QRB Discovery

www.cambridge.org/qrd

Computational drug discovery under RNA times

Mattia Bernetti[1,2*], Riccardo Aguti[1,2], Stefano Bosio[1,2], Maurizio Recanatini[2], Matteo Masetti[2] and Andrea Cavalli[1,2*]

[1]Computational and Chemical Biology, Italian Institute of Technology, 16152 Genova, Italy and [2]Department of Pharmacy and Biotechnology, Alma Mater Studiorum-University of Bologna, 40126 Bologna, Italy

Perspective

Cite this article: Bernetti M, Aguti R, Bosio S, Recanatini M, Masetti M, Cavalli A (2022). Computational drug discovery under RNA times. *QRB Discovery*, **3**: e22, 1–10
https://doi.org/10.1017/qrd.2022.20

Received: 08 June 2022
Revised: 02 November 2022
Accepted: 04 November 2022

Keywords:
Molecular Dynamics simulations; Molecular Docking; RNA selectivity; RNA-binding drugs; non-coding RNA

Author for correspondence:
*Mattia Bernetti,
E-mail: mattia.bernetti@iit.it;
Andrea Cavalli,
E-mail: andrea.cavalli@unibo.it

Abstract

RNA molecules play many functional and regulatory roles in cells, and hence, have gained considerable traction in recent times as therapeutic interventions. Within drug discovery, structure-based approaches have successfully identified potent and selective small-molecule modulators of pharmaceutically relevant protein targets. Here, we embrace the perspective of computational chemists who use these traditional approaches, and we discuss the challenges of extending these methods to target RNA molecules. In particular, we focus on recognition between RNA and small-molecule binders, on selectivity, and on the expected properties of RNA ligands.

Introduction

RNA's biological relevance has traditionally been ascribed to its role as an intermediate in the flow of genetic information from DNA to the production of functional proteins. However, it has become increasingly evident that RNA has many functional and regulatory roles in all domains of life. For instance, RNA molecules can regulate gene expression directly or by interacting with small organic molecules (miRNA, Pasquinelli *et al.,* 2005; riboswitches, Serganov and Nudler, 2013, respectively), and can exert enzymatic activity (ribozymes, Doudna and Cech, 2002). Therefore, being involved in many cellular pathways, RNA molecules offer opportunities as targets for the development of therapeutic strategies (Shortridge and Varani, 2015; Connelly *et al.,* 2016; Matsui and Corey, 2017; Warner *et al.,* 2018; Falese *et al.,* 2021). However, most of the drug discovery efforts have focused on proteins, which have long been known to modulate cellular activity. Nevertheless, in the human genome only a small fraction of the transcribed RNA is translated into proteins (Fig. 1) (Warner *et al.,* 2018; ENCODE Project Consortium, 2012; Oliver *et al.,* 2020), and only a small portion of these proteins has been successfully targeted (Warner *et al.,* 2018). The growing characterisation of noncoding RNA, therefore, offers new opportunities for novel therapeutic approaches, which may be particularly valuable when seeking alternatives in cases of drug resistance or when traditionally undruggable protein targets are encountered.

Computational approaches have reached the status of standard tools in drug discovery campaigns (Macalino *et al.,* 2015). This is particularly true for structure-based drug design (Jorgensen, 2004), where the macromolecule's structural information is used to find small molecules able to bind and modulate its activity. These computer-aided approaches have mostly been used to identify drugs for protein targets, as RNA has only recently been fully recognised as a relevant pharmaceutical target. Moreover, RNA targets are particularly challenging for standard computational approaches due to their complex structural dynamics and high charge density. Indeed, the applicability of standard approaches to RNA targets is still being debated and is a hot topic for research (Fedorova *et al.,* 2018; Warner *et al.,* 2018; Juru and Hargrove, 2021; Manigrasso *et al.,* 2021).

In this Perspective article, we put ourselves in the shoes of computational medicinal chemists who have experience in targeting proteins and who wish to extend their expertise to RNA to find potent and selective small-molecule ligands. We discuss critical aspects associated with this transition, which may require additional operations or some rethinking of standard procedures. The Perspective is divided into three sections that reflect the main challenges of computational RNA-targeted drug discovery (Fig. 2). First, we address RNA-small molecule recognition, particularly how RNA dynamics should be treated and how to identify potential binders. Second, we address small-molecule selectivity for RNA targets. Third, we address RNA binders and their expected physicochemical properties. We conclude with an outlook on future opportunities for computational medicinal chemists on the way to a wider playground, where to apply and expand their expertise.

RNA-small molecule recognition

Rational drug discovery campaigns typically start with a biomolecular target that has been pre-clinically validated. Validated targets play a critical role in a physio-pathological process and their

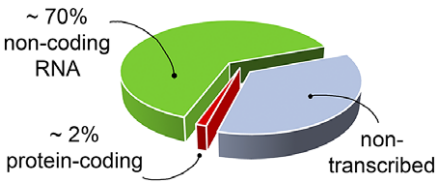

Fig. 1. The targetable portions of the human genome. More than 70% of the human genome is transcribed into RNA, but only a small portion of this encodes for, and is thus translated into, proteins (red slice) (ENCODE Project Consortium, 2012; Oliver *et al.,* 2020), of which, only a small fraction has been successfully targeted with drugs (Warner *et al.,* 2018). The possibility of targeting non-coding functional RNA molecules (green slice) could significantly increase the number of drug discovery strategies.

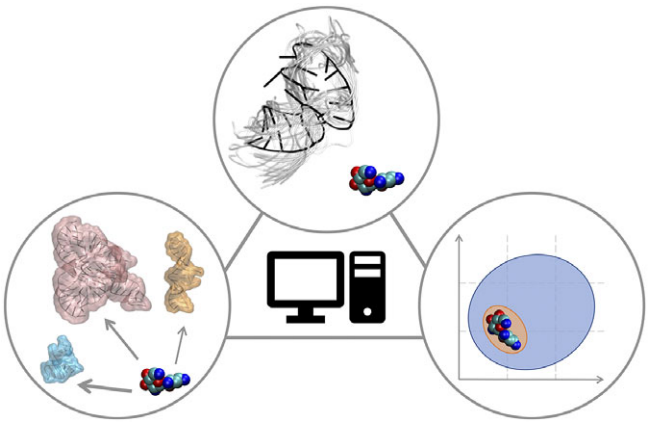

Fig. 2. RNA-targeted computational drug discovery. A schematic representation of the Perspective's three main sections: structural dynamics of the target in RNA-ligand recognition (top), target selectivity (left) and physicochemical properties of RNA binders (right).

modulation is likely to produce a therapeutic effect, hopefully within an acceptable safety window. Drug discovery pipelines, therefore, begin by identifying small-molecule hits for a validated target (Hughes *et al.,* 2011). In a structure-based context, computational strategies for this stage include fragment-based approaches (Erlanson *et al.,* 2016), de novo design (Schneider and Fechner, 2005), and, most importantly, virtual screening of small-molecule libraries (Maia *et al.,* 2020). In a virtual screening campaign, a small-molecule library undergoes molecular docking, a procedure that aims at predicting how the ligands bind to the target. This computational approach is rapid and can be used to roughly discriminate between binders and non-binders. As such, it has become a well-established strategy for the identification of small-molecule hits.

There are two particularly critical aspects involved in extending docking protocols, that have long been refined over proteins, to RNA targets: i) how to describe the target's structure, and in particular how to account for its intrinsic flexibility and structural dynamics, and ii) how to assess quantitatively the binding poses. These aspects are already critical in the context of protein targets but may be even more crucial for RNA targets. Concerning the former aspect, any docking campaign requires the target's structure as the starting point. Experimental methods to reconstruct the 3D structures at the atomistic level include X-ray crystallography, nuclear magnetic resonance (NMR), and cryo-electron microscopy (Palamini *et al.,* 2016). In the absence of experimental data, modelling approaches can be taken advantage of to reconstruct the target's structure with varying degrees of accuracy, depending on the available information (Hameduh *et al.,* 2020). In this respect, the machine-learning based approach AlphaFold (Jumper *et al.,* 2021) is achieving impressive results in protein 3D-structure prediction, however, an equivalent for RNA does not exist yet. For protein targets, X-ray crystallography experiments have long been established as an efficient method to produce high-resolution structures. Therefore, a docking effort on a protein target typically begins with a crystal structure. In contrast, a significant fraction (about 40%) of RNA structures is solved via NMR (Barnwal *et al.,* 2017), while crystal structures can be often found in the cases of larger and structurally complex RNA molecules.

