RSC Medicinal Chemistry

ROYAL SOCIETY OF CHEMISTRY

RESEARCH ARTICLE

View Article Online
View Journal

Cite this: DOI: 10.1039/d0md00167h

Systematic analysis of the interactions driving small molecule–RNA recognition†

G. Padroni, [a] N. N. Patwardhan, [a] M. Schapira[bc] and A. E. Hargrove *[a]

Received 20th May 2020,
Accepted 20th May 2020

DOI: 10.1039/d0md00167h

rsc.li/medchem

RNA molecules are becoming an important target class in drug discovery. However, the principles for designing RNA-binding small molecules are yet to be fully uncovered. In this study, we examined the Protein Data Bank (PDB) to highlight privileged interactions underlying small molecule–RNA recognition. By comparing this analysis with previously determined small molecule–protein interactions, we find that RNA recognition is driven mostly by stacking and hydrogen bonding interactions, while protein recognition is instead driven by hydrophobic effects. Furthermore, we analyze patterns of interactions to highlight potential strategies to tune RNA recognition, such as stacking and cation–π interactions that favor purine and guanine recognition, and note an unexpected paucity of backbone interactions, even for cationic ligands. Collectively, this work provides further understanding of RNA–small molecule interactions that may inform the design of small molecules targeting RNA.

1. Introduction

In cells, RNA has a myriad of functions including the transfer of genetic information, modulation of protein synthesis and interactions, and regulation of RNA processing, transcription and gene expression.[1–3] For this reason, a comprehensive knowledge of how to design compounds capable of selectively targeting RNA has the potential to extend our understanding of RNA-mediated cellular processes and increase the number of druggable targets in cells.[4,5] To date, the majority of FDA-approved small molecules designed by medicinal chemists are aimed at protein binding pockets.[6] As a result, high-throughput screening (HTS) campaigns of commercially available small molecules produce lower hit-rates when the chosen target is RNA rather than a protein.[7,8] However, identifying the principles of small molecule synthetic design for RNA recognition is not straightforward. In fact, structure-driven approaches for targeting RNA are impeded by a lack of diversity in high-resolution RNA structures, particularly larger RNA, which are notoriously challenging to obtain.[9,10] This lack of structural information for small molecule targeting is conventionally justified by the fact that RNA is highly dynamic and less prone to form binding pockets that would be observable by traditional structural methods such as X-ray diffraction and NMR.[4] Furthermore, the RNA conformational landscape can influence thermodynamic and kinetic binding properties and consequently the types of interaction driving binding events.[11] However, recent studies on Protein Data Bank (PDB) deposited RNA structures revealed that ligandable regions are present among the multiple conformations assumed by RNA, particularly in structures with more complex folds such as riboswitches.[12] These RNA binding pockets tend to have similar properties – such as volume and buriedness – to protein binding pockets despite being overall less hydrophobic. Still, an expected yet striking difference between RNA and protein binding pockets is the limited chemical diversity that is carried by the four nucleobases of RNA compared to the wider chemical diversity provided by the twenty amino acids in proteins. As a consequence, RNA binding pockets might favor interactions that are currently underrepresented in available small molecule libraries that were designed for protein targets. Preliminary evidence was found by our laboratory[13] and others[14–17] when comparing chemo-informatic parameters obtained for biologically active ligands known to bind RNA (R-BIND) or enriched in fragments with RNA preferential binding as compared to FDA-approved small molecules, which are predominantly aimed at targeting proteins. These analyses highlighted that RNA-binding small molecules can satisfy medicinal chemistry properties[4,13,16,18] but have a higher number of aromatic and/or heteroaromatic rings, an increased number of hydrogen bond acceptors/donors and nitrogen atoms, and reduced fraction of sp^3 atoms compared to FDA-approved small molecules, further suggesting that some interactions are

[a] Department of Chemistry, Duke University, Durham, North Carolina 27708, USA. E-mail: amanda.hargrove@duke.edu
[b] Structural Genomics Consortium, University of Toronto, Toronto, ON M5G 1L7, Canada
[c] Department of Pharmacology and Toxicology, University of Toronto, Toronto, ON M5S 1A8, Canada

† Electronic supplementary information (ESI) available. See DOI: 10.1039/d0md00167h

privileged in RNA recognition. A broader definition of the structural requirement of small molecules to target RNA has the potential to facilitate the discovery of active compounds by directing the synthetic effort of novel RNA-biased libraries that would more likely provide higher hit-rates in HTS and/or expedite hit-to-lead optimization.

Herein, we provide direct evidence of privileged interactions between small molecules and RNA by analyzing thousands of molecular contacts found in small molecule–RNA X-ray-determined structures available in the PDB and by comparing them with the previously determined molecular interactions between small molecules and proteins.[19] Despite the remarkably limited number of high-resolution (<2.7 Å) unique small molecule–RNA complexes deposited in the PDB compared to proteins at the time this work was conducted (37 *vs.* 11 106),[19] our analysis shows emerging trends of preferred interactions for RNA recognition in line with chemo-informatics analyses. In particular, we highlight that the contributions of hydrogen bonding and stacking interactions are predominant in RNA recognition, in contrast to protein targeting that is largely driven by hydrophobic effect. In addition, we find that nucleobases dominate binding and recognition interactions, with few backbone interactions observed. Despite observations of small molecule class-dependent interaction patterns that might be biased due to the redundancy of the RNA targets considered, we show that the trend correlates with the chemo-informatic differences highlighted between RNA bioactive binding molecules and FDA approved small molecules. Finally, we provide a detailed analysis of contact patterns between RNA and small molecules for each class of interaction and by comparing them with ligand–protein interaction, we attempt to highlight the features of the interactions that are typical of RNA binders. Collectively, this work highlights the differences in molecular recognition between small molecules targeting RNA and proteins and provides important information to consider for designing and optimizing RNA focused small molecule ligands.

2. Method

Structures deposited in the PDB containing small molecules interacting with RNA determined by X-ray crystallography and with resolution lower than 2.7 Å were extracted and considered for our analysis. This dataset includes 37 unique ligands targeting 14 different structures that comprise a series of naturally occurring and synthetic small molecules. As our analysis is predominately focused on small molecule properties, redundant RNA constructs targeted by different small molecules were retained to include a sufficient number of structures for our study. Buffer components and ions were initially excluded from the analysis. Ionic interactions that involved direct coordination between the ligand and positive ions were separately determined for each structure. Despite the limited number of small molecule–RNA complexes, our analysis includes a variety of RNA secondary and tertiary structures found in therapeutically relevant RNA targets. The secondary structures include stem, asymmetric internal loops, G-quadruplexes and pseudoknots, while tertiary contacts include kissing loops and less canonically folded riboswitches.

The molecular interactions were initially extracted in ICM by using a script that assigned heavy-atom types and distances.[20] Angular parameters were also calculated for selected interactions after initial interaction sorting. Molecular contacts were classified according to previously reported parameters to allow for comparison with determined protein–small molecule interactions and include strong and weak hydrogen bonding, interactions driven by hydrophobic effect, stacking, salt bridges, cation–π and halogen-mediated interactions (Table S1†).[19] To clarify, strong and weak hydrogen bonding classification does not refer to energetic values, but only to the identity of the heavy atoms involved based on their electronegativity. Thus, strong hydrogen bonds refer to interactions involving two electronegative atoms such as N, O and S, while weak hydrogen bonding refer to CH–O interactions. Interactions driven by hydrophobic effect were assigned for van der Waals contacts between aliphatic and aromatic carbons within 4 Å from each other. Stacking interactions were identified and classified in edge-to-face or face-to-face by calculating the planar angle between aromatic atoms (Table S1†). Potential salt-bridges and cation–π interactions were assigned when ligand positively charged atoms were within 4 Å from negative oxygen phosphate atoms or aromatic atoms, respectively. The interactions count was compared to protein–small molecules interaction previously obtained following the same approach.[19]

