scoring compounds were true ligands and all eight compounds with lowest docking energy scores were decoys.

12.4.3.2 Virtual Screening

They used their own in-house database, which included a commercially available 2,592 unique compounds. To evaluate the accuracy of binding prediction, all compounds in the database were docked into the active site and ordered by docking score. According to this list, the true positive rate (fraction of known compounds, ligands, and decoy) was plotted against the false positive (fraction of unassigned database compounds) to get a receiver operation characteristic (ROC) curve. Accuracy was calculated by measuring the area under curve (AUC) of a ROC curve value. The AUC value of ligands was 0.98 and that of decoy was 0.75, suggesting that known ligands were perfectly predicted by virtual screening and decoys were also enriched compared to random (AUC 0.5). From this result, five compounds were selected to examine their binding affinity and modes. Three compounds out of five were analogs of known ligands and two compounds were novel scaffolds. To determine the binding affinities, fluorescence assays and isothermal titration calorimetry (ITC) were used. Four out of five chosen compounds bound to GRA with affinities in the micromolar range.

12.5 Concluding Remarks

The rapidly increasing number of RNA crystal structures in the PDB, and the biological function studies on a variety of RNA structures have provided a basis for structure-based virtual screening targeting RNA. Due to several features of RNA (conformational flexibility, high negative charge, and solvation) that differ from those of proteins, researchers have observed that the conventional protein-ligand docking programs have limitations in accurately predicting RNA-ligand interactions. In some studies, active molecules have been fortuitously obtained by protein-targeted docking programs. However, continuing progress has been made in the development of docking algorithms or scoring functions optimized for RNA receptors, which makes various RNA targets more amenable for structure-based drug design. As an example, consideration of receptor flexibility is important not only for RNA-based receptors, but has also been a critical issue for protein receptors for a long time. Massive efforts have been made to incorporate protein flexibility into

docking process, but it exponentially inflates the potential search space and became impractical. Al-Hashimi's approaches to use RNA dynamic ensemble (Stelzer et al. 2011) can be practically applied to other RNA targets. RNA receptors are usually smaller than protein receptors in size, so the process to generate RNA ensembles is not extremely computationally expensive.

Structure- or ligand-based virtual screening has identified a plethora of RNA binding ligands from in-house and commercially available chemical databases, originally designed and prepared for protein targets. Actually, some commercial chemical databases possess compounds which only satisfy "Lipinski's rule of five (Lipinski et al. 1997)". In comparing the activities of small molecule RNA binders identified by virtual screening, we realized that hits with nanomolar activity were very rare, and overall their activities are in the high micromolar range. These results were probably caused by limitations of currently available chemical database, not only by those of protein-friendly docking algorithms. In protein-targeted virtual screening of chemical databases, many hits with nanomolar or submicromolar activity were identified. Thus, another possible way to improve the hit rate of RNA-targeted virtual screening is to prepare an RNA-focused chemical database. Considering the physicochemical properties of RNA binders, a modified "Lipinski's rule of five" needs to be applied to database filtering during virtual screening (Aboul-ela 2010).

Acknowledgements This work is supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (Grant 2011-0014385).

References

Abagyan R, Totrov M (1994) Biased probability Monte Carlo conformational searches and electrostatic calculations for peptides and proteins. J Mol Biol 235(3):983–1002

Aboul-ela F (2010) Strategies for the design of RNA-binding small molecules. Future Med Chem 2(1):93–119

Aboul-ela F, Karn J, Varani G (1996) Structure of HIV-1 TAR RNA in the absence of ligands reveals a novel conformation of the trinucleotide bulge. Nucleic Acids Res 24(20):3974–3981

Andronescu M, Fejes AP, Hutter F, Hoos HH, Condon A (2004) A new algorithm for RNA secondary structure design. J Mol Biol 336(3):607–624

Baranov PV, Henderson CM, Anderson CB, Gesteland RF, Atkins JF, Howard MT (2005) Programmed ribosomal frameshifting in decoding the SARS-CoV genome. Virology 332(2):498–510

Baril M, Dulude D, Gendron K, Lemay G, Brakier-Gingras L (2003) Efficiency of a programmed −1 ribosomal frameshift in the different subtypes of the human immunodeficiency virus type 1 group M. RNA 9:1246–12453