While a reliable structure is certainly a great starting point, it may however not suffice for a comprehensive description of the target of interest. Indeed, biomolecules are not frozen entities, but rather present in solution various degrees of structural flexibility. This dynamic nature not only can influence the binding of molecular partners, but structural modifications can also be triggered upon the binding of smaller molecules. Therefore, including information about the dynamics of the target during a docking procedure can depict more realistically what really occurs at the molecular level and potentially improve results. This is a long-standing issue in computational drug discovery (Feixas *et al.,* 2014). Indeed, protein flexibility is typically addressed in docking protocols through different approaches (Buonfiglio *et al.,* 2015), such as soft-docking (Ferrari *et al.,* 2004) and induced-fit docking (Sherman *et al.,* 2006). These strategies can be considered an integral part of the docking protocol. While they do not usually affect the performance of the docking calculation much, they do account for a rather limited structural flexibility of the target. An alternative approach is to include the receptor dynamics as an ensemble of multiple conformations (Huang and Zou, 2007; Amaro *et al.,* 2018). This strategy allows greater conformational changes of the target, but the docking calculation scales up linearly as the procedure must be iterated for each of the structures in the ensemble (Fig. 3).

From a practical standpoint, the ensemble is generated separately from the docking calculation using other computational methods. This ensemble docking approach has been used for protein targets (Amaro *et al.,* 2018). Compared to common protein targets, RNA molecules display marked and complex structural dynamics (Ganser *et al.,* 2019). Therefore, the ensemble docking approach appears as the natural choice and should indeed be preferred for RNA targets. The literature contains some successful examples in this direction (Stelzer *et al.,* 2011; Ganser *et al.,* 2018).

A diverse set of computational approaches can be used to generate RNA conformational ensembles. In particular, static frameworks could be employed to generate a pool of RNA structures, as it is done through the popular Fragment Assembly of RNA with Full-Atom Refinement (FARFAR) algorithm in the Rosetta software suite (Watkins *et al.,* 2020). In contrast, methods that mimic the dynamics, such as molecular dynamics (MD) simulations, can be employed in the generation of a conformational ensemble (Sponer *et al.,* 2018). MD explores the conformational dynamics of biomolecules under realistic conditions (e.g., explicit solvent, quasi-physiological ionic concentrations) and has become an indispensable tool for investigating mechanistic features at the atomistic level (De Vivo *et al.,* 2016; Decherchi and Cavalli, 2020). Notably, the results of MD simulations strongly depend on the ability of the underlying model (i.e., the force field) to capture the physics of the interactions in molecular systems. Since structural biology and drug discovery have long been focused on proteins, force fields for RNA have developed at a much slower pace (Table 1).

https://doi.org/10.1017/qrd.2022.20 Published online by Cambridge University Press

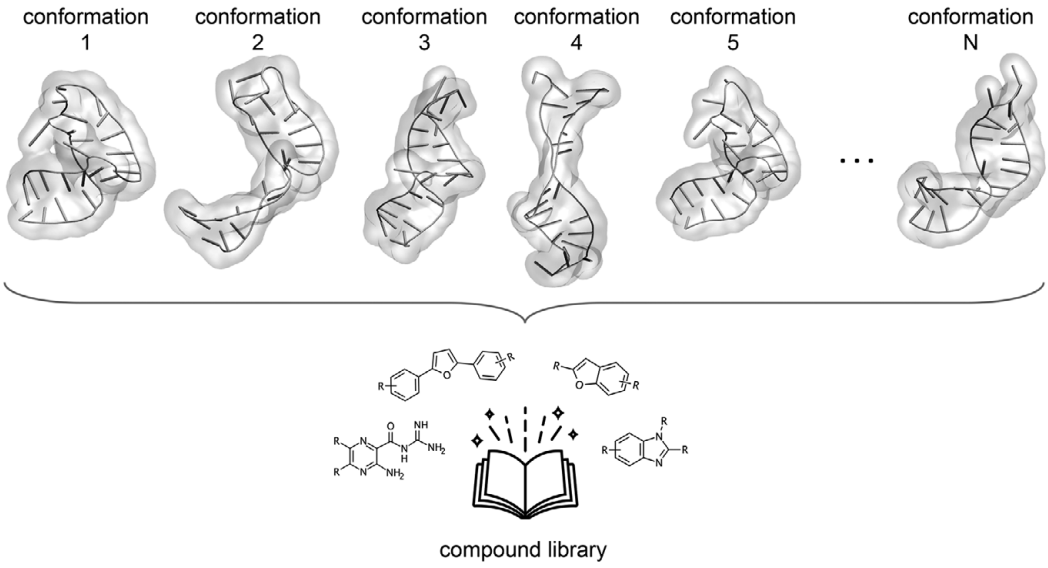

Fig. 3. Ensemble docking. An ensemble comprising multiple conformations of the target is included to take into account its structural dynamics. The docking calculation (virtual screening for large libraries) is repeated for each structure in the ensemble. The RNA structures here belong to the conformational ensemble of the transactivation response element (TAR) RNA from human immunodeficiency virus type-1 reconstructed in Salmon *et al.* (2013).

Among the Amber family of force fields, the ff99 force field combined with the bsc0 and χ_{OL3} refinements is considered state-of-the-art, as it is the most validated and widely used (Wang *et al.,* 2000; Pérez *et al.,* 2007; Zgarbová *et al.,* 2011). In particular, ff99 (Wang *et al.,* 2000) is a major branch of Amber force fields that includes parameters for both proteins and nucleic acids and was based on the previous ff94 version (Cornell *et al.,* 1995). For DNA and RNA, the main difference between ff94 and ff99 is the refinement of the sugar puckering and χ dihedral parameters. In 2007, Orozco and coworkers introduced a major correction, known as bsc0, where the α and γ dihedral angles of the nucleic acid backbone were modified to avoid the formation of nonnative γ-trans backbone dihedral states, thus reducing unrealistic helical twists in A-RNA (Pérez *et al.,* 2007). In 2011, the χ_{OL3} refinement involved a reparameterisation of the χ dihedral to prevent high-*anti* γ shifts in RNA, which led to entirely untwisted and ladder-like structures (Zgarbová *et al.,* 2011). In more recent attempts to improve the force field, dihedral reparameterisation was conducted in the Mathews group (Aytenfisu *et al.,* 2017), while dihedral, electrostatic and van der Waals parameters were considered by the Shaw group (Tan *et al.,* 2018). Recently, different schemes were also proposed, which used an additional term to better describe hydrogen-bond interactions (Fröhlking *et al.,* 2022) or introduced the grid-based energy correction map (CMAP) term in the context of RNA force fields (Chen *et al.,* 2022). The ability of the CHARMM and OPLS families of force fields to describe RNA structures is also gradually being improved (see Table 1) (Denning *et al.,* 2011; Robertson *et al.,* 2019). In addition to the force field, the thoroughness of conformational sampling is a common issue for both protein and RNA modelling. This is because the timescales that are accessible via conventional (or "plain") MD simulations are still limited (in the order of tens of microseconds for most of the research projects). Nevertheless, outstanding results have been achieved thanks to hardware advances (Pande *et al.,* 2003; Shaw *et al.,* 2014) and sophisticated enhanced sampling approaches (Abrams and Bussi, 2013; Mlýnský and Bussi, 2018) being implemented in popular instruments such as PLUMED (Tribello *et al.,* 2014). In the context of proteins, particularly challenging systems in terms of conformational sampling are intrinsically disordered proteins (Habchi *et al.,* 2014). In this respect, enhanced sampling methods were successfully employed (Granata *et al.,* 2015; Palazzesi *et al.,* 2015; Bernetti *et al.,* 2017; Masetti *et al.,* 2020), and we thus envision that they will find increasing application also on RNA molecules.