3. Results and discussion

3.1 Most frequent RNA–small molecule interactions are different from most frequent protein–ligand contacts

The overall distribution of small molecule–RNA interactions is different from small molecule–protein contacts (Fig. 1A and B). First, stacking and hydrogen bonding interactions represent the largest class of defined interactions for the RNA dataset (34.8% and 34.4%, respectively) in contrast to proteins where they are less represented (20.2% and 15.7%, respectively). Hydrophobic contacts were only the third most frequent interactions (17.8%) in the small molecule–RNA dataset, while they are the largest class in protein–small molecule complexes (47.2%). Finally, less pronounced yet relevant differences are also observed in less frequent interactions such as weak hydrogen bonding (*i.e.*, CH–O interactions) and ionic interactions (such as salt bridges, cation–π). In particular, CH–O interactions are less frequent in RNA recognition compared to proteins (6.1% *vs.* 9.6%). The overall contributions of the electrostatic interactions (cation–π + salt bridges) for RNA and protein recognition are 5.8% and 6.9%, respectively. Cation–π interactions are almost equally frequently observed as salt

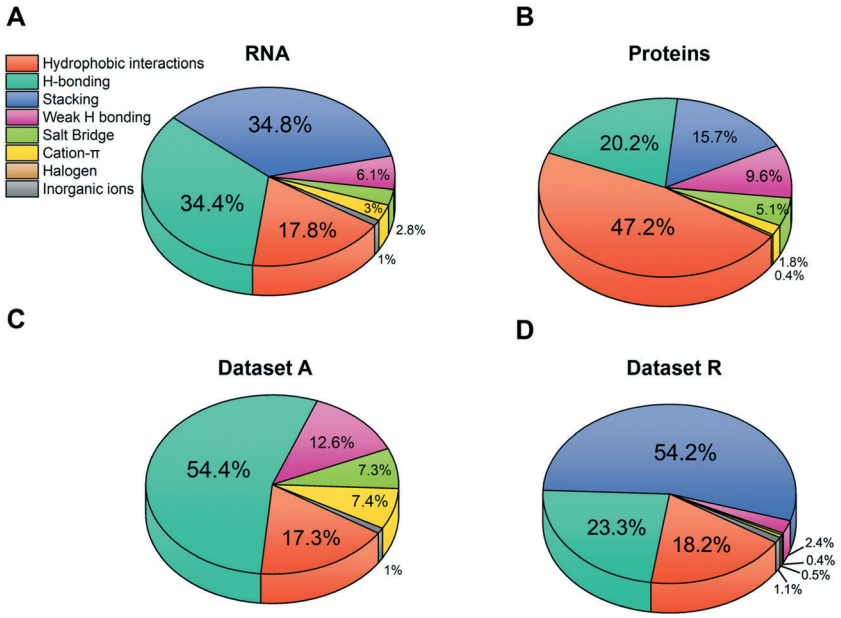

Fig. 1 Distribution of molecular interactions between (A) RNA – the totality of small molecules. (B) Protein–small molecules.[19] (C) Dataset A. (D) Dataset R (stacking interaction counts include aromatic and amide stacking).

bridges in RNA (3.0% *vs.* 2.8%), while cation–π interactions are less frequent than salt bridges in protein recognition (1.8% *vs.* 5.1%). These observations are clear reflections of the expected chemical differences between RNA and protein. The prevalence of hydrophobicity-driven contacts and higher percentages of CH–O bonding suggests that protein recognition takes place in more buried and aliphatic-rich hydrophobic binding pockets. However, it should be also considered that increasing hydrophobic effects is a common and synthetically tractable strategy for optimizing protein–ligand recognition, and thus this increased percentage might be partially due the large number of optimized ligands for protein recognition in the database.[21,22] On the other hand, the larger percentage of hydrogen bonding suggests that RNA binding pockets are more polar. Interestingly, the increased stacking interactions and the clear preference for contacts with aromatic rings such as cation–π interactions and hydrogen bonding involving heteroatoms and carbonyls of the nucleobases, indicates that the nucleobases are the major constituent of RNA binding pockets. Furthermore, the lower percentages of salt bridges that involve a direct contact between a positive and a negative charge observed for RNA is unexpected considering the overall increased net charge of RNA compared to protein due to the phosphate groups at each monomer compared to the neutral protein amide backbone and only five charged amino acid lateral chains.

Strikingly, the differences of interactions between RNA and proteins correlates well with the difference in chemo-informatics parameters observed between RNA and protein targeting small molecules.[13,17] Namely, the higher hydrogen bonding acceptor/donor content and the increased average number of aromatic and heteroaromatic rings of RNA targeting small molecules agree with the hydrogen bonding and stacking interactions being predominant in RNA recognition. This trend highlights the overlap between properties of biologically active small molecules and structural properties at the basis of RNA recognition.

We then decided to investigate the distribution of interactions after separating the ligands according to their chemical structure, particularly given that some small molecules of our analysis have chemical architectures that are distinct from conventional protein binding structures and were often not considered in the chemo-informatic analysis (Table S2†).[13] For example, aminoglycosides are largely represented in our dataset (17 out of 37 structures). These compounds are composed of multiple sugar rings linked through glycosidic bonds and widely functionalized with hydroxy and positively charged ammonium groups. Aminoglycosides bind and alter the structure of ribosomal A-site RNA, thus perturbing the recognition fidelity of tRNA for the bacterial ribosome. Although somewhat promiscuous, aminoglycosides are currently used as antibacterial agents[23] and have demonstrated distinct binding preferences.[24] Because of the distinct structural features, aminoglycosides were separated from the remaining molecules, which were mostly targeting riboswitches, to generate two datasets, namely, dataset A (aminoglycosides, ESI† 2) and dataset R (small molecules predominately targeting riboswitches, ESI† 3). Important differences are observed between these two datasets. First, the most frequent type of interaction in dataset A is strong hydrogen bonding (54.4%) followed by hydrophobicity-driven interactions (17.3%), weak hydrogen bonding (12.6%), cation–π (7.4%), salt bridges (7.3%) and coordination to inorganic ions (1%). As expected, π-stacking interactions are not observed consistently with aminoglycoside structures. Only two aminoglycoside ligands

in our dataset contained an aromatic ring, and these rings did not participate in any stacking interactions. On the other hand, stacking interactions are by far the most represented in the dataset R (54.2%) followed by strong hydrogen bonding (23.3%), hydrophobic contacts (18.2%), weak hydrogen bonding (2.4%), coordination with inorganic ions (1.1%), cation–π (0.5%) and salt bridges (0.4%). These differences clearly highlight the molecular diversity between the two datasets with ligands of dataset R being enriched in aromatic rings. Interestingly, both the datasets have similar percentages of interactions driven by the hydrophobic effect (17.3% *vs.* 18.2%), despite aminoglycosides being enriched in aliphatic carbons. Intriguingly, the chemo-informatic analysis of R-BIND and other studies captured the unique features of both dataset A and R, even though aminoglycosides were excluded from the analysis.[7,13,16] This further corroborates the fact that the structure of optimal RNA-targeting small molecules should enclose chemotypes that can form both of these privileged interactions (*i.e.*, hydrogen bonding and π-stacking). The two datasets also differ by RNA target structures, with dataset A targeting RNA secondary structures such as internal loops and dataset R targeting more complexly folded structures, which suggests that some molecular interactions are privileged for certain RNA targets. However, we caution against generalization for other structures since small molecules enriched in aromatic rings targeting RNA secondary structures such as bulges and asymmetric internal loops are well described in the literature but have not been structurally characterized.[25–29]

In general, this analysis suggests that, consistent with the architecture of known RNA ligands,[13,18] exploration of hydrogen bonding patterns and stacking interactions is a fruitful direction for improving RNA recognition by small molecules, in contrast to optimizing hydrophobicity-driven interactions as is commonly done for protein recognition.[19]