Berman HM, Bhat TN, Bourne PE, Feng Z, Gilliland G, Weissig H, Westbrook J (2000) The Protein Data Bank and the challenge of structural genomics. Nat Struct Biol 7(suppl A):957–959

Blount KF, Breaker RR (2006) Riboswitches as antibacterial drug targets. Nat Biotechnol 24(12):1558–1564

Bocobza SE, Aharoni A (2008) Switching the light on plant riboswitches. Trends Plant Sci 13(10):526–533

Böhm HJ (1994) The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure. J Comput Aided Mol Des 8(3):243–256

Brierley I, Rolley NJ, Jenner AJ, Inglis SC (1991) Mutational analysis of the RNA pseudoknot component of a coronavirus ribosomal frameshifting signal. J Mol Biol 220(4):889–902

Byun Y, Han K (2009) PseudoViewer3: generating planar drawings of large-scale RNA structures with pseudoknots. Bioinformatics 25(11):1435–1437

Case DA, Cheatham TE III, Darden T, Gohlke H, Luo R, Merz KM Jr, Onufriev A, Simmerling C, Wang B, Woods RJ (2005) The Amber biomolecular simulation programs. J Comput Chem 26(16):1668–1688

Chamorro M, Parkin N, Varmus HE (1992) An RNA pseudoknot and an optimal heptameric shift site are required for highly efficient ribosomal frameshifting on a retroviral messenger RNA. Proc Natl Acad Sci U S A 89(2):713–717

Chen Q, Shafer RH, Kuntz ID (1997) Structure-based discovery of ligands targeted to the RNA double helix. Biochemistry 36(38):11402–11407

Cruz JA, Westhof E (2009) The dynamic landscapes of RNA architecture. Cell 136(4):604–609

Daldrop P, Reyes FE, Robinson DA, Hammond CM, Lilley DM, Batey RT, Brenk R (2011) Novel ligands for a purine riboswitch discovered by RNA-ligand docking. Chem Biol 18(3):324–335

Darty K, Denise A, Ponty Y (2009) VARNA: interactive drawing and editing of the RNA secondary structure. Bioinformatics 25(1):974–1975

De Rijk P, Wuyts J, De Wachter R (2003) RnaViz 2: an improved representation of RNA secondary structure. Bioinformatics 19(2):299–300

Detering C, Varani G (2004) Validation of automated docking programs for docking and database screening against RNA drug targets. J Med Chem 47(17):4188–4201

Dinman JD, Berry MJ (2007) Regulation of termination and recoding. In: Mathews MB, Soneberg N, Hershey JWB (eds) Translational control in biology and medicine, 1st edn. Cold Spring Harbor, New York

Dulude D, Berchiche YA, Gendron K, Brakier-Gingras L, Heveker N (2006) Decreasing the frameshift efficiency translates into an equivalent reduction of the replication of the human immunodeficiency virus type 1. Virology 345:127–136

Evers D, Giegerich R (1999) RNA movies: visualizing RNA secondary structure spaces. Bioinformatics 15(1):32–37

Ewing TJ, Makino S, Skillman AG, Kuntz ID (2001) DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases. J Comput Aided Mol Des 15(5):411–428

Filikov AV, Mohan V, Vickers TA, Griffey RH, Cook PD, Abagyan RA, James TL (2000) Identification of ligands for RNA targets via structure-based virtual screening: HIV-1 TAR. J Comput Aided Mol Des 14(6):593–610

Flores SC, Wan Y, Russell R, Altman RB (2010) Predicting RNA structure by multiple template homology modeling. Pac Symp Biocomput 2010:216–227

Foloppe N, Chen IJ, Davis B, Hold A, Morley D, Howes R (2004) A structure-based strategy to identify new molecular scaffolds targeting the bacterial ribosomal A-site. Bioorg Med Chem 12:935–947

Foloppe N, Matassova N, Aboul-Ela F (2006) Towards the discovery of drug-like RNA ligands? Drug Discov Today 11(21–22):1019–1027

Frank AT, Stelzer AC, Al-Hashimi HM, Andricioaei I (2009) Constructing RNA dynamical ensembles by combining MD and motionally decoupled NMR RDCs: new insights into RNA dynamics and adaptive ligand recognition. Nucleic Acids Res 37(11):3670–3679

Fulle S, Gohlke H (2010) Molecular recognition of RNA: challenges for modelling interactions and plasticity. J Mol Recognit 23(2):220–231