Despite recent improvements, the current force fields still bear some limits. Combined with the shortcomings linked to the sampling, they can generate RNA conformational ensembles that do not entirely agree with experiments. While this is problematic in general, it may become particularly critical when using the generated structural ensembles for docking. However, experimental information can be used to generate more reliable ensembles (Pitera and Chodera, 2012; Hummer and Köfinger, 2015; Bonomi *et al.,* 2016, 2017; Cesari *et al.,* 2018; Orioli *et al.,* 2020) in both static (Shi *et al.,* 2020) and dynamic frameworks (Bottaro *et al.,* 2018). During MD simulations, for example, experimental data can be included on the fly to guide the sampling towards regions of the conformational space that are supported by experiments. Alternatively, reweighting approaches can be applied after the MD simulation to identify those conformations that better agreed with the experimental data. Both strategies have been successfully applied to reconstruct reliable conformational ensembles of RNA molecules (Borkar *et al.,* 2013; Bottaro *et al.,* 2018). Notably, the experimental data here can come from a broader range of sources than the data on the initial atomistic structures. Indeed, any experimental information that can be related to an observable computed from the biomolecule coordinates in the MD trajectory can be exploited. Thus, experimental methods that provide coarser structural information (e.g., small-angle X-ray scattering, SAXS) are extremely valuable and have been used for this purpose (Bernetti *et al.,* 2021). Finally, clustering algorithms that have become of routine use in the context of MD simulations (Bernetti *et al.,* 2020) can be of remarkable support to select representative structures from MD-generated ensembles for the subsequent docking/virtual screening stage.

Given a biomolecular target's structure or ensemble of structures, the docking procedure attempts to find plausible ligand-target bound configurations, that is, binding modes (or "poses")

https://doi.org/10.1017/qrd.2022.20 Published online by Cambridge University Press

Table 1. Main classes of RNA force fields and their major variants

Year	FF name	Composition	Main features
		Amber	
1995	ff94 (Cornell *et al.,* 1995)	ff94	
1999	ff98 (Cheatham *et al.,* 1999)	ff94 + P + χ	Improves pucker and twist; comparable to ff94
2000	ff99 (Wang *et al.,* 2000)	ff98 + P	Improves pucker; small changes
2007	parmbsc0 (Pérez *et al.,* 2007)	ff99 + $\alpha\gamma_{bsc0}$	Avoids nonnative α/γ conformations but penalises native γ ones
2011	ff99bsc0χ_{OL3} (Zgarbová *et al.,* 2011)	ff99 + $\alpha\gamma_{bsc0}$ + χ_{OL3}	Prevents high-*anti* g shifts; state-of-the-art force field for amber
2010	Amber99χ (Yildirim *et al.,* 2010)	ff99 + χ_{YIL}	Reduces ladder-like structures and the A-form inclination
2012	Amber99TOR (Yildirim *et al.,* 2012)	ff99 + χ_{YIL} + $\beta,\varepsilon,\zeta_{YIL}$ + $\alpha\gamma_{bsc0}$	Improves the description of cytidine and uridine in solution; performs suboptimally for canonical RNA
2017	Aytenfisu−Spasic−Stern−Mathews (Aytenfisu *et al.* (2017)	ff99 + $\alpha\beta\gamma\varepsilon\zeta\chi_{Mathews}$	Performs well for tetraloops; reduces intercalation events
2013	Chen−Garcia (Chen and García (2013)	ff99 + vdW_{GC} + χ_{GC}	Reduces stacking, but overstabilises hydrogen-bond interactions between bases
2018	Tan−Piana−Dirks−Shaw (Tan *et al.* (2018)	ff99bsc0χ_{OL3} + $\gamma\zeta\chi_{Shaw}$ + vdW_{Shaw} + $electrostatics_{Shaw}$	Focused on non-bonded interactions; the conformational ensembles closely reproduce experimental ones
2022	gHBfix21 (Fröhlking *et al.,* 2022)	ff99bsc0χ_{OL3} + HbFix	Improves hydrogen-bond interactions, stabilises native structures
2022	ff99OL3_CMAP1 (Chen *et al.,* 2022)	ff99bsc0χ_{OL3} + ζ/α_{CMAP}	Decreases population of incorrect structures; improves stability of tetranucleotides
		CHARMM	
1995	CHARMM22 (MacKerell *et al.,* 1995)	CHARMM22	
2000	CHARMM27 (MacKerell *et al.,* 2000)	CHARMM27	Significant improvements over CHARMM22; known issues with pair opening
2011	CHARMM36 (Denning *et al.,* 2011)	CHARMM27 + 2'-OH dihedral	Partial stabilisation of the structures; state-of-the-art force field for CHARMM
		OPLS	
1991	OPLS-AA (Pranata *et al.,* 1991)	OPLS-AA	
2019	OPLS-AA/M (Robertson *et al.,* 2019)	OPLS-AA + $\alpha\gamma\chi$ + P	Reduces intercalation events and is well-suited to describing non-canonical motifs

with overall favourable interactions between the two binding partners. We here briefly outline some popular docking software packages and refer the interested reader to a recent comprehensive overview (Zhou *et al.,* 2021). Glide (Friesner *et al.,* 2004), GOLD (Jones *et al.,* 1997) and AutoDock Vina (Trott and Olson, 2010) are software packages that were devised for protein targets and that can also be used for RNA when needed. Conversely, AutoDock (Morris *et al.,* 1998), DOCK 6 (Lang *et al.,* 2009) and ICM (Abagyan *et al.,* 1994) have been improved to be used with RNA through dedicated docking protocols, optimised ligand-sampling algorithms, or the inclusion of solvation effects either in the generation of poses or in the scoring functions. Finally, the software packages MORDOR (Guilbert and James, 2008), rDOCK (Ruiz-Carmona *et al.,* 2014), and the very recent RLDOCK (Sun *et al.,* 2020) and NLDock (Feng *et al.,* 2021) were specifically developed for RNA docking, reflecting the growing interest in RNA-oriented drug discovery. The trend that emerges from the validation of the latter approaches on diverse datasets of experimental RNA-ligand complexes, usually evaluated as the success rate in reproducing experimental binding poses, is that RNA-specific methods outperform the tools developed for proteins or generic macromolecules (Feng *et al.,* 2021; Zhou *et al.,* 2021). Notably, such trend highlights the relevance of specifically considering interaction and structural features that are peculiar of RNA-ligand binding.

The quality of the poses identified by the docking procedure can be assessed through a variety of strategies that fall under the term "scoring functions". These can be broadly classified as the follows: i) knowledge-based, when the scoring is built upon information extracted from known three-dimensional structures of target-ligand complexes; ii) physics-based, when the scoring is based on force fields or simplified empirical functions of the target-ligand interactions and iii) machine learning (ML)-based, when the scoring is evaluated through ML models trained on available experimental data (Zhou *et al.,* 2021). Scoring functions are typically included in docking software. However, there has been a rapid growth in standalone options, including the knowledge-based ITScore-NL scoring function (Feng and Huang, 2020) and the ML-based RNAPosers (Chhabra *et al.,* 2020), RNAmigos (Oliver

https://doi.org/10.1017/qrd.2022.20 Published online by Cambridge University Press

et al., 2020) and AnnapuRNA (Stefaniak and Bujnicki, 2021) scoring functions. Notably, MD simulations can also be used for scoring (Menchon *et al.,* 2018). Indeed, binding pose stability can be assessed with plain MD runs or, alternatively, binding affinities can be estimated with more computationally intensive MD-based free energy calculations (Decherchi and Cavalli, 2020). These procedures can straightforwardly be applied to RNA targets. The only caveats are the accuracy of the force field and, most importantly, the limited number of RNA-ligand complexes that can typically be managed via these approaches, which hinders their use for large-scale virtual screenings.

Selectively targeting RNA structures

Most functional proteins targeted in drug discovery campaigns fold in a well-defined native state under physiological conditions. The functional activity of proteins usually takes place in surface cavities. In structure-based approaches, structural information about these cavities is directly used to design ligands that might bind there (Pérot *et al.,* 2010). If the target of interest is well-characterised, relevant binding sites may already be known. Often, however, either the binding site is unknown or alternative binding sites may be sought (e.g., for allosteric modulation) (Kuzmanic *et al.,* 2020). The possibility of identifying binding pockets by computational means is thus integral to modern structure-based drug design. The SiteMap tool of the Schrödinger suite (Halgren, 2009), the ICM PocketFinder (An *et al.,* 2005), the NanoShaper software suite (Decherchi and Rocchia, 2013) and its dynamic extension Pocketron (La Sala *et al.,* 2017) are popular options in this regard (Pérot *et al.,* 2010). These tools have been widely employed to detect pockets in proteins. Although some of them have also been used to identify pockets in RNA molecules (Ganser *et al.,* 2018; Hewitt *et al.,* 2019; Panei *et al.,* 2022), an extensive exploration of their applicability is still missing. In proteins, suitable binding pockets have a well-defined 3D organisation of the amino acids comprised therein, which, alongside their composition and variability, display a wealth of physicochemical features (e.g., balance between hydrophobic/hydrophilic regions, solvent exposure and shape) that make the pockets rather distinctive. Taken together, these aspects encode, to a certain extent, the target selectivity that can potentially be achieved by addressing that site (Ehrt *et al.,* 2016; Smilova *et al.,* 2022). Usually, however, this information is not extensively exploited in the early stages of the drug discovery pipeline, where most of the effort is directed towards the identification of potential hits. Conversely, it is only during later phases of lead optimisation that target selectivity is fully explored, together with chemical modifications that can improve the affinity and ADMET (absorption, distribution, metabolism, excretion and toxicity) properties.