4. Specific intermolecular interactions

4.1 Stacking interactions

Aromatic stacking interactions are the most represented interactions in RNA–small molecule complexes, even though roughly half of the small molecules analysed in this work are aminoglycosides that do not contain aromatic rings and therefore do not contribute to these interactions. The higher percentages compared to small molecule–protein interactions reflect the enhanced aromatic character of RNA compared to protein, with the four nucleobases constituting binding pockets in RNA more often compared to the four aromatic side chains of tyrosine, phenylalanine, tryptophan and histidine in proteins. Interestingly, almost the totality of the aromatic interactions (764) was face-to-face π-stacking and only three contacts were classified as edge-to-face (0.004%), namely in the binding of thiamine pyrophosphate derivatives (such as **S1**) to the thiamine riboswitch (PDB ID 3D2X, 3D2Z and 2GDI, Fig. S1†). This pattern is a distinctive trait of RNA recognition compared to proteins. In fact, for protein recognition, face-to-face π-stacking and edge-to-face interactions are almost equally represented. Quantum mechanical calculations suggest that T-shaped and slipped parallel stacking between benzene rings are almost isoenergetic, indicating that other factors such as geometrical constraints due to molecular architecture might be the reason of the large prevalence of face-to-face π-stacking in RNA.[30,31] These differences underscore the fact that the chemical architecture of several nucleic acid binders favors a flat shape that allows extended aromatic surfaces to form face-to-face π-interactions with aromatic bases that even in complexly folded RNA structures tend to be orthogonal to the axis of the helices. This might also infer substantial differences in RNA binding pockets relative to proteins, that despite retaining some similar properties such as volume and buriedness, are characterized by a more homogeneous distribution of chemical functionalities. The computational prediction of nucleic acid stacking energetics is particularly challenging given the presence of multiple aromatic rings, hydrogen bonding and hydration states that are known to affect the energy of these interactions.[32] Interestingly, the distribution of stacking interactions among nucleobases is consistent with previous free energy calculations of nucleobase stacking in nucleic acid helical conformations, with the purine nucleobases having the highest interactions count (A 482, G 163) followed by the pyrimidines (U 80 and C 39, Table S3†).[33]

An extreme example of the importance of π-stacking interactions in RNA recognition is the acridine derivative **1** targeting the G-quadruplex structure TERRA (Fig. 2). This binding event is almost completely driven by stacking interactions where four molecules form a layer between two G-quadruplexes structures.[34]

We note that for amide or amide-like stacking, while highly prevalent in protein–small molecule recognition,[19] only 3 contacts are observed between small molecule and RNA, including the binding of guanidine to its riboswitch in our analysis (*vide infra*).[35]

This analysis suggests that, when designing small molecule libraries targeting RNA, extended and planar aromatic surfaces capable of stacking to multiple nucleobases might increase the binding affinity, especially for purine rich targets.[36] However, the pharmacokinetic properties of the compounds should also be considered as the presence of aromatic rings is known to affect solubility through aggregation.[37,38]

4.2 Hydrogen bonding interactions

Hydrogen bonding interactions are the second most represented (762 counts) class of contacts in RNA–small molecule interactions (Table S4†). This is not surprising considering the wide variety of hydrogen bond acceptors and donors of the RNA structure both in the sugar–phosphate backbone and nucleobases. Small molecules of our RNA dataset are also particularly enriched in hydrogen bonding donors and acceptors. This is particularly evident for dataset A where hydrogen bonding interactions are the most

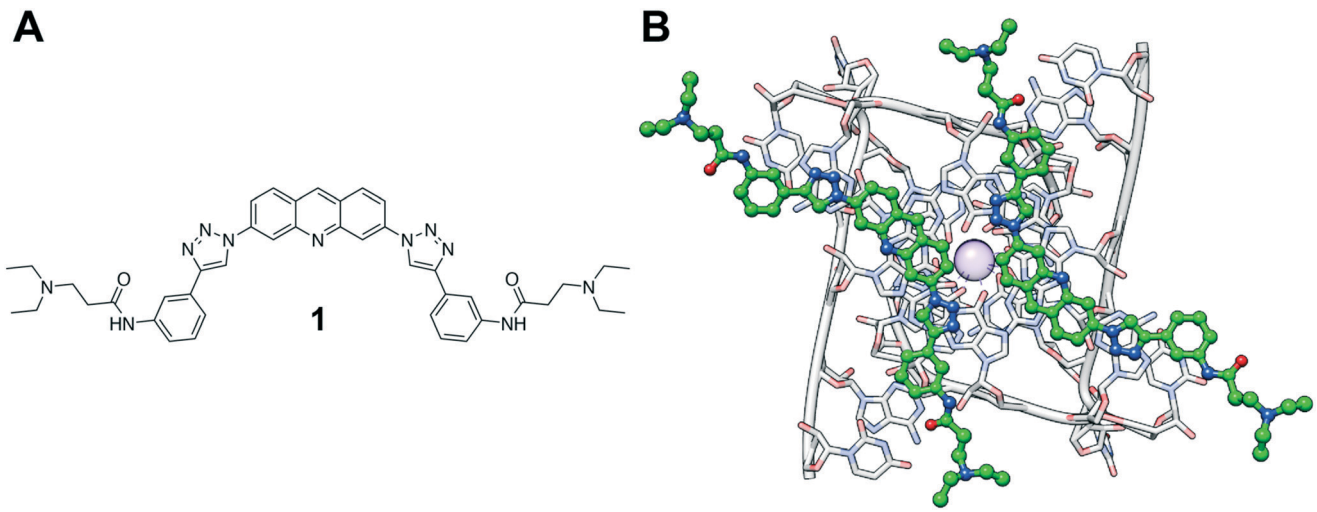

Fig. 2 (A) Chemical structure of the G-quadruplex binder acridine derivative **1**. (B) Crystal structure at 2.6 Å (PDB ID 3MIJ) of two molecules of **1** stacking in a dimeric fashion over the TERRA G-quadruplex structure.

prevalent interactions. Of the total 521 direct contacts between small molecules and RNA, NH–O are by far the most represented (250), followed by NH–N (156), OH–O (84) and OH–N (30). Interestingly, small molecule nitrogen atoms are more frequently involved either as donor and/or acceptors relative to oxygen (313 *vs.* 207), which is again consistent with the chemo-informatic observation that bioactive RNA ligands are enriched in nitrogen.[13] Only one example of direct hydrogen bonding involving a sulphur atom as acceptor was observed, namely for the thiazole ring of the thiamine pyrophosphate targeting its riboswitch.[39] However, only three structures of our dataset contains sulphur, which may indicate that sulphur-mediated interactions (also *vide infra*) are underexplored for RNA recognition. The median distances between heavy atoms are around 3 Å with the shortest median value observed for NH–O (3.17 Å) and longest for NH–N (3.44 Å) (Fig. S3†). These values are slightly higher than the one previously observed for hydrogen bonding mediating protein recognition.[19]

Hydrogen bonding events involving the nucleobases were approximately three times more frequent compared to contacts with the sugar–phosphate backbone (386 *vs.* 135). This is further evidence that nucleobases hold the majority of structural information of RNA binding pockets, at least in the structures considered in this analysis. These interactions are almost equally distributed between purines (181) and pyrimidines (205) with G being the most targeted base (120), followed by U (105), C (100) and A (61). Hydrogen bonds with nucleobases are observed both at Watson–Crick and Hoogsteen faces with both aromatic and aliphatic small molecule moieties complementing these hydrogen bond donor/acceptor patterns. For example, 2-aminopurine (**2**) binding to the guanine riboswitch uses hydrogen bonding donors and acceptors to pair with two uracil residues at their Watson–Crick side (Fig. 3A and B).[40] On the other hand, the aminoglycoside paromomycin (**3**) exploits hydroxy and ether functionalization to mimic the acceptor/donor pairing of adenine in the HIV-1 dimerization initiation site (DIS, Fig. 3C and D).[41] This is consistent with a recent analysis of hydrogen bonding interactions driving ligand–aptamer recognition in riboswitches.[42]

Of the 135 interactions observed between small molecule and RNA sugar–phosphate backbone, the majority (93) involved the negatively charged oxygen atoms of the phosphate group (Table S4†). Positively charged ammonium group hydrogen bond donors reinforced 52 of these interactions that have a median distance value ~0.2 Å shorter than the median value of the remaining NH–O interactions, while the median angle was 125° compared to 139°. This suggests that ionic charges might compensate for a less optimal geometry. Interestingly, interactions with the phosphate group are not equally distributed as the majority of these contacts involve purine nucleotides, suggesting that some factors such as local helical conformation might favour phosphate interactions.

While the 2′OH is a unique functional group of RNA that confers structural dynamics, only 21 interactions with the 2′ OH were identified (Table S4†).[43] We found that direct contact between small molecules and 2′OH are more common in structures with complex folding such as riboswitches where the structural arrangement often exposes the 2′OH to the binding pocket area as demonstrated by the guanine riboswitch (Fig. 3B).

