Structure-based design for the discovery of novel ligands targeting the FMN riboswitch

Doctoral thesis

submitted in fulfillment of the regulations
for attaining the degree of

"Doktor der Naturwissenschaften (Dr. rer. nat)"
in the subject of Pharmacy

at the Faculty of Chemistry, Pharmacy and Geoscience
at the Johannes Gutenberg University of Mainz

Thomas Wehler

born in Bad Hersfeld

Mainz 2017

Submitted at the Faculty of Chemistry, Pharmacy and Geosciences: **03.05.2017**

Dean: [redacted]

Name of 1st reviewer: [redacted]
Name of 2nd reviewer: [redacted]

Date of doctoral examination: **09.06.2017**

D77 (Dissertation Universität Mainz)

Declaration of Authorship

I, Thomas WEHLER, declare that this thesis titled, "Structure-based design for the discovery of novel ligands targeting the FMN riboswitch" and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed: [redacted]

Date: [redacted]

"Science is the key to our future, and if you don't believe in science, then you're holding everybody back."

Bill Nye

Johannes Gutenberg University Mainz

Abstract

Faculty of Chemistry, Pharmaceutical Sciences and Geoscience
Institute of Pharmacy and Biochemistry

Structure-based design for the discovery of novel ligands targeting the FMN riboswitch

by Thomas WEHLER

The FMN riboswitch is found in the 5'-UTR of bacterial mRNA and regulates gene expression of proteins involved in transport and biosynthesis of riboflavin via feedback control. This RNA structure is widespread among pathogenic bacteria and the ligand binding site is highly conserved among different species. Therefore, the FMN riboswitch represents a suitable drug target for novel antibacterial compounds.
For the identification of ligands with novel chemical scaffolds distinct from the chemical constitution of the cognate ligand FMN and its close analogues with known binding properties two successive virtual screening approaches were performed. Subsequently, an unbiased virtual screening with an in-house database of commercially available compounds and a privileged substructure search to facilitate mimicking the most important interactions between the target RNA and its cognate ligand were carried out.
The virtual screening hits were tested experimentally to confirm the *in silico* predictions. Three different methods were applied in order to determine the binding affinities of hit compounds. Whereas a method based on fluorescence quenching and ITC measurements indirectly determined the binding constants of the possible ligands, a SPR methods was developed to directly sense the interactions between the FMN riboswitch RNA and possible small molecule ligands. One compound identified by the virtual screening methods was confirmed to bind experimentally. Close analogues of the hit substance are currently under investigation to initialize a small SAR study.

Johannes Gutenberg Universität Mainz

Zusammenfassung

Fachbereich 09 - Chemie, Pharmazie und Geowissenschaften
Institut für Pharmazie und Biochemie

Strukturbasiertes Design zur Identifizierung neuartiger Liganden für den FMN Riboswitch

von Thomas WEHLER

Der FMN Riboswitch ist eine RNA-Struktur, welche sich in der 5'-UTR bakterieller mRNA befindet und die Genexpression verschiedener Proteine, die am Transport und der Biosynthese von Riboflavin beteiligt sind, im Sinne eines Feedback-Mechanismus reguliert. Diese spezialisierte RNA-Struktur findet sich in vielen pathogenen Bakterien wieder und die Ligandenbindestelle ist hochkonserviert in den verschiedenen bakteriellen Spezies. Daher eignet sich der FMN Riboswitch als Zielstruktur für neue antibakterielle Substanzen.
Für die Identifizierung von Liganden mit Strukturelementen, die sich von denen unterscheiden, die der natürliche Ligand FMN und eng verwandte Flavin-Analoga innehaben, wurden zwei aufeinanderfolgende virtuelle Screening-Methoden durchgeführt. Zunächst wurde die Eignung des Docking-Setups mit Hilfe eines Trainingssets von Verbindungen mit flavinähnlichen Strukturen durch Vergleich der vorhergesagten Bindemodi mit den kristallographisch ermittelten Strukturen validiert. Anschließend wurde einmal ein molekulares Docking mit der gesamten virtuellen Datenbank aus käuflich erwerbbaren Molekülen durchgeführt, gefolgt von einer Begrenzung der Suche auf Substanzen mit privilegierten Substrukturen, die Interaktionen zwischen FMN und RNA einfacher imitieren können.
Die Treffer des virtuellen Screenings wurden experimentell getestet, um die *in silico* vorhergesagte Bindungsaffinität zu bestätigen. Dabei wurden drei verschiedene Methoden angewendet, um die Bindungsaffinitäten dieser Moleküle zu bestimmen. Wohingegegen mit einer auf Fluoreszenzlöschung basierende Methode und ITC-Messungen die Bindungskonstanten der potentiellen Liganden auf eine indirekte Art und Weise bestimmt wurden, wurde eine SPR-basierte Methode entwickelt, um die Interaktionen zwischen dem FMN Riboswitch und den potentiellen Liganden direkt zu bestimmen. Für eine Verbindung, die aus dem virtuellen Screening kam, wurde eine Bindung zum FMN Riboswitch experimentell bestätigt. Aktuell werden strukturell eng verwandte Analoga dieser Substanz untersucht, um Struktur-Wirkungsbeziehungen herzuleiten.

Acknowledgements

First of all I would like to thank [redacted] for the opportunity to work in her group on such an interesting and challenging project. Her scientific expertise and guidance has been a great help, especially to draw the right conclusions in order to further advance the project. Of course I am grateful that she also read this manuscript carefully and her suggestions helped improving the firts draft version to this final document.
Also I would like to thank [redacted], who adopted me and allowed me to finish my work at the University of Mainz in her group, after [redacted] had moved to her new position at the University of Bergen, Norway. Furthermore, I am very grateful that the research groups of [redacted] and [redacted] for their generous gifts of plasmids containing the sequence information for the *F. nucleatum* FMN riboswitch aptamer domain and the T7 RNA polymerase respectively.
Special thanks also deserve my colleagues of the brenkgroup, [redacted] and [redacted]. We always had a nice time together and also had fruitful conversations regarding our research to mutually provide assistance with problem solving. During my last stay in Bergen I became acquainted with [redacted], who will follow-up with the work on my research topic. I also really appreciated his scientific mindset and his suggestions. Next, I would like to thank the lab technicians [redacted] in Mainz and [redacted] in Bergen for their active support in the biochemistry lab.
I also would like to thank the members (former and current) of the research group of [redacted] for tolerating me as an outsider sitting in "their" PC room most of the time and of course for the nice time we spent together. In particular I have to thank [redacted] for being a colleague and a good friend, who was always supporting me and had an open ear for me. She always had confidence in my scientific advices, too, and played an essential role in creating a pleasant working atmosphere.
Furthermore I would like to thank my family for their support. Without their help I probably would have never reached my goals. Special thanks goes to my girlfriend, [redacted], for her support and love during the past 4 years. She had to be very patient sometimes, in particular during the time I was absent from Germany and I really appreciated that she could always make for some distraction whenever needed.

Contents

List of Abbreviations

AdoCbl	**A**d**e**n**o**syl**c**o**b**a**l**amin
ADME	**A**bsorption, **d**istribution, **m**etabolism, and **e**xcretion
A. thaliana	*Arabidopsis thaliana*
ATP	**A**denosine **t**ri**p**hosphate
BSA	**B**ovine **s**erum **a**lbumin
B. subtilis	*Bacillus subtilis*
c-di-AMP	**c**yclic **di**-**a**denosyl-5′-**m**ono**p**hosphate
c-di-GMP	**c**yclic **di**-**g**uanosyl-5′-**m**ono**p**hosphate
Cpd.	**C**om**p**oun**d**
DEAE	**D**i**e**thyl**a**mino**e**thanol
DESY	**D**eutsches **E**lektronen-**Sy**nchroton
2′-dG	2′-**d**eoxy**g**uanosine
DMSO	**D**i**m**ethyl **s**ulf**o**xide
DNA	**D**eoxyribo**n**ucleic **a**cid
DTT	**D**i**t**hio**t**hreitol
E. coli	*Escherichia coli*
EDC	1-**e**thyl-3-(3-**d**imethylaminopropyl)**c**arbodiimide
EDTA	**E**thylene**d**iamine**t**etraacetic **a**cid
ESKAPE	***E**nterococcus faecium*, ***S**taphylococcus aureus*, ***K**lebsiella pneumoniae*, ***A**cinetobacter baumanii*, ***P**seudomonas aeruginosa*, ***E**nterobacter species*
FBDD	**f**ragment-**b**ased **d**rug **d**iscovery
5FDQD	5-(3-(4-**f**luorophenyl)butyl)-7,8-**d**imethylpyrido[3,4-b]**q**uinoxaline-1,3(2H,5H)-**d**ione
FMN	**F**lavin **m**ono**n**ucleotide
FMN-RS	**FMN r**ibo**s**witch
FPLC	**F**ast **p**rotein **l**iquid **c**hromatography
***glmS* ribozyme**	**G**lucosa**m**ine-6-pho**s**phate activated ribozyme
GlcN6P	**G**lu**c**osami**n**e-**6**-**p**hosphate
HA	**H**eavy **a**tom count
HCV	**H**epatitis **C** **V**irus
HDV	**H**epatitis **d**elta **v**irus
HEPES	4-(2-**h**ydroxy**e**thyl)-1-**p**iperazinyl**e**thane**s**ulfonic acid
***Hind*III**	**H**aemophilus **in**fluenzae **d**eoxyribonuclease**III**
HIV	**H**uman **I**mmunodeficiency **V**irus
HIV-1 TAR	**H**uman **I**mmunodeficiency **V**irus-1 **T**rans**a**ctivation **r**esponse element

hTR-P2b	**h**uman **T**elomerase **R**NA-P2b
HTS	**H**igh-**t**hroughput **s**creening
ICM	**I**nternal **c**oordinate **m**odelling
IPTG	**I**so**p**ropyl β-D-1-**t**hio**g**alactopyranoside
IRES	**I**nternal **r**ibosome **e**ntry **s**ite
ITC	**I**sothermal **t**itration **c**alorimetry
K_D	**d**issociation constant
kDa	**k**ilo **Da**lton
K_i	**i**nhibitory constant
LB	**L**ysogeny **b**roth
MD	**M**olecular **d**ynamics
MES	2-(*N*-**m**orpholino)**e**thane**s**ulfonic acid
MIC	**M**inimal **I**nhibitory **C**oncentration
MoCo	**Mo**lybdenum **co**factors
MOE	**M**olecular **O**perational **E**nvironment
MORDOR	**MO**lecular **R**ecognition with a **D**riven dynamics **O**ptimize**R**
mRNA	**m**essenger RNA
MRSA	**M**ethicillin-**r**esistant *Staphylococcus aureus*
NMR	**N**uclear **m**agnetic **r**esonance spectroscopy
NOESY	**N**uclear **O**verhauser **E**ffect **S**pectroscop**Y**
nt	**n**ucleo**t**ide
NTP	**N**ucleoside **t**ri**p**hosphate
$\mathbf{OD}_{600}$	**O**ptical **d**ensity at 600 nm wavelength
PAGE	**P**oly**a**crylamide **g**el **e**lectrophoresis
PEG	**P**oly**e**thyleng**l**ycol
pK_a-value	negative decadic logarithm of the **a** disscociation constant
preQ$_1$	**Pre**-**q**ueuosine$_1$
PTPP	**P**yri**t**hiamine **p**yro**p**hosphate
QM	**Q**uantum **m**echanics
ribD	gene encoding for a **rib**oflavin-specific **d**eaminase
rmsd	**r**oot-**m**ean-**s**quare **d**eviation
RNA	**R**ibo**n**ucleic **a**cid
rRNA	**r**ibosomal RNA
RS	**R**ibo**s**witch
RU	**R**esponse **u**nits
s	**s**econd(s)
SAH	**S**-**a**denosyl **h**omocysteine
SAM	**S**-**a**denosyl **m**ethionine
SAR	**S**tructure-**a**ctivity **r**elationship
SAXS	**S**mall-**a**ngle **X**-ray **s**cattering
SBDD	**S**tructure-**b**ased **d**rug **d**esign
SD	**s**tandard **d**eviation

SHAPE	**S**elective 2′-**h**ydroxyl **a**cylation analyzed by **p**rimer **e**xtension
SmaI	*Serratia **ma**rcescens* **I**
SPR	**S**urface **P**lasmon **R**esonance
STD	**S**aturation **T**ransfer **D**ifference
suppl.	**suppl**emented
tat	**T**rans-**a**ctivator of **t**ranscription
TERRA	**Te**lomeric **r**epeat-containing **R**N**A**
THF	**T**etra**h**ydro**f**olate
TOCSY	**To**tal **c**orrelation **s**pectroscop**y**
TPP	**T**hiamine **p**yro**p**hosphate
Tris	**Tris**(hydroxymethyl)aminomethane
tRNA	**t**ransfer RNA
TROSY	**T**ransverse **r**elaxation **o**ptimized **s**pectroscop**y**
T. tengcongensis	*Thermoanaerobacter tengcongensis*
UTR	**Un**translated **r**egion
vdW	**v**an **d**er **W**aals
waterLOGSY	**Water**-**L**igand **O**bserved via **G**radient **S**pectroscop**Y**
WCo	**W** (tungstate) **co**factors
XDS	**X**-ray **D**etector **S**oftware
xpt-pbuX	**X**anthine **p**hosporibosyl**t**ransferase; pbuX encodes for xanthine-specific purine permease
ypaA	gene encoding for a riboflavin transport protein

Chapter 1

Introduction

1.1 Antibiotic drug discovery and the need for new drugs

1.1.1 Antibiotics and multi-drug resistance

Since the discovery of the antibacterial action of penicillin isolated from *Penicillium* fungi in the 1920s[1], several classes of antibiotic drugs with different modes of actions have been developed. However, the corresponding mechanisms of action of all the different compound types can be summarized to only a small number when subdivided according to the targets that are addressed (Table 1.1).[2,3] A huge impact on the human health in general and an increased life span are attended by the introduction of antibiotics.[4] Particularly in the beginning of the twentieth century, antibacterial drug discovery was mainly driven by screening natural products for antibiotic effects. Later this was called the "golden era" of antibacterial drug discovery.[5] Until 1962 twenty new antibacterial compound classes entered the market. However, this development abruptly stopped thereafter and only to other new compound classes with antibacterial action were introduced two the market in more than 50 years.[6]

TABLE 1.1: Overview of the most important current antibiotic drugs on the market, their classification and their main bacterial targets. The table was compiled according to a classification from reference[3].

Bacterial target	Antibacterial compound (class)
Cell-wall synthesis	• penicillins • cephalosporins • glycopeptides • carbapenems • monobactams
Folic acid metabolism	• sulphonamides • trimethoprim

TABLE 1.1: Overview of the most important current antibiotic drugs on the market, their classification and their main bacterial targets. The table was compiled according to a classification from reference[3].

Bacterial target	Antibacterial compound (class)
Protein synthesis (ribosome)	• macrolides • chloramphenicol • clindamycin • aminoglycosides • tetracyclines
DNA gyrase	• quinolones
RNA polymerase	• ansamycines

Despite the success story behind the discovery of compounds with antibacterial activity, the effort to discover new antibiotics needs to continue. A widespread occurrence of multi-drug resistant bacteria has evoked a looming antibiotic-resistance crisis and has made infections that had already been curable deadly again. Unsuccessful antibacterial discovery programs accompanied the increasing number of resistant pathogens and this in turn accelerated the negative process.[7] The term "post-antibiotic era" has already evolved from this situation.[8] One reason for the low output of drug discovery programs aiming to identify novel antibacterial compounds is the economical point of view. New antibiotics yield lower revenues than other drugs. Prescriptions typically last only for weeks and new drug approvals are most often directed towards severe bacterial infections restricting their use.[9]
There is evidence that a relationship between extensive use of antibacterial compounds and increasing numbers of resistant strains exists. Reductions in the prescription frequency and a more restricted use can lower the rate of resistance emergence.[10] A variety of molecular mechanisms are known to confer to intrinsic or acquired resistance of bacterial strains. Reduced permeability of the outer membrane, drug efflux pumps, mutations of the target site or direct modifications of the antibiotics, such as the β-lactamase-catalyzed degradation of penicillins, are among the most common resistance mechanisms.[11] The so-called ESKAPE pathogens are responsible for the majority of hospital-acquired infections in industrial countries and especially the Gram-negative representatives of these pathogens become progressively resistant.

1.1.2 The need for new drugs

Apparently, there is an urgent need for novel antibiotic scaffolds and novel targets to overcome the resistance problem. The two novel classes of antibiotics which were

introduced to the market after the golden era of antibiotic drug discovery, oxazolidinones and cyclic lipopeptides, both are only active against Gram-positive bacteria like MRSA.[6] Other new drugs approved in the field of antibiotics after the early 1960s were mainly derivatives of the major antibiotic classes which were designed by chemical tailoring and left the core intact. This resulted in novel antibiotics with altered activity spectra. The development of second, third or fourth generation compounds does not hide the fact that medicinal chemistry modifications will probably not solve the resistance problem indefinitely.[12] New targets for example can be explored by structure-based drug design approaches. These new targets could also be found among RNA structures. In the past the ribosomal RNA for example has already been widely used as a target structure for antibiotics.[13]
Riboswitches offer another opportunity for the exploration of novel targets in antibiotic drug discovery. These gene regulatory elements in bacterial mRNAs function as structured receptors and recognize low-molecular weight ligands.[14] Several known antibacterial compounds act by targeting riboswitch structures.[14–17] Therefore, recent studies have begun to encourage drug discovery based on this type of targets.

1.2 Riboswitches

1.2.1 Basic principle of riboswitches

Riboswitches are non-coding RNAs mostly found in bacteria.[18] These *cis*-acting gene regulatory RNA elements[19] are typically found in the 5′-untranslated regions (5′-UTR) of bacterial mRNA.[20] They are generally composed of two distinct structural domains, an aptamer domain, which is responsible for the specific interaction with a small molecule ligand, and a downstream expression platform, which controls the expression of downstream genes.[21] Riboswitches usually control the expression of genes for metabolite transport and biosynthesis, whereas the specifically binding ligand most often is a product or precursor of the biosynthetic pathway they control. Therefore, gene regulation by riboswitches often is a kind of feedback control mechanism.[18] If the concentration of the recognized small molecule ligand exceeds a certain concentration threshold, the aptamer domain acts as a sensor and adopts a certain conformation with a stable assembly of secondary and tertiary structural elements, whereas in the unbound state some nucleotides form alternative structures with parts of the expression platform (Figure 1.1).[22] Hence, the conformational change is relayed to the expression platform conferring major conformational changes to this domain and thereby the downstream gene expression is modulated. The ligand-dependent interchange between these conformations ("riboswitching") is the basic principle of riboswitch activity and explains the name.[18,23,24] Since the expression platform is the mediator that dictates the gene expression level, it is less conserved among different species than the aptamer domain to enable differential responses upon ligand binding in different cellular contextes.[25]

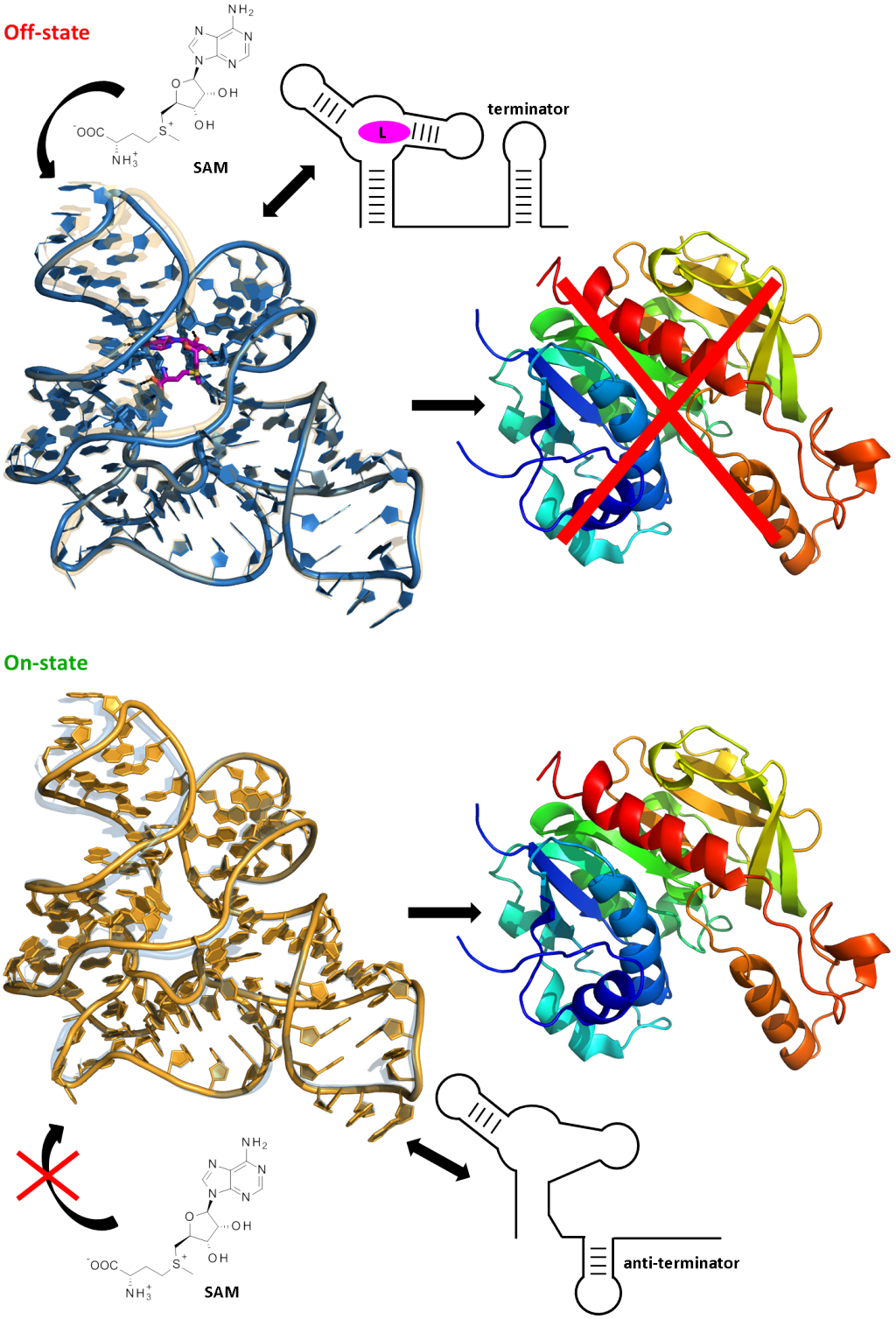


FIGURE 1.1: The SAM-I riboswitch (A94G variant) of *Thermotoga maritima* as an example of a typical transcriptionally controlled "off"-riboswitch. In the SAM-bound structure (pdb-code: 3iqr, upper panel, *blue cartoon*) a conformational change in the expression platform results in the formation of a terminator hairpin and transcription is repressed. Proteins involved in bacterial methionine biosynthesis are less synthesized. In the metabolite-free state (pdb-code: 3iqp, lower panel, *orange cartoon*) the expression platform adopts a different conformation and by formation of an anti-terminator hairpin protein synthesis continues. The ligand SAM is displayed with purple carbon atoms and *black dashed lines* indicate polar interactions within the binding pocket. The ligand-bound and ligand-free aptamer domains differ only marginally. For clarification purposes the formation of the terminator- and anti-terminator hairpins are illustrated in addition.

Depending on the conformation of the riboswitch RNA gene expression is either repressed ("off"-state) or ongoing ("on"-state). There are two general types of riboswitches. An OFF-switch prevents gene expression in the metabolite-bound state, whereas in the ligand-free state gene expression remains unaltered. This is the most common form[23], probably because from a genetic point of view the demand for mechanisms preventing excess amounts of metabolite is higher.[26] ON-switches operate in the opposite way, and albeit being less common, some prominent examples exist among certain glycine, adenine or lysine riboswitches.[23]
Riboswitches are subdivided into almost 30 different classes. A riboswitch class is defined by the natural ligand and fold of the RNA. The recognized ligand and the global riboswitch architecture are not correlated and therefore one and the same ligand can be involved in several riboswitch classes. In the case of the SAM riboswitch for example four different classes (I, II, III and I/IV) exist.[27–30] For almost all relevant riboswitch classes at least one three-dimensional structure has been determined.[18,24,31] The different small molecule ligands targeting riboswitch RNA include purines, coenzymes, sugars and ligands as small as anorganic ions (Table 1.2).[18,31,32]

TABLE 1.2: Known riboswitch ligands grouped according to their compound class. This table was adapted from Sherwood and Henkin.[32]

Coenzymes	**Nucleotides**	**Amino acids**	**Sugars**	**Ions**
AdoCbl	Adenine	Glutamine	GlcN6P	Magnesium
FMN	Guanine	Glycine		Fluoride
MoCo/WCo	$PreQ_1$	Lysine		Manganese
SAM	ATP			Nickel/cobalt
SAH	c-di-GMP			
THF	c-di-AMP			
TPP	2′-dG			

Riboswitches are capable of recognizing their cognate ligands via the aptamer domain with high affinities and a remarkable specificity.[16] K_D-values for riboswitch-ligand interactions range from mid micromolar for GlcN6P binding to the *glmS* ribozyme to picomolar values as observed for some TPP, FMN or c-di-GMP riboswitches.[33] Since riboswitch ligands are structurally diverse, there are different strategies to ensure selective binding and discrimination against closely related ligands[34]:

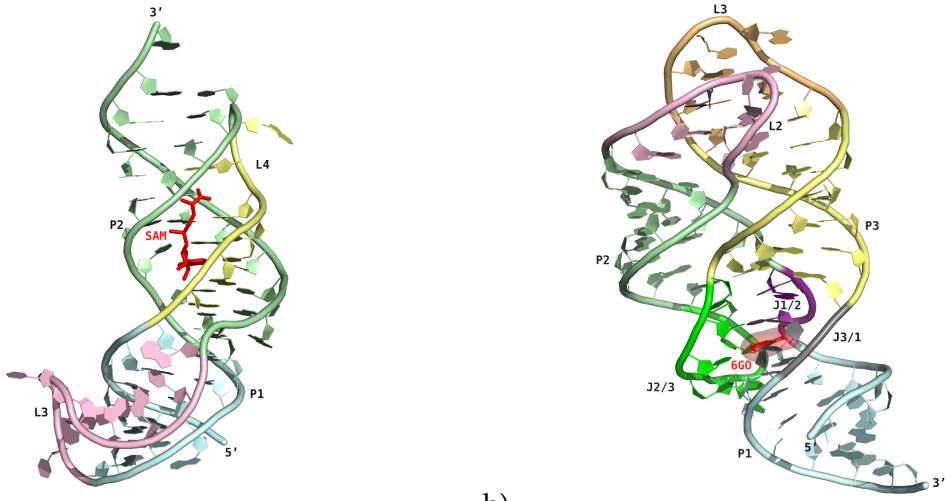
Shape complementarity: Especially for small ligands, such as glycine, a tight binding pocket discriminates against larger ligands[35] and full encapsulation of the ligands increases specificity.[36]

Specific hydrogen bonding: In riboswitches recognizing nucleobases like adenine and guanine all heteroatoms are involved in intermolecular interactions.[36,37]

Stacking interactions: π-π-stacking between aromatic rings of ligands and RNA bases is critical for high affinity and selectivity in some riboswitch classes, for instance the TPP[38], the FMN[39], and the preQ$_1$ riboswitches.[40]

Recognition of charged groups: Negatively charged phosphate groups or carboxylate groups are often coordinated by Mg^{2+}-ions (TPP[38], FMN[39] and GlcN6P[41]) and K^+-ions (lysine[35]), respectively. Positively charged groups are also recognized by specific electrostatic forces. The sulfonium group of SAM makes electrostatic contacts with oxygen atoms of conserved uridines.[28]

Overall, the architectures of the different riboswitch classes are very diverse and rarely related to each other. However, the folding architectures can be subdivided into two large groups. Riboswitch RNA structures are globally folded in pseudoknotted or junctional shapes. Pseudoknotted riboswitches, such as the preQ$_1$-I, preQ$_1$-II and SAM-II classes, are based on a knot-shaped conformation formed through base pairing between a loop of a RNA stem-loop and an outside region (Figure 1.2 a). Ligands interact with pseudoknots either within the junctional region (e.g. preQ$_1$-I[40]) or are positioned in such a way that they are surrounded by the loop nucleotides (e.g. SAM-II[28]).



a) b)

FIGURE 1.2: Riboswitch classes can be subdivided into two major groups according to their folding principle. In both riboswitch structures stem(P)-loops(L) and junctions(J) are colour-coded and labeled, respectively. a) The SAM-II riboswitch class (pdb-code: 2qwy) is an example of a pseudoknotted riboswitch. SAM is trapped by nucleotides of an outside loop (L4), which is involved in base-pairing to stem-loop P2. b) The guanine riboswitch bound to the ligand 6-O-methylguanine (pdb-code: 3fo6) contains a three-way multihelical junction. Ligand binding occurs at the region connecting all three helices by interaction with nucleotides of J1/2, J2/3 and J3/1.

Junctional riboswitches contain a multihelical junction which connects various numbers of helices. These helices can stack on each other in the junctional region or form helical bundles by packing side-to-side above and below the junction (Figure 1.2 b).

Typically, ligands are bound within the junctional region or in regions directly adjacent to the junction. Several riboswitches, including SAM-I, lysine, *glmS*, and THF contain both multihelical junctions and pseudoknots.[34]

1.2.2 Riboswitch mechanisms

Different mechanisms of gene expression control exist as well as different mechanisms by which the ligand sensing is controlled. Five different gene expression control mechanisms with further subcategories are known:[18]

1. transcription control
    - transcription termination
    - rho-dependent transcription termination
    - transcription antitermination
2. translation control
    - translation attenuation
    - translation activation
3. ribozyme cleavage and mRNA degradation
4. alternative self-splicing and translation activation
5. *trans* translation repression

Most riboswitches control gene expression by interference at the transcriptional level or by influencing translation initiation.[18,23,32] Whether a particular riboswitch acts via transcriptional or translational control, is also dependent on the taxonomic group the respective bacterial strain belongs to. Among the often Gram-positive Firmicutes including for instance *Clostridia* and *Bacilli* transcription attenuation is predominant, whereas in all other bacterial groups the translational control is more common.[23,24] Transcriptional riboswitches usually act as OFF-switches (Figure 1.3) . A ligand concentration exceeding a certain threshold results in the formation of a GC-rich intrinsic terminator stem, which is followed by a poly-U stretch. This in turn triggers the stabilization of the terminator stem and the RNA polymerase (RNAP) dissociates from its DNA template. The resulting incomplete RNA transcript then prevents the production of a functional protein. In the absence of ligand binding, the RNA folds into an anti-terminator stem and sequences that would be part of the terminator stem are trapped. Thus, expression of downstream genes continues (Figure 1.3 a).[18,23,32] Transcription antitermination proceeds the opposite, since ligand addition prevents formation of the terminator stem. An example would be the *ydhL* adenine riboswitch aptamer domain of *Bacillus subtilis* (Figure 1.3 a).[42] The third mechanism

of transcriptional control utilizes rho-dependent terminators instead of intrinsic terminators. Sufficiently high concentrations of metabolite foster a RNA conformation that permits rho to interact with the nascent RNA via the *rut* (Rho utilization) binding site. This results in transcription termination. In contrast, at lower ligand concentrations Rho binding sites are sequestered and transcription continues.[43] Translational riboswitches typically block translation initiation upon ligand-induced structural rearrangements. Thus, an OFF-switch is more common. In lieu of the terminator helix in transcriptionally acting riboswitches, a sequestrator helix is involved in base-pairing with the Shine-Dalgarno (SD) sequence and thereby occludes the access to the ribosome binding site (RBS) of the downstream gene. In the absence of ligand an alternative antisequestrator structure is formed, which increases the accessibility to the RBS (Figure 1.3 b) which is sequestered in the presence of a ligand.[18,23,32] Exceptions are known as well for translational riboswitches, where ligand binding prevents the sequestering sequence to bind to the SD sequence and thereby enhances the accessibility to the RBS (Figure 1.3 b).[44]

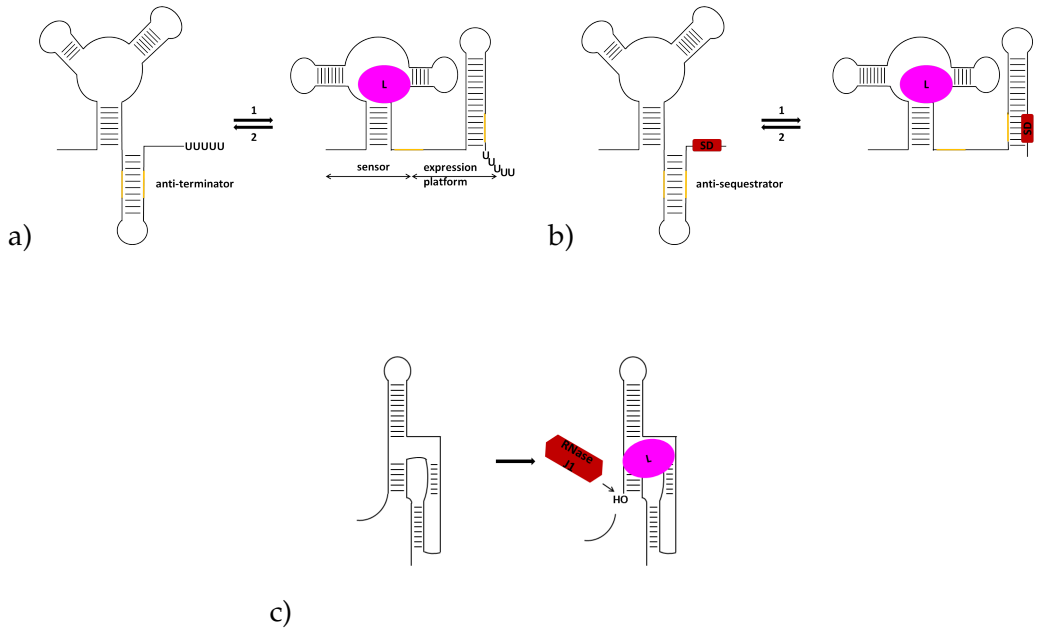


FIGURE 1.3: Common riboswitch control mechanisms (adapted from Mayer and coworkers[16]). a) Transcription control by an OFF-switch (1) operates by secondary structure changes upon cognate ligand binding that lead to the formation of a terminator hairpin stabilized by base pairing of the switching sequence (yellow) with parts of the terminator hairpin sequence. A stretch of uridines follows the terminator hairpin sequence. ON-switches (2) function by an abolition of the terminator hairpin interactions upon metabolite binding. The switching sequence is part of base-pairing interactions that result in an anti-terminator sequence. b) Translation control by OFF- and ON-switches proceeds in a similar fashion. In OFF-Riboswitches inhibiting the translation initiation (1) the SD-sequence becomes sequestered via base-pairing. ON-switches (2) permit translation initiation by formation of an anti-sequestrator hairpin. c) The *glmS* ribozyme has a unique regulatory mechanism. Upon binding of GlcN6P the autocleavage of the ribozyme is fostered. The resulting free 5′ end is prone to degradation by RNase J1.

The three other riboswitch control mechanisms are far less frequent. In the case of the *glmS* ribozyme mRNA degradation is the mechanism of gene expression regulation. Ligand binding activates ribozyme activity and this in turn enhances the RNase J1-mediated RNA degradation starting at the freshly generated 5′ end (Figure 1.3 c).[18,32,45] To date only one riboswitch is known to operate not only as typical *cis*-acting regulatory element, but also to function in *trans*. In *Listeria monocytogenes* certain SAM riboswitches operate by an antisense mechanism. After SAM-dependent transcription attenuation, these riboswitches pair with the 5′-UTR of the mRNA encoding for a virulence factor thereby attenuating translation.[18,46] An additional mechanism was revealed in *Clostridium difficile*. The c-di-GMP riboswitch controls a ribozyme in an allosteric fashion. Ligand binding induces conformational changes allowing the self-excision of a group I self-splicing intron and translation of the specific mRNA is activated.[18,47]

Riboswitches can sense ligand binding by either a thermodynamic or kinetic control mechanism.[18,44] Transcriptional riboswitches usually regulate gene expression by kinetic control and require ligand concentrations higher than their K_D to initiate the ligand-dependent pathway of gene expression regulation.[34] Ligand concentrations are often $\sim$ 100-1000-fold above the dissociation constant to achieve half-maximal transcription termination (T_{50}) during gene expression regulation of FMN, SAM-I and lysine riboswitches for instance.[24] Since stabilization of the terminator stem is an irreversible process, the regulatory decision has to be made prior to that.[32] In transcriptional regulation the time frame for reaching equilibrium binding between RNA and cognate ligand is very short. The RNA-folding kinetics compete temporally with ligand binding kinetics and the extent of transcription termination is largely dependent on the association rate of the metabolite. Therefore, kinetic control often takes place, when the rate of metabolite binding exceeds the rate of RNA folding.[48] In a kinetic regime the riboswitch needs to fold and bind its ligand co-transcriptionally to influence transcription.[44]

In contrast, riboswitches under thermodynamic control operate in the equilibrium state of ligand binding. In translational riboswitches the extent of mRNA translation is a dynamic response to the fraction of bound riboswitch.[48] Such riboswitches can therefore be considered as true switches, because translation is turned on and off upon association and dissociation of the metabolite.[34] Signal interaction with translational riboswitches can be both co- and posttranscriptionally.[32] Albeit translational regulation can provide a faster response to the environmental stimuli, the disadvantage here is the energetic cost of folding. The thermodynamics of RNA folding have to compete with the thermodynamics of ligand binding. Hence, the mRNA conformation in the ligand-bound state has to be more stable than other conformations.[48] Unnecessary transcripts due to inactive mRNA conformations are conceivable.[32] Besides the role of RNA folding, the decision whether the process

of "riboswitching" is kinetically or thermodynamically driven depends on the transcription rate of the RNA polymerase. Several factors including NTP- and Mg^{2+}-concentration as well as the DNA template can cause pausing of the transcription elongation. Depending on the extra time gained at this pauses, the speed of the RNA polymerase determines whether a riboswitch is kinetically or thermodynamically controlled.[48,49]

1.2.3 Distribution of riboswitches in different bacterial groups

Riboswitches have been identified in the genomes of archaea, fungi and plants, but most riboswitches have been found in bacteria. Especially in Gram-positive bacteria riboswitches are widespread (Table 1.3). Throughout the taxonomic groups of bacteria the AdoCbl riboswitch occurs most frequently, but TPP, FMN and SAM-I riboswitches are also very common in many groups. The amino acid-sensing riboswitches adopt an intermediate position by being widespread among certain taxonomic groups while being absent from others, whereas purine, GlcN6P and $preQ_1$ riboswitches are found only in a few groups. The highest abundance of riboswitches is given among the Firmicutes and Fusobacteria. Also γ-Proteobacteria, such as *E. coli* carry various riboswitch classes, but the individual species usually carry fewer classes in total and less representatives per class. In contrast, riboswitches are very rare in Chlamydiae and Spirochetes.[23] Most strikingly, riboswitches control genes essential for survival and virulence in many pathogenic bacteria, which underlines that riboswitches might be suitable antibacterial drug targets.[14]
Despite the fact that the vast majority of riboswitches identified were found in prokaryotic organisms, the occurrence of riboswitches in eukaryotes was proven. The first confirmed eukaryotic riboswitch was the TPP riboswitch of the plant *Arabidopsis thaliana*.[50] For this riboswitch also a TPP-bound crystal structure was determined (pdb-code: 3d2g).[51] In addition the TPP riboswitch was also found in the 5'-UTR introns of fungal genes.[52] The presence of introns in eukaryotes necessitates a different mechanism of gene expression control. TPP riboswitches regulate genes via alternative splicing. Binding of TPP to its aptamer domain compromises the "normal" splicing by making a formerly hidden splice site accessible to the proteins of the spliceosome. Alternatively spliced mRNA thereupon contains internal stop codons that can cause premature translation abortion or destabilized transcripts are produced.[18] Less validated is the presence of AdoCbl, $preQ_1$, and SAM-II riboswitches in eukaryotes. Sequences very similar to the ones of known functional riboswitches have been found. However, a closer inspection of the sequences found revealed that they were not functional.[53] Furthermore, a novel riboswitch candidate was examined in the fungus *Aspergillus nidulans*, where arginine binding controls arginase activity.[54]

TABLE 1.3: The distribution of 15 classes of riboswitches in pathogenic bacteria. Riboswitch classes are grouped according to the type of their cognate ligand. Numbers indicate the predicted riboswitch count per bacterial strain, whereas the numbers in parentheses represent the number of genes regulated by these riboswitches. The data table was compiled with data and adapted from[55].
[a] Amino acids

	AAs[a]		**Coenzymes**									**Nucleobases**			**Sugars**
	Glycine	**Lysine**	**AdoCbl**	**FMN**	**Mo(W)Co**	**SAM-I**	**SAM-II**	**SAM-III**	**SAM-IV**	**SAH**	**TPP**	**c-di-GMP**	**preQ$_1$-1**	**Purine**	**GlcN6P**
Acinetobacter baumanii	1(1)		1(2)	1(1)							1(1)				
Bacillus anthracis	1(1)	4(4)	1(1)	2(5)		17(36)					7(19)	1(1)	2(5)	6(9)	1(1)
Brucella melitensis	1(3)		4(8)	1(1)			1(1)				2(11)				
Clostridium difficile	2(2)	1(1)	3(25)	2(5)		5(5)					2(7)	12(21)		6(9)	1(1)
Enterococcus faecalis		1(1)	2(5)	1(1)				1(1)			2(5)		2(3)	1(2)	1(1)
Escherichia coli		1(1)	1(2)	1(1)	1(5)						3(11)				
Francisella tularensis				1(5)			1(1)				1(1)		1(1)		
Haemophilus influenzae	1(1)	1(1)		1(1)	1(1)						3(11)				
Helicobacter pylori											1(2)				
Klebsiella pneumoniae		1(1)	2(21)	1(1)	1(6)						4(14)				
Listeria monocytogenes	1(3)	1(1)	2(20)	1(1)		7(14)					2(5)		1(6)	2(3)	1(3)
Mycobacterium tuberculosis	2(3)		2(4)						1(1)		2(6)				
Pseudomonas aeruginosa			5(24)	1(2)						1(1)	1(1)				
Salmonella enterica			2(22)	1(1)	1(1)						3(11)				
Shigella flexneri		1(1)	1(1)	1(1)	1(5)						3(12)				
Staphylococcus aureus	1(3)	2(2)		2(5)		4(6)					2(7)		1(3)	1(4)	1(1)
Streptococcus pneumoniae	1(1)			2(5)				1(1)			4(11)			1(2)	
Vibrio cholerae	1(1)	3(3)	1(1)	1(1)	1(5)						2(7)			2(2)	
Yersinia pestis			1(2)	1(1)	1(5)						3(8)				

1.2.4 Selected riboswitch ligands with antibacterial activity

Several riboswitches are considered to be possible targets for novel antibiotics.[14,16,17] For some riboswitches even validation by compounds known to exert their antibiotic activity due to interaction with the riboswitch RNA is present. The most advanced target validation has been achieved for the purine[56], the TPP[57], and the FMN riboswitch (Figure 1.4).[58–61] To serve as a potential drug target, the riboswitch should occur in pathogenic bacteria and downregulate the expression of genes essential for survival or virulence.[14] Furthermore, a druggable binding site is required to allow binding of drug-like small molecules mimicking metabolite binding.

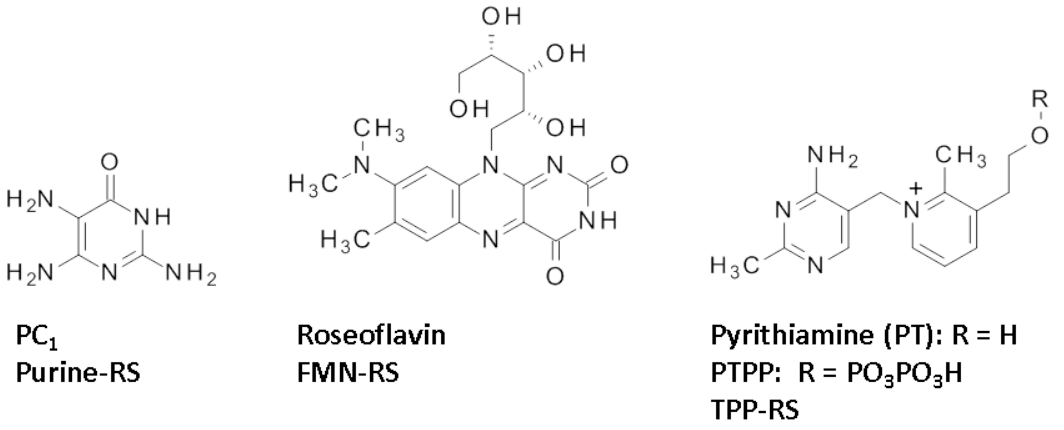


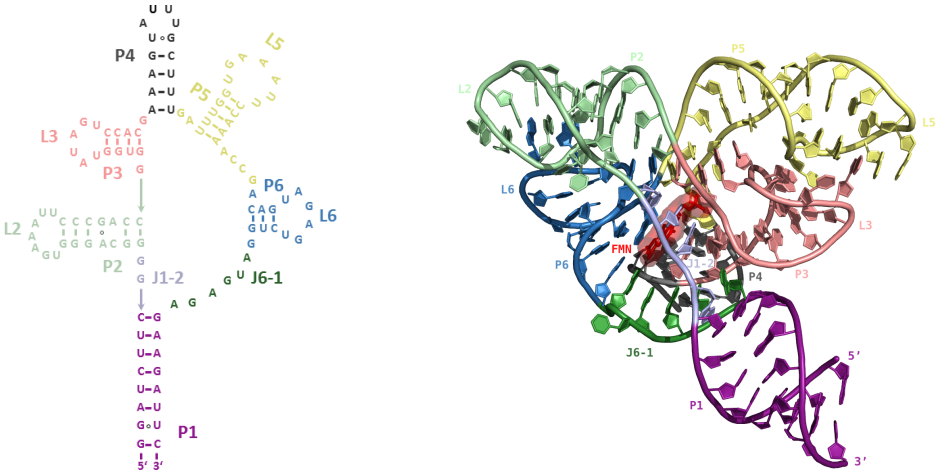
FIGURE 1.4: Ligands for three different riboswitches together with their chemical name and their respective target riboswitch.

1.3 The FMN riboswitch

1.3.1 Overall structure of the FMN riboswitch and recognition principle

Already in the year 1999 a highly conserved RNA structure, the *RFN* element, was reported to exhibit regulatory function on genes related to riboflavin biosynthesis. In Gram-positive bacteria it was observed that this *RFN* element is located upstream of the regulated genes.[62] A mechanism, by which the *RFN* element regulates gene expression, was proposed using comparative analyses of genes and regulatory elements in 58 types of bacteria. Potential hairpins were found downstream of all *RFN* elements and it was suggested that upon FMN binding alternative structures are formed by interactions of a base stem of the *RFN* element with the regulatory hairpin. Mainly in Gram-positive bacteria, such as *Thermotoga maritima*, transcription termination was proposed as regulation mechanism based on the sequence analysis. In Gram-negative bacteria highly conserved sequences downstream the *RFN* element strongly indicated a translation attenuation mechanism upon FMN binding via sequestration of the RBS.[63] Structural probing studies with the 5′ UTRs of the *ribD* and *ypaA* genes from *B. subtilis* confirmed the *RFN* element being a FMN

riboswitch. The very high selectivity and affinity ($K^D \approx 10$ nM) for its cognate ligand was determined as well. Although only these two genes were investigated, already the two most common riboswitch regulation mechanisms, translational and transcriptional control, were observed. Structural changes occurred in the region occupied by the SD-sequence upon FMN binding to the *ypaA* mRNA, whereas transcription termination resulted from FMN interaction with the *ribD* mRNA.[20]
The best studied model system for the determination of the binding modes of various ligands for the FMN riboswitch is the aptamer domain of the *Fusobacterium nucleatum* FMN riboswitch controlling the regulation of a riboflavin transporter gene (*impX*). All 13 crystal structures available to date originate from this RNA-type. The resolution of each of these structures is about 3.0 Å. In addition to several FMN-bound structures[39,64], structures with two other flavins, riboflavin (pdb-code: 3f4g) and roseoflavin (pdb-code: 3f4h), bound to the RNA target have been determined (Figure 1.12).[39] Most recently, co-crystallization with a novel ligand of completely different molecular structure, ribocil, was achieved (pdb-codes: 5kx9 and 5c45, Figure 1.12).[61,65] The *F. nucleatum* FMN riboswitch aptamer domain used for crystallization of all the published three-dimensional structures and throughout this work consists of 112 nucleotides when employed as a full-length product.[39] Though, some structures were obtained by annealing two strands of RNA with slight modifications to facilitate crystal packaging.[39,61] The length of the aptamer domain varies from one bacterial species to another. For instance, the entire aptamer domain of a FMN riboswitch influencing the gene expression of the *ribD* gene in *B. subtilis* has more than 160 nucleotides.[49,64]
Concerning the different mechanisms by which riboswitches can regulate gene expression and which were described in the previous section (1.2.2), it can be emphasized that the FMN riboswitch most likely acts as a kinetically-driven riboswitch effecting transcription attenuation when FMN exceeds a threshold concentration. Although the K_D-value for the interaction between aptamer domain and FMN was ≈ 10 nM in a fluorescence-based assay, much higher FMN concentrations were necessary to result in half-maximal transcription termination ($T_{50} > 1$ µM).[49]
Crystallization experiments revealed that the FMN riboswitch belongs to the group of junctional riboswitches. The overall secondary structure is composed of a six-helical junction (Figure 1.5). Two peripheral domains built by helices P2 and P6 and P3 and P5, respectively, branch off at the junction. Within each peripheral domain stabilizing tertiary contacts through either loop-loop or loop-helix interactions exist. Thus, an overall butterfly-like fold can be observed. The non-peripheral stems P1 and P4 spread in opposite directions from the bottom of the junction. Remarkably, both peripheral domains are very symmetrical to each other. The ligand binding site is positioned centrally within the junctional region. Hence, FMN is surrounded by all six stems forming the aptamer domain (Figure 1.5 b and Figure 1.7 a).[39]



a) b)

FIGURE 1.5: The structure of the *F. nucleatum* FMN riboswitch. a) Secondary structure diagram. All six stem-loops of the multihelical junction are colour-coded. In addition the two most important joining regions (J1-2 and J6-1) involved in FMN recognition are highlighted. b) The crystal structure of the FMN riboswitch (pdb-code: 3f2q) indicates the butterfly-like fold of the RNA. FMN residing in the middle of the junctional region is displayed with *red sticks* and a transparent surface. The same colouring scheme as for the secondary structure diagram is used.

Multiple hydrogen bonds and stacking interactions are the basis for the high affinity binding of FMN to the target site (Figure 1.6). The planar isoalloxazine heterocycle is kept in place by face-to-face π-π-stacking interactions to A48 and A85. The uracil-like edge of FMN forms Watson-Crick-like hydrogen bonds with A99. One hydroxyl group of the open ribose side chain donates a hydrogen bond to the carbonyl oxygen of G11 and the phosphate oxygens interact via multiple hydrogen bonds with Watson-Crick edges of the conserved guanines G10, G11, G32, G62, and G84. Additionally, one interaction between the phosphate group and the RNA is bridged by a binding site Mg^{2+}-ion. Hence, one phosphate oxygen of FMN is coordinated with the N7-atom of G33.[39] The Mg^{2+}-ion forms water-mediated contacts to three guanines of the P3 helix (G32, G33 and G41) and to one guanine of the P5 helix (G62) (Figure 1.7 b). Since the FMN analogues riboflavin and roseoflavin (Figure 1.12) are lacking the phosphate group, the magnesium ion was absent in the respective crystal structures.[39]

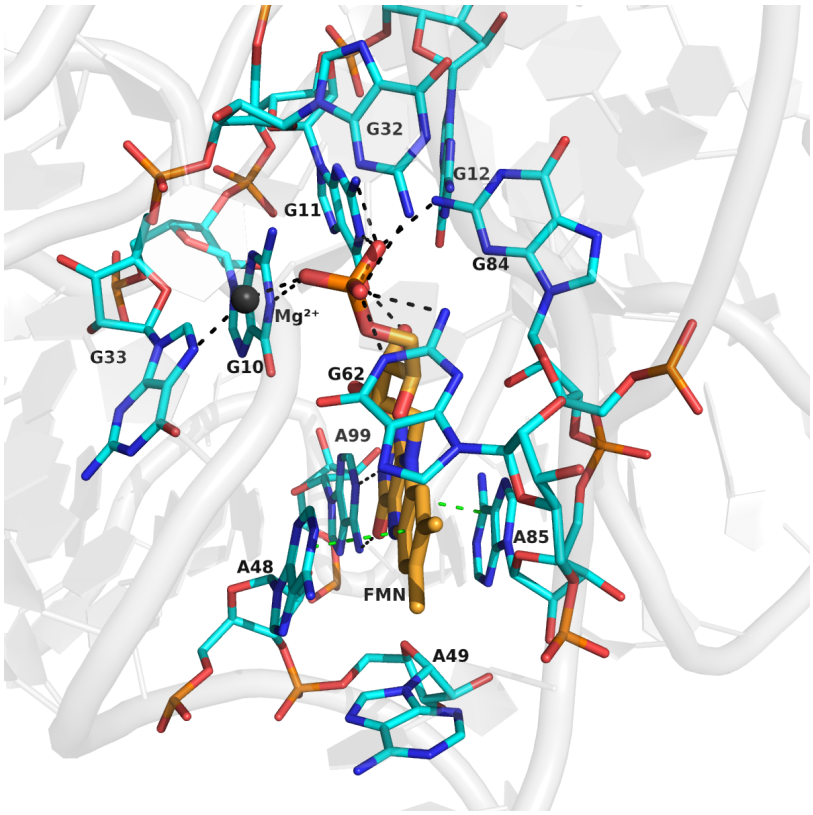


FIGURE 1.6: The ligand FMN is displayed with orange carbon atoms in stick representation, and the RNA bases involved in binding interactions are shown with cyan carbon atoms. Multiple hydrogen bonds are mediated via the phosphate group of FMN displayed by *black dashed lines*. A magnesium ion displayed as *black sphere* is also involved in the coordination of the phosphate group. The *green dashed lines* represent π-π-stacking between FMN and A48/A85.

The extensive interaction network between FMN and the RNA bases enables a strong discrimination between FMN and closely related ligands. Fluorescence quenching assays with the isolated, *in vitro* transcribed aptamer domain and flavin ligands revealed, that the lack of the phosphate group in riboflavin reduces binding affinity by three orders of magnitude (37.5 nM to $\approx$ 40 µM), whereas removal of the ribityl moiety from riboflavin resulting in lumiflavin only marginally decreases the binding affinity further. A large drop in binding affinity is probably due to the loss of multiple interactions to conserved guanines of the P3-P5 peripheral domain and the J1-2 junctional region.[39,58]

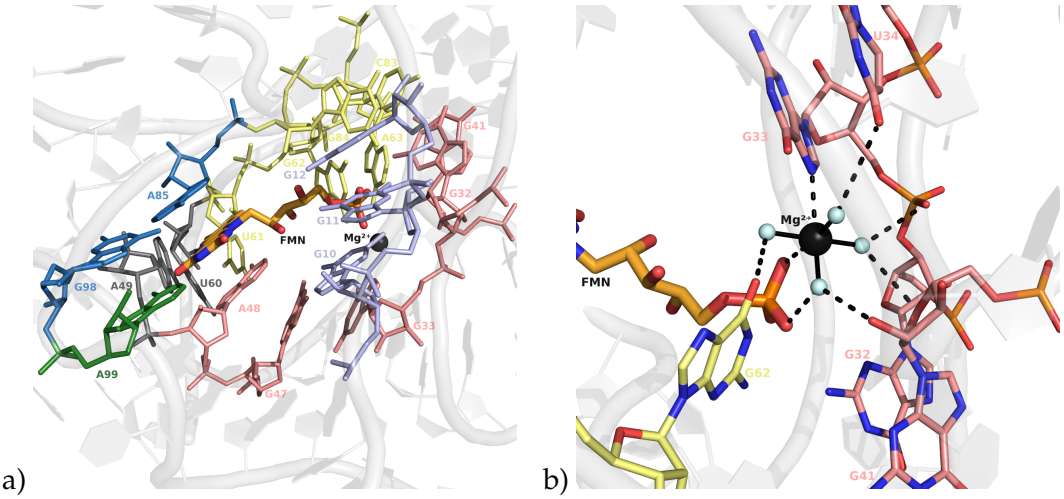


FIGURE 1.7: Binding site residues of the FMN riboswitch RNA. The FMN-bound crystal structure determined by Vicens et al. (pdb-code: 2yie,[64]) was used to prepare the figure. a) FMN displayed with orange carbon atoms is bound within the junctional region of the riboswitch. Contacts and interactions to almost all stem-loops and junctional regions except for helices P2 and P6 are formed. The binding site residues are colour-coded in the same fashion as in figure 1.5 and the binding site Mg^{2+}-ion is depicted as *black sphere*. b) The magnesium ion is involved in a direct contact to one of the phosphate oxygens and in several water-mediated contacts. Again, the same colouring scheme was applied to the RNA residues. Conserved water molecules are illustrated as *palecyan spheres* and the contacts between RNA and water molecules are depicted as *black dashes*. Coordination bonds between Mg^{2+} and water molecules are displayed as *black lines*.

Footprinting assays indicated relatively rigid peripheral domains. The restriction patterns upon addition of different RNases were comparable with and without addition of FMN. Therefore, the FMN-induced small changes suggested a largely preformed conformation of the binding site.[39] To further support the hypothesis of only weak conformational changes in the RNA to accommodate FMN binding, SAXS-analyses and SHAPE mapping[66] were carried out on either the *ribD* FMN riboswitch of *B. subtilis* alone or together with the *impX* FMN riboswitch of *F. nucleatum*.[64,67] SAXS data confirmed that the riboswitch aptamer domain of *B. subtilis* adopts the same global fold in the presence and absence of FMN over a wide range of different Mg^{2+}-concentrations.[67] Experimental data from SHAPE mapping were in good agreement with these findings. In both the *B. subtilis* and the *F. nucleatum* FMN riboswitch the reactivity of the nucleotides widely remained unchanged, regardless of the presence of ligand. The major conformational changes were observed for the nucleotides within the joining regions of the RNA. These bases which are all located 4 Å around FMN displayed a decreased reactivity indicating a more stable conformation.[64]

The ligand-free FMN riboswitch structure of the *F. nucleatum* aptamer domain (pdb-code: 2yif) confirmed the SHAPE data. A superposition of two FMN-bound, riboflavin-, roseoflavin-, and ribocil-bound structures with a free riboswitch structure indicates the largest movements within the joining regions J3-4 and J4-5. Here, the

largest conformational difference between the free state RNA and all other well-aligned structures were found (Figure 1.8).

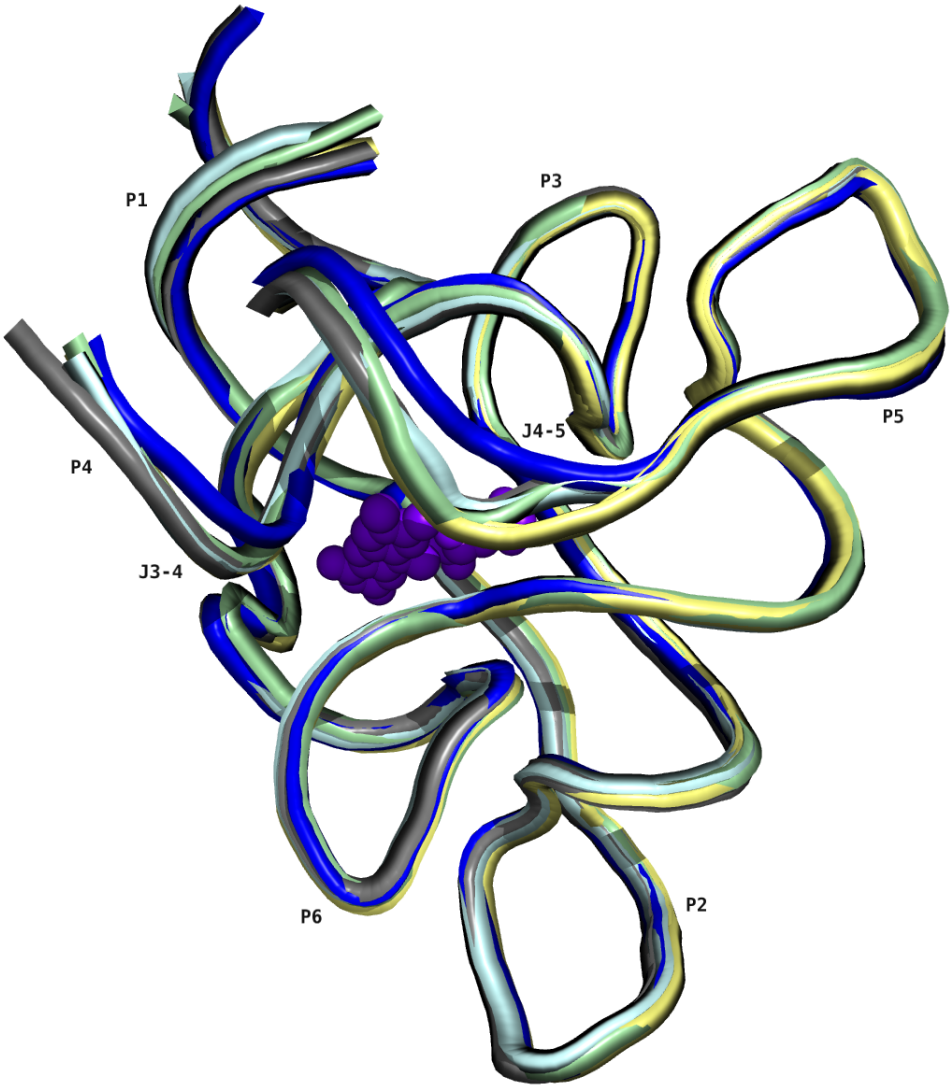


FIGURE 1.8: Overlay of the *F. nucleatum* FMN riboswitch free form (pdb-code: 2yif, dark-blue) with several ligand bound structures reveals the high degree of preorganization in the aptamer domain. Roseoflavin- (pdb-code: 3f4h) and riboflavin-bound (pdb-code: 3f4g) structures are displayed in pale-yellow colour. FMN-bound structures (pdb-codes: 3f2q and 2yie) are shown in pale-cyan and grey colour, respectively. The ribocil-bound structure (pdb-code: 5c45) is displayed in pale-green and FMN atoms are represented as *purple spheres*. For clarity only the RNA backbone is shown.

In the joining region J3-J4 A48 undergoes large conformational changes (Figure 1.9). Instead of stacking to A49 like in the FMN-free structure the heterocycle flips about 90° in the FMN-bound structure to enable stacking interactions with the isoalloxazine ring. As a consequence the Watson-Crick base pairs A49-U60 and A48-U61 become disrupted. Additionally, G62 in the joining region J4-5 moves towards the phosphate of FMN to form two hydrogen bonding contacts (Figure 1.9).[64]

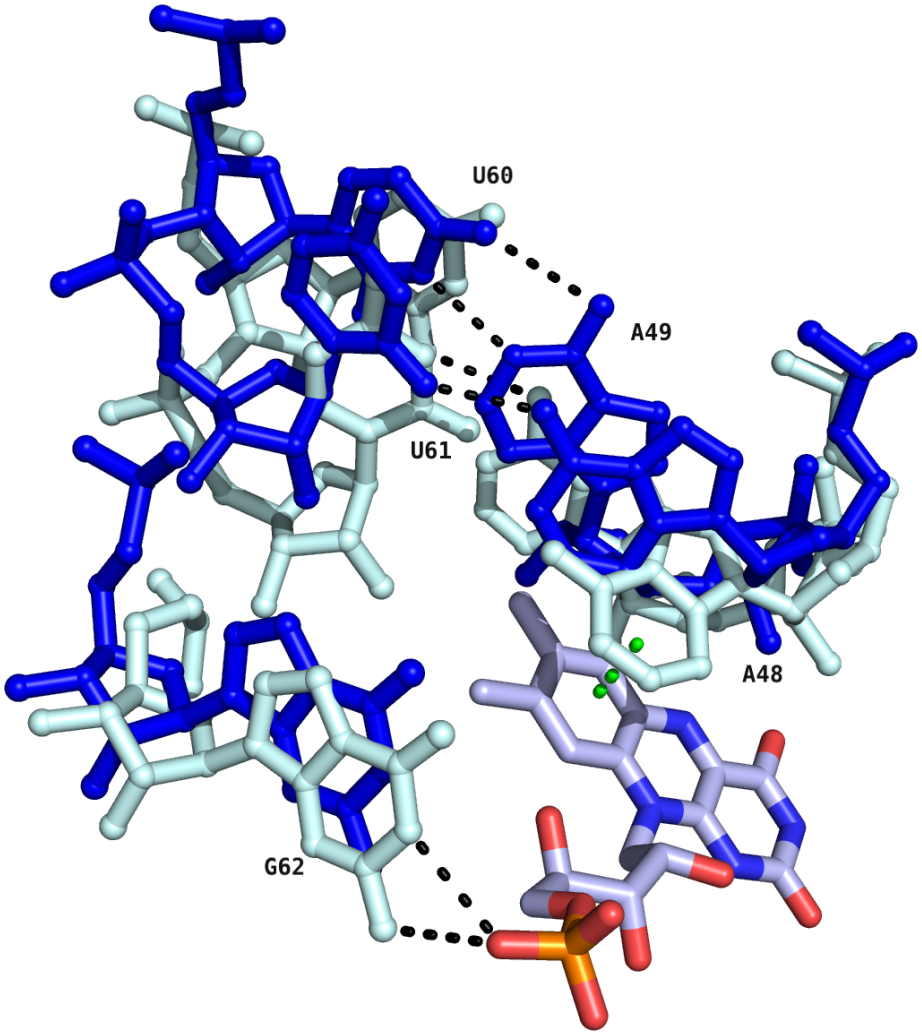


FIGURE 1.9: Conformational changes of RNA bases within J3-4 and J4-5 at the FMN binding site. The free form is displayed in dark-blue colour (pdb-code: 2yif) and the FMN-bound structure (pdb-code: 2yie) in pale-cyan colour. FMN is coloured with light-blue carbon atoms. Hydrogen bonds are illustrated as *black dashed lines* and stacking interactions as *green dashes*.

The crystal structures of the FMN riboswitch bound to the flavin analogues roseoflavin (pdb-code: 3f4h) and riboflavin (pdb-code: 3f4e) only slightly differ from the FMN-bound crystal structure (Figure 1.10).[39] Though, there are some ligand-dependent conformational changes that are observed. Especially nucleotides in the phosphate-binding region undergo spatial adjustments near the ligands without a phosphate group. The hydrophobic edges become repositioned and the flexible ribityl moiety adopts different conformations in all flavin-bound structures. Especially nucleotides around the isoalloxazine ring of FMN superpositioned on each other with low rmsd-values when aligning the riboflavin-bound with the FMN-bound crystal structure. The atomic positions of one of the adenines (A48) which are involved in stacking interactions to the isoalloxazine ring deviated little. Due to fact

that face-to-face π-π-stacking interactions are usually parallel displaced with a certain distance tolerance[68] stacking interactions between A48 and FMN would still be possible when investigating the spatial arrangement of FMN in the riboflavin-bound crystal structure. The largest conformational differences are observed for the residues G62, G10 and U61. In the FMN-bound structure especially G10 and G62 are moved inwards in the direction of the phosphate position to accommodate for hydrogen bonding interactions with the phosphate group. G10 and G62 are part of the residues involved in the hydrogen bonding network between the binding site and the phosphate group. The more pronounced deviation for the atomic positions of U61 near the hydrophobic edge of the flavins in the presence of either riboflavin and FMN does not seem to be of great importance, since no strong interactions are expected between this residue and the hydrophobic dimethylbenzene ring.
The FMN riboswitch RNA also provides large openings (Figure 1.11 a) for drug-like compounds. A broad activity against various types of bacteria might be possible. Aptamer domains are usually highly conserved in riboswitch structures (e.g.[21]), but especially the binding site residues are highly conserved among different bacterial strains in the case of the FMN riboswitch. All bases that are involved in either stacking or hydrogen bonding interactions belong to the group of the conserved residues (Figure 1.11 b).[39]
Taken together the results indicate the amenability of the FMN-binding pocket to a structure-based drug design approach towards the identification of antibacterial compounds. While the riboswitch RNA has sufficient inherent plasticity to recognize different ligands, no large conformational changes upon ligand binding were observed thereby facilitating predictions on novel ligands. This is supported by the fact that the FMN riboswitch has already been validated as drug target[58,60,61], which underlines that a structure-based drug design could be valuable to discover novel FMN riboswitch ligands.