Moving into the context of RNA, molecules with functional roles in cellular pathways display a markedly heterogeneous structural complexity. Indeed, functional RNA molecules vary from short hairpin loops and miRNA, to tRNA, riboswitches and larger ribozymes, up to the scale of the ribosome (Ganser *et al.,* 2019). From a structure-based drug discovery perspective, we can broadly distinguish two main scenarios depending on the complexity of the RNA target's molecular structure.

In the first scenario, relatively simple secondary structure elements can be identified as hot spots for small-molecule binding. These short stem-loop motifs include for instance apical loops, bulges and internal loops, and have been extensively studied in this regard (Liu *et al.,* 2004; Disney and Childs-Disney, 2007; Meyer and Hergenrother, 2009). Here, the structure-based strategy is made possible by the unpaired nucleobases, which can arrange in characteristic structures and thus offer the possibility of small-molecule binding (Juru and Hargrove, 2021). However, their pockets are usually shallow or relatively small, which makes it challenging to find high-affinity binders that produce specific interactions (Fig. 4) (Warner *et al.,* 2018; Juru and Hargrove, 2021). Indeed, shallow pockets pose a biophysical limit to how potent non-irreversible binders can be. Furthermore, this strategy may prove particularly arduous in terms of selectivity because similar secondary structure elements can be found across diverse RNA molecules. Indeed, only modest potency and selectivity have been achieved in reported studies focused on these types of RNA molecules (Warner *et al.,* 2018). However, this strategy may be particularly valuable and viable in the absence of a more complex tertiary structure, which offers more characteristic architectures for potential drug binding (Juru and Hargrove, 2021).

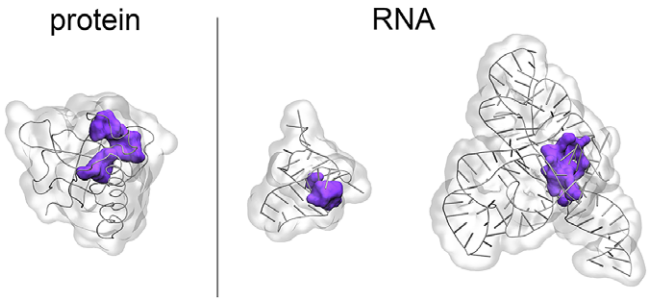

Fig. 4. Protein and RNA binding pockets. Binding pockets in proteins (left, riboflavin kinase, PDBID: 1NB9) are typically highly structured. In RNAs, structured pockets like those in proteins are found in highly folded structures (right, FMN riboswitch, PDBID: 3F4G). In contrast, relatively simple RNA structures (middle, HIV TAR, PDBID: 1QD3) usually offer shallow or relatively small pockets, which are more challenging to target with small molecules. The pockets shown herein (in violet) were identified with the NanoShaper software (Decherchi and Rocchia, 2013). The two RNA structures were chosen as representatives of good and intermediate quality pockets from the examples reported by Warner *et al.* (2018): in this work, pocket quality was estimated using the ICM tool PocketFinder (An *et al.,* 2005), where pockets of larger size and buriedness resulted in higher quality.

The second scenario involves RNA molecules achieving a higher level of folding and thus complex 3D structures. Although RNA molecules have less chemical variety than proteins (4 nucleotides vs 21 amino acids), this complex folding can nevertheless produce distinctive cavities reminiscent of protein-like binding pockets (Warner *et al.,* 2018; Hewitt *et al.,* 2019). These cavities can be suitable for computational approaches already established in the context of protein targets. In such cases, the determinants of RNA-ligand recognition are likely to be similar to those of protein-ligands, with no need to develop alternative RNA-centric approaches (Fedorova *et al.,* 2018). Riboswitches and ribozymes are remarkable examples of highly folded RNA molecules with complex tertiary structures (Zafferani and Hargrove, 2021). Since those interact with metabolites or substrates, they are already prone to ligand recognition in pre-formed binding pockets. Multi-junctions and pseudoknots, in general, have also been suggested as promising RNA species for RNA-targeted drug discovery because their great structural complexity is suitable for pocket formation (Warner *et al.,* 2018). Despite this, to the best of our knowledge, the literature contains just one report of a successful campaign of rational drug design using classical, established medicinal chemistry protocols (Fedorova *et al.,* 2018). This work used high-throughput experimental assays to identify hits. However, computational approaches could also be employed for hit

https://doi.org/10.1017/qrd.2022.20 Published online by Cambridge University Press

identification of these RNA molecules, so compound libraries could potentially be investigated on a much larger scale.

Given this overall picture, at variance with protein targets, the discourse around selectivity becomes more urgent already at an earlier stage when targeting RNA molecules. This is because the choice of a certain class of RNA targets can greatly impact the level of selectivity that can be achieved. While the second scenario discussed for RNA appears to hold more promise for identifying selective binders to modulate RNA activity, this may however preclude opportunities to develop effective drugs for pathological conditions mediated by structurally simpler RNAs (Juru and Hargrove, 2021). We, therefore, encourage researchers to be open to both scenarios while considering their respective implications for selectivity.

Properties of RNA ligands

To identify drug candidates, it is essential to know the physicochemical features that ligands should possess in order to bind to a particular class of biomolecular targets. In computational drug discovery, this knowledge can be instrumental to design libraries for virtual screening or to guide the lead optimisation stage of a candidate. For proteins, drug discovery usually aims to identify small organic molecules with physicochemical profiles that meet the criteria of oral drugs, including solubility, bioavailability, cell and tissue permeability, chemical stability and absence of toxicity. In this respect, Lipinski's rule of five is the established guiding principle for rational drug design (Lipinski, 2004). Indeed, through a retrospective analysis of approved drugs and drug candidates, Lipinski's rule of five empirically set the drug-likeness boundaries for physicochemical parameters including molecular weight, lipophilicity and number of hydrogen-bond donors and acceptors. This, in conjunction with the wealth of knowledge accumulated through years of experience in the field, has led to the identification of a rather defined chemical space that is characteristic of protein-targeting small organic molecules. Similarly, the knowledge gained in decades of successes and failures in drug discovery campaigns has allowed to compile a list of undesirable chemical features that for several reasons (mostly non-specific interference with biological assays) should not be possessed by drugs, the so-called PAINS (Baell and Holloway, 2010). In structure-based virtual screening, Lipinski's rule of five and PAINS filters are typically applied before the molecular docking. This is to avoid wasting time on performing docking calculations of molecules that will likely be discarded anyway, regardless of their ability to bind to the target.