Finally, 241 interactions (~30% of the total number of hydrogen bonds) involve water residues, which in our analysis are considered as part of the receptor. Water molecules and hydration pattern are essential for RNA thermodynamic stability, and high count of this type of interaction also suggests their importance in mediating interactions with small molecules. However, it is difficult to predict the net energetic advantages obtained from these interactions as binding free energy of hydrogen bonds is often not sufficient to compensate

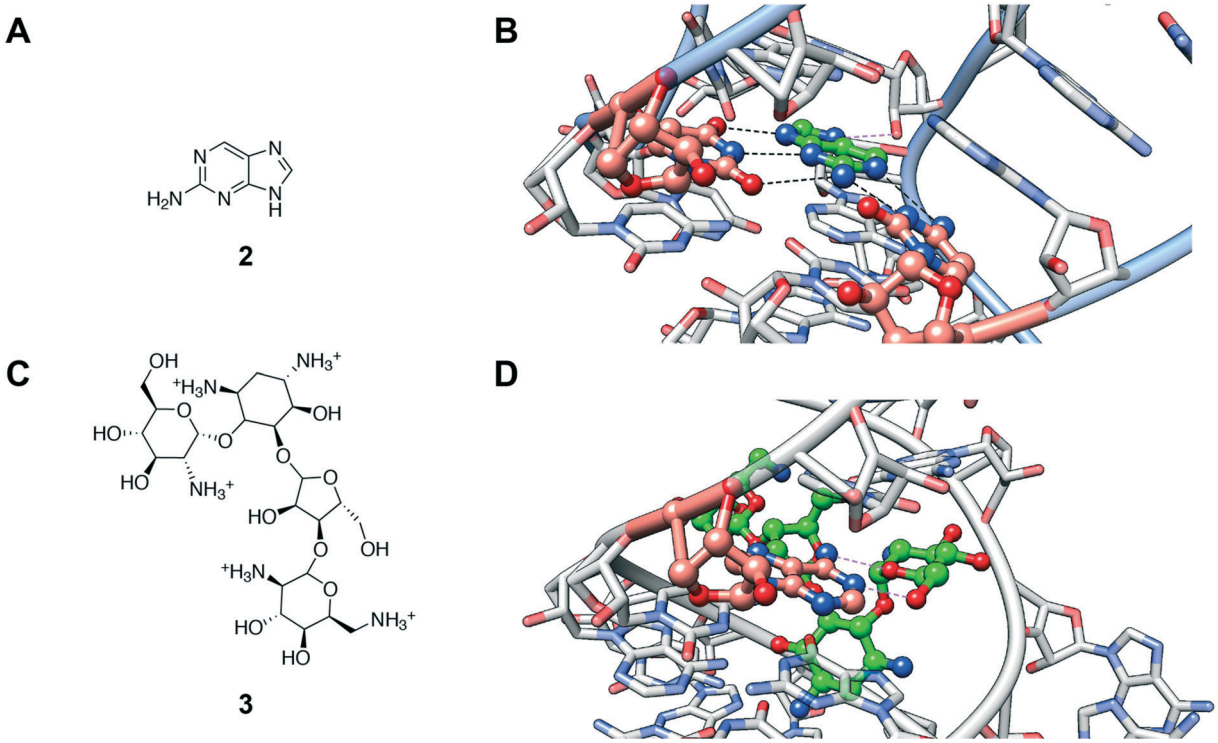

Fig. 3 Hydrogen bonding patterns used by small molecules to interact with nucleobases (nucleobases interacting with the small molecules are highlighted in salmon). (A) Structure of 2-aminopurine (**2**). (B) Binding of **2** to the purine riboswitch (PDB ID 3G4M). (C) Structure of paromomycin (**3**). (D) Binding of **3** to the HIV-1 DIS RNA (PDB ID 3C44).

for the lack of desolvation,[44] and solvent exposed hydrogen bonds have less contribution to the overall energetic profile.[45,46] Interactions between small molecules and water are often observed whenever the geometry between ligands and RNA receptor are not optimal and the presence of a water bridge might compensate for this geometric penalty. For example, in the binding of paromomycin to the A-site of rRNA, the contact between atom O6 of the 2-deoxystreptamine core (2-DOS) and atom O4 of U1406 is mediated by a water molecule (Fig. S2A and B†).[47] Interestingly, synthetic derivatization of paromomycin with an aromatic ring linked through an amino aliphatic chain to the sugar core (C2″) induced an enhanced binding geometry allowing for a direct contact between O6 of 2-DOS and O4 of U1406 while inducing optimal pairing between U1406 and U1495 (Fig. S2C and D†).[48] The combination of improved pharmacokinetics properties induced by the chemical modification and this optimized binding mode resulted in increased potency towards a wider variety of Gram positive and negative bacteria, demonstrating the importance of optimized hydrogen bonding directionality in RNA recognition.

This analysis suggests that hydrogen bonding plays a major role in RNA recognition and that optimization of these contacts can lead to small molecules with enhanced affinity and selectivity. However, optimizing the directionality of hydrogen bonds is often difficult especially for highly flexible molecules. On the other hand, small molecules containing aromatic chemotypes with hydrogen bond donors/acceptors that are complementary to Watson–Crick or Hoogsteen faces of nucleobases as well as computational methods to identify and exploit these interactions are rapidly emerging.[14,49–51]

4.3 Interactions driven by hydrophobic effects

In contrast to protein–ligand recognition, hydrophobicity-driven contacts are only the third most prevalent type of interaction in RNA. The majority are contacts between aliphatic and aromatic carbons. Of the 395 total interactions measured, only 56 (14% relative to all hydrophobicity-driven interactions) engage with sugar atoms, with the large majority exploiting the surface of the bases (Table S5†). G is by far the nucleotide with the most contacts. Interestingly, the guanine nucleobase is reported as the most hydrophilic of the natural bases.[52,53] This highlights that predicting hydrophobicity effects in RNA recognition might be difficult without high-resolution structures and considering that hydrophobicity driven contacts appear to contribute less to the net free energy of binding relative to other type of interactions (*i.e.*, hydrogen bonding, stacking, water displacement *etc.*). This is in contrast to protein recognition, where efficient ligands are often obtained by increasing hydrophobic contacts rather than hydrogen bonding and ionic interactions. It is also easier to optimize hydrophobic contacts that rely less on the bond directionality compared to hydrogen bonding.[19] The fact that enhanced recognition events driven by the introduction of small aliphatic motifs such as the magic methyl are rare in nucleic acid recognition and predominately influence binding selectivity over affinity further supports this

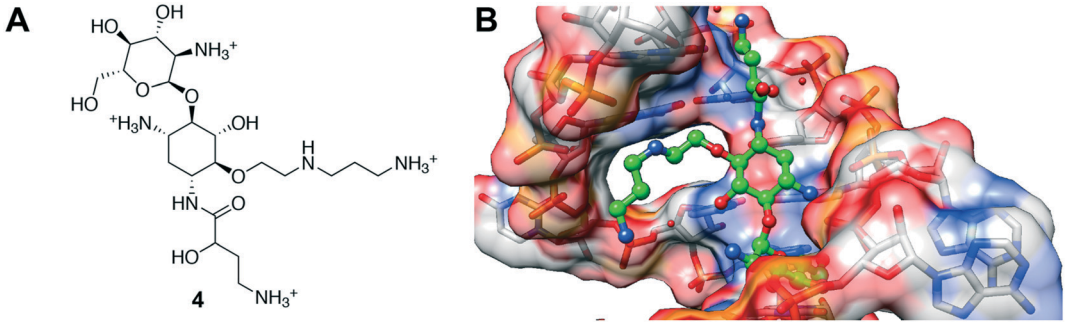

Fig. 4 (A) Structure of the neamine analog 4. (B) Depiction of the contact between the methylene groups of 4 and the ribosomal A site (PDB ID 2F4, surface is coloured according to atom identity. Grey = carbon, orange = phosphorus, blue = nitrogen, red = oxygen).

notion.[54–56] Interestingly, methyl groups are instead widely used by Nature to modulate nucleic shape and indeed interactions with proteins.[57] This might suggest that extended hydrophobic contacts are *de facto* disfavoured in RNA recognition by small molecules, where hydrophobic and structured binding pockets are rare and hydrophobic moieties are more likely to produce a substantial structural rearrangement through steric clashes rather than fit in favourable cavities.

