From DEPARTMENT OF BIOSCIENCES AND NUTRITION
Karolinska Institutet, Stockholm, Sweden

BIOMOLECULAR SIMULATIONS, FROM RNA TO PROTEIN: THERMODYNAMIC AND DYNAMIC ASPECTS

Olof Allnér

Karolinska Institutet

Stockholm 2012

Published by Karolinska Institutet. Printed by Reproprint

ISBN 978-91-7457-801-0

To the Less Fortunate

ABSTRACT

The process of transforming the information stored in the DNA of genes into functional RNA molecules and proteins via transcription and translation is the most fundamental process of all known life. Even though these processes involve large macromolecules and dynamics on long time scales they all ultimately rely on atomic level interactions between nucleic acids or amino acids. Only a few experimental techniques are available that can study the large systems involved in atomic detail. Computer simulations, modeling biological macromolecules, are therefore an important tool in investigating fundamental biological processes. In this thesis, Molecular Dynamics (MD) simulations have been used to study the translation of mRNA by tRNA and the function of the regulatory riboswitches. The thesis also covers the improvement of methodology by the development of a new representation of the important Mg^{2+} ions and an improvement of the understanding of the connection between MD and experimental NMR data.

In Paper I, the effect of post transcriptional modifications of the tRNA anti codon on the decoding of mRNA in the ribosome is studied. All atom MD simulations have been performed of the ribosomal A site with and without modifications present, including extensive free energy calculations. The results show two mechanism by which the decoding is affected: The further reach provided by the modifications allows an alternative outer conformation to be formed for the non cognate base pairs, and the modifications results in increased “catalytic” contacts between tRNA, mRNA and the ribosome.

In Paper II, the folding mechanism of the *add* A riboswitch is studied under different ionic conditions and with and without the ligand bound. In addition to standard simulations, we simulated the unfolding by umbrella sampling of distance between the L2 and L3 loops. In the results, no significant effect of Mg^{2+} or Na^{+} ion environments or ligand presence can be seen. But a consistent mechanism with the P3 stem being more flexible than P2 is observed. More data might however be needed to draw general conclusions.

In Paper III, the parameters describe Mg^{2+} ions in MD simulations are improved by optimizing to kinetic data of the H_2O exchange. Data from NMR relaxation experiments was used as optimization goal. The newly developed parameters do not only display better kinetic properties, but also better agreement with experimental structural data.

In Paper IV, the dynamical data, obtained from NMR relaxation experiment of a protein is related to dynamics seen in an MD simulation. The analysis provides important information for the interpretation of experimental data and the development of simulation methods. The results show, among other things, that significant parts of the entropy are not seen by NMR due to a limited time window and inability to account for correlation of motions.

LIST OF PUBLICATIONS

I. OLOF ALLNÉR and Lennart Nilsson. *Nucleotide modifications and tRNA anticodon - mRNA codon interactions on the ribosome*, 2011, RNA, 17(12):2177-2188

II. OLOF ALLNÉR, Lennart Nilsson, Alessandra Villa. *Effects of Ion Environment and Ligand on the Kissing Loops in the add Adenine-Riboswitch*, 2012, manuscript

III. OLOF ALLNÉR, Lennart Nilsson, Alessandra Villa. *Magnesium Ion-Water Coordination and Exchange in Biomolecular Simulations*, 2012, Journal of Chemical Theory and Computation, 8(4):1493-1502

IV. OLOF ALLNÉR and Lennart Nilsson. *Motions and Entropies in Proteins as Seen in NMR Relaxation and Molecular Dynamics Simulations*, 2012, manuscript

CONTENTS

LIST OF ABBREVIATIONS

16S	RNA molecule in 30S
23S	Large RNA molecule in 50S
30S	Small ribosomal subunit
50S	Large ribosomal subunit
5S	Small RNA molecule in 50S
70S	Complete bacterial ribosome
aa	aminoacyl
AMBER	Assisted Model Building with Energy Refinement
ASL	Anti Stem Loop
DNA	DeoxyRibonecleic Acid
EF	Elongation Factor
FEP	Free Energy Perturbation
FMN	Flavin MonoNucleotide
FPP	Farnesyl diphosphate
GDP	Guanosine DiPhosphate
GPU	Graphical Processing Unit
GROMACS	GROningen MAchine for Chemical Simulations
GRX	Glutaredoxin
GTP	Guanosine TriPhosphate
GTPase	Enzyme hydrolyzing GTP
IF	Initiation Factor
MC	Monte Carlo
MD	Molecular Dynamics
mRNA	messenger RiboNucleic Acid
NAMD	Not just Another Molecular Dynamics program
NMR	Nuclear magnetic Resonance
NOE	Nuclear Overhauser Effect
OPLS	Optimized Potentials for Liquid Simulations
PDB	Protein Data Bank
PME	Particle Mesh Ewald
PMF	Potential of Mean Force
PTC	Peptidyl Transfer Center
RF	Release Factor
RNA	RiboNucleic Acid
rRNA	ribosomal RiboNucleic Acid
S^2	Generalized Order Parameter
tRNA	transfer RiboNucleic Acid
UTR	UnTranslated Region
WHAM	Weighted Histogram Analysis Method