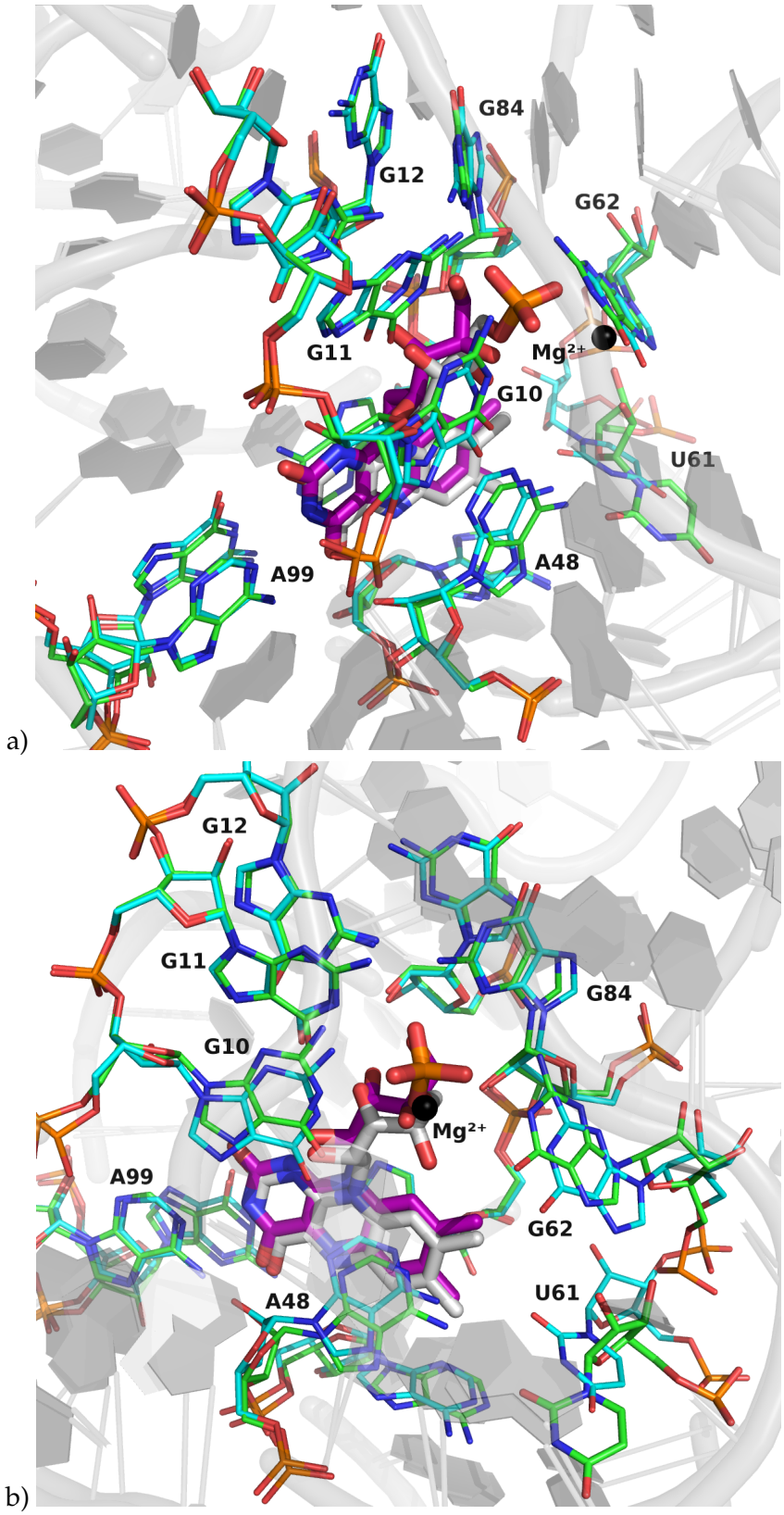
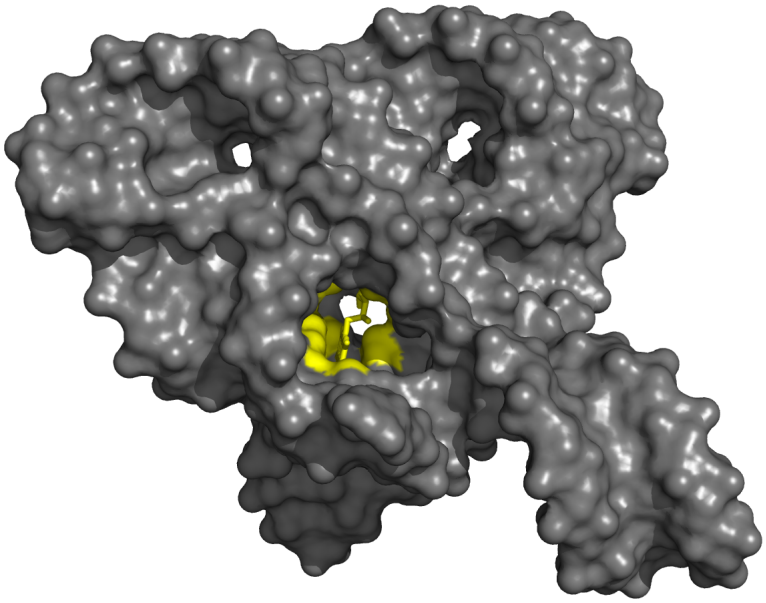
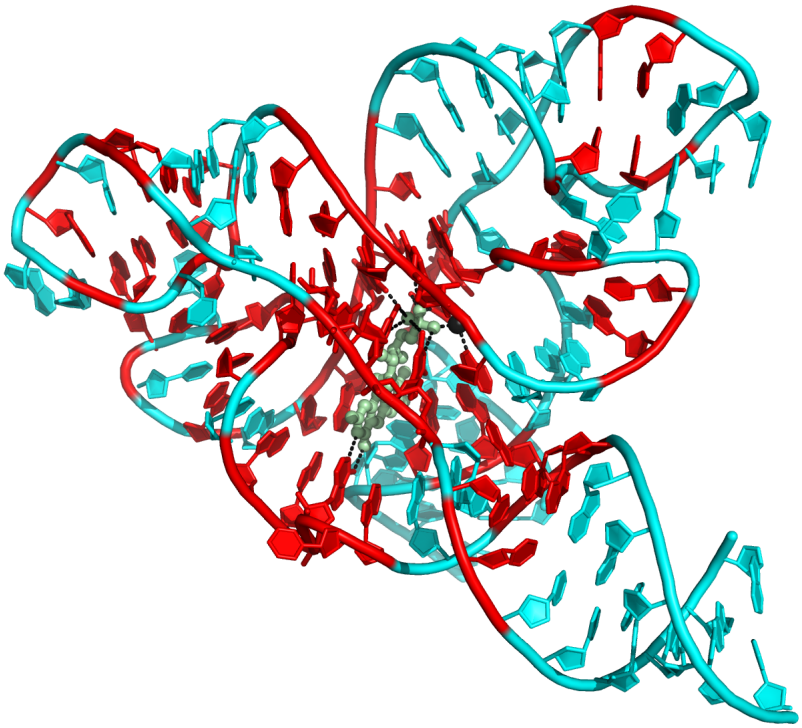


FIGURE 1.10: Overlay of the FMN-bound crystal structure (pdb-code: 3f4e) and the riboflavin-bound crystal structure (pdb-code: 3f4h). Residues of the binding pocket are depicted as *sticks* with green (3f4e) and cyan carbon atoms (3f4h), respectively. The ligands are displayed with *offset sticks* and purple (riboflavin) and white carbon atoms (FMN), respectively. A magnesium ion crucial for FMN recognition is displayed as *black sphere*. a) Positioning indicating the well-superposed A99 in both crystal structures. b) Closer inspection of the phosphate-binding sites revealing the most significant conformational differences.



a)



b)

FIGURE 1.11: Structural features of the FMN riboswitch. a) The surface representation of the FMN-bound structure reveals the large openings of the binding pocket highlighted by a yellow-coloured surface around FMN atoms shown as *yellow sticks*. b) RNA bases with a high sequence conservation among different bacteria are displayed in red colour, other bases are coloured cyan. FMN is shown in pale-green colour with balls and sticks and the binding site Mg^{2+}-ion is illustrated as *black sphere*. *Black dashed lines* indicate the hydrogen bonds between FMN and the RNA.

1.3.2 FMN riboswitch as drug target for antibiotics

Three compounds with antibiotic activity were identified to date, where FMN riboswitch binding has been validated as at least one mode of action. Two of these compounds, roseoflavin and 5FDQD (Figure 1.12), are similar to FMN.
Roseoflavin, a natural pigment isolated from *Streptomyces davawensis*, has been known a long time for its antimicrobial properties.[69] In-line probing assays were employed to detect roseoflavin binding to the *ribD* FMN riboswitch of *B. subtilis*. The characteristic spontaneous RNA cleavage patterns were similar to the ones previously observed for FMN with concentration-dependent cleavage rate changes of binding site residues. It was assumed that the dimethylamino substituent on the isoalloxazine ring replacing the methyl group of riboflavin permits more favorable interactions, since the K_D-value was $\approx$ 100 nM and therefore the affinity was significantly higher than the binding affinity of riboflavin to this particular aptamer domain ($K_D \approx 3$ µM). Reporter gene assays were successfully used to confirm that the expression of a gene under the control of the FMN riboswitch is repressed in *B. subtilis* cells when roseoflavin is added. Roseoflavin-resistant bacterial strains possessed mutations within the *ribD* FMN riboswitch aptamer domain leading to a disruption of crucial interactions.[58] Suggesting that the FMN riboswitch is the target for roseoflavin it was further postulated that roseoflavin phosphate might be the biologically relevant antimicrobial compound, since the flavokinase encoded by the *ribC* gene in *B. subtilis* does not discriminate between roseoflavin and riboflavin.[70]

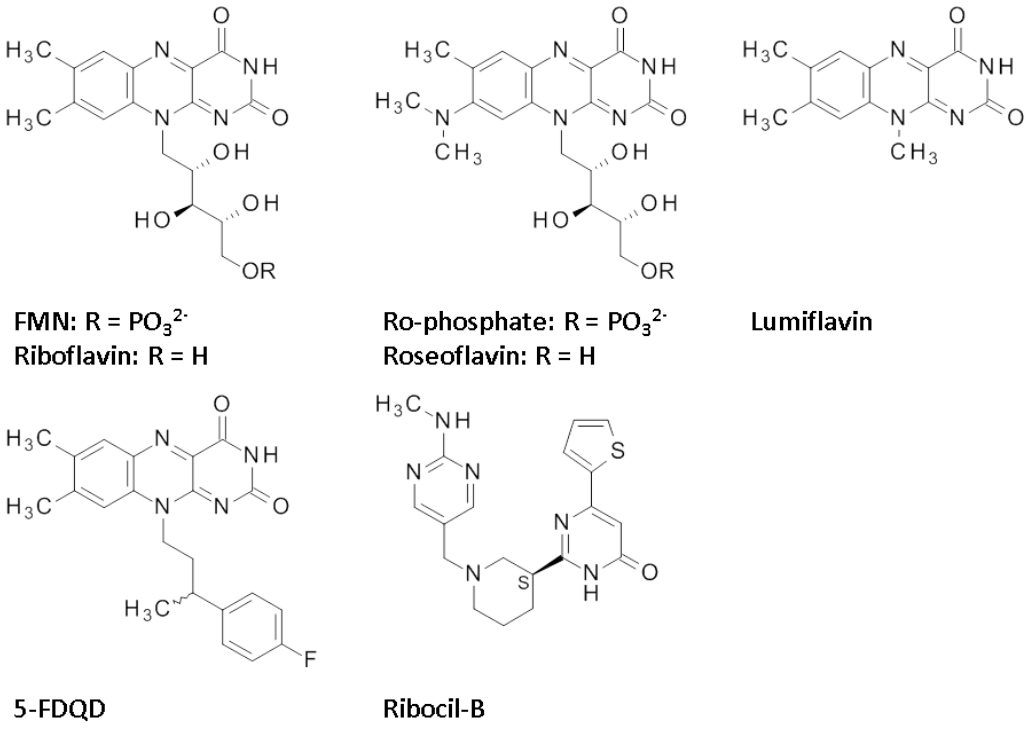


FIGURE 1.12: FMN riboswitch ligands with antibacterial activity at least partially attributed to interaction with the FMN riboswitch aptamer domain.

In another bacterial strain, *Listeria monocytogenes*, growth inhibition was also achieved by addition of roseoflavin. These bacteria are not able to synthesize riboflavin *de novo*. Gene expression of a riboflavin uptake transporter was downregulated in the presence of roseoflavin. Target validation was achieved by site-directed mutagenesis of the FMN riboswitch sequence resulting in roseoflavin-resistant strains. However, by a riboswitch-independent mechanism the virulence of *L. monocytogenes* was increased. Roseoflavin competes with riboflavin for transport or several enzymatic reactions and is likely to influence targets other than the riboswitch as well. This could prove deleterious for a drug candidate.[71] Briefly afterwards it was identified that roseoflavin also is a substrate of human flavokinase and potentially renders flavoenzymes inactive giving rise to toxicity issues related to this compound.[72]
The flavin analogue 5FDQD (Figure 1.12) was identified via medicinal chemistry optimization of the known flavin analogues that interact with the FMN riboswitch. Mice with a *Clostridium difficile* infection of the intestine were cured with an efficacy comparable to the usually administered antibiotics vancomycin and fidaxomicin. In-line probing assays were utilized to determine, whether binding to the FMN riboswitch is likely to be the mechanism of antimicrobial activity. FMN and 5FDQD induced nearly identical structural changes in the *B. subtilis ribD* FMN aptamer domain and 5FDQD reached an equivalent high affinity (7.5 vs. 6.4 nM). *In vitro* transcription termination assays also predicted a similar potency as FMN to induce riboswitch function. The high affinity of 5FDQD was rather unexpected, since the aryl-alkyl side chain is unable to form an extensive hydrogen bonding network like the ribityl moiety and the phosphate group of FMN. Besides the high *in vivo* efficacy, pharmacokinetic parameters were desirable as well and 5FDQD affected the normal cecal flora less than vancomycin for example. A narrower spectrum can be expected. It has to be further elaborated, if the binding affinity to the FMN riboswitch RNA is exclusively contributing to the antimicrobial effects.[60]
Most recently, a chemically distinct FMN riboswitch ligand was identified by a phenotypic screen of a library of diverse compounds. The growth inhibitory activity of each of the compounds against an *E. coli* strain was investigated. After it had been elucidated that the antimicrobial activity of this compound named ribocil was reversed upon addition of exogenous riboflavin, the target was identified by whole-genome sequencing of resistant strains. All resistant strains had mutations within the FMN riboswitch RNA of the *ribB* gene of *E. coli* encoding for 3,4-dihydroxy-2-butanone 4-phosphate synthase involved in riboflavin biosynthesis. Ribocil bound to the *E. coli* FMN riboswitch aptamer with a very high affinity. In a fluorescence-based assay the fluorescence of FMN was efficiently restored, demonstrating the competitive binding to the target RNA, while the K_D was determined to be 16 nM.[61]
Co-crystallization of ribocil with the *F. nucleatum* FMN riboswitch aptamer revealed that, albeit a racemic mixture of ribocil was used for crystallization, the (*S*)-isomer ribocil-B (Figure 1.12) apparently is the eutomer, since only this enantiomer was bound to the riboswitch. Interestingly, the observed conformation permits much

less hydrogen bonds but a more extensive network of stacking interactions when compared to the biologically active conformation of FMN (Figure 1.13). Within the binding pocket ribocil adopts an U-shaped conformation and the pyrimidonyl oxygen is the key hydrogen bond acceptor interacting with the exocyclic amino function of A99 and the C2'-OH of A48. The pyrimidonyl ring is involved in face-to-face stacking with A85, whereas the thiophenyl group stacks face-to-face with A48 and edge-to-face with A49.[61] One methylene group of the piperidinyl ring is postulated to act as weak hydrogen bond donor to the carbonyl oxygen of G11, although more often CH groups bound to N or O atoms in aromatic heterocycles act as donors.[73] Lastly, the methylamino-pyrimidinyl is involved in face-to-face π-π-stacking to G62.[61] Even though the chemical structure of ribocil differs significantly from the structure of FMN, ribocil is still positioned inside the junctional region of the six RNA stems interacting with most of the key residues FMN also is interacting with.

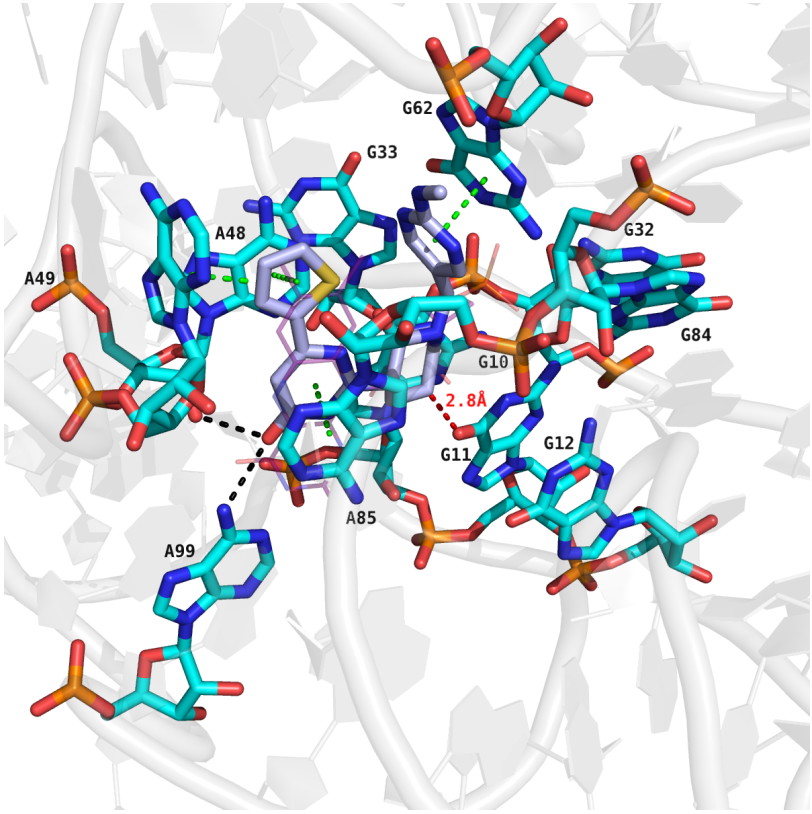


FIGURE 1.13: Binding mode of the antibacterial compound ribocil. The binding site residues (pdb-code: 5c45) are represented as sticks with cyan-coloured carbon atoms. Ribocil is displayed with light-blue carbon atoms, whereas FMN is depicted as *transparent sticks* with purple carbon atoms. The π-π-stacking interactions are shown with *green dashed lines* and the hydrogen bonds are displayed as *black dashed lines*. A methylene group of the piperidinyl moiety might act as a weak H-bond donor to the carbonyl oxygen of G11 according to Howe et al.[61], which is highlighted by the *red dashed lines*.

A determination of the microbiological activity against an *E. coli* strain further proved, that ribocil-B is the biologically active form. The mechanism of action seemed to

be selectively mediated via interaction with the FMN riboswitch, since the suppression of bacterial growth was fully reversed upon addition of exogenous riboflavin. Limited off-target effects are therefore expected. Furthermore, the pharmacological effect of a riboswitch-mediated inhibition of riboflavin biosynthesis was demonstrated in a murine *E. coli* septicaemia model.[61] Most recently, with a new crystal structure of a close analogue of ribocil carrying a methoxy substitution instead of a methylamino group (ribocil-D) the previously described binding mode of ribocil was reaffirmed (pdb-code: 5kx9). Initial SAR efforts were undertaken to enhance the antibacterial activity of the lead compound ribocil. The hydroxy-pyrimidine piperidine core was identified to be essential. Only hydrogen bond donor substituents at the piperidine ring potentially allowing for an interaction with G11 did not lead to a significant loss of activity. The methylamino-pyrimidine substituent is supposed to be the position where the largest changes are permitted.[65]

1.4 Structure-based approaches for RNA ligand discovery

1.4.1 Role of RNA in biology and disease

For a long time RNA was primarily seen as the linker between the genetic information carried by DNA molecules and proteins, whereupon mRNAs are operating as transitory templates, tRNAs are serving as connecting piece between nucleotide and amino acid sequences and rRNAs are building an integral part of the protein synthesis machinery.[74,75] Over the last years of RNA research it became more and more evident that non-coding RNAs (ncRNA) other than tRNA and rRNA are relevant for many processes in living cells. They can be involved in gene silencing on the transcriptional level[76] or control gene expression by specifically binding small molecules accompanied by conformational changes influencing the read-out by the RNA polymerase.[20,77] Especially long non-coding RNAs can modulate protein activities or alter the protein localization by specifically interacting with them.[78] Furthermore, RNA structures, such as ribozymes, can exhibit catalytic activity by either self-cleavage[79] or by cleaving other RNAs in a *trans*-acting fashion.[80]
Certain non-coding RNAs are also involved in the development of various diseases. By mis-splicing or other mechanisms RNA has a major influence on certain diseases like cancer, neurodegenerative disorders, diabetes and genetic disorders.[81,82] Additionally, many diseases are caused by RNA viruses. They are a serious health threat, since viruses with RNA genomes underlie more rapid mutations and new variants are selected faster. Many emerging viruses originating from an animal host are RNA viruses.[83] Among the most prominent candidates of these type of viruses are HIV, HCV and influenza viruses.[84]

1.4.2 Targeting RNA structures with small molecules

RNA exhibits some attractive features and combines properties of the other two major macromolecule types that could be addressed by drugs, DNA and proteins. Similar to DNA the number of building blocks is smaller than for proteins and the regular A-form helix of RNA is often disturbed by unpaired regions, which permits the formation of more complex three-dimensional structures than in DNA and the creation of well-defined binding pockets similar to proteins.[85] Large RNAs are mainly composed of stacked stems that are linked together by tertiary interactions and functional RNA molecules recognize their ligands by specific atomic interactions and shape complementarity, especially by base-pairing and base-stacking.[86]
The best evidence for RNA being a valid drug target is found among the ribosomal RNAs, since the majority of antibacterials exert their function by affecting the ribosomal protein synthesis machinery. In addition, these drugs have been introduced to the clinic since around the 1950s.[12] These types of antibiotics interacting with either the small (30S) or the large (50S) ribosomal subunit of procaryotes include aminoglycosides, macrolides, tetracyclines, chloramphenicol, and oxazolidinones (Table 1.1).[87] Aminoglycosides for example interact with a scrutinized RNA drug target, the 16S A-site rRNA in the small ribosomal subunit. They cause a mis-incorporation of amino acids in the growing polypeptide chain.[88] Successful integration of these antibiotics into therapies of infectious diseases proves the possibility to exploit RNA targets for the discovery of new drugs.
For many other functional RNAs it was also proven that they could possibly serve as drug targets. For example, HIV-1 offers even two RNAs, which could be addressed by small molecule ligands (Figure 1.17 and Figure 1.16). The HIV-1 transactivation response (TAR) element has been recognized as a potential target for the treatment of HIV by inhibiting the interaction with the Tat protein with small molecule ligands.[89,90] Further, a special tRNA, $tRNA^{Lys3}$, is the natural primer for the reverse transcription and several ligands have been reported.[91] Another potential antiviral target is the HCV IRES RNA, more precisely subdomain IIa. This is another target for small molecule inhibitors interfering with HCV translation.[92] The influenza A RNA promoter is another potential antiviral drug target.[93] Furthermore, for several riboswitches, RNA structures that could possibly represent novel targets for antibacterial compounds, small molecule ligands have been identified.[14,15,17,58] Even catalytic RNA structures, such as the group I intron RNA, are attractive target structures for therapeutics, because they predominantly occur in key genes of pathogenic microorganisms. Aminoglycoside antibiotics for example also act as inhibitors of catalytic RNA and could therefore serve as starting point for drug design.[94,95] There is not only evidence for suitable RNA targets in the viral and bacterial genome. TERRA molecules are required for telomere lengthening in cancer cells. These molecules fold into G-quadruplexes and could therefore be specifically targeted to fight certain cancer types.[96] Recently, the design of a small molecule directed against an oncogenic miRNA involved in breast cancer development was reported.[97]

Special considerations have to be taken into account when identifying drugs targeting functional RNAs due to the unique properties of RNA. Generally speaking, RNA is a negatively charged polyanion and often surrounded by a dense and well-ordered water shell and positively charged metal ions. Unsurprisingly, RNA ligands tend to be strongly polar to displace this shell and often have positive net charges. This illustrates the particular challenges accompanied with drug design towards RNA, since small, charged and polar compounds exhibit worse permeation through membranes and are more prone to non-specific interactions with the negatively charged backbone.[98,99] While a major contribution to affinity in protein-ligand interactions are hydrophobic interactions, the small molecules have to interact with RNA primarily via π-π-stacking interactions, hydrogen bonding, and electrostatic interactions.[100] The hypothesis that the properties of the substrate determine the properties of suitable ligands[98] can be rationalized by comparing the properties of antibacterial compounds interacting either with protein or RNA targets. Antibacterials that modulate protein functions most often comply with the "Rule-of-five" guidelines[101], whereas antibacterials interfering with ribosomal RNA are often larger than 500 Da and have higher numbers of hydrogen bond donors and acceptors. Some aminoglycosides for example have logP-values < -8.5 and polar surface areas (PSA) > 130 $Å^2$ predicting unfavourable ADME properties.[99,102,103] Nevertheless, the closer inspection of databases with protein-binding ligand sets and RNA-binding ligand sets still reveals large overlaps. RNA ligands indeed have on average a larger number of hydrogen-bond donors and acceptors compared to protein ligands (Figure 1.14). However, the ligand sets have the tendency to be skewed by the presence of large, positively charged antibiotics derived from natural products.[98] In addition to these large compounds, they also contain small drug-like molecules with desirable properties demonstrating that those can also bind to RNA molecules.

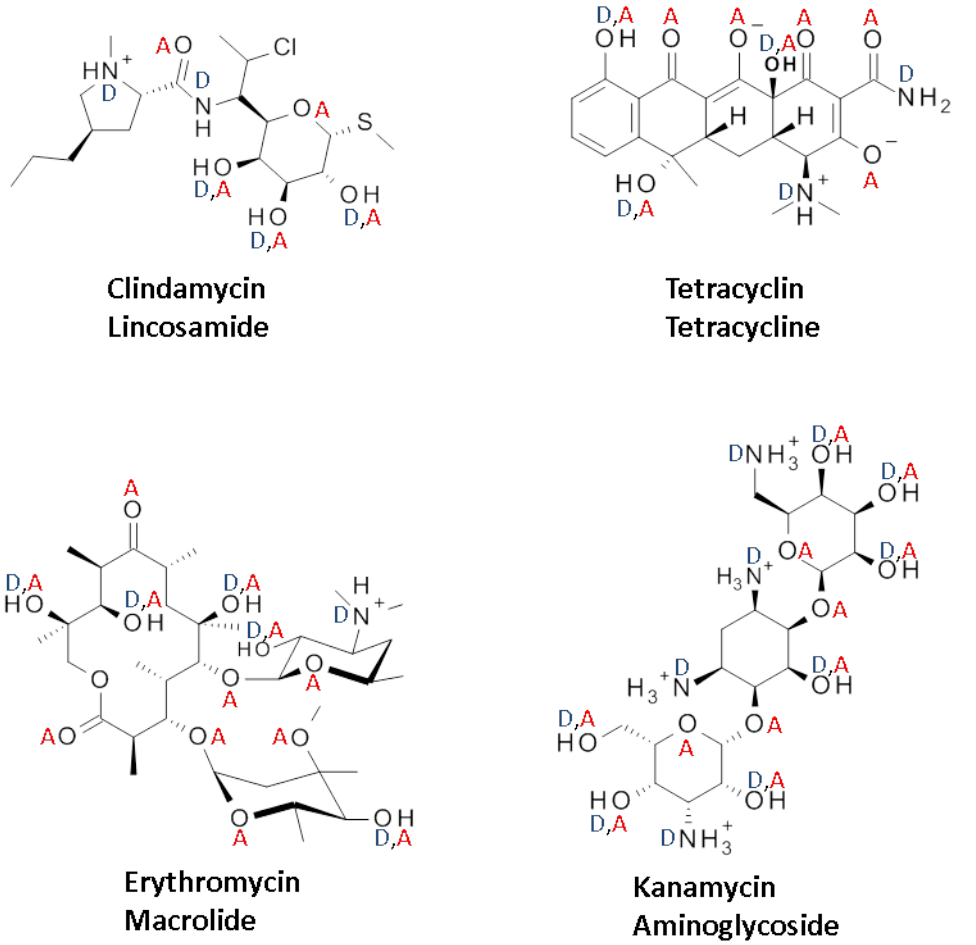


FIGURE 1.14: Examples of antibiotics exhibiting their antibiotic activity by interference with rRNA. The molecular structures, the chemical name and the corresponding antibiotic class are displayed. Acceptor (A, red colour) and donor (D, blue colour) are indicated to reflect the polarity of the representatives. Acceptor and donor positions were determined using the Marvin program suite (v5.11.3, Oct2012, ChemAxon, Budapest, Hungary, `https//www.chemaxon.com/`).

1.4.3 General process of structure-based drug design

Structure-based drug design is an iterative process and finding a lead compound often proceeds through multiple cycles before an optimized lead compound can be developed beyond a preclinical testing.[104] The first cycle of structure-based drug design starts with the choice of a suitable drug target. Ideally a crystal or NMR structure is available and the target macromolecule can be purified for example by cloning into a vector and expression in competent *E. coli* cells. With the structural information molecular docking often is applied to place database compounds into the binding site and select the best compounds according to their scores and ranks, which are subsequently tested in biochemical assays to confirm binding affinity. Following the hit conformation, the structure-based drug design proceeds through the second cycle. The new cycle usually starts with the structure determination of the target-ligand complex to reveal the main interaction sites and possible optimization

sites aiming to increase the potency of the initial hit. Finally, in subsequent cycles the synthesis of optimized leads often is carried out, ideally together with further structure determination. The cycles are repeated until optimized compounds with sufficient affinity and specificity for the target are obtained.

1.4.4 Molecular docking for the discovery of novel ligands for RNA targets

The identification of candidate compounds in a drug discovery process often starts with *in silico* methods. Molecular docking has become a valuable tool to identify novel chemical scaffolds binding to a target of interest and the aim is to predict the binding mode of potential novel ligands and estimate their binding affinities. Several docking algorithms and scoring functions have been developed and several thorough reviews and textbooks are available in general.[105–107] Docking has been extensively applied to proteins with success. In contrast, docking studies for RNA targets are quite rare and there are no drugs in clinical use that target RNA structures and have been developed by SBDD approaches.[108] However, with the steadily increasing number of available crystal structures containing RNA and a small molecule ligand, structure-based approaches are becoming feasible. In September 2016, 303 crystal structures containing RNA and a small molecule ligand ($\leq$ 1000 Da) have already been deposited in the protein data bank (PDB[109]).[99]
The general principles of recognizing ligands by RNA do not differ from those in protein-ligand binding. Nevertheless, some special considerations have to be taken into account when applying molecular docking methods for the prediction of RNA-ligand binding:

1. Due to the highly charged character water molecules and ions are often found at the binding site and have to be considered during docking.[110]

2. RNA structures are highly flexible and often undergo conformational changes upon ligand binding.[111]

3. Re-parameterization is necessary for scoring functions, especially when parameterization for atom types was carried out on protein-ligand complexes.[110]

Based on these considerations two options to apply docking programs for the structure-based drug design for RNA targets are conceivable. Either docking methods and scoring functions that were originally developed for protein-ligand docking are used with adaptations or completely new methods and scoring functions are developed that are tailored for the special requirements of RNA. All major docking programs except for FlexX[112] were applied to RNA-ligand docking so far.[99] Retrospective validation comparing the predicted binding modes with available binding data was carried out with satisfying results for all the docking programs that were employed for RNA-ligand docking. Prospective predictions to analyze the enrichment of known ligands from a large data set of compounds were performed with

AutoDock[113], DOCK[114–118] and ICM[119] among the programs originally developed for protein-ligand docking and with MORDOR[120] and RiboDOCK[121] among the docking tools specialized for dealing with RNA targets.[99]
Collectively, the performance of the docking programs especially tailored for RNA-ligand docking were quite encouraging. MORDOR accounts for the high flexibility of RNA, since the ligand is energy minimized within the binding pocket while receptor and ligand are kept flexible. A force-field based scoring function is used to rank the ligands. In 74% of all test cases the binding mode was predicted correctly, which was expressed by a root-mean-square deviation (rmsd) between docked and crystallographically determined binding mode $\leq$ 2.5 Å.[120] MORDOR was used in a screening for ligands of the potential anticancer target human telomerase RNA (hTRP2b). Different structural motifs were confirmed to bind to this RNA preferentially while exhibiting weak affinities in the high micromolar range.[122] RiboDock was tailored for RNA-ligand docking by the implementation of an empirical scoring function which includes special terms for common RNA-binding groups, such as charged guanidinium groups and a π-π-stacking term to better account for the special RNA environment, in addition to hydrogen-bond terms or terms for lipophilic interactions.[121] Compared to the study using MORDOR less RNA-ligand complexes were available at that time for the retrospective studies and the performance was slightly worse with 50% of the binding modes being correctly predicted.[99] The discovery of new ligands for the bacterial ribosomal A-site RNA was also reported.[123]
Most extensively studied in the field of docking programs originally developed for protein-ligand docking and adapted to RNA-ligand docking are AutoDock, DOCK and ICM. AutoDock was adapted to various extents to better account for the differences in RNA-ligand recognition compared to protein-ligand binding. Solvation parameters for RNA atoms were defined based on similar atom types in amino acids and the overall negative charge of the backbone was reduced to simulate the presence of counter ions and avoid overestimation of electrostatic energies. In 60% of the test cases a binding mode within 2.5 Å rmsd compared to the one identified in the crystal structure was predicted.[124] Later on the flexibility of the RNA was taken into account alongside with the properties of the first solvation shell. An aminoglycoside test set was re-docked with a low rmsd of 1.4 Å on average.[125] Five selective compounds binding to the stem loop 3 (SL3) RNA of the packaging element ψ of HIV-1 were identified by applying DOCK and AutoDock.[126,127] ICM was used in combination with DOCK by applying an adapted scoring function for the identification of HIV-1 TAR RNA ligands. Weighted terms for electrostatic, solvation and vdW energies as well as hydrogen bonding were included together with a penalty term accounting for the loss of rotational flexibility of ligands. The weighting factors were derived from known RNA-small molecule complexes.[128] Improvements were obtained by docking against dynamic ensembles of RNA generated from NMR

structures and MD simulations with known ligands to account for induced conformational changes upon ligand binding. In more than 50% of the test cases the predicted binding mode matched with the experimentally determined binding mode as judged by a low rmsd. In a prospective study six novel compounds selectively inhibited the TAR-Tat peptide interaction and one compound, netilmicin, even exhibited *in vivo* antiviral activity.[129] The program DOCK was optimized in different ways to be suited for RNA-ligand docking, depending on the version. In DOCK 6 the anchor-and-grow sampling algorithm for the placement of the ligands was modified and the AMBER force-field scoring function was supplemented by sophisticated molecular mechanics-based solvation models. For ligands with < 7 rotatable bonds in 80% of the test cases the overlay between docked and crystallographically determined binding mode was very well.[118] In the former versions of DOCK the scoring function itself was left unaltered. The AMBER atom parameters for the phosphorus atoms in the backbone were adjusted to account for the high negative charge of the RNA. To reflect a RNA environment in aqueous solutions the partial charge of the phosphorus atoms in the backbone of the RNA was increased by 1 unit, thus mimicking the presence of sodium counter ions.[124]
DOCK 3.5.54 with minor modifications regarding the partial charges to account for non-specifically bound counter ions has already been applied to screen for riboswitch ligands. Encouraging retrospective and prospective results were obtained for the model system *Bacillus subtilis xpt-pbuX* guanine riboswitch carrying a C74U mutation (= GRA) to become adenine responsive. The predicted binding modes of three known ligands were very close to those determined crystallographically with rmsd-values < 0.34 Å. Furthermore, the known ligands were enriched among the top-scoring ligands and separated well from the decoys included into the test set. Docking of a large database of compounds led to the identification of novel ligands that were either analogues of known ligands or even contained novel chemical scaffolds. Binding affinities could be determined for four out of five top-scoring novel ligands and finally for three ligands a three-dimensional structure was obtained.[130] The crystallographically determined binding mode was compared with the predicted binding mode proving a high quality of the prediction. This is an good indicator that these compounds were not revealed coincidentally, but rather due to a reliable and efficient docking approach.[131]
Collectively, these studies show that RNA-ligand docking is a promising tool for the identification of novel small molecule ligands interacting with functional RNA species. Both, specialized docking algorithms developed for RNA and adapted protein-ligand docking programs resulted in success rates $\geq$ 50% when predicting the correct binding modes. However, the data sets used to validate the docking approaches were rather small and there is still a lot of data that has to be generated to fully compare the performances of RNA-ligand-docking with its protein counterpart.[99] Advantageous for the present study is that DOCK 3.5.54 was successfully applied to discover novel riboswitch ligands, albeit the binding site of the

GRA riboswitch was rather small and deeply buried making this target easier to address than shallow RNA pockets on the outside. Specific problems solely occurring in the RNA-ligand docking field were not encountered in this model study with a riboswitch.[99,130] Since riboswitch RNAs represent a potential class of novel drug targets (see 1.2), this is an incentive to further confirm the hypothesis made on the basis of this model system that novel riboswitch ligands can be identified by virtual screening approaches.

1.4.5 Fragment Screening

In connection with a structure-based drug design approach also fragment screening provides a valuable tool for the retrieval of novel ligands for a given target structure. Relatively small libraries of low-molecular weight compounds (typically 150-300 Da) are screened for weak-affinity ligands. At the cost of a lower potency of the initial hits a fragment-based approach leads to higher hit rates in drug discovery than HTS for instance. The weakly binding fragments are amenable to various optimization procedures, such as merging fragments, linking fragments that bind to different subpockets of a target site, or fragment growing to explore additional areas of the binding site.[132–135]
Fragment hits generally obey a "Rule of Three". They usually have a molecular weight < 300 Da, < 3 hydrogen bond donors and acceptors and logP-values < 3. In addition, the number of freely rotatable bonds is often < 3 and the polar surface area (PSA) is < 60 Å^2.[136] These general guidelines are also valid when applying fragment-based drug discovery to RNA targets. Three different options are conceivable to assemble RNA-directed fragment libraries:[99]

1. Assembly of RNA-binding fragment libraries by literature search
2. Kinase-directed libraries
3. RNA-focused libraries based on privileged substructures

The first approach to derive a focused library is to search published literature for small molecules with a sufficiently high binding affinity to RNA ($K_D \leq 50$ µM) and subsequent fragmentation of these compounds preferably between their ring systems. A fragment library with 102 compounds for example was assembled by this approach and validated by screening against the bacterial ribosomal A-site.[137] Due to the fact that the pharmacophores of ligands with affinity towards the bacterial ribosomal A-site (Figure 1.15 b) and the ATP binding site of kinases (Figure 1.15 a) are very similar, kinase-directed libraries might be suitable for fragment screening against RNA targets. By molecular docking of a kinase library against the ribosomal A-site this was virtually confirmed, but still has to be proven by further experimental validation.[110] All three-dimensional structures of aminoglycosides bound to the ribosomal A-site RNA of bacteria deposited in the PDB[109] reveal two main

features of the binding mode. Aminoglycosides all stack on a conserved guanine (G1491) and form two hydrogen bonds with the Watson-Crick edge of a conserved adenine (A1408).[138] In kinase inhibitors the spatial arrangement of these different pharmacophore features is shared with the one of aminoglycosides binding to the ribosomal A-site RNA.[110] An additional option is the assembly of a focused small molecule library based on privileged substructures. There are several scaffolds that are known to interact with different types of functional RNAs. Small molecules with a *bis*-benzimidazole as core scaffold are capable of competing with aminoglycosides for binding to bacterial rRNA and can also interact with the hepatitis C internal ribosomal entry site (IRES) or CUG repeats.[139] A similarity search for fragment-like molecules with these scaffolds can be applied to compile a fragment library.[99]

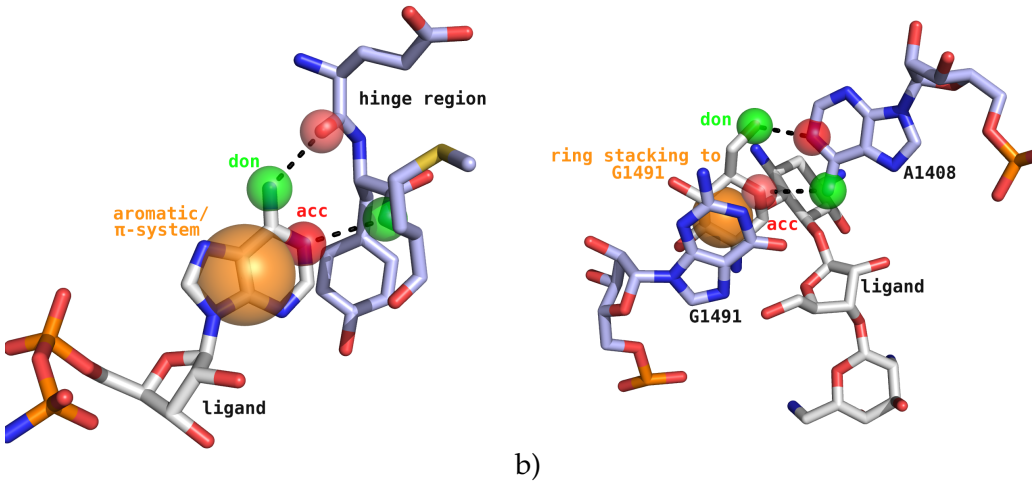


FIGURE 1.15: Shared pharmacophore of kinase- and ribosomal A-site-binding ligands. The exemplary ligands (white carbon atoms) and the binding site residues (light-blue carbon atoms) are shown in *stick mode*. Hydrogen bonds are depicted as *black dashed lines*. a) ATP analogue binding to the hinge region of lymphocyte-specific kinase (pdb-code: 1qpc). Kinase inhibitors contain an aromatic ring (*orange sphere*) and a hydrogen-bond donor and acceptor pair (*green* and *red spheres*) interacting with the hinge region. b) The aminoglycoside paromomycin bound to the ribosomal A-site (pdb-code: 1j7t). An aromatic ring (*orange sphere*) stacks on G1491, a hydrogen-bond donor interacts with N1 of A1408 (*green* and *red spheres*, resp.), and a hydrogen-bond acceptor interacts with N6 of A1408 (*red* and *green spheres*, resp.). The figure was taken and adapted from a review by Wehler and Brenk.[99]

For screening of a fragment library with compounds of presumably low affinity methods with a high throughput and sufficient sensitivity are required. The most studies reporting on fragment screening against RNA targets used NMR-based methods to detect ligand binding. Numerous examples are available, where NMR-studies with fragment libraries led to the identification of small molecules binding to various types of RNA.[99] It can be differentiated between two basic detection principles. In ligand-detected methods, such as STD, waterLOGSY and NOESY, the protons involved in binding to the RNA are traced, whereas target-detected methods monitor the induced changes of RNA chemical shifts in the binding site region upon

the altered environment in the presence of bound ligand. When employing target-detected methods the chemical shifts of either the imino protons of G and U or the aromatic H5/H6 protons of U and C are traced during the NMR-studies. Here, also three common setups are possible: 1D-^{1}H-NMR, TROSY, and TOCSY.[140] Further alternatives to detect fragment binding to RNA targets include suitable mass spectrometry (MS)-based methods allowing rapid characterization of screening compounds, since even mixtures of compounds can be tested in one single experiment without the need for labeling or any other pre-processing.[99] MS-based studies were developed for the identification of HIV-1 transactivation response element (TAR) RNA and 16S ribosomal A-site RNA for example. The combination of fragment screening and virtual screening is also an alternative for the initial fragment hit identification, since the overall performances of fragment docking and docking larger molecules are comparable.[141] In principle surface plasmon resonance (SPR) represents another method for the direct measurement of macromolecule-ligand interactions in order to perform a fragment screening.[99] For the purpose of fragment screening SPR is routinely used in the case of proteins.[142] Although compounds can be rapidly screened, the feasibility of SPR studies on RNA targets still has to be elucidated.

1.4.6 RNA-binding ligands discovered by structure-based drug design

1.4.6.1 RNA-binding ligands discovered by molecular docking

RNA-ligand docking was successfully implemented in the discovery of new ligands for several RNA targets (Figure 1.16). Compounds interrupting the formation of the TAR RNA-Tat complex in HIV-1 were revealed by using adopted versions of DOCK and ICM and for some of them binding to the proposed binding site was confirmed with NMR experiments.[90,128] In another screening for TAR ligands considering the target flexibility by docking into an ensemble of NMR-derived as well as MD-derived conformations using ICM, six compounds were identified as hits with K_i-values from 10 nM-170 µM and one compound even exhibiting *in vivo* activity suppressing replication of a HIV-1 cell line.[129]
With the aid of the consecutive application of the docking engines DOCK and AutoDock ligands binding to RNA tetraloops and to stem loop 3 (SL3) RNA of the HIV-1 packaging element ψ were searched. In the case of the RNA tetraloop one compound was confirmed to bind to its target by NMR (Figure 1.16).[143] For the SL3 RNA one molecule with selectivity for the SL3 RNA against other stem loops in ψ RNA was identified (Figure 1.16).[126]

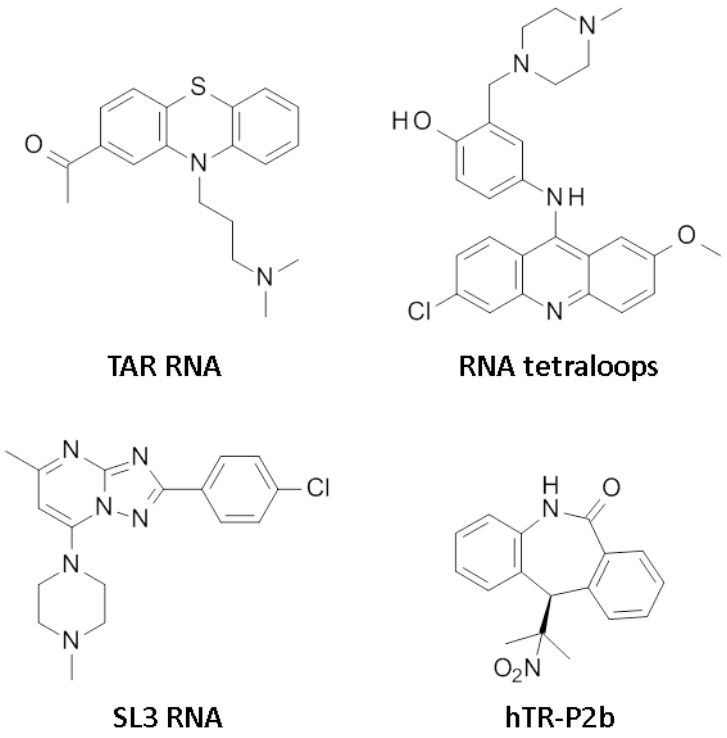


FIGURE 1.16: Selected ligands discovered by RNA-ligand docking. The corresponding target name is written below the compound. This figure was adapted from Wehler and Brenk.[99]

MORDOR was used for the discovery of novel ligands binding to hTR-P2b, a part of the human telomerase RNA (Figure 1.16).[122] After a small database of 3,000 molecules was docked first into the binding site and binding was confirmed by NMR, subsequently a similarity search was conducted in a much larger database. Finally, the most promising candidate compounds were tested and 24 compounds bound selectively to hTR-P2b RNA with dissociation constants ranging from 84 µM to 1.5 mM.

1.4.6.2 RNA-binding ligands discovered by fragment screening

Similar to molecular docking approaches fragment-based drug design has proven to be an alternative for the identification of novel ligands for various RNA types as well. Lead structures were successfully discovered for several RNA structures (Figure 1.17).[99] HIV-1 TAR was not only addressed by a molecular docking approach, but also by means of an FBDD approach.[144] Disruption of the Tat-TAR interaction was tested by a competitive fluorimetric assay with a labelled Tat protein. 2,4,6-triaminoquinazoline (Figure 1.17) exhibited the highest affinity with an IC_{50}-value of 40 µM. Further studies with ^{1}H-NMR revealed that this compound has two binding sites within the TAR RNA.

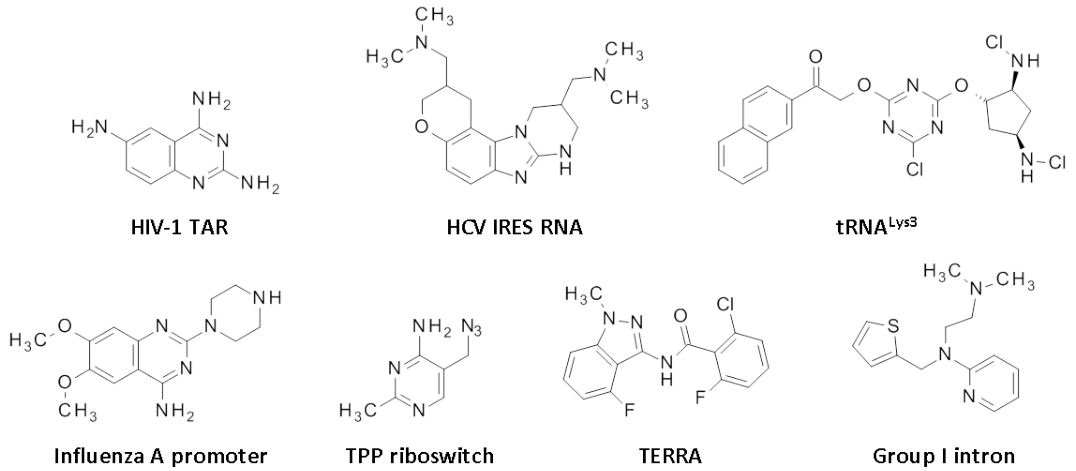


FIGURE 1.17: Novel small molecule ligands of various functional RNA species identified by FBDD approaches over the past years. Most of these ligands exhibit affinities in the µM-range towards their drug targets. This figure was adapted from Wehler and Brenk.[99]

Another potential novel antiviral target fragment-based methods were applied to is the subdomain IIa of the HCV IRES RNA. Starting with screening of fragment-like molecules ligands by a mass-spectrometric method structure-activity relationship studies were performed by MS, too.[145] The optimization finally resulted in a ligand with submicromolar affinity and activity in replicon assays (Figure 1.17).

Further, the P4-P6 domain of the *Tetrahymena thermophila* group I intron was exploited in a specialized fragment screening approach.[99] The so-called SHAPES approach, a NMR-method used to detect binding of fragments derived from scaffolds most commonly found in known drug molecules, was applied.[146] Indeed several compounds interfered with the target RNA, but only one small molecule, methapyrilene, specifically interacted with the group I intron when counterscreened against a control RNA.[147]

Ligands for the antiviral target tRNA^{Lys3} were detected by screening a small focused fragment library with 1D-^{1}H-NMR.[148] The binding sites of the two initial hits were determined by 2D HMQC NMR with ^{15}N-labelled RNA and this revealed different binding sites in the T and D stem, respectively. Both fragments were linked resulting in a compound with higher affinity.[91] Additional linker optimization led to a compound with a K_D-value of 1.4 µM and selectivity for one of the binding sites (Figure 1.17).[149]

^{1}H-1D-NMR spectroscopy was used to screen fragment molecules for their binding affinity to the influenza A virus RNA promoter.[150] Seven primary hits were revealed with 6,7-dimethoxy-2-(1-piperazinyl)-4-quinazolinamine (DPQ, Figure 1.17) having the highest affinity (K_D = 50.5 µM) and exerting antiviral activity in cell-based assays as well. Binding of this compound was further substantiated by a structure determination via NMR (pdb-code: 2lwk).

Novel ligands for the telomeric RNA G-quadruplexes (TERRA) were discovered by screening 355 fluorinated fragments with ^{19}F-NMR spectroscopy.[96] Primary hits

were validated by further ^{19}F-NMR and STD experiments. The highest affinity ligand had a K_D-value of $\approx$ 120 µM (Figure 1.17).
Integration of FBDD into the development of novel riboswitch ligands is illustrated by studies with the *thiM* TPP riboswitch of *E. coli*.[151,152] Initially the fragments were screened using equilibrium dialysis with [^{3}H]thiamine as reporter ligand. The primary hits were then confirmed by ligand-detected NMR methods and binding affinities of 17 out of 20 compounds could be determined by ITC. The best fragments showed affinities towards the TPP riboswitch in the two-digit µM-range and crystal structures of four fragments bound to the RNA were determined (Figure 1.17).[153]

Chapter 2

Objectives

The aim of this work was to exploit the FMN riboswitch structure for the discovery of structurally diverse ligands by virtual screening. For several reasons it was considered as reasonable to address this RNA structure by a structure-based drug design approach. The FMN riboswitch belongs to the most abundant riboswitch structures in the genome of various pathogenic bacteria. Especially in Gram-positive pathogenic bacteria, such as *Clostridium difficile*, *Staphylococcus aureus* and *Streptococcus pneumoniae* often several genes are regulated by one single FMN riboswitch.[55] Any compound binding to the FMN riboswitch aptamer domain with high affinity and inducing the same conformational changes within the expression platform as FMN would affect the expression of several proteins involved in riboflavin uptake and biosynthesis in the bacterial cell. In a situation of a high availability of riboflavin the FMN riboswitch function is desired for the bacteria. However, a potent ligand mimicking FMN and competing with FMN for aptamer binding, would maintain the suppression of gene expression even in the absence of sufficient riboflavin levels in the bacterial cell. A total depletion of riboflavin levels promises bactericidal effects and the FMN riboswitch aptamer domain therefore represents a promising drug target. The large openings to the binding site indicate the accessibility of the binding site. The globally preformed binding site in the absence of a ligand and despite this inherent plasticity to accommodate for different ligands emphasizes that this RNA structure is capable of possibly interacting with ligands other than FMN.[39,64] Validation with compounds like roseoflavin[58], 5-FDQD[60] and ribocil[61] which exert antibiotic activity via interaction with the FMN riboswitch underlines that it is a possible drug target and especially ribocil with its drug-like properties (Figure 1.12) proves the druggability of this target RNA. Particularly, the high sequence conservation among the binding site residues directly involved in interactions to the cognate ligand, leads to the assumption that antibacterial compounds interfering with the FMN riboswitch have a potential for broad-spectrum activity.[64]

In contrast to the prediction of a reasonable binding pose, the accurate prediction of a binding affinity by molecular docking is beyond the scope of almost all docking programs.[154] Hence, it was the intention to measure binding affinities of virtually binding ligands during an adequate experimental validation. More than one detection principle should be applied to confirm the binding affinity and avoid being

misled by assay interference and artefacts. Initial virtual screening hits usually exhibit affinities ranging from 1-500 µM.[155] Thus, robust assays able to even detect weak affinity ligands are necessary to identify and confirm novel ligands. Due to the presumably low affinities of initial screening hits indirect measurements with a displacement of FMN from the binding site were employed first. For this purpose a known method based on the inherent fluorescence of FMN was adapted to be applied as competitive assay.[39] As an orthogonal method less prone to interference than fluorescence intensity measurements an ITC displacement assay was developed. Aiming to directly measure the interactions between riboswitch RNA and small molecules without the presence of FMN a SPR-based method was developed. Finally, it was attempted to cocrystallize the identified novel ligands to determine their binding modes.

Chapter 3

Materials and Methods

3.1 Virtual ligand screening / molecular docking

3.1.1 Receptor preparation and sphere set generation

The coordinates of the receptor molecule serving as input for the virtual screening approach were obtained from the PDB (`http://rcsb.org/pdb/home/home.do`).[109] First, all polar and non-polar hydrogen atoms were added to the pdb-file 3f2q[39] containing the three-dimensional structure of the FMN riboswitch bound to its cognate ligand flavin mononucleotide (FMN) using the program Moloc[156] (Gerber Molecular Design). In the presence of the reference ligand the positions of these hydrogen atoms were energy minimized using the MAB all atom force field as implemented in Moloc while all other atoms were kept rigid. After the energy minimization step all atoms that were not part of the final receptor setup, including Mg^{2+}-atoms and water atoms, were removed. Finally, only the RNA atoms were left to determine the receptor and the binding site. A sphere set was subsequently used to define the ligand binding site (Figure 3.1). It was based on a cubic grid enveloping the coordinates of the original ligand. The sphere set serves two purposes during the docking procedure:

1. it defines the binding site and corresponds to the area within the binding site with a reduced dielectric constant ε to calculate the electrostatic potential of the binding site

2. during ligand docking spheres are matching points for ligand atoms to help generating binding modes

Hence, it is possible to use different sized sphere sets for receptor preparation and docking. In the case of the FMN riboswitch the sphere set for the docking calculations was smaller compared to the one defining the binding site. The smaller sphere set was chosen to force placement close to A99, since the binding site was rather large and it would have been difficult to retrieve hits addressing both the binding region of the heterocycle and the binding region of the ribityl moiety and the phosphate group of FMN at the same time.

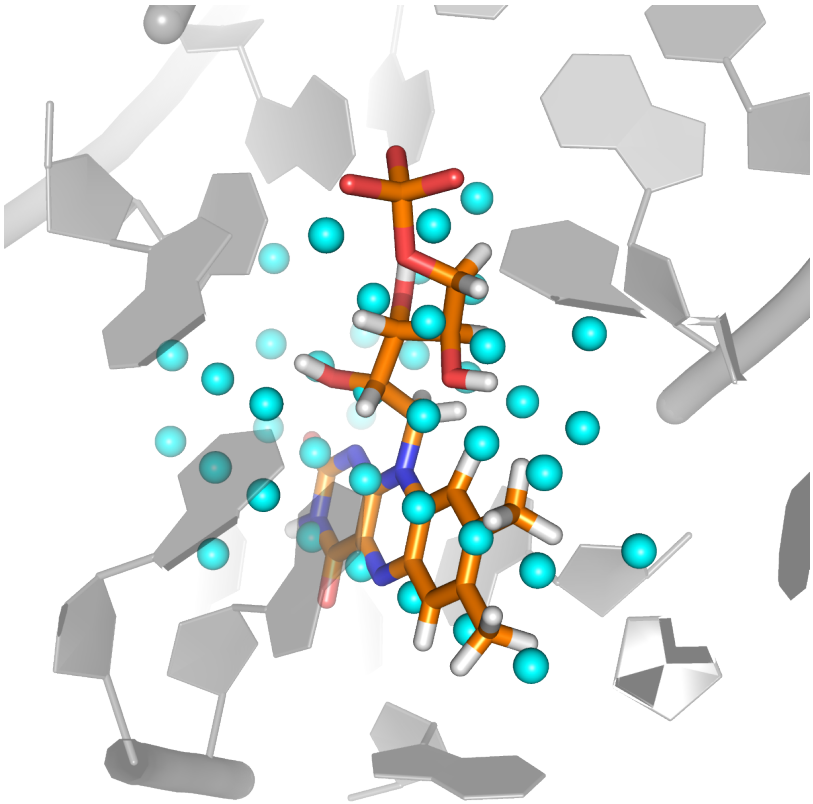


FIGURE 3.1: This is an example of a sphere set generated for the determination of the binding site. FMN as reference ligand is depicted with orange carbon atoms. Cyan-coloured spheres are arranged on a simple cubic lattice around the ligand and fill out the binding site.

The distance of neighboring points in the sphere set during receptor preparation was set to 2.4 Å and they extended up to 4 Å around the original ligand atoms. The minimum distance of a sphere set point to any receptor atom was 1.7 Å. For docking against the FMN riboswitch the smaller sphere set was built around the isoalloxazine heterocycle of FMN with grid points lying up to 2.5 Å around the ligand atoms and a distance of 1.5 Å between neighboring grid points (Figure 3.2).
Partial charges of all receptor atoms kept for docking were assigned using the AMBER force field parameters[157] and were modified to result in a net charge of zero per nucleotide, thereby mimicking the presence of neutralizing counter ions.[124] The DISTMAP utility of DOCK was used to generate the contact map of a grid box wrapping the sphere points to produce precalculated "scores". During docking, each ligand atom received the score of the nearest point and the total contact score reflected the shape complementarity of receptor and ligand.
DelPhi[158] was used to calculate the electrostatic map of the receptor atoms, which is necessary to obtain the electrostatic energy term during docking. An internal dielectric constant $\varepsilon = 2$ was used, whereas outside of the binding site this value was kept at $\varepsilon = 78$, reflecting the dielectric constant of water.

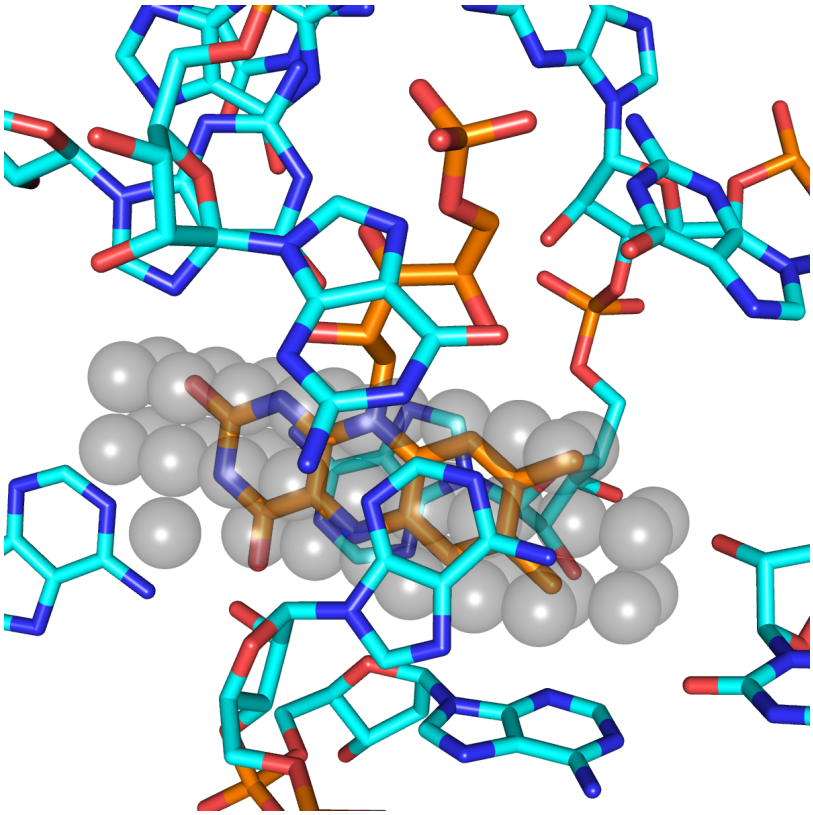


FIGURE 3.2: Grey spheres depict the sphere points used as matching points during docking. Binding site residues of the nucleic acid are displayed with cyan carbon atoms and the ligand is represented with orange carbon atoms.

The vdW energy map was calculated using the DOCK utility CHEMGRID[159] and SOLVMAP served for the calculation of a map for the solvent-excluded volume. Both energy maps were necessary to calculate the vdW and the desolvation energy terms of the scoring function.
The initial receptor setup was modified after testing a training set of compounds (3.1.2) to account for the most crucial hydrogen bond between FMN and the riboswitch RNA. For this purpose the partial charges of the exocyclic N6-amino group of adenine 99 (A99) were adjusted to generate a more polarized N-H-bond facilitating hydrogen bonding. While keeping the net charge of A99 constant, the partial charge of the N6-atom was reduced by 0.3 units and the partial charges of both corresponding hydrogens were increased by 0.15 units each.

3.1.2 Training set of compounds for testing the receptor setup

First a redocking of FMN was performed to test the suitability of the receptor setup and determine a rmsd-value between the docked and the crystallographically determined binding mode of the ligand to the RNA structure. Redocking was also performed with riboflavin and roseoflavin as ligands. The rmsd-values between the experimentally determined binding mode of the co-crystallized ligand and the docked pose were calculated using the python script rmsd.py (`https://github.com/`

`rasbt/protein-science/tree/master/scripts-and-tools/rmsd`). All heavy atoms were aligned for the calculation of rmsd-values. The receptor setup was further validated by comparing the performance of known active compounds such as riboflavin, roseoflavin and lumiflavin and manually created flavin analogues lacking the uracil-like edge of the isoalloxazine heterocycle crucial for the hydrogen bonding pattern to A99 (Table 4.1). Discrimination between ligands and non-binders by correct receptor preparations was subsequently tested carrying out initial docking studies on this small training set.

3.1.3 Small molecule preparation

3.1.3.1 In-house database of purchasable compounds

An in-house database of commercially available screening compounds from initially 16 different suppliers (Table 3.1) was assembled. After obtaining the compounds from the different suppliers in sd-format, they were uploaded in a MySQL database in SMILES format.[160] Duplicates were removed and thus 4,777,515 unique compounds remained as starting point for the docking procedure. Simultaneously, physicochemical properties (molecular weight, number of heavy atoms etc.) were calculated[161] and uploaded together with the smiles codes using in-house scripts based on the OEToolkit (Openeye, Santa Fe, NM). In the same procedure unwanted groups were identified and assigned to the screening compounds. The SlogP-value was calculated within the program MOE v.2012.10 (Chemical Computing Group, Montreal, QC, Canada).

TABLE 3.1: Overview of the suppliers and the corresponding compound numbers that were used to generate the in-house database of commercially available screening compounds.

Supplier	**Number of compounds per supplier**
Apollo Scientific	14,439
Asinex Ltd.	427,492
ChemBridge Corporation	446,637
Chemdiv	1,313,948
Enamine	1,290,053
InterBioScreen Ltd. (IBS)	503,339
Key Organics Ltd.	54,746
Life Chemicals Inc.	199,941
Maybridge	54,318
Otava Ltd.	254,617
Peakdale	14,644
Sigma Aldrich	10,000
Specs	197,486
TimTec LLS	917,293
TOSLab	17,584
Vitas-M Laboratory Ltd.	1,070,933

3.1.3.2 Compilation of a ligand database suitable for pharmacophore searching and docking

The protocol for the preparation of the possible ligands for docking was similar to the one described previously[162]. First, the screening subset was assembled by filtering for compounds with desired drug-like properties (Table 4.2).
For the remaining molecules possible protomers, tautomeric states and stereoisomers were calculated using in-house scripts based on the OEToolkit (Openeye, Santa Fe, NM). This led to a hierarchical setup of the compound database, since starting from one entry different protomers were built, followed by the calculation of different tautomerization states starting from the individual protomers and finally followed by the calculation of stereoisomers starting from individual tautomers.
Subsequently, the filtered compounds were subjected to a pharmacophore search described later (3.1.4) to further reduce the number of molecules during the docking procedure. Three-dimensional coordinates and low energy conformations of the small molecules were built using OMEGA2[163] (Openeye, Santa Fe, NM). Partial charges and desolvation energies for the ligands were calculated with the program AMSOL[115,164] (`http//comp.chem.umn.edu/amsol/`). For the bulk solvent the dielectric constant of water was assumed and for the receptor environment the dielectric constant of hexadecane (ε = 2.06) was used during the calculation of the desolvation energy terms as it was done previously.[115]
Finally, the multiple conformers were aligned on their ring systems and stored in a hierarchical format (flexibase format) together with the information calculated by AMSOL. This step was carried out with the DOCK utility mol2db.

3.1.4 Pharmacophore search

After the in-house database was filtered for desired molecular properties, a pharmacophore search was applied as second filter step. The pharmacophore model was based on the binding mode of FMN to the riboswitch RNA derived from the same crystal structure as used for docking. This is a common in structure-based design approaches as reviewed here[165]. Development of the pharmacophore model and the pharmacophore search were carried out using MOE v.2012.10 (Chemical Computing Group, Montreal, QC, Canada).
The SMILES strings of all the stereoisomers passing the first filter step were converted into a three-dimensional format (sd-file), whereby multiple low energy conformers of the compounds were calculated using OMEGA2 (Openeye, Santa Fe, NM). Only the molecules with at least one conformation meeting the criteria of the pharmacophore search were kept for docking.

3.1.5 Molecular docking

Receptor preparation and validation of the system were initially carried out using the DOCK version 3.6. The whole dataset was docked using the former version DOCK 3.5.54 due to its higher robustness when screening large datasets of molecules. Both versions are very similar and highly comparable except for slight differences in the calculation of the desolvation energy map. After the desolvation energy map had been recalculated with the SOLVMAP utility of the former DOCK version, docking was performed using DOCK 3.5.54[114,115,166] with the following sampling parameters:

- receptor and ligand binsize were set to 0.5 Å
- receptor and ligand overlap binsize were set to 0.4 Å
- the distance tolerance of matching pairs of receptor and ligand spheres was set to 1.2 Å
- the number of energy minimizations for the docked molecules was 20

Only the best-scoring representation of the docked molecules in terms of the most favourable docking pose with a certain protonation and tautomerization state was stored in the resulting files. The scoring function used for ranking the docked ligands according to their predicted binding energies contained terms for electrostatic and vdW interaction energies as well as a correction for desolvation energies:[115]

$$E = E_{elec} + E_{vdW} + \Delta G_{solv} \tag{3.1}$$

where E_{elec} represents the electrostatic interaction energy term, E_{vdW} the van der Waals contribution, ΔG_{solv} displays the desolvation energy for the ligand when moving from the bulk solvent to the binding site.

3.1.6 Docking analysis and molecular graphics

The docking poses of the successfully docked compounds were stored in a MOE database (proprietary format .mdb) together with their corresponding scores in kJ/mol. Predicted binding modes were then visualized and filtered with the pharmacophore model while keeping the pose fixed. All docking poses were ranked according to their total scores as predictor of the ΔG-value for the interaction between ligand and receptor and as a second step ranked by comparing the ligand efficiency (LE) using the following formula:

$$\mathrm{LE} = -1 * \left(\frac{\mathrm{Score}}{\mathrm{HA}} \right) \tag{3.2}$$

The absolute score for the predicted pose is divided through the heavy atom count (HA) in the particular molecule.

All diagrams of molecular structures were generated in PyMOL[167] (`https//pymol.org`) and two-dimensional chemical depictions were generated using Accelrys Draw 4.1 (Biovia, San Diego, CA).