Gallego J, Varani G (2001) Targeting RNA with small-molecule drugs: therapeutic promise and chemical challenges. Acc Chem Res 34(10):836–843

Gardner PP, Daub J, Tate JG, Nawrocki EP, Kolbe DL, Lindgreen S, Wilkinson AC, Finn RD, Griffiths-Jones S, Eddy SR, Bateman A (2009) Rfam: updates to the RNA families database. Nucleic Acids Res 37(Database issue):D136–D140

Gareiss PC, Miller BL (2009) Ribosomal frameshifting: an emerging drug target for HIV. Curr Opin Investig Drugs 10(2):121–128

Gilbert SD, Batey RT (2005) Riboswitches: natural SELEXion. Cell Mol Life Sci 62(21):2401–2404

Gilbert SD, Mediatore SJ, Batey RT (2006) Modified pyrimidines specifically bind the purine riboswitch. J Am Chem Soc 128(44):14214–14215

Guilbert C, James TL (2008) Docking to RNA via root-mean-square-deviation-driven energy minimization with flexible ligands and flexible targets. J Chem Inf Model 48(6):1257–1268

Halperin I, Ma B, Wolfson H, Nussinov R (2002) Principles of docking: an overview of search algorithms and a guide to scoring functions. Proteins 47(4):409–443

Han K, Kim D, Kim HJ (1999) A vector-based method for drawing RNA secondary structure. Bioinformatics 15(4):286–297

Hermann T (2005) Drugs targeting the ribosome. Curr Opin Struct Biol 15(3):355–366

Hofacker IL (2003) Vienna RNA secondary structure server. Nucleic Acids Res 31(13):3429–3431

Huey R, Morris GM, Olson AJ, Goodsell DS (2007) A semiempirical free energy force field with charge-based desolvation. J Comput Chem 28(6):1145–1152

Ippolito JA, Steitz TA (1998) A 1.3-Å resolution crystal structure of the HIV-1 trans-activation response region RNA stem reveals a metal ion-dependent bulge conformation. Proc Natl Acad Sci U S A 95(17):9819–9824

Jacks T, Varmus HE (1985) Expression of the Rous sarcoma virus pol gene by ribosomal frameshifting. Science 230(4731):1237–1242

Jacobs JL, Belew AT, Rakauskaite R, Dinman JD (2007) Identification of functional, endogenous programmed −1 ribosomal frameshift signals in the genome of Saccharomyces cerevisiae. Nucleic Acids Res 35(1):165–174

Jain AN (2003) Surflex: fully automatic flexible molecular docking using a molecular similarity-based search engine. J Med Chem 46(4):499–511

Jones G, Willett P, Glen RC (1995) Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. J Mol Biol 245(1):43–53

Jonikas MA, Radmer RJ, Laederach A, Das R, Pearlman S, Herschlag D, Altman RB (2009) Coarse-grained modeling of large RNA molecules with knowledge-based potentials and structural filters. RNA 15(2):189–199

Jossinet F, Westhof E (2005) Sequence to structure (S2S): display, manipulate and interconnect RNA data from sequence to structure. Bioinformatics 21(15):3320–3321

Jossinet F, Ludwig TE, Westhof E (2010) Assemble: an interactive graphical tool to analyze and build RNA architectures at the 2D and 3D levels. Bioinformatics 26(16):2057–2059

Kang H, Tinoco I Jr (1997) A mutant RNA pseudoknot that promotes ribosomal frameshifting in mouse mammary tumor virus. Nucleic Acids Res 25(10):1943–1949

Kang X, Shafer RH, Kuntz ID (2004) Calculation of ligand-nucleic acid binding free energies with the generalized-born model in DOCK. Biopolymers 73(2):192–204

Kellenberger E, Rodrigo J, Muller P, Rognan D (2004) Comparative evaluation of eight docking tools for docking and virtual screening accuracy. Proteins 57(2):225–242

Kim JN, Blount KF, Puskarz I, Lim J, Link KH, Breaker R (2009) Design and antimicrobial action of purine analogues that bind Guanine riboswitches. ACS Chem Biol 4(11):915–927

Kitchen DB, Decornez H, Furr JR, Bajorath J (2004) Docking and scoring in virtual screening for drug discovery: methods and applications. Nat Rev Drug Discov 3(11):935–949

Knowles DJ, Foloppe N, Matassova NB, Murchie AI (2002) The bacterial ribosome, a promising focus for structure-based drug design. Curr Opin Pharmacol 2(5):501–506