For RNA targets, the chemical space of small-molecule binders has not been fully characterised yet. Since relatively few small organic molecule binders of RNA targets are known, their expected properties are not yet established and are a hot research topic (Warner *et al.,* 2018; Juru and Hargrove, 2021). Early identified ligands that acted by binding RNA had a positive net charge and were able to intercalate between RNA bases (Thomas and Hergenrother, 2008; Guan and Disney, 2012). However, such physicochemical properties cause non-specific binding on the negatively charged RNA backbone, yielding low selectivity. For this reason, ligands may display a relatively high level of toxicity, therefore often resulting non-viable. Recently, research efforts based on the analysis of ligands with activities towards RNA were directed to the characterisation of the physicochemical space of RNA small-molecule binders (Morgan *et al.,* 2017, 2019; Haniff *et al.,* 2020; Rizvi *et al.,* 2020). The overall picture that is gradually emerging points to an RNA-privileged chemical space. The most

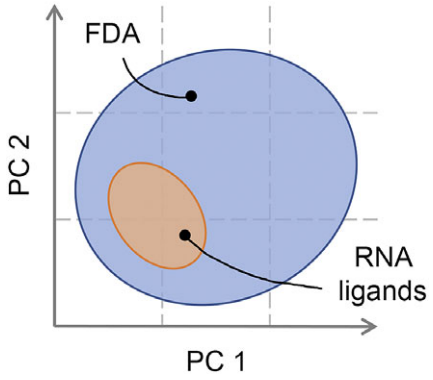

Fig. 5. The chemical space of RNA-binding ligands. Bioactive RNA-targeted compounds populate a region in chemical space (here projected along two hypothetical principal components of cheminformatic parameters) occupied by FDA-approved drugs, which mostly target proteins (Juru and Hargrove, 2021). Therefore, while RNA ligands have particular structural and shape properties, they can also possess the typical drug-like properties.

peculiar structural features identified in RNA-targeting ligands appear to be a higher nitrogen count, a lower oxygen count, an enrichment in aromatic rings, fewer stereocenters and fewer sp3-hybridised carbon atoms (Morgan *et al.,* 2019; Haniff *et al.,* 2020). Additionally, in contrast to the more heterogeneous spatial arrangements of the approved drugs, a prevalence of rod- and planar-like shapes (Wirth and Sauer, 2011) has also been observed. Interestingly, we note how, as a whole, these features are reminiscent of the nucleobases. Most remarkably, this RNA-privileged chemical space appears to be a subset of the space traditionally occupied by protein-binding ligands and, more generally, by orally administered drugs (Fig. 5) (Juru and Hargrove, 2021). An interesting implication of this latter aspect is that there is a real potential for these molecules to be RNA-targeting therapeutics. From a more practical perspective, the identified characteristics represent indispensable instruments for computational medicinal chemists to refine screening libraries and to guide the optimisation of promising binders.

While it is essential to take advantage of this knowledge, the number of known RNA binders is still limited, so the exploration and definition of the chemical space of RNA-binding ligands has only just begun. Therefore, further advances may come into play and refine the current picture in the (possibly near) future. Indeed, as a further level of complexity, recently discovered potent and selective ligands of an RNA ribozyme contained chemical groups that would usually be classified as PAINS (Fedorova *et al.,* 2018). This highlights how critical it can be to apply rules of thumb, which were developed to target proteins, for RNA targeting, since the ligands' chemical space is still under construction. Given our limited knowledge, it is therefore important to build screening libraries by taking a non-RNA-biased approach, increasing chemical diversity and maintaining drug-likeness.

In conclusion, we are still in the process of comprehensively characterising the physicochemical properties of RNA binders. Therefore, while the boundaries of their features are gradually becoming clearer, room for potential expansion in this respect should nevertheless be contemplated.

Outlook and concluding remarks

The growing recognition of RNAs as promising pharmaceutical targets requires a mindset change in drug discovery. In this

https://doi.org/10.1017/qrd.2022.20 Published online by Cambridge University Press

Perspective, we discussed the three main challenges of computational RNA-targeted drug discovery: i) the prominent role of target flexibility in predicting small-molecule binding; ii) the importance of achieving binding selectivity; and iii) the knowledge of the chemical space expected for RNA-binding drugs.

We have discussed how the currently available computational procedures, which have been optimised and refined over decades of efforts directed to protein targets, can be used in the novel and partially unexplored context of RNA targets, and how to take appropriate precautions. For example, recent scoring functions were specifically developed to describe the interaction of small molecules with RNA, which means that well-established docking protocols can be easily adapted to RNA targets. Moreover, we discussed how MD simulations, which are extensively used for lead optimisation in protein-based drug discovery, will become an essential tool for considering RNA target flexibility in the earlier stages of screening campaigns. Furthermore, experimental data can increase the reliability of RNA conformational ensembles reconstructed via MD. Interestingly, we note how experimental information often has a dual function when targeting RNA, since the experimental data can be used as the source of starting structures for the simulations and as a guide to refine the conformational ensembles. Moreover, MD simulations (possibly combined with enhanced sampling methods and appropriate analysis tools) may help identify suitable pockets to focus on in the search for specific interactions. In this respect, by showing a greater amount of structural complexity, RNA motifs such as riboswitches, ribozymes, multi-junctions and pseudoknots are intrinsically more inclined to pocket formation, and thus they are better suited for RNA-targeted discovery of small molecule drugs. Given the above and relative to traditional protein-based drug discovery, computational medicinal chemists may need a more skilled background in statistical mechanics and simulative methods in order to take full advantage of these approaches.

Once the chemical space of RNA-targeting drugs has been defined, machine learning and artificial intelligence, which are revolutionising several aspects of conventional drug discovery (Vamathevan *et al.,* 2019), will be critical to developing novel active compounds. Indeed, the most recently developed scoring functions for docking tend to be based on machine learning approaches, and artificial intelligence is already being leveraged to advance the area of force field improvement.

The landscape of currently available small-molecule drugs targeting RNA is somewhat limited. Current drugs are antibiotics targeting ribosomal RNA (rRNA), such as the synthetic oxazolidinone linezolid that acts by binding a highly structured pocket, and the very recent Roche and PTC Therapeutics' risdiplam, used in the treatment of spinal muscular atrophy (SMA), which acts by stabilising the interaction between an RNA splice site and a small nuclear ribonucleoprotein (snRNP) (Sheridan, 2021). Furthermore, Merck's ribocil inhibits bacterial growth by binding to a bacterial riboswitch, however, this small molecule is on hold at the preclinical stage due to the rapid development of bacterial resistance and unlikely will be pursued further (Warner *et al.,* 2018). Despite being limited in number, all these examples support the idea that RNA is a legitimate target of small molecules and highlights the potential of focusing on RNA molecules with high structural complexity to achieve high affinity and selectivity.

Finally, a mention is here required on other types of RNA molecules (not extensively covered in this contribution) as pharmaceutical targets, which are currently in the market/clinical trials for major unmet medical needs. Remarkable examples thereof are the recently approved risdiplam for SMA and the Novartis's branaplam, under clinical trial for both SMA and Huntington disease, which act as splicing modulators by binding to pre-mRNA (Childs-Disney *et al.,* 2022).

In conclusion, time has come for computational drug discovery to embrace the potential of RNA to become an established drug target shortly. Indeed, thanks to the growing interest in discovering small molecules that target RNA, the field is gradually producing useful resources, such as the recent HARIBOSS database of RNA-small molecule structures (Panei *et al.,* 2022). In the same spirit, efforts by the computational community in sharing simulation inputs via dedicated resources (e.g., the PLUMED-NEST initiative (Bonomi *et al.,* 2019)), and sharing computational practises via Jupyter Notebooks (Kluyver *et al.,* 2016), are likely to accelerate the expansion of more complex and sectorial computational skills to successful drug discovery.

Acknowledgements. Giovanni Bussi, Angelo Favia and Thorben Fröhlking are gratefully acknowledged for carefully reading the manuscript and providing useful suggestions.

Open peer review. To view the open peer review materials for this article, please visit http://doi.org/10.1017/qrd.2022.20.

References

Abagyan R, Totrov M and Kuznetsov D (1994) ICM—A new method for protein modeling and design: Applications to docking and structure prediction from the distorted native conformation. *Journal of Computational Chemistry* **15**(5), 488–506.

Abrams C and Bussi G (2013) Enhanced sampling in molecular dynamics using Metadynamics, replica-exchange, and temperature-acceleration. *Entropy* **16** (1), 163–199.

Amaro RE, Baudry J, Chodera J, Demir Ö, McCammon JA, Miao Y and Smith JC (2018) Ensemble docking in drug discovery. *Biophysical Journal* **114**(10), 2271–2278.

An J, Totrov M and Abagyan R (2005) Pocketome via comprehensive identification and classification of ligand binding envelopes. *Molecular & Cellular Proteomics* **4**(6), 752–761.

Aytenfisu AH, Spasic A, Grossfield A, Stern HA and Mathews DH (2017) Revised RNA dihedral parameters for the Amber force field improve RNA molecular dynamics. *Journal of Chemical Theory and Computation* **13**(2), 900–915.

Baell JB and Holloway GA (2010) New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. *Journal of Medicinal Chemistry* **53**(7), 2719–2740.