3.2 Virtual screening / substructure search

In order to focus on structures that fulfill the minimal requirements to interact with the key residues of the riboswitch RNA via hydrogen bonding in the same manner as the uracil-like edge of the isoalloxazine heterocycle a substructure search was conducted with the same receptor setup for docking as the one used previously in the whole database screening (3.1.3.2). Queries with SMARTS strings - a line notation for substructural patterns in molecules similar to SMILES codes but with the ability to define wildcard atoms and bonds - were used to identify the substructures (Figure 4.5).
Additionally, core fragments were extracted from the substructures (Figure 3.3) by dividing the structures into rings plus the directly attached polar group, whereas all halogen atoms or carbon atoms that are not part of a polar functional group or ring system were discarded (Figure 3.3).[161] The main core fragments were evaluated together with the docking results by visually inspecting the "Top-Scorer" per main core fragment.

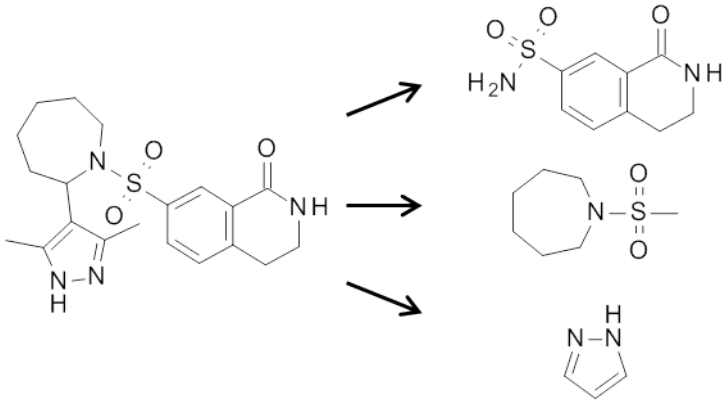


FIGURE 3.3: Example of core fragment extraction. In this example the compound structure is divided into three core fragments. Polar functional groups like the sulfur-containing function are added to both adjacent rings.

3.3 Synthesis and purification of the FMN riboswitch aptamer domain

For all transformations into *E. coli*-cells, plasmid amplifications and protein expressions described in the following subsections LB(-Agar) medium with the following composition (LB-Lennox) was used for growing bacteria:

- Tryptone 10 g/l
- Yeast extract 5 g/l
- NaCl 5 g/l
- (Agar) 15 g/l

3.3.1 Expression and purification of the T7 RNA polymerase

Generally, the RNA synthesis steps described in this chapter were performed by *in vitro* transcription using T7 RNA polymerase from commercially available sources (New England Biolabs, Ipswich, MA) for small-scale transcriptions or produced by recombinant protein expression for large-scale transcriptions.
The plasmid DNA of a pET vector system with ampicillin resistance gene, T7 tag and N-terminal His-tag containing the T7 RNA polymerase sequence information was generously supplied by the research group of Prof. Dr. Mark Helm (Johannes Gutenberg University, Mainz, Germany). According to the protocol of the manufacturer, the plasmid was transformed into competent *E. coli* Rosetta (DE3) pLysS cells (Merck Millipore). Clones were selected on LB agar plates supplemented with 100 µg/ml ampicillin and 30 µg/ml chloramphenicol and a single colony was grown in an 5-ml overnight bacterial culture at 37 °C in LB medium containing the antibiotics in the same concentrations. The plasmid DNA was purified afterwards using a Miniprep kit (Qiagen, Hilden, Germany) and the correct sequence was verified by DNA sequencing (GATC Biotech, Konstanz, Germany).
The purification of the His-tagged T7 RNA polymerase was carried out according to a previously published protocol.[168] Shortly, a 25-ml preculture of LB medium containing ampicillin and chloramphenicol was inoculated first with a bacterial glycerol stock overnight at 37 °C. The preculture was used to inoculate 3 L of LB medium at 37 °C and RNA polymerase expression was induced by addition of IPTG to a final concentration of 0.5 mM, after the OD_{600} was approximately 0.6. The bacterial culture was allowed to grow for additional 3-4 h at 37 °C before harvesting the cells. This was done by centrifuging the cells at 2,800 x *g*. Subsequently, the pellet was resuspended in 20 ml buffer A (Table 3.2). Cells were lysed by sonication in an ice bath with 25 pulses of 30 seconds and the cell lysate was centrifuged for 50 minutes at 4 °C and 15,000 x *g*. The supernatant was applied to a HisTrap column (GE Healthcare Life Sciences) on a Äkta pure FPLC system (GE Healthcare Life Sciences) with a flow rate of 1 ml/min. The protein fractions were collected by applying elution buffer B (Table 3.2) to the column. Fractions containing the protein of interest were pooled together and concentrated using 30 kDa molecular weight cut-off centrifugal filters (Amicon®, Merck Millipore). The purity was verified by running SDS-PAGE and the concentration was determined measuring the absorbance at 280 nm wavelength using a Nanodrop (Thermo Fisher Scientific, Waltham, MA). Protein aliquots were mixed in a 1:1 volume ratio with glycerol and stored at -20 °C for further use.

TABLE 3.2: The composition of the lysis buffer A and the elution buffer B for the purification of T7 RNA polymerase.

Lysis buffer A	**Elution buffer B**
50 mM Tris-HCl (pH 8,0)	50 mM Tris-HCl (pH 8,0)
100 mM NaCl	100 mM NaCl
5 mM β-mercaptoethanol	β-mercaptoethanol
5% glycerol	5% glycerol
1 mM imidazole	**300 mM** imidazole

3.3.2 Synthesis and purification of FMN riboswitch RNA for binding assays by *in vitro* transcription

The aptamer domain of the 112-mer *Fusobacterium nucleatum* FMN riboswitch for the binding assays was transcribed *in vitro* from the pUT7 plasmid template generously supplied by the laboratory of Alexander Serganov.[169] In this template the sequence of interest was imbedded by the T7 promoter sequence on the 5′ end and a hammerhead ribozyme sequence on the 3′ end to avoid heterogeneity of the 3′ end of the RNA product, followed by a *Hind*III recognition site to linearize the plasmid (Figure 3.4).

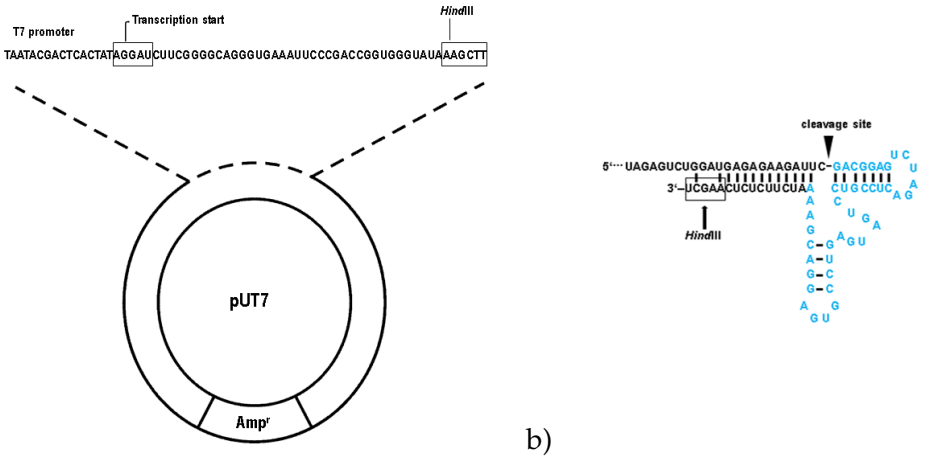


FIGURE 3.4: a) The pUT7 plasmid used for transcription of the riboswitch with indications for the T7 promoter and *HindIII* restriction site. b) The RNA after *in vitro* transcription of the *Hind*III linearized plasmid with the hammerhead ribozyme joined to the 3′ end. The hammerhead ribozyme sequence is indicated with cyan-coloured bases. A box displays the former *Hind*III restriction site.

3.3.2.1 DNA template preparation

The DNA template preparation was performed according to the protocol of Pikovskaya et al. with slight modifications.[169] In the following paragraphs the most important steps are described. First, the plasmid DNA was transformed into the chloramphenicol-resistant *E. coli* strain XL10-Gold® (Agilent, Santa Clara, CA) according

to the protocol of the manufacturer and clones were selected on LB agar plates supplemented with 100 µg/ml ampicilin and 30 µg/ml chloramphenicol. Afterwards, 5-ml bacterial cultures in LB medium containing ampicillin and chloramphenicol were grown overnight and the plasmid DNA was purified using a Miniprep kit (Qiagen, Hilden, Germany) and the correct sequence was verified by DNA sequencing (GATC Biotech, Konstanz, Germany).
For large-scale synthesis of *F. nucleatum* FMN riboswitch RNA bacteria were grown in 3 L of LB medium supplemented with 100 µg/ml ampicilin and 30 µg/ml chloramphenicol at 37 °C overnight and typically $\sim$ 10 mg purified plasmid DNA template was obtained using a Gigaprep kit (Qiagen, Hilden, Germany). Subsequently, the DNA template was linearized for *in vitro* transcription by restricting 10 mg of plasmid DNA with 2,000 U of *Hind*III in 10 ml of restriction enzyme buffer at 37 °C for $\sim$ 4 h. The completion of the reaction was verified by 1% agarose gel electrophoresis.
For removal of the restriction enzyme a phenol/chloroform extraction was performed by shaking the restricted DNA in the presence of 10 ml of a 1:1 (V/V) mixture of phenol and chloroform in 50-ml tubes and then centrifuging at 15,000 x *g* for 2 min. The upper aqueous phase was collected into fresh tubes and the extraction was repeated with chloroform alone. Finally, the plasmid DNA was precipitated with ethanol in 50-ml tubes. The aqueous phase was mixed with 1/10 volume of 3 M sodium acetate (pH 5.2) and 3 volumes of pure ethanol. The mixture was chilled at -20 °C overnight and centrifuged at 14,000 x *g* for 30 minutes at 4 °C. After discarding the supernatant, the pellet was washed with $\sim$ 10 ml of 70% ethanol and centrifuged again for 30 minutes at the same centrifugal force. The pellet was air-dried and dissolved in a suitable volume of autoclaved water (typically to concentrations of $\sim$ 2.5 mg/ml).

3.3.2.2 *In vitro* transcription

Large-scale *in vitro* transcriptions typically were performed in a final volume of 25 ml. The general procedure has been described several times in the literature[169,170] and was adapted to the available equipment in the laboratory. The reaction mixture contained 100 mM Tris-HCl (pH 7.9), 30 mM DTT, 2 mM spermidine, 0.01% Triton X-100, 20 mM $MgCl_2$, 4 mM of each NTP (Sigma aldrich), 100 µg/ml DNA template from the restriction digest described above and 0.05 mg/ml T7 RNA polymerase (prepared in-house). Transcriptions were allowed to proceed 4h at 37 °C and incubated for another 45 min with $MgCl_2$ adjusted to a concentration of 50 mM to ensure efficient ribozyme cleavage from the RNA.[169] An estimation of the transcription yield and cleavage efficiency was obtained by small-scale denaturing PAGE electrophoresis. Hereupon, the finished transcription reactions were stored at -20 °C overnight.
The transcription reactions were thawed and magnesium pyrophosphate precipitated during the reaction was removed by centrifuging the mixture twice at

4,000 x *g* for $\sim$ 20 min and filtering the isolated supernatant through 0.45-µm and 0.2-µm disposable filtration units. Afterwards, the mixture was concentrated by 10 kDa molecular weight cut-off centrifugal filter units (Amicon®, Merck Millipore) to final volumes of 2-4 ml. The concentrated transcription reactions were subsequently purified by denaturing PAGE on gels containing 10% polyacrylamide (19:1), 44.5 mM Tris-borate (pH 8.3), 1mM EDTA and 8 M urea. UV shadowing at λ = 260 nm was used to visualize the corresponding product bands[171], the product band was cut out and the collected gel slices of 4 parallel gels were transferred into a sterile, plastic syringe and forced through the syringe tip into a disposable 50-ml centrifuge tube. Thereafter, the RNA was electro-eluted with a Model 422 Electro-Eluter (Bio-Rad, Hercules, CA) using 3 kDa molecular weight cut-off membrane caps in a buffer containing 89 mM Tris-borate (pH 8.3) and 2 mM EDTA for 4-5 h. The eluate was buffer-exchanged three times into RNA storage buffer (10 mM K-HEPES (pH 7.5) and 0.1% sodium azide) and concentrated to volumes $\leq$ 500 µl using 10 kDa molecular weight cut-off centrifugal filter units (Amicon®, Merck Millipore). Finally, the RNA was quantified by UV absorption calculating the molar concentrations with a molar extinction coefficient $\varepsilon_{260\text{nm,FMN-RS}}$ = 1,071,400 $\frac{1}{\text{M} \times \text{cm}}$.

3.3.3 Synthesis and purification of FMN riboswitch RNA for crystallization and SPR by *in vitro* transcription

For crystallization as well as for labeling the riboswitch RNA with a biotin tag to perform certain SPR experiments, the RNA was synthesized by annealing two separate strands (referred to as strand A/B in this thesis) engineered to ensure crystal contacts as previously reported:[39,61]

Strand A (54 nt): 5′-GGAUCUUCGGGGCAGGGUGAAAUUCCCGACCGGUGGU AUAGUCCACGAAAGCUU-3′

Strand B (56 nt): 5′-GCUUUGAUUUGGUGAAAUUCCAAAACCGACAGUAGAG UCUGGAUGAGAGAAGAUUC-3′

The *in vitro* transcribed RNA was produced from a single-stranded synthetic DNA template (IBA Lifescience, Göttingen, Germany) where only the promoter sequence is double-stranded.[172] To circumvent the addition of one or more nontemplated nucleotides to the 3′ end of the transcribed RNA two C2′-methoxy modifications at the 5′ end of the DNA template were incorporated.[173] Finally, a template strand containing the "bottom" strand of the T7 promoter in 3′ direction to the desired RNA sequence was annealed to the "top" strand DNA bearing the T7 promoter (Figure 3.5) and transcription by T7 RNA polymerase was initiated.

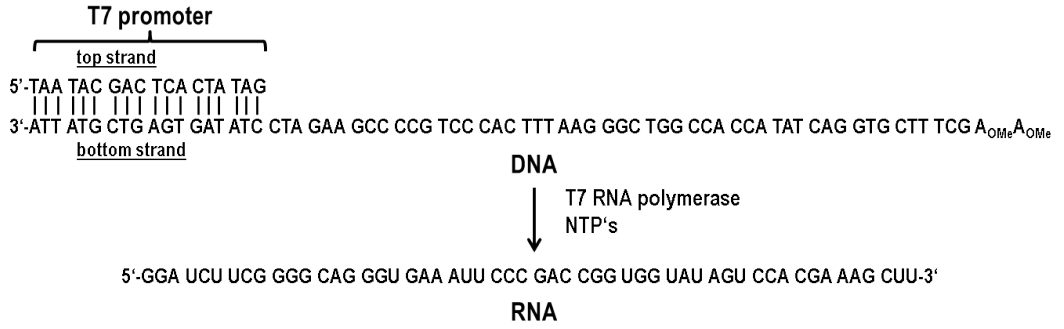


FIGURE 3.5: Transcription from a synthetic DNA template exemplified for strand A. The DNA with a sequence encoding the RNA for strand A fused to the T7 promoter having two methylated adenine residues at the 5′ end is annealed to a DNA complementary to the promoter sequence.

The lyophilized top and bottom DNA strands were dissolved in autoclaved water resulting in concentrations of 100 µM. In order to produce the transcription template for strand A the top strand was mixed with the bottom strand A in equimolar amounts. The transcription template for strand B was prepared by mixing equimolar amounts of the same top strand with bottom strand B. Both mixtures were heated to 95 °C for 1 minute and snap cooled on ice to anneal the DNA strands.[173]

Top strand: 5′-TAATACGACTCACTATAG-3′

Bottom strand A: 5′- 2′-O-Methyl-A-2′-O-Methyl-A-GCTTTCGTGGACTATACCAC CGGTCGGGAATTTCACCCTGCCCCGAAGATCCTATAGTGAG TCGTATTA-3′

Bottom strand B: 5′- 2′-O-Methyl-G-2′-O-Methyl-A-ATCTTCTCTCATCCAGACTC TACTGTCGGTTTTGGAATTTCACCAAATCAAAGCTATAGTGA GTCGTATTA-3′

The transcription reaction itself typically was carried out in a final volume of 12.5 ml containing 100 mM Tris-HCl (pH 7.9), 30 mM DTT, 2 mM spermidine, 0.01% Triton-X 100, 20 mM $MgCl_2$, 4 mM of each NTP (Sigma aldrich), 100 mg/ml PEG8000, 0.05 mg/ml T7 RNA polymerase (prepared in-house) and 400 nM of the particular annealed template as previously described.[169,170] One modification was the addition of PEG8000 to increase the transcription efficiency.[174] Subsequently, the reactions were incubated for at least 2 h at 37 °C and precipitated magnesium pyrophosphate was removed by centrifugation and filtration as described in the previous section (3.3.2.2). The transcription mixtures were concentrated by spinning down the samples with 3 kDa molecular weight cut-off centrifugal filter units (Amicon®, Merck Millipore) to volumes of $\sim$ 3 ml.[169]

Further processing of the *in vitro* transcription mixture was performed as described by Golden et al.[173] First, the transcription reactions were purified by denaturing PAGE on gels containing 10% polyacrylamide (19:1), 44.5 mM Tris-borate (pH 8.3),

1 mM EDTA and 8 M urea. Again, the RNA was visualized by UV shadowing and the product bands were cut out using a clean scalpel. The gel slices were transferred to a plastic syringe and forced through a syringe tip into a disposable 50-ml centrifuge tube. The crushed gel matrices were subjected to a freeze-thaw procedure by placing them into the -80 °C freezer for approximately 30 minutes and thawing them back to room temperature on the bench. Extraction of the RNA was performed by gently rocking the crushed gels overnight at 4 °C in a buffer containing 10 mM Tris-HCl (pH 7.5), 1 mM EDTA and 250 mM NaCl (TEN buffer). To recover the RNA, the RNA extracts were filtered through a sterile 0.2-µm filtration unit and concentrated firstly by using 3 kDa molecular weight cut-off centrifugal filter units (Amicon®, Merck Millipore) and secondly by ethanol precipitation overnight at -20 °C after addition of 1/10 volume of 3 M sodium acetate (pH 5.2) and 3 volumes of ethanol. The RNA was pelleted as described above (3.3.2.2). Both pelleted RNA strands were resuspended in a buffer containing 10 mM K-HEPES (pH 7.5) and 0.1% sodium azide (RNA storage buffer) for long-term storage in a typical volume of 500 µl. The concentrations of the RNA strands were measured by UV absorbance with the following molar extinction coefficients: $\varepsilon_{260\text{nm,strand A}}$= 508,300 $\frac{1}{\text{M} \times \text{cm}}$ and $\varepsilon_{260\text{nm,strand B}}$= 543,400 $\frac{1}{\text{M} \times \text{cm}}$.

3.3.4 Synthesis and purification of the FMN riboswitch RNA with a 36-nt long insert for SPR measurements

Besides the different methods described in the following section (3.4) to directly biotinylate the FMN riboswitch RNA and immobilize the RNA on a streptavidin-coated sensor chip, an approach via a DNA-RNA hybrid was also tested. This had been previously described in a publication about SPR measurements with the AdoCbl riboswitch.[175]

For sufficient base pairing at least 25 nucleotides were necessary that interact with a complementary 5'-end biotinylated DNA oligonucleotide that finally was attached to the sensor chip. The aptamer domain of the FMN ribowitch was elongated on the basis of the approach described for the AdoCbl riboswitch[175] to build a template for hybridization of the RNA with the DNA oligonucleotide without perturbing the global folding of the RNA. The inserted sequence was based on the sequence of the expression platform naturally occurring in the *F. nucleatum* FMN riboswitch and some base positions were slightly modified. Finally, the definite sequence of the RNA to be inserted directly following the 3'-end of the 112-nt aptamer domain was chosen by the help of the RNAfold web server[176] (`rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi`). Hereby, it was predicted, whether the inserted sequence would change the global fold in the aptamer domain and would influence FMN binding to the RNA. The fold prediction with the lowest energy was compared with the fold prediction for the wild type aptamer domain. The following sequence was selected as insert:

5′-GCACAUUGACAUUUUUAACUUGGAGUAAACAUUUUC-3′

Starting point for the purification of the elongated aptamer domain of the FMN riboswitch was a pUC19 plasmid carrying the T7 RNA polymerase promoter followed by the aptamer domain, the 36-nt insert and the hammerhead ribozyme sequence. The sequence was embedded by a 5′ SmaI restriction site and a 3′ HindIII restriction site. The plasmid carrying the sequence information was purchased from GenScript (GenScript, Piscataway, NJ). Transformation into competent *E. coli* cells, amplification of the plasmid DNA, *in vitro* transcription, denaturing polyacrylamide gel electrophoresis and electroelution were carried out in exactly the same fashion as described in a previous subsection (3.3.2). The concentration of the purified elongated riboswitch RNA was determined by measuring the UV absorbance at 260 nm with a molar extinction coefficient ε = 1,400,300 $\frac{1}{\mathrm{M} \times \mathrm{cm}}$.
In order to hybridize the last 25 nucleotides of the 3′ end of the FMN riboswitch RNA to an DNA oligonucleotide carrying a biotin tag, the appropriate 5′-end biotinylated oligonucleotide was purchased from Eurofins Scientific (Eurofins, Luxemburg, Luxemburg). The PAGE-purified ordered product had the following sequence:

5′-Biotin-GAAAATGTTTACTCCAAGTTAAAAA-3′

3.4 Biotinylation of the *F. nucleatum* FMN riboswitch RNA

3.4.1 Biotinylation of the 3′ end of the FMN riboswitch RNA by periodate-chemistry

The RNA structures that were biotinylated for SPR experiments were labelled by selective periodate-mediated oxidation of the RNA's 3′-terminal ribose *cis*-diol.[177,178] Both the full-length *F. nucleatum* FMN riboswitch RNA as well as strand B of the two-piece RNA construct used for crystallization were biotinylated at the 3′ end.
Dephosphorylation of the 2′,3′-cyclic phosphate at the 3′ end of the full-length aptamer domain was accomplished by reaction with T4 polynucleotide kinase (New England Biolabs, Ipswich, MA) resulting in a *cis*-diol at the 3′ end under previously published conditions.[178] Dephosphorylation reactions were typically carried out in 50-µl aliquots with the following concentrations:

- 100 mM imidazole/HCl pH 6.0
- 10 mM $MgCl_2$
- 0.1 mM ATP
- 20 µg/ml BSA
- 500 pmol RNA template
- 10 U T4 polynucleotide kinase

This reaction was incubated for 6 h at 37 °C while gently mixing. Afterwards, the 3'-dephosphorylated RNA was extracted from the mixture by vigorously shaking it with an equal volume of phenol-chloroform-isoamyl alcohol (25:24:1) saturated with TE buffer at pH 8.0 (Sigma aldrich). Phase separation was accomplished by centrifuging at 15,000 x *g* for 2 minutes. The aqueous layer was extracted a second time with chloroform-isoamyl alcohol mixture (24:1, Sigma aldrich) alone by vigorously shaking both layers and separating them again by centrifugation. Finally, the extracted RNA was purified by desalting on a 7 kDa molecular weight cut-off Zeba® Spin Desalting column (Thermo Fisher Scientific, Waltham, MA) of 5-ml size for larger volumes ($\geq$ 500 µl and $\leq$ 2 ml) or by desalting on 6 kDa molecular weight cut-off Bio-Spin® (Bio-Rad, Hercules, CA) for small volumes ($\leq$ 100 µl) into a buffer containing 10 mM K-HEPES (pH 7.5) and 0.1% sodium azide (RNA storage buffer). The extracted and purified RNA was concentrated by ethanol precipitation. For this purpose 1/10 volume of 3M sodium acetate (pH 5.2) and 2.5 volumes of ethanol were added to the RNA solution. The mixture was put in the -20 °C freezer for 30 minutes and spun at 4 °C and 14,000 x *g*. Afterwards, the pellet was washed by addition of 300 µl cold ethanol 70% (V/V) and performance of the same centrifugation step. The pellet of the dephosphorylated RNA was air-dried for 10 minutes at 37 °C and redissolved in a suitable volume of autoclaved water.
Both the full-length aptamer domain as well as strand B of the two-piece riboswitch construct used for crystallization now were accessible for the oxidation step with sodium *meta*periodate by having a *cis*-diol at the 3' end. The reaction conditions taken from[177] during the oxidation step were as follows (final concentrations are indicated):

- 5 µM RNA template
- 0.1 M sodium acetate (pH 5.3)
- 2.5 mM sodium *meta*periodate

Incubation of the reaction mixture proceeded for 50 minutes on ice. The dialdehyde-functionalized RNA products were concentrated by ethanol precipitation under the same conditions as described in the previous paragraph for the dephosphorylated FMN riboswitch RNA aptamer domain.
The pellet of each of the purified, 3'-oxidized RNA species was dissolved in 100 µl of a 10 mM aqueous solution of EZ-Link® Hydrazide-PEG-Biotin (Thermo Fisher Scientific, Waltham, MA) and incubated at 37 °C for 2 h while gently shaking the mixture. Subsequently, the hydrazone produced in the former reaction was further stabilized by reduction with sodium borohydride. To the reaction mixture 100 µl of 0.2 M sodium borohydride and 200 µl of 1 M Tris-HCl (pH 8.2) were added and incubation on ice was carried out for 30 minutes. Immediately, unreacted biotinylation reagent and excess sodium borohydride were removed by desalting into RNA storage buffer using 7 kDa molecular weight cut-off Zeba® Spin columns (Thermo

Fisher Scientific, Waltham, MA), after the total volume was adjusted to 800 µl with autoclaved water. The desalted, 3'-biotinylated RNA was concentrated by ethanol precipitation following the procedure as already described in this section. Finally, the air-dried RNA pellet was dissolved in autoclaved water typically resulting in concentrations of $\sim$ 40-50 µM for the full-length aptamer domain riboswitch RNA and $\sim$ 100 µM for strand B of the two-piece construct. The same molar extinction coefficients allowing the calculation of the molar RNA concentrations by measuring the UV absorption at 260 nm could be used as described in the previous sections (3.3.2.2, 3.3.3).

3.4.2 Biotinylation of the 3' end of the full-length *F. nucleatum* FMN riboswitch by enzymatic labeling with T4 RNA ligase

The full-length aptamer domain of the FMN riboswitch was alternatively labelled with a biotin tag at the 3' end by an enzymatic ligation of a biotinylated donor nucleotide and the RNA template as acceptor substrate with the help of T4 RNA ligase.[179] Dephosphorylation of the template RNA to provide an acceptor for the reaction without the cyclic-2',3' phosphate was accomplished in the same way as described in the previous section (3.4.1).[178] 500 pmol RNA template in a total volume of 50 µl was used in the dephosphorylation reaction.
3' end biotinylation was carried out using the commercially available Pierce® RNA 3' End Biotinylation Kit (Thermo Fisher Scientific, Waltham, MA) according to the manufacturer's protocol with slight adjustments. Concisely, the dephosphorylated riboswitch RNA template was heated in the presence of 25% DMSO to 85 °C for 4 minutes to relax secondary structures and ligation was performed overnight at 16 °C with the following final concentrations in a final volume of $\sim$ 100 µl:

- RNA template 1.67 µM
- T4 RNA ligase 1.33 U/µl
- biotinylated cytidine (bis)phosphate 33.3 µM

At the end of the ligation reaction the RNA ligase was extracted once by shaking vigorously for 1-2 minutes with an equal volume of chloroform-isoamyl alcohol (24:1, Sigma aldrich) and separating both phases by centrifugation at 15,000 x *g* for 2 minutes. For the removal of excess biotinylated nucleotide a desalting step was performed. By means of a 6 kDa molecular weight cut-off Bio-Spin® column (Bio-Rad, Hercules, CA) the biotinylated riboswitch RNA was desalted into RNA storage buffer. Subsequently, the RNA sample was concentrated by ethanol precipitation. The precipitation procedure was slightly different compared to the one described for the periodate-based labelling. To the desalted and biotinylated RNA 1/10 volume of 5 M NaCl, 1% (V/V) of a 20 mg/ml glycogen solution and 3 volumes of ethanol were added and the precipitation was allowed to proceed for 1 h at -20 °C.

The mixture was centrifuged for 30 minutes at 4 °C and 14,000 x *g* and the pellet was washed with 300 µl of ice-cold 70% ethanol. After discarding the supernatant, the pellet was air-dried at 37 °C for 10 minutes. Finally, the pellet of the 3'-biotinylated RNA was dissolved in 25 µl autoclaved water ($\sim$ 5 µM). The molar extinction coefficient for the calculation of the molar RNA concentrations using UV absorbance at 260 nm was $\varepsilon_{260\text{nm}}$ = 1,080,300 $\frac{1}{\text{M} \times \text{cm}}$.

3.4.3 Biotinylation of the full-length *F. nucleatum* FMN riboswitch by conjugation to the 5′ phosphate via carbodiimide reaction

For the full-length aptamer domain of the FMN riboswitch RNA an alternative approach was applied to selectively label the 5′ end with a biotin tag.[180] Similar to the periodate-based labeling of the 3′ end, EZ-Link® Hydrazide-PEG-Biotin (Thermo Fisher Scientific, Waltham, MA) served as biotinylation reagent, after activation of the 5′ phosphate group with EDC.
The biotinylation of the 5′ end was done in a similar manner as described previously.[180] First, an aliquot of 350 µl with a concentration of $\sim$ 85 µM of the unmodified riboswitch RNA in RNA storage buffer was subjected to a buffer exchange into a buffer suitable for the EDC-reaction. Amicon® Ultra-0.5 ml centrifugal devices with a 10 kDA molecular weight cut-off were used three times to exchange the RNA into a buffer containing 10 mM sodium phosphate (pH 7.2), 150 mM NaCl and 10 mM EDTA (EDC reaction buffer). The RNA concentration was adjusted with EDC reaction buffer, so that 7.5 µl were containing $\sim$ 100 µg unmodified RNA. Into a microcentrifuge tube 1.25 mg of EDC were weighed out and the 7.5-µl RNA solution prepared before was added. Immediately, 5 µl of the biotinylation reagent stock solution was added and the contents were mixed briefly. Finally, 20 µl of imidazole/HCl (pH 6.0) was added resulting in final concentrations for EDC and the biotinylation reagent of 20 mM and 40 mM, respectively. The reaction proceeded overnight at room temperature while gently shaking the mixture.
For the removal of excess EDC and non-reacted biotinylation reagent the reaction mixture was desalted into a buffer containing 10 mM K-HEPES (pH 7.5) and 0.1% sodium azide (RNA storage buffer) by using 6 kDA molecular weight cut-off Bio-Spin® columns (Bio-Rad, Hercules, CA), after the volume had been adjusted by the addition of an equal amount of autoclaved water to the reaction mixture. Subsequently, the biotinylated RNA was concentrated by ethanol precipitation performing the procedure, that was already described for the periodate-based biotinylation of the 3′ end (3.4.1). The air-dried pellet was redissolved in autoclaved water, typically resulting in concentrations of $\sim$ 50 µM. The molar extinction coefficient for the calculation of the RNA concentration by UV absorbance at 260 nm was unchanged compared to the unmodified full-length *F. nucleatum* FMN riboswitch aptamer domain.

3.4.4 Detection of the biotinylation and estimation of the biotinylation efficiency

Two complementary methods were used to detect the successful biotinylation of the different RNA species and to estimate the extent of the biotinylation efficiency. The first approach was based on a shift of the electrophoretic mobility of the RNA upon interaction with streptavidin tracked on a native agarose gel and the second approach was based on the detection of dot-blotted RNA using a commercially available kit (Biotin Chromogenic Detection Kit, Thermo Fisher Scientific, Waltham, MA).

3.4.4.1 Streptavidin retardation experiment

Streptavidin protein (Sigma aldrich) was dissolved in a buffer composed of 10 mM Tris-HCl (pH 7.4), 2.5 mM $MgCl_2$ and 100 mM NaCl. About 200 ng of RNA was incubated with different amounts of streptavidin ranging from 0.1 µg to 2 µg in the same buffer that was used for dissolving streptavidin. After 15 minutes of incubation the samples were analyzed on a native agarose gel containing 0.9% agarose, 89 mM Tris-borate (pH 8.3) and 2 mM EDTA. Biotinylated RNA without the addition of streptavidin and non-biotinylated RNA both served as control samples.[178]
Afterwards, the agarose gel was subjected to methylene blue staining. The gel was agitated for 15 minutes in a solution of methylene blue (0.1%) in 0.5 M sodium acetate (pH 5.3) and destained in deionized water for several hours.

3.4.4.2 Chromogenic detection of biotinylated RNA

The manufacturer's protocol accompanying the Biotin Chromogenic Detection Kit (Thermo Fisher Scientific, Waltham, MA) was slightly amended for the detection of the biotinylation efficiency of the riboswitch RNA samples. A positively charged nylon membrane (BrightStar®-Plus, Thermo Fisher Scientific, Waltham, MA) was hydrated in TE buffer (10 mM Tris-HCl (pH 7.4) and 1 mM EDTA) for at least 10 minutes. Biotinylated control RNA was obtained from the biotinylation kit used for the enzymatic labeling of the 3′ end (3.4.2). The control RNA (sequence: 5′ → 3′: UCCUGCUUCAACAGUGCUUGGACGGAAC-Biotin) was diluted in autoclaved water to a concentration of 1 ng/µl (referred to as 100%). Further dilutions of the control RNA were made in the same way (75%, 50%, 25%). Biotinylated riboswitch RNA as test RNA was also diluted in autoclaved water to a concentration of 1 ng/µl.
2 µl of each sample were spotted on the equilibrated nylon membrane and allowed to dry for 30-45 minutes. The RNA was immobilized by UV cross-linking. A handheld UV lamp at 254 nm was placed $\sim$ 0.5 cm above the membrane for 5 minutes. Finally, the detection procedure on the membrane using alkaline phosphatase-conjugated streptavidin cleaving its substrate and resulting in a insoluble purple precipitate was performed according to the manufacturer's protocol.

3.5 Binding assays

3.5.1 Fluorescence quenching assay

The fluorescence measurements were carried out on a Tecan Infinite® 200Pro multimode microplate reader (Tecan, Männedorf, Switzerland). Studies based on monitoring the fluorescence quenching of FMN upon interaction with the FMN riboswitch RNA have been described earlier.[39,49] Except for the respective concentrations used the protocol initially developed from Serganov et al. was utilized here.[39] The sample buffer consisted of 50 mM Tris-HCl (pH 7.4), 100 mM KCl and 2 mM $MgCl_2$. Emission spectra of the intrinsically fluorescent riboswitch ligand FMN were collected from 480 to 620 nm at an excitation wavelength of 450 nM. For the binding affinity measurements, the aptamer domain of the full-length *F. nucleatum* FMN riboswitch was titrated against 40 nM constant FMN ligand concentration. The FMN concentration was determined by measuring the UV absorbance at 450 nm, assuming a molar extinction coefficient of ε = 12,500 $\frac{1}{\mathrm{M} \times \mathrm{cm}}$.[181] Prior to the preparation of the dilution series, the RNA stock solution stored in 10 mM K-HEPES (pH 7.5) and sodium azide 0.1% was refolded by diluting the stock in 10X sample buffer and autoclaved water and heating the samples to $\sim$ 90 °C for 3 minutes and placing the mixture 10 minutes on ice directly afterwards.
Ten different concentrations of RNA prepared by a 1:1 dilution series starting from 5 or 10 µM (final assay concentrations) each were premixed with 40 nM FMN (final assay concentration) in the sample buffer mentioned above. Finally, a total volume of usually 80 µl solution was transferred to 96-well half-area black flat plates (Greiner Bio-One, Frickenhausen, Germany) and the mixture was equilibrated for 10 minutes to several hours. Then, the intensity of the fluorescence emission was measured 3-5 times at 530 nm with excitation at 450 nm at room temperature. Buffer solution and 40 nM FMN served as control wells for 100% and 0% fluorescence quenching, respectively. The buffer fluorescence was subtracted, the data were normalized to the free ligand fluorescence and the data were finally fitted to the following equation[39]:

$$F = 1 + (f - 1)\frac{(L_0 + R_0 + K_D) - \sqrt{(L_0 + R_0 + K_D)^2 - 4L_0R_0}}{2L_0} \tag{3.3}$$

where F is the normalized fluorescence intensity, L_0 and R_0 are the total ligand and RNA concentrations, K_D is the apparent dissociation constant and f is the residual fluorescence intensity at the saturated concentration of ligand. The evaluation of the results was done with the help of the program QtiPlot (Bucharest, Romania).
Hit compounds from the virtual screening approach applied to the FMN riboswitch were tested in competition experiments by adding test compounds to the mixture of RNA and FMN. Due to different solubilities of the screening hits, the concentrations used were ranging from 50 µM to 1 mM. All of the compounds were diluted 1:20 from a stock solution in DMSO when added to the RNA-ligand mixture, resulting in 5% final DMSO concentration. The dilution series of riboswitch RNA was prepared

in the same manner as described in the previous paragraph. In a final volume of typically 80 µl ten different FMN riboswitch RNA dilutions, 40 nM FMN and 5% (V/V) of the competing ligand were mixed together. In parallel control experiments were prepared with the same RNA and FMN concentrations and addition of 5% pure DMSO. Again, the fluorescence emission intensity at 530 nm with excitation at 450 nm was measured to determine the binding affinity constant K_{D}. A dissociation binding constant for the competing ligand was calculated from the apparent K_{D} of FMN in the absence and presence of the test compound using the following formula[182,183]:

$$K_D^{complig} = \frac{c_{complig}}{\left(\frac{K_{Dapp}^{FMN+lig}}{K_{Dapp}^{FMN}} - 1\right)} \tag{3.4}$$

where $K_D^{complig}$ is the dissociation constant for the competitor ligand binding to the target RNA, $c_{complig}$ is the concentration of the competitor ligand used in the assay and $K_{Dapp}^{FMN+lig}$ and K_{Dapp}^{FMN} are the dissociation constants for the FMN-riboswitch interaction in the presence and absence of competitor ligand, respectively.

3.5.2 Competition ITC measurements for the full-length *F. nucleatum* FMN riboswitch aptamer domain

ITC measurements were carried out at 30 °C, a differential power of 5 µcal/s and 1000 rpm stirring speed using the MicroCal® iTC$_{200}$ (GE Healthcare Lifesciences). For that purpose FMN was titrated against unmodified, full-length and *in vitro* transcribed FMN riboswitch RNA. The sample buffer contained 50 mM K-HEPES (pH 7.5), 100 mM KCl and 2 mM $MgCl_2$ (ITC exchange buffer). Thus, pH and concentrations of cations were kept constant in comparison to the fluorescence-based assay.
Initially, the riboswitch RNA solution stored in a buffer containing 10 mM K-HEPES (pH 7.5) and 0.1% sodium azide was prepared by refolding the sample via heating up to > 90 °C for 3 minutes and placing on ice for 10 minutes. After the refolding was completed, the RNA was dialyzed by five times buffer-exchanging into the sample buffer using 15-ml centrifugal filter units (Amicon®, Merck Millipore) with a 10 kDa molecular weight cut-off. In the last step the RNA solution was spun down to $\sim$ 500 µl resulting in concentrations of > 200 µM. The flow-through of buffer-exchange steps 4 and 5 was used as the buffer for diluting the RNA and FMN prior to use and preparing the instrument. A FMN stock solution was prepared by dissolving solid FMN (Sigma aldrich) directly in ITC exchange buffer and determining the concentration via UV absorbance measurements at 450 nm (3.5.1).
For three consecutive measurements the RNA solution inserted into the cell was prepared by diluting the concentrated RNA stock solution in sample buffer down to a final concentration of 15 µM in a volume of 1 ml. The FMN solution for the syringe was prepared by diluting the FMN stock solution in sample buffer to a final concentration of 105 µM (7-fold ligand excess) and a total volume of 300 µl. Depending on

the solubility in the ITC sample buffer, which was determined in pretests prior to the start of the ITC measurements, stock solutions of compounds in DMSO were added to give final concentrations of 50 µM - 250 µM in both the cell and syringe solutions. The final DMSO content was dependent on the initial DMSO stock concentration of the screening compound and its solubility in sample buffer and ranging from 0.5 to 5%. The concentrations of compounds were exactly matched when preparing the cell and syringe solutions to avoid strong heat signals due to mismatching buffers. For the reference measurement in the absence of competition ligands DMSO only was used in the cell and syringe solutions in the same amount as used for testing the particular compound. The sample solutions both were properly vacuum degassed for several minutes before starting the experiment. Following an initial injection of 0.2 µl 28-31 injections of 1.2 µl were performed.
The resulting calorimetric data were analyzed using the software accompanying the calorimetry system (MicroCal Origin) and fitted to a single-site binding model. The obtained fitting parameters for the association constant and the enthalpic and entropic terms of the interaction between FMN and RNA were averaged. For the calculation of the binding constants the MicroCal Origin software assumes the following model:

$$\Delta H = 0.5 V_{cell} \Delta H^0 \left(1 + \frac{1 - \frac{X_t}{M_t} - \frac{n}{K_a M_t}}{\sqrt{(1 + \frac{X_t}{M_t} - \frac{n}{K_a M_t})^2 - 4 M_t}} \right) \tag{3.5}$$

where V_{cell} is the volume of the cell, K_a is the apparent association constant of FMN, X_t is the concentration of FMN in the cell, M_t is the concentration of RNA in the cell, n the stoichiometry of the interaction, and ΔH is the normalized enthalpy of binding.
A full characterization of all binding parameters, including the free binding energy ΔG^0 and the dissociation constant, was obtained with the following equation:

$$\Delta G^0 = \Delta H^0 - T \Delta S^0 = -RTlnK_D \tag{3.6}$$

where R is the ideal gas constant, T the temperature in Kelvin, ΔH^0 is the standard molar enthalpy, ΔS^0 is the standard molar entropy and ΔG^0 is the standard molar free energy released upon ligand binding. The K_D-values for the competition ligands were calculated similar to the previous section (3.5.1) from the apparent binding constant of FMN in the absence and presence of the screening compound using the equation stated in the previous section (Equation 3.4[182,183]). The reaction enthalpy for the interaction between riboswitch RNA and the competitive ligand was analyzed in an analogous fashion:[184]

$$\Delta H^0_{complig} = \Delta H^0_{FMN} - \Delta H^0_{FMN+lig} \left(1 + \frac{1}{K_{complig} \times c_{lig}} \right) \tag{3.7}$$

with $\Delta H^0_{complig}$ representing the molar enthalpy set free upon binding of the competitive ligand, ΔH^0_{FMN} and $\Delta H^0_{FMN+lig}$ representing the molar enthalpies of FMN binding to the riboswitch in the absence and presence of competitive ligand, $K_{complig}$ being the association constant for the interaction of RNA and competitive ligand and c_{lig} expressing the total concentration of competitive ligand in the ITC cell.

3.5.3 SPR measurements for the FMN riboswitch aptamer domain

SPR measurements were carried out on a Biacore® T200 system (GE Healthcare Lifesciences) at 25 °C using streptavidin(SA)-coated sensor chips (GE Healthcare Lifesciences). The running buffer during the measurement consisted of 50 mM Tris-HCl (pH 7.5), 100 mM KCl and 2 mM $MgCl_2$. Analytes were always run on two flow cells within the sensor - a sample and a reference flow cell. On the reference flow cell the 5′-biotinylated DNA oligonucleotide (sequence 5′ → 3′: Biotin-GAAAATGTTTACTCCAAGTTAAAAA) , which was initially purchased from Eurofins Scientific in order to hybridize with the elongated *F. nucleatum* FMN riboswitch aptamer domain (see 3.3.4) and therefore was available in-house, was immobilized to detect non-specific binding events to nucleic acid structures.

3.5.3.1 Immobilization and kinetic measurements using SPR with the full-length FMN riboswitch aptamer domain

The RNA sample for the immobilization of the biotinylated riboswitch RNA on the sensor chip was prepared similar to a protocol that had been published previously and had been employed to study the kinetics of binding of adenosylcobalamin (AdoCbl) to the *btu*B riboswitch from *E. coli*.[175]
The refolding protocol for the preparation of 200 µl of a 1-µM refolded RNA sample suitable for immobilization on the SA sensor chip was as follows:

1. 100 µl of a RNA solution diluted down to 2 µM in 50 mM Tris-HCl (pH 7.5) was heated to 95 °C in a heating block for 1 minute
2. The RNA sample was snap-cooled on ice for 5 minutes
3. 100 µl of a solution containing 50 mM Tris-HCl (pH 7.5), 200 mM KCl and 4 mM $MgCl_2$ were added to the RNA sample
4. The RNA sample was further kept at 37 °C for 15 minutes to allow complete refolding

The stock solution of the reference, the biotinylated DNA oligonucleotide, had a concentration of 50 µM in autoclaved water. Since no refolding procedure was applicable to this nucleic acid, a 1-µM solution in SPR running buffer was prepared by diluting the stock solution directly in autoclaved water and a 10X running buffer. FMN was dissolved in SPR running buffer to obtain an analyte stock solution. The

concentration was determined by measuring the UV absorbance in running buffer at 450 nm using the molar extinction coefficient of ε = 12,500 $\frac{1}{\mathrm{M} \times \mathrm{cm}}$ (3.5.1).
Before the immobilization was started, three consecutive conditioning cycles to remove any unbound streptavidin from the matrix were carried out by applying 1 M NaCl and 50 mM sodium hydroxide for 60 s with a flow rate of 30 µl/min each time. Subsequently, the 1-µM solution of the biotinylated DNA oligonucleotide was immobilized on the reference surface with a low flow rate of 1 µl/min for 300 s. This yielded an immobilization level of $\sim$ 2,000 RU. For the immobilization of the riboswitch RNA on the sample surface a lower concentration was initially chosen because of the molecular weight dependency of the immobilization levels[175] and the four times higher molecular weight of the RNA molecule compared to the biotinylated oligonucleotide. Immobilization was performed stepwise starting with a concentration of 100 nM for 600 s and a flow rate of 1 µl/min. The immobilization level was 1,360 RU for this step. Additional 300 RU could be immobilized with completely the same setup. It was ensured that the streptavidin binding sites all were occupied by immobilizing a solution of 200 nM riboswitch RNA for the same time frame and with the same flow rate. The overall immobilization level of $\sim$ 1,650 RU was more or less unchanged.
The kinetic measurements for determining the binding characteristics of the FMN-RNA interaction were carried out at a flow rate of 10 µl/min with an association phase of 60 s and a dissociation phase of 300 s, followed by a stabilization period of 60 s before the next analyte injection. Briefly, immobilized sensor surfaces were equilibrated by running start-up cycles eight times with buffer only. A dilution series of 1:5 dilutions of FMN with concentrations between 16 nM and 50 µM was prepared and measured starting from zero concentration (buffer only), proceeding stepwise to the highest concentration. Afterwards, the concentration series was measured backwards from the highest to zero concentration of FMN. The resulting kinetic data (K_D, rate constants k_a and k_d) were fitted to a 1:1 interaction model using the Biacore® evaluation software that accompanied the instrument and compared to the results from the evaluation of the K_D-values using the steady-state affinity equations. Data were double-referenced by subtracting the responses from the reference cell and the zero concentration.
The following rate equation for an 1:1 interaction model was used by the software:[185]

$$\frac{dR}{dt} = k_a C R_{max} - (k_a C + k_d) R \tag{3.8}$$

where *dR/dt* is the net rate of complex formation, k_a the association rate constant in $\mathrm{M}^{-1}\mathrm{s}^{-1}$, k_d the dissociation rate constant in s^{-1}, C the analyte cocentration, R_{max} the maximum analyte binding capacity in RU and R the response measured in RU.
To evaluate the results via determination of the steady-state affinity, a plot of the response at equilibrium R_eq against the analyte concentration C was fitted to this equation:[185]

$$R_{eq} = \frac{K_A C R_{max}}{K_A C + 1} \tag{3.9}$$

The value of the equilibrium association constant K_A was therefore obtained, directly resulting in the K_D-value via:

$$K_D = \frac{1}{K_A} \tag{3.10}$$

3.5.3.2 Immobilization and kinetic measurements using SPR with the elongated FMN riboswitch aptamer domain

The lyophilized, biotinylated 25-nt long DNA oligonucleotide was dissolved in autoclaved water to a stock concentration of 50 µM. A 10-µM stock solution of the biotinylated DNA oligo was prepared directly before the hybridization was started. The PAGE-purified and electroeluted RNA was concentrated by using 10 kDa molecular weight cut-off centrifugal filter devices (Amicon®, Merck Millipore) to volumes of ≈ 30-50 µl with concentrations of ≥ 50 µM. FMN was dissolved to a 10 mM stock solution in 1X SPR running buffer. The sample preparation was conducted according to a previously published protocol:[175]

Biotinylated DNA oligo: The biotinylated DNA oligonucleotide was diluted to a final concentration of 1µM in 50 mM Tris-HCl (pH 7.5), 100 mM KCl and 2 mM $MgCl_2$ with 10X SPR running buffer and autoclaved water in a 50 µl total volume.

RNA-DNA hybrid: The PAGE-purified elongated riboswitch RNA was diluted to a 0.1-µM solution in 50 mM Tris-HCl (pH 7.5) and heated to 90 °C for 1 minute. The biotinylated DNA in a solution containing KCl and $MgCl_2$ was added to the RNA to a final concentration 10-fold higher than the RNA (1 µM) resulting in a RNA-DNA hybrid in 1X SPR running buffer. This final mixture was further kept at 55 °C for 2 min, followed by an incubation at 37 °C for 30 minutes to fold and anneal the construct (Figure 3.6). For one immobilization ≈ 50 µl sample solution was needed.

Pure FMN riboswitch RNA: About 30 µM RNA was prepared in 50 mM Tris-HCl (pH 7.5) and autoclaved water and heated to 90 °C for 1 min. After KCl and $MgCl_2$ had been added to their final concentrations, the sample was further incubated at 37 °C over 15 minutes to fold the RNA.

The streptavidin sensor chips were conditioned by applying a solution of 1 M NaCl and 50 mM NaOH at a flow rate of 30 µl/min for 60 seconds onto the reference and the sample flow cell. This step was repeated three times. On the reference flow cell the biotinylated DNA oligonucleotide alone was immobilized by applying the 1-µM solution at a flow rate of 10 µl/min for 5 minutes.

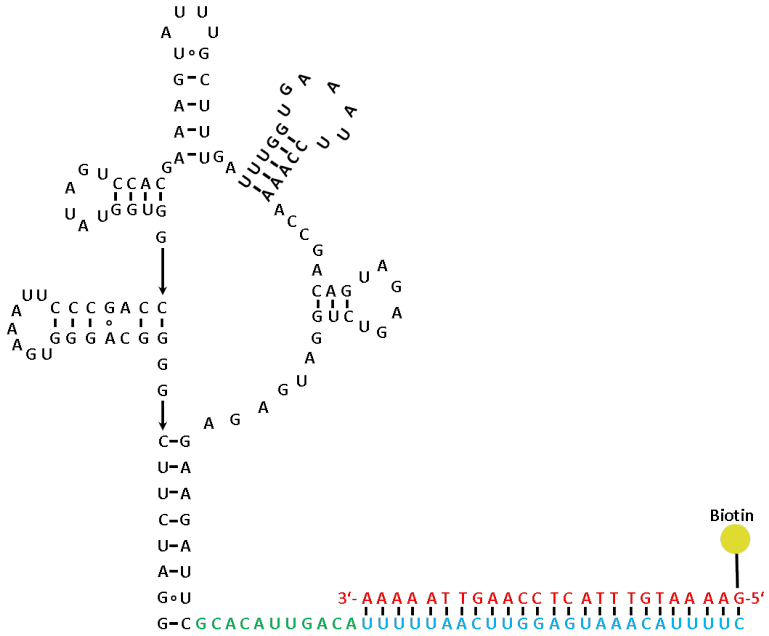


FIGURE 3.6: The immobilization of the FMN riboswitch to the streptavidin sensor surface is achieved by hybridizing a 25-nt long DNA oligonucleotide (red) with a biotin (yellow) at its 5′ end to the complementary nucleotides of the 3′ end of the elongated aptamer (blue). For proper folding 11-nt (green) directly following the aptamer domain at its 3′ end are not involved in the hybridization as a "safety gap". Watson-Crick base pairing is indicated by *bold black lines* between nucleotides.

Immobilization on the sample flow cell was performed in two steps. First, the DNA-RNA hybrid was applied with the same flow rate and over the same time course as the biotinylated DNA oligo was applied to the reference cell. Then the flow rate was reduced to 1 µl/min. Over 10 minutes the pure RNA sample was added to the flow cell to increase the density of RNA bound to the surface by utilizing excess DNA oligonucleotide from the first part of the immobilization. Typical immobilization levels for both flow cells were about 2,500-3,000 response units. Application of either the reference cell solution of the biotinylated DNA oligonucleotide alone or the pure RNA solution on the sample flow cell was repeated, if the immobilization levels differed too much.

For the determination of the kinetic constants a concentration series of FMN solutions was prepared. Seven different concentrations were applied to both surfaces: 30 nM, 100 nM, 250 nM, 500 nM, 1 µM, 5 µM and 30 µM. For the conditioning of the flow cells eight start-up cycles were performed with buffer only. The concentration series was run from zero concentration to the highest concentration of FMN back and forth. An association time of 180 s was chosen for the different FMN dilutions, whereas a longer time period of 600 s was used for the dissociation phase. The flow rate was set to 75 µl/min. Following the dissociation a stabilization period of 120 s was added, before the flow cells were regenerated by applying 1 M NaCl for 180 s at a flow rate of 30 µl/min to completely remove FMN from the previous

injection. Data were fitted as described previously with the Biacore® evaluation software (3.5.3.1) after subtracting the buffer response and the reference sensor response ("double-referencing"[186]).

3.5.3.3 Immobilization and kinetic measurements using SPR with the two-piece FMN riboswitch aptamer domain

The two-piece RNA construct was folded in the presence of FMN to ensure correct pairing of the two independent strands A and B adding up to the full-length FMN riboswitch aptamer domain. 100-µM stock solutions of strand A and 3'-biotinylated strand B served as starting point. A solution of 150 µl was prepared for the immobilization on the sample surface by diluting both strands and the FMN stock solutions in 10X running buffer and autoclaved water to result in final concentrations of 5 µM of both RNA strands and 12.5 µM of FMN in 50 mM Tris-HCl (pH 7.5), 100 mM KCl and 2 mM $MgCl_2$. Reannealing was accomplished by incubating both strands at 37 °C for 30 minutes in the presence of FMN, followed by an incubation on ice for 15 minutes (see 3.6.1). The solution of the biotinylated DNA oligonucleotide for immobilization onto the reference surface was prepared in the same way as described for the SPR measurements using the full-length aptamer (3.5.3.1) resulting in the same concentration.
Both, sample and reference cell were loaded with their nucleic acid ligands by using the function of the Biacore® control software to aim for a certain immobilization level, after the conditioning cycles had been performed (3.5.3.1). The program automatically adjusts the flow rate and contact times with the surface depending on parameters like the steepness of the loading curve and the RU level after a first immobilization pulse. The target level was set to 5,000 RU. However, for both flow cells an overall immobilization level of $\sim$ 2,200 RU could be achieved.
For the determination of the kinetic constants a similar FMN dilution series as described in the previous section was prepared with six different concentrations ranging from 20 µM down to 6.4 nM in 1:5 steps. The association and dissociation times remained at 60 s and 300 s, respectively, with a flow rate of 10 µl/min and a stabilization period of 60 s. Before the first kinetic measurements eight start-up cycles were run to equilibrate the sensor surfaces. The concentration series was either run once for all concentrations starting from zero to 20 µM measuring the highest concentration twice or all concentrations were run twice starting from zero concentration, proceeding stepwise to the highest concentration and measuring the concentration series backwards directly after that. Again, the kinetic constants were derived from the Biacore® evaluation software (Equations 3.8, 3.9, 3.10).
Binding affinities of small molecule screening compounds were also tested using the SPR-approach with the two-piece riboswitch RNA construct. The measurements were done in a 1X running buffer supplemented with 5% DMSO. Solubility limits of the compounds in running buffer were evaluated in pretests. Finally, the compounds were tested at concentrations between 50 and 250 µM. The sample solutions were

prepared by mixing stock solutions of the test compounds in DMSO that were 20-fold more concentrated compared to the final assay conditions with 1.05 X running buffer. This led to solutions containing 50 mM Tris-HCl, 100 mM KCl and 2 mM $MgCl_2$ and 5% compound ($\equiv$ DMSO). A FMN control solution was prepared in the same manner with DMSO only. A concentration of 5 µM was chosen for the control sample to saturate the RNA binding sites. DMSO has a very high bulk response when used in SPR measurements. Due to this fact a solvent correction procedure was carried out, which was accessible via the Biacore ® control software. Even when working very precisely, it is almost impossible to match the sample solution and the running buffer perfectly. Secondly, the responses from small molecules interacting with the immobilized RNA were presumably much lower than the actual bulk shift caused by mismatching DMSO concentrations. Namely, the theoretical maximum response R_{max} of the analyte binding to the RNA is molecular weight-dependent:

$$R_{max}(RU) = \frac{Molecular\ weight\ of\ analyte}{Molecular\ weight\ of\ RNA} \times R_L(RU) \times S_m \tag{3.11}$$

where R_L is the theroretical ligand immobilization level and S_m is the stoichiometry of binding. Eight different solutions with concentrations ranging from 5.8 to 4.5% DMSO were prepared for the solvent correction procedure. Within the Biacore ® control software the measured responses are plotted automatically versus the DMSO concentration, so that during the measurement of the samples the DMSO related response was subtracted from the total response to derive response related to analyte binding only.

Again, eight start-up cycles were run to equilibrate the system with the adjusted buffer conditions. The negative control (running buffer suppl. with 5% DMSO) and all test compound solutions were measured twice. Every tenth sample was a FMN control sample to check the integrity of the sensor surface and if there was still a sufficient signal upon binding. The discrimination between binding compounds and inactive ones was done by comparing the binding level to the one of the positive (FMN 5 µM) and negative (running buffer only) control samples. Full kinetic characterization was not performed for the screening compounds, since due to limiting solubilities the highest possible concentrations used presumably did not lead to saturation of the RNA binding sites. An additional control experiment, that was carried out, was the full kinetic characterization of FMN binding to the riboswitch RNA in the presence of 5% DMSO after the small molecule screening to exclude any negative impact of the screening compounds on the ligand activity on the surface.

3.6 Crystallography and structure determination

3.6.1 Crystallization of the two-piece *F. nucleatum* FMN riboswitch RNA with FMN

Crystals of the *F. nucleatum* FMN riboswitch aptamer construct formed by annealing two strands prior to crystallization and bound to FMN were grown using the hanging drop method in 24-well ComboPlates (Greiner Bio-One, Frickenhausen, Germany) sealed with 18 mm diameter cover slides at 20 °C as reported earlier.[39,61] Both strands (for sequence see 3.3.3) were stored in a buffer containing 50 mM K-HEPES (pH 7.5) and 0.1% sodium azide after *in vitro* transcription and purification by denaturing PAGE. The presence of $MgCl_2$ is a prerequisite of FMN binding to the riboswitch aptamer domain. Before the crystallization experiment were started, both RNA strands were buffer-exchanged five times into reannealing buffer consisting of either 50 mM K-HEPES (pH 7.5), 100 mM KCl and 2 mM $MgCl_2$[39] or 10 mM cacodylic acid (pH adjusted to 6.8 with NaOH), 100 mM sodium acetate and 4 mM $MgCl_2$[61] and concentrated during the final exchange to concentrations of 500-800 µM using 3 kDA Amicon®Ultra-0.5 ml centrifugal filters. The ligand FMN was directly dissolved in the respective reannealing buffers to prepare the stock solutions. The sample solutions for the crystallization were made by diluting both RNA strands and FMN in the respective reannealing buffers down to final concentrations of 400 µM for the RNA and 1 mM for FMN. Reannealing was accomplished by incubating this mixture for 30 minutes at 37 °C and placing it 15 minutes on ice afterwards. Crystals were then grown in drops consisting of 1 µl sample solution with 1 µl mother liquor for several days. The crystal trays were subsequently stored at a constant temperature of 20 °C.

In the case of the sample solution containing K-HEPES the mother liquor contained 0.1 M MES-sodium (pH 6.5), 0.1 M $MgCl_2$ and 13% PEG 4000 with a reservoir volume of 750 µl, whereas the mother liquor contained 0.1 M sodium acetate (pH 5.0), 0.2 M $MgCl_2$ and 7% PEG 4000 with a reservoir volume of 750 µl when using the sample solution with sodium cacodylate. Attempts to cocrystallize the FMN riboswitch with diverse virtual screening hits were done with DMSO concentrations of 2%. 50-fold stock concentrations were therefore mixed with 400 µM of each RNA strand and incubated in the same way as with FMN before setting up the crystal trays. Crystals were then either grown in drops generated by mixing 1 µl RNA with 1 µl mother liquor or in a 3 µl/ 3 µl-ratio. The cryoprotectant solution for harvesting the crystals consisted of 0.1 M MES-sodium (pH 6.5), 0.2 M $MgCl_2$, 10% PEG 4000 and 20% glycerol. Raw data for crystals containing the two-piece riboswitch RNA construct co-crystallized with FMN were collected on beamline P13 at the DESY institute (Hamburg, Germany) and the data were integrated by the help of the XDS package (`xds.mpimf-heidelberg.mpg.de`).[187] Pointless within the CCP4 program suite (`www.ccp4.ac.uk`) was used to scale the data.[188]

Chapter 4

Virtual screening for ligands of the *F. nucleatum* FMN riboswitch

4.1 Abstract

A large set of compounds was virtually screened to identify potential novel ligands of the FMN riboswitch with chemical structures distinct from the flavin analogues closely resembling FMN using the software DOCK. The approach was validated by comparison of the docking results for known binders with virtually created non-binders lacking the most crucial functionalities of the heterocycle of FMN for forming the hydrogen bond interactions to the aptamer domain. Candidate compounds were finally selected by visual inspection for experimental testing to confirm the docking predictions (Chapter 5). In a second virtual screening process, privileged substructures were screened against the aptamer domain of the FMN riboswitch to facilitate the search for potential ligands and avoid inspection of a large number of false positive predictions. Substructures were regarded as privileged if they had a structural similarity to the core scaffold of FMN interacting with A99 in the riboswitch binding site, the uracil-like edge.

4.2 Results

4.2.1 Screening of a training set of known binders and virtually created non-binders

So-called sphere sets are required by DOCK to define the binding site boundaries. Here, these spheres were a set of dummy atoms based on a cubic grid wrapped around the coordinates of the original ligand and bordered by the receptor atoms. The ring atoms of the screened compounds were placed in the vicinity of these spheres, since the ring systems of the compound are used as anchor points during docking.[114,189] For a successful docking approach the sphere set has to be chosen carefully. The sphere sets for receptor preparation and for the matching during the docking are not necessarily of the same size. Due to the properties of the binding site

and the ligand FMN, a smaller sphere set for matching the poses was chosen (Figure 3.2).
A large contribution to binding energy between the *F. nucleatum* FMN riboswitch and its ligand was attributed to the network of hydrogen bonds formed by the phosphate group of FMN (Figure 1.6).[39] There is also an additional electrostatic interaction between a conserved Mg^{2+}-ion and the phosphate. However, for docking we focused on the uracil-like edge of the isoalloxazine heterocycle, that is involved in base-pairing to A99 similar to standard A-U Watson-Crick base pairs. This was decided for several reasons:

1. Compounds with high shape similarity to FMN would include molecules with a high molecular weight and rotational flexibility due to the space occupied by the ribityl group linking heterocycle and phosphate.

2. Scores tend to scale poorly with increasing mass and number of rotatable bonds, complicating the interpretation of results.[105] FMN itself has a molecular weight of 455 Da and 7 freely rotatable bonds.

3. As shown previously for phosphosugar-metabolizing enzymes[190] or inhibitors of 6-phosphogluconate dehydrogenase[191] malonic acid-based compounds and carboxylic acids in general are often screened as phosphate mimics to resemble the interactions of the phosphate group. The virtual screening approach was aiming for orally available drug-like small molecules. A sphere set filling the whole, large binding site of the FMN riboswitch would have resulted in compounds with a high likelihood of being constantly charged and having a high molecular weight at the same time. Such molecules easily violate the "Rule-of-Five" as predictor of oral bioavailabilty.[101]

For the purpose of assessing the accuracy of binding mode predictions and the discrimination between known active ligands for the FMN riboswitch and manually created non-binders a training set of flavin analogues was assembled (Table 4.1). Three substances included in the training set were known actives. Lumiflavin exhibited binding affinity in the two-digit µM-range in a fluorescence-based assay and was only slightly less potent than riboflavin in the same assay.[39] Compound **A** and Compound **B** were manually created and both supposed to bind to the riboswitch. Compound **A** was lacking the hydroxyl groups and carrying a n-pentyl substituent instead of the ribityl moiety. Based on the small differences between lumiflavin and riboflavin a small contribution of the hydroxyl groups to binding affinity of flavin analogues was expected. Compound **B** was only lacking the two methyl group substituents on the isoalloxazine ring and therefore little influence on the overall binding pattern was expected. Compounds **C**-**E** were all expected to be non-binders or more precisely to be compounds with less favourable binding properties. In all three flavin analogues the double hydrogen bonding to A99 was virtually impossible, which should negatively affect the docking score.

The initial receptor setup composed of a sphere set generated according to a standard protocol[114,189] with an assignment of partial charges based on an AMBER look-up table with fixed charges per atom type could not discriminate between the different flavin analogues in the training set (Table 4.1). Riboflavin and roseoflavin both had considerably negative docking scores. However, the predicted binding mode for roseoflavin clearly deviated from the one that was predicted for riboflavin (Figure 4.1) even though it was a known ligand and the structure of roseoflavin bound to this RNA had been solved earlier.[39,58] In particular, the predicted binding mode of roseoflavin did not reveal any directed electrostatic interactions between ligand and RNA. Furthermore, the docking score for Compound **D** from the training set (Table 4.1) without any polar groups that could interact with A99 in a productive manner was slightly worse compared to that of the known ligands like riboflavin.

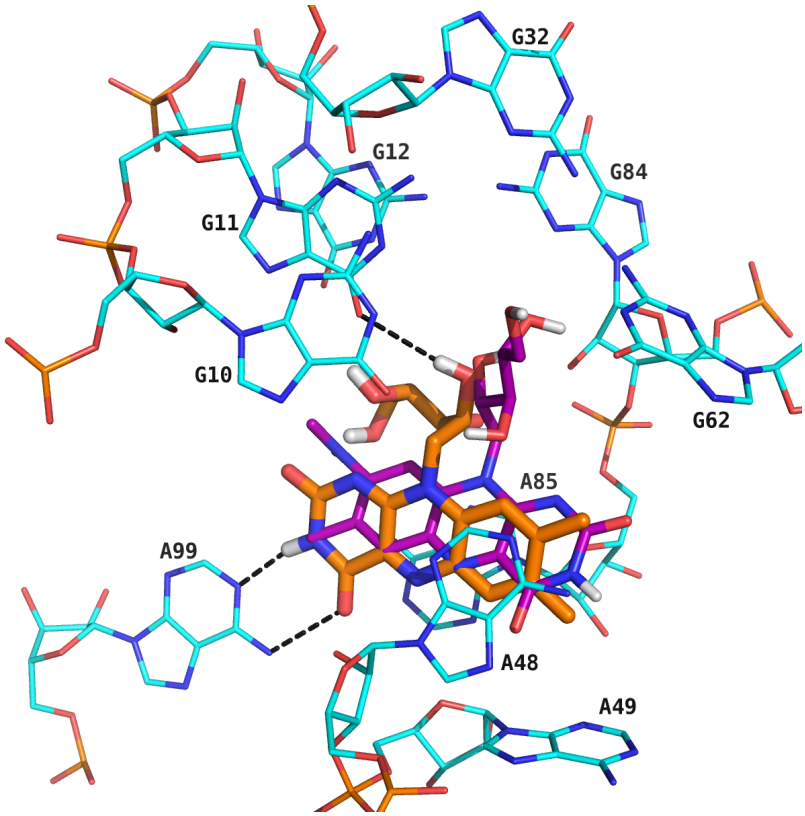


FIGURE 4.1: Predicted binding modes of riboflavin (orange carbon atoms) and roseoflavin (purple carbon atoms) after docking with the initial receptor setup. The binding site residues are displayed with cyan carbon atoms and hydrogen bonds as *black dashes*.

Since the double hydrogen bonding interaction with A99 was supposed to be important for FMN and its analogues binding tightly to the riboswitch pocket, the partial charges of this residue were adjusted to locally create a higher polarization of the N-H bond to facilitate hydrogen bonding. This procedure is common when using this docking algorithm and has been applied earlier.[192] Not every hydrogen bond in a protein or a nucleic acid has the same strength, rather the strength is context-dependent. The micro-environment polarity impacts the hydrogen bond strength

and buried hydrogen bonds that are less solvent-exposed tend to be stronger.[73,193] A99 indeed is buried in a less solvent-exposed surrounding and the NH_2-group is directed towards the pocket, so that this hydrogen bond is located in an environment with a lower dielectric constant. To give consideration to this, the partial charges were adjusted. The underlying force field (AMBER) used a fixed partial charge for each atom type for the calculation of the electrostatic grid points during receptor preparation. Thus, the parametrization tables were adjusted manually for A99.
The adjustment of the partial charges by 0.3 units supported a better discrimination between known binders and compounds with clearly reduced ability to interact with the FMN riboswitch. Normally, the AMBER assigned partial charges are -0.692 for the nitrogen atom and +0.347 for each hydrogen atom of the NH_2-group of an adenine residue. After modification of the parametrization table, the N6-atom of A99 was assigned a partial charge of -0.992 and both attached hydrogen were assigned a partial charge of +0.497. The sum of partial charges therefore remained zero, reflecting the status of RNA in solution, where the polyanionic macromolecule is surrounded by positively charged counter ions to keep electroneutrality.[124] All of the three following requirements were given in the final receptor setup:

1. A reasonable rmsd-value between co-crystallized ligand and docked pose
2. Predicted binding mode close to the crystallographically determined for the known actives riboflavin and roseoflavin, and reasonable binding mode for lumiflavin including the double hydrogen bonding pattern to A99
3. Sufficient discrimination between the known actives and the supposed non-binders

A successful pose prediction by a docking program is usually measured by the root-mean-square deviation (RMSD) between the heavy atom positions of the ligand conformation observed in the crystal structure and the top-scoring pose. In most cases solutions with rmsd-values below 2 Å are considered as indicative for successful docking run, depending on the size and molecular complexity of the ligand.[194,195] With the adjusted receptor setup a rmsd-value of 1.271 Å could be obtained for riboflavin when superimposing the heavy atoms with the crystallographically determined binding mode (Table 4.1). For roseoflavin a higher rmsd-value of 2.550 Å was obtained (Table 4.1). The phosphate group was disregarded for the development of the sphere set used for docking. Hence, riboflavin was used for redocking purposes instead of FMN itself. Particularly, the heterocycle of the crystallographically determined binding mode and the predicted binding pose aligned well (Figure 4.2 a) with a low rmsd for these atom positions. Although the rmsd-value was twice as high for roseoflavin in comparison to riboflavin, especially the heterocyclic molecular part was predicted to be well-aligned with the isoalloxazine ring of the crystallographically determined structure (Figure 4.2 b).