Kramer B, Rarey M, Lengauer T (1999) Evaluation of the FLEXX incremental construction algorithm for protein-ligand docking. Proteins 37(2):228–241

Lang PT, Brozell SR, Mukherjee S, Pettersen EF, Meng EC, Thomas V, Rizzo RC, Case DA, James TL, Kuntz ID (2009) DOCK 6: combining techniques to model RNA-small molecule complexes. RNA 15(6):1219–1230

Lee ER, Blount KF, Breaker RR (2009) Roseoflavin is a natural antibacterial compound that binds to FMN riboswitches and regulates gene expression. RNA Biol 6(2):187–194
Li Y, Shen J, Sun X, Li W, Liu G, Tang Y (2010) Accuracy assessment of protein-based docking programs against RNA targets. J Chem Inf Model 50(6):1134–1146
Lind KE, Du Z, Fujinaga K, Peterlin BM, James TL (2002) Structure-based computational database screening, in vitro assay, and NMR assessment of compounds that target TAR RNA. Chem Biol 9(2):185–193
Lipinski CA, Lombardo F, Dominy BW, Feeney PJ (1997) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Deliv Rev 23:3–25
Mandal M, Boese B, Barrick JE, Winkler WC, Breaker RR (2003) Riboswitches control fundamental biochemical pathways in Bacillus subtilis and other bacteria. Cell 113(5):577–586
Martin YC (1992) 3D database searching in drug design. J Med Chem 35(12):2145–2154
Martinez HM, Maizel JV Jr, Shapiro BA (2008) RNA2D3D: a program for generating, viewing, and comparing 3-dimensional models of RNA. J Biomol Struct Dyn 25(6):669–683
Matzura O, Wennborg A (1996) RNAdraw: an integrated program for RNA secondary structure calculation and analysis under 32-bit Microsoft Windows. Comput Appl Biosci 12(3): 247–249
McInnes C (2007) Virtual screening strategies in drug discovery. Curr Opin Chem Biol 11(5): 494–502
Miranda-Ríos J, Navarro M, Soberón M (2001) A conserved RNA structure (thi box) is involved in regulation of thiamin biosynthetic gene expression in bacteria. Proc Natl Acad Sci U S A 98(17):9736–9741
Mironov AS, Gusarov I, Rafikov R, Lopez LE, Shatalin K, Kreneva RA, Perumov DA, Nudler E (2002) Sensing small molecules by nascent RNA: a mechanism to control transcription in bacteria. Cell 111(5):747–756
Moitessier N, Westhof E, Hanessian S (2006) Docking of aminoglycosides to hydrated and flexible RNA. J Med Chem 49(3):1023–1033
Morley SD, Afshar M (2004) Validation of an empirical RNA-ligand scoring function for fast flexible docking using Ribodock. J Comput Aided Mol Des 18:189–208
Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, Olson AJ (1998) Automated docking using a lamarckian genetic algorithm and an empirical binding free energy function. J Comput Chem 19:1639–1662
Neu HC (1992) The crisis in antibiotic resistance. Science 257(5073):1064–1073
Nix J, Sussman D, Wilson C (2000) The 1.3 Å crystal structure of a biotin-binding pseudoknot and the basis for RNA molecular recognition. J Mol Biol 296(5):1235–1244
Parisien M, Major F (2008) The MC-fold and MC-Sym pipeline infers RNA structure from sequence data. Nature 452(7183):51–55
Park SJ, Jung YH, Kim YG, Park HJ (2008) Identification of novel ligands for the RNA pseudoknot that regulate −1 ribosomal frameshifting. Bioorg Med Chem 16(8):4676–4684
Park SJ, Kim YG, Park HJ (2011) Identification of RNA pseudoknot-binding ligand that inhibits the −1 ribosomal frameshifting of SARS-Coronavirus by structure-based virtual screening. J Am Chem Soc 133(26):10094–10100
Paulus C, Hellebrand S, Tessmer U, Wolf H, Kräusslich HG, Wagner R (1999) Competitive inhibition of human immunodeficiency virus type-1 protease by the Gag-Pol transframe protein. J Biol Chem 30:21539–21543
Perrin L, Telenti A (1998) HIV treatment failure: testing for HIV resistance in clinical practice. Science 280(5371):1871–1873
Pinto IG, Guilbert C, Ulyanov NB, Stearns J, James TL (2008) Discovery of ligands for a novel target, the human telomerase RNA, based on flexible-target virtual screening and NMR. J Med Chem 51:7205–7215
Plant EP, Pérez-Alvarado GC, Jacobs JL, Mukhopadhyay B, Hennig M, Dinman JD (2005) A three-stemmed mRNA pseudoknot in the SARS coronavirus frameshift signal. PLoS Biol 3(6):e172