Barnwal RP, Yang F and Varani G (2017) Applications of NMR to structure determination of RNAs large and small. *Archives of Biochemistry and Biophysics* **628**, 42–56.

Bernetti M, Bertazzo M and Masetti M (2020) Data-driven molecular dynamics: A multifaceted challenge. *Pharmaceuticals* **13**(9), 253.

Bernetti M, Hall KB and Bussi G (2021) Reweighting of molecular simulations with explicit-solvent SAXS restraints elucidates ion-dependent RNA ensembles. *Nucleic Acids Research* **49**(14), e84.

Bernetti M, Masetti M, Pietrucci F, Blackledge M, Jensen MR, Recanatini M, Mollica L and Cavalli A (2017) Structural and kinetic characterization of the intrinsically disordered protein SeV NTAIL through enhanced sampling simulations. *The Journal of Physical Chemistry B* **121**(41), 9572–9582.

Bonomi M, Bussi G, Camilloni C, Tribello GA, Banáš P, Barducci A, Bernetti M, Bolhuis PG, Bottaro S and Branduardi D (2019) Promoting transparency and reproducibility in enhanced molecular simulations. *Nature Methods* **16**(8), 670–673.

Bonomi M, Camilloni C, Cavalli A and Vendruscolo M (2016) Metainference: A Bayesian inference method for heterogeneous systems. *Science Advances* **2**(1), e1501177.

Bonomi M, **Heller GT**, **Camilloni C and Vendruscolo M** (2017) Principles of protein structural ensemble determination. *Current Opinion in Structural Biology* **42**, 106–116.

Borkar AN, **De Simone A**, **Montalvao RW and Vendruscolo M** (2013) A method of determining RNA conformational ensembles using structure-based calculations of residual dipolar couplings. *The Journal of Chemical Physics* **138**(21), 06B604_1.

Bottaro S, **Bussi G**, **Kennedy SD**, **Turner DH and Lindorff-Larsen K** (2018) Conformational ensembles of RNA oligonucleotides from integrating NMR and molecular simulations. *Science Advances* **4**(5), eaar8521.

Buonfiglio R, **Recanatini M and Masetti M** (2015) Protein flexibility in drug discovery: From theory to computation. *ChemMedChem* **10**(7), 1141–1148.

Cesari A, **Reißer S and Bussi G** (2018) Using the maximum entropy principle to combine simulations and solution experiments. *Computation* **6**(1), 15.

Cheatham III T, **Piotr Cieplak E and Kollman PA** (1999) A modified version of the Cornell et al. force field with improved sugar pucker phases and helical repeat. *Journal of Biomolecular Structure and Dynamics* **16**(4), 845–862.

Chen AA and García AE (2013) High-resolution reversible folding of Hyperstable RNA Tetraloops using molecular dynamics simulations. *Proceedings of the National Academy of Sciences* **110**(42), 16820–16825.

Chen J, **Liu H**, **Cui X**, **Li Z and Chen H-F** (2022) RNA-specific force field optimization with CMAP and reweighting. *Journal of Chemical Information and Modeling* **62**, 372–385.

Chhabra S, **Xie J and Frank AT** (2020) RNAPosers: Machine learning classifiers for ribonucleic acid–ligand poses. *The Journal of Physical Chemistry B* **124** (22), 4436–4445.

Childs-Disney JL, **Yang X**, **Gibaut QMR**, **Tong Y**, **Batey RT and Disney MD** (2022) Targeting RNA structures with small molecules. *Nature Reviews Drug Discovery* **21**, 736–762.

Connelly CM, **Moon MH and Schneekloth Jr JS** (2016) The emerging role of RNA as a therapeutic target for small molecules. *Cell Chemical Biology* **23**(9), 1077–1090.

Cornell WD, **Cieplak P**, **Bayly CI**, **Gould IR**, **Merz KM**, **Ferguson DM**, **Spellmeyer DC**, **Fox T**, **Caldwell JW and Kollman PA** (1995) A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *Journal of the American Chemical Society* **117**(19), 5179–5197.

De Vivo M, **Masetti M**, **Bottegoni G and Cavalli A** (2016) Role of molecular dynamics and related methods in drug discovery. *Journal of Medicinal Chemistry* **59**(9), 4035–4061.

Decherchi S and Cavalli A (2020) Thermodynamics and kinetics of drug-target binding by molecular simulation. *Chemical Reviews* **120**(23), 12788–12833.

Decherchi S and Rocchia W (2013) A general and robust ray-casting-based algorithm for triangulating surfaces at the nanoscale. *PLoS One* **8**(4), e59744.

Denning EJ, **Priyakumar UD**, **Nilsson L and Mackerell Jr AD** (2011) Impact of 2′-hydroxyl sampling on the conformational properties of RNA: Update of the CHARMM all-atom additive force field for RNA. *Journal of Computational Chemistry* **32**(9), 1929–1943.

Disney MD and Childs-Disney JL (2007) Using selection to identify and chemical microarray to study the RNA internal loops recognized by 6′-N-acylated kanamycin a. *Chembiochem* **8**(6), 649–656.

Doudna JA and Cech TR (2002) The chemical repertoire of natural ribozymes. *Nature* **418**(6894), 222–228.

Ehrt C, **Brinkjost T and Koch O** (2016) Impact of binding site comparisons on medicinal chemistry and rational molecular design. *Journal of Medicinal Chemistry* **59**(9), 4121–4151.

ENCODE Project Consortium (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**(7414), 57.

Erlanson DA, **Fesik SW**, **Hubbard RE**, **Jahnke W and Jhoti H** (2016) Twenty years on: The impact of fragments on drug discovery. *Nature Reviews Drug Discovery* **15**(9), 605–619.

Falese JP, **Donlic A and Hargrove AE** (2021) Targeting RNA with small molecules: From fundamental principles towards the clinic. *Chemical Society Reviews* **50**(4), 2224–2243.

Fedorova O, **Jagdmann GE**, **Adams RL**, **Yuan L**, **Van Zandt MC and Pyle AM** (2018) Small molecules that target group II introns are potent antifungal agents. *Nature Chemical Biology* **14**(12), 1073–1078.

Feixas F, **Lindert S**, **Sinko W and McCammon JA** (2014) Exploring the role of receptor flexibility in structure-based drug discovery. *Biophysical Chemistry* **186**, 31–45.

Feng Y and Huang S-Y (2020) ITScore-NL: An iterative knowledge-based scoring function for nucleic acid–ligand interactions. *Journal of Chemical Information and Modeling* **60**(12), 6698–6708.

Feng Y, **Zhang K**, **Wu Q**, **and Huang S-Y** (2021) NLDock: A fast nucleic acid–ligand docking algorithm for modeling RNA/DNA–ligand complexes. *Journal of Chemical Information and Modeling* **61**(9), 4771–4782.

Ferrari AM, **Wei BQ**, **Costantino L and Shoichet BK** (2004) Soft docking and multiple receptor conformations in virtual screening. *Journal of Medicinal Chemistry* **47**(21), 5076–5084.

Friesner RA, **Banks JL**, **Murphy RB**, **Halgren TA**, **Klicic JJ**, **Mainz DT**, **Repasky MP**, **Knoll EH**, **Shelley M and Perry JK** (2004) Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *Journal of Medicinal Chemistry* **47**(7), 1739–1749.

Fröhlking T, **Mlýnský V**, **Janeček M**, **Kührová P**, **Krepl M**, **Banáš P**, **Šponer J and Bussi G** (2022) Automatic learning of hydrogen-bond fixes in the AMBER RNA force field. *Journal of Chemical Theory and Computation* **18** (7), 4490–4502. https://doi.org/10.1021/acs.jctc.2c00200

Ganser LR, **Kelly ML**, **Herschlag D and Al-Hashimi HM** (2019) The roles of structural dynamics in the cellular functions of RNAs. *Nature Reviews Molecular Cell Biology* **20**(8), 474–489.

Ganser LR, **Lee J**, **Rangadurai A**, **Merriman DK**, **Kelly ML**, **Kansal AD**, **Sathyamoorthy B and Al-Hashimi HM** (2018) High-performance virtual screening by targeting a high-resolution RNA dynamic ensemble. *Nature Structural & Molecular Biology* **25**(5), 425–434.

Granata D, **Baftizadeh F**, **Habchi J**, **Galvagnion C**, **De Simone A**, **Camilloni C**, **Laio A and Vendruscolo M** (2015) The inverted free energy landscape of an intrinsically disordered peptide by simulations and experiments. *Scientific Reports* **5**(1), 1–15.