TABLE 4.1: Structures of the training set compounds and their docking scores in the final receptor setup used for the virtual screening.
n.d. = not determined

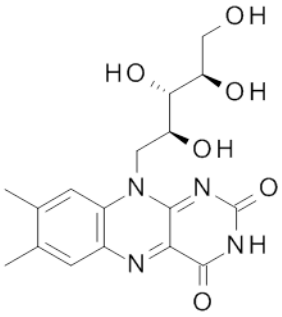
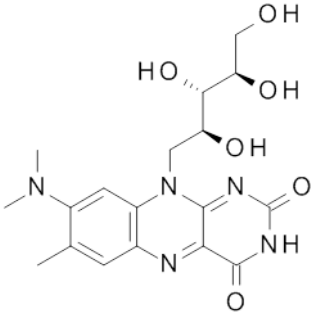
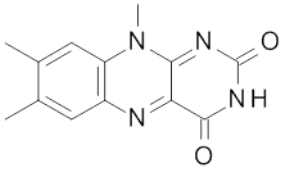
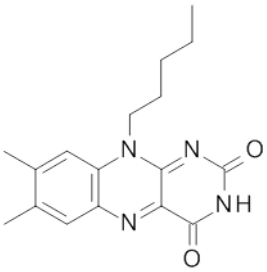
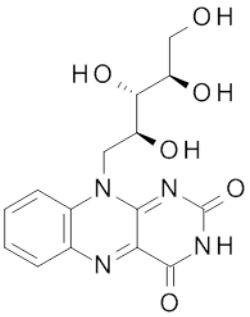
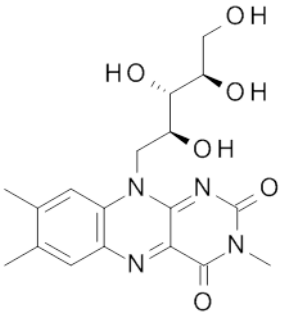
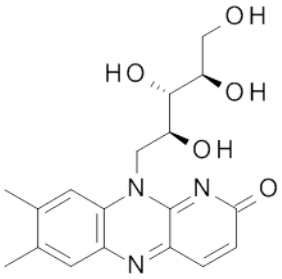
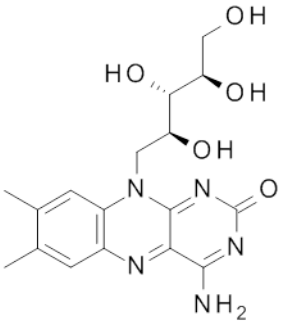
Compound	Structure	Total score[$\frac{kJ}{mol}$]	RMSD[Å]
Binders			
Riboflavin		-51.47	1.271
Roseoflavin		-49.05	2.550
Lumiflavin		-41.41	n.d.
Compound **A**		-45.78	n.d.
Compound **B**		-48.13	n.d.

TABLE 4.1: Structures of the training set compounds and their docking scores in the final receptor setup used for the virtual screening.
n.d. = not determined

Compound	Structure	Total score[$\frac{kJ}{mol}$]	RMSD[Å]
Non-binders			
Compound **C**		-42.46	n.d.
Compound **D**		-44.59	n.d.
Compound **E**		-43.18	n.d.

The second requirement for a reasonable prospective docking study that was only fulfilled by the final receptor setup with adjusted partial charges was the reflection of the hydrogen bonding pattern to A99 in all predicted binding modes of the flavin analogues with known activity towards the riboswitch RNA. The heterocycles of lumiflavin, roseoflavin and riboflavin all overlaid very well allowing for the formation of the desired hydrogen bonds (Figure 4.2 c)
Furthermore, the discrimination between known binders and the artificially compiled non-binders could be slightly increased by adjusting the partial charges to a difference of up to ≈ 7 kJ/mol (riboflavin compared to compound **D**, Table 4.1). The scores for all of the compounds from the training set were highly negative (Table 4.1), mainly due to the high energy contribution of the vdW-term used in the scoring function. Even with disrupted hydrogen bonding patterns all of the training set

compounds still displayed a high shape complementarity to the binding site.

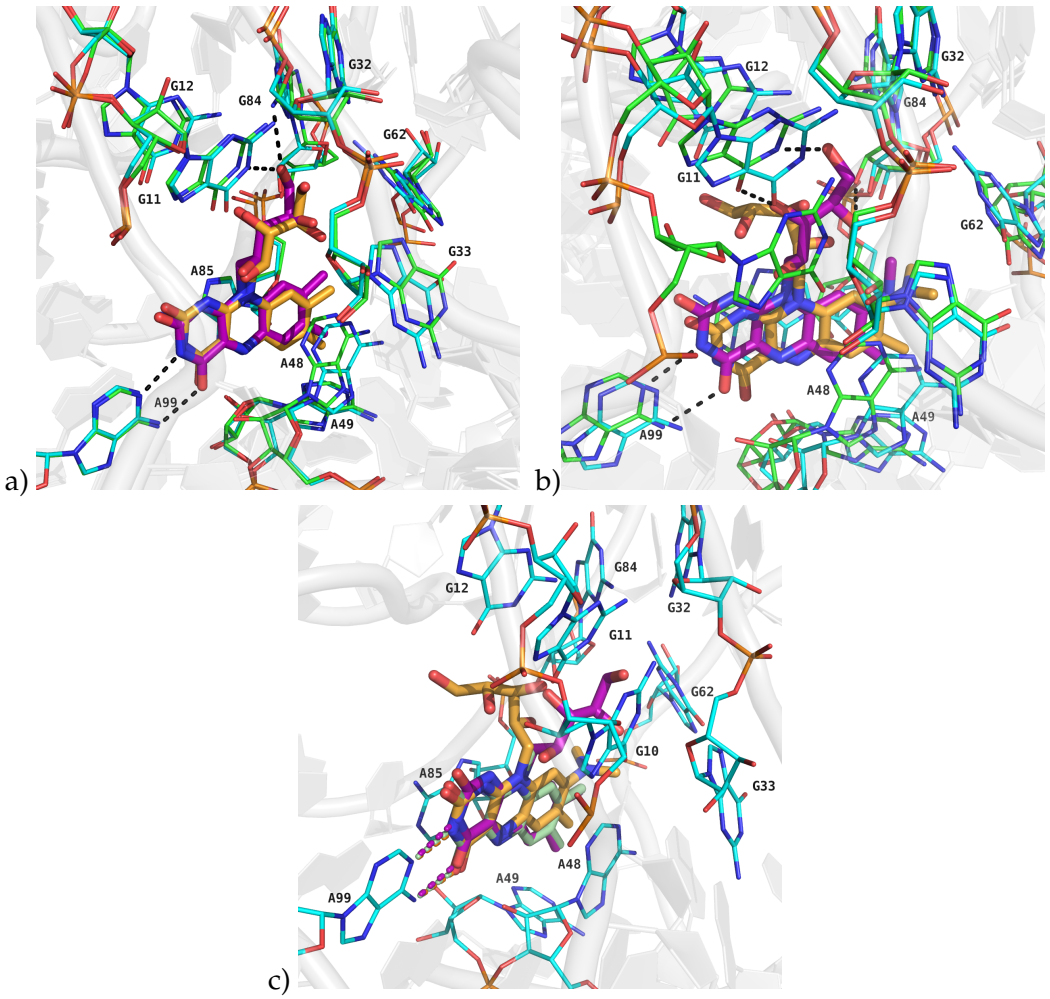


FIGURE 4.2: a) Overlay of the predicted binding mode for riboflavin (orange carbon atoms) and its crystallographically determined binding mode (pdb-code: 3f4g, purple carbon atoms) after alignment of the receptor used for docking (cyan carbon atoms) with the riboflavin-bound FMN riboswitch (green carbon atoms). Hydrogen bonds formed according to crystallographic data are displayed as *black dashes*. b) Overlay of the predicted binding mode for roseoflavin (orange carbon atoms) and its crystallographically determined binding mode (pdb-code: 3f4h, purple carbon atoms) after alignment of the receptor used for docking (cyan carbon atoms) with the roseoflavin-bound FMN riboswitch (green carbon atoms). Hydrogen bonds formed according to crystallographic data are displayed as *black dashes*. c) The predicted binding modes of the three flavin analogues lumiflavin, roseoflavin and riboflavin are depicted with pale-green, orange and purple carbon atoms, respectively. All predicted poses reflect the crucial hydrogen bond pattern (dashed lines colour-coded according to the particular ligand) to the binding site residues displayed with cyan-coloured carbon atoms.

As expected, the known binders riboflavin and roseoflavin ranked best, since they possess all features necessary for binding of a flavin to the FMN riboswitch except for the phosphate group. Lumiflavin and the N3-methyl analogue of riboflavin were ranked worst. Although lumiflavin can form the double hydrogen bond pattern to A99, it is lacking the hydroxyl group in the side chain for an additional hydrogen bond interaction and fills only one part of the binding site. Compound **C** formed

only one hydrogen bond with A99 in its predicted binding pose, but the hydrogen bond in the ribityl moiety was retained. Compound **D** with its unsaturated CC-bond instead of the uracil-like edge of FMN had the lowest electrostatic interaction energy, because none of the hydrogen bonds to the crucial adenine could be observed in the predicted binding mode. Surprisingly, compound **E** did not have the least negative docking score, although the donor/acceptor properties had been interchanged at the ring positions 3 and 4 of the heterocycle. The compound was placed in a completely different position within the pocket in the corresponding top-scoring docking solution compared to the other training set compounds. Collectively, the adaption for the partial charges of A99 led to an improved docking performance.

4.2.2 Filtering the database for desired properties

Our in-house database of compounds was filtered for compounds with lead-like properties (Table 4.2). This filter step aimed at limiting the molecular complexity and focusing on compounds with lead-like properties, as already been described in previous protocols.[161] Binding site properties and properties of the natural ligands were taken into account when determining the selection criteria (Table 4.2). After the first filter step employed, about 800,000 molecules of the original database were left.

TABLE 4.2: Selection criteria to filter the in-house database for molecules with lead-like properties. The criteria were subdivided into three categories: the absence of unwanted functionalities, lead-like properties and limited complexity.

Selection criteria	**Definition**
Absence of unwanted functionalities	No unwanted groups
Lead-like properties	Σ of heavy atoms: $\geq$12 and $\leq$30
	$\geq$1 hydrogen bond donor
	$\geq$1 hydrogen bond acceptor
	Σ H-bond functionalities: $\geq$2 and $\leq$12
	SlogP between -3.5 and +3.5
	Net charge: -1, 0 or +1
	$\geq$1 ring system
	Molecular weight $\leq$500 Da
Limited complexity	$\leq$7 freely rotatable bonds
	fused rings $\geq$1

Briefly, the application of these criteria ensured that candidate compounds do not violate the requirements of the binding site (e.g. number of hydrogen bond donors/acceptors) and show typical properties of lead-like molecules that underlie more stringent rules than the initial "Rule-of-Five".[196,197] All compounds with unwanted groups e.g. reactive groups like michael-acceptors and epoxides were removed from

the candidate list. The number of freely rotatable bonds was restricted to reach enhanced docking accuracy and reliability of the results. Earlier studies have revealed, that the docking accuracy of DOCK drops substantially for ligands with ≥ 8 freely rotatable bonds.[198]

4.2.3 The pharmacophore model

The pharmacophore model employed in the subsequent filter step was based on the binding mode observed for the heterocycle of FMN, since compounds were sought that bind into this subpocket. In addition to an excluded volume covering all receptor atoms to abandon any possible steric clash, six other pharmacophore features were included (Figure 4.3). The area covered by the dimethyl-substituted benzene ring of FMN was defined as hydrophobic region in the size of a sphere with 2.1 Å radius. Directly adjacent to this region an aromatic region was defined with a radius of 2.4 Å to enable π-π-stacking interactions to both of the adenines A48 and A85. The spheres corresponding to the uracil-like edge of FMN had a radius of 1.6 Å around the N3-donor function and the carbonyl oxygen at the 4-position of FMN.

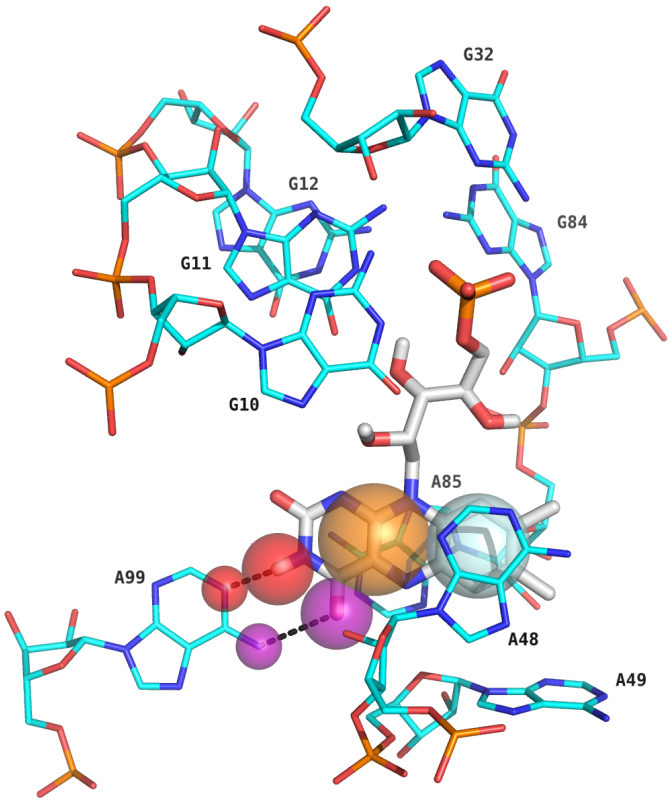


FIGURE 4.3: The pharmacophore hypothesis for FMN riboswitch ligands is depicted as colour-coded spheres. The volumes of the spheres define the positions of the corresponding electronic and steric features in ligand conformations. Binding site residues are displayed with cyan carbon atoms, whereas FMN is displayed with light grey carbon atoms. The pale-cyan sphere represents a hydrophobic region and the orange sphere represents an aromatic region to enable π-π-stacking. Red and purple spheres indicate the hydrogen bond donor/acceptor-pairs of FMN and A99.

Additionally, the associated donor and acceptor functions in the adenine 99 residue of the receptor were defined as pharmacophore features with radii of 1.0 Å each. The sphere radii were kept small to ensure a hydrogen bond interaction with appropriate distances and angles. In order to fulfill the pharmacophore hypothesis, the excluded volume was applied as an essential feature. In addition, all four features associated with the double hydrogen bond between FMN and A99 were determined as mandatory features, and a total of six out of the overall seven features were required to pass the pharmacophore search.
After several millions of different low energy conformations arising from the $\approx$ 800,000 molecules passing the first filter step had been calculated (Chapter 2, 3.1.4), only the conformation of a particular molecule with the lowest rmsd-value between its matching position and the pharmacophore query points was added to the result table of the pharmacophore search. A total of 73,293 molecules passed this second filter step. These molecules were subsequently subjected to docking. This corresponds to around 1.5% of the original count of library molecules.

4.2.4 Screening of a large compound library

Altogether a total of about 50,000 compounds were successfully docked into the binding pocket. For the remaining molecules the program could either not generate a reasonable pose or simply failed in the calculation for reasons like issues with the input data. The most promising candidate compounds were selected by considering the total score and the calculated predicted ligand efficiency (Equation 3.2). Predicted docking solutions with scores of $\leq$ -38 kJ/mol as threshold were visually inspected. This threshold was chosen due to the fact that the flavin analogue lumiflavin meeting all the minimal criteria of binding to the FMN riboswitch and filling the part of the receptor binding site used for matching had a docking score in this range with the final receptor setup when using DOCK 3.5.54. Further, a LE of at least 1.3 kJ$\times$mol$^{-1}\times$HA^{-1} and the predicted hydrogen bond interactions to A99 similar to those from FMN were required.
Based on these criteria 20 compounds were purchased (Table 4.3). In addition four further compounds were selected (compounds **21**-**24**). These contained scaffolds that occurred more than once among the hit compounds and were predicted to still interact with A99 in the RNA binding site. Even though different protomers, tautomerization states and stereoisomers had been considered during docking, it was only possible to obtain racemic mixtures.
On the basis of four selected compounds (Figure 4.4) the choice of the purchased compounds by visual inspection is exemplified in the following. The four compounds all have in common that they were predicted to form Watson-Crick like base pairs to A99, and in addition also should form different interactions with further residues in the binding site.

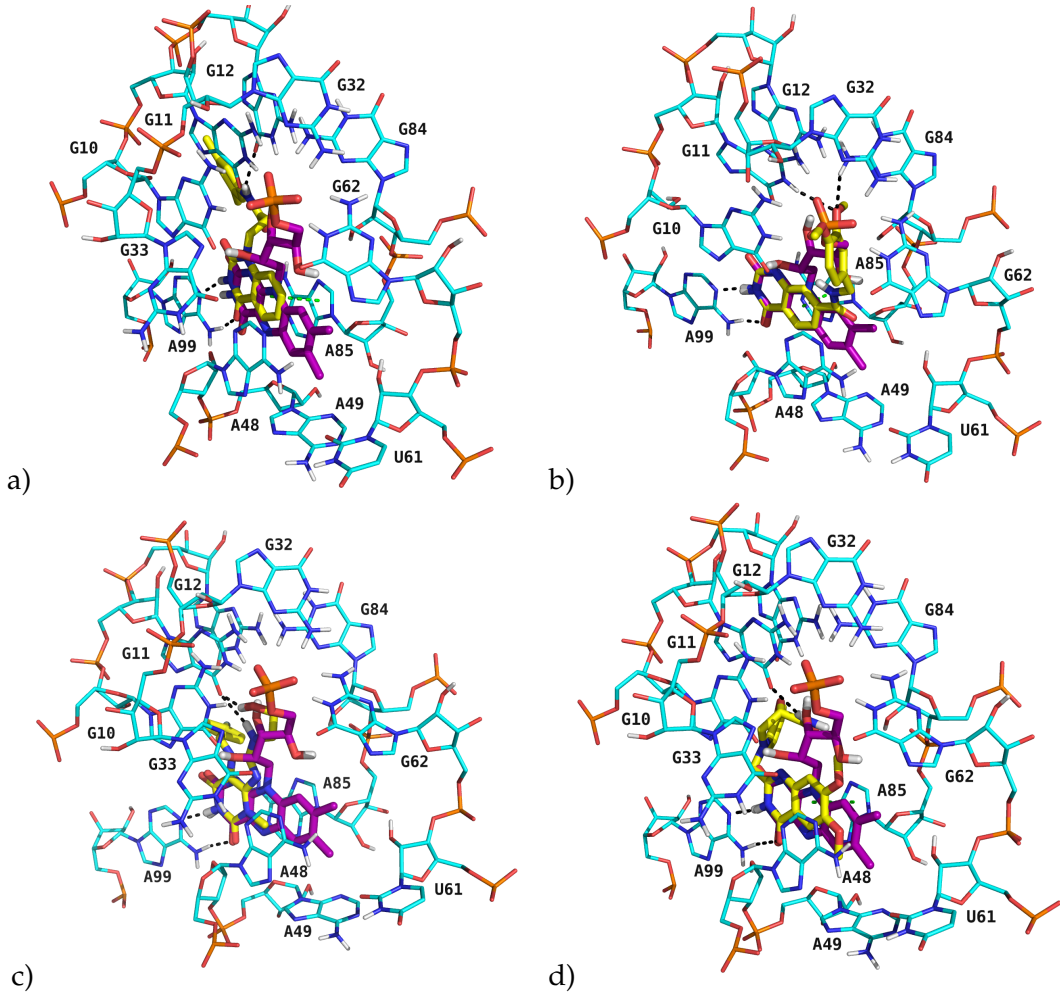


FIGURE 4.4: The predicted binding modes for four selected screening hits (yellow carbon atoms) from a virtual screening of the FMN riboswitch are shown. Key binding site nucleotides are shown with cyan carbon atoms and FMN is displayed with purple carbon atoms as stick representation. Predicted hydrogen bonds are indicated as *black dashed lines*. Stacking interactions are indicated as *green dashed lines*. a-d) Predicted binding modes for compounds **8** (a), **11** (b), **15** (c) and **16** (d).

Compound **8** (Figure 4.4 a) formed π-π-stacking interactions to A85 and allowed for an additional hydrogen bond interaction to the carbonyl function at the 6-position of guanine 12 via its amide linker according to the binding mode prediction.
In the predicted binding mode for compound **11** (Figure 4.4 b) also the potentially formed π-π-stacking interactions to A85 were observed. In addition the compound showed a relatively good shape complementarity to the binding site, since it filled almost the same region as the natural ligand FMN. One methoxy group could serve as an acceptor function for the amino group of G84 and the N1-proton of G11.
For compound **15** (Figure 4.4 c) no π-π-stacking were predicted because of a different position of the heterocycle. However, the ethanolamine group might serve as a bidentate hydrogen bond donor function in the binding site and was predicted to form interactions with the carbonyl group at the 6-position of G11.

The predicted binding mode of compound **16** (Figure 4.4 d) is also worth to be illustrated. Here, the side chain fills similar space as the ribityl moiety of FMN thereby increasing the shape complementarity with the binding site. The terminal amide function might donate a hydrogen bond to the carbonyl group at the 6-position of G11 in a similar manner as observed for the open ribose chain in FMN. Furthermore π-π-stacking to A85 was predicted for compound **20**.

4.2.5 Screening for privileged substructures

In an alternative approach, the in-house database was filtered for compounds containing scaffolds which appear to be privileged for binding to the FMN riboswitch RNA by applying four different SMARTS pattern representing the privileged substructures (Figure 4.5) . A total of 3,936 unique compounds could be assigned to at least one of the SMARTS pattern.

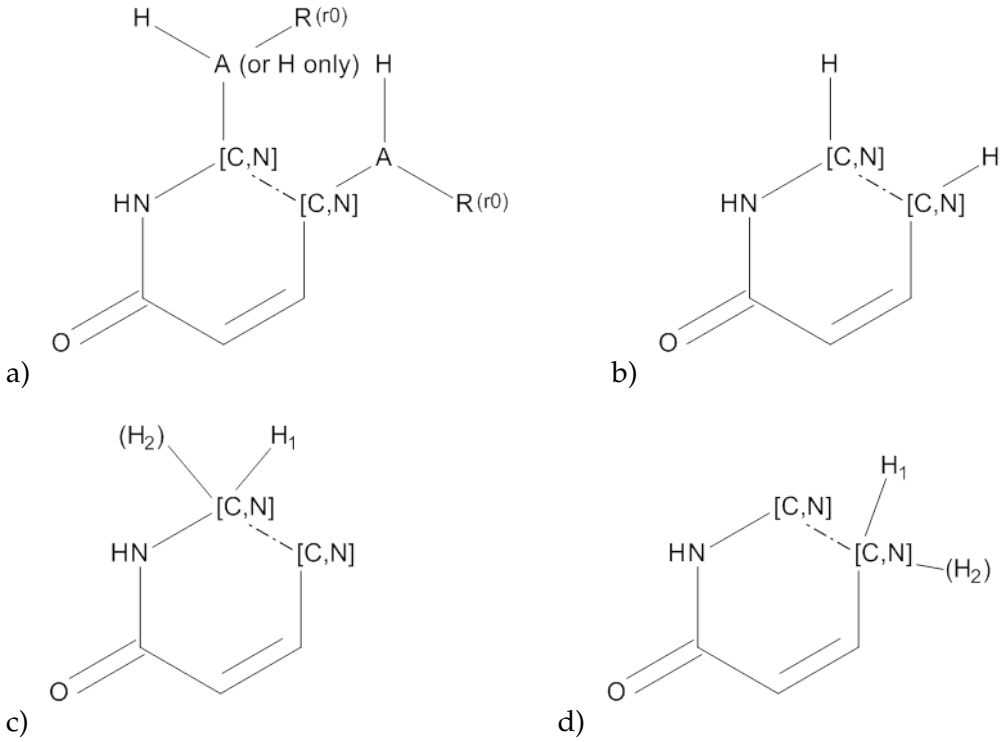


FIGURE 4.5: The 2D-structures of the four different molecular patterns, that were used for the substructure search of the in-house database. a) Display of the SMARTS string c1[C,c](=O)[NH,nH][C,c,n,N]C([1H,!R])~[C,c,n,N]([1H,!R])c1; b) Display of the SMARTS string c1[C,c](=O)[NH,nH][CH,cH,nH,NH]~[CH,cH,nH,NH]c1; c) Display of the SMARTS string c1[C,c](=O)[NH,nH][CH2,CH,cH,nH,NH]~[C,c,n,N]c1; d) Display of the SMARTS string c1[C,c](=O)[NH,nH][C,c,n,N]~[CH,CH2,cH,nH,NH]c1.
(A = any atom except for hydrogen; R(r0) = any atom not part of a ring system; [C,N] = aliphatic or aromatic carbon or nitrogen atom; — · — · — · = any bond)

Those compounds were subsequently filtered for lead-like criteria (Table 4.2), with the only change compared to the whole database screening that a molecular weight cut-off of 450 Da was used. Subsequently, the remaining compounds were docked

into the riboswitch binding site. A total of 3,523 molecules from the database were successfully placed into the binding site of the FMN riboswitch. To aid analysis of the results, the docked compounds were grouped by common core fragments (see 3.2, Figure 3.3 and Table A.1).
After the core fragments had been extracted and the resulting docking scores as well as the generated binding poses had been obtained, compounds were selected for experimental testing. The compounds were shortlisted for purchase based on the following criteria:

1. A total score of -30 kJ/mol or less
2. LE of at least 1.5 $kJ \times mol^{-1} \times HA^{-1}$
3. Watson-Crick-like base-pairing to A99

The predicted binding affinity was expected to be lower for the virtual screening hits of the substructure search compared to the screening of the whole database because of a shift of the molecular weight distribution towards smaller molecular weight compounds due to the adjusted selection criteria. This influenced especially the van-der-Waals term of the scoring function. The LE cut-off was tightened, as a high ligand efficiency (> 1.5 $kJ \times mol^{-1} \times HA^{-1}$) is desirable for a lead optimization.[199]
In total, 21 small molecules were purchased from different suppliers (Table 4.4). As already mentioned in the previous section (4.2.4), all the purchased compounds were only available as racemic mixtures. However, the result table contains the hits in the enantiomeric form with the best pose and score to reflect that this point is of importance. Compound **28b** was purchased in addition to the other 20 compounds as an analogue of compound **28**, even though it was not part of the original in-house database.
In the following paragraphs representative examples of predicted binding modes of substructure search hits are described. Especially binding modes revealing distinctive features besides the minimal requirements to be regarded as screening hits are presented.
Compound **28** contains a three-membered heterocyclic system with similarity to the flavin analogues (Table 4.4) and therefore was predicted to accommodate for both π-π-stacking interactions to A48 and A85. One additional hydrogen bond was postulated to be formed upon binding of this compound. The NH-donor function of the central pyrrole core could interact weakly with the ether function of the ribose sugar of A49 as acceptor of the hydrogen bond (Figure 4.6).

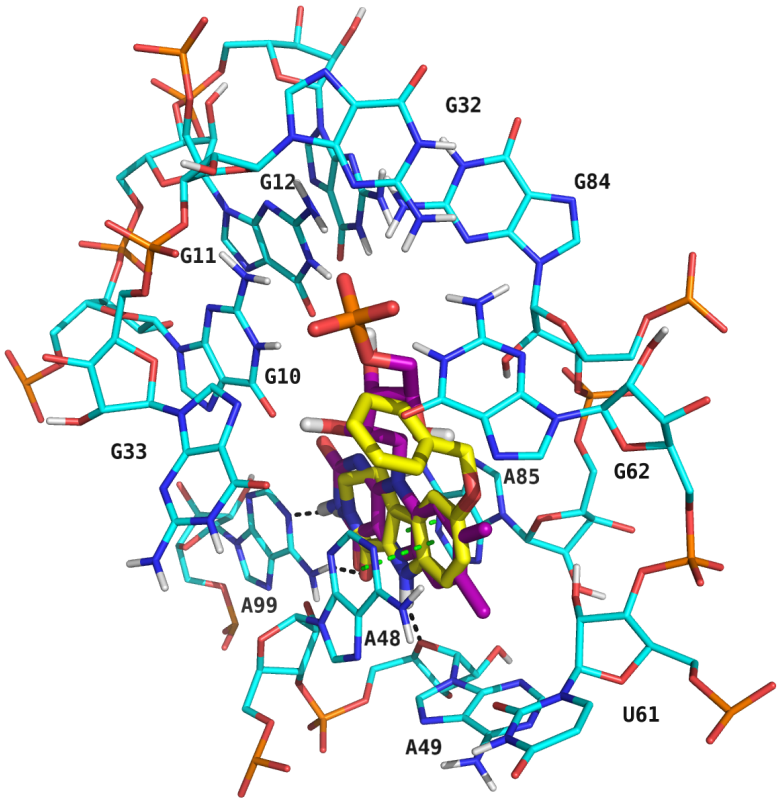


FIGURE 4.6: Overlay of the predicted binding pose of compound **28** (yellow carbon atoms) and FMN with light-grey carbon atoms and transparent sticks surrounded by the binding site nucleotides with cyan-coloured carbon atoms. Interactions between RNA and compound **28** are displayed as *black* and *green dashed lines* for hydrogen bonds and stacking interactions, respectively.

Compound **38** had an interesting predicted binding mode because of the possible two additional hydrogen bonding interactions mediated via the primary amide group in the molecule (Figure 4.7 b). The two-membered heterocycle overlapped nicely with the heterocycle of FMN and permitted the desired hydrogen bonds to A99 as well as π-π-stacking to A85. Furthermore, the terminal amide function was supposed to donate a hydrogen bond to each carbonyl oxygen of the guanine residues G11 and G12. FMN also interacts with the O6-position of G11 via hydrogen bonding mediated by a hydroxyl group and to G12 with its phosphate group.
The binding mode of compound **35** was comparable to the one described above for compound **38** (Figure 4.7 c). Both compounds are very closely related. However, the nitrophenyl ring pointed in a backpocket of the FMN riboswitch RNA, which is not occupied by FMN. The hydroxyl group of the ribose sugar of G93 was proposed to interact with the nitro group of compound **35** serving as hydrogen bond acceptor.
Compound **25** was the compound with the most negative predicted docking score among the purchased hits from the privileged substructure search (Table 4.4) and represented an interesting postulated binding mode. The double hydrogen bonding pattern to A99 was not mediated by an uracil-like edge as observed in the flavin analogues and most of the screening hits. Instead, it was mediated by the two nitrogen

atoms of a 2,3-dihydropyridazin-3-one (Figure 4.7 d). The planar, three-membered heterocycle was predicted to enable stacking interactions to A48 and A85. High shape complementarity was observed in the binding pocket, since the flexible side chain of compound **25** in the predicted binding conformation was positioned similar to the ribityl moiety of FMN. One additional hydrogen bond was predicted for compound **25** besides the double hydrogen bonding pattern to A99. The nitrogen atom of the terminal pyridine ring was proposed to serve as hydrogen bond acceptor and thereby to interact with the exocyclic amino function of G32.

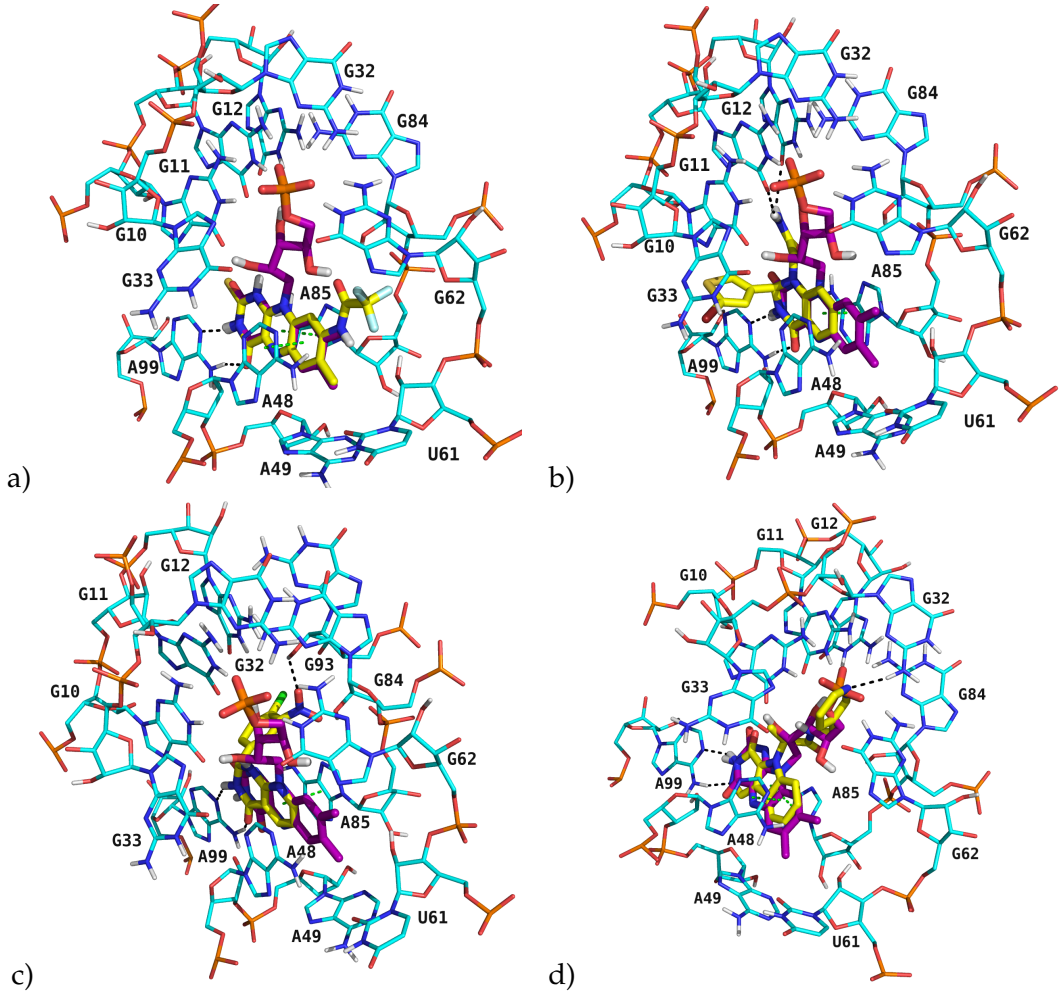


FIGURE 4.7: The predicted binding modes for four selected privileged substructure hits from a virtual screening of the FMN riboswitch are shown. Key binding site nucleotides are shown in cyan carbon atoms and the FMN-ligand is displayed with purple carbon atoms in stick representation. Predictions of the binding modes are colour-coded with yellow carbon atoms and predicted hydrogen bonds are indicated as *black dashed lines*. Stacking interactions are indicated as *green dashed lines*. a-d) Predicted binding modes for the compounds **30** (a), **38** (b), **35** (c) and **25** (d).

TABLE 4.3: Virtual screening hits selected for experimental testing ranked according to their score. The compounds are numbered consecutively and with their internal ID. Assay concentrations used in the fluorescence quenching assay as first hit validation method as well as DMSO stock concentrations are indicated.
(1) Protonation states/enantiomers generating the best docking pose
(2) Fragments of screening hits given
(3) Scores in kJ/mol
(4) Predicted LE in $kJ \times mol^{-1} \times HA^{-1}$
(5) DMSO stock concentration and assay concentration [mM]
NA = not applicable, ins. = insoluble

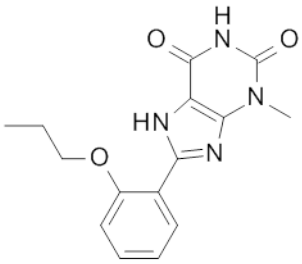
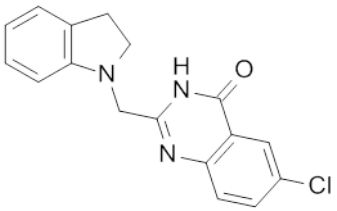
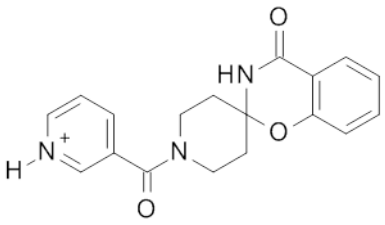
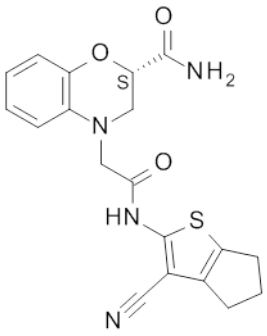
No.	Internal ID	Structure	Score(3)	LE(4)	Stock(5)	Assay(5)
1	Vitas 1		-44.94	2.04	200	0.25
2	Enamine 1		-41.99	1.91	ins.	NA
3	Asinex 1(1)		-41.39	1.72	50	0.5
4	Enamine 2(1)		-41.33	1.53	ins.	NA

TABLE 4.3: continued

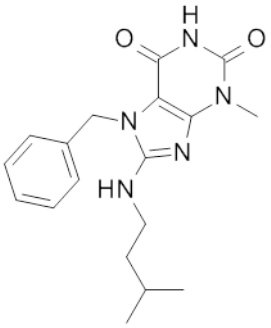
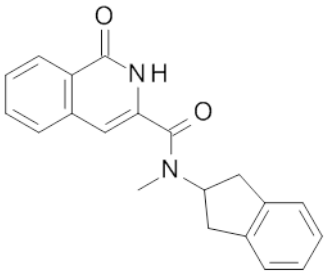
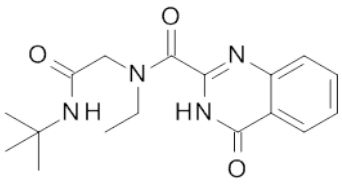
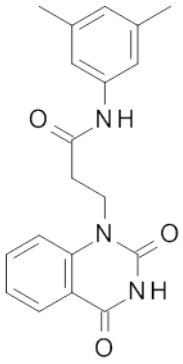
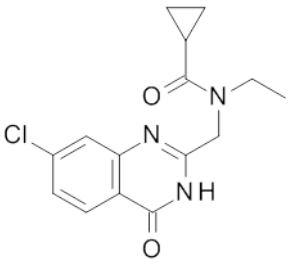
No.	Internal ID	Structure	Score[3]	LE[4]	Stock[5]	Assay[5]
5	Vitas 2		-41.15	1.65	200	1
6	Enamine 3		-41.07	1.71	200	1
7	Enamine 4		-40.86	1.70	200	0.1
8	Enamine 5		-40.66	1.63	200	1
9	Enamine 6		-40.45	1.93	200	2

TABLE 4.3: continued

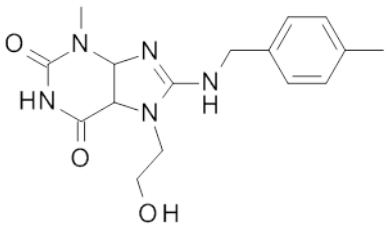
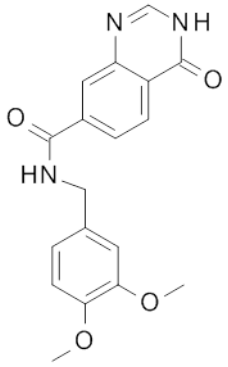
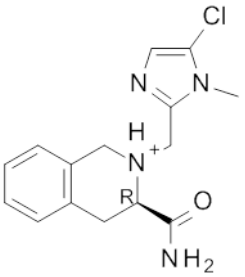
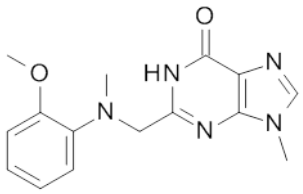
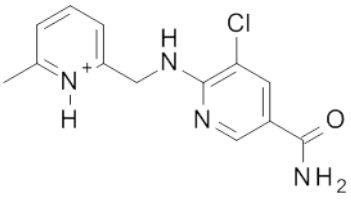
No.	Internal ID	Structure	Score[3]	LE[4]	Stock[5]	Assay[5]
10	Vitas 3		-40.35	1.68	25	0.1
11	Vitas 4		-40.22	1.61	200	0.5
12	Enamine 7[1]		-40.15	1.92	200	2
13	Enamine 8		-40.07	1.82	200	0.5
14	Enamine 9		-39.90	2.10	100	0.5

TABLE 4.3: continued

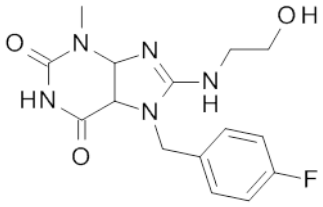
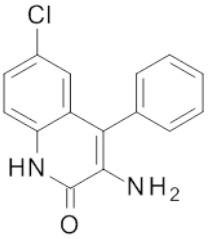
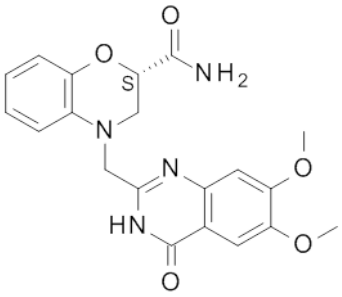
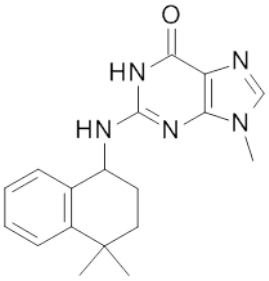
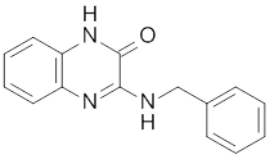
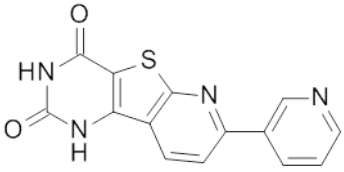
No.	Internal ID	Structure	Score[3]	LE[4]	Stock[5]	Assay[5]
15	Asinex 2		-39.88	1.66	200	1
16	Vitas 5		-39.84	2.10	200	0.5
17	Enamine 10[1]		-39.71	1.37	ins.	NA
18	Enamine 11		-39.17	1.63	200	0.5
19	Vitas 6		-39.10	2.06	200	1
20	Vitas 7		-38.90	1.85	ins.	NA

TABLE 4.3: continued

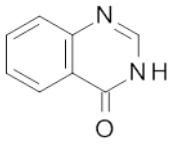
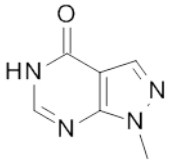
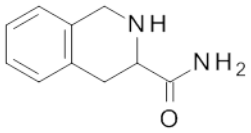
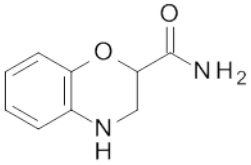
No.	Internal ID	Structure	Score[3]	LE[4]	Stock[5]	Assay[5]
21	Vitas 8[2]		NA	NA	200	1
22	Vitas 9[2]		NA	NA	100	1
23	Vitas 10[2]		NA	NA	100	5
24	Enamine 12[2]		NA	NA	200	10

TABLE 4.4: Virtual screening hits from the substructure search selected for experimental testing ranked according to their score. The compounds are numbered consecutively and with their internal ID and the calculated ligand efficiencies indicated as well.
(1) Enantiomer with the best performance in the docking approach
(2) Analogue of compound **28**
(3) Scores in kJ/mol
(4) Predicted LE in $kJ \times mol^{-1} \times HA^{-1}$
(5) DMSO stock concentration and assay concentration [mM]

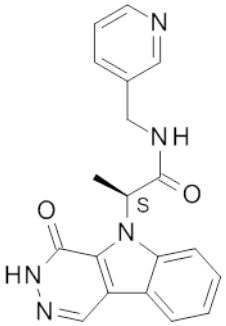
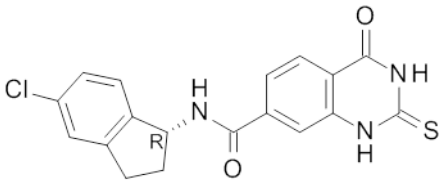
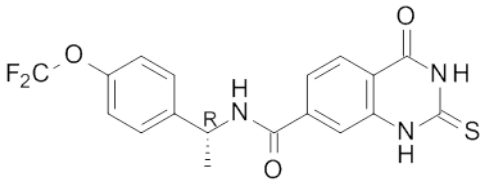
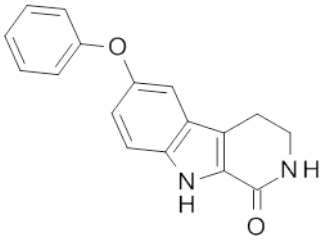
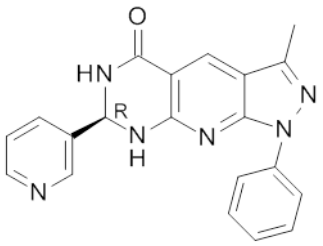
No.	Internal ID	Structure	Score(3)	LE(4)	Stock(5)	Assay(5)
25	Chd 2(1)		-48.70	1.87	50	0.75
26	Ena 5(1)		-47.56	1.90	50	0.05
27	Ena 4(1)		-47.21	1.75	200	0.25
28	Vit 1		-45.58	2.07	200	0.25
29	Specs 3(1)		-45.08	1.67	25	0.1

TABLE 4.4: continued

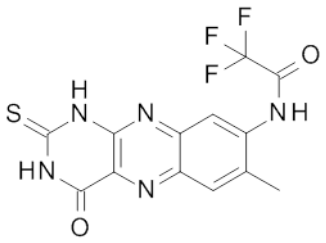
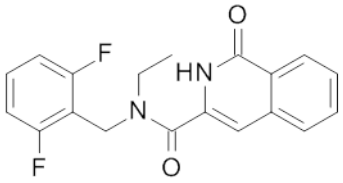
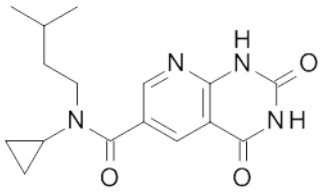
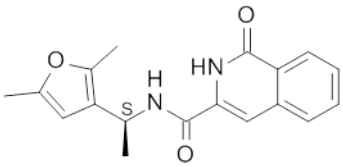
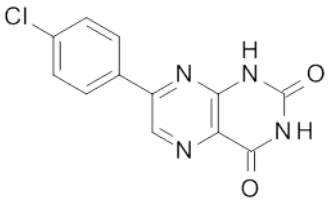
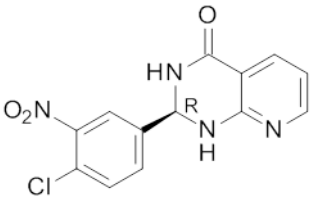
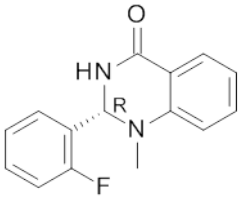
No.	Internal ID	Structure	Score[3]	LE[4]	Stock[5]	Assay[5]
30	Chd 1		-42.87	1.79	100	0.25
31	Ena 7		-42.15	1.69	200	0.5
32	Ena 10		-40.36	1.75	20	0.5
33	Ena 6[1]		-40.03	1.74	200	0.5
34	Specs 2		-37.73	1.99	5	0.1
35	Ena 8[1]		-37.55	1.79	50	0.25
36	Ena 1[1]		-37.31	1.96	200	1

TABLE 4.4: continued

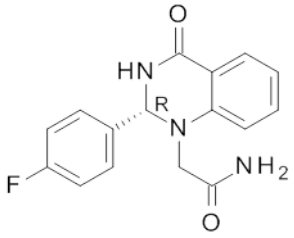
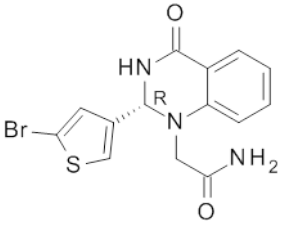
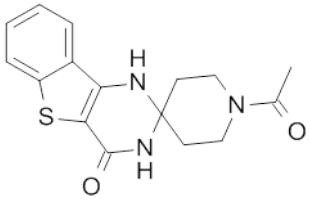
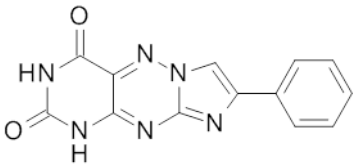
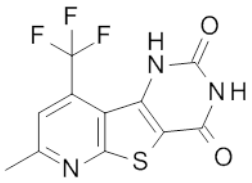
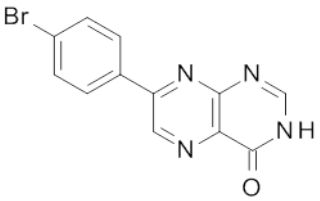
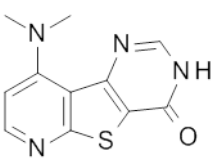
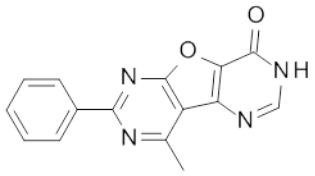
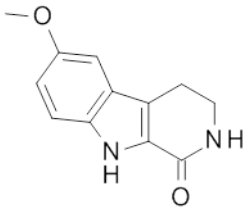
No.	Internal ID	Structure	Score[3]	LE[4]	Stock[5]	Assay[5]
37	Ena 2[1]		-37.29	1.70	200	1
38	Ena 3[1]		-36.93	1.76	200	1
39	Ena 9		-36.75	1.67	40	0.5
40	Vit 2		-35.68	1.70	25	0.25
41	Chbr 1		-34.96	1.75	20	0.5
42	Specs 5		-34.14	1.90	4	0.2
43	Specs 1		-32.74	1.93	12.5	0.3

TABLE 4.4: continued

No.	Internal ID	Structure	Score[3]	LE[4]	Stock[5]	Assay[5]
44	Specs 4		-32.55	1.55	25	0.25
28b	Vit 3[2]		NA	NA	100	1

Chapter 5

Experimental testing of the virtual screening hits

5.1 Abstract

The virtual screening approach described in the previous chapter aimed at the discovery of novel molecular scaffolds that possibly bind to the aptamer domain of the *F. nucleatum* FMN riboswitch. After each of both screenings, the whole database screening and the substructure search, 20 or more individual compounds had been purchased. An initial experiment to validate the virtual screening results was based on the inherent fluorescence of the natural ligand FMN. This method was chosen due to the fact, that its throughput is relatively high and the consumption of purified RNA is rather low. Subsequently, an competitive ITC assay was employed to validate these initial results and to test compounds that were not amenable to the fluorescence-based experimental method. Attempts were also made to develop a SPR method to directly measure the interactions between lower affinity ligands and the riboswitch RNA and the suitability to use a biotinylated two-piece RNA construct for this purpose was shown. Crystals grew under suitable conditions when co-crystallizing FMN together with the RNA. However, so far no crystal structures of the FMN riboswitch complexed with screening hits could be determined.

5.1.1 Validation of the fluorescence quenching assay

The fluorescence quenching assay was based on the intrinsic fluorescent properties of FMN. When a dilution series of the *F. nucleatum* FMN riboswitch RNA aptamer domain is titrated against a constant FMN concentration the fluorescence intensity decreases due to FMN getting buried in the binding site, thus altering its fluorescence properties.[39,49]

First, the assay was validated by measuring the binding affinity of FMN in the absence of a screening hit (Figure 5.1). The best results were obtained with 40 nM ligand concentration (vs. 60 nM in the published assay). The binding affinity was calculated by determining the fluorescence intensities of FMN at 530 nm, after the fluorophore had been excited at 450 nm. The K_D-value determined by plotting the normalized fluorescence intensity of FMN against the RNA concentration and fitting

the data to the appropriate equation (Equation 3.3) was 210.6 ± 34.1 nM, a value in good agreement with the reported results (37.5 nM[39]). The residual normalized fluorescence intensity at saturating RNA concentrations still was about 35% of the initial fluorescence in the absence of RNA (Figure 5.1). This might be attributed either to residual fluorescence of FMN in the bound state or to differently folded RNA species in the mixture with some of them being incapable of ligand recognition so that the measured RNA concentration and the amount of active RNA did not conform to each other. At least the ITC experiments described in a later section of this chapter are indicative of misfolded RNA species and this might also be an explanation for the tendency to higher K_D-values of the interaction between the *F. nucleatum* FMN riboswitch and FMN throughout this work compared to literature values.

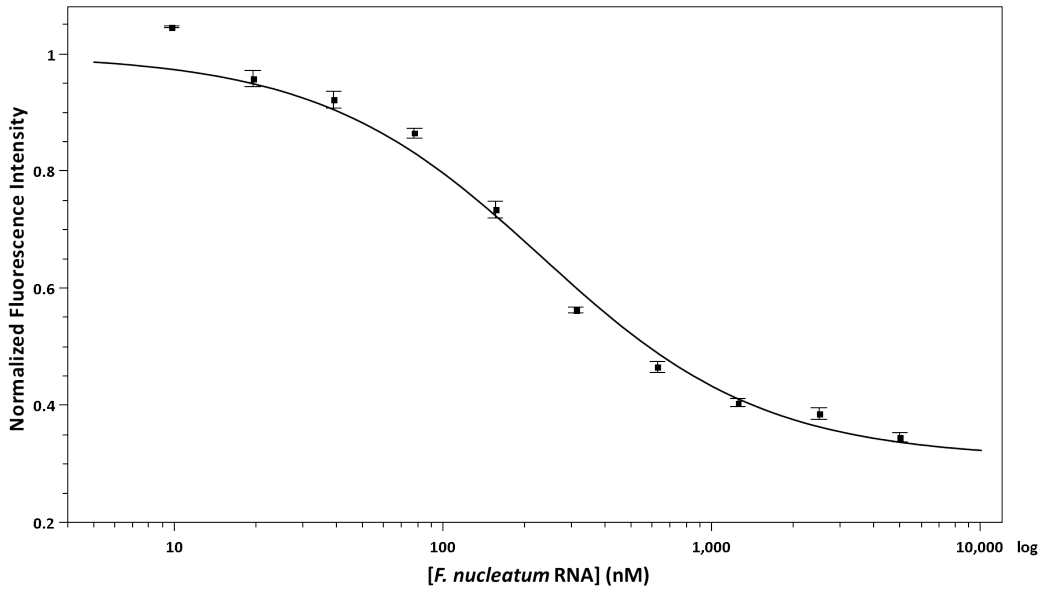


FIGURE 5.1: The buffer-subtracted normalized fluorescence intensity of FMN plotted against the RNA concentration averaged over four consecutive measurements. A *bold black line* represents the best fit to the binding model. Error bars indicate the standard deviation of the normalized fluorescence intensity.

The virtual screening hits were subsequently tested for their binding affinity to the FMN riboswitch via the fluorescence quenching assay. For this purpose a displacement method was chosen. Hence, the K_D-value for the interaction between RNA and its cognate ligand FMN was determined in the absence and presence of test compounds. The theory behind this was that upon binding of a competitor substance to the target RNA, the apparent dissociation constant (K_{Dapp}) for FMN binding to the riboswitch increases, since both compounds compete for the same binding site (Figure 5.2). The difference in K_{Dapp}-values in the presence and absence of a competitor then allows for the determination of the dissociation constant of the test compound (Equation 3.4).

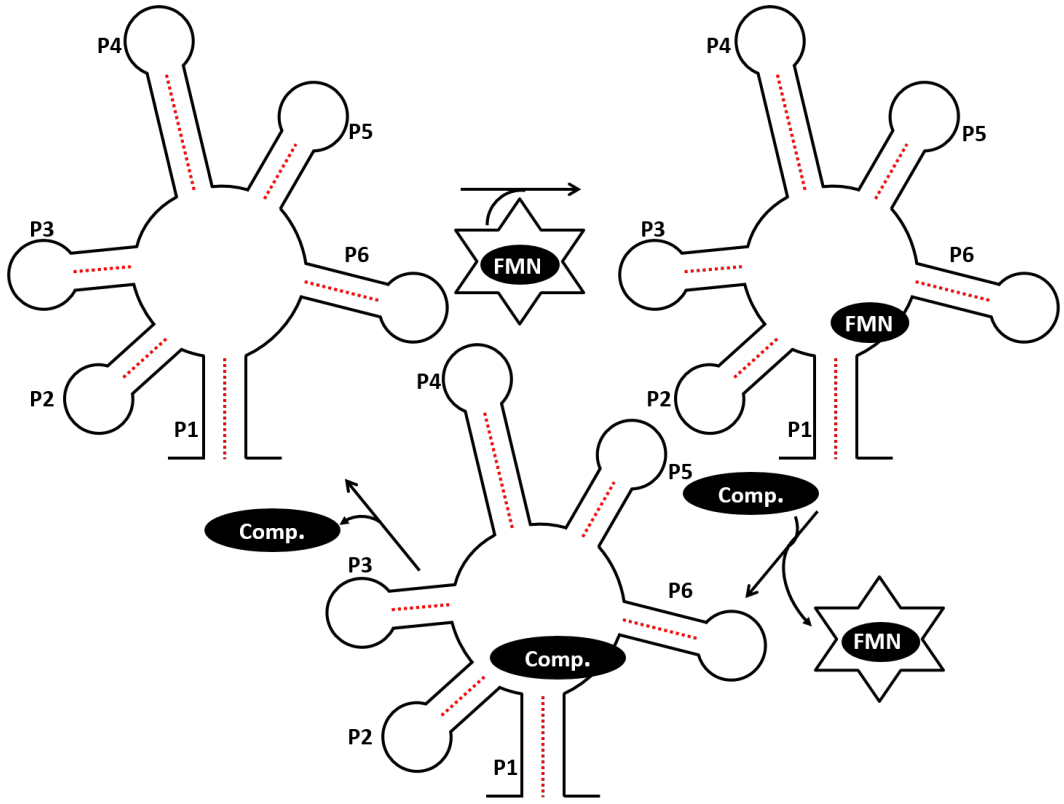


FIGURE 5.2: Schematic circle showing the competition between a screening hit (Comp.) and FMN for the riboswitch binding site. The *star shape* around FMN indicates its fluorescence when it is in the unbound state. P1-P6 represent the six different stems the *F. nucleatum* FMN riboswitch is composed of. To illustrate the restored fluorescence of FMN after displacement FMN is displaced here. During the performance of the assay the competitor was added at the same time and could also bind directly to free RNA rather than displacing FMN first.

5.1.2 Fluorescence quenching assay of the virtual screening hits

Initially, the screening hits of the first molecular docking approach employed on the whole in-house database were subjected to the fluorescence quenching assay (Table 4.3). It turned out that the compounds **2**, **4**, **17** and **20** were not sufficiently soluble in DMSO (Table 4.3). Therefore, these four compounds were not further pursued.
Two other screening hits, compound **1** and compound **5** (Table 4.3), strongly interfered with the assay due to autofluorescent properties.
The remaining 18 compounds could be tested using the fluorescence quenching assay. None of the purchased compounds increased the apparent K_D-value of the riboswitch-ligand interaction with the natural ligand FMN at least twofold at a compound concentration of at least 100 µM. Exemplary, the assay performance of compound **6** is highlighted here. Compound **6** was one of the top-scoring compounds with a total docking score of -41.07 kJ/mol (Table 4.3) and did not face problems with either aqueous solubility or assay interference. The assay results in the presence and absence of the test compound at a concentration of 1 mM indicated, that compound **6** did not affect the interaction (Figure 5.3) between RNA and FMN. Both binding curves almost overlapped, had the same shape and the fitted K_D-values were 220.6

± 24.4 nM in the absence and 293.6 ± 28.6 nM in the presence of this compound when averaged over 5 consecutive measurements.

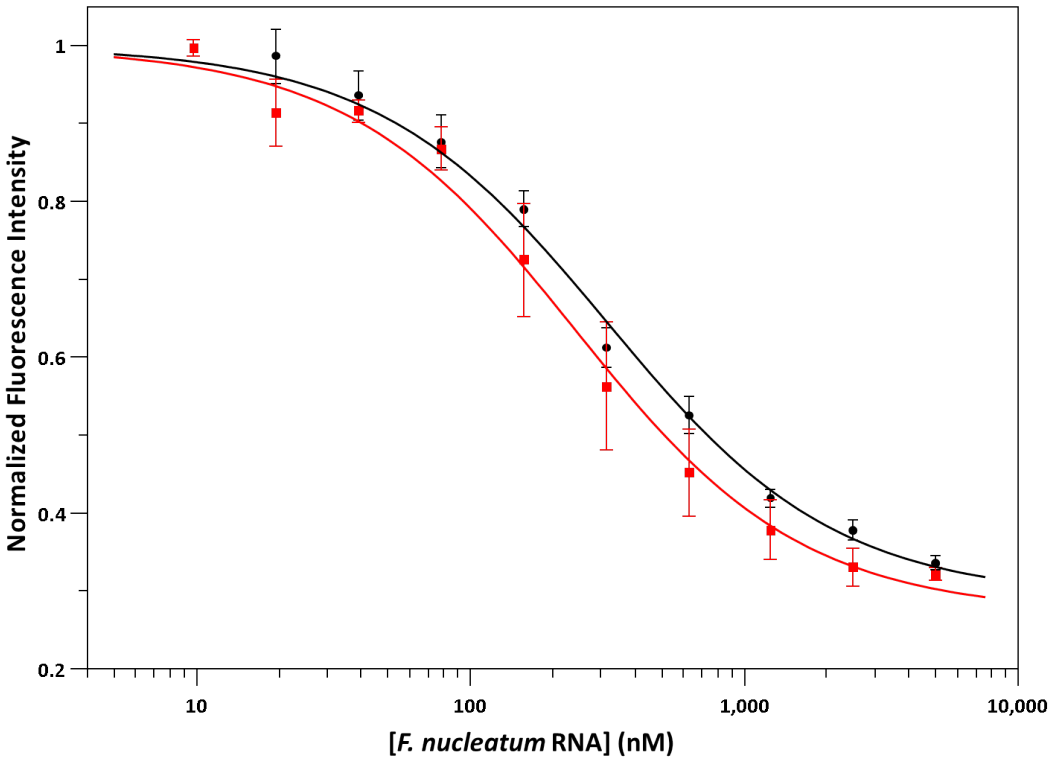


FIGURE 5.3: The normalized and buffer-subtracted fluorescence intensities were plotted against the *F. nucleatum* RNA concentration in a semilogarithmic plot in the absence and presence of compound **6**. A *bold black line* represents the best fit for the interaction in the presence of the particular compound and the *bold red line* represents the best fit for the control reaction averaged over five measurements. Error bars indicate the standard deviations of the normalized fluorescence intensities.

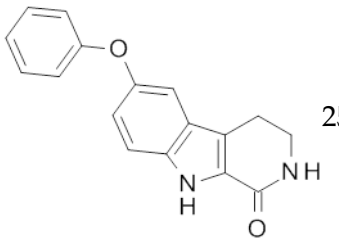
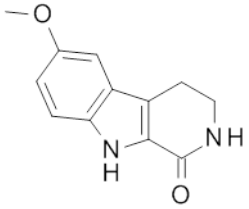
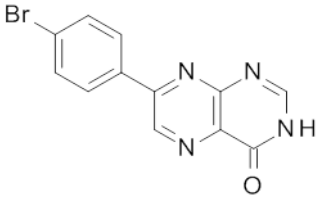
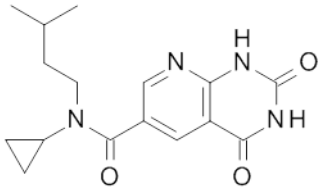
Conclusively, the initial round of testing with the fluorescence-based assay did not identify hit compounds that induced a shift in the binding curves of the interaction between FMN and riboswitch RNA. This applies at least to the compounds that were actually tested successfully. The two compounds interfering strongly with the assay (compounds **1** and **5**, 5.1.2) were still left to be tested by orthogonal assay methods. However, these negative initial results led to the decision to carry out the second virtual screening approach with the results being described in the previous chapter (4.2.5). Thus, the hits of the privileged substructure-based virtual screening were subjected to the fluorescence quenching assay afterwards.

Comparable to the results of the fluorescence quenching assay with the screening hits from the first virtual screening not all compounds were amenable to testing by this method. Even more compounds interfered with the assay. The compounds **30**, **41**, **40**, **34** and **44** all showed high fluorescence read-outs and therefore no K_D-values for the FMN-RNA interaction could be determined in their presence.

Two virtual screening hits, compounds **28** and **42**, from the privileged substructure search showed the largest increase in the apparent dissociation constant between the RNA and FMN when present in the reaction mixtures. The associated K_D-value for a competitor compound with binding affinity to the FMN riboswitch was calculated by an equation originally developed for ITC displacement assays by Zhang and Zhang (Equation 3.4).[182]

In the presence of compound **42**, the K_{Dapp} for the interaction between FMN and riboswitch RNA increased about twofold (Table 5.1). The binding curve was shifted to the right-hand side upon addition of this compound, too (Figure 5.4 a). The K_D-value was calculated to be $\approx$ 142 µM for the competitor compound **42**.

TABLE 5.1: Summary of the fluorescence quenching assay results for four selected compounds described in the results part. The apparent K_D-values for FMN binding to the riboswitch RNA in the absence (-) and presence (+) of the respective test compounds are indicated together with a calculated K_D-value for the competitor substance when applicable. NA = not applicable

Cpd.	Structure	K_{Dapp}(-)[nM]	K_{Dapp}(+)[nM]	$K_D^{complig}$[µM]
28		256.4.4 ± 62.1	490.4 ± 108.2	274.0
28b		168.1 ± 13.2	238.9 ± 50.9	NA
42		138.5 ± 47.7	333.8 ± 55.6	141.8
32		167.0 ± 15.1	226.5 ± 40.4	NA

The second compound that triggered a pronounced increase of the K_{Dapp}-value for the control interaction between RNA and cognate ligand of about twofold was compound **28**. After fitting the data of both curves in the absence and presence of compound **28**, the K_D-value for this competitor compound was calculated to $\approx$ 274 µM (Table 5.1). The inspection of the binding curves in the presence and absence of compound **28** also clearly indicated the shift to higher RNA concentrations (Figure 5.4 b).

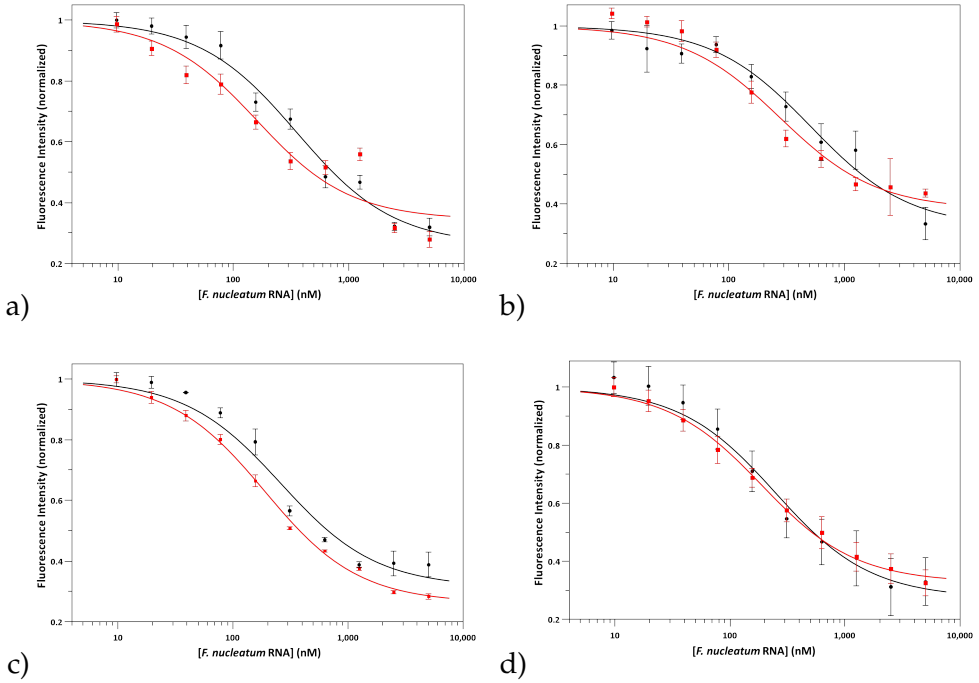


FIGURE 5.4: The normalized and buffer subtracted fluorescence intensities were plotted against the *F. nucleatum* RNA in a semilogarithmic plot. A solid, black line represents the best fit for the interaction in the presence of the particular compound and the solid red line represents the best fit for the control reaction averaged over five meaurements. Error bars indicate the standard deviations of the normalized fluorescence intensities. a) Results for compound **42** averaged over 9 measurements. b) Results for compound **28** averaged over 15 measurements. c) Results for compound compound **28b** averaged over 4 measurements. d) Results for compound **32** averaged over 40 measurements.