Despite this consideration, there are examples of the use of hydrophobicity-driven contacts that moderately enhanced binding affinity when coupled with other interactions, such as salt bridges. For example, it was demonstrated through molecular dynamics and binding free energy calculations that the enhanced affinity of a synthetically modified aminoglycoside analogue of neamine (**4**) incorporating aliphatic and positively charged anchors was mainly due to the van der Waals and hydrophobic component of the solvation free energy deriving from the methylene chain pointing down or along the major groove (Fig. 4).[58]

Hydrophobicity-driven interactions mediated by sulphur are currently underexplored in RNA–small molecule recognition as demonstrated by only three ligands of our dataset containing this element, which include thiamine pyrophosphate derivatives (PDB 3D2X and 2GDI) and biotin (PDB 1F27). Interactions between an aliphatic side chain containing sulphur and aromatic rings provide higher stabilization compared to carbon–carbon interactions and are widely exploited in protein recognition.[59] In nucleic acid recognition, sulphur has mostly been explored when incorporated into aromatic rings as it increases the hydrophobicity of the ring[60] and might play an important role in promoting preorganized conformations through sigma-hole intramolecular interactions[61] as well as potentially accept hydrogen bonding.[62,63] Because of these inherent properties, exploring sulphur chemical space might be a fruitful strategy for exploring hydrophobicity-driven contacts as well as hydrogen bonding in RNA recognition.

4.4 Weak-hydrogen bond

Weak hydrogen bonding formed between CH and oxygen atoms was the fourth most frequent interaction. The importance of these interactions is well documented both in protein and nucleic acid structure and recognition.[64,65] The median length of this bond is 3.4 Å in agreement with protein recognition. The median angle calculated between direct contact between RNA–small molecule small donor (C), the hydrogen (H) and the acceptor (O) is slightly closer to optimality (180°) for RNA compared to proteins (147° *vs.* 130°).

The large majority of these contacts involve aliphatic carbons of small molecules interacting with oxygen atoms of the phosphate, sugar or base of RNA. This is in contrast with protein ligands that predominantly form weak hydrogen bonds between aromatic hydrogens atoms of the ligands and oxygen atoms on the proteins. In our dataset, weak hydrogen bonding mostly involves aminoglycosides as indicated by the increased percentage (12.6% *vs.* 2.4%) of weak hydrogen bonding in dataset A compared to dataset R. The negatively charged oxygen atom of the phosphate group is one of the most frequent acceptors (21) (Table S6†). The carbonyl oxygens of G were the most frequent acceptor (28 counts) confirming the notion of CH–O hydrogen bonding as secondary interactions that are often accompanied by stronger NH–O bonds.[64]

Interestingly, potential CH–O interactions involving the oxygen of the 2′OH involved almost exclusively aromatic carbons. Furthermore, as observed for strong hydrogen bonding, ~30% of CH–O interactions involve water.

One interesting example of weak hydrogen bonding emerging from our analysis includes the interaction between the CH of the methyl group of the thiazole ring of oxythiamine (**5**) and the O6 of G and the phosphate oxygen of a C residue of the thiamine pyrophosphate riboswitch (PDB 3D2X, Fig. 5). Similarly, the methyl substituent CH in pyrithiamine (PDB 3D2V) also favourably interact with the oxygen of a bridging water molecule located in the riboswitch binding pocket through a CH–O interaction. While the impact of this methyl substituent is not yet clear, it is a frequently recurring motif in molecular scaffolds and analogues binding to the TPP riboswitch. This suggests that these weaker interactions might play a relatively important role in riboswitch recognition and folding.[66,67]

While understanding the actual contributions to the binding of these weaker interactions is not straightforward,

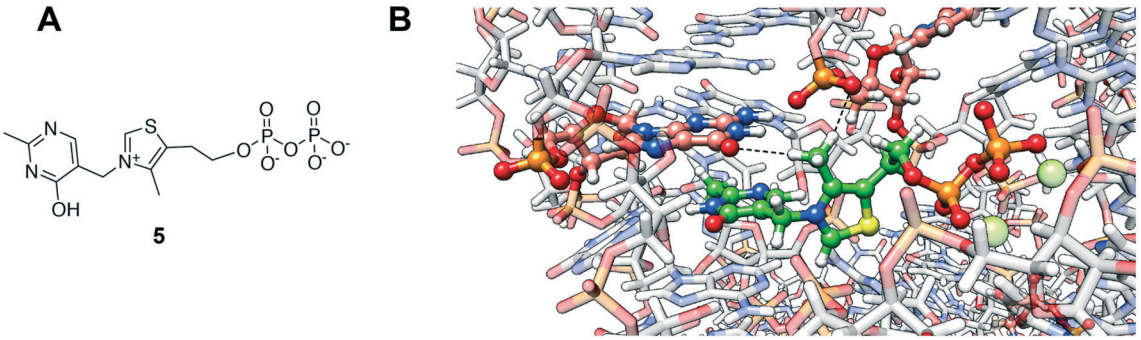

Fig. 5 (A) Structure of oxythiamine pyrophosphate (**5**). (B) Depiction of the weak hydrogen bonding interaction of the methyl group of **5** to 06G and phosphate oxygen of specific nucleotides (in salmon) of the TPP riboswitch (PDB ID 3D2X).

this analysis suggests that weak hydrogen bonding either through aliphatic or aromatic carbon can be exploited for optimizing small molecule–RNA recognition.

4.5 Interactions involving ionic species

4.5.1 Cation–π. Cation–π contacts are essentially electrostatic interactions as they form between a positively charged atom, normally an ammonium group, and the negative electronic cloud of an aromatic system. These are the fifth most represented type of interaction in RNA recognition (3.0%). Because of the RNA structure, the positive charge is exclusively found on the ligand. This is the case of aminoglycosides that are enriched with charged ammonium groups at physiological pH as demonstrated by the enhanced percentage of cation–π in dataset A (7.4%) compared to dataset R (0.4%). On the other hand, in protein recognition, the positive charge is predominately found on amino acid side chains to interact with an aromatic ring of the ligand.[19] G is by far the most frequent nucleotide (39 counts) for this type of interaction (Table S7†) and this is consistent with G enhanced polarizability and previous observations of cation–π between proteins and nucleic acids.[68]

A prominent example of cation–π in RNA recognition is the binding of guanidinium (**6**) to the guanidine II riboswitch (Fig. 6).[35] The guanidinium cation is stacked upon the nucleobase G6, making a cation–π interaction, and it is almost completely enclosed in a box delimited by C8 and A7 (not shown). This example highlights the importance of the guanidinium group in RNA recognition due to the ability to form a variety of interactions such as cation–π, salt bridges, hydrogen bonding and stacking. For this reason, the guanidinium is also largely exploited by proteins for nucleic acid recognition and several nucleic acid small molecule classes have demonstrated enhanced activity when containing this functionalization.[28,69]

4.5.2 Salt bridges. Ionic interactions such as salt bridges between positive charges of small molecules and the negative charges of the phosphate groups in the backbone are considered crucial for small molecule affinity towards RNA. Despite this, they only represent 2.8% of interactions in our dataset (Fig. 1 and Table S8†). Predicting the importance of the salt bridges to optimize the binding of small molecules can be difficult. In fact, these interactions are directional, and the energetic gain might vary according to the position of these contacts. For example, ionic interactions that occur in more buried binding pockets contribute to binding affinity more significantly compared to solvent exposed contacts because of the lower penalty paid by desolvation of the charge.[19,70] In general, compounds that contain multiple positive charges are expected to bind more promiscuously. However, the difficulty in predicting the importance of ionic interactions is demonstrated by the unexpected structural behaviors observed when reducing the number of charges from five to four in neamine and geneticin aminoglycosides, which resulted in nonspecific and sub stoichiometric

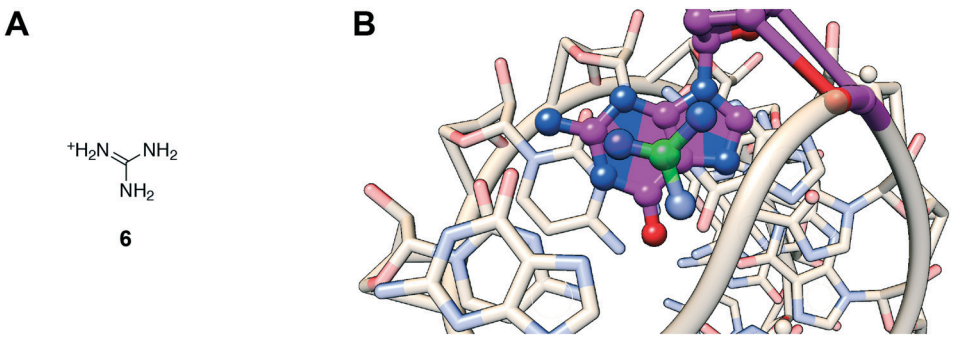

Fig. 6 (A) Structure of guanidine (**6**). (B) Depiction of the stacking of **6** to a G residue (in purple) of the guanidine II riboswitch (PDB ID 5NEF).

binding.[41] This might be correlated to the architecture of aminoglycosides where neighbouring moieties can influence the equilibrium of protonation states at physiological conditions.[71] This analysis underscores that as expected positive charges are fundamental for RNA small molecules binders (23 out of the 37 ligands of this analysis have at least one positive charge). However, the low counts of ionic interactions in our dataset suggest that highly charged ligands are not the only avenue for RNA recognition, which is promising for achieving binding selectivity for therapeutically relevant RNA constructs and optimal medicinal chemistry properties, including cellular permeation, for RNA ligands.