Guan L and Disney MD (2012) Recent advances in developing small molecules targeting RNA. *ACS Chemical Biology* **7**(1), 73–86.

Guilbert C and James TL (2008) Docking to RNA via root-Mean-Square-deviation-driven energy minimization with flexible ligands and flexible targets. *Journal of Chemical Information and Modeling* **48**(6), 1257–1268.

Habchi J, **Tompa P**, **Longhi S and Uversky VN** (2014) Introducing protein intrinsic disorder. *Chemical Reviews* **114**(13), 6561–6588.

Halgren TA (2009) Identifying and characterizing binding sites and assessing druggability. *Journal of Chemical Information and Modeling* **49**(2), 377–389.

Hameduh T, **Haddad Y**, **Adam V and Heger Z** (2020) Homology modeling in the time of collective and artificial intelligence. *Computational and Structural Biotechnology Journal* **18**, 3494–3506.

Haniff HS, **Knerr L**, **Liu X**, **Crynen G**, **Boström J**, **Abegg D**, **Adibekian A**, **Lekah E**, **Wang KW and Cameron MD** (2020) Design of a small molecule that stimulates vascular endothelial growth factor a enabled by screening RNA fold–small molecule interactions. *Nature Chemistry* **12**(10), 952–961.

Hewitt WM, **Calabrese DR and Schneekloth Jr JS** (2019) Evidence for Ligandable sites in structured RNA throughout the protein data Bank. *Bioorganic & Medicinal Chemistry* **27**(11), 2253–2260.

Huang S-Y and Zou X (2007) Ensemble docking of multiple protein structures: Considering protein structural variations in molecular docking. *Proteins: Structure, Function, and Bioinformatics* **66**(2), 399–421.

Hughes JP, **Stephen Rees SBK and Philpott KL** (2011) Principles of early drug discovery. *British Journal of Pharmacology* **162**(6), 1239–1249.

Hummer G and Köfinger J (2015) Bayesian ensemble refinement by replica simulations and reweighting. *The Journal of Chemical Physics* **143**(24), 12B634_1.

Jones G, **Willett P**, **Glen RC**, **Leach AR and Taylor R** (1997) Development and validation of a genetic algorithm for flexible docking. *Journal of Molecular Biology* **267**(3), 727–748.

Jorgensen WL (2004) The many roles of computation in drug discovery. *Science* **303**(5665), 1813–1818.

Jumper J, **Evans R**, **Pritzel A**, **Green T**, **Figurnov M**, **Ronneberger O**, **Tunyasuvunakool K**, **Bates R**, **Žídek A and Potapenko A** (2021) Highly accurate protein structure prediction with AlphaFold. *Nature* **596**(7873), 583–589.

Juru AU and Hargrove AE (2021) Frameworks for targeting RNA with small molecules. *Journal of Biological Chemistry* **296**, 100191.

https://doi.org/10.1017/qrd.2022.20 Published online by Cambridge University Press

Kluyver T, **Ragan-Kelley B**, **Pérez F**, **Granger B**, **Bussonnier M**, **Frederic J**, **Kelley K**, **Hamrick J**, **Grout J**, **Corlay S**, **Ivanov, P.**, **Avila, D.**, **Abdalla, S. and Willing, C.**, (2016) Jupyter development team. Jupyter Notebooks – a Publishing Format for Reproducible Computational Workflows. In *20th International Conference on Electronic Publishing(01/01/16)*; Loizides, F., Scmidt, B., Eds.; IOS Press, 2016; pp 87–90.

Kuzmanic A, **Bowman GR**, **Juarez-Jimenez J**, **Michel J and Gervasio FL** (2020) Investigating cryptic binding sites by molecular dynamics simulations. *Accounts of Chemical Research* **53**(3), 654–661.

La Sala G, **Decherchi S**, **De Vivo M and Rocchia W** (2017) Allosteric communication networks in proteins revealed through pocket crosstalk analysis. *ACS Central Science* **3**(9), 949–960.

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

Lipinski CA (2004) Lead-and drug-like compounds: The rule-of-five revolution. *Drug Discovery Today: Technologies* **1**(4), 337–341.

Liu X, **Thomas JR and Hergenrother PJ** (2004) Deoxystreptamine dimers bind to RNA hairpin loops. *Journal of the American Chemical Society* **126**(30), 9196–9197.

Macalino SJY, **Gosu V**, **Hong S and Choi S** (2015) Role of computer-aided drug Design in Modern Drug Discovery. *Archives of Pharmacal Research* **38**(9), 1686–1701.

MacKerell Jr AD, **Banavali N and Foloppe N** (2000) Development and current status of the CHARMM force field for nucleic acids. *Biopolymers: Original Research on Biomolecules* **56**(4), 257–265.

MacKerell Jr AD, **Wiorkiewicz-Kuczera J and Karplus M** (1995) An all-atom empirical energy function for the simulation of nucleic acids. *Journal of the American Chemical Society* **117**(48), 11946–11975.

Maia EHB, **Assis LC**, **De Oliveira TA**, **Da Silva AM and Taranto AG** (2020) Structure-based virtual screening: From classical to artificial intelligence. *Frontiers in Chemistry* **8**, 343.

Manigrasso J, **Marcia M and De Vivo M** (2021) Computer-aided design of RNA-targeted small molecules: A growing need in drug discovery. *Chem* **7**(11), 2965–2988.

Masetti M, **Bernetti M and Cavalli A** (2020) Enhanced molecular dynamics simulations of intrinsically disordered proteins. *Methods in Molecular Biology* **2141**, 391–411

Matsui M and Corey DR (2017) Non-coding RNAs as drug targets. *Nature Reviews Drug Discovery* **16**(3), 167–179.

Menchon G, **Maveyraud L and Czaplicki G** (2018) Molecular dynamics as a tool for virtual ligand screening. *Methods in Molecular Biology* **1762**, 145–178.

Meyer ST and Hergenrother PJ (2009) Small molecule ligands for bulged RNA secondary structures. *Organic Letters* **11**(18), 4052–4055.

Mlýnský V and Bussi G (2018) Exploring RNA structure and dynamics through enhanced sampling simulations. *Current Opinion in Structural Biology* **49**, 63–71.

Morgan BS, **Forte JE**, **Culver RN**, **Zhang Y and Hargrove AE** (2017) Discovery of key physicochemical, structural, and spatial properties of RNA-targeted bioactive ligands. *Angewandte Chemie International Edition* **56**(43), 13498–13502.

Morgan BS, **Sanaba BG**, **Donlic A**, **Karloff DB**, **Forte JE**, **Zhang Y and Hargrove AE** (2019) R-BIND: An interactive database for exploring and developing RNA-targeted chemical probes. *ACS Chemical Biology* **14**(12), 2691–2700.

Morris GM, **Goodsell DS**, **Halliday RS**, **Huey R**, **Hart WE**, **Belew RK and Olson AJ** (1998) Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *Journal of Computational Chemistry* **19**(14), 1639–1662.

Oliver C, **Mallet V**, **Gendron RS**, **Reinharz V**, **Hamilton WL**, **Moitessier N and Waldispühl J** (2020) Augmented base pairing networks encode RNA-small molecule binding preferences. *Nucleic Acids Research* **48**(14), 7690–7699.

Orioli S, **Larsen AH**, **Bottaro S and Lindorff-Larsen K** (2020) How to learn from inconsistencies: Integrating molecular simulations with experimental data. *Progress in Molecular Biology and Translational Science* **170**, 123–176.

Palamini M, **Canciani A and Forneris F** (2016) Identifying and visualizing macromolecular flexibility in structural biology. *Frontiers in Molecular Biosciences* **3**, 47.

Palazzesi F, **Prakash MK**, **Bonomi M and Barducci A** (2015) Accuracy of current all-atom force-fields in modeling protein disordered states. *Journal of Chemical Theory and Computation* **11**(1), 2–7.

Pande VS, **Baker I**, **Chapman J**, **Elmer SP**, **Khaliq S**, **Larson SM**, **Rhee YM**, **Shirts MR**, **Snow CD and Sorin EJ** (2003) Atomistic protein folding simulations on the submillisecond time scale using worldwide distributed computing. *Biopolymers: Original Research on Biomolecules* **68**(1), 91–109.