In the list of the purchased screening hits the close analogue of compound **28**, compound **28b**, was also included. Compound **28b** possesses a methyl group instead of the additional phenyl ring of the diphenyl ether partial structure in compound **28** (Table 4.4). Results of the fluorescence quenching assay indicated only negligible effects on the riboswitch-ligand interaction with a small increase of the K_{Dapp} (Table 5.1 and Figure 5.4 c). The difference was much smaller than a twofold increase and therefore the compound was considered as a non-binder despite the close structural relationship to compound **28** which led to a more pronounced increase of the K_{Dapp}-value.

For compound **32** initial assays revealed that this compound possibly exhibits activity against the *F. nucleatum* FMN riboswitch. The apparent dissociation constant

increased about twofold in its presence. However, the first results in the presence of compound **32** could not be confirmed by further testing and rather ambiguous results were obtained. Therefore, the experiments were repeated in large numbers to recuperate more confidence in the results. Averaged over 40 measurements only marginal changes of the apparent K_D-values in the absence and presence of compound **32** were detected (Table 5.1). Similar to compound **28b** it was therefore not possible to calculate a K_D-value for this compound. Also the binding curves (Figure 5.4 d) indicated that compound **32** did not cause a shift to higher RNA concentrations. This led to the conclusion that compound **32** is a non-binder.

5.2 Results of the ITC displacement assay

The ITC displacement assay was carried out to confirm hit substances arising from the fluorescence quenching assay and to determine the dissociation constants for compounds that interfered strongly with the former assay due to their autofluorescent properties. A direct ITC approach was expected to be inadequate, since the possible ligands presumably had a weak affinity towards the target structure and complete sigmoidal-shaped binding curves would only result when using prohibitive high concentrations of RNA.

5.2.1 Experimental results of the feasibility study

Before experiments were performed in the presence of virtual screening hit compounds in a competitive manner, direct measurements by titrating FMN into the cell containing riboswitch RNA only were carried out. Generally, ITC measurements are suitable to measure interactions between small molecules and RNA including riboswitches. Often ligand binding induces folding events coupled to large enthalpies that can be easily identified as evolving heat signals.[200]

The results of the control measurements in the absence of test compounds correlated well with previous experiments. According to different amounts of DMSO needed to solubilize the particular compounds, different control reactions were necessary to account for small changes in the heat of dilution. With increasing DMSO concentrations stronger dilution effects occurred, which became mainly apparent during the last injections of ligand. In the first attempts the screening hits were typically diluted 1:50 in the ITC assay buffer. For that reason the results for the control reaction are exemplified with a 2% DMSO concentration. Due to the multiple hydrogen bonding interactions, large enthalpies of -79.24 $\pm$ 7.55 kJ/mol were found averaged over 25 titrations (Figure 5.5). The entropic penalty was rather high with +40.39 $\pm$ 7.92 kJ/mol and applying the Gibbs-Helmholtz equation (3.6) for the calculation of the binding free energy resulted in ΔG = -38.85 kJ/mol. The *n*-value, a measurement of the stoichiometry of the present interaction, was determined to be 0.55 on average (25 titrations).

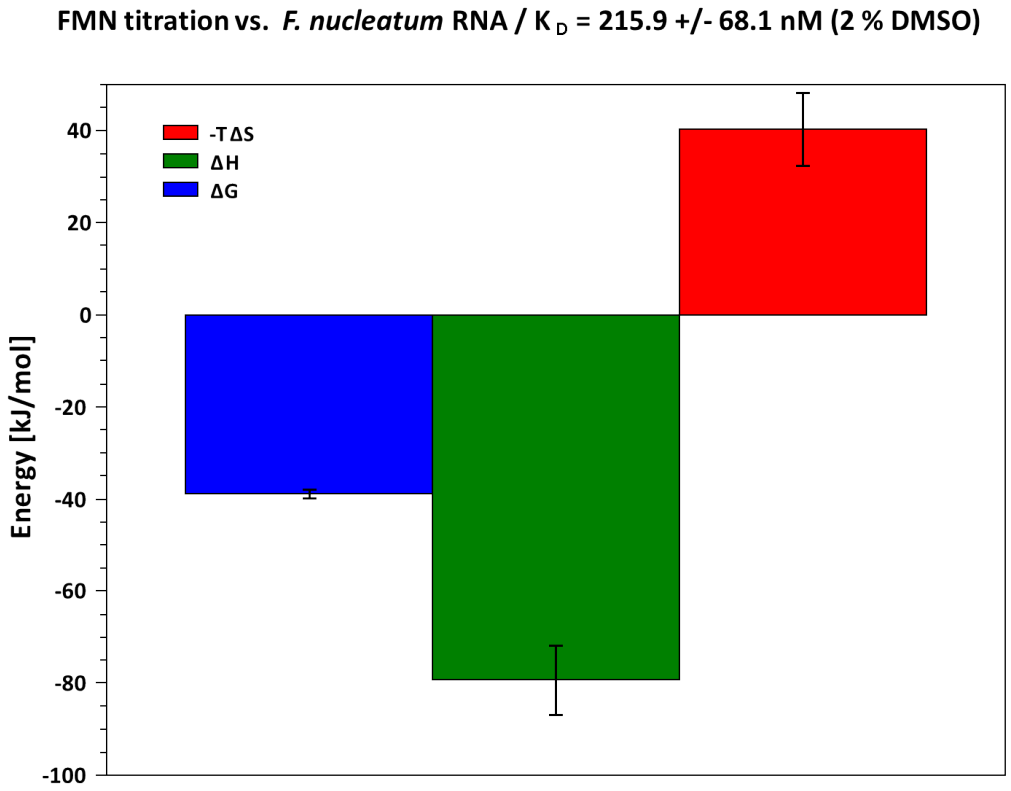


FIGURE 5.5: Bar plot illustrating the energetic contributions to ligand binding over 25 measurements. ΔG is represented as blue bar, ΔH is represented as green bar and the -TΔS term is shown as red bar. Error bars indicate the standard error of the mean.

The average K_D-value of 215.9 $\pm$ 68.1 nM obtained for the interaction of the *F. nucleatum* FMN riboswitch and its natural ligand was close to the one that had been determined by the fluorescence quenching assay earlier ($\approx$ 210 nM, 5.1.1).
For an average measurement with three repetitive titrations during one day, prepared from the same stock solutions, the deviations in terms of measured K_D-values and heat released upon binding of FMN to the riboswitch RNA were less pronounced and the data could be fitted well to a 1:1 binding model (Figure 5.6). When investigating the enthalpy plotted versus the molar ratio, it again became evident that possibly not 100% of the RNA species were competent for ligand binding. Saturation occurred at molar ratios of about 1 and even at the end of the titration the molar ratio was smaller than 2 with the chosen setup. Nevertheless, the similar binding affinity resulting from this method in comparison to the fluorescence-based assay led to the assumption that a change of the affinity from FMN to the RNA in the presence of test compounds should be detectable by ITC experiments.

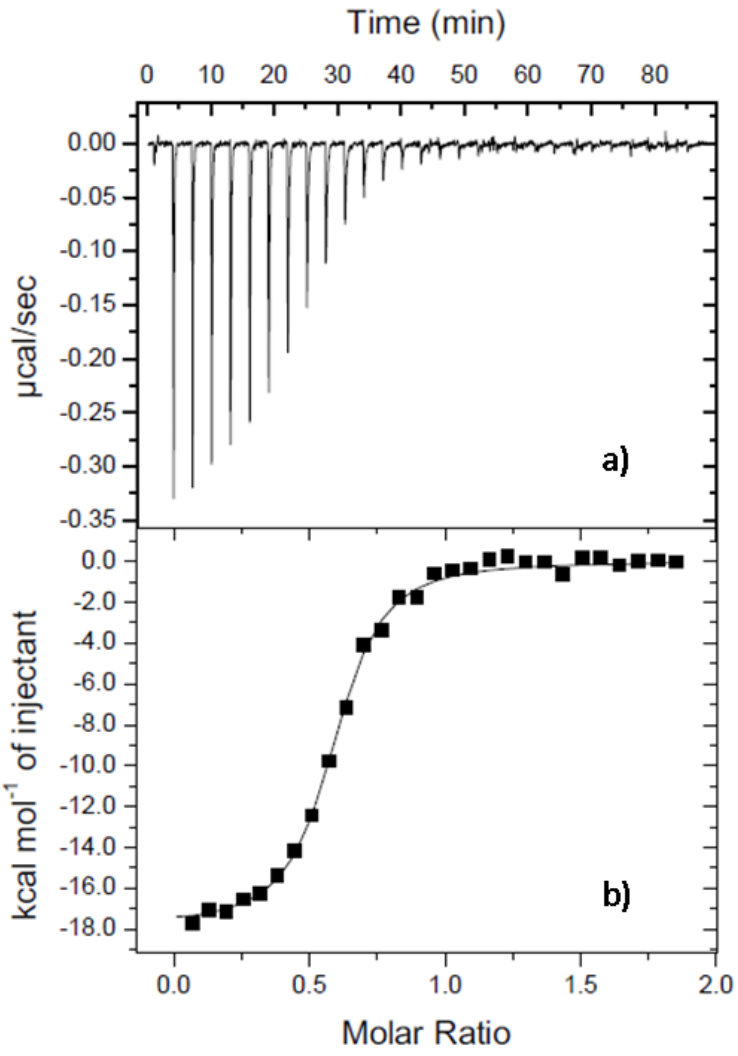


FIGURE 5.6: a) Plot of power versus time for a titration of 105 µM of a FMN solution into 15 µM of FMN riboswitch RNA. b) The injection enthalpy plotted versus the molar ratio. Fitting the data resulted in ΔH = -17.85 $\pm$ 0.18 kcal/mol, K_A = 4.57 $\times$ 10^6 M^{-1}, ΔS = -0.0284 $\frac{\text{kcal}}{\text{mol}\times\text{K}}$ and n = 0.575.

5.2.2 Results of the ITC displacement experiments

The buffers used for the ITC experiments and the fluorescence quenching assay only differed in terms of the buffering component. Tris was changed against HEPES. Previous work on investigating riboswitch-ligand interactions by ITC had also been conducted in HEPES-based buffer systems.[200] Despite the knowledge already gained regarding the solubilities of the virtual screening hits in aqueous buffers, the initial solubility tests were repeated. The high stirring speed during the ITC measurements might influence the observations and in contrast to the dark 96-well plates in the fluorescence assays precipitation was much easier detected in the larger volumes necessary for ITC. Due to this fact, the concentrations for some of the compounds had

to be adjusted compared to the fluorescence-based assay (Table 5.2).
Out of the nine compounds that were subjected to the ITC displacement assay, eight compounds did not alter the binding affinity of FMN to the FMN riboswitch. A detailed overview of the K_D-values determined in the presence and absence of test compound as well as the different enthalpies, entropies and free energies calculated for the FMN-RNA interaction is given in the corresponding table (Table 5.2). This in turn means that even in the presence of the most promising candidate compounds from the fluorescence quenching assay, compounds **42** and **28**, the measured dissociation constants did not change upon addition of the respective screening substance when considering the error size of the determined K_D-values. For example, for compound **42** very similar energetic terms for the binding interaction (Table 5.2) in the presence and absence of compound were obtained. In addition, also a very similar enthalpy value was determined in the presence of compound, again clearly indicating the compound as a non-binder. Both binding curves had the same slope and superimposed very well (Figure 5.7 a) when averaged over three consecutive titrations each. For the other initial hit compound of the fluorescence quenching assay, compound **28**, similar results were obtained (Figure 5.7 b). For compound **30** the measurements in the presence of compound failed and no fit for the binding curves could be obtained. The reason for that was unclear.
In contrast to the results of the initial hits identified by the fluorescence quenching assay, one of the screening hits arising from the substructure search that was not amenable to the fluorescence-based assay was active in the ITC displacement assay. In the absence of the compound **44** the dissociation constant for the FMN-riboswitch interaction was determined to be 120.9 $\pm$ 34.6 nM. The K_D-value in the presence of compound **44** increased to 244.0 $\pm$ 20.9 nM resulting in a dissociation constant of $\approx$ 240 µM for compound **44** binding to the *F. nucleatum* FMN riboswitch aptamer domain. Furthermore, the increase of the K_{Dapp} for the FMN-RNA interaction ($\Delta\Delta K_{Dapp}$=123.1 nM) was > 3 SDs (=103.8 nM) of the mean dissociation constant for the control interaction over these consecutive measurements with compound **44**, which serves as an indicator of a significant change.[201] In the presence of this compound, the steepness of the FMN binding curve was lowered compared to the uninhibited interaction between FMN and the RNA structure (Figure 5.7 c). This fact underlines the weaker affinity of FMN to the RNA in the presence of compound **44**.[202] Furthermore, the binding affinities differed the most in the presence of compound **44** compared to all other compounds that were subjected to the ITC displacement assay. The apparent dissociation constant in the presence of any test compound was increased less than 3 SDs in all other cases.

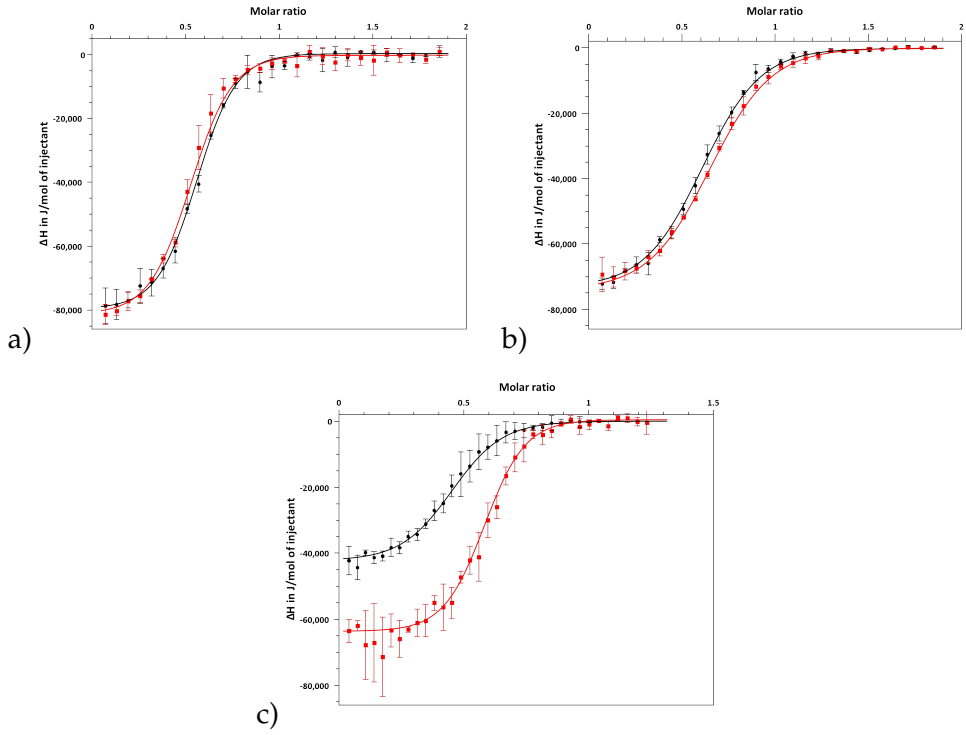


FIGURE 5.7: ΔH released upon binding of FMN to the RNA was plotted against the molar ratio of the ligand and RNA concentrations. For the illustration of the results a simple sigmoidal fit was carried out within the program QtiPlot. The particular resulting binding constants of the table below (Table 5.2) were calculated directly with the software accompanying the instrument. The *bold red line* represents the fit for the control data set and the *bold black line* the fit for the data in the presence of each screening compound. a) Results for compound **42** averaged over three measurements. b) Results for compound **28** averaged over three measurements. c) Results for compound **44** averaged over three measurements.

Without the addition of compound **44** a ΔH-value of -67.08 ± 6.96 kJ/mol was determined, whereas this value decreased to -43.61 ± 1.85 kJ/mol in its presence. This corresponds to a considerable drop of about 35%. These findings are consistent with a binding event in a competitive manner. Already from the beginning of the titration some binding sites in the riboswitch binding pocket were blocked by the competitor and therefore less heat was released upon FMN binding. Since binding to the FMN riboswitch RNA was detected for compound **44**, a calculation of all energetic terms of binding was possible (Equation 3.6 and 3.7). Thus, the binding free energy for compound **44** was calculated to be -21 kJ/mol at a ΔH-value of -39.34 kJ/mol and an entropic penalty -TΔS of +18.34 kJ/mol.

TABLE 5.2: The binding affinities and the energetic terms for FMN binding to the riboswitch RNA derived from the fit to a 1:1-binding model are displayed. Each row represents the average results for three measurements in the presence (+) and absence (-) of the compounds using the same batch of RNA for control measurement and measurement in the presence of a test compound. The change of the apparent dissociation constant is indicated as x-fold increase of the apparent $K_{D(FMN)}$-value in the absence of compound.
[a] apparent dissociation constants for the FMN-riboswitch RNA interaction are given in nM concentration
[b] enthalpy, entropy and free energy of binding are given in kJ/mol

	- Test cpd. absent					**+ Test cpd. present**					
Cpd./[µM]	$K_{D(FMN)}$[a]	**n-value**	ΔH[b]	$T\Delta S$[b]	ΔG[b]	$K_{D(FMN)}$[a]	**n-value**	**x-fold ($K_{D(FMN)}$)**	ΔH[b]	**-$T\Delta S$**[b]	ΔG[b]
25/250	293.8±26.0	0.475±0.025	-86.65±1.60	-48.72±1.78	-37.93±0.21	302.9±30.6	0.482±0.018	1.03	-79.86±4.72	-42.00±4.82	-37.86±0.25
28/200	397.5±15.8	0.651±0,007	-72.63±3.26	-35.50±3.17	-37.13±0.09	385.4±17.4	0.605±0.025	0.97	-74.25±2.31	-37.10±2.33	-37.15±0.14
30/750	295.1±16.7	0.659±0.038	-76.83±2.56	-38.92±2.71	-37.91±0.16	n.d.	n.d.	n.d.	n.d.	n.d.	n.d
32/250	261.6±45.3	0.610±0.024	-78.95±3.72	-40.69±4.14	-38.26±0.41	252.0±9.2	0.629±0.006	0.96	-70.61±0.46	-32.33±0.41	-38.28±0.05
34/50	105.7±20.2	0.526±0.019	-91.71±2.26	-51.15±2.72	-40.56±0.47	122.9±7.75	0.589±0.008	1.16	-75.48±2.38	-35.37±2.56	-40.11±0.18
40/150	320.5±12.8	0.592±0.013	-79.56±1.44	-41.90±1.33	-37.66±0.11	374.2±23.8	0.632±0.016	1.17	-64.78±2.80	-27.47±2.64	-37.31±0.17
41/250	314.1±14.0	0.642±0.015	-68.17±1.63	-30.42±1.67	-37.75±0.11	358.4±17.6	0.677±0.010	1.14	-55.45±0.63	-18.04±0.75	-37.41±0.12
42/100	159.4±37.1	0.360±0.024	-88.44±3.69	-48.94±4.03	-39.50±0.56	162.4±29.5	0.347±0.025	1.02	-82.70±2.31	-43.23±2.52	-39.47±0.45
42/200	213.7±29.3	0.501±0.026	-82.59±3.11	-43.82±3.29	-38.77±0.31	241.8 ±67.9	0.544±0.014	1.13	-80.47±4.84	-41.92±5.24	-38.55±0.75
44/250	**120.9**±34.6	0.609±0.009	-67.08±6.96	-26.88±7.67	-40.21±0.71	**244.0**±20.9	0.532±0.017	**2.03**	-43.61±1.85	-5.21±1.68	-38.39±0.22

5.2.3 Follow-up ITC displacement experiments with analogues of the hit compound

Based on the ITC results for compound **44**, it was decided to purchase close analogues of compound **44** (Table 5.3, the compounds **45-49**), with the aim of starting a small structure-activity relationship study. More precisely, it was tried to obtain compounds with the same three-membered heterocycle as compound **44** to test, whether or not different substitutions are tolerated by the FMN riboswitch binding pocket. The part of compound **44** that was predicted to interact via hydrogen bonding with A99 was left unchanged (Figure 5.8). In the predicted binding mode of compound **44** the pyrimidine-4(3*H*)-one part forms hydrogen bonds with A99. In addition, face-to-face π-stacking to A85 is predicted (Figure 5.8).

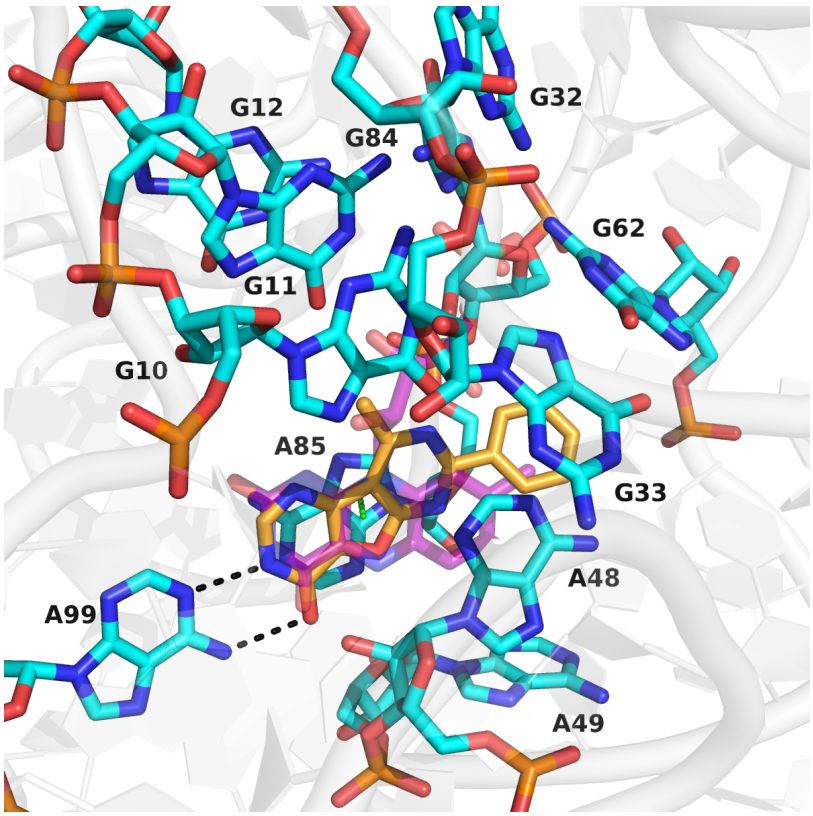
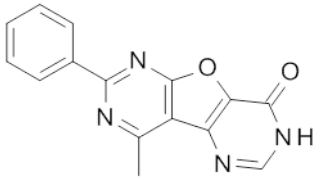
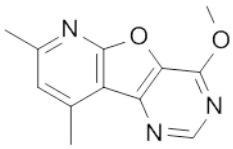
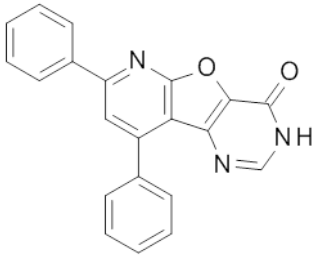
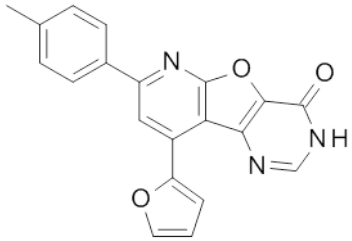
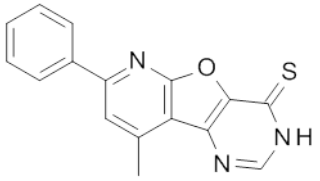
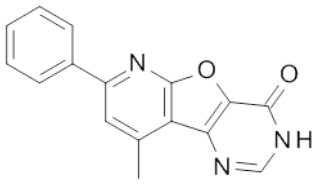


FIGURE 5.8: The predicted binding mode of compound **44** in the FMN riboswitch RNA binding pocket. Nucleic acid residues are shown as *grey cartoon* with binding site residues being displayed as *sticks* with cyan-coloured carbon atoms. Compound **44** is represented with orange-coloured carbon atoms overlaid with FMN (*transparent sticks*, purple carbon atoms) for better orientation. Hydrogen bonds are displayed as *black dashed lines* while stacking interactions are shown with *green dashed lines*.

However, analogues with exactly the same heterocyclic core of compound **44** were not available at that time. The five different analogues purchased to conduct the small SAR-study all had a pyrido[3′,2′:4,5]furo[3,2-d]pyrimidine-4(3*H*)-one core containing one less nitrogen atom than the parent compound. However, this additional nitrogen atom was not predicted to interact with the RNA binding pocket. The analogous compound **45** was chosen as a negative control. Due to the O-methylation in

the pyrimidine ring it was expected to not strongly interact with the riboswitch and therefore to be inactive in the assay.

TABLE 5.3: Five close analogues of compound **44** were purchased after identifying the parent compound as a hit substance in the ITC displacement assay. DMSO stock concentrations and ITC assay concentrations are both indicated.

Cpd.	Structure	DMSO stock[mM]	Assay-conc.[µM]
44		25	250
45		25	100
46		25	50
47		5	250
48		25	100
49		25	50

The fitted data for the analogues of compound **44** in the ITC displacement assay revealed some worth mentioning results (Table 5.4). As expected, in the presence of compound **45** the K_D-value did not increase and also the slope and shape of the binding curves indicated that no binding occurred (Figure 5.9 a). Despite the high structural similarity of the compounds **46**-**49** to compound **44** binding could not be detected using the ITC approach (Figure 5.9 b-e) at first glance.

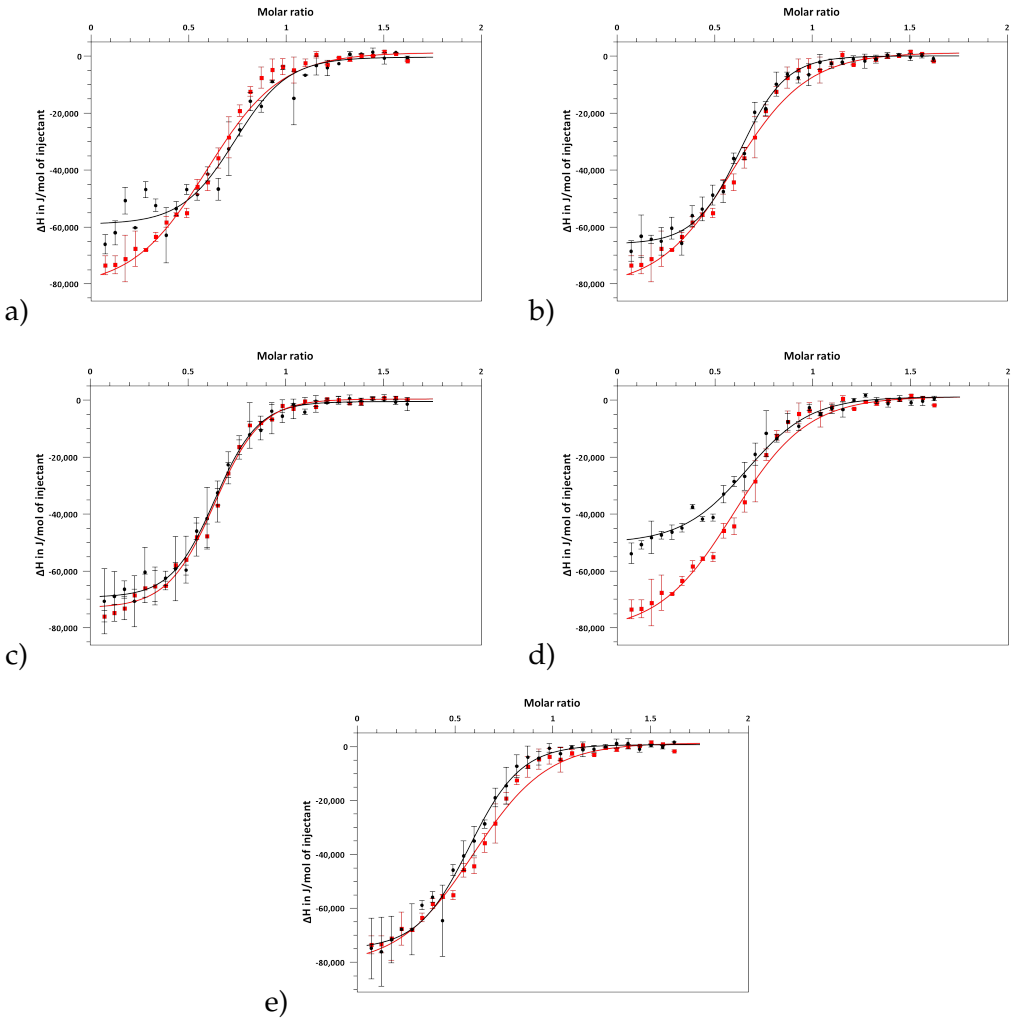


FIGURE 5.9: ΔH released upon binding of FMN to the RNA was plotted against the molar ratio of the ligand and RNA concentrations. For the illustration of the results a simple sigmoidal fit was carried out within the program QtiPlot. The particular resulting binding constants of the table below (Table 5.2) were calculated directly with the software accompanying the instrument. The *bold red line* represents the fit for the control data set and the *bold black line* the fit for the data in the presence of each screening compound. a) Results for compound **45** averaged over three measurements. b) Results for compound **46** averaged over three measurements. c) Results for compound **47** averaged over three measurements. d) Results for compound **48** averaged over three measurements. e) Results for compound **49** averaged over three measurements.

The explanation for compounds **46** and **47** might be the additional aromatic substitution of the pyridine ring in the heteroaromatic core. It could be possible that the spatial arrangement of the aromatic ring systems in these molecules lowered the shape complementarity with the binding site. All of the values determined - K_D, ΔH and -TΔS - were very close to each other both in the presence and absence of compounds **46** and **47**, respectively. At first sight also the results for compound **48** were not striking. In the presence of compound **48** the K_D-value only increased $\approx$ 20% from 233.1 $\pm$ 29.4 nM to 284.8 $\pm$ 53.1 nM and therefore $\Delta\Delta K_{Dapp}$ was < 3 SDs in contrast to the findings for the parent compound **44**. However, the binding enthalpy decreased about 30% in the presence of compound **48** from -73.20 $\pm$ 5.41 kJ/mol to -51.88 $\pm$ 2.70 kJ/mol when tracking the interaction between the riboswitch RNA and its cognate ligand FMN in the ITC displacement assay, suggesting binding of this compound to the target RNA (Table 5.4).

TABLE 5.4: The binding affinities and the energetic terms for FMN binding to the riboswitch RNA derived from the fit to a 1:1-binding model are displayed. Each row represents the average results for three measurements in the presence (+) and absence (-) of the compounds using the same batch of RNA for control measurement and measurement in the presence of a test compound. The change of the apparent dissociation constant is indicated as x-fold increase of the $K_{D(FMN)}$ in the absence of compound.
[a] apparent dissociation constants for the FMN-riboswitch RNA interaction are given in nM concentration
[b] enthalpy, entropy and free energy of binding are given in kJ/mol

	- Test cpd. absent					**+ Test cpd. present**					
Cpd./[µM]	$K_{D(FMN)}$ [a]	**n-value**	ΔH [b]	$T\Delta S$ [b]	ΔG [b]	$K_{D(FMN)}$ [a]	**n-value**	**x-fold ($K_{D(FMN)}$)**	ΔH [b]	$-T\Delta S$ [b]	ΔG [b]
45	233.1±29.4	0.595±0.136	-73.20±5.42	-34.67±5.77	-38.53±0.35	187.0±37.4	0.564±0.025	0.80	-59.76±2.88	-20.62±3.13	-39.14±0.51
46	233.1±29.4	0.595±0.136	-73.20±5.42	-34.67±5.77	-38.53±0.35	223.7±42.7	0.495±0.016	0.96	-67.89±3.49	-29.20±3.73	-38.69±0.50
47	184.6±57.6	0.501±0.013	-73.40±1.07	-34.14±1.62	-39.26±0.87	181.6±63.7	0.496±0.023	0.98	-71.21±8.43	-31.91±9.22	-39.31±1.04
48	233.1±29.4	0.595±0.136	-73.20±5.42	-34.67±5.77	-38.53±0.35	284.8±53.1	0.508±0.035	1.22	-51.88±2.70	-13.84±3.21	-38.04±0.51
49	233.1±29.4	0.595±0.136	-73.20±5.42	-34.67±5.77	-38.53±0.35	258.2±84.1	0.449±0.042	1.11	-77.24±9.48	-38.84±10.33	-38.40±0.86

5.3 Direct biotinylation of the FMN riboswitch RNA

Data resulting from the former assays described in the previous sections of this chapter were derived from displacement assays. The fluorescence quenching assay and the ITC displacement assay were both used to determine binding data for the interaction between the *F. nucleatum* riboswitch RNA and FMN, whereas the binding constants for the possible new ligands were determined indirectly. With the aim to utilize a direct method to identify new ligands for the FMN riboswitch a SPR approach was subjected to trials. Thus, a further orthogonal binding assay was employed to possibly identify new ligands that were not detectable by the former methods and to confirm hit compounds of either the fluorescence-based method or the ITC displacement assay. All compounds that did not interfere with the fluorescence-based assay and did not affect the FMN-RNA interaction by increasing the K_{Dapp} were excluded from the SPR assay.
In order to carry out SPR assays one of the reactants had to be immobilized on a sensor chip. It was decided to immobilize the biotinylated target RNA on a streptavidin-coated sensor chip. Different approaches were tested to specifically label the full-length aptamer domain of the FMN riboswitch RNA with a biotin tag (Chapter 2, 3.4) on either the 5′ end or the 3′ end. To specifically label the 3′ end and by one method the 5′ end three different methods were carried out. The 3′-end biotinylation was achieved either by selective periodate-mediated oxidation of the terminal 3′ end ribose and subsequent reaction with a biotinylated nucleophile[177,178] or by transfer of a biotinylated nucleotide derivative with T4 RNA ligase to the 3′ end.[179] Biotinylation of the 5′ end was accomplished by activation of the terminal phosphate group at the 5′ end with EDC and subsequent reaction with a biotinylated nucleophile.[180] The highest labelling efficiency was reached by labelling the 3′ end of the riboswitch RNA using periodate chemistry and a pegylated biotin hydrazide as tag. Native agarose gels were initially used to qualitatively confirm a successful biotinylation.[178] The agarose gels were set up as gel shift assays. The velocity of the RNA running on the gel was decelerated upon addition of streptavidin, especially when using higher amounts of protein (2 and 5 µg). A smeared band that was running with a higher molecular weight became observable due to the interaction (Figure 5.10). The biotinylation itself did not alter the mobility of the riboswitch RNA on the gel. A second method was applied to more quantitatively measure the labelling efficiency. For that purpose, intensities of the purple precipitation upon an enzymatic reaction between a streptavidin-conjugate cleaving its chromogenic substrate when bound to dot-blotted biotinylated RNA were considered. A fully biotinylated RNA served as control at a concentration of 0.5 ng/µl. The intensity of this spot was referred to a biotinylation of 100% in the following. The spot intensities for 3′-end biotinylated riboswitch RNA modified via periodate-chemistry and 5′-end biotinylated riboswitch RNA modified via EDC-reaction were compared to the control spot with fully biotinylated RNA. For the 3′-end biotinylated RNA a similar intensity as

for the 100% control was detected at a concentration of 0.5 ng/µl implying an almost complete reaction, whereas the 5′-end biotinylated RNA could only be detected at four times higher RNA concentrations. Thus, the 5′ end was much less efficiently labelled with the biotin-hydrazide tag (Figure 5.10).

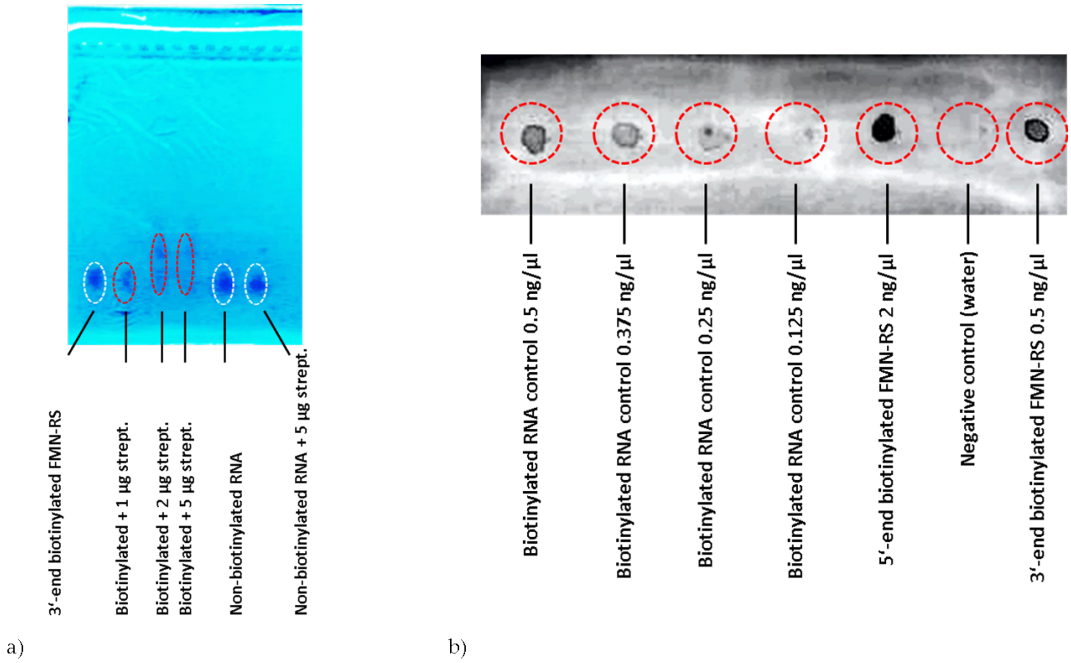


FIGURE 5.10: The 3′ end of the FMN riboswitch RNA was efficiently labelled with a biotin tag by periodate-based chemistry. Biotinylation was verified by two different experiments. a) Results of the gel shift assay. 3′-biotinylated FMN-RS RNA was applied to the agarose gel in the absence and in the presence of 1-5 µg streptavidin. Control lanes contained non-biotinylated riboswitch RNA ± streptavidin. b) Results of the enzymatic reaction initiated by a RNA-streptavidin complex immobilized on a nylon membrane. Four different concentrations of fully biotinylated control RNA were immobilized together with a negative control without RNA. 3′-end biotinylated as well as 5′-end biotinylated riboswitch RNA were compared to the control RNA.

In an alternative approach, strand B of the two-piece RNA construct was biotinylated to immobilize the annealed FMN riboswitch RNA on the streptavidin sensor surface during SPR experiments. The results for the biotinylation of the 3′ end were similar compared to the full-length aptamer domain (data not shown). Chromogenic detection of dot-blotted, biotinylated RNA revealed sufficient labelling in a high yield (data not shown).

To prove that the 3′-end linker did not interfere with FMN binding the ability of the modified RNA to interact with its natural ligand was tested using the fluorescence quenching assay. During the fluorescence quenching assay the procedure remained unchanged compared to the non-biotinylated RNA. This setup was corresponding to a FMN concentration of 40 nM and a RNA concentration series starting with 5 µM. With a K_D-value of 354.3 nM ± 31.9 nM in a single measurement only a slight loss of affinity was observed (Figure 5.11).

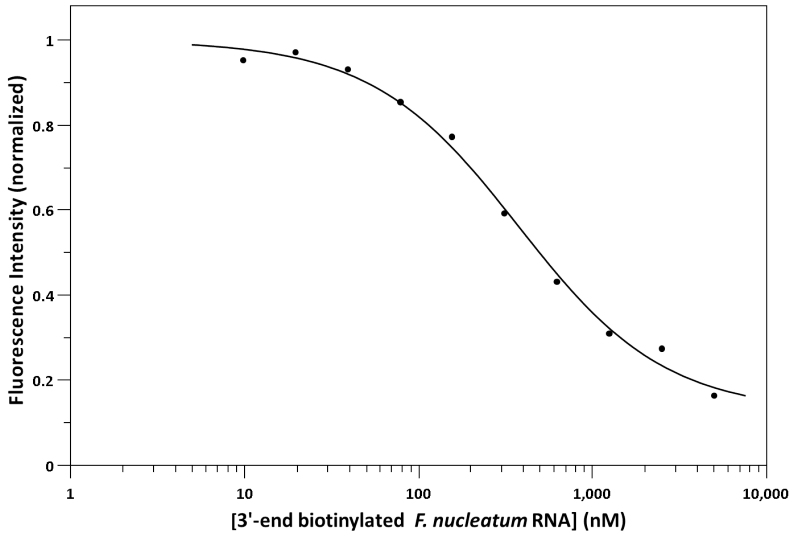


FIGURE 5.11: The normalized and buffer-subtracted fluorescence intensities were plotted against the *F. nucleatum* RNA in a semilogarithmic plot. A *solid, black line* represents the best fit for the interaction between FMN and the 3′-end biotinylated RNA.

5.4 Experimental testing of the virtual screening hits by SPR

In total, three different approaches were employed to perform SPR measurements. As first option the 3′-end biotinylated full-length aptamer domain of the *F. nucleatum* FMN riboswitch was directly immobilized on the sensor chip. The second approach used an indirect method to immobilize the target RNA onto the sensor surface via hybridization of the elongated 3′ end of the RNA to a complementary 5′-end biotinylated DNA oligonucleotide. As last approach the two-piece FMN riboswitch RNA construct previously used for crystallization was employed for SPR measurements, after the 3′ end of strand B had been biotinylated.

5.4.1 SPR measurements with the 3′-end biotinylated full-length aptamer domain of the FMN riboswitch

The 3′-end labelled full-length aptamer domain was immobilized to an immobilization level as described in chapter 2 (3.5.3.1), whereupon the level of immobilization was tracked by the response units measured during the loading step. A 5′-end biotinylated DNA oligonucleotide with a length of 25 nucleotides served as reference (see 3.3.4) and was immobilized to a similar level on the reference cell as the 3′-end biotinylated RNA on the sample cell. Comparing the binding levels of FMN expressed as relative response at the end of the association phase, a concentration dependence was observed for the sample flow cell but not the reference flow cell (Figure 5.12).

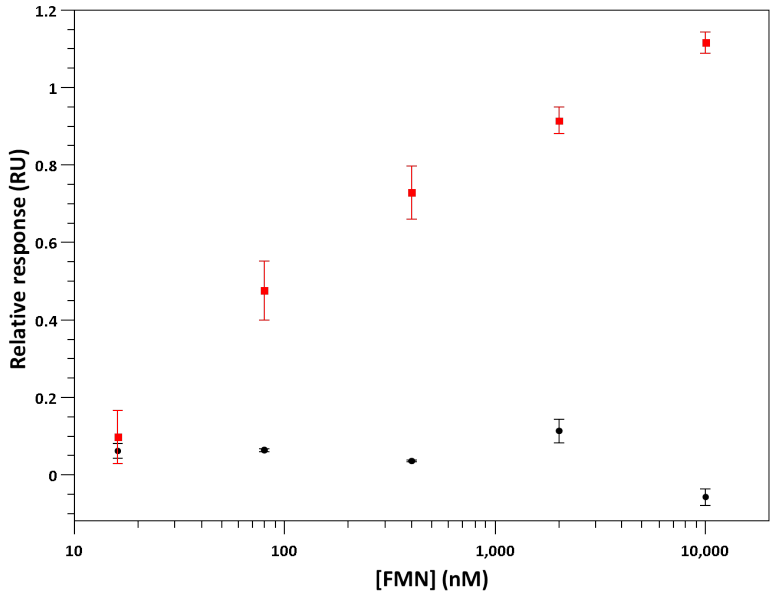


FIGURE 5.12: The relative responses averaged over two measurements with error bars indicating the standard error are plotted against the FMN concentrations for a representative SPR run using the 3′-end biotinylated full-length aptamer domain. *Black circles* indicate the data points corresponding to the reference surface responses after buffer subtraction and *red squares* represent the data points for the sample surface after reference and buffer subtraction.

Concentration series with FMN showed a concentration-dependent increase of the relative responses (Figure 5.13). However, it was not possible to determine the kinetic constants with an adequate accuracy as judged by the internal data integrity check within the Biacore® evaluation software, probably because the maximum response was too small. Nevertheless, it can be concluded that the concentration-dependent responses likely arose from specific interactions. Despite the inaccuracy in the determination kinetic data could be fitted omitting the highest concentration of FMN to get an idea whether the determined binding affinity of FMN agreed with previous data from the other methods. Especially for the concentrations above the expected K_D-value, steady-state plateau was reached in the association phase, indicating that the chosen concentration window was large enough. All of the analyte fully dissociated during the dissociation phase (Figure 5.13).

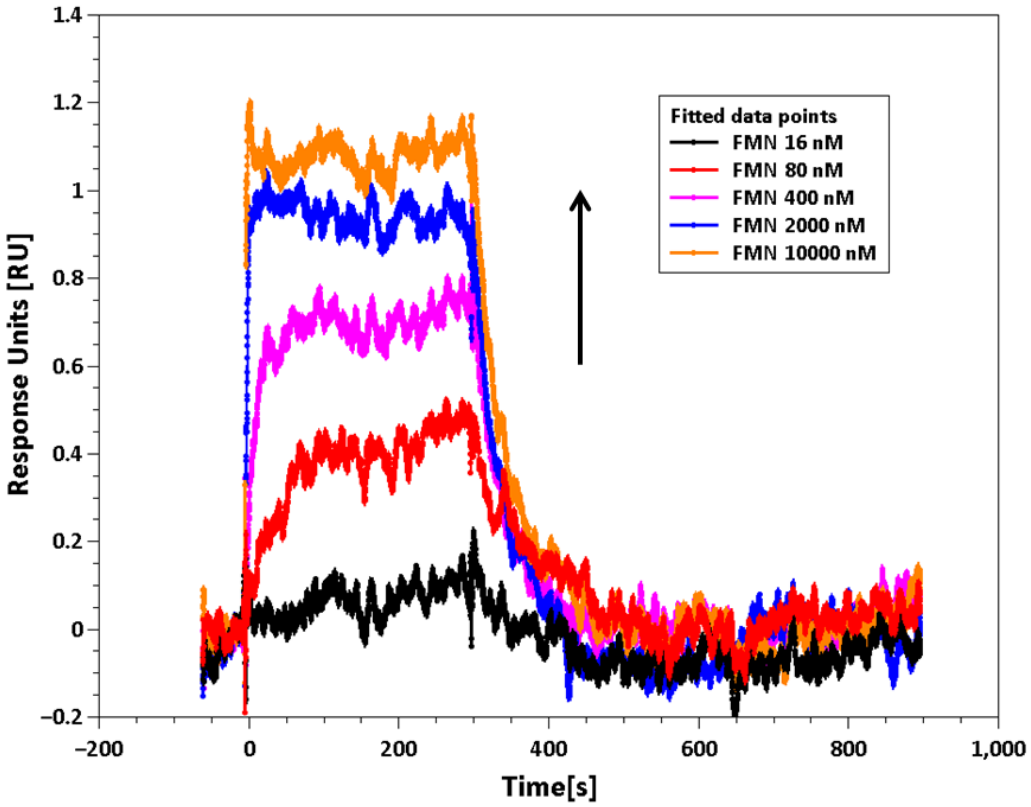


FIGURE 5.13: SPR control measurements with the 3′-end labelled full-length riboswitch RNA. The fitted sensorgrams averaged over two cycles for every concentration are depicted. Relative responses are plotted against the time in seconds with the association phase (analyte injection) starting at zero s and dissociation phase ending after 900 s. Reference cell response and zero concentration response were subtracted. Different FMN concentrations are colour-coded according to the *rectangular box* and the *black arrow* refers to the increasing concentrations from bottom to the top.

Subsequently, the kinetic data were inspected closer. The maximum response was fitted to be very low with only ≈ 1 RU. Via the relationship between k_a and k_d a K_D of 74.29 nM was calculated roughly indicating the same affinity between FMN and the riboswitch RNA as previously determined by the former two assay methods (Table 5.5). It has to be considered that there is a larger uncertainty accompanied with this SPR method and the determined K_D compared to the other methods. After the evaluation of the data the software directly provided a "quality control". This control implied a good buffer match and kinetic constants within the instrument's specifications, but difficulties in determining the constants correctly.

TABLE 5.5: The association and the dissociation rate constants were fitted from the sensorgrams with the evaluation software together with the K_D-value and the maximum response upon ligand binding (R_{max}).

Association rate constant (k_a)	**Dissociation rate constant (k_d)**	**Dissociation constant (K_D)**	**R_{max}**
2.927×10^{5} $M^{-1}s^{-1}$	$2.174\times10^{-2}s^{-1}$	74.29 nM	0.9434 RU

As second method to determine the K_D-value the steady-state affinities were evaluated by fitting data of the plot displaying relative responses versus FMN concentration. Relative to the very low overall response upon FMN binding to the RNA the standard errors between two measurements with the same analyte concentration were quite high and the data points lay far away from the fitted saturation binding curve. The relatively low quality of the fit was also reflected in the calculated K_D-value. Again, the particular value was quite close to the expected values with 107.2 nM, but the standard error was as high as $\pm$ 50 nM (Figure 5.14).

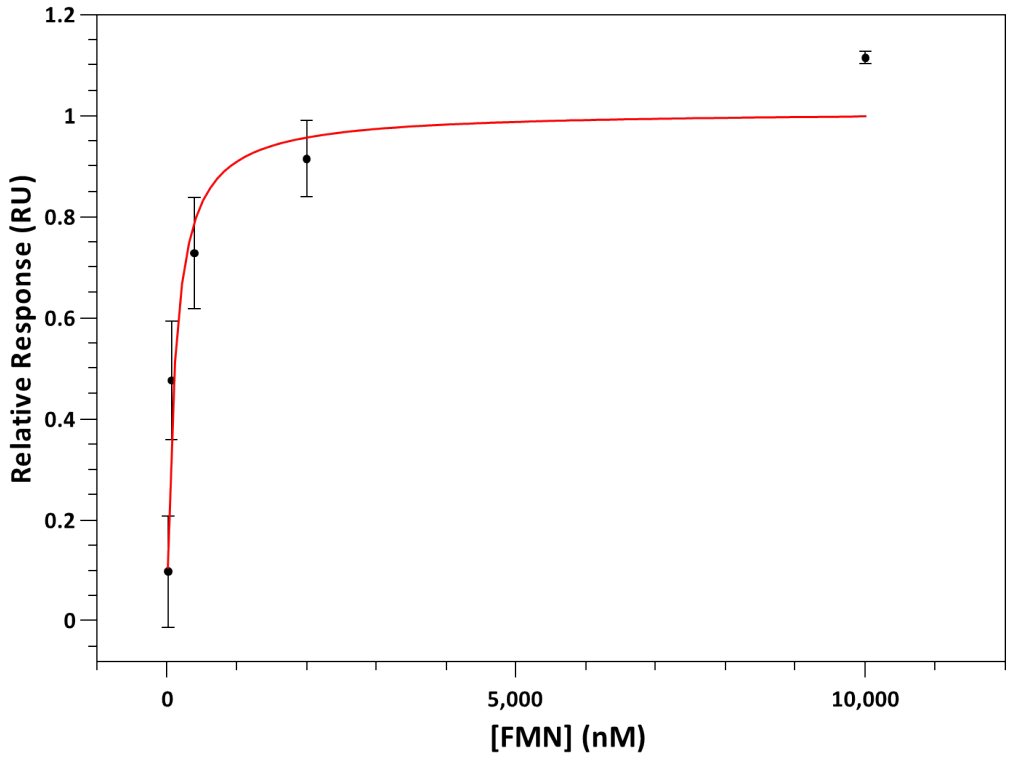


FIGURE 5.14: The relative responses in the steady state (after double-referencing) at the end of the association phase for the interaction between the 3'-end biotinylated full-length aptamer domain and FMN were plotted against the FMN concentration and fitted to a 1:1 binding model. A *bold red line* indicates the best fit. The coordinates for the fitting curve were extracted from the evaluation results by the software. Error bars represent the standard deviation for the relative responses averaged over two measurements.

Since the kinetic constants could not be determined accurately, it was decided to move on with other SPR approaches. A fast decline of the residual activity was also observed with responses falling below the noise level. Furthermore, it was difficult to reproduce the results over several cycles.

5.4.2 SPR measurements with the elongated aptamer domain of the FMN riboswitch

The second SPR approach relied on an indirect immobilization of the FMN riboswitch aptamer domain that was prolonged by 36 nucleotides by hybridization to a 5′-end biotinylated DNA oligonucleotide complementary to the 3′ end of this elongated RNA. A prolongation of the initial aptamer domain was conducted to avoid a disturbed global folding of the aptamer domain due to the hybridization. For the purpose of immobilization a two-step approach was used. First, a RNA-DNA-hybrid was immobilized on the sample cell surface followed by an additional injection of free riboswitch RNA. Free biotinylated DNA oligonucleotide was immobilized on the reference cell to a similar level as described for the first approach (5.4.1).
In the same way as done before with the 3′-end biotinylated full-length aptamer domain of the FMN riboswitch RNA a fluorescence quenching assay was conducted to ensure that the inserted 36 nucleotides do not perturb the functionality of the RNA. Again, a concentration series of ten 1:1 dilutions RNA starting with 5 µM as highest concentration was titrated against 40 nM FMN. No remarkable changes of the RNA performance were detected in a single experiment. A slightly higher K_D was observed with 326.7 $\pm$ 58.6 nM, but still a full binding curve with clear concentration dependence was obtained (Figure 5.15 a).

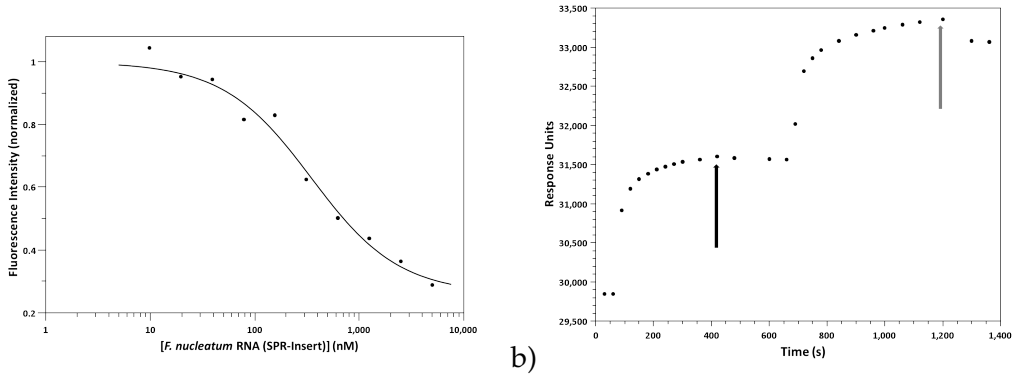


FIGURE 5.15: a) The normalized and buffer-subtracted fluorescence intensities were plotted against the concentration of the *F. nucleatum* FMN riboswitch RNA carrying the insert in a semilogarithmic plot. A *solid black line* represents the best fit for the FMN-RNA interaction. b) The elongated aptamer domain containing an insert at the 3′ end is immobilized on the streptavidin sensor chip in two steps. A *black arrow* indicates the the injection of the DNA-RNA hybrid and a *grey arrow* represents the addition of pure RNA to increase the surface density by binding free DNA oligonucleotide.

The initial DNA-RNA hybrid was immobilized to ≈ 1,700 RU and the immobilization level was increased to a total of about 3,500 RU after injection of the pure RNA sample to the hybrid, indicating an interaction between the elongated riboswitch RNA and free complementary biotinylated DNA oligonucleotide (Figure 5.15 b).
Kinetic constants could not be determined and the immobilization was not stable over time. The baseline response level for the immobilized DNA-RNA hybrid clearly decreased with every kinetic cycle. Since the drift was much less pronounced for the biotinylated DNA oligonucleotide, it was concluded that the indirect immobilization of the RNA via Watson-Crick base pairing might explain the strong drift, which was higher than the strongest response level upon saturating concentrations of FMN.
To circumvent this problem, it was tried to re-immobilize the pure RNA sample to the sample flow cell with the beginning of every new cycle. Here, a high concentration of FMN riboswitch RNA was chosen and a low flow rate to ensure the substitution of dissociated RNA from the heteroduplex and obtain a stable baseline. However, the baseline levels as starting points for the determination of association and dissociation phases were still differing, because the response did not decrease by exactly the same amount during one single cycle. The maximum response reached is dependent on the immobilization level (Chapter 2, Equation 3.11) and therefore no concentration-dependent responses were observed with these unsteady immobilization levels. With the fluctuating baseline it was even more difficult to interpret the binding data than with the steady baseline drop over time. For that reason this approach was abandoned.

5.4.3 SPR measurements with the two-piece FMN riboswitch RNA construct and 3'-end labelled strand B

There was evidence in the literature that the annealing of both strands A and B in the presence of the ligand FMN results in an aptamer domain with the ability of binding FMN, since several crystal structures had been determined using this approach.[39,61]
Based on the results for the 3'-end biotinylated full-length RNA aptamer domain, it was assumed that also for this two-piece construct a label should not interfere with binding. Therefore, the functionality and performance of the construct was tested directly on the sensor chip surface without further control experiments.
The immobilization on the streptavidin-coated sensor surfaces proceeded as desired. 2,157 RU biotinylated DNA oligonucleotide was loaded on the reference surface and 2,272 RU FMN riboswitch RNA was loaded on the sample surface. Thus, both sensor surfaces were loaded to a similar density of nucleic acid, which enhanced the reliability of the obtained binding data. Exemplary sensorgrams for the determination of the kinetic constants with a FMN concentration series ran twice showed that the responses measured were concentration-dependent regardless of the order they were injected. Sensorgrams for one and the same concentration superimposed very well and the fitted curves also matched the sensorgrams (Figure 5.16). The sensorgrams

also indicated a 300 s dissociation time as sufficient to ensure complete dissociation of the ligand.

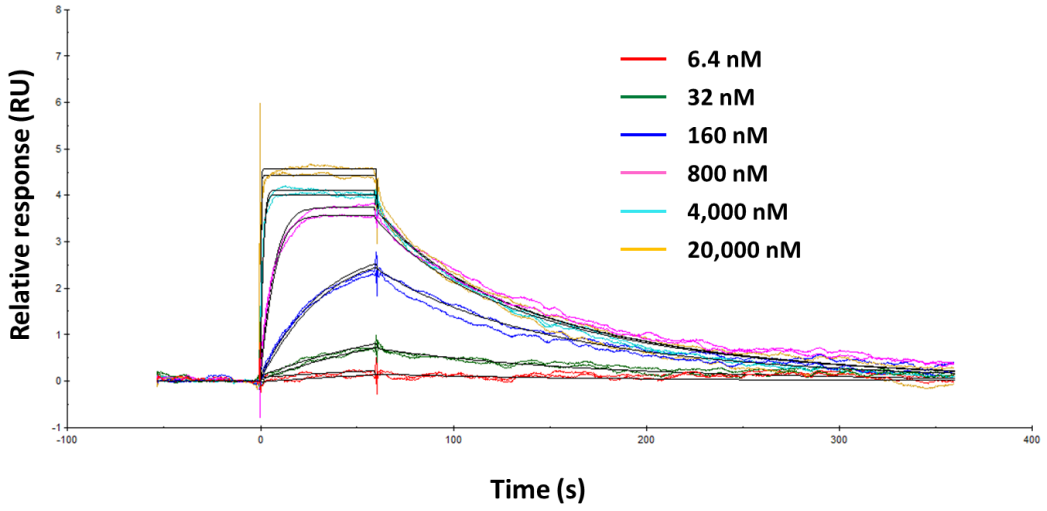


FIGURE 5.16: SPR measurements with a FMN concentration series applied to a sample cell the two-piece aptamer domain construct was immobilized to. Sensorgrams with the relative responses plotted against the time are depicted in a colour-coded manner. Two full cycles for each concentration are shown. The *solid black lines* indicate the best fit to a 1:1 binding model for each curve.

An internal data integrity check within the software confirmed that the kinetic constants (Table 5.6) could be fitted and reliable data were generated. Together with the good overlay of the fitted curves with the sensorgrams and the much lower uncertainties in the output data compared to the biotinylated full-length RNA expressed as standard errors this enhanced reliability was substantiated.

TABLE 5.6: Kinetic rate constants, K_D-value and maximum response obtained by kinetic fitting of the sensorgrams to a 1:1 interaction model within the Biacore® evaluation software.

Association rate constant (k_a)	**Dissociation rate constant (k_d)**	**Dissociation constant (K_D)**	**R_{max}**
2.479×10^{5} $M^{-1}s^{-1}$	$1.490\times10^{-2}s^{-1}$	60.11 nM	3.745 RU

In the present case, the most important kinetic constant for a validation of the feasibility of this method to investigate the binding affinity of possible novel ligands was the dissociation constant. Therefore, in addition to kinetic fitting of the sensorgrams, the K_D-value was determined by fitting the relative responses in the steady-state to the RNA concentrations. Minor differences for the calculated K_D-values were observed depending on whether fitting was done using the complete sensorgram or using the relative responses measured at equilibrium at the end of the association phase. The K_D-value in the equilibrium binding state was required to be consistent during several runs, before testing the binding affinities of the virtual screening hits.

Therefore, it was tested with four different runs on two subsequent days, whether the generated data was reproducible. Herewith, the response of each concentration was measured six times. The mean relative responses for each concentration from 6.4 nM to 20 µM lay within the mean fitted curves for the reference and buffer subtracted relative responses plotted against the FMN concentration (Figure 5.17). For the riboswitch-FMN interaction a slightly higher K_D-value of 150.23 $\pm$ 12.06 nM was determined by applying steady-state analysis in comparison to kinetic fitting (60.11 nM, Table 5.6). The maximum response upon binding of saturating FMN concentrations was 4.398 $\pm$ 0.400 RU. A tendency for lower responses over time expressed as decreasing R_{max}-values was detected when the same sensor surface was used for 48 hours during four independent runs. During the last of the four runs the affinity for FMN dropped indicating that the immobilized RNA looses some activity over time, which can be substantiated by the falling baseline levels before starting the next cycle. The baseline levels decreased much more for the sample cell with the biotinylated RNA than for the reference cell.

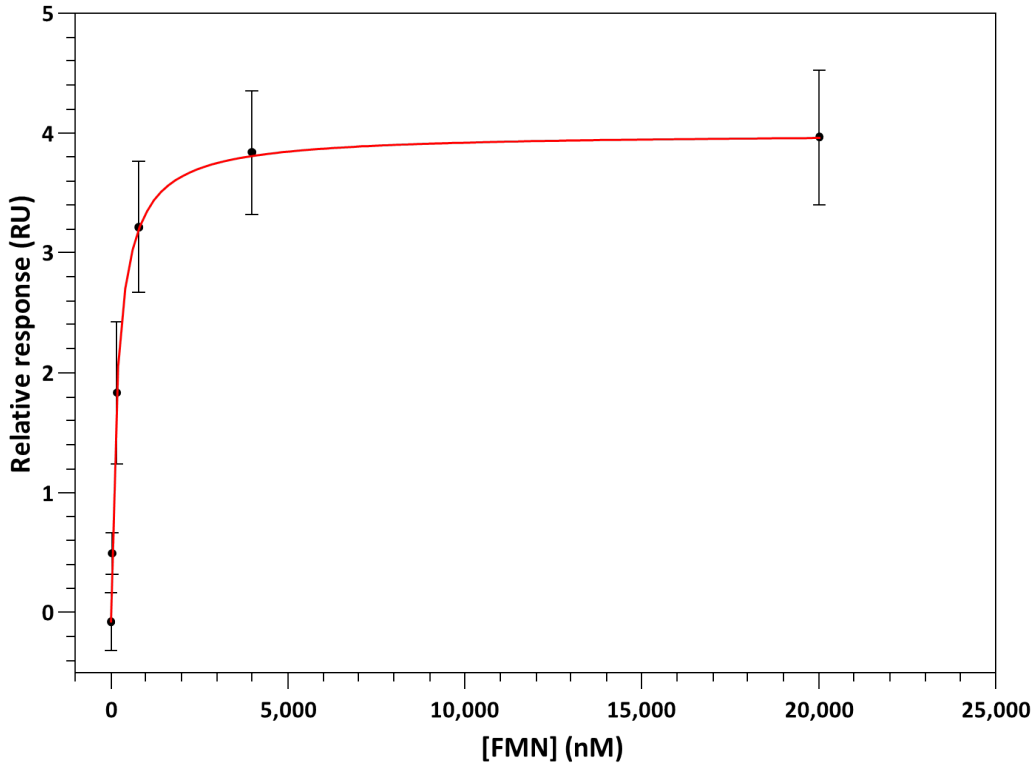


FIGURE 5.17: The relative responses in the steady state (after double-referencing) at the end of the association phase were plotted against the FMN concentration and fitted to a 1:1 binding model for the interaction between FMN and the two-piece RNA construct. A *bold red line* indicates the best fit averaged over six consecutive measurements. The coordinates for the fitting curve were extracted from the evaluation results by the software. Error bars represent the standard deviation for the relative responses averaged over six measurements at the six indicated concentrations of the dilution series.

After the feasibility of this approach was proven, the virtual screening compounds were tested by injecting them over the flow cells with the riboswitch RNA and the biotinylated DNA oligonucleotide, respectively. The performance of the freshly immobilized riboswitch RNA was retested by applying a FMN concentration series first. Due to the fact that SPR running buffer containing 5% DMSO was necessary to keep the screening compounds solubilized during the run, also a single run with the standard FMN concentration series in the presence of DMSO was conducted. The binding curves as well as the fitted kinetic constants did not reveal any major changes for both control experiments and the surface therefore was regarded as ready-to-use.
The virtual screening hits that were included in the SPR experiments were selected for different reasons. For example, compound **25** was chosen, because it could not be tested in the fluorescence-based assay and it was the top-scoring compound among the virtual screening hits arising from the substructure search (Table 4.4). Compound **28** possibly showed binding in the the fluorescence quenching assay (5.1.2), however these results could not be confirmed in the ITC assay. Reasons for the inclusion of further compounds are discussed in the next chapter (6.2.4). The close analogues of compound **44** (compounds **45**-**49**) as part of the small structure-activity relationship study were selected as substitutes for compound **44**, which was no longer available at that time, to validate the results of the ITC displacement assay. A complete list of compounds included in the assay is given in table 5.7:

TABLE 5.7: List of the compounds tested in the SPR assay using the two-piece aptamer domain.

Screening compound	Concentration in assay [µM]
45	100
46	50
47	250
48	50/100
49	25/50
41	250
25	250
47	100
34	50
42	100
40	150
45	200

To monitor the surface performance during the assay, a control sample with a saturating FMN concentration (5 µM) was included in the middle and at the end of every run. The FMN control consistently produced the highest responses (Figure 5.18) meaning that the control samples in the middle of the run and at the very end

of the run were not compromised by carry-over through non-specific binding compounds or any other left samples with screening compounds. Thus, the obtained results appeared to be valid.

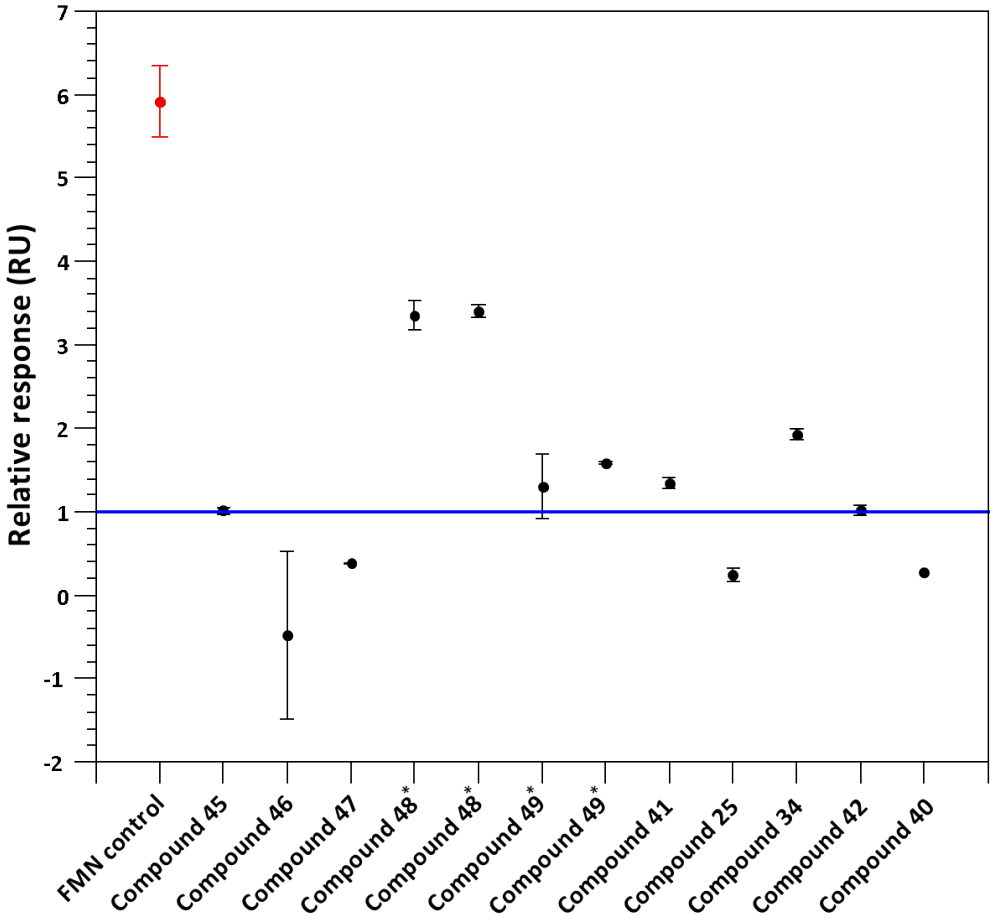


FIGURE 5.18: Virtual screening hits were applied to the immobilized two-piece aptamer domain produced by annealing strand A and B. The relative responses of the compounds are plotted against the corresponding compound name at the end of the association phase after subtraction of the buffer response. Data belong to one single measurement run while each compound was injected twice. A control for testing the surface integrity with saturating concentrations of FMN (5 µM) was run five times - at the beginning of the experiment, at the end and between testing the screening hits. The mean response is indicated as *red circle* for the control. The *solid blue line* indicates the initial cut-off applied to separate active from inactive compounds.
(* = Compound **48** and **49** were tested in two different concentrations. They are depicted in the plot in the following order: compound **48** 50 and 100 µM and compound **49** 25 and 50 µM.)