4.5.3 Interactions involving inorganic positive ions. Positive ions such as Mg^{2+} and K^{+} are fundamental for regulating RNA tertiary folding through ionic interaction and the water molecules in their solvation shell.[72] Our analysis suggests that ions might also mediate the binding of small molecules through the same forces. In particular, we found 23 direct interactions between small molecules and Mg^{2+} and K^{+} ions (Table S9†).

For example, metal coordination is largely exploited by the thiamine pyrophosphate (TPP) riboswitch in which the two phosphate groups of the pyrophosphate are bound by a pair of hexa-coordinated Mg^{2+} ions with octahedral ligation geometry that coordinate the ligands to RNA through water-mediated hydrogen bonds with RNA (Fig. S1†).[39] Through these interactions, the energetic cost associated with the charge repulsion between the negative charges of the ligand and the RNA are neutralized. This analysis corroborates the notion that ionic strengths and buffer components not only affect RNA conformation but are also important mediators between RNA and small molecules that should be considered when choosing the conditions for HTS.

4.6 Interactions involving halogens

Interactions involving halogen atoms such as halogen bonding and multipolar interactions are usually less frequent in small molecule–macromolecule interactions, nonetheless they can have an important role in determining the binding affinity and modulating the pharmacokinetic properties of the compounds.[22] Halogen bonds are interactions defined between the positive electrostatic potential of a halogen (sigma hole) and electronegative acceptor such as oxygen, sulphur and nitrogen atoms. Because of the lack of polarizability of the C–F bond, this interaction can take place only with heavier halogen such as chlorine, bromine and iodine. On the other hand, multipolar interactions are contacts that involve halogen (more frequently fluorine) with a weakly electrophilic group such as the amide carbon of the protein backbone.

In our dataset, only two molecules contained fluorine or chlorine atoms and none of these interactions were observed. Because of this limited amount of data, it is not possible to predict the importance of these interactions in RNA recognition. However, a wider analysis determined to understand the effect of halogen-mediated interactions in nucleic acid structures deposited in the PDB also revealed the lack of small molecule ligands utilizing these interactions for targeting nucleic acids.[73] This analysis also underscored that for the few interactions measured, non-optimal geometries were found suggesting that the weak energetic gain derived from these interactions is usually overcome by more efficient interactions such as hydrogen bonding and stacking.

5. Conclusion

We are currently at the early stages of developing therapeutics targeting RNA. Chemo-informatic analyses that compare RNA-targeting ligands and FDA-approved small molecules support the notion that there are privileged structural characteristics for RNA recognition. Our systematic analysis provides direct evidence of this by highlighting the most frequent interactions used by small molecules for RNA recognition based on currently available complexes structurally determined by X-ray crystallography in the PDB. The comparison with protein recognition interactions highlights a distinct binding signature for RNA recognition, with hydrogen bonding and stacking being more prominent than hydrophobicity-driven interactions. Unfortunately, the lack of complete activity, thermodynamic and kinetic data currently available for these ligands precludes further extrapolation of the molecular recognition features for achieving high efficiency molecules. While the field awaits additional studies containing full characterization, the information gathered from this survey can be used qualitatively for optimizing RNA–ligand interactions both in scaffold based-drug design and library curation for HTS. For example, these analyses suggest that optimization of interactions driven by hydrophobic effects might be more difficult compared to proteins and focusing on hydrogen bonding and stacking optimization is potentially more fruitful. Additionally, our analysis indicates that nucleobases rather than backbone are most prominently interacting with small molecules, and stacking interaction optimization might be more effective for targeting purine-rich RNA motifs, while cation–π interaction can be useful for targeting G residues. Furthermore, the ranking of interactions provided might also serve for improving docking scoring functions that can be used for developing software packages that specifically address RNA recognition and result in more accurate and successful virtual screening campaigns. Altogether, we foresee that this information will aid in establishing the principles driving RNA recognition by small molecules.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

We would like to thank all members of the Hargrove lab for providing valuable feedback on this manuscript. We thank Dr. Gary J. Kapral for discussions and support on examining the

Protein Data Bank for suitable RNA:SM structures. The authors acknowledge financial support for this work from Duke University (AEH, GP, NNP), the US National Institute of General Medical Sciences (NIGMS) Maximizing Investigator's Research Award (MIRA) Grant/Award Number (R35GM124785) (AEH, GP), the US National Institute of Allergy and Infectious Diseases (U54 AI150470) (AEH, NNP), and the National Science Foundation (CAREER 1750375) (AEH). MS gratefully acknowledges support from the Natural Sciences and Engineering Research Council of Canada (Grant RGPIN-2019-04416).

References

1 S. A. Mortimer, M. A. Kidwell and J. A. Doudna, Insights into RNA Structure and Function from Genome-Wide Studies, *Nat. Rev. Genet.*, 2014, **15**, 469.

2 F. P. Marchese, I. Raimondi and M. Huarte, The Multidimensional Mechanisms of Long Noncoding RNA Function, *Genome Biol.*, 2017, **18**(1), 206.

3 A. E. Dahlberg, The Functional Role of Ribosomal RNA in Protein Synthesis, *Cell*, 1989, **57**(4), 525–529.

4 K. D. Warner, C. E. Hajdin and K. M. Weeks, Principles for Targeting RNA with Drug-like Small Molecules, *Nat. Rev. Drug Discovery*, 2018, **17**, 547–558.

5 A. Donlic and A. E. Hargrove, Targeting RNA in Mammalian Systems with Small Molecules, *Wiley Interdiscip. Rev.: RNA*, 2018, **9**(4), e1477.

6 R. Santos, O. Ursu, A. Gaulton, A. P. Bento, R. S. Donadi, C. G. Bologa, A. Karlsson, B. Al-Lazikani, A. Hersey and T. I. Oprea, *et al.* A Comprehensive Map of Molecular Drug Targets, *Nat. Rev. Drug Discovery*, 2016, **16**, 19.

7 T. Tran and M. D. Disney, Identifying the Preferred RNA Motifs and Chemotypes That Interact by Probing Millions of Combinations, *Nat. Commun.*, 2012, **3**, 1125.

8 S. G. Rzuczek, M. R. Southern and M. D. Disney, Studying a Drug-like, RNA-Focused Small Molecule Library Identifies Compounds That Inhibit RNA Toxicity in Myotonic Dystrophy, *ACS Chem. Biol.*, 2015, **10**(12), 2706–2715.

9 J. H. Cate and J. A. Doudna, Solving large RNA structures by X-ray crystallography, *Methods Enzymol.*, 2000, **317**, 169–180.

10 R. P. Barnwal, F. Yang and G. Varani, Applications of NMR to Structure Determination of RNAs Large and Small, *Arch. Biochem. Biophys.*, 2017, **628**, 42–56.

11 A. Umuhire Juru, N. N. Patwardhan and A. E. Hargrove, Understanding the Contributions of Conformational Changes, Thermodynamics, and Kinetics of RNA–Small Molecule Interactions, *ACS Chem. Biol.*, 2019, **14**(5), 824–838.