Panei FP, **Torchet R**, **Menager H**, **Gkeka P and Bonomi M** (2022) HARIBOSS: A curated database of RNA-small molecules structures to aid rational drug design. *Bioinformatics* **38**, 4185–4193. https://doi.org/10.1093/bioinformatics/btac483

Pasquinelli AE, **Hunter S and Bracht J** (2005) MicroRNAs: A developing story. *Current Opinion in Genetics & Development* **15**(2), 200–205.

Pérez A, **Marchán I**, **Svozil D**, **Sponer J**, **Cheatham III TE**, **Laughton CA and Orozco M** (2007) Refinement of the AMBER force field for nucleic acids: Improving the description of α/γ conformers. *Biophysical Journal* **92**(11), 3817–3829.

Pérot S, **Sperandio O**, **Miteva MA**, **Camproux A-C and Villoutreix BO** (2010) Druggable pockets and binding site centric chemical space: A paradigm shift in drug discovery. *Drug Discovery Today* **15**(15–16), 656–667.

Pitera JW and Chodera JD (2012) On the use of experimental observations to bias simulated ensembles. *Journal of Chemical Theory and Computation* **8**(10), 3445–3451.

Pranata J, **Wierschke SG and Jorgensen WL** (1991) OPLS potential functions for nucleotide bases. Relative association constants of hydrogen-Bonded Base pairs in chloroform. *Journal of the American Chemical Society* **113**(8), 2810–2819.

Rizvi NF, **Maria Jr JPS Nahvi A**, **Klappenbach J**, **Klein DJ**, **Curran PJ**, **Richards MP**, **Chamberlin C**, **Saradjian P and Burchard J** (2020) Targeting RNA with small molecules: Identification of selective, RNA-binding small molecules occupying drug-like chemical space. *SLAS DISCOVERY: Advancing the Science of Drug Discovery* **25**(4), 384–396.

Robertson MJ, **Qian Y**, **Robinson MC**, **Tirado-Rives J and Jorgensen WL** (2019) Development and testing of the OPLS-AA/M force field for RNA. *Journal of Chemical Theory and Computation* **15**(4), 2734–2742.

Ruiz-Carmona S, **Alvarez-Garcia D**, **Nicolas Foloppe ABG-D**, **Juhos S**, **Schmidtke P**, **Barril X**, **Hubbard RE and Morley SD** (2014) RDock: A fast, versatile and open source program for docking ligands to proteins and nucleic acids. *PLoS Computational Biology* **10**(4), e1003571.

Salmon L, **Bascom G**, **Andricioaei I and Al-Hashimi HM** (2013) A general method for constructing atomic-resolution RNA ensembles using NMR residual dipolar couplings: The basis for Interhelical motions revealed. *Journal of the American Chemical Society* **135**(14), 5457–5466.

Schneider G and Fechner U (2005) Computer-based de novo Design of Drug-like Molecules. *Nature Reviews Drug Discovery* **4**(8), 649–663.

Serganov A and Nudler E (2013) A decade of riboswitches. *Cell* **152**(1–2), 17–24.

Shaw DE, **Grossman JP**, **Bank JA**, **Batson B**, **Butts JA**, **Chao JC**, **Deneroff MM**, **Dror RO**, **Even A and Fenton CH** (2014) *Anton 2: Raising the Bar for Performance and Programmability in a Special-Purpose Molecular Dynamics Supercomputer*. New Orleans, LA, USA: IEEE, pp. 41–53.

Sheridan C (2021) First small-molecule drug targeting RNA gains momentum. *Nature Biotechnology* **39**(1), 6–9.

Sherman W, **Day T**, **Jacobson MP**, **Friesner RA and Farid R** (2006) Novel procedure for modeling ligand/receptor induced fit effects. *Journal of Medicinal Chemistry* **49**(2), 534–553.

Shi H, **Rangadurai A**, **Assi HA**, **Roy R**, **Case DA**, **Herschlag D**, **Yesselman JD and Al-Hashimi HM** (2020) Rapid and accurate determination of atomistic RNA dynamic ensemble models using NMR and structure prediction. *Nature Communications* **11**(1), 1–14.

Shortridge MD and Varani G (2015) Structure based approaches for targeting non-coding RNAs with small molecules. *Current Opinion in Structural Biology* **30**, 79–88.

Smilova MD, **Curran PR**, **Radoux CJ**, **von Delft F**, **Cole JC**, **Bradley AR and Marsden BD** (2022) Fragment hotspot mapping to identify selectivity-determining regions between related proteins. *Journal of Chemical Information and Modeling* **62**, 284–294.

https://doi.org/10.1017/qrd.2022.20 Published online by Cambridge University Press

Sponer J, **Bussi G**, **Krepl M**, **Banáš P**, **Bottaro S**, **Cunha RA**, **Gil-Ley A**, **Pinamonti G**, **Poblete S and Jurečka P** (2018) RNA structural dynamics as captured by molecular simulations: A comprehensive overview. *Chemical Reviews* **118**(8), 4177–4338.

Stefaniak F and Bujnicki JM (2021) AnnapuRNA: A scoring function for predicting RNA-small molecule binding poses. *PLoS Computational Biology* **17**(2), e1008309.

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

Sun L-Z, **Jiang Y**, **Zhou Y and Chen S-J** (2020) RLDOCK: A new method for predicting RNA–ligand interactions. *Journal of Chemical Theory and Computation* **16**(11), 7173–7183.

Tan D, **Piana S**, **Dirks RM and Shaw DE** (2018) RNA force field with accuracy comparable to state-of-the-art protein force fields. *Proceedings of the National Academy of Sciences* **115**(7), E1346–E1355.

Thomas JR and Hergenrother PJ (2008) Targeting RNA with small molecules. *Chemical Reviews* **108**(4), 1171–1224.

Tribello GA, **Bonomi M**, **Branduardi D**, **Camilloni C and Bussi G** (2014) PLUMED 2: New feathers for an old bird. *Computer Physics Communications* **185**(2), 604–613.

Trott O and Olson AJ (2010) AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry* **31**(2), 455–461.

Vamathevan J, **Clark D**, **Czodrowski P**, **Dunham I**, **Ferran E**, **Lee G**, **Li B**, **Madabhushi A**, **Shah P and Spitzer M** (2019) Applications of machine learning in drug discovery and development. *Nature Reviews Drug Discovery* **18**(6), 463–477.

Wang J, **Cieplak P and Kollman PA** (2000) How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *Journal of Computational Chemistry* **21**(12), 1049–1074.

Warner KD, **Hajdin CE and Weeks KM** (2018) Principles for targeting RNA with drug-like small molecules. *Nature Reviews Drug Discovery* **17**(8), 547–558.

Watkins AM, **Rangan R and Das R** (2020) FARFAR2: Improved de novo Rosetta prediction of complex global RNA folds. *Structure* **28**(8), 963–976.

Wirth M and Sauer WHB (2011) Bioactive molecules: Perfectly shaped for their target? *Molecular Informatics* **30**(8), 677–688.

Yildirim I, **Kennedy SD**, **Stern HA**, **Hart JM**, **Kierzek R and Turner DH** (2012) Revision of AMBER torsional parameters for RNA improves free energy predictions for tetramer duplexes with GC and IGiC Base pairs. *Journal of Chemical Theory and Computation* **8**(1), 172–181.

Yildirim I, **Stern HA**, **Kennedy SD**, **Tubbs JD and Turner DH** (2010) Reparameterization of RNA χ torsion parameters for the AMBER force field and comparison to NMR spectra for cytidine and uridine. *Journal of Chemical Theory and Computation* **6**(5), 1520–1531.

Zafferani M and Hargrove AE (2021) Small molecule targeting of biologically relevant RNA tertiary and quaternary structures. *Cell Chemical Biology* **28**(5), 594–609.

Zgarbová M, **Otyepka M**, **Šponer J**, **Mládek A**, **Banáš P**, **Thomas E, Cheatham and Jurecka P** (2011) Refinement of the Cornell et al. nucleic acids force field based on reference quantum chemical calculations of Glycosidic torsion profiles. *Journal of Chemical Theory and Computation* **7**(9), 2886–2902.

Zhou Y, **Jiang Y and Chen S-J** (2022) RNA–ligand molecular docking: Advances and challenges. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **12**, e1571.

https://doi.org/10.1017/qrd.2022.20 Published online by Cambridge University Press