Compound **45** was expected to be inactive, since it does not contain the hydrogen bond donor and acceptor features in the same spatial arrangement predicted to be crucial for compound **44** binding to the riboswitch. Indeed, at 100 µM a response of only $\approx$ 1 RU was observed. Compounds **46** and **47** resulted in similar low RUs and were therefore also not considered to specifically bind to the FMN riboswitch.

The compounds **25**, **42** and **40** also showed no binding. Compound **34** had an almost twofold higher relative response with 1.924 RU than its close analogue compound **42**, but compound **42** was one of the two compounds exhibiting activity in the fluorescence-based assay. In the ITC displacement assay both compounds did not alter the affinity of the FMN riboswitch to its cognate ligand. Both molecules were expected to be recognized by the riboswitch binding site with very similar binding modes and so it was doubted that the response corresponding to compound **34** indicated the compound as a true binder. Due to this fact the classification for all the compounds with responses between 1 and 2 RU (Compounds **34**, **41** and **49**) were unclear at this point.
Compound **48** showed the best performance of all the analogues of compound **44** during the ITC assays. To lower the possibility that this compound was incidentally assigned a true binder, compound **48** and the closest related compound **49** were applied to the SPR assay in two different concentrations. In both cases, the concentrations were close to each other so that only marginal differences in the observed responses were expected. Indeed, the responses were consistent with compound **48** displaying the highest responses of about 3.4 RU (Figure 5.18) . Compared to the response for FMN at saturating concentrations this was not much more than 50% of the binding level suggesting that at the highest possible concentrations, 100 and 50 µM, respectively, saturation was not achieved. Based on the dissociation constant determined for the parent compound, compound **44**, in the ITC displacement assay and the high structural similarity, the affinity for compound **48** was expected to be in the range of 100-200 µM.
The next step in the identification of compound **48** as a SPR hit substance was the investigation of the sensorgrams, especially the shape of the sensorgrams in comparison to the ones obtained for FMN binding to the riboswitch RNA. For instance, in the association and dissociation phase an almost linear increase and decline, respectively, are indicative of artefacts (e.g. due to mass limited transport) that have to be distinguished from true binding events. Without any curvature in the association and dissociation phase reliable data cannot be obtained.[203] During the SPR measurements the samples containing 5 µM FMN were designated as control measurements, however it was impossible to track the sensorgrams of "controls" in the evaluation software and therefore sensorgrams cannot be shown here. Thus, for the purpose of comparing sensorgram types, a sample containing 4 µM FMN from the assessment of kinetic constants with FMN solutions also containing 5% DMSO was used. These sensorgram was expected to behave very similar. The sensorgram of the control began with a oscillating baseline, but with the beginning of the injection the sensorgram reflected a typical binding event (Figure 5.19). The addition of the compound **44** analogues, compounds **48** and **49** did not result in a concentration-dependent signal, a prerequisite for specific binding. The chosen concentrations might have been too close to each other when differing by only a factor of two.

Rather for compound **48** than for compound **49** a slightly higher response was observed at the higher concentrations (1.58 vs. 1.30 RU for compound **49**). Therefore more concentrations were necessary to elucidate, whether or not those analogues of compound **44** specifically bind to the RNA.

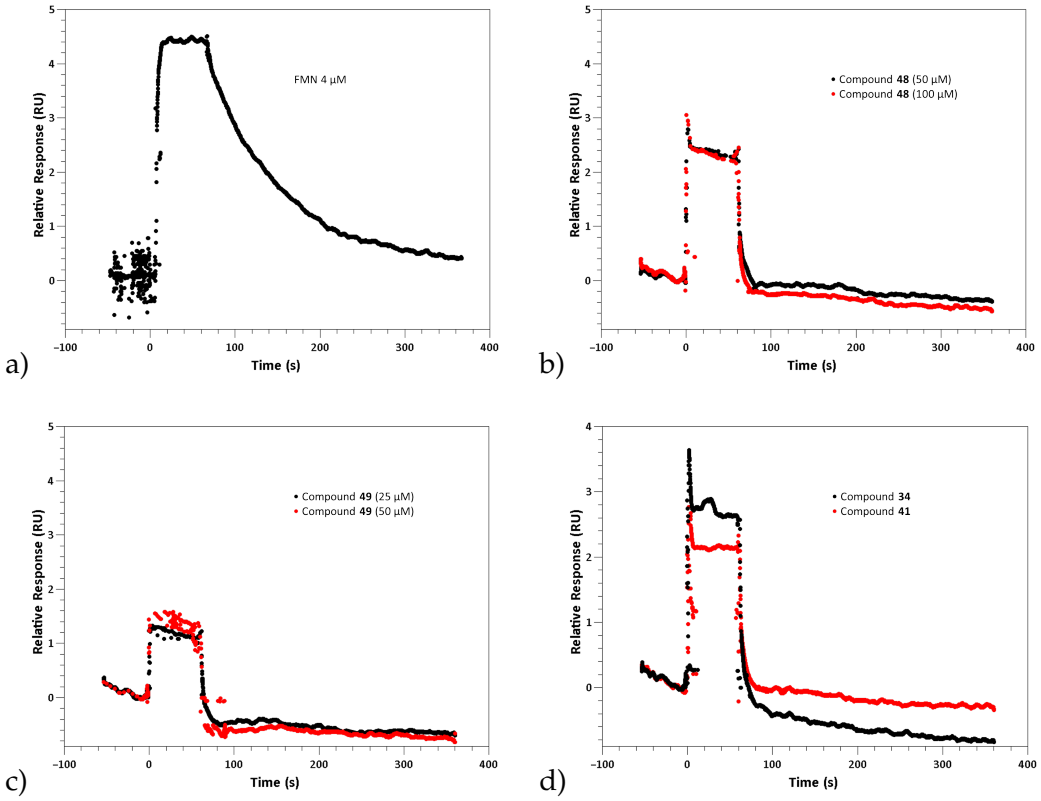


FIGURE 5.19: Sensorgrams for some of the virtual screening compounds applied to a surface the two-piece aptamer domain RNA was immobilized to are displayed. The sensorgrams are an average of two measurements and not fitted to a binding model, because only 1-2 concentrations were used. Adjustment of the sensorgrams was performed by subtracting the effect of DMSO by a solvent correction method with different DMSO concentrations and they are blank (buffer) and reference-subtracted, too. a) Sensorgram of a saturating FMN concentration binding to the FMN riboswitch. b) Sensorgrams for two different concentrations of compound **48** interacting with the sensor surface. c) Sensorgrams for two different concentrations of compound **49** interacting with the sensor surface. d) Overlay of the sensorgrams for the compounds **34** and **41**.

With the aim to investigate the response upon injection of the most promising candidate compounds **48** and **49** concentration series with 1:5 dilutions were prepared. Compound **48** was tested in concentrations ranging from 100 µM to 32 nM and compound **49** accordingly in concentrations from 50 µM to 16 nM. Directly afterwards a quality control was run with a concentration series of FMN ranging from 20 µM to 6.4 nM to ensure that the surface remained capable of binding and there were no negative influences by the compounds. The results clearly suggested that the kinetic assay for the screening compounds was carried out on a functional surface throughout the whole assay. With the six FMN concentrations used for the control experiment the typical sigmoidal shape of a 1:1-binding curve was observed when

the responses were plotted against the concentration on a logarithmic scale (Figure 5.20 b) and saturation binding was also achieved. The kinetic constants of FMN binding to the FMN riboswitch had been elucidated previously (Table 5.6), so that this time a simple sigmoidal fit was carried out within QtiPlot to illustrate the data. The dissociation constant K_D was calculated to be 115.8 $\pm$ 20.2 nM in that case. This value was in good correlation with all of the other control experiments performed and described earlier in this section.
In contrast, for the possible riboswitch RNA ligands the binding level data points did not describe a full saturation binding curve (Figure 5.20 a). Especially for compound **49** only marginal differences were observed for the responses when either 16 nM or 50 µM compound were applied and it has to be considered that the absence of concentration dependence might be an indication of non-specific binding effects.[142] In the case of compound **48**, the application of the four highest concentrations at least resulted in steadily increasing responses with a very pronounced increase from 20 to 100 µM compound **48**. However, this disproportional sharp increase indicates that the last value is probably an outlier.

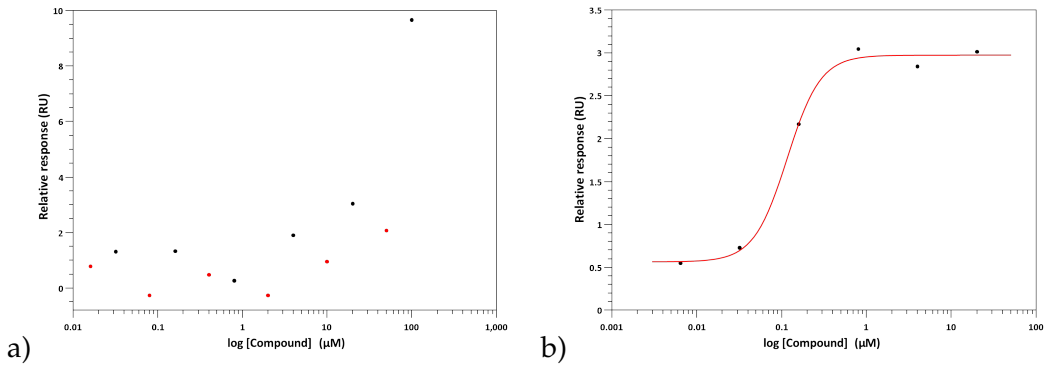


FIGURE 5.20: The relative responses after reference and buffer subtraction for the binding of six concentrations of the compounds **48** and **49** to the two-piece RNA construct are plotted against the concentration in µM on a semilogarithmic scale. All concentrations were measured once. Data points were collected directly at the end of the association phase. The FMN concentration series was tested in a different run and therefore the responses span a larger range for the analogues, which was actually not attributed to a higher binding, but to a more negative buffer response. a) Binding levels of compound **48** (*black circles*) and compound **49** (*red circles*) are plotted vs. the logarithm of the respective concentrations. b) Concentration-dependent binding levels of the FMN dilution series are indicated as *black circles*. The *bold red line* represents the sigmoidal fit (QtiPlot) for the binding curve.

Since saturation was detected for neither compound **48** nor compound **49**, the evaluation software could not determine the binding constants of the particular interactions. Albeit having a large error, the kinetic constants could be calculated for compound **48**. With a dissociation constant of $\approx$ 35 µM the binding affinity was probably overestimated by this fit. For obtaining the right-hand branch of the saturation binding curve concentrations of up to ten-fold of the K_D were necessary. Due to solubility issues it was not possible to measure the responses of a solution containing 1 mM

compound **48**.

Inspection of the sensorgrams corresponding to the respective concentrations of compound **48** was carried out in addition for an easier recognition of artefacts that might have caused misinterpretation of the data (Figure 5.21). For the concentrations in the µM-range the shape of the sensorgrams was altered. Especially at concentrations of 4 and 100 µM of compound **48** there was some curvature during the association phase, indicating that the steady state was not reached (Figure 5.21). The sensorgram for the 20 µM was compromised by a strong drift directly after the injection had begun (data not shown). Thus, the response went down further compared to the other concentrations.

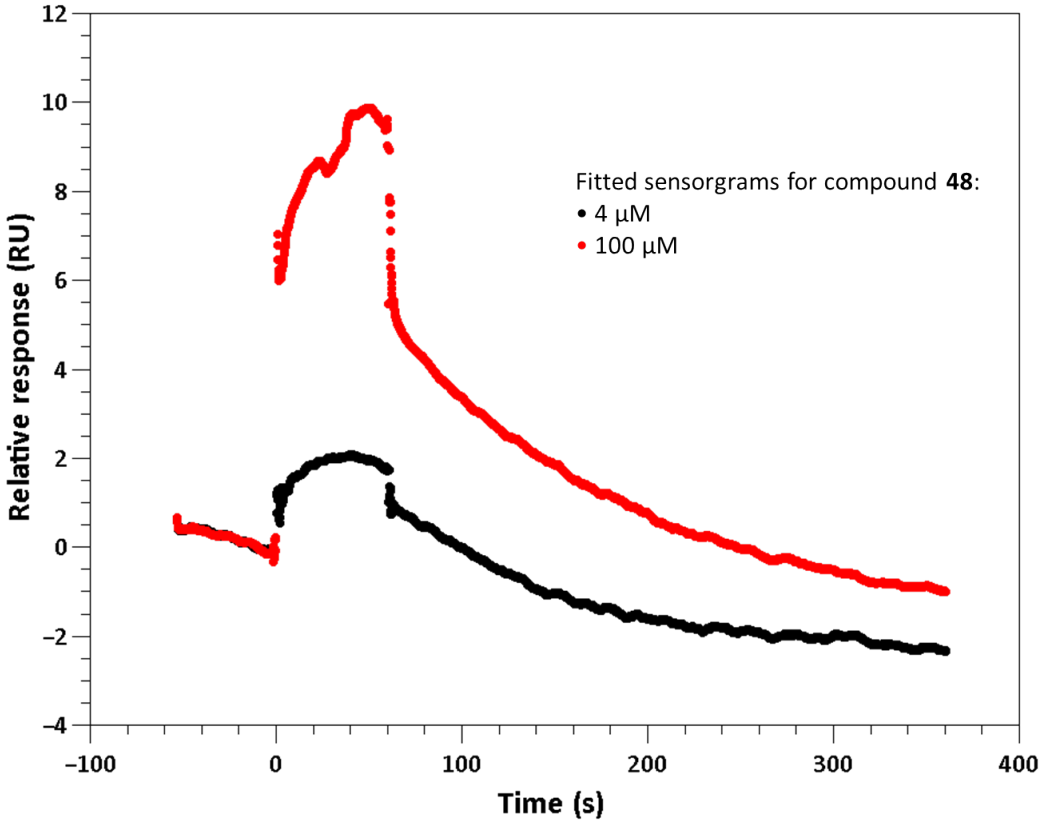


FIGURE 5.21: The sensorgrams corresponding to the binding of 4 and 100 µM compound **48** to the two-piece aptamer domain are shown. Relative responses are plotted against the time in s. Fitting curves are not displayed for the reason that the internal quality control within the Biacore® evaluation software could not reliably determine the binding constants.

The experimental artefact explains the above-average increase in the binding level plot from 20 to 100 µM concentration of compound **48** (Figure 5.20). A large bulk drift due to buffer mismatch possibly elicited the drop in the response directly before the dissociation curvature developed. Other reasons could be the influences of more complex binding than caused by a simple 1:1 interaction.[203] Under the reserve that the reproducibility has to be proven, the present sensorgrams indicated a

concentration-dependent response upon binding of compound **48** to the FMN riboswitch RNA. Even for the highest concentration no saturation of binding was achieved. Therefore, equilibrium binding analysis was impossible. The sensorgrams underlined that higher compound concentrations would be necessary to reliably determine a dissociation binding constant for compound **48** by this method.

5.5 Crystallization trials with the annealed two-piece riboswitch RNA construct

5.5.1 Crystallization of the FMN riboswitch RNA with its cognate ligand FMN

Before co-crystallization trials with compounds arising from the virtual screening process were initialized, conditions were tested, at which structures already had been successfully solved via X-ray crystallography with FMN bound to the riboswitch RNA in the past.[39,61] The first entry of FMN bound to the FMN riboswitch in the protein data bank (pdb: 3f2q) was obtained by co-crystallizing the *in vitro* transcribed aptamer domain of the *F. nucleatum* FMN riboswitch with FMN while using the hanging drop method with a precipitant solution containing 0.1 M MES-sodium (pH 6.5), 0.1 M $MgCl_2$ and 10% PEG 4000 (pdb: 3f2q). Crystals diffracted with a resolution of 2.95 Å.[39] Successful crystallization was also achieved with the annealed riboswitch construct by this group (pdb: 3f4e). With the same precipitant solution composition a comparable resolution of 3.05 Å was obtained.
For the annealed strands that were *in vitro* transcribed from their synthetic DNA-templates (Chapter 2, 3.3.3) it was easier to achieve the large amounts of purified RNA in a high concentration necessary for crystallization. Therefore this approach was pursued. Crystals grew in the presence of FMN under varying concentrations of $MgCl_2$ and PEG 4000. A common observation in all the drops containing crystals was precipitation directly starting, after the crystal drops had been set up (Figure 5.22). Finally, after waiting for at least five days crystals began to grow around the borders of the precipitate. The two conditions yielding the best results had a similar composition of the reservoir solution compared to the published conditions. The buffer component MES-sodium had the same concentration throughout all wells of the crystallization tray. Less growth and smaller crystals were obtained with a lower magnesium concentration and a higher PEG concentration compared to a magnesium concentration of 0.2 M and 10% PEG 4000. However, the crystals grew more individually in this case and a three-dimensional shape was observed, whereas with 0.2 M $MgCl_2$ and 10% PEG 4000 plates and spherulites were observed and several crystals aggregated together.
Data for two crystals grown with the precipitant solution consisting of 0.1 M MES-sodium (pH 6.5), 0.1 M $MgCl_2$ and 13% PEG 4000 were collected and for one crystal after data refinement it was concluded, that the space group as well as the unit cell

parameters were in good agreement with the previously published results for FMN binding to the riboswitch RNA (Table 5.8).[39] However, the crystals diffracted only up to 6.05 Å necessitating further optimization.

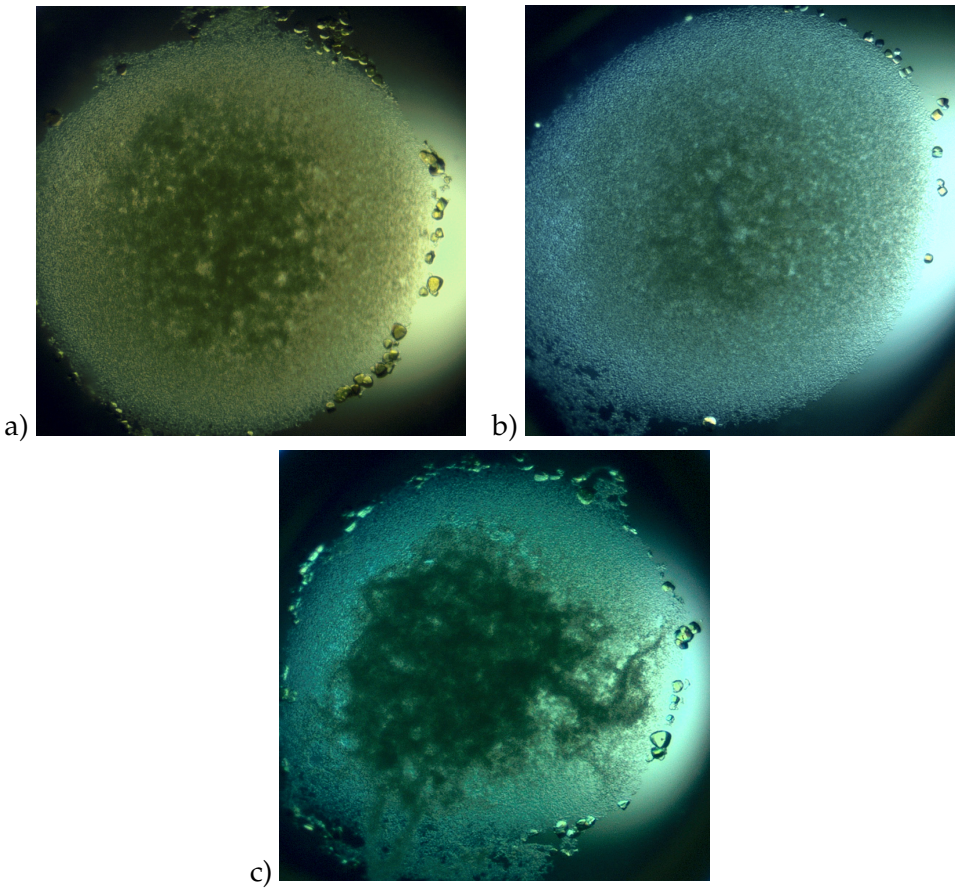


FIGURE 5.22: All of the displayed crystallization drops resulted from crystal growth by the hanging drop method, after both RNA strands had been reannealed in a buffer containing 50 mM K-HEPES (pH 7.5), 100 mM KCl and 2 mM $MgCl_2$ in the presence of 1 mM FMN. a) Crystal growth in the presence of a reservoir solution containing 0.1 M MES-sodium (pH 6.5), 0.2 M $MgCl_2$ and 10% PEG 4000. b) Crystal growth in the presence of a reservoir solution containing 0.1 M MES-sodium (pH 6.5), 0.1 M $MgCl_2$ and 13% PEG 4000. c) Crystal growth in the presence of a reservoir solution containing 0.1 M MES-sodium (pH 6.5), 0.2 M $MgCl_2$ and 13% PEG 4000, after reannealing in the presence of 2% DMSO.

TABLE 5.8: The collected data for the crystallization experiment with the annealed two-piece riboswitch RNA construct are compared to the experimental data deposited in the protein data bank.

	3F4E	**Collected data**
Space group	$P3_121$	$P3_121$
Cell dimensions a = b, c (Å); $\alpha = \beta$, γ (°)	71.47, 140.55; 90, 120	69.66, 138.60; 90, 120
Wavelength (Å)	1.08	1.00
Resolution range (Å)	20-3.05	55.32-6.05

5.5.2 Crystallization of the two-piece riboswitch RNA construct with screening compounds

Up to this time point there was only two structures deposited in the protein data bank (pdb:5c45 and 5kx9, resp.) with a small molecule other than a flavin-containing compound bound to the FMN riboswitch of *F. nucleatum*.[61,65] The small molecule ribocil was added prior to the annealing step of the RNA strands for cocrystallization in a very similar fashion to FMN in the presence of 2% DMSO. Control experiments were performed to elucidate, whether the presence of DMSO affects the crystal growth with FMN as ligand in conditions crystals were obtained with previously. Indeed, for some crystallization drops the previous results were confirmed and crystals grew even when DMSO was added to the system. Alongside to precipitation crystals grew at the borders of the precipitate (Figure 5.22 c) with a reservoir solution containing 0.1 M MES-sodium (pH 6.5), 0.2 M $MgCl_2$ and 13% PEG 4000. The shape of most of the crystals was less well-defined compared to the crystals grown without the addition of DMSO.
Next, the reannealing step of the RNA strands was adapted to the one described for the co-crystallization of ribocil. Here, reannealing of the nucleic acid strands was carried out in a buffer containing cacodylic acid, sodium acetate and magnesium chloride (Chapter 2, 3.6.1) and a final concentration of 2% DMSO when ribocil was added.[61,65] The reservoir solution also had a completely different composition with a 1.5 units lower pH and sodium acetate as buffer component (0.1 M sodium acetate (pH 5.0), 0.2 M $MgCl_2$ and 7-11% PEG 4000). Initial attempts were made to co-crystallize for example the SPR screening hit compound **48**. Precipitation throughout all the wells occurred and no single crystals were obtained. Due to solubility issues compound **48** was only applied in concentrations of 100 µM each, whereas the highly active compound ribocil was used in a final concentration of 1 mM for the crystallization experiments.[61] Therefore it remains difficult to judge, whether the conditions were suboptimal or the concentrations of the screening hits were just too low to generate ligand-bound crystals.

5.6 Summary of the experimental testing results

The virtual screening hits of both docking approaches, an unbiased large database screening and a search for substructures with scaffolds similar to the uracil-like edge of FMN, were experimentally tested to confirm their binding affinity to the FMN riboswitch. In order to determine the K_D-values of the virtual screening hits, a fluorescence quenching assay exploiting the intrinsic fluorescence of FMN was used in a displacement method. Two compounds, compounds **28** and **42** (Table 5.1), showed activity, but could not be validated as true binders with the ITC displacement assay used as follow-up assay. In this ITC displacement method one compound, compound **44**, increased the apparent dissociation constant of the FMN-RNA interaction

≈ twofold when present in the reaction mixture, which corresponded to an increase > 3 SDs, and therefore was identified as FMN riboswitch ligand (Table 5.2). Aiming to start small SAR-studies close analogues of compound **44**, the compounds **45-49** (Table 5.3), were purchased. However, binding of analogues could not be confirmed with the ITC displacement method (Table 5.4). Furthermore, a SPR approach to determine ligand binding was established. One of the analogues of compound **44**, compound **48**, was active in the SPR-based assay (Figure 5.18), which supports the hypothesis that a novel scaffold binding to the target RNA was identified in this study.

Chapter 6

Discussion

6.1 Abstract

The overall goal of this thesis was to discover novel FMN riboswitch ligands. For that purpose, a suitable virtual screening protocol was derived to guide the selection of promising candidates and different binding assays were established to experimentally confirm the predicted binding to the target structure. This resulted in one confirmed hit compound and the initiation of preliminary SAR-studies with closely related analogues. The following points that stand out from this study will now be discussed in more detail:

1. General strengths and drawbacks of the chosen assay methods,
2. problems associated with each established assay and an interpretation of the assay results,
3. interpretation of the challenges accompanying both virtual screenings, and
4. evidence for the classification of compound **44** as a promising hit.

6.2 Discussion of the established assays and the experimental results

6.2.1 General interpretation of the experimental data

The discussion of the experimental data starts with careful considerations on the established assay methods employed to confirm virtual screening hits. In the course of this study, novel ligands binding to the aptamer domain of the *F. nucleatum* FMN riboswitch were identified. All three established assays, the fluorescence quenching assay, the ITC displacement assay and the SPR-based method, were generally suitable to measure the interaction between small molecules and the FMN riboswitch RNA, as they all resulted in dissociation constants between 100-300 nM for the interaction between the aptamer domain and FMN. In the fluorescence quenching assay an average K_D-value of $\approx$ 211 nM was determined for the control measurements in the absence of test compounds (Figure 5.1), whereas the K_D-value was almost exactly

the same in the ITC assay ($\approx$ 216 nM) and saturation binding was also achieved (Figure 5.6). Finally, the SPR measurements employing the two-piece riboswitch RNA construct originally developed for crystallization resulted in a K_D-value of $\approx$ 150 nM for the riboswitch-FMN interaction when plotting the relative responses in the steady-state equilibrium against the FMN concentration (Figure 5.17).
The most promising results were obtained for compound **44** and its analogue compound **48**. Compound **44** generated promising results in the ITC displacement assay and its close structural analogue compound **48** performed best in the SPR-based assay. Both, the binding affinity of FMN to its target structure and ΔH upon titration of FMN against the riboswitch RNA, were considerably decreased in the presence of compound **44** (Figure 5.7 c and Table 5.2). Although the affinity of compound **44** was not directly confirmed, the very closely related compound **48** exhibited activity in the SPR measurements. Compound **44** was no longer available at that time as ready-made final product and it is very likely that both substances behave similar against the target structure, since marginal structural differences between compound **48** and its parent compound were present (Table 5.3). However, the aqueous solubility of compound **48** (Table 5.3) was even worse compared to its parent compound and therefore the displacement properties in the ITC experiment were less pronounced (Figure 5.9 d and Table 5.4). The concentration of compound **48** used was even below its expected K_D-value of $\approx$ 200 µM. The solubility of compound **49** is even lower than for compound **48**, which might explain that binding could not be determined with the assay, albeit both structures are almost identical.
Taken together, these results suggest that two novel FMN riboswitch ligands were found. They contain a furopyrimidine scaffold fused to either a pyrimidino (compound **44**) or a pyrido (compound **48**) scaffold (Figure 6.2). Compound **44** and its analogue compound **48** both were predicted to interact with A99 in a similar manner as the uracil-like edge of FMN. Therefore, this could serve as a starting point for further SAR-studies.
During the hit conformation and validation, the two major problems were the low affinity of the ligands and the poor solubility in aqueous solutions. The low affinity was expected in preparation of the assay methods. Virtual screening methods are most often used to identify novel chemical scaffolds rather than to directly retrieve compounds with nanomolar activity. A survey of the hit identification affinity cut-offs in structure-based design approaches revealed that they are typically ranging from 1-500 µM.[155] In the present study, compounds at the higher end of this range (100-500 µM) were considered as hits. This can be explained by the fact that the lack of the phosphate group in riboflavin leads to an affinity drop in the size of three orders of magnitude in comparison to FMN. In the same fluorescence-based assay as used in this study, the K_D was $\approx$ 39.8 µM compared to 37.5 nM for FMN.[39] Since riboflavin fits nicely into the binding pocket, it was expected that based on the methods employed during the study, an active compound with high shape complementarity and favourable electrostatic interactions would exhibit affinities in the

three-digit µM-range as well. The low affinity still remained a problematic feature of the virtual screening hits when being experimentally tested. The ITC displacement assay and the fluorescence quenching assay were both based on the test compound displacing FMN from the binding site. Thus, the binding affinity of the screening hits can only be determined indirectly by measuring a change of the apparent dissociation constant of FMN binding to the RNA. It is self-evident, that a higher affinity of the test compounds would have caused a larger change of the apparent K_{D} of FMN binding to the riboswitch RNA. In the present case, the largest differences in the apparent K_{D}-value were a duplication when adding one of the test compounds. Even after several repeats and by inspecting the errors of the respective method it remained difficult to decipher for some test compounds, whether or not a significant change or a high noise level affected the interaction.
These findings are directly associated with the second problematic feature of most of the screening compounds that have been tested to confirm binding, the solubility. In the equation used to calculate the dissociation constant of the competitor compound (3.4), the competitor concentration is found in the numerator of the fraction. Due to the expected low affinities and thus high $K_D^{complig}$-values, with a higher concentration of competitor substance a larger shift of the binding curve belonging to the control interaction would result. Therefore, it was attempted to apply the test compounds in the highest possible concentration they were still soluble in an aqueous solution supplemented with up to 5% DMSO. Assuming affinities between 100-500 µM a concentration of the low-affinity competitor at least matching the expected K_{D}-value is ideal.[184] Screening hits that were derived from the whole database screening had an overall higher solubility throughout the assay. About 70% of all purchased compounds were amenable to being experimentally tested at concentrations $\geq$ 500 µM. For the compounds derived from the substructure search this value fell below 50% of all compounds. The substructure search with its stricter structural requirements for the molecular pattern interacting with A99 and a higher rate of compounds addressing the face-to-face stacking interactions with adenines A48 and A85 altogether resulted in compounds with higher similarity to the reference ligand FMN. Riboflavin itself alongside with the other isoalloxazine-containing analogues has an intrinsically low water solubility due to the lack of the solubility-promoting phosphate group.[204] The high molecular planarity could serve as an explanation for the poor solubility in water. It is now well-established that the disruption of molecular symmetry and planarity can have tremendous influences on the water solubility of the compounds.[205] Even among the subset of compounds purchased for experimental testing molecule pairs proving this hypothesis can be found. The bicyclic core predicted to be involved in the interaction to A99 of the riboswitch RNA in compound **1** and **5** was a xanthine core in both cases. Compound **5** with its different substitution pattern at the xanthine core is overall less planar with larger dihedral angles (Figure 6.1) and was soluble to a concentration of 1 mM in the assay buffers. Compound **1** in contrast was only soluble up to 250 µM.

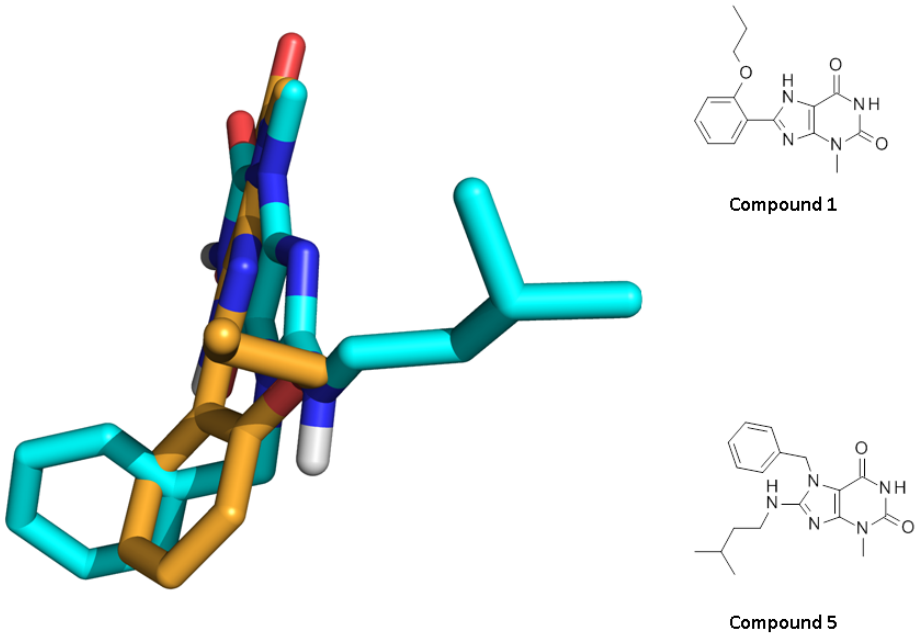


FIGURE 6.1: Overlay of the docked poses for the compounds **1** (orange carbon atoms) and **5** (cyan carbon atoms). The substituents at the xanthine core of compound **5** disrupt the planarity of the compounds stronger than in compound **1**. A 2D depiction of both molecules is displayed in addition on the right-hand side.

The three methods used to detect binding of screening compounds to the riboswitch RNA can be regarded as complementary to each other. Initially, it was planned that docking hits are confirmed as true binders, if they exhibit activity in two independent assay methods. But as already mentioned above, the low affinities made it difficult to detect activity especially in the competitive assay methods and the properties of some hit compounds limited their applicability to all three types of methods. Due to this fact it was almost impossible to retrieve compounds that are actives throughout all assay methods. None of the assay methods employed on the FMN riboswitch can be seen as a stand-alone method to screen large sets of molecules for binding affinity to the target RNA.

The fluorescence quenching assay and the ITC experiments both were indirect methods. Therefore, the binding affinity of the screening hits can only be derived by observing the interaction between the high-affinity ligand FMN and the target RNA. For a direct determination of binding constants between the RNA and the novel ligands prohibitive high concentrations of RNA of up to several mM would have been necessary in the fluorescence quenching assay for screening hits with intrinsic fluorescence exhibiting three-digit µM affinity, when the RNA is titrated against a constant ligand concentration.

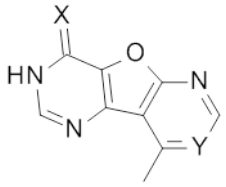


FIGURE 6.2: A 2D depiction of the common molecular structure of compound **44** and analogous compounds **48** and **49**. X = O or S; Y = C or N

In terms of establishing a screening cascade for further studies, the fluorescence-based assay should be used first, followed by additional complementary assay methods. The fluorescence-based assay has a higher throughput than the ITC displacement assay and measurements can start faster compared to the SPR-measurements, where an immobilization is necessary as first step. Restrictions due to compound properties on the other hand are a disadvantage of the assay based on the inherent fluorescence of FMN. For the ITC measurements, there is in principle no limitation on the molecular structures of the possible ligands, since it is a label-free method. Furthermore, active compounds are slightly easier to detect in the ITC displacement assay. Not only the alterations in the apparent K_D-values of FMN interacting with the riboswitch in the absence and presence of the test compounds can be evaluated. A decreased free binding enthalpy is an additional indicator of competitive displacement. However, the throughput of the ITC assay is lower and the RNA consumption is considerably higher, which makes the method more expensive and also time-consuming because of the need for repeated purifications. An advantage of both methods is that compounds showing activity are likely to specifically bind to the riboswitch binding pocket.

The SPR measurements represent a direct method to determine the binding constants of the screening hits and at the same time also the rate constants can be determined. Certainly, this is an advantage of the method. However, the RNA has to be labelled for using this method, which might have an influence on the RNA folding, even though activity was retained throughout this study. Generally, this method also provides a fast throughput. Within a few hours after diluting all the samples all the purchased compounds can be tested at one single concentration to roughly estimate the classification into binders and non-binders and to select compounds for further studies. Yet, with the current assay setup the surface remains not stable over a long time period and the immobilization has to be repeated on fresh sensor chips, which makes the method very expensive. Even though it is a direct method, SPR is also a very sensitive method and therefore the detection of nonspecific binding events is higher in comparison to both indirect methods. Care has to be taken for interpretation of the results and orthogonal methods are necessary for validation.

The most striking result of the control experiments carried out before each round

of testing screening compounds by the respective assay methods with FMN and riboswitch RNA alone was the fact that the determined K_D-values were consistently 5-10 times higher than previously reported for the interaction between the 112-nt aptamer domain of the *F. nucleatum* FMN riboswitch and its cognate ligand. Albeit the plasmid DNA carrying the sequence information for the RNA was obtained from the same research group, where a K_D-value of 37.5 nM was determined by the fluorescence quenching assay[39], in the present study values of 150-300 nM were obtained when using this assay. Most of the ITC experiments confirmed these findings and also the SPR measurements indicated a similar result with a tendency to slightly higher affinities in comparison to the two competitive methods. The altered behaviour was mainly attributed to a not fully active RNA.

It can be assumed that not all of the RNA species were completely refolded after purification via denaturing PAGE. This was revealed by all three of the assay methods. During the fluorescence quenching assay the fluorescence reached the lower plateau of the normalized fluorescence at higher RNA concentrations (Figure 5.1). However, the FMN fluorescence at the highest RNA concentration was still much higher than the fluorescence corresponding to the negative control containing only buffer, indicating that possibly not all of the RNA species were capable of binding FMN and altering its fluorescence properties. Due to a too high RNA consumption RNA concentrations higher than 10 µM were not tested to elucidate whether this would further increase the quenching. Since saturation of riboswitch binding sites was achieved during the fluorescence quenching assays, it is also possible that the residual FMN fluorescence in the bound state is as high as one third of the free fluorescence. The hypothesis that inactive and misfolded RNA species led to these observations is supported more strongly by the results of the two other assay methods. With the ITC displacement assays it was easier to observe and assign the folding problems. The interaction between the riboswitch RNA and FMN clearly follows a 1:1-stoichiometry. Throughout all the measurements the *n*-value obtained from the ITC measurements was in a range between 0.5 and 0.7. The RNA and FMN concentrations were calculated by measuring the absorbances and using the respective molar extinction coefficients. Inaccurate concentration determination as reason for too low stoichiometries was therefore excluded. Thus, the presence of improperly folded RNA species was the most probable reason for unexpected stoichiometries. This can often be observed for larger RNAs.[206] Several measurements with one batch of purified RNA were sometimes conducted over 3-4 days and the stoichiometry did not further decrease over time. Therefore, a simple contamination with for example RNAses was also excluded as reason for these observations. SPR measurements were also indicative of inactive RNA species in the reaction mixtures. The maximum response upon binding of saturating FMN concentrations to the sample sensor surface immobilized with riboswitch RNA was far below the theoretical maximum response dependent on the molecular weights of ligand and analyte and the

overall immobilization level in all different SPR setups tested (Figure 5.12 and Figure 5.13 for the full-length aptamer and Figure 5.16 and Figure 5.17 for the two-piece construct, respectively). Since labelling of the RNA was necessary for carrying out SPR measurements, the label itself could compromise the global fold of the RNA.
Apparently, for further studies the folding problem has to be addressed and different folding protocols should be tested, since it will be less difficult to identify compounds with binding affinity to the FMN riboswitch RNA in the presence of completely active RNA species. Several parameters could influence a proper refolding. It is for example conceivable that the refolding process is dependent on the RNA concentration. There are studies, where RNA folding was not only dependent on the ion concentration, but also on the RNA concentration or more precisely on the molar ratio of divalent ions and RNA.[207] Based on this observation, the Mg^{2+}-concentration would have to be adjusted.
To ultimately validate a correct prediction of the binding mode by the docking protocol employed on the FMN riboswitch, confirmation by X-ray crystallography would be desirable.[162] Structure determination by either X-ray crystallography or NMR still remains elusive to ensure that the compounds specifically bind to the riboswitch binding pocket. All attempts to co-crystallize the novel ligands with the riboswitch RNA failed so far.

6.2.2 Interpretation of the fluorescence quenching results

Due to the inherent fluorescent properties of FMN it was obvious to exploit this property for the development of a binding assay. The main two compound properties evoking difficulties associated with testing the virtual screening hits, the affinity and solubility, have already been described extensively in the previous subsection. In the case of the fluorescence quenching assays, it was especially difficult to ensure dissolvation of the compounds throughout the whole assay. Pretests determining the solubility borders in pure assay buffer were performed in transparent tubes and these tubes were subsequently spun down by centrifugation to visually inspect any precipitate. However, a delayed precipitation was not detectable within the black 96-well microplates. The addition of RNA could change the properties of the solutions. Magnesium ions for example are incorporated during the refolding process of the FMN riboswitch RNA and the altered free concentration of Mg^{2+}-ions could influence the solubilities. Predictions without experimental validation are difficult, since salt effects in both directions are possible depending on the compound of interest and type of salt.[208] This substantiates the hypothesis that some compounds might precipitate less in a reaction mixture without the addition of the riboswitch RNA. All compounds were applied in the highest concentrations possible without provoking precipitation in assay buffer alone and without exceeding a DMSO concentration of 5%. Even in such small amounts stacking interactions can be decreased by DMSO and local disorder can be increased. In studies with the

HIV-1 TAR RNA as well as the bacterial ribosomal A-site RNA also using fluorescence intensity measurements as readout, 0.3-4-fold reductions of measured binding affinities were observed with 5-10% DMSO present in the reaction mixtures. Significant increases of the dissociation constant were mainly observed with 10% DMSO, whereas the addition of 5% DMSO caused less perturbations.[209] Due to this fact an increase of the DMSO concentration to enhance the solubility of the compounds and increase the applied compound concentrations was not an option for the fluorescence-based assays. In addition, with DMSO concentrations up to 10% the affinity of FMN to its target RNA could possibly also be lowered to a K_D-value of ≈ 1 µM assuming a 4-fold decrease of the binding affinity. This would have necessitated higher RNA concentrations to perform the assay thereby hampering a high throughput of this method.

As already mentioned (6.2.1) the fluorescence quenching assay was not an universally applicable method and was suffering from limitations due to inherent fluorescence of some of the screening compounds. More than 20% of the purchased compounds derived from the substructure-based virtual screening were not amenable to this kind of assay, whereas only two out of 20 compounds were affected by this issue during the experimental validation of the screening hits arising from the larger screening. This was foreseeable, since the privileged substructure search enforced the identification of compounds with higher structural similarity to FMN with a high degree of planarity and rigidity to enable analogous interactions with the RNA binding pocket like the isoalloxazine heterocycle. Often the transition from less planar to more rigid molecules is accompanied by stronger inherent fluorescence properties.[210] Actually, five compounds arising from the virtual screening with privileged substructures - **30**, **41**, **40**, **34** and **44** (Table 4.4) - exhibited autofluorescent properties. Direct measurements of the fluorescence quenching based on the inherent fluorescence of the these virtual screening hits were not carried out, since expected K_D-values of around 100 µM would have necessitated RNA concentrations of up to 1 mM. However, this was not feasible. Those five compounds, except for compound **34**, have in common that they consist of a three-membered heterocyclic ring system with rigid and flat properties. Enhanced planarity and rigidity of molecules usually is an indicator of high fluorescence quantum yields.[210] Overall, the substructure search increased the probability of identifying structural hits with higher similarity to FMN itself and thereby the higher rate of autofluorescent hit compounds can be explained. Interestingly, compound **42** interfered less with the assay setup compared to compound **34**, even though both compounds share a closely related structure. A possible mechanism by which the fluorescence of compound **42** might become diminished in comparison to compound **34** is the internal heavy atom effect. Introduction of a heavy atom into the molecular system can perturb the whole fluorescent properties. Both compounds are substituted by large halogen substitutions, but the larger bromine atom can perturb the fluorescent system stronger than the chlorine atom in compound **34**.[211]

Another main source of fluorescence assay interference, the inner-filter effect, was unlikely to be observed with the setup used throughout this work, at least in terms of identifying false positives as hit compounds. Secondary inner-filter effects occur when excitation or emission light is attenuated by absorption of the test compound in the respective wavelength area. This would lower the fluorescence quantum yield of FMN, but less quenching and thereby higher fluorescence of FMN have been the endpoints of the measurement. The inner-filter effect would therefore mainly affect the false-negatives hit rate.[212] However, particular measurements aiming to decipher possible secondary inner-filter effects of the tested virtual screening hits were not carried out.

The screening for privileged substructures resulted in two hit compounds with calculated binding affinities below 500 µM: Compounds **28** and **42** (Table 5.1). Both compounds were not validated as a true binder by the downstream assay methods. Compound **42** was applied in a concentration of 200 µM in the fluorescence-quenching assay and later in concentrations of 100 and 200 µM in the ITC displacement assay, because problems with the solubility at the higher concentrations were observed during the ITC, where delayed precipitation was much easier detectable. After the titration the sample was directly removed from the sample cell and precipitate was visible in transparent glassware and tubes. During the ITC experiments both the sample cell solution and syringe solution contained the same amounts of **42** and no alterations in the binding affinities of FMN to its target structure were observed in the presence of the compound. Presumably, the ITC experiments were more robust and less sensitive to particles within the solutions than the fluorescence measurements. In turbid solutions reflections from particles can result in light scattering interfering with the fluorescence spectra and disturbing the evaluation. Depending on the proximity of excitation and emission wavelengths significantly higher or lower read-outs than expected are possible and this is not easily predictable.[213] Hence, compound **42** should be classified as a false positive hit of the fluorescence quenching assay. Compound **28** might have been wrongly classified as an active substance for the same reason. However, the concentration used throughout the ITC displacement assay was only slightly lower than during the fluorescence-based assay (200 vs. 250 µM). The binding affinity to the FMN riboswitch RNA was not confirmed by the orthogonal ITC assay method and due to the higher robustness of the ITC assay and the higher explanatory power of this method the compound was also categorized as a false positive hit of the fluorescence-based assay.

6.2.3 Interpretation of the ITC displacement assay results

The ITC displacement assay was expected to be more robust than the fluorescence-based assay and easier to interpret. Previous results published for ITC measurements with the purine riboswitch supported this hypothesis.[214] The aptamer domain of this adenine riboswitch for example was well characterized thermodynamically and kinetically by ITC. Here, a large favourable enthalpy attenuated by a relatively strong entropic term was found resulting in a free energy of binding of around 10 kcal/mol for 2,6-diaminopurine. The entropic penalty was attributed to conformational changes within the RNA to accommodate ligand binding in this study. Similar results were expected for the study with the FMN riboswitch aptamer domain. Even though a ligand-induced conformational stabilization of the joining regions around the binding pocket is observed upon ligand binding (Figure 1.8), a less negative $T\Delta S$ term was expected, since the aptamer domain is known to be globally pre-organized in the absence of ligand.[64] An entropic penalty during FMN binding to the riboswitch aptamer domain was suggested as well. The experimentally determined entropic penalty ($-T\Delta S$) indeed was as high as +9.65 kcal/mol (Figure 5.5), but clearly less pronounced compared to the adenine riboswitch for example, where presumably stronger local rearrangements upon ligand binding caused an entropic penalty of +29.6 kcal/mol.[214] Metabolites binding to riboswitch aptamer domains are typically up to 90% buried in the binding site.[24] Therefore, particularly the highly flexible ribose chain of FMN has to be conformationally restricted upon ligand binding presumably leading to unfavourable entropies.

Due to the fact that the buffers in the sample cell and syringe have to be matched exactly to avoid large heat of dilutions long dialysis experiments or several rounds of buffer exchange were necessary to adjust the ligand and analyte solutions.[215] Against this background, not all virtual screening hit substances were tested in the ITC displacement assay; all compounds that did not interfere with the fluorescence quenching assay and at the same time were not indicative of being active were excluded from further experiments.

An advantage of the ITC method is the possibility to reuse the riboswitch RNA samples almost without loss of activity. After completion of the FMN titration into the sample cell containing the aptamer domain of the *F. nucleatum* FMN riboswitch, the sample can be recovered from the sample cell and buffer-exchanged into 8 M urea to completely unfold the RNA and release FMN from the binding pocket. The success of the method can be tracked easily in the case of FMN because of its inherent yellow colour. Concentrating the samples recovered from the ITC sample cell resulted in an intensified yellow colour as long as FMN was bound to its target RNA. The addition of urea on the other hand led to the migration of FMN into the filtrate and disappearance of the colour. Subsequently, the RNA sample could be refolded as before and applied to new ITC measurements. In the present study, this procedure was carried out without major differences in the performance of the reused RNA samples compared to the first use of the particular batches. The ITC displacement assays with

the virtual screening hits for example (5.2.2) were carried out with batches of freshly refolded purified RNA samples and reused batches, respectively. Measurements for one and the same compound during one day were of course not performed with a mixture of different batches. The K_{Dapp}-values were ranging from 100-400 nM (Table 5.2). However, these variations were independent of the re-use of a certain batch and rather dependent on other factors, such as the refolding process for example. Theoretically, the solutions of the fluorescence-based assay could be reused as well. However, it is less convenient and feasible to recover all the solutions from a 96-well plate than to collect sample solutions of three consecutive ITC measurements.

The control experiments to determine the thermodynamic parameters of the interaction between FMN and the riboswitch RNA without the addition of a competitor ligand revealed a highly favourable reaction enthalpy (ΔH) at 30 °C of -18.94 $\pm$ 1.80 kcal/mol counterbalanced by a relatively large entropic penalty of +9.65 $\pm$ 1.89 kcal/mol. Overall, FMN binding consistently yielded a free binding energy of -9.29 $\pm$ 0.23 kcal/mol (Figure 5.5). The resulting thermodynamic values are in good agreement with previously reported ITC data for the interaction between FMN and the 141-nt aptamer domain of the *ribD* FMN riboswitch of *B. subtilis*. Here, the enthalpic term ΔH was determined to be -14.5 kcal/mol and the entropic contribution -TΔS was +5.7 kcal/mol also resulting in a free energy of binding of around -9 kcal/mol.[67]

SAXS data revealed that the FMN riboswitch does not require metabolite-binding and only minimal amounts of Mg^{2+}-ions for global folding.[67] In the case of the FMN riboswitch the high entropic penalty is therefore at least partially associated with the ligand properties. The flexible ribityl side chain of FMN with its several polar functions has to be desolvated and rigidified upon transition from the free to the bound state.

Among all of the screening compounds tested experimentally by ITC displacement assays, compound **44** was the only substance clearly increasing the apparent dissociation constant of FMN binding to the RNA. The K_D-value of the control interaction between riboswitch RNA and FMN was duplicated in the presence of 250 µM of compound **44** (Table 5.2). In addition, the binding enthalpy was decreased to about 65% of the heat evolved in the absence of compound **44** together with a lowered entropic penalty for the transition of FMN from the free to the bound state. This drop of binding enthalpy was also considered to be significant, since the increase of ΔH from -67.08 kJ/mol to -43.61 kJ/mol in the presence of compound **44** (= increase of 24 kJ/mol) was > 3 SDs ($\approx$ 21 kJ/mol). The entropic term of the interaction even decreased to about 20% of the initial value in the absence of the competitor. The change of all these thermodynamic parameters gives evidence that compound **44** is a true binder to the FMN riboswitch. Due to the occupation of the binding sites by compound **44**, less heat was released upon binding of FMN to the riboswitch RNA, which was especially apparent at the beginning of the titration (Figure 5.7c). Furthermore, to keep the compound solubilized in the ITC assay only 2% DMSO were

necessary and the risk of detecting alterations associated with disruption caused by the cosolvent were rather low. To detect binding of a weak-affinity ligand, it is usually sufficient that a concentration of competitor close to its K_D is added to the mixture in an ITC displacement assay.[184] This requirement was met here. The K_D-value for compound **44** was calculated to be 244 µM, which is almost exactly the concentration used throughout the assay.

Despite their very close structural similarity, the purchased analogues of compound **44** were less soluble in aqueous solutions than the parent compound. There were no large changes in the binding affinities expected when comparing especially the compounds **48** and **49** with compound **44**. The concentrations of the analogues (50-100 µM) in the ITC assay were far below the expected K_D-values which implies the difficulty to identify a binding event. At the time of the investigations compounds with exactly the same structure of the polynuclear aromatic ring system were not available. As already discussed earlier, the solubilizing effect of heteroatoms can be additive and this could explain the differences in the solubilities between the analogous compounds and the parent compound **44**.[216]

The best ITC results for all of the analogues were obtained for compound **48** (see 5.4 and 5.9). The apparent K_D-value for the interaction between FMN and the riboswitch RNA only increased about 22%. This deviation corresponded to a borderline situation when considering the typical data variation band, which is defined as the mean $\pm$ 3 SDs, to choose a suitable cut-off to distinguish between active and inactive compounds.[201] In this test series, 88.2 nM corresponded to the threefold of the SD. Therefore, a significant signal was expected to cause at least an increase of the apparent dissociation constant from $\approx$ 233 to $\approx$ 321 nM. Though, in the presence of compound **48** the K_D-value was only increased to $\approx$ 285 nM (Table 5.4). Therefore, the increase of the apparent K_D-value was below this cut-off. However, the heat evolved upon binding of FMN to the RNA decreased to about 70% of the initial value in the absence of compound **48**. These findings were very close to the enthalpic change $\Delta\Delta H$ observed in the presence of compound **44**. It was already discussed that this might be an indicator of binding sites being occupied by this compound when the heat of interaction upon titration of FMN is diminished.

The categorization of compound **49** on the contrary is more difficult. Data of the ITC displacement assay do not indicate binding of this compound to the FMN riboswitch RNA despite being almost identical to compound **48**. In principle the oxygen in compound **49** should serve as a stronger hydrogen bond acceptor in the interaction between A99 and the ligand in comparison to the sulfur atom of compound **48** due to the higher electronegativity.[217] Therefore, a lower solubility of compound **49** in aqueous buffers was regarded as reason for the inactivity of this ligand in the ITC displacement assay.

The thione sulfur atom of compound **48** has to participate in the interactions between the compound and A99 by acting as the H-bond acceptor. Admittedly, sulfur atoms

are less often involved in hydrogen bonds than oxygen atoms and most often not included in reviews surveying common molecular interactions[73,218] The general belief for quite a long time was that hydrogen bonds to oxygen atoms are stronger than to sulfur due to stronger electrostatic effects. However, sulfur-containing hydrogen bonds are not only observed involving cysteines and methionines in proteins[219], but also in Watson-Crick base-pairing. Especially at the wobble positions of tRNAs 2-thiouridine (s^2U) and 4-thiouridine (s^4U) are found among a lot of other base modifications and have a great influence on biological functions. Wobble base pairs for example involving the sulfur atom of 2-thiouridine as hydrogen bond acceptor are known.[220] It is assumed that enhanced dispersion and flexibility around the sulfur atoms compensate for the loss in the electrostatic contribution.[221]

6.2.4 Interpretation of the SPR measurements

In the present study, three different approaches were tested to immobilize the aptamer domain of the *F. nucleatum* FMN riboswitch RNA. For two of these approaches the biotinylation of the target RNA was necessary before initialization of the SPR-based assay, while the RNA-DNA hybrid approach made use of specific base-pairing of an elongated FMN riboswitch RNA with a commercially available biotinylated DNA oligonucleotide. Especially when using periodate-based oxidation of the 3′ end good yields were achieved (Figure 5.10). Biotinylation of the full-length FMN riboswitch aptamer domain (Figure 5.11) and the insertion of 36 nucleotides to the 3′ end of the aptamer domain for hybridization with a complementary DNA oligonucleotide (Figure 5.15 a) revealed no loss of activity and resulted in roughly the same affinities for FMN as unlabelled riboswitch RNA when applied to a fluorescence quenching assay. Common to all of these methods was a low response upon ligand binding to the immobilized RNA. Therefore, in the control measurements using saturating concentrations of FMN, it was difficult to separate responses associated with a binding event from the noise level. This was probably partially attributed to inactive and misfolded conformations of the immobilized RNA or to misfolding due to the immobilization itself. Changes of the refractive index are exploited to follow an interaction and those are dependent on the thickness of the layer on the surface for example.[222] The thickness of the layer in turn changes more dramatically with higher concentrations and a high-molecular weight component binding.
Both, the direct, terminal biotinylation of the full-length aptamer domain (Figure 5.13 and Figure 5.14) and the indirect immobilization procedure via hybridization of an unpaired overhang at the 3′ end of the RNA with a biotinylated, complementary DNA oligonucleotide, did not generate reliable binding data. Reports on the successful application of each procedure to obtain kinetic data of riboswitch-ligand interactions have been published.[175,223] This is an indicator that none of these methods has proven its universal applicability to study such interactions. However, in the case of the RNA-DNA hybrid approach it is also possible that an optimized linker to elongate the aptamer domain would enable the determination of kinetic constants,

as the linker sequence was chosen solely based on a computational prediction of the resulting riboswitch fold (see 3.3.4). Comparative data for SPR measurements tracking the binding of FMN to its riboswitch target structure are not available. The aptamer domain of the AdoCbl-sensitive *btub* riboswitch of *E. coli* was elongated with a part of the expression platform at the 3′ end and hybridized with a complementary, biotinylated DNA oligonucleotide for the purpose of immobilization. A concentration series of the natural ligand was applied to determine the dissociation constant as well as the rate constants for the interaction.[175] The kinetics of $preQ_1$ and $preQ_0$ binding to different variants of the $preQ_1$ riboswitch aptamer domain of *T. tengcongensis* were determined by SPR measurements after direct biotinylation of the RNA and immobilization on streptavidin-coated SPR chips. Here, RNA was biotinylated at the 5′ end and the immobilization levels were typically between 3,800 and 4,000 RU.[223] Various attempts were made to reach similar immobilization levels with the biotinylated full-length FMN riboswitch RNA, such as repeating the loading step with a low flow rate or injecting very high concentrations, albeit with no success. The lower immobilization level is probably connected to the different sensor surfaces that were used. Instead of a ready-made streptavidin-coated sensor surface amine-coupling was used to immobilize streptavidin on a CM5 (carboxymethyl group-functionalized) dextran chip prior to capturing the RNA of interest.[223] Probably, a higher surface density of streptavidin was obtained by this approach.
The analysis of the results using the hybridization strategy can be hampered by the partial dissociation of the RNA-DNA hybrid during an analysis cycle. This will happen to all non-covalently captured ligands. However, the interaction between biotin and streptavidin is of a strength close to that of a covalent bond.[185] An advantage of the hybridization approach is the possibility to theoretically reuse one chip for the capturing of several RNA variants (e.g. mutants) with the same 3′ end in an "one-chip-for-all" strategy.[224] The interaction between a hybrid of 25-nt long complementary RNA and DNA strands is highly specific, but over a certain time period dissociation will occur. For the evaluation of the interaction between a small molecule and the immobilized RNA the dissociation has to be as small as possible. In the present study the free energy of the RNA-DNA hybridization was too low and this implicates that an optimized linker with stronger affinity to its complementary DNA oligonucleotide would be necessary to investigate the riboswitch-ligand interactions with this approach. The feasibility of SPR experiments with such constructs hybridized to a DNA oligonucleotide is dependent on the required sequence and a balance between facilitated hybrid duplex formation and unhindered global folding of the actual RNA. Here, the sequence of the elongated aptamer domain of the *F. nucleatum* riboswitch was based on the related sequence of the expression platform and modified to obtain a correct folding of the aptamer domain. However, the global folding of the RNA was only predicted and not experimentally verified. Thus, it is difficult to distinguish whether the method failed to generate reliable data due to improper folding or the lack of stable RNA immobilization.

Direct biotinylation of the $preQ_1$ riboswitch aptamer domain was carried out at the 5′ end of the RNA. In the correctly folded 112-nt long aptamer domain of the FMN riboswitch RNA both terminal ends base pair with each other (e.g. pdb-code: 3f2q[39]). Therefore, a different behaviour of the labelled RNA upon biotinylation of either the 3′ end or the 5′ end was not assumed, whereas in the case of the class I $preQ_1$ riboswitch both termini were situated at opposite ends in the correctly folded aptamer RNA (pdb-code: 3q50[223]). A higher immobilization level was achieved with the $preQ_1$ riboswitch in comparison to the FMN riboswitch. Moreover, the aptamer domains tested in these experiments had lengths of only ≈ 35 bases. Thus, binding of a small molecule changes the thickness of the surface layer and thereby the refractive index stronger in such cases than binding of FMN to the three times larger FMN riboswitch aptamer domain. This might explain, why detection of binding was possible between the $preQ_1$ riboswitch RNA and its metabolite ligands by SPR measurements with a directly biotinylated full-length aptamer domain in contrast to the FMN riboswitch.

The direct biotinylation resulted in an assay setup, where binding of FMN to the riboswitch RNA was generally detectable. Nevertheless, this approach did not deliver desirable results for FMN. Kinetic constants could be determined and were close to previously reported association and dissociation rate constants and K_D-values. Data on the kinetics of FMN binding to various aptamer domains of different bacterial species were previously derived from stopped-flow fluorescence measurements.[225] Though both methods are based on different detection principles, they should result in comparable data (Table 6.1), as they were based on exactly the same riboswitch sequence. The kinetic constants determined in the SPR measurements carried out in this study differ only about a factor of two from the previously determined values (Table 6.1).

TABLE 6.1: The kinetic constants of FMN binding to the *F. nucleatum* FMN riboswitch aptamer domain derived from SPR measurements employed on the biotinylated, full-length RNA and the two-piece annealed aptamer domain construct are compared to previous reports from Rode et al.[225] using stopped-flow fluorescence measurements.

Method	$\mathbf{k}_a$ **[10^5 M^{-1} s^{-1}]**	$k_\mathbf{d}$ **[10^{-2} s^{-1}]**	$K_\mathbf{D}$ **[10^{-9} M]**
Full-length (SPR)	2.93	2.17	74.29
Two-piece (SPR)	2.48	1.49	60.11
Full-length (fluorescence)	1.50	3.79	252.00

Collectively, the SPR approach employed on the biotinylated full-length aptamer domain of the *F. nucleatum* FMN riboswitch gave insights into the binding kinetics of FMN binding to its target RNA. However, difficulties with reproducing and evaluating these data properly together with a bad fit (Figure 5.14) when plotting the equilibrium relative responses against the analyte concentration necessitated an

alternative approach.
The most reliable SPR data were obtained when using the two-piece FMN riboswitch construct initially developed to facilitate crystallization. For this purpose the bottom strand B (see 3.3.3) was biotinylated at the 3′ end. A better performance and easier identification of differential responses when following the sensorgrams of the FMN concentration series in comparison to the biotinylated full-length aptamer domain were essentially accomplished by a higher maximum response of $\sim$ 4 RU, whereas with the former approach responses between zero and one RU were achieved. Theoretically, the dissociation of both RNA strands during the measurement could also be an issue when using this method similar to the RNA-DNA hybrid approach. Indeed, some dissociation or at least a drifting baseline was observed. An average reduction of the response level of 3 RU in relation to the overall response level in the sample flow cell after immobilization of ($\approx$ 33,000 RU) was observed per kinetic cycle. This corresponds to a reduction of only 0.15% of the total RNA immobilized from one cycle to the next. Therefore, the dissociation rate constant of the RNA duplex and the velocity of nuclease-mediated degradation were low enough to enable a full kinetic characterization of FMN binding. However, the overall activity of the riboswitch RNA during the SPR assay was still only 16% emanating of the theoretical R_{max} of 25 RU.
The differences in the overall performance of both of the approaches using either the full-length aptamer domain or the two-piece RNA construct were not attributed to a perturbed folding pathway by the 3′-end biotinylation. Since the sequence and global fold of 3′-end biotinylated full-length aptamer domain of the FMN riboswitch and the annealed two-piece construct engineered for crystallization do not substantially differ (only a modification of the 3′ terminus of strand A to facilitate crystal contacts[39]), the general structures of both labelled nucleic acids were not assumed to differ. An explanation for the better performance of the two-piece RNA construct might be the simplified folding before the SPR experiments, if two independent shorter RNA strands fold around the ligand during the annealing process. The refolding of the full-length 112-nt long aptamer domain after denaturing PAGE and biotinylation apparently requires optimized conditions to provide a similar level of active riboswitch RNA for immobilization compared to the biotinylated two-piece riboswitch construct.
The reliability of the generated data for the interaction between the FMN riboswitch RNA and its cognate ligand can easily be manifested by the closeness of the fitted to the raw sensorgrams and also by the overlay of two cycles corresponding to the injection of the same FMN concentration (Figure 5.16). Furthermore, the determined kinetic constants of the two-piece RNA construct only marginally deviated from the values obtained with the full-length RNA both using SPR and stopped-flow kinetic measurements (Table 6.1). In addition, the internal quality control within the Biacore® evaluation software indicated that the calculated binding constants were

within the specification range of the instrument, and also could be determined unambiguously without large errors. As opposed to the biotinylated full-length FMN riboswitch RNA with the two-piece RNA construct reproducibility was achieved, too. Several rounds of testing the control interaction between FMN and the immobilized RNA resulted in unaltered binding constants, even if the maximum response upon addition of saturating FMN concentrations decreased over time due to gradual loss of immobilized RNA. After 48 hours the performance of the two-piece riboswitch RNA construct became worse suggesting that measurements should be performed within two days following the immobilization.

As stable and reliable data was generated by this SPR approach, it was decided to further utilize the two-piece *F. nucleatum* FMN riboswitch for the analysis of the virtual screening hits and their derivatives. In the following, the reasons for the inclusion of certain compounds to the SPR-based analysis is given. Compound **42** was active in the fluorescence-based assay, albeit problems with the solubility during the assay occurred. Moreover, the initial results were not confirmed by the ITC displacement assay. Compound **34** is a close analogue of compound **42**, but exhibited inherent fluorescence. It was assumed that the compounds either both bind to the FMN riboswitch or both do not bind to the FMN riboswitch. A third method was sought to render a judgement on the binding properties of both screening hits. Compounds **41** and **40** were both chosen for the same reason. They had autofluorescent properties and could therefore only be tested in the ITC displacement assay, where they only marginally increased the K_{Dapp}-value for the FMN-RNA interaction. However, their presence slightly diminished the heat released upon binding of FMN to the riboswitch by up to 17% (compound **41**, Table 5.2) and for that reason an orthogonal assay method was applied. Compound **44** was the hit substance of the ITC displacement assay (5.2.2). It could not be included in the series of compounds applied to the SPR assay, because the compound was no longer available.

The injection of various screening compounds revealed only compound **48** as a compound with a response offset from all the other compounds tested during the SPR assay indicating binding to the immobilized RNA. Relative responses caused by injecting of compound **48** to the sample surface (3.4 RU) were almost exactly in the middle between the responses resulting from the application of 5 µM FMN (5.9 RU), a saturating FMN concentration ($> 10\times K_{\mathrm{D}}$), and the average response generated by all the other test compounds (0.9 RU).

In addition, the sensorgrams for compound **48** indicated specific interactions (Figure 5.19). Therefore, it was decided to further investigate the binding properties of this compound by the present SPR approach. It was assumed that a full characterization of the kinetic constants for compound **48** by measuring the relative responses of a concentration series would be impossible due to the solubility problems accompanying this compound. A concentration series of serial 1:5 dilutions of compound **48** was prepared to create similar conditions as used in the determination of kinetic constants for the control interaction with FMN. Thus, some lower concentrations in

the nM-range were included that were below 0.01-fold the expected K_D-value. Only for concentrations > 4 µM a reliable change in the response was detected (Figure 5.20). A detailed inspection of the corresponding sensorgrams revealed that even at the highest concentration of compound **48** curvature during the association phase was preserved. Upon addition of 100 µM of compound **48** a plateau by the end of the injection was not reached (Figure 5.21) and therefore an equilibrium binding analysis with a saturation curve to elucidate the K_D-value was not possible.[203] These findings support the hypothesis that concentrations in the low mM-range would be necessary to fully saturate the sensor surface upon addition of compound **48** as analyte. However, quantitative determination of the interaction did not yield reliable data and due to this fact a K_D-value could not be calculated in the present case. Taken together, the SPR measurements with compound **48** indicate that this derivative of compound **44** is a ligand of the *F. nucleatum* FMN riboswitch RNA. However, based on the available data only qualitative information on the binding properties was gained and no kinetic constants were obtained. The dissociation binding constant for compound **48** interacting with the FMN riboswitch RNA could neither be determined by the ITC displacement assay nor by the SPR-based assay.

Acting on the assumption that compound **48** truly presents a novel binder of the FMN riboswitch, the data obtained for the SPR assay also fitted well to the data resulting from the ITC displacement assay of the screening compounds excluding compound **44**. In the ITC displacement assay the only compound with indications for being a competitively binding ligand of the riboswitch RNA had been compound **48**, albeit the effects had been rather weak (6.2.3). It was even more difficult to detect binding below the expected K_D-value by displacing a strong binding ligand like FMN compared to investigate the binding properties of weak binders by directly applying them to the target RNA without the need for any displacement. However, the smaller heat evolved upon the titration of FMN into the RNA solution and the small increase of K_{Dapp} in the presence of compound **48** during the ITC displacement method further minimizes the probability of having only detected an artefact.