1 INTRODUCTION

"Somewhere, something incredible is waiting to be known."
- Carl Sagan (1934-1996), American scientist.

Most tasks in a living cell are performed by RNA and proteins. Since the function of these molecules is closely related to their structure, the knowledge of their shape and dynamical behavior is essential to the understanding of biology. While some RNA and proteins function by themselves, many form large macromolecular complexes resulting in an incredible large diversity of shapes and sizes of these assemblies. Similar to individual RNA molecules and proteins, the complexes are not static in their functional state. Instead, they undergo functionally important, conformational changes, often triggered by ligand binding, in order to perform their specific biological activity. Examples of systems with a close connection between structure and function are the RNA-protein complex of the ribosome and the regulatory RNA-riboswitches. The science in which structure, dynamics and function is studied is called structural biology and it is the main theme of this thesis

X-ray crystallography and Nuclear Magnetic Resonance (NMR) are the two most important experimental methods used to determine the three-dimensional (3D) structure of proteins at atomic resolution. X-ray crystallography has been an extremely useful tool for determining the structure of complex macromolecules. But to form crystals, the system needs to be in a single conformational state and X-ray crystallography consequently only provides a static picture, making information on dynamics or the structure of flexible regions hard to capture. The other method, NMR, has the advantage of studying molecules in solution and of being able to capture the dynamics, thus allowing the system to be in all its functional states. However, this technique is restricted to studying relatively small systems with high solubility in water.

Computer simulations can overcome many of the limitations faced by the experimental techniques and have become a very important tool to compliment experiments in structural biology. Molecular Dynamics (MD) simulations provide a detailed view of the structure and dynamics of biomolecules on an atomic level. While limited in size and time of the studied systems, the exponential growth in computer power constantly expands the size of systems and length of processes that are achievable to simulate.

The first part of this thesis introduces the biological systems and processes studied in the four papers representing the original work of this thesis. The process in which the information in DNA is transformed into functional RNA and proteins is briefly introduced and the riboswitch and ribosome is described in more detail. The second part of the thesis discusses the methods used during my work. The basic equations and algorithms behind an MD simulation are presented together with some important choices that have to be made before starting a simulation. The main limitations of the methods and the connection to experimental techniques are also covered briefly. In the last part, the main results are summarized followed by the complete papers and manuscripts.

2 BIOLOGICAL CONTEXT

"However many ways there may be of being alive, it is certain that there are vastly more ways of being dead."
- Richard Dawkins (1941-), English biologist.

Until the structure of DNA was discovered in 1953,[1-3] the way cells store all genetic information, needed to make up all living organisms, in the four letter code of DNA was unknown. The progress in knowledge has since then been astounding. After fifty years the complete genome for many organisms, including humans, has been sequenced. Although immensely more complex than anything humanity has ever assembled, all data needed create life forms such as humans can be fit on a single DVD disc. While this does not make biology easy to comprehend or study, it does set a clear limit on the amount of information needed to create life and it comforts us with that biology is at least not infinitely complex.

2.1 THE CENTRAL DOGMA

To create life from the genetic information stored in DNA, a polymer of four nucleotides. DNA is *transcribed* into RNA, a polymer similar to DNA but with an added hydroxyl group on the sugar and the uracil nucleotide instead of thymine. The RNA is in turn *translated* into amino acids that form functional proteins. Proteins are the main building block of living organisms and constitute most structural and functional parts of cells. This chain of events is central to all known life and has been named the central dogma of molecular biology (Figure 1).