Puglisi JD, Tan R, Calnan BJ, Frankel AD, Williamson JR (1992) Conformation of the TAR RNA-arginine complex by NMR spectroscopy. Science 257(5066):76–80

Renner S, Ludwig V, Boden O, Scheffer U, Göbel M, Schneider G (2005) New inhibitors of the tat-tar RNA interaction found with a "Fuzzy" Pharmacophore model. Chembiochem 6:1119–1125

Rother M, Rother K, Puton T, Bujnicki JM (2011) ModeRNA: a tool for comparative modeling of RNA 3D structure. Nucleic Acids Res 39(10):4007–4022

Seifert MH, Lang M (2008) Essential factors for successful virtual screening. Mini Rev Med Chem 8(1):63–72

Serganov A (2010) Determination of riboswitch structures: light at the end of the tunnel? RNA Biol 7:98–103

Sharma S, Ding F, Dokholyan NV (2008) iFoldRNA: three-dimensional RNA structure prediction and folding. Bioinformatics 24(17):1951–1952

Stahl M (2000) Modifications of the scoring function in FlexX for virtual screening applications. Persp Drug Discov Des 20:83–98

Stelzer AC, Frank AT, Kratz JD, Swanson MD, Gonzalez-Hernandez MJ, Lee J, Andricioaei I, Markovitz DM, Al-Hashimi HM (2011) Discovery of selective bioactive small molecules by targeting an RNA dynamic ensemble. Nat Chem Biol 7(8):553–559

Su MC, Chang CT, Chu CH, Tsai CH, Chang KY (2005) An atypical RNA pseudoknot stimulator and an upstream attenuation signal for −1 ribosomal frameshifting of SARS coronavirus. Nucleic Acids Res 33(13):4265–4275

Sudarsan N, Barrick JE, Breaker RR (2003) Metabolite-binding RNA domains are present in the genes of eukaryotes. RNA 9(6):644–647

Suess B, Weigand JE (2008) Engineered riboswitches: overview, problems and trends. RNA Biol 5:24–29

Tan RKZ, Petrov AS, Harvey SC (2006) YUP: a molecular simulation program for coarse-grained and multiscaled models. J Chem Theory Comput 2(3):529–540

ten Dam E, Brierley I, Inglis S, Pleij C (1994) Identification and analysis of the pseudoknot-containing gag-pro ribosomal frameshift signal of simian retrovirus-1. Nucleic Acids Res 22(12):2304–2310

Tzeng TH, Tu CL, Bruenn JA (1992) Ribosomal frameshifting requires a pseudoknot in the Saccharomyces cerevisiae double-stranded RNA virus. J Virol 66(2):999–1006

Villoutreix BO, Eudes R, Miteva MA (2009) Structure-based virtual ligand screening: recent success stories. Comb Chem High Throughput Screen 12(10):1000–1016

Whitty A, Kumaravel G (2006) Between a rock and a hard place? Nat Chem Biol 2(3):112–118

Winkler W, Nahvi A, Breaker RR (2002) Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. Nature 419(6910):952–956

Yang M (2005) Discoveries of Tat-TAR interaction inhibitors for HIV-1. Curr Drug Targets Infect Disord 5(4):433–444

Yang H, Jossinet F, Leontis N, Chen L, Westbrook J, Berman H, Westhof E (2003) Tools for the automatic identification and classification of RNA base pairs. Nucleic Acids Res 31(13):3450–3460

Zhang Q, Stelzer AC, Fisher CK, Al-Hashimi HM (2007) Visualizing spatially correlated dynamics that directs RNA conformational transitions. Nature 450(7173):1263–1267

Zhou R, Friesner RA, Ghosh A, Rizzo RC, Jorgensen WL, Levy RM (2001) New linear interaction method for binding affinity calculations using a continuum solvent model. J Phys Chem B 105:10388–10397

Zwieb C, Müller F (1997) Three-dimensional comparative modeling of RNA. Nucleic Acids Symp Ser 36:69–71