12 W. M. Hewitt, D. R. Calabrese and J. S. Schneekloth, Evidence for Ligandable Sites in Structured RNA throughout the Protein Data Bank, *Bioorg. Med. Chem.*, 2019, **27**(11), 2253–2260.

13 B. S. Morgan, J. E. Forte, R. N. Culver, Y. Zhang and A. E. Hargrove, Discovery of Key Physicochemical, Structural, and Spatial Properties of RNA-Targeted Bioactive Ligands, *Angew. Chem.*, 2017, **129**(43), 13683–13687.

14 H. S. Haniff, A. Graves and M. D. Disney, Selective Small Molecule Recognition of RNA Base Pairs, *ACS Comb. Sci.*, 2018, **20**(8), 482–491.

15 J. L. Childs-Disney and M. D. Disney, Approaches to Validate and Manipulate RNA Targets with Small Molecules in Cells, *Annu. Rev. Pharmacol. Toxicol.*, 2016, **56**(1), 123–140.

16 N. F. Rizvi, J. P. Santa Maria Jr and A. Nahvi, *et al.* Targeting RNA with Small Molecules: Identification of Selective, RNA-Binding Small Molecules Occupying Drug-Like Chemical Space, *SLAS Discovery*, 2020, **25**(4), 384–396.

17 H. S. Haniff, L. Knerr, X. Liu, G. Crynen, J. Bostrom, D. Abegg, A. Adibekian, M. Lemurell and M. D. Disney, Design of a Small Molecule That Stimulates VEGFA Informed from an Expanded Encyclopedia of RNA Fold-Small Molecule Interactions, Hafeez S. Haniff, *ChemRxiv*, 2019, Pre-Print DOI: 10.26434/chemrxiv.7967291.v1.

18 B. S. Morgan, J. E. Forte and A. E. Hargrove, Insights into the Development of Chemical Probes for RNA, *Nucleic Acids Res.*, 2018, **46**(16), 8025–8037.

19 R. Ferreira de Freitas and M. Schapira, A Systematic Analysis of Atomic Protein–Ligand Interactions in the PDB, *MedChemComm*, 2017, **8**(10), 1970–1981.

20 R. Abagyan, M. Totrov and D. Kuznetsov, ICM—A New Method for Protein Modeling and Design: Applications to Docking and Structure Prediction from the Distorted Native Conformation, *J. Comput. Chem.*, 1994, **15**(5), 488–506.

21 C. S. Leung, S. S. F. Leung, J. Tirado-Rives and W. L. Jorgensen, Methyl Effects on Protein–Ligand Binding, *J. Med. Chem.*, 2012, **55**(9), 4489–4500.

22 C. Bissantz, B. Kuhn and M. Stahl, A Medicinal Chemist's Guide to Molecular Interactions, *J. Med. Chem.*, 2010, **53**(14), 5061–5084.

23 M. Chittapragada, S. Roberts and Y. W. Ham, Aminoglycosides: Molecular Insights on the Recognition of RNA and Aminoglycoside Mimics, *Perspect. Med. Chem.*, 2009, **3**, S2381.

24 C. S. Eubanks, J. E. Forte, G. J. Kapral and A. E. Hargrove, Small Molecule-Based Pattern Recognition To Classify RNA Structure, *J. Am. Chem. Soc.*, 2017, **139**(1), 409–416.

25 A. Di Giorgio and M. Duca, Synthetic Small-Molecule RNA Ligands: Future Prospects as Therapeutic Agents, *MedChemComm*, 2019, **10**(8), 1242–1255.

26 W.-Y. Yang, F. He, R. L. Strack, S. Y. Oh, M. Frazer, S. R. Jaffrey, P. K. Todd and M. D. Disney, Small Molecule Recognition and Tools to Study Modulation of r(CGG)Exp in Fragile X-Associated Tremor Ataxia Syndrome, *ACS Chem. Biol.*, 2016, **11**(9), 2456–2465.

27 N. M. Bell, A. L'Hernault, P. Murat, J. E. Richards, A. M. L. Lever and S. Balasubramanian, Targeting RNA–Protein Interactions within the Human Immunodeficiency Virus Type 1 Lifecycle, *Biochemistry*, 2013, **52**(51), 9269–9274.

28 E. A. Jefferson, P. P. Seth, D. E. Robinson, D. K. Winter, A. Miyaji, S. A. Osgood, E. E. Swayze and L. M. Risen, Biaryl Guanidine Inhibitors of in Vitro HCV-IRES Activity, *Bioorg. Med. Chem. Lett.*, 2004, **14**(20), 5139–5143.

29 E. Khan, S. Biswas, S. K. Mishra, R. Mishra, S. Samanta, A. Mishra, A. Tawani and A. Kumar, Rationally Designed Small Molecules Targeting Toxic CAG Repeat RNA That Causes Huntington's Disease (HD) and Spinocerebellar Ataxia (SCAs), *Biochimie*, 2019, **163**, 21–32.

30 S. Tsuzuki, K. Honda, T. Uchimaru, M. Mikami and K. Tanabe, Origin of Attraction and Directionality of the π/π Interaction: Model Chemistry Calculations of Benzene Dimer Interaction, *J. Am. Chem. Soc.*, 2002, **124**(1), 104–112.

31 C. R. Martinez and B. L. Iverson, Rethinking the Term "Pi-Stacking", *Chem. Sci.*, 2012, **3**(7), 2191–2201.

32 B. Milovanović, M. Kojić, M. Petković and M. Etinski, New Insight into Uracil Stacking in Water from Ab Initio Molecular Dynamics, *J. Chem. Theory Comput.*, 2018, **14**(5), 2621–2632.

33 K. M. Guckian, B. A. Schweitzer, R. X.-F. Ren, C. J. Sheils, D. C. Tahmassebi and E. T. Kool, Factors Contributing to Aromatic Stacking in Water: Evaluation in the Context of DNA, *J. Am. Chem. Soc.*, 2000, **122**(10), 2213–2222.

34 G. W. Collie, S. Sparapani, G. N. Parkinson and S. Neidle, Structural Basis of Telomeric RNA Quadruplex–Acridine Ligand Recognition, *J. Am. Chem. Soc.*, 2011, **133**(8), 2721–2728.

35 L. Huang, J. Wang and D. M. J. Lilley, The Structure of the Guanidine-II Riboswitch, *Cell Chem. Biol.*, 2017, **24**(6), 695–702.e2.

36 N. W. Luedtke, Q. Liu and Y. Tor, RNA–Ligand Interactions: Affinity and Specificity of Aminoglycoside Dimers and Acridine Conjugates to the HIV-1 Rev Response Element, *Biochemistry*, 2003, **42**(39), 11391–11403.

37 T. J. Ritchie, S. J. F. Macdonald, S. Peace, S. D. Pickett and C. N. Luscombe, Increasing Small Molecule Drug Developability in Sub-Optimal Chemical Space, *MedChemComm*, 2013, **4**(4), 673–680.

38 V. S. Anokhina, J. D. McAnany, J. H. Ciesla, T. A. Hilimire, N. Santoso, H. Miao and B. L. Miller, Enhancing the Ligand Efficiency of Anti-HIV Compounds Targeting Frameshift-Stimulating RNA, *Bioorg. Med. Chem.*, 2019, **27**(13), 2972–2977.

39 A. Serganov, A. Polonskaia, A. T. Phan, R. R. Breaker and D. J. Patel, Structural Basis for Gene Regulation by a Thiamine Pyrophosphate-Sensing Riboswitch, *Nature*, 2006, **441**(7097), 1167–1171.

40 S. D. Gilbert, F. E. Reyes, A. L. Edwards and R. T. Batey, Adaptive Ligand Binding by the Purine Riboswitch in the Recognition of Guanine and Adenine Analogs, *Structure*, 2009, **17**(6), 857–868.

41 B. François, R. J. M. Russell, J. B. Murray, F. Aboul-ela, B. Masquida, Q. Vicens and E. Westhof, Crystal Structures of Complexes between Aminoglycosides and Decoding A Site Oligonucleotides: Role of the Number of Rings and Positive Charges in the Specific Binding Leading to Miscoding, *Nucleic Acids Res.*, 2005, **33**(17), 5677–5690.

42 P. P. Seelam, A. Mitra and P. Sharma, Pairing Interactions between Nucleobases and Ligands in Aptamer:Ligand Complexes of Riboswitches: Crystal Structure Analysis, Classification, Optimal Structures, and Accurate Interaction Energies, *RNA*, 2019, **25**(10), 1274–1290.