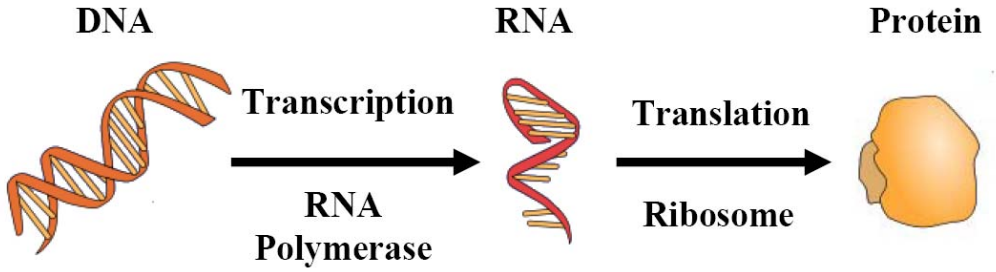

Figure 1. The Central dogma of molecular biology. Genetic information stored in DNA is transcribed into RNA by RNA Polymerase. The RNA is then translated into proteins by the ribosome.

During the first phase, transcription, a complimentary RNA chain is produced from the DNA template. The transcription process is carried out with the help of enzymes called RNA polymerase. Bacteria contain a single type of RNA polymerase and the RNA molecule is produced in a relatively straight forward process. It is *initiated* when RNA polymerase reads a promoter region of DNA, RNA polymerase then synthesizes the RNA during the *chain elongation* stage, *terminates* the transcription when reaching a terminator region and finally releases both the DNA template and the completed RNA molecule. In eucaryotic cells, the process of transcription is much more complex with three RNA polymerases, polymerase I, II, and III and it involves post transcriptional modification of mRNA (splicing) before leaving the nucleus.

Protein coding mRNA is not the only RNA product of DNA transcription. The product can also be RNA that folds into functional (parts of) molecules needed for the upcoming protein synthesis such as the ribosome (rRNA), transfer-RNA (tRNA) and the regulatory riboswitches present in procaryotes. These molecules have functions similar to those of proteins and are believed to be relics from a time before the evolution of amino-acids and proteins, the so called RNA-world.

In the *translation* process, the protein coding mRNA is picked up by the ribosome in the cytoplasm where it is translated into proteins via the genetic code (Figure 4). The amino acids used in the protein synthesis are first attached to a tRNA molecule, which then waits for its corresponding mRNA code to show up in the ribosome. When a match occurs, the tRNA binds to the ribosome and releases its amino acid. In this way, the entire sequence of nucleotides in the mRNA read according to the genetic code while an amino acids is transferred from the tRNA to the nascent peptide chain for each accepted codon. The protein then folds into its functional form on the way out from the ribosome.

The biological processes studied in this thesis are the decoding of mRNA by the tRNA in the ribosome (Paper I) and the folding of the functional form of a riboswitch (Paper II). The structure and function of these two biological systems will therefore be described in more detail in the following pages. The studied systems both stem from bacteria and the focus of the discussion will be on prokaryotic systems and mechanisms. There are some key differences of these two systems in eukaryotic and prokaryotic organisms making them important targets for antibiotic agents and their potential role in the fight against antibiotic resistance among bacteria will also be highlighted.

2.2 RIBOSWITCHES

While most gene regulation in eukaryotic organism is performed by proteins, genetic regulation by RNA is widespread in bacteria. One common form of riboregulation in bacteria is the use of RNA sequences encoded within mRNA that directly affect the expression of genes encoded in the full transcript (called cis-acting elements because they act on the same molecule they are coded in). These regulatory elements are known as *riboswitches*. They are defined as regions of mRNA that bind metabolites or metal ions as ligands and regulate mRNA expression by forming alternative structures that promotes or inhibits the translation or transcription of the associated gene in response to this ligand binding.[4, 5] To date, it has been discovered that a variety of ligands are sensed by riboswitches; including magnesium ions, nucleic acid precursors, enzyme cofactors, and amino acid residues. Riboswitches are most often located in the 5'

untranslated region (5'-UTR), a stretch of RNA that precedes the start of protein coding mRNA regions, of bacterial mRNA (Figure 2).