As already described previously, the derivative of compound **44**, compound **49**, is predestined to build even stronger Watson-Crick-like base pairing with A99 than compound **48**, since it even closer resembles the uracil-like molecule part. Addition of the highest possible concentration of compound **49** resulted in a weaker response of only 1.58 RU compared to its sulfur-analogue. In the first screening of several virtual screening hits by SPR measurements this response was higher than the cut-off set to distinguish between binders and non-binders among the screening compounds in the SPR assay (Figure 5.18). Testing a concentration series of compound **49** was subjected to the same restrictions as the kinetic determination of compound **48**. Again, assuming a K_D-value in the range of 100-300 µM illustrates that saturation binding was not obtainable with the actually added maximum concentration of this compound. Therefore, the explanatory power of this kinetic assay was rather weak.

Although a slight increase of the relative responses was observed for increasing concentrations of compound **49** between 2 and 50 µM, the responses were generally small. With a maximum relative response of 2.07 RU the highest concentration of compound **49** resulted in an even lower response than the injection of 25 µM of compound **48** (3.04 RU), even though for both compounds very similar K_D-values were assumed. Based on the present performance of compound **49** in both the ITC displacement assay and the SPR measurements it can not be assigned to the group of riboswitch ligands. Under the chosen terms of the assays ligand binding was not detected.

6.2.5 Discussion of the crystallization results

In principle reproducibility of the crystallization results under previously reported conditions for the co-crystallization of the annealed two-piece riboswitch RNA construct in the presence of FMN was proven. Crystals were obtained, but they diffracted poorly. Co-crystallization of the possible novel riboswitch ligands was also hampered by the poor solubility of the compounds. Probably a new screening for conditions, where smaller RNA concentrations are sufficient to crystallize the nucleic acids and thus the solubility of the ligands will be higher, is necessary to obtain crystal structures with small molecules bound to the FMN riboswitch aptamer domain.

6.3 Discussion of the virtual screening results

After the experimental results of the virtual screening compounds were obtained by the different assays used throughout the study, it is worth to discuss the predictive power of the chosen virtual screening protocols in terms of the suitability for molecular docking to discover novel FMN riboswitch ligands. In particular the questions will be addressed, which reasons might have caused failure to identify novel ligands with the unbiased large database screening and why only one hit compound in both screening was retrieved, even though it has to be considered that docking screens do not guarantee the discovery of confirmed hits. Furthermore, the higher probability of receiving a hit from the privileged substructure search rather than from screening the whole in-house library will be discussed.

6.3.1 Application of DOCK to identify new FMN riboswitch ligands by molecular docking

In addition to HTS campaigns, virtual screening approaches are more and more applied to identify novel binders for a given target structure. Strikingly, the application of virtual screening for this purpose is often accompanied with a higher hit rate as opposed to their HTS counterparts.[226] Virtual screening approaches have often served as basis for the identification of novel chemotypes that interact with

a certain target.[227] Certainly, most of the SBDD approaches in the past have been applied in the protein field. However, the purine riboswitch is one target structure of the same target class as the FMN riboswitch used in the present study, where novel ligands had been discovered by RNA-ligand docking.[130] The identification of ligands with novel chemotypes interacting with the FMN riboswitch by structure-based drug design was not accomplished up to now. Novel FMN riboswitch ligands with antibiotic activity were either close analogues of the natural ligand FMN[58,60] or discovered incidentally by a phenotypic screening.[61] Since riboswitch binding sites are perfectly adapted to allow recognition of their cognate ligand[34], it is intuitive to design artificial ligands as similar as possible to the natural ligand. Though, the natural ligand most often is an ubiquitous molecule and therefore could possibly interact with (human) off-target structures. This in turn requires the development of new ligands with a perfect balance between similarity and dissimilarity to the naturally recognized ligand.[17]

Docking algorithms previously applied to protein-ligand docking had been adapted to the special requirements of RNA and especially DOCK 3.5.54 has proven its applicability for the identification of new riboswitch ligands.[130] Therefore, it was decided to use this approach for docking against the FMN riboswitch RNA. The scoring function within the program is force-field based with three energetic terms contributing to the overall score. These energetic terms include energetic terms for vdW and electrostatic interactions as well as a term accounting for the desolvation penalty upon ligand binding. For RNA-ligand interactions, these terms are also valid and therefore only modified lookup tables with RNA atom types and partial charges are required. For the work here, the partial charges were assigned by using the AMBER force field, which had been proven to work for RNA molecules as well[228], and also was successfully integrated in RNA-ligand docking previously.[130]

Balancing the time-consumption on virtual screening and the accuracy of the prediction DOCK 3.5.54 is based on a methodology common to most of the docking algorithms: different ligand conformations are sampled, while the receptor is kept rigid.[114] In recent years, some progress was achieved with respect to keeping the receptor flexible during docking.[229] However, it was decided to perform the screening based on a certain FMN-bound structure (pdb-code: 3f2q) in the present study. The FMN riboswitch does not undergo large conformational changes upon ligand binding and adopts a highly pre-organized structure.[39] Thus, starting the docking process with an unbound structure or an average structure between bound and unbound state was regarded as unnecessary.

6.3.2 Finding the correct setup for the virtual screening process

A small sphere set focusing on the interactions of FMN with A99 was used to dock the ligands to the riboswitch binding site (Figure 3.2). This was mainly based on two considerations. Firstly, it was assumed that interactions to A99 are important for tight binding FMN riboswitch ligands. Besides, novel ligands with drug-like

properties were sought that are not as large and polar as FMN. In the following both considerations are further explicated.

For the flavin analogues, it is known that the phosphate group of FMN is crucial for high affinity to the FMN riboswitch. Yet still having rather high affinity, riboflavin exhibits a loss of binding affinity by nearly three orders of magnitude compared to FMN as observed in structural probing studies with the 165-nt *ribD* FMN riboswitch aptamer domain of *B. subtilis* ($K_D \approx 3$ µM vs. < 10 nM for FMN).[20] Between lumiflavin and riboflavin (Table 4.1) there is less difference in affinity as determined using the fluorescence quenching assay (59 µM vs. 39.8 µM).[39] This led to the assumption that the combination of favourable stacking interactions and the Watson-Crick-like base-pairing between the uracil-like edge of flavin analogues and A99 is the other key determinant for a strong interaction between the FMN riboswitch and its ligands.

The natural ligand coenzyme FMN is composed of a rigid, planar isoalloxazine core, the flexible ribityl moiety, and a phosphate group. These subunits are directed towards different structural domains.[39] Hence, the FMN binding site can be subdivided into two main subsites, the isoalloxazine ring recognition site and the phosphate-binding region (Figure 1.6), and both of them are bridged by a flexible linker in the ligand with presumably small contributions to the binding energy. Interactions of FMN to the adenines A48, A85 and A99 were all formed within the isoalloxazine-recognizing subsite and therefore it was straightforward to prepare the sphere set based on this subsite only.

More evidence for interactions to A99 being crucial for tight-binding riboswitch ligands was gained by comparing the crystal structures of different flavins bound to the target RNA, since a high degree of preo-rganization was observed especially in this subsite of the riboswitch binding pocket (Figure 1.8). The largely pre-organized structure of the FMN riboswitch does not mean that no conformational accommodations emerge when different ligands bind to the structure (Figure 1.9). The binding site rather adapts for the different flavin analogs through residue movements and structural rearrangements (Figure 1.10). A similar behaviour has already been observed for the *Thi* box riboswitches recognizing different analogues of thiamine pyrophosphate.[230] As expected from the very close structural proximity of FMN and riboflavin, the major conformational changes occur at the residues around the ribityl moiety and the phosphate-binding region (Figure 1.10).[39] It would have been more difficult to decide, which ligand-bound structure of the FMN riboswitch to use as starting point for the docking studies, if the entire binding site should have been covered by the spheres. The close conformational similarity of the riboflavin-bound and the FMN-bound structure in the region interacting with the heterocyclic system legitimized the choice for the FMN-bound structure as starting point. Though, it has to be underlined that ligand atoms are allowed to be placed in regions other than the sphere set used for matching.

We were seeking to identify drug-like molecules that obey the guessed rules concerning the molecular properties of drug candidates with oral bioavailability. Beyond the well-known "Rule-of-Five" with limits for molecular weight, logP and H-bond donors and acceptors, it is also the count of rotatable bonds and the polar surface area that predict the bioavailability. With increasing molecular weight, the number of rotatable bonds often increases simultaneously and is a negative predictor for the bioavailability.[231] The first filter applied to reduce the number of database molecules for the virtual screening was based on common cut-offs for physicochemical properties (Table 4.2). For the interaction with the phosphate-binding area a charged moiety increasing the polar surface area was believed to be favourable, at least until the antibiotic compound ribocil was discovered as a novel FMN riboswitch ligand.[61,65] Addressing the isoalloxazine binding region at the same time would require an enhanced molecular weight and molecular complexity. In the case of FMN itself, absorption and permeation are not an issue. Bacteria are capable of riboflavin biosynthesis and transport by genes that are in turn regulated on either the transcriptional or translational level by riboswitches[63] and human FMN absorption is mediated via a specific transporter in the gut.[232] The decision to place the sphere set used as matching points for docking only in the region of the heterocyclic core of FMN was based on these considerations.

The initial binding site setup was not sufficiently discriminating between compounds bearing both of the structurally required motifs for recognition of ligands in the subsite recognizing the heterocycle, the planar isoalloxazine ring system intercalating between A48 and A85 and the uracil-like edge, and compounds exclusively providing the planar ring system (structures provided in Table 4.1). This can be explained by the single energetic contributions that add up to the overall docking score. Regardless of slightly different desolvation penalties, the score based on the vdW energies between ligand and target was more than four-fold higher than the electrostatic contribution for all of the training set compounds. Thus, a worse electrostatic energy contribution in the set of non-binders among the training set compounds was probably outweighed by a better arrangement of the heterocycle and the side chain in the binding pocket with better geometries of the π-π-stacking interactions. The three distinct terms of the scoring function are all scalable and therefore fine-tuning would have been possible by scaling down the influence of the vdW-term. However, most of the recognition processes between riboswitch RNAs and their cognate ligands are mediated by heteroatom functionalities forming specific electrostatic interactions with the RNA and planar groups being sandwiched between two purine bases to allow for stacking interactions.[18] This interplay of interaction types constitutes the usually high affinity interactions of riboswitch-ligand complexes. Due to this fact it was decided to increase the discriminative power of the docking setup by adjusting the standard partial charges similar as it has been described earlier[192] and not by scaling the energetic terms of the scoring function. The partial charges

in the lookup files used to define the binding site provide average values for the respective atom types of every adenine, cytosine, guanine and uracil within the RNA. Standard force fields, such as the AMBER force field used in this study, describe the electrostatic interactions as fixed, atom-centered charges.[233] In a high dielectric medium like water polarization strongly affects the geometry of intermolecular recognition. To better account for this, polarizable force fields have been developed in the past.[234] These considerations contributed to the decision to adjust partial charges in this study.

As a first indicator of the predictive power of the docking setup, rmsd-values < 2 Å were obtained for riboflavin with the final receptor setup containing the adjusted partial charges for A99 (Table 4.1). It has to be emphasized that due to the high inherent flexibility of especially the ribityl moiety and due to the fact that almost the entire binding site is filled by the ligands, this is a quite good outcome. In contrast to riboflavin, the rmsd-value for roseoflavin was > 2 Å, which is usually considered as an undesirable outcome of a retrospective docking validation.[194,195] However, roseoflavin was docked into the FMN-bound receptor structure and even though no large structural rearrangements occur when different closely related ligands bind to the FMN riboswitch[64] the roseoflavin-bound structure deviates stronger from the receptor used for docking (e.g. position of A99) than the riboflavin-bound structure (Figure 4.2 and Figure 1.10). The isoalloxazine ring of the crystallographic and the docked binding mode still aligned well and the largest difference emerged from a different conformation of the ribityl side chain. A certain binding mode of this highly flexible part is unequally more difficult to predict and therefore the predicted binding mode of roseoflavin still was considered as desirable. The slightly worse docking score compared to riboflavin (Table 4.1) could be attributed to a different hydrogen bonding pattern accompanying the altered conformation of the ribityl moiety.

The test set contained compounds of very high structural similarity (Table 4.1) to FMN. Still, all of the test set compounds including the manually created non-binders reached a favourable docking score with the modified receptor setup. But the discrimination between binders and non-binders remarkably increased to 6-8 kJ/mol difference in total docking scores when adjusted partial charges were used. Except for the much smaller lumiflavin with less shape complementarity all four other training set compounds grouped as binders were placed first to fourth. Lumiflavin was shown to interact with the FMN riboswitch, since its inherent fluorescence is quenched upon addition of increasing RNA concentrations as observed also for FMN and riboflavin.[39] Thus, in the absence of structural data, the binding mode of lumiflavin can be assumed to be very similar to the ones for the other flavins in terms of heterocycle recognition. The predicted binding mode indeed was deviating little from the one predicted for riboflavin and FMN (Figure 4.2). Compound **D** of the training set completely lacked any polar functionality substituting the uracil-like edge and indeed exhibited the worst electrostatic score among all of the training set

compounds. This points out that the adjustment of the partial charges in A99 helped in predicting the correct binding mode for FMN and its analogues and to achieve a separation between binders and non-binders.

6.3.3 Docking of a large database of compounds

At this point it shall be discussed, why two different approaches were chosen for virtual screening. The experimental verification of the screening hits is further explicated in the corresponding sections discussing the binding assay results. The virtual screening on privileged substructures derived from the core scaffold of FMN interacting with the isoalloxazine-binding subsite of the riboswitch was biased by definition towards screening compounds with higher similarity to the initial ligand and therefore it was meaningful to also conduct an unbiased virtual screening to identify novel compounds, while taking advantage of a larger chemical space. In the following the possible reasons for discrepancies between predictions and the experimental validation are considered.
Before the different compounds had been chosen for further testing by binding assays, it was ensured that in the predicted binding mode the double hydrogen bond between the compound and A99 was formed and of course that an adequate affinity in terms of a negative docking score was predicted. Every compound purchased for testing met these criteria. Though, during a closer inspection of the predicted binding modes it was revealed that despite meeting the mandatory criteria, some of the predicted binding modes among the purchased compounds had unfavourable features, too (Figure 6.3). The reasons for apparently not having chosen the most suitable compounds could be attributed to inappropriate filter steps applied to the database of compounds for instance. The first filter step applied to reduce the number of compounds being subjected to virtual screening was rather straightforward. Compounds without drug-like properties and properties inappropriate for generating reliable data by a docking process were excluded from further processing. Alterations to this procedure are not expected to result in better data.
The pharmacophore search on the other hand was considered as a possible error source. Even though all of the compounds passing this filter had an energetically favourable conformer that was capable of interacting with A99 in the same manner as the uracil-like edge of FMN and all but one feature of the pharmacophore query had to be met by the compounds, the query should probably have been stricter. An aromatic group interacting with A48 and A85 (Figure 4.3) and an additional adjacent hydrophobic moiety were as well parts of the pharmacophore query. However, in the present pharmacophore search, compounds having the essential substructure for interacting with A99 and no substructure capable of forming face-to-face stacking interactions were also allowed to pass the filter. A lot of different conformers were subjected to the pharmacophore search and some of them probably passed the filter step in a conformation different from the favoured conformation with the best interaction scores during the screening process. Since this finding is per se inherent

to the method, keeping the aromatic feature involved in stacking interactions essential would have resulted in an enrichment of compounds with predicted binding modes more similar to the FMN binding mode among the top-scoring compounds and in a choice of more suitable compounds for the purchase. The size of the stacking free energies ΔG_{stack} may vary on the constitution of the small molecule ligand, but one might assume that they have a significant energetic contribution. Stacking interactions of isolated nucleoside pairs in RNA for example were computed to be the strongest when adenines are involved ($\Delta G_{stack} \approx$ -2 kcal/mol).[235] Other flavins known to bind to the FMN riboswitch like riboflavin and lumiflavin provide the same interaction network as required by such a stricter pharmacophore hypothesis and those compounds exhibit µM affinities.[39] Thus, it was expected that compounds forming the same interaction pattern, the double hydrogen bonding to A99 and π-π-stacking with the two adenines A48 and A85, could bind with comparable affinity as riboflavin and lumiflavin to the target RNA. Albeit the pharmacophore feature corresponding to the aromatic moiety involved in stacking interactions was not a mandatory criterion to be met by the potential ligands in the current pharmacophore model, it has to be emphasized that most of the selected virtual screening hits nonetheless were predicted to form the desired stacking interactions based on the present pharmacophore model in this study, except for the compounds **5** and **10** (Figure 6.3 b).

The position of the additional hydrophobic moiety of the novel ligands substituting the dimethylbenzene substructure of the heterocyclic system was also not localized well by the pharmacophore search. Since not all query points had to be matched during the pharmacophore search by the compounds, there were certainly some compounds that did not form π-π-stacking interactions, yet matched the hydrophobic feature due to a flexible linker between the uracil-like edge and this position. With a highly flexible linker the accuracy of the predicted conformation is likely to be reduced.

Four different types of problematic features among the predicted binding modes of the virtual screening hits were revealed upon closer inspection. Firstly, the predicted binding modes for compounds like **2** and **8** (Table 4.3), had the unfavourable property that hydrophobic rings were directed towards the solvent around the RNA (Figure 6.3 a). It would be more advantageous, if such ring systems would be buried deeply into the binding pocket[236] or be involved in additional stacking interactions to the RNA bases. Though, bulky hydrophobic substituents exposed to solvent do not have to be desolvated for binding and their energetic contribution ΔG would therefore be zero. Thus, such predicted conformations are less meaningful in a drug discovery process, where it is usual that hydrophobic substituents are not solvent-exposed. In addition to their low impact on increasing the binding affinity, such solvent-exposed hydrophobic substituents lower aqueous solubility, which is undesirable, too. Similar unfavourable predicted binding modes were observed for compounds **6** and **13** (not shown).

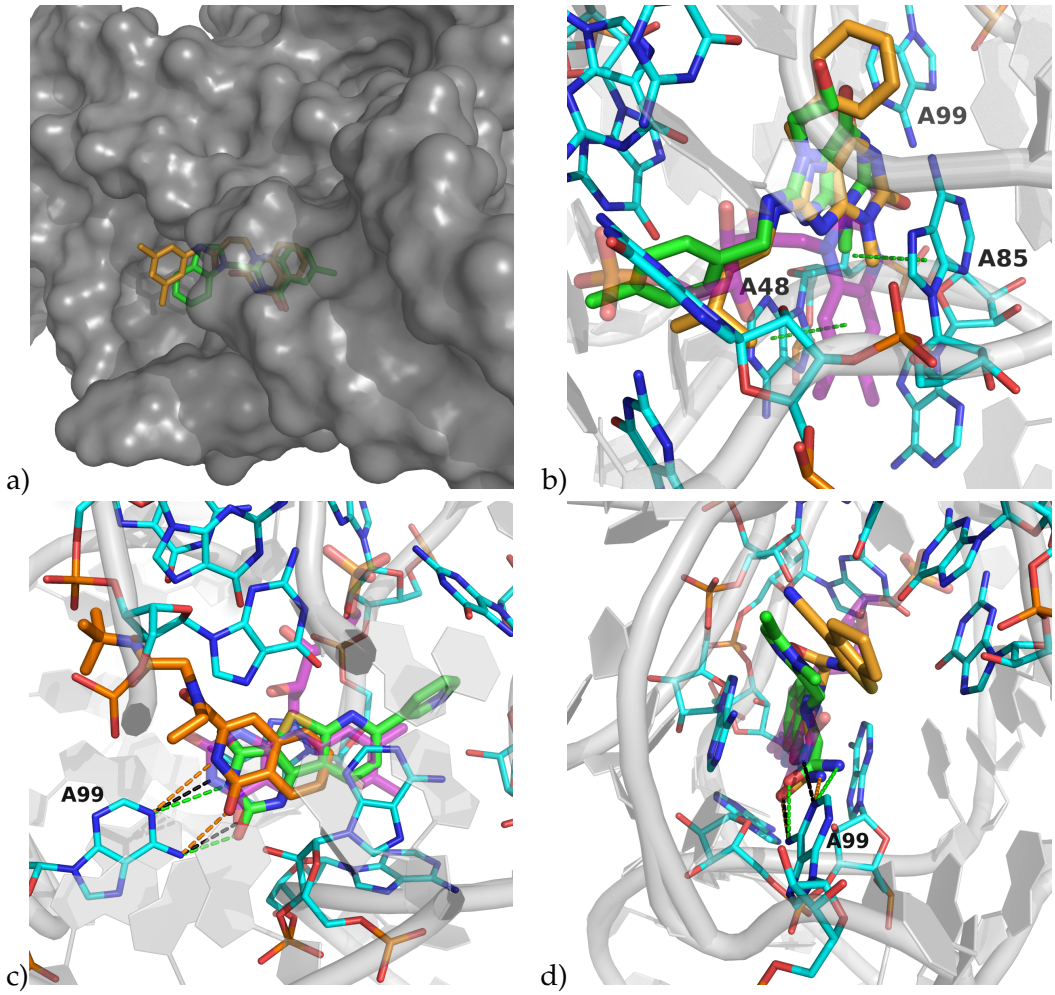


FIGURE 6.3: Exemplary binding modes with suboptimal geometries of the predicted poses that might explain the failure to confirm hit substances in subsequent assays. The FMN riboswitch RNA is shown as *grey-coloured cartoon*, the binding site residues are depicted with cyan-coloured carbon atoms as stick representation and the reference ligand FMN is represented with carbon atoms in purple colour as *transparent sticks*. a) Hydrophobic ends of compound **2** (green carbon atoms) and compound **8** (orange carbon atoms) are directed towards the outside of the target structure. b) Compound **5** (orange carbon atoms) and compound **10** (green carbon atoms) both are not involved in stacking interactions facing either A48 or A85. Stacking interactions between FMN and both adenines are displayed as *green dashed lines*. c) Compound **7** (orange carbon atoms) and compound **20** (green carbon atoms) undergo hydrogen bonding to A99 with geometries deviating from the optimal lengths and angles. d) Flexible amide functionalities of compounds **4** (orange carbon atoms) and **24** (green carbon atoms) are predicted to interact with A99 via hydrogen bonding while geometries are deviating from the optimal ones.

Secondly, for the compounds **5** and **10** in the predicted binding modes none of the two stacking interactions that the isoalloxazine ring system of FMN forms was proposed (Figure 6.3 b). The reason and the consequences of this had already been described above. Both compounds bear two substituents at the heterocyclic core predicted to interact with A99 and in each case the longer flexible side chains superimpose the ribityl chain of FMN in the predicted binding modes. However, these side chains are more apolar than the open ribose chain and therefore cannot provide additional hydrogen bonding to either G11 similar to one of the hydroxyl groups

of FMN and riboflavin or to the many other conserved guanines of the phosphate-binding region. The hydrophobic region of these compounds therefore does not occupy the region the hydrophobic edge of FMN fills within the binding site and the present conformations of the predicted binding modes do not enable a strong binding to the FMN riboswitch except for the double hydrogen bonding to A99.
Thirdly, some compounds were predicted to form the required hydrogen bond network, but only with imperfect geometries. Most of the data for typical angles and hydrogen bond lengths are of course derived from crystal structures of proteins. The median length for the NH·····O=C hydrogen bonds is 2.9 Å.[73] This value was also measured between the amino function of A99 and the carbonyl oxygen of the uracil-like edge of FMN (Figure 6.4). Further, in standard Watson-Crick base-pairing N ····· O separations in the range of 2.86-2.91 Å are commonly found.[237] NH-donors are weaker donor functions than OH-donors for instance[238] and every further deviation from a near-linear D-H-A (D = Donor atom, A = Acceptor atom) angle rapidly diminishes the hydrogen bond strength. In the case of the compounds **7** and **20** for example it was observed, that the predicted binding modes, albeit meeting the distance criterion for this hydrogen bond type, deviate from the perfect geometry in terms of deviating angles (Figure 6.3 c). This would negatively influence the interaction strength. The other hydrogen bond type =N·····H-N of the double hydrogen bonding interaction between FMN and A99 has a median distance between donor and acceptor atom of 3.05 Å.[238] The N ····· N separation in standard Watson-Crick base-pairing was found to be 2.95 Å. The measured distance of 3.0 Å for the FMN-RNA interaction lies well within this ideal range (Figure 6.4). However, in the case of compound **7** (3.6 Å) and compound **20** (3.7 Å) the hydrogen bond length was remarkably longer and both angles deviated from the optimum, indicating an attenuation of the hydrogen bonding properties. For compound **19** similar observations were made (not shown).
Fourthly, a subset of purchased compounds contained a flexible primary amide linker substituting for the uracil-like edge of FMN. In the compounds **4** and **24** a primary amide is a substituent at an sp^3-hybridized carbon atom, leading to rotational flexibility and twisting of the predicted hydrogen bonds (Figure 6.3 d). This in turn would again provoke hydrogen bonds with properties deviating from the optimal geometries.
Overall, at least 11 compounds were presented in the above paragraphs, where a second inspection of the predicted binding modes could help identifying suboptimal geometries that might have caused failure to confirm binding affinity in the experimental section. Almost half of all purchased compounds could show very weak or no affinity due to suboptimal geometries, which could be attributed to an inability of the scoring function used in differentiating between inactive and active poses (see also 6.3.4).

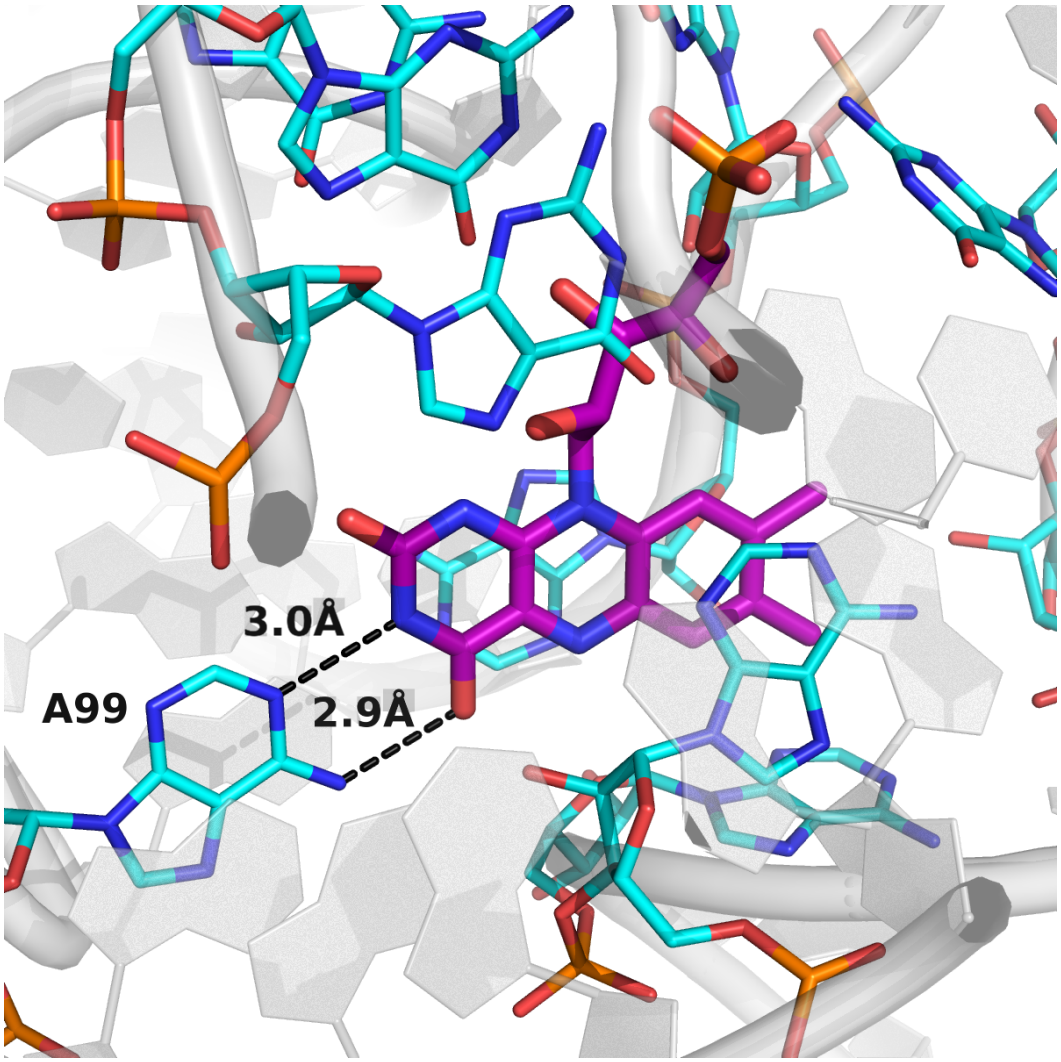
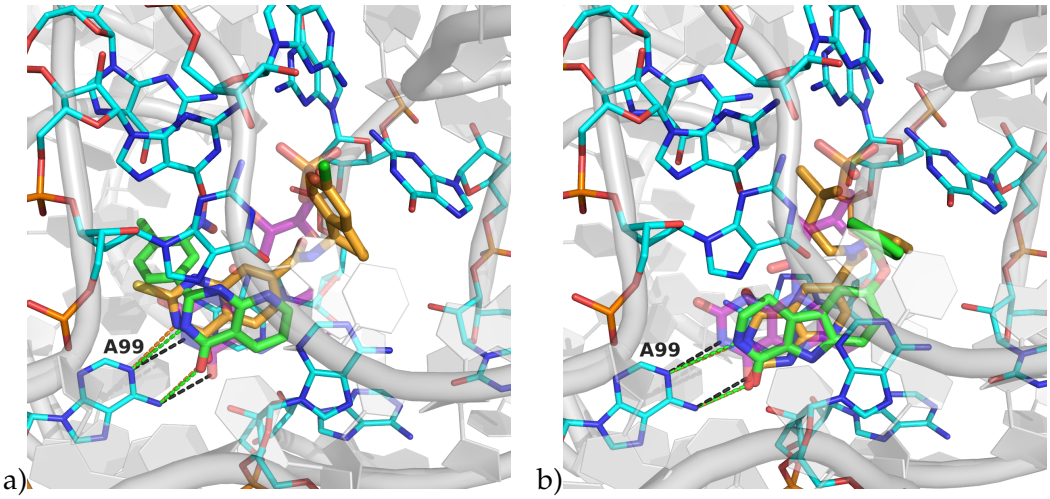


FIGURE 6.4: Hydrogen bond lengths and geometry of the interactions between the uracil-like edge of FMN and A99. The FMN riboswitch RNA is shown as *transparent grey cartoon* with the binding site residues being displayed as stick representation with cyan-coloured carbon atoms. FMN is depicted with purple carbon atoms and the hydrogen bonds between the uracil-like edge of FMN and A99 are displayed as *black dashed lines*. The lengths are given in Å as indicated.

6.3.4 Docking of a dataset of privileged substructure

The virtual screening hit confirmed in the ITC assay and its close analogue exhibiting activity in the SPR assay arose from the substructure search. This serves as the best proof of the higher likelihood of discovering a novel FMN riboswitch ligand with the second molecular docking approach. By means of the SMARTS patterns chosen for the substructure search, cyclic molecular substructures interacting with A99 of the riboswitch binding site were preferred. Constraining the donor and acceptor functionalities of the screening compounds to a molecular pattern resembling the uracil-like edge embedded into a six-membered ring reduced the number of docking solutions with unfavourable hydrogen bonding patterns exhibiting suboptimal geometries.

In both virtual screening approaches the predicted ligand efficiency based on the computed docking score was one of the selection criteria for the choice of the compounds that were purchased for experimental testing. A slightly stricter cut-off of 1.5 kJ$\times$mol$^{-1}\times$HA^{-1} was applied in the privileged substructure search. Despite its relatively high molecular weight and a number of heavy atoms of 27 riboflavin had a very good predicted ligand efficiency with 1.9 kJ$\times$mol$^{-1}\times$HA^{-1} as well when calculating the LE with the predicted docking score of -51.47 kJ/mol using the final receptor setup. Lumiflavin even had a predicted LE of 2.15 kJ$\times$mol$^{-1}\times$HA^{-1}. It would have been plausible to apply a LE cut-off in the range of riboflavin, since this known ligand exhibits an experimentally determined binding affinity in the µM-range and the docking process should serve as a rough estimate of the actual binding affinity. However, this appeared as a too strict limitation as only seven of the purchased compounds would have passed this filter. The cut-off of the predicted ligand efficiency was chosen to lie in between typical cut-offs for ligand efficiencies of drug-like and fragment-like compound subsets, respectively.[239] Thus, compounds that were predicted to be more efficient than average drug-like molecules thereby maintaining the opportunity for potential optimization procedures were chosen.
Except for compound **28b**, all purchased hit compounds of the substructure search (Table 4.4) were chosen for their favourable docking scores as well as a sufficiently high predicted ligand efficiency to possibly be a valuable starting point during a lead optimization process, and for their desired Watson-Crick-like base pairing to A99 (4.2.5). Compound **28b** was selected in addition due to two different reasons. Firstly, the parent compound **28** was predicted to exhibit a rather high affinity towards the riboswitch RNA (rank 4, Table 4.4). A relatively high binding affinity was predicted, since compound **28** contains a three-membered heterocyclic system similar to the one of flavin analogues which also score high (Table 4.1). A48 and A85 were predicted to be addressed via face-to-face stacking interactions in a similar fashion. Furthermore, an additional hydrogen bond was postulated for compound **28** (Figure 4.6). The NH-donor function of the central pyrrole was in close proximity to the ether oxygen of the ribose in A49 (Figure 4.6). Beyond the predicted high ligand efficiency and the proposed extra hydrogen bond, it was speculated whether the phenolether structure is essential for the binding properties of compound **28**. In the predicted binding mode of compound **28**, the phenyl ring was not arranged to accommodate for π-π-stacking interactions for instance. Therefore, it was supposed that a methyl ether in compound **28b** instead of a phenol ether-function should not negatively affect binding to the FMN riboswitch.



a) b)

FIGURE 6.5: Binding modes of privileged substructure search hits with suboptimal geometries of the hydrogen bonds. The FMN riboswitch RNA is shown as *grey-coloured cartoon*, the binding site residues are depicted with cyan-coloured carbon atoms as stick representation and the reference ligand FMN is represented with purple-coloured carbon atoms as *transparent sticks*. a) Binding modes of compounds **26** (orange carbon atoms) and **35** (green carbon atoms) compared to the binding mode of FMN. b) The predicted poses of the compounds **32** (orange carbon atoms) and **28** (green carbon atoms) overlaid with the binding mode of the reference ligand FMN.

Among the 20 compounds (excluding compound **28b** as close analogue of compound **28**) that were purchased upon the results of the substructure search considerably less of the problems discussed in the above section (6.3.3) occurred. Three of the leading reasons for suboptimal geometries found among the first screening hit structures arising from the screening the whole in-house library of compounds were not recovered. The freely rotatable primary amide serving as donor and acceptor function was per se not observable due to the restrictions based on the SMARTS pattern as described above. Additionally, in all docking solutions the top-scoring pose was predicted to undergo at least one π-π-stacking interaction to the adenines A48 and A85 and furthermore, there were no flexible alkyl side chains predicted to be solvent-exposed, protruding from the binding pocket. The percentage of predicted binding modes with suboptimal hydrogen bonding geometries decreased from 50% in the larger virtual screening to 20%. Based on the results of the substructure search, only four compounds were selected whose predicted binding modes contained hydrogen bonds to A99 with non-ideal geometry (Figure 6.5). Compound **26** was one of the compounds for which still suboptimal hydrogen bonding properties were observed. However, the deviations were only marginal compared to the predicted hydrogen bonds for FMN interacting with A99. The hydrogen bond between the carbonyl oxygen acceptor function of compound **26** and adenine 99 was prolonged by 0.3 Å to 3.2 Å compared to FMN and the =N:⋯H-N hydrogen bond was 3.1 Å long, which is only 0.05 Å longer than the median for such hydrogen bonds.[238] The angles between donor atom, hydrogen and acceptor atom also were not strongly deviating from the ones observed for the natural ligand FMN binding to the FMN

riboswitch RNA. These differences should not preclude the interaction of the heterocycle in compound **26** with A99. In the case of compound **35** the hydrogen bond geometry of the NH-donor function of the pyrido[2,3-d]pyrimidin-4(*1H*)-one was altered compared to the interactions of the uracil-like edge. With 3.4 Å, the hydrogen bond was by 0.4 Å longer, which is already outside the normal distribution of such hydrogen bond lengths.[73] In the case of the compounds **32** and **28** even higher deviations from the optimal hydrogen bond lengths for the =N:·····H-N hydrogen bond were observed. Only depending on the energy cut-off applied they became invisible when inspecting the binding modes visually with MOE (Chemical Computing Group, Montreal, QC, Canada) indicating that the hydrogen bond would be significantly weaker in the predicted binding conformation compared to the hydrogen bonds of the FMN-RNA interaction.

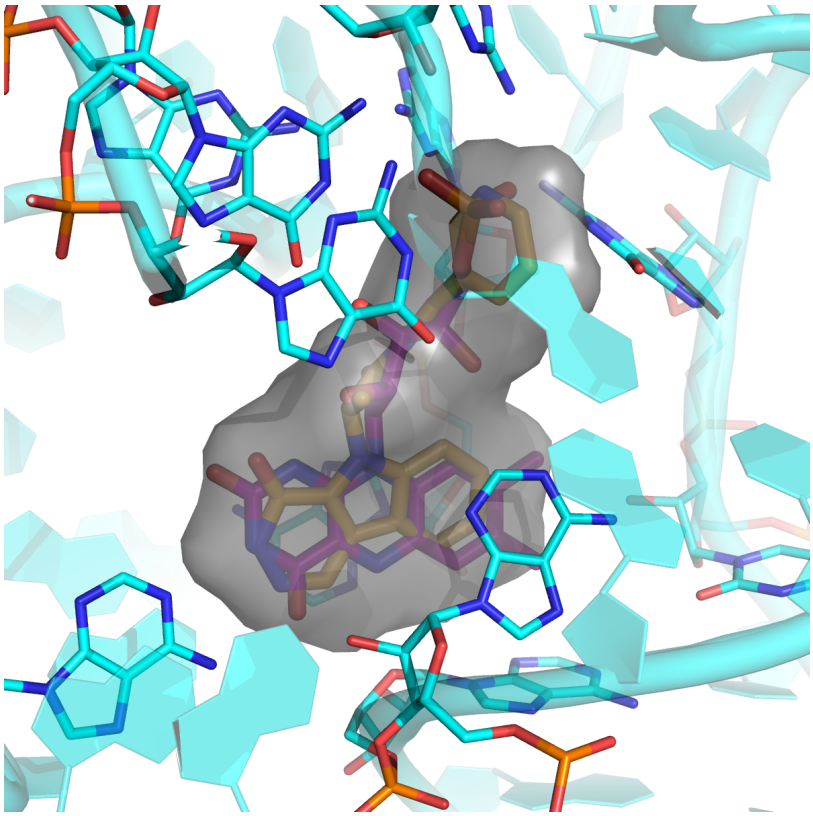


FIGURE 6.6: The predicted binding mode of compound **25** with orange carbon atoms is displayed as an overlay with the crystallographically determined binding mode of FMN (purple carbon atoms). Binding site nucleotides are represented as *cyan sticks* within the riboswitch RNA (*cyan-coloured cartoon*). The grey surface around the atom positions of compound **25** indicates an almost entirely wrapped natural ligand FMN, so that compound **25** fills the whole binding site.

An additional advantage of the substructure-based virtual screening was the relatively small number of docking solutions. The only filters necessary were a cut-off for the overall docking score of -30 kJ/mol and a ligand efficiency of $\geq$ 1.5 kJ$\times$mol$^{-1}\times$HA^{-1}. For that reason it was possible to visually inspect all of the remaining docking poses and to differentiate easier between promising candidates

and compounds whose predicted binding modes evoked doubts. In the first screening, some promising candidates probably might have got lost, since they were not ranked among the best-scoring solutions, but still had a reasonably good docking score.

Most of the 20 compounds that were chosen for experimental testing did not span the whole area occupied by FMN. This is not very surprising due to the fact that the sphere set used for matching (Figure 3.2) also was focusing on the isoalloxazine binding site. Although the compounds are allowed to protrude the sphere set during matching, only the largest screening compounds can meet the requirements of the main pharmacophore and interact in the phosphate-binding region at the same time as the binding site is rather large. In principle, additional interactions with the nucleotides in the phosphate-binding region of the FMN riboswitch would be desirable as well. At least the selection criteria for the first filter step applied to the database did not exclude compounds interacting outside the sphere set area with the RNA, since the molecular weight and the number of heavy atoms were not restricted to fragment-sized small molecules. Only two compounds, compounds **25** and **26**, filled almost the entire binding site (Figure 6.6). This is also reflected by the docking scores for both compounds (Table 4.4). Both compounds took rank number one and two, which delineates their predicted good shape complementarity with the binding pocket. Especially the favourable vdW energetic contribution to the total score supported the determination of a good shape complementarity by visual inspection. Compound **25** in addition was predicted to form an additional hydrogen bond with the exocyclic amino group of G32 similar to a phosphate-mediated hydrogen bond observable in the interaction network between FMN and riboswitch RNA (Figure 1.6). Most recently, a novel riboswitch ligand with an aromatic ring displacing the phosphate moiety was reported[60], which made this binding mode observed here quite promising. However, it turned out in the experimental testing that both of the virtual screening hits with the best predicted docking score were inactive, too.

For the binding affinity of the isoalloxazine-containing flavins it is crucial, that the dioxo-tautomer of the uracil-like ring segment is the main tautomer. Since the substructure search was based on SMARTS pattern (Figure 4.5) derived from the uracil-like edge of FMN, it was likely that for some virtual screening hits a distribution of different tautomers have to be considered as well when discussing the results. Generally speaking, different protomeric, tautomeric and stereoisomeric states are well-known to affect the outcome of docking studies and therefore accurate predictions are necessary.[240] The dioxo-form of FMN is predominant in aqueous solution (compare with free uracil[241]) at physiological pH-values and considering the given pK_a-value of ≈ 10[242].

Compound **30** (Table 4.4) contains an especially interesting heterocycle with very close relation to the isoalloxazine of FMN. The substitution of one oxo-group with

a sulfur atom might change the tautomerization state properties and therefore compromise the accuracy of the the binding affinity prediction for compound **30**. Initial calculations with the Marvin program suite (v5.11.3, Oct2012, ChemAxon, Budapest, Hungary, `https//www.chemaxon.com/`) support the hypothesis of a more complicated situation by proposing four different, equivalent tautomers at physiological pH compared to a predominant form for the oxo-analogue. This could be a possible reason for failure to experimentally confirm this compound as FMN riboswitch ligand. Besides for compound **30** also for compound **25** (Table 4.4) tautomerism could have influenced the accuracy of the prediction. To enable the proposed binding mode, it is crucial that the oxo-form and not the hydroxy-tautomer of the 2,3-dihydropyridazin-3-one of compound **25** is available and the pyridine ring is deprotonated. According to calculations the oxo-tautomer exists almost to 100% at a pH of 7.4 - a pH that was applied in all assays and experiments. With a calculated pK_a of 4.8 the pyridine ring should be deprotonated to almost 100%. Both calculations were done by using the Marvin program suite (v5.11.3, Oct2012, ChemAxon, Budapest, Hungary, `https//www.chemaxon.com/`). The isomer of compound **25** yielding the top-scoring pose for this particular molecule could therefore considered to be prevalent in aqueous solution. Despite the promising predicted binding mode and the dominant tautomer in aqueous solution being the one suggested to interact with the binding site, experimental validation of compound **25** as a riboswitch ligand was unsuccessful.

A unique binding mode was predicted for compound **35** (Figure 4.7 c). For this compound, a nitro group was predicted to serve as hydrogen bond acceptor in a back-pocket of the FMN riboswitch which is usually not occupied by flavin analogues. It has been known for a long time, that nitro groups can act as proton acceptors in hydrogen bonding.[243] However, nitro groups are rather weak hydrogen bond acceptors and often face toxicity problems.[244,245] This has to be considered when trying to improve the properties of such compounds in an optimization process. Nevertheless, it was decided that it was worth to include this compound in the list of purchased compounds because of the unique binding mode with implications for additional favourable interactions.

Even though it was illustrated in the last paragraphs that in the substructure search reasonable poses were generated with predicted binding modes capable of properly interacting with the binding site, it has to be emphasized that only one compound was active in the assays. At least one can partially argue that a higher probability of finding a hit from the second screening set was expected, but still both virtual screening methods suffered from a common problem. Different docking engines often succeed in predicting a correct binding mode, but predicting binding energies remains a much larger challenge.[229,246–248] An optimized scoring function has to differentiate between correct and incorrect poses to appropriately rank the compounds. Due to otherwise too large computional demands especially in large screenings, a lot

of simplifications and assumptions are made by the docking programs in the evaluation of the modelled complexes.[105] Collectively, this indicates shortcomings of the used scoring functions that have to be addressed in the future as reasons for failure to discover more virtual screening hits with an affirmed binding affinity in the experimental testing. Possible solutions will be briefly discussed in the section regarding the future perspectives on computational drug design towards riboswitch RNA in the following chapter (see 7.2).
Four out of the five analogues of compound **44**, the compounds **46**-**49**, were included in the SAR studies after hit confirmation by the ITC displacement assays with the aim to serve as "positive" control. Affinity towards the FMN riboswitch RNA was expected for all of these compounds, but the ITC experiments with compounds **46** and **47** were disappointing. This can be rationalized by inspection of the predicted binding mode for compound **44** (Figure 5.8).

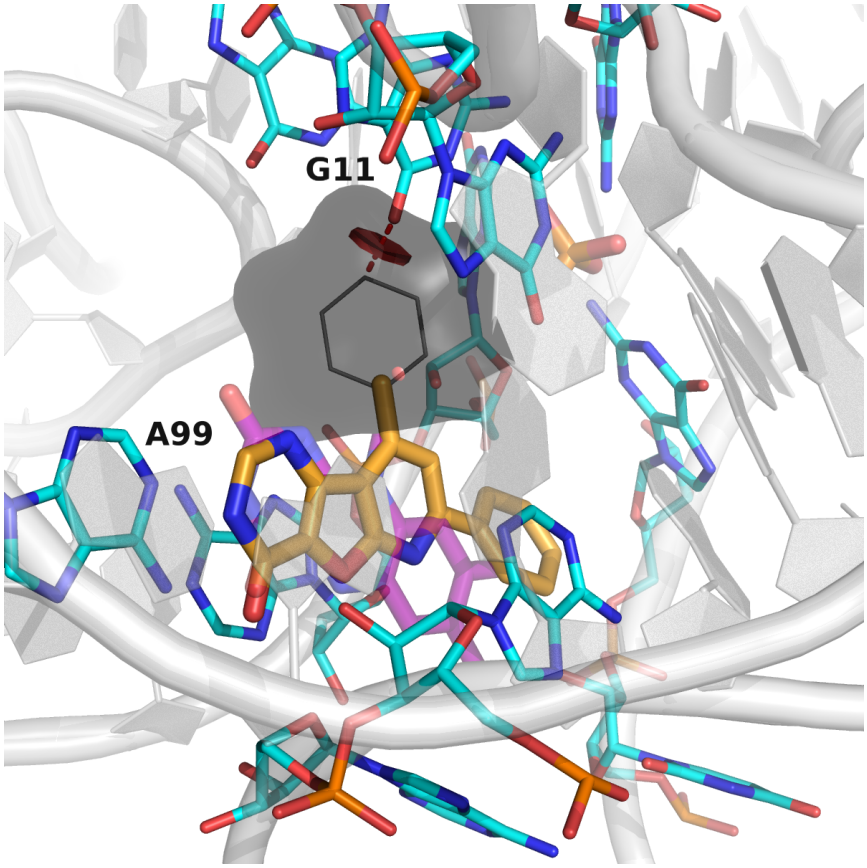


FIGURE 6.7: Hypothetic binding mode for compound **46** (orange carbon atoms) is represented here. The RNA target structure is shown as *grey cartoon* with binding site residues displayed as *sticks* and cyan carbon atoms, too. In the background for better orientation the cocrystallized ligand FMN is shown as *sticks* with purple carbon atoms. Assuming a correct binding mode prediction and optimized orientation of the central heterocycle in compound **44** and its analogues the space occupied by an additional aromatic substitution (compounds **46** and **47**) is indicated by the dark-grey, transparent surface. A strong steric clash between the additional phenyl ring and the carbonyl oxygen of G11 is highlighted by a *red disk* and a *red dashed line*.

Compound **44** overall is quite planar and therefore well-suited for stacking interactions to either one or both conserved adenines A48 and A85 and to form the required Watson-Crick-like base pairing. It is conceivable that no major conformational changes of the three-membered heterocyclic core are allowed to retain the binding capability. Overlaying the core scaffolds of compound **44** and compound **46** in the binding pocket revealed that the phenyl substituent probably is directed into a region of the binding pocket less accessible for a large substituent (Figure 6.7). The minimal distance between one of the carbon atoms of the phenyl substituent and the carbonyl oxygen atom of G11 was only 1.7 Å. With the additional hydrogen atom attached the vdW radii overlap resulting in a steric clash. For compound **47** a similar explanation seems likely.

Chapter 7

Conclusion and future perspectives

7.1 Conclusion

The central question of this thesis is, whether a virtual screening approach is suitable to predict ligands for the FMN riboswitch containing novel chemical scaffolds. The different chapters dealt with different steps during the first cycle of a structure-based drug design process (see 1.4.3).[104] Molecular docking was used to score and rank compounds from a large database according to their electrostatic and steric interactions with the binding site nucleotides. The most promising compounds were tested in biochemical assays that had either been developed by other research groups (fluorescence quenching assays) or that had been established as part of this thesis. One compound was identified as hit compound with micromolar affinity. Herewith, the study ended at the transition from the first to the second cycle of a virtual screening process.

7.1.1 Virtual ligand screening

Most of the prospective virtual screening studies identify only weakly active compounds (usually in the range of 1-500 µM). Except for flavin analogues and ribocil, which was identified by a phenotypic screening, little information is known about possible ligands for the FMN riboswitch. Therefore, it was justified to search for compounds with very weak affinities.[249] Even though novel riboswitch ligands had already been identified by a virtual screening method earlier[130] and the docking protocol used here mainly conformed to the one used for the purine riboswitch, the initial predictions were not satisfying. It is mainly the task of a high-quality scoring function to identify the correct binding mode of a ligand out of different docking solutions. But especially when docking large compound libraries, these scoring functions have to be very fast and therefore neglect many contributions to binding free energy. Scoring functions should help recognize solutions displaying good steric and electrostatic complementarity between receptor and ligand.[250]. Despite the assumed applicability to all types of targets, comparative studies in the past have stated that a universally accurate, generic scoring function is out of reach.[251] To overcome this problem, re-parameterization of scoring functions is a common strategy, and tweaking of the parameters leads often to a more target-biased, balanced

scoring function.[251] In several studies this improved the prediction accuracy of the docking process.[252–255] Here, tweaking of the AMBER atom type parameters was necessary to better satisfy the binding site requirements and this finally resulted in the correct binding mode prediction for FMN with a desirably low rmsd-value.
For the majority of screening compounds in a screening library the likelihood of exhibiting the desired binding affinity to a certain target is very low. Target-focused libraries around a certain scaffold with one or more positions used to attach various other chemical moieties are often introduced to increase the success rate of the virtual screening approach.[256] However, there is a need for a correct balance between diversity of library molecules and focus on the target-related requirements to still discover novel chemical scaffolds.[257] The results of the present study support the hypothesis that focused libraries might help increasing the identification of bioactive compounds. Only the substructure-based search revealed novel ligands. Basis for the focused compound library in the screening process for novel FMN riboswitch ligands was only the ring involved in hydrogen bonding to A99 and the substitutions were amenable to diverse moieties. Albeit closer resembling the chemical structure of FMN, the substructure search hits still represent novel chemical scaffolds. The confirmed virtual screening hit, compound **44**, for example, contained a three-membered heterocycle different from the isoalloxazine ring system of FMN.

7.1.2 Experimental testing

A suitable experimental follow-up is crucial for hit validation. The virtual screening for novel FMN riboswitch ligands revealed two common issues accompanying the validation of hit substances: poor aqueous solubility and autofluorescent properties.[258] Light scattering and autofluorescence compromised the evaluation of the results for several compounds. Also with more universally applicable fluorescence-based assays, that are independent of the inherent fluorescence of FMN, autofluorescence would still be an issue.[259] The solubility prediction is usually based on the logP-value and although the most hydrophobic compounds had been deleted before the molecular docking process by applying a threshold for this partition coefficient, this prediction is seemingly far from perfect. Besides, predictors directly calculating the aqueous solubility, expressed as logS-value, which are for instance based on the different atom contributions[260], are implemented in molecular modelling software (e.g. ChemAxon, Budapest, Hungary, `https//www.chemaxon.com/`). However, similar to logP-predictions, these predictions can also suffer from the fact that certain atom types have the the same contribution to solubility, regardless of the overall structure of the molecule
From fragment screening it is known that testing an almost identical subset of screening compounds by two distinct assay methods not necessarily results in the same hit structures.[261] Although the docking procedure applied to the FMN riboswitch was not aiming to identify fragment structures, it was handling hit compounds with

affinities > 100 µM. Particularly for compounds exhibiting borderline activity a careful selection of methods to confirm binding is of utmost importance. There are a variety of reasons for potential assay artefacts and secondary assays are mandatory to confirm binding, especially by direct biophysical binding assays and orthogonal assays.[155] The development of a direct binding assay and at the same time orthogonal methods based on different detection principles was accomplished here and this definitely facilitated the removal of false positive hits. Those were not validated by one of the orthogonal methods.
The results of the experimental hit validation emphasize that a universal assay method for ligands interacting with the FMN riboswitch has not yet been found. Fluorescence intensity measurements are usually highly sensitive, but often are lacking reproducibility and are prone to interference. Additionally, binding affinity was only detected by a competitive interaction and therefore the detection of very low affinities is impeded. The density of the immobilized target molecule layer and the remaining activity limit the generally good ability to detect weak binders by the SPR approach. Even the economic feasibility of the present SPR approach is limited, since the immobilized RNA can only be used within two days after immobilization on an expensive sensor chip. The ITC displacement assay in turn generated stable responses, but is also lacking economic feasibility. Essentially, purification of sufficient amounts for a large number of repetitions would be too time-consuming.

7.2 Perspectives in computational drug design towards the discovery of novel FMN riboswitch compounds

All compounds from the unbiased virtual screening consisting of a diverse set of library compounds were inactive in the assay carried out for hit conformation. This clearly indicates shortcomings of the used scoring function and emphasizes that this issue has to be addressed, too, in the future.
The scoring function determines the prediction accuracy of the docking program. As a centrepiece of the docking algorithm, the scoring function is needed to optimize the ligand orientation and to estimate binding affinity of the docked molecule. Scoring functions are usually good in generating the correct binding mode, whilst the correct estimation of binding affinities remains more difficult. The application of a scoring function basically is an assessment of receptor-ligand complementarity.[262] One option could certainly be testing the performance of different scoring functions, since the force field based scoring functions utilized by DOCK is not the only scoring function available. However, all of the scoring functions rely on a few general parameters that are transferable to various systems and each of them have their strengths and drawbacks.[99]
An improvement of the binding affinity prediction using the force field-based scoring function could be the introduction of environment-specific force fields. There are environmental polarization effects influencing the force field charges, which in turn

affects the electrostatic energies. But also the Lennard-Jones parameters contributing to the vdW energetic term are sensitive to the local environment of the atom. Instead of parameterization based on look-up tables, QM calculations would be necessary to provide the case-sensitive environment-specific force fields.[263] However, this would require sufficient computational resources. It is more common to incorporate quantum mechanical calculations into the force field parameterization of small molecules. Parameterization of the macromolecule would be necessary by using the same methodology. In the last few years, large-scale QM calculations became more and more routine and can already be performed with single desktop computers on large proteins.[264,265] A lot of work has to be done and it remains elusive, whether such environment-specific force fields are able to replace the general, transferable force-field. Existing transferable force-fields are improving as well by including specialized parameters for the molecules of interest.

Even though already two virtual screening runs have been conducted and the aptamer domain of the FMN riboswitch does not undergo large structural arrangements upon binding of different ligands, new docking studies are still valuable. Most of the potential antibacterial drugs exhibiting their effects by targeting riboswitches are structurally related to the cognate ligands as discussed in a recent review from Mayer and coworkers.[16] High structural similarity on the one hand increases the probability of compounds that truly interact in an agonistic fashion with the aptamer domain of a riboswitch RNA[17], but on the other hand riboswitch ligands are occurring ubiquitously also in the human body. A lot of flavoproteins are FMN-dependent and it is known that roseoflavin for example not only interacts in a riboswitch-mediated mechanism and targets many flavoproteins in bacteria as well, which might pose the risk for significant off-target effects, because these flavoproteins are present in almost all organisms.[266] To overcome the problem of enriching compounds with a higher similarity to FMN among the top-rank order, an excellent starting point for further docking studies is the recently identified novel ligand ribocil. This compound has a high binding affinity in the low nM-range towards the *E. coli* FMN riboswitch aptamer domain.[61] The ribocil-bound structure of the FMN riboswitch has already been determined by X-ray crystallography (pdb-codes: 5c45[61], 5kx9[65]) and this revealed a unique binding mode, very distinct from the binding mode of FMN to the target RNA (Figure 1.13). The development of a modified receptor setup for docking and the generation of a new pharmacophore model pave the way for a new database screening based on this binding mode. There is a high chance to identify new chemotypes due to the large differences in the structure and binding modes of ribocil and FMN. Furthermore, the performance of cross-docking studies might help interpreting the value of the virtual screening results. If the docking scores of FMN are very similar to each other when docked into the FMN-bound and the ribocil-bound crystal structure and the same is valid for ribocil when docked into the FMN-bound and the ribocil-bound structure, this reduces uncertainties and increases the explanatory power of the virtual screening.[267] Albeit

previous findings do not suggest high flexibility, good cross-docking results would support this theory and retrospectively determine the reliability of the data from the initial virtual screenings.
Docking is sometimes applied to lead optimization as well. However, if rank-ordering a list of a diverse compound set is already a goal beyond the currently available docking methods, then ranking closely related derivatives is it even more. The free energy difference between a ligand of high potency ($\approx$ 50 nM) and weak potency ($\approx$ 100 µM) is only about 4.5 kcal/mol.[154] For the follow-up and to improve the value of the virtual screening campaign accurate binding affinity predictions sensitive to very small changes like single substitutions are of utmost importance. For small group and linker refinement free energy perturbations could guide the optimization of the compounds without simply inspecting the predicted binding modes and synthesis relying on the predicted affinities by the docking program.[268] Apart from the classical scoring functions machine-learning approaches could also help in guiding the optimization process via docking in the follow-up of a large screen. At the latest after the binding mode of the identified novel binders had been confirmed, a training set of complexes with known binding modes would be available to teach the scoring functions, which features of the interactions are the most important ones.[269] This might also help in the choice of the right substituents to optimize properties of the virtual screening hits in general.
Computational methods other than molecular docking are also conceivable to guide the selection of optimized ligands for either purchase or synthesis. So-called medicinal chemistry transformations such as functional group addition, ring modification and terminal group exchange can be performed with programs suited for structure-based drug design applications like MOE (Chemical Computing Group, Toronto, Canada). When a database of linkers and substituents is provided, the computer program will assist in the choice of the most promising derivatives to follow up with. Collectively, starting points for optimization are identified and many options for future experiments to gain deeper insights into the discovery of novel FMN riboswitch ligands are imminent.

7.3 Perspectives for the experimental testing of novel FMN riboswitch ligands

The search for new or improved biochemical assay methods to elucidate the binding affinities of small molecules interacting with the FMN riboswitch still is necessary. Experimental results have proven that the FMN riboswitch is a drug target, which is amenable to structure-based drug design and provides desirable binding site properties. Without optimization of the initial hit compounds and improvement of the solubility it will be difficult to quantify the binding properties. To date a dissociation constant was reliably achieved for one compound, but these data have not been

achieved yet for its derivatives. In addition, for this compound hit validation by an orthogonal assay still has to be performed.
A high throughput assay for fast identification of compounds with binding affinity and universal applicability to all compound types would be desirable. The methods employed throughout this study all had their strengths and drawbacks and compound-dependent or independent limitations. A fluorescence polarization assay is well suited for high throughput applications and often involved in small molecule screening[259] and the general feasibility in a riboswitch environment has already been proven. For a high-throughput screen targeting the SAM-I riboswitch RNA fluorescent analogues of the natural ligand S-adenosylmethionine (SAM) were designed. The fluorophore was introduced through a functionalized linker without diminishing the binding capability and a displacement fluorescence polarization assay was developed thereupon.[270] A suitable fluorescent dye with a red-shift could limit the probability of compound interferences, if the excitation and emission wavelengths are sufficiently high. Although the intrinsic fluorescent properties of FMN were obvious to be exploited in a fluorescence-based assay throughout this study, this alternative should be considered.[259]
To further reduce the probability of finding nonspecific binding ligands counter screens could be performed in the future. No binding should be detected for example with other riboswitch RNAs in the screening methods.
Another important issue that has to be addressed when continuing with FMN riboswitch-ligand discovery is to solve the problems associated with the global folding of the riboswitch RNA. The presence of inactive RNA species to an extent of almost 50% complicated the evaluation and analysis of all the experimental methods used to confirm the virtual screening results. In particular, with the completely solution-based methods, the ITC displacement assay and the fluorescence quenching assay, the situation became evident. Due to the immobilization of the target structure and the specific labelling that was necessary, the situation is more complex for the SPR-based method and a full activity was not necessarily expected. Two different approaches are available to address this problem:

1. Purification by native, non-denaturing techniques to keep the co-transcriptional fold and avoid harsh conditions.[177]

2. Improvement of the renaturation after denaturing purification with varying conditions.[271]

Sample renaturation is often achieved by heat-denaturing RNA samples and renaturing them by either snap-cooling, as it was done during this work, or over a shallow temperature gradient depending on the nature of the RNA. Common strategies to improve the outcome of renaturation procedures following denaturing purification with the addition of 8 M urea include the dilution of the RNA prior to heating, testing whether Mg^{2+}-ions should be added during the heating step, comparing different buffer components, comparing different monovalent cations, using additives

like arginine, cooling by placing the sample on ice or by diluting with prechilled buffers.[272] Another possibility is to refold denatured RNA by gradual dialysis from a buffer with urea into buffer without urea and thereby slowly reducing the amount of denaturing agent.[273] In general, it is very difficult to obtain active RNA in a single, active conformation, and avoiding the denaturation step can assist in obtaining the desired conformation.[274] Two procedures would be possible to purify the aptamer domain of the FMN riboswitch by native purification. Firstly, at the 3' end of the desired sequence a *glmS* ribozyme followed by a protein that can be immobilized on a nickel column via interaction with the histidine-tagged protein could be inserted in the vector used for transcription. After elution from the column and ribozyme cleavage the purified RNA would be ready for further usage.[275] The second approach can be implemented faster and tested first for its ability to achieve a large population of correctly folded RNA species. Instead of purifying the *in vitro* transcribed samples by PAGE, fast performance liquid chromatography (FPLC) could be used. Both a weak anion-exchange chromatography with a DEAE-sepharose matrix[276] and size-exclusion chromatography[277] with Superdex columns are conceivable.

Another short term goal is the structure determination of the experimentally validated virtual screening hits by X-ray crystallography or NMR studies. For claiming a prospective docking study being successful the question has to be answered, whether the ligands were discovered for the right reasons. A successful docking campaign should enrich actives among the top-ranked molecules and predict a correct binding pose for the co-crystallized ligands. Experimental testing is necessary to validate the screening hits as true binders, whereupon orthogonal assay methods should be used to lower the chance of identifying artefacts. Finally, for a complete study the predicted binding mode has to be compared with the experimental structure determination to decipher, whether ligands were just identified by coincidence or for the right reason.[131]

Especially the lead compound **44** has to be retested to confirm and validate the compound as a ligand for the FMN riboswitch RNA. Either the purchasability of the compound is reobtained in the meantime or complete resynthesis will be necessary. The synthesis route has already been described a few years ago.[278] Due to the non-availability of compound **44** the binding affinity has not yet been experimentally confirmed by an orthogonal method. Similar to its analogues, compound **44** is amenable to the SPR assay and it will be exciting to see whether the predicted binding mode conforms with the crystallographically determined binding mode.

Initialization of the synthesis of compound **44** and analogues is not only interesting for the fact that this compound was not available any longer. Derivatives with optimized physicochemical properties and/or improved potency could also emanate from the integration of chemical synthesis into this project. Especially the bad aqueous solubility combined with the low potency of the initial hits ran like a thread through the whole project and prevented a self-contained analysis of the compound

properties. With a higher solubility changes upon compound addition in the displacement assays and the SPR measurements would be less difficult to detect.
After optimization and verification of potential novel binders it would be desirable to not only prove their binding affinity against the isolated, purified RNA, but rather also to determine, whether these compounds are functional in interfering with the riboswitch structure. The *F. nucleatum* FMN riboswitch is known to regulate downstream genes on the transcriptional level and therefore directly monitoring the *in vitro* transcription by a suitable assay would be suitable to quantify the efficacy of the compound bound to the aptamer domain. A fluorescence-based dual molecular beacon assay for example was successfully implemented to monitor ligand binding to the purine riboswitch *pbuE* of *Bacillus subtilis* regulating the purine base efflux pump.[279] Reporter gene assays are even less sensitive to artefacts in comparison to the molecular beacon assay described above. By construction of plasmid-based reporter genes with several related FMN riboswitches from other bacterial organisms the specificity of the compound for the target can be demonstrated directly and at the same time the potential to become a promising antibacterial compound can be tested. The GFP gene for example could be placed under the control of a *F. nucleatum* impx promoter and the FMN riboswitch 5' sequence or related riboswitch sequences in a similar way as already been described in the target validation for ribocil.[61] Compounds specifically interacting with the aptamer domain would cause a strong reduction of the GFP expression, which is easily detectable. Efficacy of the compounds in a cell-based context will also be of great importance, although it is definitely a long term goal and would be meaningful, if all *in vitro* measurements confirmed the binding of lead structures to the FMN riboswitch.
Alongside NMR-based detection methods, SPR measurements are most widely used for fragment screening approaches at the beginning of the identification of novel ligands for protein targets.[280,281] The SPR-approach established in this study could be further improved to develop a fragment screening process. Fragments also rather weakly interact with their drug targets, but they display some advantages over larger drug-like molecules. Smaller fragments are less likely to contain interfering moieties that could hinder specific ligand-macromolecule interactions and often better fit into well-defined subsites of the target. Improvements by means of fragment elaboration therefore often result in lead structures with higher affinity.[133] In addition, the introduction of solubilizing groups is also feasible during the fragment optimization process. In contrast to protein targets, fragment screenings with RNA structures using SPR have not yet been published.[99]

Appendix A

Virtual screening for ligands of the *F. nucleatum* FMN riboswitch

TABLE A.1: The main core fragments are represented together with the number of occurrences among the compounds matching the molecular patterns of the substructure search. Only core fragments with at least 20 occurrences are displayed. The core fragments are ranked according to the number of occurrences from left to right and top to bottom.

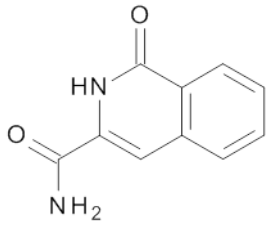
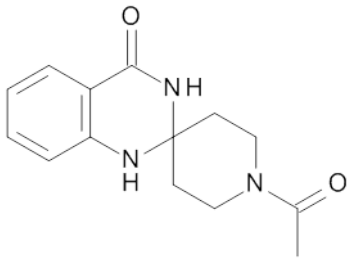
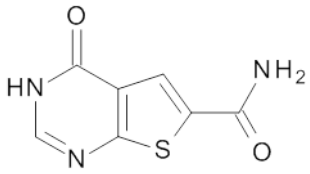
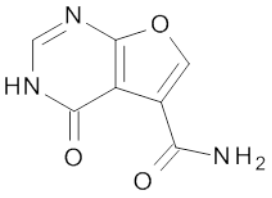
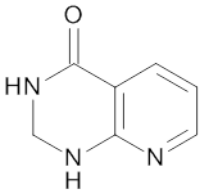
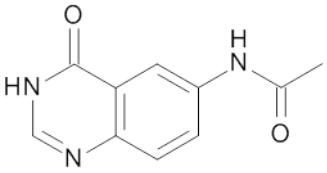
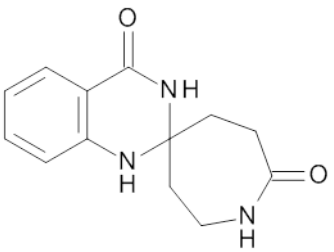
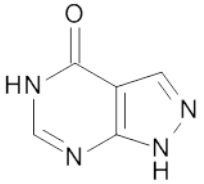
Core fragment	Occurrence	Core fragment	Occurrence
	**376**		**222**
	**146**		**139**
	**126**		**120**
	**101**		**93**

TABLE A.1: continued.