43 P. Auffinger and E. Westhof, Rules Governing the Orientation of the 2′-Hydroxyl Group in RNA1 1Edited by I. Tinoco, *J. Mol. Biol.*, 1997, **274**(1), 54–63.

44 D. H. Williams and M. S. Westwell, Aspects of Weak Interactions, *Chem. Soc. Rev.*, 1998, **27**(1), 57–64.

45 B. K. Shoichet, No Free Energy Lunch, *Nat. Biotechnol.*, 2007, **25**, 1109.

46 D. Chen, N. Oezguen, P. Urvil, C. Ferguson, S. M. Dann and T. C. Savidge, Regulation of Protein-Ligand Binding Affinity by Hydrogen Bond Pairing, *Sci. Adv.*, 2016, **2**(3), e1501240.

47 Q. Vicens and E. Westhof, Crystal Structure of Paromomycin Docked into the Eubacterial Ribosomal Decoding A Site, *Structure*, 2001, **9**(8), 647–658.

48 B. François, J. Szychowski, S. S. Adhikari, K. Pachamuthu, E. E. Swayze, R. H. Griffey, M. T. Migawa, E. Westhof and S. Hanessian, Antibacterial Aminoglycosides with a Modified Mode of Binding to the Ribosomal-RNA Decoding Site, *Angew. Chem., Int. Ed.*, 2004, **43**(48), 6735–6738.

49 J. F. Arambula, S. R. Ramisetty, A. M. Baranger and S. C. Zimmerman, A Simple Ligand That Selectively Targets CUG Trinucleotide Repeats and Inhibits MBNL Protein Binding, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**(38), 16068–16073.

50 J. Sztuba-Solinska, S. R. Shenoy, P. Gareiss, L. R. H. Krumpe, S. F. J. Le Grice, B. R. O'Keefe and J. S. Schneekloth, Identification of Biologically Active, HIV TAR RNA-Binding Small Molecules Using Small Molecule Microarrays, *J. Am. Chem. Soc.*, 2014, **136**(23), 8402–8410.

51 C. G. Oliver, R. S. Gendron, N. Moitessier, V. Mallet, V. Reinharz and J. Waldispühl Extended RNA Base Pairing Networks Imprint Small Molecule Binding Preferences, *bioRxiv*, 2019, p. 701326.

52 J. L. Miller and P. A. Kollman, Solvation Free Energies of the Nucleic Acid Bases, *J. Phys. Chem.*, 1996, **100**(20), 8587–8594.

53 P. Shih, L. G. Pedersen, P. R. Gibbs and R. Wolfenden, Hydrophobicities of the Nucleic Acid Bases: Distribution Coefficients from Water to Cyclohexane1 1Edited by I. Tinoco, *J. Mol. Biol.*, 1998, **280**(3), 421–430.

54 S. P. Velagapudi, S. J. Seedhouse, J. French and M. D. Disney, Defining the RNA Internal Loops Preferred by Benzimidazole Derivatives via 2D Combinatorial Screening and Computational Analysis, *J. Am. Chem. Soc.*, 2011, **133**(26), 10111–10118.

55 K. Aman, G. Padroni, J. A. Parkinson, T. Welte and G. A. Burley, Structural and Kinetic Profiling of Allosteric Modulation of Duplex DNA Induced by DNA-Binding Polyamide Analogues, *Chem. – Eur. J.*, 2019, **25**(11), 2757–2763.

56 P. Guo, A. A. Farahat, A. Paul, N. K. Harika, D. W. Boykin and W. D. Wilson, Compound Shape Effects in Minor Groove Binding Affinity and Specificity for Mixed Sequence DNA, *J. Am. Chem. Soc.*, 2018, **140**(44), 14761–14769.

57 Y. Chen, T. Hong, S. Wang, J. Mo, T. Tian and X. Zhou, Epigenetic Modification of Nucleic Acids: From Basic Studies to Medical Applications, *Chem. Soc. Rev.*, 2017, **46**(10), 2844–2872.
58 J. B. Murray, S. O. Meroueh, R. J. M. Russell, G. Lentzen, J. Haddad and S. Mobashery, Interactions of Designer Antibiotics and the Bacterial Ribosomal Aminoacyl-TRNA Site, *Chem. Biol.*, 2006, **13**(2), 129–138.
59 B. R. Beno, K.-S. Yeung, M. D. Bartberger, L. D. Pennington and N. A. Meanwell, A Survey of the Role of Noncovalent Sulfur Interactions in Drug Design, *J. Med. Chem.*, 2015, **58**(11), 4383–4438.
60 G. Padroni, J. A. Parkinson, K. R. Fox and G. A. Burley, Structural Basis of DNA Duplex Distortion Induced by Thiazole-Containing Hairpin Polyamides, *Nucleic Acids Res.*, 2018, **46**(1), 42–53.
61 P. Guo, A. Paul, A. Kumar, A. A. Farahat, D. Kumar, S. Wang, D. W. Boykin and W. D. Wilson, The Thiophene "Sigma-Hole" as a Concept for Preorganized, Specific Recognition of G·C Base Pairs in the DNA Minor Groove, *Chem. – Eur. J.*, 2016, **22**(43), 15404–15412.
62 S. Foister, M. A. Marques, R. M. Doss and P. B. Dervan, Shape Selective Recognition of T·A Base Pairs by Hairpin Polyamides Containing N-Terminal 3-Methoxy (and 3-Chloro) Thiophene Residues, *Bioorg. Med. Chem.*, 2003, **11**(20), 4333–4340.
63 M. A. Marques, R. M. M. Doss, A. R. R. Urbach and P. B. B. Dervan, Toward an Understanding of the Chemical Etiology for DNA Minor-Groove Recognition by Polyamides, *Helv. Chim. Acta*, 2002, **85**(12), 4485–4517.
64 S. Horowitz and R. C. Trievel, Carbon-Oxygen Hydrogen Bonding in Biological Structure and Function, *J. Biol. Chem.*, 2012, **287**(50), 41576–41582.
65 Y. Mandel-Gutfreund, H. Margalit, R. L. Jernigan and V. B. Zhurkin, A Role for CH…O Interactions in Protein-DNA Recognition1 1Edited by B. Honig, *J. Mol. Biol.*, 1998, **277**(5), 1129–1140.
66 E. Cressina, L. Chen, C. Abell, F. J. Leeper and A. G. Smith, Fragment Screening against the Thiamine Pyrophosphate Riboswitch ThiM, *Chem. Sci.*, 2011, **2**(1), 157–165.
67 L. Chen, E. Cressina, N. Dixon, K. Erixon, K. Agyei-Owusu, J. Micklefield, A. G. Smith, C. Abell and F. J. Leeper, Probing Riboswitch-Ligand Interactions Using Thiamine Pyrophosphate Analogues, *Org. Biomol. Chem.*, 2012, **10**(30), 5924–5931.
68 M. Rooman, J. Liévin, E. Buisine and R. Wintjens, Cation–π/H-Bond Stair Motifs at Protein–DNA Interfaces, *J. Mol. Biol.*, 2002, **319**(1), 67–76.
69 N. N. Patwardhan, L. R. Ganser, G. J. Kapral, C. S. Eubanks, J. Lee, B. Sathyamoorthy, H. M. Al-Hashimi and A. E. Hargrove, Amiloride as a New RNA-Binding Scaffold with Activity against HIV-1 TAR, *Med. Chem. Commun.*, 2017, **8**(5), 1022–1036.
70 C. D. Waldburger, J. F. Schildbach and R. T. Sauer, Are Buried Salt Bridges Important for Protein Stability and Conformational Specificity?, *Nat. Struct. Biol.*, 1995, **2**(2), 122–128.
71 H. Wang and Y. Tor, Electrostatic Interactions in RNA Aminoglycosides Binding, *J. Am. Chem. Soc.*, 1997, **119**(37), 8734–8735.
72 D. E. Draper, D. Grilley and A. M. Soto, Ions and RNA Folding, *Annu. Rev. Biophys. Biomol. Struct.*, 2005, **34**(1), 221–243.
73 M. H. Kolář and O. Tabarrini, Halogen Bonding in Nucleic Acid Complexes, *J. Med. Chem.*, 2017, **60**(21), 8681–8690.