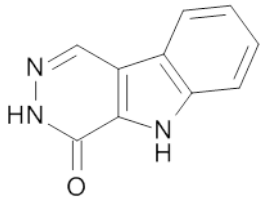
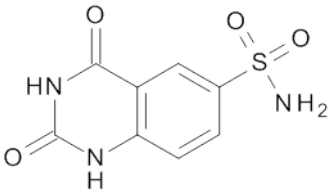
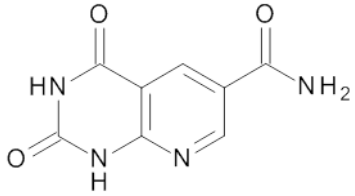
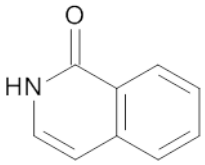
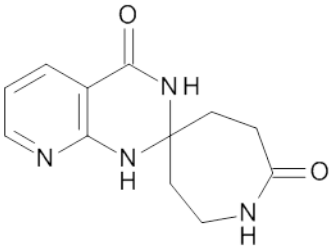
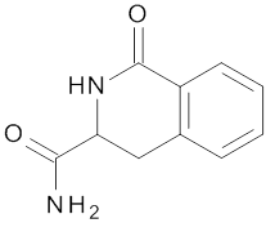
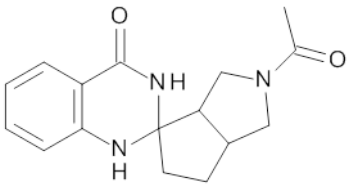
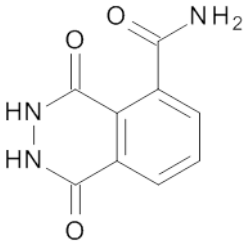
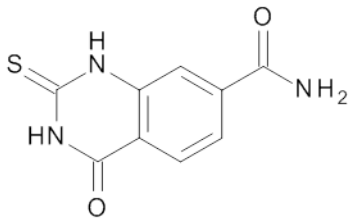
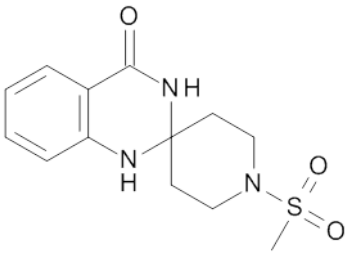
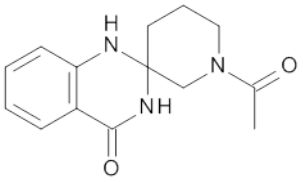
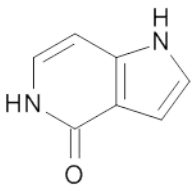
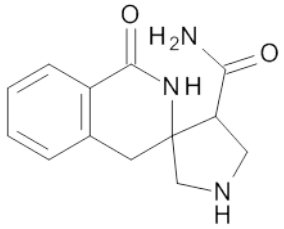
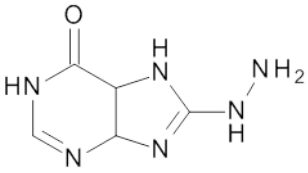
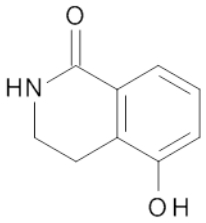
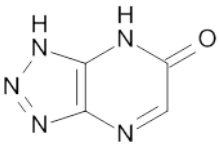
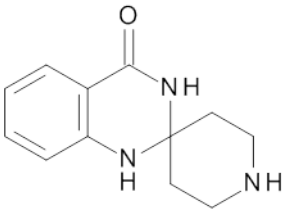
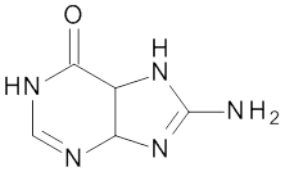
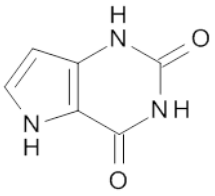
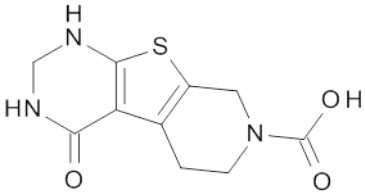
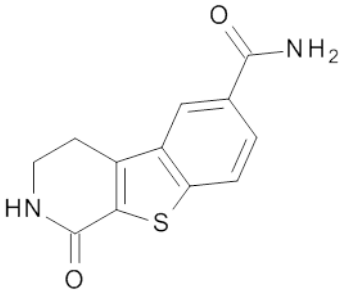
Core fragment	Occurrence	Core fragment	Occurrence
	**82**		**62**
	**61**		**58**
	**54**		**49**
	**47**		**44**
	**40**		**38**
	**34**		**33**

TABLE A.1: continued.

Core fragment	Occurrence	Core fragment	Occurrence
	**31**		**31**
	**30**		**29**
	**29**		**29**
	**27**		**22**
	**22**		

Bibliography

[1] A. Fleming. "On the Antibacterial Action of Cultures of a Penicillium, with Special Reference to their Use in the Isolation of B. influenzae". In: *Br. J. Exp. Pathol.* 10 (1929), pp. 226–236.

[2] A. Coates, Y. Hu, R. Bax, and C. Page. "The future challenges facing the development of new antimicrobial drugs". In: *Nat. Rev. Drug Discovery* 1 (2002), pp. 895–910.

[3] K. Lewis. "Platforms for antibiotic discovery". In: *Nat. Rev. Drug Discovery* 12 (2013), pp. 371–387.

[4] AL. Demain. "Antibiotic: Natural Products Essential to Human Health". In: *Med. Res. Rev.* 29.6 (2009), pp. 821–842.

[5] JA. Jones, KG. Virga, G. Gumina, and KE. Hevener. "Recent advances in the rational design and optimization of antibacterial agents". In: *MedChemComm* 7 (2016), pp. 1694–1715.

[6] ARM. Coates, G. Halls, and Y. Hu. "Novel classes of antibiotics or more of the same?" In: *Br. J. Pharmacol.* 163 (2011), pp. 184–194.

[7] ED. Brown and GD. Wright. "Antibacterial drug discovery in the resistance era". In: *Nature* 529 (2016), pp. 336–343.

[8] AJ. Alanis. "Resistance to Antibiotics: Are We in the Post-Antibiotic Era?" In: *Arch. Med. Res.* 36 (2005), pp. 697–705.

[9] MA. Fischbach and CT. Walsh. "Antibiotics For Emerging Pathogens". In: *Science* 325 (2009), pp. 1089–1093.

[10] PM. Hawkey. "The growing burden of antibacterial resistance". In: *J. Antimicrob. Chemother.* 62 (2008).

[11] JMA. Blair, MA. Webber, A. Baylay, DO. Ogbolu, and LJV. Piddock. "Molecular mechanisms of antibiotic resistance". In: *Nat. Rev. Microbiol.* 13 (2015), pp. 42–51.

[12] CT. Walsh and TA. Wencewicz. "Prospects for new antibiotics: a molecule-centered perspective". In: *J. Antibiot. (Tokyo)* 67 (2014), pp. 7–22.

[13] J. Poehlsgaard and S. Douthwaite. "The bacterial ribosome as a target for antibiotics". In: *Nat. Rev. Microbiol.* 3 (2005), pp. 870–881.

[14] KF. Blount and RR. Breaker. "Riboswitches as antibacterial drug targets." In: *Nat. Biotechnol.* 24 (2006), pp. 1558–1564.

[15] KE. Deigan and AR. Ferré-D'Amaré. "Riboswitches: discovery of drugs that target bacterial gene-regulatory RNAs". In: *Acc. Chem. Res.* 44 (2011), pp. 1329–1338.

[16] CE. Lünse, A. Schüller, and G. Mayer. "The promise of riboswitches as potential antibacterial drug targets". In: *Int. J. Med. Microbiol.* 304 (2014), pp. 79–92.

[17] D. Matzner and G. Mayer. "(Dis)similar Analogues of Riboswitch Metabolites as Antibacterial Lead Compounds". In: *J. Med. Chem.* 58 (2015), 3275–3286.

[18] A. Serganov and E. Nudler. "A Decade of Riboswitches". In: *Cell* 152 (2013), pp. 17–24.

[19] RK. Montagne and RT. Batey. "Riboswitches: Emerging Themes in RNA Structure and Function". In: *Annu. Rev. Biophys.* 37 (2008), pp. 117–133.

[20] WC. Winkler, S. Cohen-Chalamish, and RR. Breaker. "An mRNA structure that controls gene expression by binding FMN". In: *Proc Natl Acad Sci U.S.A.* 99 (2002), pp. 15908–15913.

[21] WC. Winkler and RR. Breaker. "Genetic Control by Metabolite-Binding Riboswitches". In: *Chembiochem* 4 (2003), pp. 1024–1032.

[22] WC Winkler and RR. Breaker. "Regulation of bacterial gene expression by ribowitches". In: *Annu. Rev. Microbiol.* 59 (2005), pp. 487–517.

[23] JE Barrick and RR Breaker. "The distributions, mechanisms, and structures of metabolite-binding riboswitches". In: *Genome Biol.* 8 (2007), R239.

[24] AD. Garst, AL. Edwards, and RT. Batey. "Riboswitches: Structures and Mechanisms". In: *Perspectives in Biology* (2011).

[25] RR. Breaker. "Riboswitches and the RNA World". In: *Cold Spring Harbor Perspect. Biol.* 4 (2012), pp. 1–15.

[26] M. Mandal and RR. Breaker. "Gene regulation by riboswitches". In: *Nat. Rev. Mol. Cell Biol.* 5 (2004), pp. 451–463.

[27] RK. Montagne and RT. Batey. "Structure of the *S*-adenosylmethionine riboswitch regulatory mRNA element." In: *Nature* 441 (2006), pp. 1172–1175.

[28] SD. Gilbert, RP. Rambo, D. Van Tyne, and RT. Batey. "Structure of the SAM-I riboswitch bound to *S*-adenosylmethionine". In: *Nat. Struct. Mol. Biol.* 15 (2008), pp. 177–182.

[29] C. Lu, AM. Smith, RT. Fuchs, F. Ding, K. Rajashankar, TM. Henkin, and A. Ke. "Crystal structures of the SAM-III/S($_{MK}$) riboswitch reveal the SAM-dependent translation inhibition mechanism." In: *Nat. Struct. Mol. Biol.* 15 (2008), pp. 1076–1083.

[30] JJ. Trausch, Z. Xu, AL. Edwards, FE. Reyes, PE. Ross, R. Knight, and RT. Batey. "Structural basis for diversity in the SAM clan of riboswitches". In: *Proc Natl Acad Sci U S A* 111 (2014), pp. 6624–6629.

[31] RR. Breaker. "Prospects for riboswitch discovery and analysis." In: *Mol. Cell* 43 (2011), 867–879.

[32] AV. Sherwood and TM. Henkin. "Riboswitch-Mediated Gene Regulation: Novel RNA Architectures Dictate Gene Expression Responses". In: *Annu. Rev. Microbiol.* 70 (2016), pp. 361–374.

[33] RR. Breaker. "Riboswitches and the RNA World". In: *Cold Spring Harbor Perspect. Biol.* 4 (2012).

[34] A. Peselis and A. Serganov. "Themes and variations in riboswitch structure and function". In: *Biochim. Biophys. Acta* 1839 (2014), pp. 908–918.

[35] A. Serganov, L. Huang, and DJ. Patel. "Structural insights into amino acid binding and gene control by a lysine riboswitch". In: *Nature* 455 (2008), pp. 1263–1267.

[36] RT. Batey, SD. Gilbert, and RK. Montagne. "Structure of a natural guanine-responsive riboswitch complexed with the metabolite hypoxanthine". In: *Nature* 432 (2004), pp. 411–415.

[37] A. Serganov, YR. Yuan, O. Pikovskaya, A. Polonskaia, L. Malinina, AT. Phan, C. Hobartner, R. Micura, RR. Breaker, and DJ. Patel. "Structural basis for discriminative regulation of gene expression by adenine- and guanine-sensing mRNAs". In: *Chem. Biol.* 11 (2004), pp. 1729–1741.

[38] A. Serganov, A. Polonskaia, AT. Phan, RR. Breaker, and DJ. Patel. "Structural basis for gene regulation by a thiamine pyrophosphate-sensing riboswitch". In: *Nature* 441 (2006), pp. 1167–1171.

[39] A. Serganov, L. Huang, and DJ. Patel. "Coenzyme recognition and gene regulation by a flavin mononucleotide riboswitch." In: *Nature* 458 (2009), pp. 233–237.

[40] M. Kang, R. Peterson, and J. Feigon. "Structural Insights into riboswitch control of the biosynthesis of queuosine, a modified nucleotide found in the anticodon of tRNA". In: *Mol. Cell* 33 (2009), pp. 784–790.

[41] DJ. Klein and AR. Ferre-D'Amare. "Structural basis of glmS ribozyme activation by glucosamine-6-phosphate". In: *Science* 313 (2006), pp. 1752–1756.

[42] M. Mandal and RR. Breaker. "Adenine riboswitches and gene activation by disruption of a transcription terminator". In: *Nat. Struct. Mol. Biol.* 11 (2004), pp. 29–35.

[43] K. Hollands, A. Sevostiyanova, and EA. Groisman. "Unusually long-lived pause required for regulation of a Rho-dependent transcription terminator". In: *Proc. Natl. Acad. Sci. U.S.A.* 19 (2014), pp. 1999–2007.

[44] J-F. Lemay, G. Desnoyers, S. Blouin, B. Heppell, L. Bastet, P. St-Pierre, E. Masse, and DA: Lafonatine. "Comparative Study between Transcriptionally- and Translationally-Acting Adenine Riboswitches Reveals Key Differences in Riboswitch Regulatory Mechanisms". In: *PLos Genet.* 7 (2011).

[45] JA. Collins, I. Irnov, S. Baker, and WC. Winkler. "Mechanism of mRNA destabilization by the *glmS* ribozyme". In: *Genes Dev.* 21 (2007), pp. 3356–3368.

[46] E. Loh, O. Dussurget, J. Gripenland, K. Vaitkevicius, T. Tiensuu, P. Mandin, F. Repoila, C. Buchrieser, P. Cossart, and J. Johansson. "A trans-acting riboswitch controls expression of the virulence regulator PrfA in Listeria monocytogenes". In: *Cell* 139 (2009), pp. 770–779.

[47] AG. Chen, N. Sudarsan, and RR. Breaker. "Mechanism for gene control by a natural allosteric group I ribozyme". In: *RNA* 17 (2011), pp. 1967–1972.

[48] RL. Coppins, KB. Hall, and EA. Groisman. "The intricate world of riboswitches". In: *Curr. Opin. Microbiol.* 10 (2007), pp. 176–181.

[49] JK. Wickiser, WC. Winkler, RR. Breaker, and DM. Crothers. "The speed of RNA transcription and metabolite binding kinetics operate an FMN riboswitch". In: *Mol. Cell* 18 (2005), pp. 49–60.

[50] N. Sudarsan, JE. Barrick, and RR. Breaker. "Metabolite-binding RNA domains are present in genes of eukaryotes". In: *RNA* 9 (2003), pp. 644–647.

[51] S. Thore, C. Frick, and N. Ban. "Structural Basis of Thiamine Pyrophosphate Analogues Binding to the Eukaryotic Riboswitch". In: *J. Am. Chem. Soc.* 130 (2008), pp. 8116–8117.

[52] T. Kubodera, M. Watanabe, K. Yoshiuchi, N. Yamashita, A. Nishimura, S. Nakai, K. Gomi, and H. Hanamoto. "Thiamine-regulated gene expression of Aspergillus oryzae thiA requires splicing of the intron containing a riboswitch-like domain in the 5'-UTR". In: *FEBS Lett.* 555 (2003), pp. 516–520.

[53] S. Griffiths-Jones, S. Moxon, M. Marshall, A. Khanna, SR. Eddy, and A. Bateman. "Rfam: annotating non-coding RNAs in complete genomes". In: *Nucleic Acids Res.* 1 (2005), pp. 121–124.

[54] P. Borsuk, A. Przykorska, K. Blachnio, M. Koper, JM. Pawlowicz, M. Pekala, and P. Weglenski. "L-arginine influence the structure and function of arginase mRNA in Aspergillus nidulans". In: *Biol. Chem.* 338 (2007), pp. 135–144.

[55] ER. Lee, KF. Blount, and RR. Breaker. *Metabolite-sensing Riboswitches as Antibacterial Drug Targets in: Emerging Trends in Antibacterial Discovery: Answering the Call to Arms*. Ed. by AA. Miller and PF. Miller. Caister Academic Press, 2011.

[56] J. Mulhbacher, E. Brouilette, M. Allard, L-C. Fortier, F. Malouin, and DA. Lafontaine. "Novel Riboswitch Ligand Analogs as Selective Inhibitors of Guanine-Related Metabolic Pathways". In: *PLoS Pathog.* 6 (2010).

[57] N. Sudarsan, S. Cohen-Chalamish, S. Nakamura, GM. Emilsson, and RR. Breaker. "Thiamine Pyrophosphate Riboswitches Are Targets for the Antimicrobial Compound Pyrithiamine". In: *Chem. Biol. (Cambridge, MA, U.S.)* 12 (2005), pp. 1325–1335.

[58] ER. Lee, KF. Blount, and RR. Breaker. "Roseoflavin is a natural antibacterial compound that bind to FMN riboswitches and regulates gene expression". In: *RNA Biol.* 6 (2009), pp. 187–194.

[59] DB. Pedrolli, A. Matern, J. Wang, M. Ester, K. Siedler, R. Breaker, and M. Mack. "A highly specialized flavin mononucleotide riboswitch responds differently to similar ligands and confers roseoflavin resistance to *Streptomyces davawensis*". In: *Nucleic Acids Res.* 40 (2012), pp. 8662–8673.

[60] KF. Blount, C. Megyola, M. Plummer, D. Osterman, T. O'Connell, P. Aristoff, R. Quinne, A. Chrusciel, TJ. Poel, HJ. Schostarez, CA. Stewart, DP. Walker, PGM. Wuts, and RR. Breaker. "Novel Riboswitch-Binding Flavin Analog That Protects Mice against Clostridium difficile Infection without Inhibiting Cecal Flora". In: *Antimicrob. Agents Chemother.* 59 (2015), 5736–5746.

[61] JA. Howe, H. Wang, TO. Fischmann, CJ. Balibar, L. Xiao, AM. Galgoci, JC. Malinverni, T. Mayhood, A. Villafania, A. Nahvi, N. Murgolo, CM. Barbieri, PA. Mann, D. Carr, E. Xia, P. Zuck, D. Riley, RE. Painter, SS. Walker, B. Sherborne, R. de Jesus, W. Pan, MA. Plotkin, J. Wu, D. Rindgen, J. Cummings, CG. Garlisi, R. Zhang, PR. Sheth, CJ. Gill, H. Tang, and T. Roemer. "Selective small-molecule inhibition of an RNA structural element." In: *Nature* 526 (2015), 672–677.

[62] MS. Gelfand, AA. Mironov, J. Jomantas, YI. Kozlov, and DA. Perumov. "A conserved RNA structure element involved in the regulation of bacterial riboflavin synthesis genes". In: *Trends in Genetics* 15 (1999), pp. 439–442.

[63] AG. Vitreschak, DA. Rodionov, AA. Mironov, and MS. Gelfand. "Regulation of riboflavin biosynthesis and transport genes in bacteria by transcriptional and translational attenuation". In: *Nucleic Acids Res.* 30 (2002), pp. 3141–3151.

[64] Q. Vicens, E. Mondragón, and RT. Batey. "Molecular sensing by the aptamer domain of the FMN riboswitch: a general model for ligand binding by conformational selection". In: *Nucleic Acids Research* 39 (2011), pp. 8586–8598.

[65] JA. Howe, L. Xiao, TO. Fischmann, H. Wang, H. Tang, A. Villafania, R. Zhang, CM. Barbieri, and T. Roemer. "Atomic resolution mechanistics studies of ribocil: A highly selective unnatural ligand mimic of the E.coli FMN riboswitch." In: *RNA Biol.* 13 (2016), pp. 946–954.

[66] EJ. Merino, KA. Wilkinson, JA. Coughla, and KM. Weeks. "RNA structure analysis at single nucleotide resolution by selective 2'-hydroxyl acylation and primer extension (SHAPE)". In: *J. Am. Chem. Soc.* 127 (2005), pp. 4223–4231.

[67] NJ. Baird and AR. Ferre-D'Amare. "Idiosyncratically tuned switching behavior of riboswitch aptamer domains revealed by comparative small-angle X-ray scattering analysis". In: *RNA* 16 (2010), pp. 598–609.

[68] EA. Meyer, RK. Castellano, and F. Diederich. "Interactions with Aromatic Rings in Chemical and Biological Recognition". In: *Angew. Chem. Int. Ed.* 42 (2003), pp. 1210–1250.

[69] S. Otani, M. Takatsu, M. Nakano, S. Kasai, and R. Miura. "Letter: Roseoflavin, a new antimicrobial pigment from Streptomyces". In: *J. Antibiot. (Tokyo)* 27 (1974), pp. 86–87.

[70] S. Grill, S. Busenbender, M. Pfeiffer, U. Köhler, and M. Mack. "The bifunctional flavokinase/flavin adenin dinucleotide synthetase from *Streptomyces davawensis* produces inactive flavin cofactors and is not involved involved in resistance to the antibiotic roseoflavin". In: *J. Bacteriol.* 190 (2008), pp. 1546–1553.

[71] M. Mansjö and J. Johansson. "The Riboflavin analog roseoflavin targets and FMN-riboswitch and blocks *Listeria monocytogenes* growth, but also stimulates virulence gene-expression and infection". In: *RNA Biol.* 8 (2011), pp. 674–680.

[72] DB. Pedrolli, S. Nakanishi, M. Barile, M. Mansurova, EC. Carmona, A. Lux, W. Gartner, and M. Mack. "The antibiotics roseoflavin and 8-demethyl-8-amino-riboflavin from *Streptomyces davawensis* are metabolized by human flavokinase and human FAD synthethase". In: *Biochem. Pharmacol.* 82 (2011), pp. 1853–1859.

[73] C. Bissantz, B. Kuhn, and M. Stahl. "A Medicinal Chemist's Guide to Molecular Interactions". In: *J. Med. Chem.* 53 (2010), pp. 5061–5084.

[74] H. Lodish, A. Berk, S. Lawrence Zipursky, P. Matsudaira, D. Baltimore, and J. Darnell. *Molecular Cell Biology: The Three Roles of RNA in Protein Synthesis*. Ed. by H. Lodish, A. Berk, S. Lawrence Zipursky, P. Matsudaira, D. Baltimore, and J. Darnell. 4th. W.H. Freeman and Company, 2000.

[75] FHC. Crick. "On Protein Synthesis". In: *Symp. Soc. Exp. Biol.* 12 (1958), pp. 138–163.

[76] MS. Weinberg and KV. Morris. "Transcriptional gene silencing in humans". In: *Nucleic Acids Res.* 44.14 (2016), pp. 6505–6517.

[77] A. Nahvi, N. Sudarsan, MS. Ebert, X. Zou, KL. Brown, and RR. Breaker. "Genetic Control by a Metabolite Binding mRNA". In: *Chem. Biol. (Cambridge, MA, U.S.)* 9 (2002), pp. 1043–1049.

[78] JE. Wilusz, H. Sunwoo, and DL. Spector. "Long noncoding RNAs: functional surprises from the RNA world". In: *Genes Dev.* 23 (2009), pp. 1494–1504.

[79] J. Perreault, Z. Weinberg, A. Roth, O. Popescu, P. Chartrand, G. Ferbeyre, and RR. Breaker. "Identification of Hammerhead Ribozymes in All Domains of Life Reveals Novel Structural Variations". In: *PLoS Comput Biol* 7 (2011).

[80] E. Kikovska, NE. Mikkelsen, and LA. Kirsebom. "The naturally trans-acting ribozyme RNase P RNA has leadzyme properties". In: *Nucleic Acids Res* 33 (2005), pp. 6920–6930.

[81] TA. Cooper, L. Wan, and G. Dreyfuss. "RNA and Disease". In: *Cell* 136 (2009), pp. 777–793.

[82] MM. Scotti and MS. Swanson. "RNA mis-splicing in disease". In: *Nat. Rev. Genet.* 17 (2016), pp. 19–32.

[83] CR. Howard and NF. Fletcher. "Emerging virus diseases: can we ever expect the unexpected?" In: *Emerging Microbes & Infections* 1 (2012).

[84] P. Poltronieri, B. Sun, and M. Mallardo. "RNA Viruses: RNA Roles in Pathogenesis, Coreplication and Viral Load". In: *Curr. Genomics* 16 (2015), pp. 327–335.

[85] JR. Thomas and PJ. Hergenrother. "Targeting RNA with Small Molecules". In: *Chem. Rev.* 108.4 (2008), pp. 1171–1224.

[86] NJ. Reiter, CW. Chan, and A. Mondragon. "Emerging structural themes in large RNA molecules". In: *Curr. Opin. Struct. Biol.* 21 (2011), pp. 319–326.

[87] L. Guan and MD. Disney. "Recent advances in developing small molecules targeting RNA". In: *ACS Chem. Biol.* 7 (2012), pp. 73–86.

[88] S. Yoshizawa, D. Fourmy, and JD. Puglisi. "Recognition of the Codon-Anticodon Helix by Ribosomal RNA". In: *Science* 285 (1999), pp. 1722–1725.

[89] A. Davidson, DW. Begley, C. Lau, and G. Varani. "A small molecule probe induces a conformation in HIV TAR RNA capable of binding drug-like fragments". In: *J. Mol. Biol.* 410 (2011), pp. 984–996.

[90] KE. Lind, Z. Du, K. Fujinaga, BM. Peterlin, and TL. James. "Structure-based computational database screening, in vitro assay, and NMR assessment of compounds that target TAR RNA". In: *Chem. Biol. (Cambridge, MA, U.S.)* 9 (2002), pp. 707–712.

[91] F. Chung, C. Tisne, T. Lecourt, B. Seijo, F. Dardel, and L. Micouin. "Design of $tRNA^{Lys3}$ Ligand: Fragment Evolution and Linker Selection Guided by NMR Spectroscopy". In: *Chemistry* 15 (2009), pp. 7109–7116.

[92] SM. Dibrov, J. Parsons, M. Carnevali, S. Zhou, KD. Rynearson, K. Ding, E. Garcia Sega, ND. Brunn, MA. Boerneke, MP. Castaldi, and T. Hermann. "Hepatitis C Virus Translation Inhibitors Targeting the Internal Ribosomal Entry Site". In: *Journal of Medicinal Chemistry* 57 (2014), pp. 1694–1707.

[93] A. Bottini, SK. De, B. Wu, C. Tang, G. Varani, and M. Pellecchia. "Targeting Influenza A virus RNA Promoter". In: *Chem. Biol. Drug Des.* 86 (2015), pp. 663–673.

[94] T. Hermann and E. Westhof. "RNA as a drug target: chemical, modelling, and evolutionary tools". In: *Curr. Opin. Biotechnol.* 9 (1998), pp. 66–73.

[95] J. Mulhbacher, P. St-Pierre, and DA. Lafontaine. "Therapeutic applications of ribozymes and riboswitches". In: *Curr. Opin. Pharmacol.* 10 (2010), pp. 551–556.

[96] M. Garavis, B. Lopez-Mendez, A. Somoza, J. Oyarzabal, C. Dalvit, A. Villasante, R. Campos-Olivas, and C. Gonzalez. "Discovery of Selective Ligands for Telomeric RNA G-quadruplexes (TERRA) through ^{19}F-NMR Based Fragment Screening". In: *ACS Chem. Biol.* (2014), pp. 1559–1566.

[97] SP. Velagapudi, MD. Cameron, CL. Haga, LH. Rosenberg, M. Lafitte, DR. Duckett, DG. Phinney, and MD. Disney. "Design of a small molecule against an oncogenic noncoding RNA". In: *Proc. Natl. Acad. Sci. U.S.A.* 113.21 (2016), pp. 5898–5903.

[98] F. Aboul-ela. "Strategies for the design of RNA-binding small molecules". In: *Future Med. Chem.* 2 (2010), pp. 93–119.

[99] T. Wehler and R. Brenk. "Structure-Based Discovery of Small Molecules Binding to RNA". In: *Topics in Medicinal Chemistry: RNA therapeutics* (2017).

[100] MD. Disney, I. Yildirim, and JL. Childs-Disney. "Methods to enable the design of bioactive small molecules target RNA". In: *Org. Biomol. Chem.* 12 (2014), pp. 1029–1039.

[101] CA. Lipinski, F. Lombardo, BW. Dominy, and PJ. Feeney. "Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings". In: *Adv. Drug Delivery Rev.* 46 (2001), pp. 3–26.

[102] G. Mugumbate and JP. Overington. "The relationship between target class and the physicochemical properties of antibacterial drugs". In: *Bioorg. Med. Chem.* 23 (2015), pp. 5218–5224.

[103] R. O'Shea and HE Moser. "Physicochemical properties of antibacterial compounds: implications for drug discovery". In: *J. Med. Chem.* 51 (2008), pp. 2871–2878.

[104] AC. Anderson. "The Process of Structure-Based Drug Design". In: *Chem. Biol.* 10 (2003), pp. 787–797.

[105] DB. Kitchen, H. Decornez, JR. Furr, and J. Bajorath. "Docking and scoring in virtual screening for drug discovery". In: *Nat. Rev. Drug Discovery* 3 (2004), pp. 935–949.

[106] G. Klebe. "Virtual ligand screening: strategies, perspectives and limitations". In: *Drug Discovery Today* 11 (2006), pp. 580–594.

[107] D. Rognan. *Docking Methods for Virtual Screening: Principles and Recent Advances in: Virtual Screening: Principles, Challenges, and Practical Guidelines*. Ed. by C. Sotriffer. Wiley VCH, 2011.

[108] S. Fulle and H. Gohlke. "Molecular recognition of RNA: challenges for modelling interactions and plasticity". In: *J. Mol. Recognit.* 23 (2009), pp. 220–231.

[109] HM. Berman, J. Westbrook, Z. Feng, G. Gilliland, TN. Bhat, H. Weissig, IN. Shindyalov, and PE. Bourne. "The Protein Data Bank". In: *Nucleic Acids Res.* 28 (2000), pp. 235–242.

[110] N. Foloppe, N. Matassova, and F. Aboul-ela. "Towards the discovery of drug-like RNA ligands?" In: *Drug Discovery Today* 11 (2006), pp. 1019–1027.

[111] JA. Cruz and E. Westhof. "The Dynamic Landscapes of RNA Architecture". In: *Cell* 136 (2009), pp. 604–609.

[112] M. Rarey, B. Kramer, T. Lengauer, and G. Klebe. "A fast flexible docking method using an incremental construction algorithm". In: *J. Mol. Biol.* 261 (1996), pp. 470–489.

[113] GM. Morris, DS. Goodsell, RS. Halliday, R. Huey, WE. Hart, RK. Belew, and AJ. Olson. "Automated Docking Using a Lamarckian Genetic Algorithm and an Empirical Binding Free Energy Function". In: *Journal of Computational Chemistry* 19 (1998), pp. 1639–1662.

[114] DM. Lorber and BK. Shoichet. "Flexible ligand docking using conformational ensembles." In: *Protein Sci.* 7 (1998), pp. 938–950.

[115] BQ. Wei, WA. Baase, LH. Weaver, BW. Matthews, and BK. Shoichet. "A Model Binding Site for Testing Scoring Functions in Molecular Docking". In: *J. Mol. Biol.* 322 (2002), pp. 339–355.

[116] TJ. Ewing, S. Makino, AG. Skillman, and ID. Kuntz. "DOCK 4.0: search strategies for automated molecular docking of flexible molecule ddatabase." In: *J. Comput.-Aided Mol. Des.* 15 (2001), pp. 411–428.

[117] DT. Moustakas, PT. Lang, S. Pegg, E. Pettersen, ID. Kuntz, N. Brooijmans, and RC. Rizzo. "Development and validation of a modular, extensible docking program: DOCK 5." In: *J. Comput.-Aided Mol. Des.* 20 (2006), pp. 601–619.

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

[119] 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." In: *J. Comput. Chem.* 15 (1994), pp. 488–506.

[120] C. Guilbert and TL. James. "Docking to RNA via Root-Mean-Square-Deviation-Driven Energy Minimization with Flexible Ligands and Flexible Targets". In: *J. Chem. Inf. Model.* 48 (2008), pp. 1257–1268.

[121] SD. Morley and M. Afshar. "Validation of an empirical RNA-ligand scoring function for fast flexible docking using Ribodock". In: *J. Comput.-Aided Mol. Des.* 18 (2004), pp. 189–208.

[122] IG. Pinto, C. Guilbert, NB. Ulyanov, J. Stearns, and TL. James. "Discovery of Ligands for a Novel Target, the Human Telomerase RNA, Based on Flexible-Target Virtual Screening and NMR". In: *J. Med. Chem.* 51 (2008), pp. 7205–7215.

[123] N. Foloppe, I-J. Chen, B. Davis, A. Hold, D. Morley, and R. Howes. "A structure-based strategy to identify new molecular scaffold targeting the bacterial ribosomal A-site". In: *Bioorg. Med. Chem.* 12 (2004), pp. 935–947.

[124] C. Detering and G. Varani. "Validation of automated docking programs for docking and database screening against RNA drug targets." In: *J. Med. Chem.* 47 (2004), pp. 4188–4201.

[125] N. Moitessier, E. Westhof, and S. Hanessian. "Docking of Aminoglycosides to Hydrated and Flexible RNA". In: *J. Med. Chem.* 49 (2006), pp. 1023–1033.

[126] DM. Warui and AM. Baranger. "Identification of specific small molecule ligands for stem loop 3 ribonucleic acid of the packaging signal Psi of human immunodeficiency virus-1". In: *J. Med. Chem.* 52 (2009), pp. 5462–5473.

[127] DM. Warui and AM. Baranger. "Identification of small molecule inhibitors of the HIV-1 nucleocapsid-stem-loop 3 RNA complex." In: *J. Med. Chem.* 55 (2012), pp. 4132–4141.

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

[129] AC. Stelzer, AT. Frank, JD. Kratz, MD. Swanson, MJ. Gonzalez-Hernandez, J. Lee, I. Andricioaei, DM. Markovitz, and HM. Al-Hashimi. "Discovery of Selective Bioactive Small Molecules by Targeting an RNA Dynamic Ensemble". In: *Nat. Chem. Biol.* 7 (2011), pp. 553–559.

[130] P. Daldrop, FE. Reyes, DA. Robinson, CM. Hammond, DM. Lilley, RT. Batey, and R. Brenk. "Novel Ligands for a Purine Riboswitch Discovered by RNA-Ligand Docking". In: *Chemistry & Biology* 18 (2011), pp. 324–335.

[131] P. Kolb and JJ. Irwin. "Docking screens: right for the right reasons?" In: *Curr. Top. Med. Chem.* 9 (2009), pp. 755–770.

[132] M. Baker. "Fragment-based lead discovery grows up." In: *Nat. Rev. Drug Discovery* 12 (2012), pp. 5–7.

[133] DA. Erlanson, RS. McDowell, and T. O'Brien. "Fragment-based drug discovery". In: *J. Med. Chem.* 47 (2004), pp. 3463–3482.

[134] M. Congreve, G. Chessari, D. Tisi, and AJ. Woodhead. "Recent developments in fragment-based drug discovery". In: *Journal of Medicinal Chemistry* 51 (2008), pp. 3661–3680.

[135] RA. Carr, M. Congreve, CW. Murray, and DC. Rees. "Fragment-based lead discovery: leads by design". In: *Drug Discovery Today* 10 (2005), pp. 987–992.

[136] M. Congreve, M. Carr, C. Murray, and H. Jhoti. "A "rule of three" for fragment-based lead discovery?" In: *Drug Discovery Today* 8 (2003), pp. 876–877.

[137] K. Bodoor, V. Boyapati, V. Gopu, M. Boisdore, K. Allam, J. Miller, WD. Treleaven, T. Weldeghiorgis, and F. Aboul-ela. "Design and implementation of an ribonucleic acid (RNA) directed fragment library". In: *J. Med. Chem.* 52 (2009), pp. 3753–3761.

[138] B. Francois. "Crystal structures of complexes between aminoglycoamino and decoding A site oligonucleotides: role of the number of ring and positive charges in the specific binding leading to miscoding". In: *Nucleic Acids Res.* 33 (2005), pp. 5677–5690.

[139] SP. Velagapudi, A. Pushechnikov, LP. Labuda, JM. French, and MD. Disney. "Probing a 2-aminobenzimidazole library for binding to RNA internal loops via two-dimensional combinatorial screening". In: *ACS Chem. Biol.* 7 (2012), pp. 1902–1909.

[140] R. Moumne, M. Catala, V. Larue, L. Micouin, and C. Tisne. "Fragment-based design of small RNA binbind: promising developments and contribution of NMR". In: *Biochimie* 94 (2012), pp. 1607–1619.

[141] ML. Verdonk, I. Giangreco, RJ. Hall, O. Korb, PN. Mortenson, and CW. Murray. "Docking performance of fragments and druglike compounds". In: *J. Med. Chem.* 54 (2011), pp. 5422–5431.

[142] CA. Shepherd, AL. Hopkins, and I. Navratilova. "Fragment screening by SPR and advanced application to GPCRs". In: *Prog Biophys Mol Biol* 116 (2014), pp. 113–123.

[143] Z. Yan, S. Sikri, DL. Beveridge, and AM. Baranger. "Identification of an aminoacridine derivative that binds to RNA tetraloops". In: *J. Med. Chem.* 50 (2007), pp. 4096–4104.

[144] M. Zeiger, S. Stark, E. Kalden, B. Ackermann, J. Ferner, U. Scheffer, F. Shoja-Bazargani, V. Erdel, H. Schwalbe, and MW. Göbel. "Fragment based search for small molecule inhibitors of HIV-1 Tat-TAR". In: *Bioorg. Med. Chem. Lett.* 24 (2014), pp. 5576–5580.

[145] PP. Seth, A. Miyaji, EA. Jefferson, KA. Sannes-Lowery, SA. Osgood, SS. Propp, R. Ranken, C. Massire, R. Sampath, DJ. Ecker, EE. Swayze, and RH. Griffey. "SAR by MS: discovery of a new class of RNA-binding small molecules for the hepatitis C virus: internal ribosome entry site IIa subdomain." In: *J. Med. Chem.* 48 (2005), pp. 7099–7102.

[146] J. Fejzo, CA. Lepre, JW. Peng, GW. Bemis, M. Ajay, A. Murcko, and JM. Moore. "The SHAPES strategy: an NMR-based approach for lead generation in drug discovery." In: *Chem. Biol.* 6 (1999), pp. 755–769.

[147] EC. Johnson, VA. Feher, JW. Peng, JM. Moore, and JR. Williamson. "Application of NMR SHAPES screening to an RNA target." In: *J. Am. Chem. Soc.* 125 (2003), pp. 15724–15725.

[148] F. Chung, C. Tisne, T. Lecourt, F. Dardel, and L. Micouin. "NMR-guided fragment-based approach for the design of $tRNA^{Lys3}$ ligands dedesign by a fragment-based approach." In: *Angew. Chem. Int. Ed.* 46 (2007), pp. 4489–4491.

[149] R. Moumne, V. Larue, B. Seijo, T. Lecourt, L. Micouin, and C. Tisne. "Tether influence on the binding properties of $tRNA^{Lys3}$ ligands designed by a fragment-based approach." In: *Org. Biomol. Chem.* 8 (2010), pp. 1154–1159.

[150] M-K. Lee, A. Bottini, M. Kim, MF. Bardaro, Z. Zhang, M. Pellecchia, B-S. Choi, and G. Varani. "A novel small-molecule binds to the influenza A virus RNA promoter and inhibits viral replication." In: *Chem. Commun.* 50 (2014), pp. 368–370.

[151] L. Chen, E. Cressina, FJ. Leeper, AG. Smith, and C. Abell. "A fragment-based approach to identifying ligands for riboswitches". In: *ACS Chem. Biol.* 5 (2010), pp. 355–358.

[152] E. Cressina, L. Chen, M. Moulin, FJ. Leeper, C. Abell, and AG. Smith. "Identification of novel ligands for thiamine pyrophosphate (TPP) riboswitches." In: *Biochem. Soc. Trans.* 39 (2011), pp. 652–657.

[153] KD. Warner and AR. Ferré-D'Amaré. "Crystallographic analysis of TPP riboswitch binding by small molecule ligands discovered through fragment-based drug discovery approaches". In: *Methods Enzymol.* 549 (2014), pp. 221–233.

[154] AR. Leach, BK. Shoichet, and CE. Peishoff. "Prediction of Protein-Ligand Interactions. Docking and Scoring: Successes and Gaps". In: *J. Med. Chem.* 49 (2006), pp. 5851–5855.

[155] T. Zhu, S. Cao, P-C. Su, R. Patel, D. Shah, HB. Chokshi, R. Szukala, ME. Johnson, and KE. Hevener. "Hit Identification and Optimization in Virtual Screening: Practical Recommendations Based Upon a Critical Literature Analysis". In: *J. Med. Chem.* 56 (2013), pp. 6560–6572.

[156] PR. Gerber and K. Müller. "MAB, a generally applicable molecular force field for structure modelling in medicinal chemistry." In: *J. Comput. Aided Mol. Des.* 9 (1995), pp. 251–268.

[157] SJ. Weiner, PA. Kollman, DA. Case, U. Chandra Singh, C. Ghio, G. Alagona, S. Profeta Jr., and P. Weiner. "A New Force Field for Molecular Mechanical Simulation of Nucleic Acids and Proteins". In: *J. Am. Chem. Soc.* 106 (1984), pp. 765–784.

[158] A. Nicholls and B. Honig. "A rapid finite difference algorithm, utilizing successive over-relaxation to solve the Poisson-Boltzmann equation". In: *J. Comput. Chem.* 12 (1991), pp. 435–445.

[159] EC. Meng, BD. Shoichet, and ID. Kuntz. "Automated docking with grid-based energy evaluation". In: *J. Comput. Chem.* 13 (1992), pp. 505–524.

[160] D. Weininger. "SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules". In: *J. Chem. Inf. Model.* 28 (1988), pp. 31–36.

[161] R. Brenk, A. Schipani, D. James, A. Krasowski, IH. Gilbert, J. Frearson, and PG. Wyatt. "Lessons learnt from assembling screening libraries for drug discovery for neglected diseases". In: *ChemMedChem* 3 (2008), pp. 435–444.

[162] CP. Mpamhanga, D Spinks, LB. Tulloch, EJ. Shanks, DA Robinson, IT. Collie, AH. Fairlamb, PG. Wyatt, JA. Frearson, WN. Hunter, IH. Gilbert, and R. Brenk. "One Scaffold, Three Binding Modes: Novel and Selective Pteridine Reductase 1 Inhibitors Derived from Fragment Hits Discovered by Virtual Screening". In: *J. Med. Chem.* 52 (2009), pp. 4454–4465.

[163] PD. Hawkins, GA. Skillman, GL. Warren, BA. Ellingson, and MT. Stahl. "Conformer Generation with OMEGA: Algorithm and Validation Using High Quality Structures from the Protein Databank and Cambridge Structural Database". In: *J. Chem. Inf. Model.* 50 (2010), pp. 572–584.

[164] R. Brenk, SW. Vetter, SE. Boyce, DB. Goodin, and BK. Shoichet. "Probing molecular docking in a charged model binding site." In: *J. Mol. Biol.* 357 (2006), pp. 1449–1470.

[165] T. Kaserer, KR. Beck, M. Akram, A. Odermatt, and D. Schuster. "Pharmacophore Models and Pharmacophore-Based Virtual Screening: Concepts and Applications Exemplified on Hydroxysteroid Dehydrogenases". In: *Molecules* 20 (2015), pp. 22799–22832.

[166] MM. Mysinger and BK. Shoichet. "Rapid context-dependent ligand desolvation in molecular docking." In: *J. Chem. Inf. Model.* 50 (2010), pp. 1561–1573.

[167] LLC Schrödinger. "The PyMOL Molecular Graphics System, Version 1.3". The PyMOL Molecular Graphics System, Version 1.3, Schrödinger LLC.

[168] B. He, M. Rong, D. Lyakhov, H. Gartenstein, G. Diaz, R. Castagna, WT. McAllister, and RK. Durbin. "Rapid Mutagenesis and Purification of Phage RNA Polymerases". In: *Protein Expression Purif.* 9 (1997), pp. 142–151.

[169] O. Pikovskaya, AA. Serganov, A. Polonskaia, A. Serganov, and DJ. Patel. "Preparation and crystallization of riboswitch-ligand complexes". In: *Methods Mol. Biol.* 540 (2009), pp. 115–128.

[170] AL. Edwards, AD. Garst, and RT. Batey. "Determining Structures of RNA Aptamers and Riboswitches by X-Ray Crystallography". In: *Methods Mol. Biol.* 535 (2009), pp. 135–163.

[171] SM. Hassur and HW. Jr. Whitlock. "UV Shadowing - A New and Convenient Method for the Location of Ultraviolet-Absorbing Species in Polyacrylamide Gels". In: *Anal. Biochem.* 59 (1974), pp. 162–164.

[172] JF. Milligan and OC. Uhlenbeck. "Synthesis of small RNAs using T7 RNA polymerase". In: *Methods Enzymol.* 180 (1989), pp. 51–62.

[173] BL. Golden. "Preparation and Crystallization of RNA". In: *Methods Mol. Biol.* 363 (2007), pp. 239–257.

[174] C. Kao, M. Zheng, and S. Rudisser. "A simple and efficient method to reduce nontemplated nucleotide addition at the 3' terminus of RNAs transcribed by T7 RNA polymerase". In: *RNA* 5 (1999), pp. 1268–1272.

[175] MF. Schaffer, PK. Choudhary, and RK. Sigel. "The AdoCbl-riboswitch interaction investigated by in-line probing and surface plasmon resonance spectroscopy (SPR)." In: *Methods Enzymol.* 549 (2014), pp. 467–488.

[176] AR. Gruber, R. Lorenz, SH. Bernhart, R. Neuböck, and IL. Hofacker. "The Vienna RNA Websuite". In: *Nucleic Acids Res.* 36 (2008), W70–W74.

[177] AJ. Rinaldi, KC. Suddala, and NG. Walter. "Native purification and labeling of RNA for Single Molecule Fluorescence Studies". In: *Methods Mol. Biol.* 1240 (2015), pp. 63–95.

[178] DK. Willkomm and RK. Hartmann. "3'-Terminal Attachment of Fluorescent Dyes and Biotin". In: *Handbook of RNA Biochemistry* (2014). Ed. by RK. Hartmann, A. Bindereif, A. Schön, and E. Westhof.

[179] RW. Richardson and RI. Gumport. "Biotin and fluorescent labeling of RNA using T4 RNA ligase". In: *Nucleic Acids Res.* 11 (1983), pp. 6167–6184.

[180] GT. Hermanson. *Bioconjugate Techniques - Chapter 27: Nucleic Acid and Oligonucleotide Modification and Conjugation*. Ed. by GT. Hermanson. 2nd ed. Elsevier, 2008.

[181] A. Mukherjee, J. Walker, KB. Weyant, and CM. Schroeder. "Characterization of Flavin-Based Fluorescent Proteins: An Emerging Class of Fluorescent Reporters". In: *PLoS One* 8 (2013).

[182] Y-L. Zhang and Z-Y. Zhang. "Low-Affinity Binding Determined by Titration Calorimetry Using a High-Affinity Coupling Ligand: A Thermodynamic Study of Ligand Binding to Protein Tyrosine Phosphatase 1B". In: *Anal. Biochem.* 261 (1998), pp. 139–148.

[183] EC. Hulme and MA. Trevethick. "Ligand binding assay at equilibrium: validation and interpretation". In: *Br. J. Pharmacol.* 161 (2010), pp. 1219–1237.

[184] BW. Turnbull. *Divided we fall? Studying low-affinity fragments of ligands by ITC*. Tech. rep. GE Healthcare Life Sciences, 2005.

[185] *Biacore Assay Handbook*. 29-0194-00 AA. GE Healthcare. May 2012.

[186] Y. Liu and DW. Wilson. "Quantitative Analysis of Small Molecule-Nucleic Acid Interactions With a Biosensor Surface and Surface Plasmon Resonance Detection". In: *Methods Mol. Biol.* 613 (2010), pp. 1–23.

[187] W. Kabsch. "Integration, scaling, space-group assignment and post-refinement". In: *Acta Crystallogr D Biol Crystallogr.* 66(Pt2) (2010), pp. 133–144.

[188] PR. Evans. "An introduction to data reduction: space-group determination, scaling and intensity statistics". In: *Acta Crystallogr D Biol Crystallogr.* 67(Pt 4) (2011), pp. 282–292.

[189] BK. Shoichet, DL. Bodian, and ID. Kuntz. "Molecular Docking Using Shape Descriptors". In: *J. Comput. Chem.* 13 (1992), pp. 380–397.

[190] S. Desvergnes, S. Courtiol-Legourd, R. Daher, M. Dabrowski, and L. Salmon. "Synthesis and evaluation of malonate-based inhibitors of phosphosugar-metabolizing enzymes: Class II fructose-1,6-bis-phosphate aldolases, type I phosphomannose isomerase, and phosphogluco isomerase". In: *Bioorganic & Medicinal Chemistry* 20 (2012), pp. 1511–1520.

[191] GF. Ruda, G. Campbell, VP. Alibu, MP. Barrett, R. Brenk, and IH. Gilbert. "Virtual fragment screening for novel inhibitors of 6-phosphogluconate dehydrogenase". In: *Bioorg. Med. Chem.* 18 (2010), pp. 5056–5062.

[192] R. Brenk, JJ. Irwin, and BK. Shoichet. "Here Be Dragons: Docking and Screening in an Uncharted Region of Chemical Space". In: *J. Biomol. Screening* 10 (2005), pp. 667–674.

[193] J. Gao, DA. Bosco, ET. Powers, and JW. Kelly. "Localized thermodynamic coupling between hydrogen bonding and microenvironment polarity significantly stabilizes proteins." In: *Nat. Struct. Mol. Biol.* 16 (2009), pp. 684–690.

[194] KE. Hevener, W. Zhao, DM. Ball, K. Babaoglu, J. Qi, SW. White, and RE. Lee. "Validation of Molecular Docking Programs for Virtual Screening against Dihydropteroate Synthase". In: *J. Chem. Inf. Model.* 49 (2009), pp. 444–460.

[195] JC. Cole, CW. Murray, JW. Nissink, RD. Taylor, and R. Taylor. "Comparing protein-ligand docking programs is difficult". In: *Proteins: Structure, Function, and Bioinformatics* 60 (2005), pp. 325–332.

[196] TI. Oprea, AM. Davis, SJ. Teague, and PD. Leeson. "Is There a Difference between Leads and Drugs? A Historical Perspective". In: *J. Chem. Inf. Comput. Sci.* 41 (2001), pp. 1308–1315.

[197] MM. Hann, AR. Leach, and G. Harper. "Molecular complexity and its impact on the probability of finding leads for drug discovery". In: *J. Chem. Inf. Comput. Sci.* 41 (2001), pp. 856–864.

[198] JA. Erickson, M. Jalaie, DH. Robertson, RA. Lewis, and M. Vieth. "Lessons in Molecular Recognition: The Effects of Ligand and Protein Flexibility on Molecular Docking Accuracy". In: *J. Med. Chem.* 47 (2004), pp. 45–55.

[199] AL. Hopkins, GM. Keserü, PD. Leeson, DC. Rees, and CH. Reynolds. "The role of ligand efficiency metrics in drug discovery". In: *Nat. Rev. Drug Discovery* 13 (2014), pp. 105–121.

[200] SD. Gilbert and RT. Batey. "Monitoring RNA–Ligand Interactions Using Isothermal Titration Calorimetry". In: *Methods Mol. Biol.* 540 (2009), pp. 97–114.

[201] J-H. Zhang, TDY. Chung, and KR. Oldenburg. "A Simple Statistical Parameter for Use in Evaluation and Validation of High Throughput Screening Assays". In: *J. Biomol. Screening* 4 (1999), pp. 67–73.

[202] A. Velazquez Campoy and E. Freire. "ITC in the post-genomic era...? Priceless". In: *Biophys. Chem.* 115 (2005), pp. 115–124.

[203] RL. Rich and DG. Myszka. "Survey of the year 2007 commercial optical biosensor literature". In: *J. Mol. Recognit.* 21 (2008), pp. 355–400.

[204] IV. Terekhova, M. Kozbial, RS. Kumeev, and GA. Alper. "Inclusion Complex Formation Between Modified Cyclodextrins and Riboflavin and Alloxazine in Aqueous Solution". In: *J. Solution Chem.* 40 (2011), pp. 1435–1446.

[205] M. Ishikawa and Y. Hashimoto. "Improvement in Aqueous Solubility in Small Molecule Drug Discovery Programs by Disruption of Molecular Planarity and Symmetry". In: *J. Med. Chem.* 54 (2011), pp. 1539–1554.

[206] NN. Salim and AL. Feig. "Isothermal Titration Calorimetry of RNA". In: *Methods* 47 (2009), pp. 198–205.

[207] M. Wu and I. Jr. Tinoco. "RNA folding causes secondary structure rearrangement". In: *Proc Natl Acad Sci U.S.A.* 95 (1998), pp. 11555–11560.

[208] A. Al-Maaieh and DR. Flanagan. "Salt Effects On Caffeine Solubility, Distribution, And Self-Association". In: *J. Pharm. Sci.* 91 (2002), pp. 1000–1008.

[209] J. Lee, CE. Vogt, M. McBrairty, and HM. Al-Hashimi. "Influence of Dimethylsulfoxide on RNA Structure and Ligand Binding". In: *Anal. Chem.* 85 (2013), pp. 9692–9698.

[210] N. Nijegorodov, PVC. Luhanga, JS. Nkoma, and DP. Winkoun. "The influence of planarity, rigidity and internal heavy atom upon fluorescence parameters and the intersystem crossing rate constant in molecules with biphenyl basis". In: *Spectrochim. Acta, Part A* 64 (2006), pp. 1–5.

[211] JD. Winefordner, PA. St. John, and WJ. McCarthy. *Fluorescence Assay in Biology and Medicine*. Ed. by S. Udenfriend. Vol. II. Academic Press, 1969.

[212] JF. Eccleston, JP. Hutchinson, and DM. Jameson. *Progress in Medicinal Chemistry: Fluorescence-Based Assays*. Ed. by F.D. King and G. Lawton. Vol. 43. Elsevier, 2005.

[213] DE. Duggan, RL. Bowman, BB. Brodie, and S. Udenfriend. "A Spectrophotofluorometric Study of Compounds of Biological Interest". In: *Arch. Biochem. Biophys.* 68 (1957), pp. 1–14.

[214] SD. Gilbert, CD. Stoddard, SJ. Wise, and RT. Batey. "Thermodynamic and Kinetic Characterization of Ligand Binding to the Purine Riboswitch Aptamer Domain". In: *J. Mol. Biol.* 359 (2006), pp. 754–768.

[215] MR. Duff Jr., J. Grubbs, and EE. Howell. "Isothermal Titration Calorimetry for Measuring Macromolecule-Ligand Affinity". In: *Journal of Visualized Experiments* 55 (2011), p. 2796.

[216] RS. Pearlman, H. Yalkowsky, and S. Banerjee. "Water Solubilities of Polynuclear Aromatic and Heteroaromatic Compounds". In: *J. Phys. Chem. Ref. Data* 13 (1984), pp. 555–562.

[217] JM. Berg, JL. Tymoczko, and L. Stryer. *Biochemistry*. Ed. by JM. Berg, JL. Tymoczko, and L. Stryer. 5th ed. W.H. Freeman and Company, 2002.

[218] BR. Beno, K-S. Yeung, MD. Bartberger, LD. Pennington, and NA. Meanwell. "A Survey of the Role of Noncovalent Sulfur Interactions in Drug Design". In: *Journal of Medicinal Chemistry* 58 (2015), pp. 4383–4438.

[219] P. Zhou, F. Tian, F. Lv, and Z. Shang. "Geometric characteristics of hydrogen bonds involving sulfur atoms in proteins." In: *Proteins: Struct., Funct., Bioinf.* 76 (2008), pp. 151–163.

[220] E. Sochacka, RH. Szczepanowski, M. Cypryk, M. Sobsczak, M. Janicka, K. Kraszewska, P. Bartos, A. Chwialkowska, and B. Nawrot. "2-Thiouracil deprived of thiocarbonyl function preferentially base pairs with guanine rather than adenine in RNA and DNA duplexes". In: *Nucleic Acids Res* 43 (2015), pp. 2499–2512.

[221] F. Wennmohs, V. Staemmler, and M. Schindler. "Theoretical investigation of weak hydrogen bonds to sulfur". In: *J. Chem. Phys.* 119 (2003), pp. 3208–3218.

[222] M. Piliarik, H. Vaisocherova, and J. Homola. "Surface Plasmon Resonance Biosensing". In: *Methods Mol. Biol.* 503 (2009), pp. 65–88.

[223] JL. Jenkins, J. Krucinska, RM. McCarty, V. Bandarian, and JE. Wedekind. "Comparison of a $PreQ_1$ Riboswitch Aptamer in Metabolite-bound and Free States with Implications for Gene Regulation". In: *J. Biol. Chem.* 286 (2011), pp. 24626–24637.

[224] Y. Yang, Q. Wang, and D. Guo. "A Novel Strategy for Analyzing RNA-Protein Interactions by Surface Plasmon Resonance Biosensor". In: *Mol. Biotechnol.* 40 (2008), pp. 87–93.

[225] AB. Rode, T. Endoh, and N. Sugimoto. "Tuning Riboswitch-Mediated Gene Regulation by Rational Control of Aptamer Ligand Binding Properties". In: *Angew. Chem. Int. Ed.* 54 (2015), pp. 505–509.

[226] E. Lionta, G. Spyrou, DK. Vassilatis, and Z. Cournia. "Structure-Based Virtual Screening for Drug Discovery: Principles, Applications and Recent Advances". In: *Curr. Top. Med. Chem.* 16 (2014), pp. 1923–1938.

[227] D. Stumpfe, P. Ripphausen, and J. Bajorath. "Virtual compound screening in drug discovery". In: *Future Med. Chem.* 4 (2012), pp. 593–602.

[228] NJ. Deng and P. Cieplak. "Free energy profile of RNA hairpins: a molecular dynamics simulation study." In: *Biophys. J.* 98 (2010), pp. 627–636.

[229] S-Y. Huang and X. Zou. "Advances and Challenges in Protein-Ligand Docking". In: *Int. J. Mol. Sci.* 11 (2010), pp. 3016–3034.

[230] TE. Edwards and AR. Ferre-D'Amare. "Crystal structures of the thi-box riboswitch bound to thiamine pyrophosphate analogs reveal adaptive RNA-small molecule recognition". In: *Structure* 14 (2006), pp. 1459–1468.

[231] DF. Veber, SR. Johnson, HY. Cheng, BR. Smith, KW. Ward, and KD. Kopple. "Molecular properties that influence the oral bioavailability of drug candidates." In: *J. Med. Chem.* 45 (2002), pp. 2615–2623.

[232] WJ. Jusko and G. Levy. "Absorption, Metabolism, and Excretion of Riboflavin-5'-phosphate in Man". In: *J. Pharm. Sci.* 56 (1967), pp. 58–62.

[233] TA. Halgren and W. Damm. "Polarizable force field". In: *Curr. Opin. Struct. Biol.* 11 (2001), pp. 236–242.

[234] HA. Stern, GA. Kaminski, JL. Banks, R. Zhou, BJ. Berne, and RA. Friesner. "Fluctuating charge, polarizable dipole, and combined models: Parameterization from *ab initio* quantum chemistry". In: *J. Phys. Chem. B* 103 (1999), pp. 4730–4737.

[235] RF. Brown, CT. Andrews, and AH. Elcock. "Stacking free energies of all DNA and RNA nucleoside pairs and dinucleoside-monophosphates computed using recently revised AMBER parameters and compared with experiment". In: *J. Chem. Theory Comput.* 11 (2015), pp. 2315–2328.

[236] GM. Whitesides and VM. Krishnamurty. "Designing ligands to bind proteins". In: *Q. Rev. Biophys.* 38 (2005), pp. 385–395.

[237] GA. Jeffrey and H. Maluszynska. "A survey of the geometry of hydrogen bonds in the crystal structures of barbiturates, purines and pyrimidines". In: *J. Mol. Struct.* 147 (1986), pp. 127–142.

[238] P. Gilli, L. Pretto, V. Bertolasi, and G. Gilli. "Predicting Hydrogen-Bond Strengths from Acid-Base Molecular Properties. The pK_a Slide Rule: Toward the Solution of a Long-Lasting Problem". In: *Acc. Chem. Res.* 42 (2009), pp. 33–44.

[239] S. Schultes, C. de Graaf, EEJ. Haaksma, IJP. de Esch, R. Leurs, and O. Krämer. "Ligand efficiency as a guide in fragment hit selection and optimization". In: *Drug Discovery Today* 7.3 (2010), pp. 157–162.

[240] T. ten Brink and TE. Exner. "Influence of protonation, tautomeric, and stereoisomeric states on protein-ligand docking results." In: *J. Chem. Inf. Model.* 49 (2009), pp. 1535–1546.

[241] B. Stanovnik, M. Tisler, AR. Katritzky, and OV. Denisko. *The Tautomerism of Heterocycles: Substituent Tautomerism of Six-Membered Ring Heterocycles*. Ed. by AR. Katritzky. Vol. 91. Elsevier, 2011, pp. 1–108.

[242] SG. Schulman. "pH Dependence of Fluorescence of Riboflavin and Related Isoalloxazine Derivatives". In: *J. Pharm. Sci.* 60 (1971), pp. 628–631.

[243] WF. Baitinger, R. von Schleyer, TSSR. Murty, and L. Robinson. "Nitro Groups as Proton Acceptor in Hydrogen Bonding". In: *Tetrahedron* 20 (1964), pp. 1635–1647.

[244] GF. Smith. *Progress in Medicinal Chemistry: Designing Drugs to Avoid Toxicity*. Ed. by G. Lawton and DR. Witty. Elsevier, 2011.

[245] UA. Boelsterli, HK. Ho, S. Zhou, and KY. Leow. "Bioactivation and hepatotoxicity of nitroaromatic drugs." In: *Curr. Drug Metab.* 7 (2006), pp. 715–727.

[246] X-Y. Meng, H-X. Zhang, M. Mezei, and M. Cui. "Molecular Docking: A powerful approach for structure-based drug discovery". In: *Curr. Comput.-Aided Drug Des.* 7 (2011), pp. 146–157.

[247] P. Ferrara, H. Gohlke, DJ. Price, G. Klebe, and CL 3rd Brooks. "Assessing scoring functions for protein-ligand interactions". In: *J. Med. Chem.* 47 (2004), pp. 3032–3047.

[248] GL. Warren, CW. Andrews, AM. Capelli, B. Clarke, J. LaLonde, LH. Lambert, M. Lindvall, N. Nevins, SF. Semus, S. Senger, G. Tedesco, ID. Wall, JM. Woolven, CE. Peishoff, and MS. Head. "A critical assessment of docking programs and scoring functions". In: *J. Med. Chem.* 49 (2006), pp. 5912–5931.

[249] P. Ripphausen, B. Nisius, L. Peltason, and J. Bajorath. "Quo Vadis, Virtual Screening? A Comprehensive Survey of Prospective Applications". In: *J. Med. Chem.* 53 (2010), pp. 8461–8467.

[250] M. Stahl and M. Rarey. "Detailed Analysis of Scoring Functions for Virtual Screening". In: *J. Med. Chem.* 44 (2001), pp. 1035–1042.

[251] T. Cheng, Q. Li, Z. Zhou, Y. Wang, and SH. Bryant. "Structure-Based Virtual Screening for Drug Discovery: a Problem-Centric Review". In: *AAPS J.* 14.1 (Mar. 2012), pp. 133–141.

[252] R. Teramoto and H. Fukunishi. "Supervised Scoring Models with Docked Ligand Conformations for Structure-Based Virtual Screening". In: *J. Chem. Inf. Model.* 47 (2007), pp. 1858–1867.

[253] MHJ. Seifert. "Optimizing the Signal-to-Noise Ratio of Scoring Functions for Protein-Ligand Docking". In: *J. Chem. Inf. Model.* 48 (2008), pp. 602–612.

[254] TA. Pham and AN. Jain. "Customizing scoring functions for docking". In: *J. Comput.-Aided Mol. Des.* 22 (May 2008), pp. 269–286.

[255] C. Catana and PFW. Stouten. "Novel, Customizable Scoring Functions, Parameterized Using N-PLS, for Structure-Based Drug Discovery". In: *J. Chem. Inf. Model.* 47 (2007), pp. 85–91.

[256] M. Naderi, C. Alvin, Y. Ding, S. Mukhopadhyay, and M. Brylinski. "A graph-based approach to construct target-focused libraries for virtual screening". In: *Journal of Chemoinformatics* 8 (2016).

[257] AR. Leach and MM. Hann. "The *in silico* world of virtual libraries". In: *Drug Discovery Today* 5 (2000), pp. 326–336.

[258] S. Forli. "Charting a Path to Success in Virtual Screening". In: *Molecules* 20 (2015), pp. 18732–18758.

[259] WA. Lea and A. Simeonov. "Fluorescence Polarization Assays in Small Molecule Screening". In: *Expert Opin. Drug Discovery* 6 (2011), pp. 17–32.

[260] TJ. Hou, K. Xia, W. Zhang, and XJ. Xu. "ADME Evaluation in Drug Discovery. 4. Prediction of Aqueous Solubility Based on Atom Contribution Approach". In: *J. Chem. Inf. Comput. Sci.* 44 (2004), pp. 266–275.

[261] J. Wielens, SJ. Headey, DI. Rhodes, RJ. Mulder, O. Dolezal, JJ. Deadman, J. Newman, DK. Chalmers, MW. Parker, TS. Peat, and MJ. Scanlon. "Parallel Screening of Low Molecular Weight Fragment Libraries - Do Differences in Methodology Affect Hit Identification?" In: *Journal of Biomolecular Screening* 18 (2013), pp. 147–159.

[262] T. Schulz-Gasch and M. Stahl. "Scoring functions for protein-ligand interactions: a critical perspective". In: *Drug Discovery Today* 1 (2004), pp. 231–239.

[263] DJ. Cole, JZ. Vilseck, J. Tirado-Rives, MC. Panye, and WL. Jorgensen. "Biomolecular Force Field Parameterization via Atoms-in-Molecule Electron Density Partitioning". In: *J. Chem. Theory Comput.* 12 (2016), pp. 2312–2323.

[264] G. Lever, DJ. Cole, R. Lonsdale, KE. Ranaghan, DJ. Wales, AJ. Mulholland, C-K. Skylaris, and MC. Payne. "Large-Scale Density Functional Theory Transition State Searching in Enzymes". In: *J. Phys. Chem. Lett.* 5 (2014), pp. 3614–3619.

[265] IS. Ufimtsev, N. Luehr, and TJ. Martinez. "Charge Transfer and Polarization in Solvated Proteins from Ab Initio Molecular Dynamics". In: *J. Phys. Chem. Lett.* 2 (2011), pp. 1789–1793.

[266] S. Langer, M. Hashimoto, B. Hobl, T. Mathes, and M. Mack. "Flavoproteins Are Potential Targets for the Antibiotic Roseoflavin in *Escherichia coli*". In: *J. Bacteriol.* 195 (2013), pp. 4037–4045.

[267] M. Rueda, G. Bottegoni, and R. Abagyan. "Recipes for the Selection of Experimental Protein Conformations for Virtual Screening". In: *J. Chem. Inf. Model.* 50 (2010), pp. 186–193.

[268] WL. Jorgensen. "Efficient Drug Lead Discovery and Optimization". In: *Acc. Chem. Res.* 42 (2009), pp. 724–733.

[269] Q. Ul Ain, A. Aleksandrova, FD. Roessler, and PJ. Ballester. "Machine-learning scoring functions to improve structure-based binding affinity prediction and virtual screening". In: *Wiley Interdisciplinary Reviews. Computational Molecular Science* 5 (2015), pp. 405–424.

[270] SF. Hickey and MC. Hammond. "Structure-Guided Design of Fluorescent *S*-Adenosylmethionine Analogs for a High-Throughput Screen to Target SAM-I Riboswitch RNAs". In: *Cell Chemical Biology* 21 (2014), pp. 345–356.

[271] A. Ke and JA. Doudna. "Crystallization of RNA and RNA-protein complexes". In: *Methods* 34 (2004), pp. 408–414.

[272] FE. Reyes, CE. Schwartz, JA. Tainer, and RP. Rambo. *Methods in Enzymology: Riboswitch Discovery, Structure and Function*. Ed. by DH. Burke-Aguero. Vol. 549. Elsevier, 2014.

[273] SR. Price, C. Oubridge, G. Varani, and K. Nagai. *RNA-Protein Interactions. A Practical Approach*. Ed. by CWJ. Smith. Oxford University Press, 1998.

[274] OC. Uhlenbeck. "Keeping RNA happy". In: *RNA* 1 (1995), pp. 4–6.

[275] RT. Batey and JS. Kieft. "Improved native affinity purification of RNA". In: *RNA* 13 (2007), pp. 1384–1389.

[276] LE. Easton, Y. Shibata, and PJ. Lukavsky. "Rapid, nondenaturing RNA purification using weak anion exchange fast performance liquid chromatography". In: *RNA* 16 (2010), pp. 647–653.

[277] I. Kim, SA. McKenna, EV. Puglisi, and JD. Puglisi. "Rapid purification of RNAs using fast performance liquid chromatography (FPLC)". In: *RNA* 13 (2007), pp. 289–294.

[278] MIA. Moneam, AA. Geies, GM. El-Naggar, and SM. Mousa. "Synthesis of New Furo[2,3-d]pyrimidines and Pyrimido[4',5':4,5]furo[2,3-d]pyrimidines". In: *J. Chin. Chem. Soc.* 51 (2004), pp. 1357–1366.

[279] R. Chinnappan, A. Dube, J-F. Lemay, and DA. Lafontaine. "Fluorescence monitoring of riboswitch transcription regulation using a dual molecular beacon assay". In: *Nucleic Acids Res.* 41 (2013), pp. 106–114.

[280] S. Perspicace, D. Banner, J. Benz, F. Müller, D. Schlatter, and W. Huber. "Fragment-based screening using surface plasmon resonance technology". In: *Journal of Biomolecular Screening* 14 (2009), pp. 337–349.

[281] I. Navratilova and AL. Hopkins. "Fragment Screening by Surface Plasmon Resonance". In: *ACS Med. Chem. Lett.* 1 (2010), pp. 44–48.