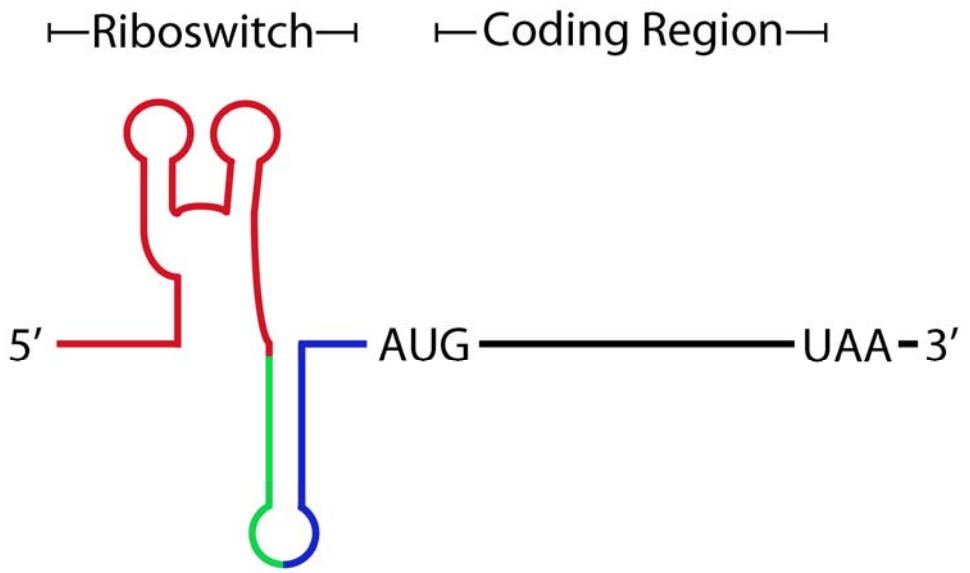

Figure 2. Schematic representation of an mRNA transcript including a riboswitch. The riboswitch is situated upstream of the coding region of mRNA with its aptamer domain in red, expression platform in blue and the common switching sequence in green. AUG and UAA are start and stop codons respectively.

2.2.1 Basic Structure and Function

The function of riboswitches is closely tied to the ability of RNA to form a diversity of structures. Unlike DNA, RNA with its additional hydroxyl group on the sugar is able to form a wider range of structures than the double helix of DNA. For example, a single strand of RNA can fold back on itself to form a hairpin, which is composed of a helix capped by a loop. In large RNAs, secondary structural elements such as helices and hairpins pack together into a specific pattern, to form intricate tertiary interactions such as the so called kissing loop interactions and base triplets and tetrads. These processes depend on positive ions to counter act the close packing of the negative charge of the phosphate groups of the RNA backbone (see section 2.4 for more details).

Riboswitches are composed of two domains: the aptamer domain and the expression platform.[6] The aptamer domain (red in Figure 2) acts as a receptor that recognizes and binds the ligand. The expression platform (blue in Figure 2) acts directly on gene expression through its ability to toggle between two different secondary structures in response to ligand binding in the aptamer domain. Common to both domains is the so called the switching sequence (green in Figure 2). It is its binding, to either the aptamer domain or the expression platform that ultimately decides the expression outcome of the mRNA. Specifically, if metabolite binding to the riboswitch stabilizes the interactions of the switching sequence to the aptamer domain, the expression platform undergoes a conformation change, affecting its expression.

The actual regulation is different among riboswitches and can be performed at both the transcription and translation stage of gene expression.[7] Riboswitches can signal transcriptional repression with a switching sequence that directs formation of a *transcriptional terminator*, a short stem-loop structure that signals RNA polymerase to abort transcription. The other main type of regulation targets the translation. These

riboswitches regulate the translational initiation by utilizing a switching sequence that expose or hide a ribosomal binding site called the Shine-Dalgarno sequence. (see section 2.3.3.1 for more details).

2.2.2 Purine Riboswitches

Riboswitches are usually organized into families and classes according to two features: the type of ligand they bind, and their secondary structure.[8] A family of riboswitches is typically a group of RNAs related by the ligands they recognize. Within a family, there may be distinct classes of riboswitches, each class distinguished by a common sequence pattern that usually defines the ligand-binding pocket, as well as features required for folding the RNA into a three-dimensional shape. An exception to this classification scheme is the purine riboswitch family, to which the *add* A-riboswitch (studied in Paper II and used as a model system in Paper III), belongs. This group of RNA shares a common secondary structure but can recognize multiple distinct ligands.[9]

The purine family includes riboswitches that bind to adenine, guanine, and 2'-deoxyguanosine. Because its members recognize multiple ligands, this riboswitch family serves as an excellent model for understanding the mechanisms of ligand recognition. The global architecture of the RNA in a purine riboswitch is defined by the organization of the three conserved helices joined by a central junction region (Figure 3). Two of the RNA helices (P2 and P3) are situated on top of the third (P1), forming a Y-shaped structure and the terminal loops of P2 and P3 are held together by tertiary kissing loop interactions including a base quadroplex.[10]

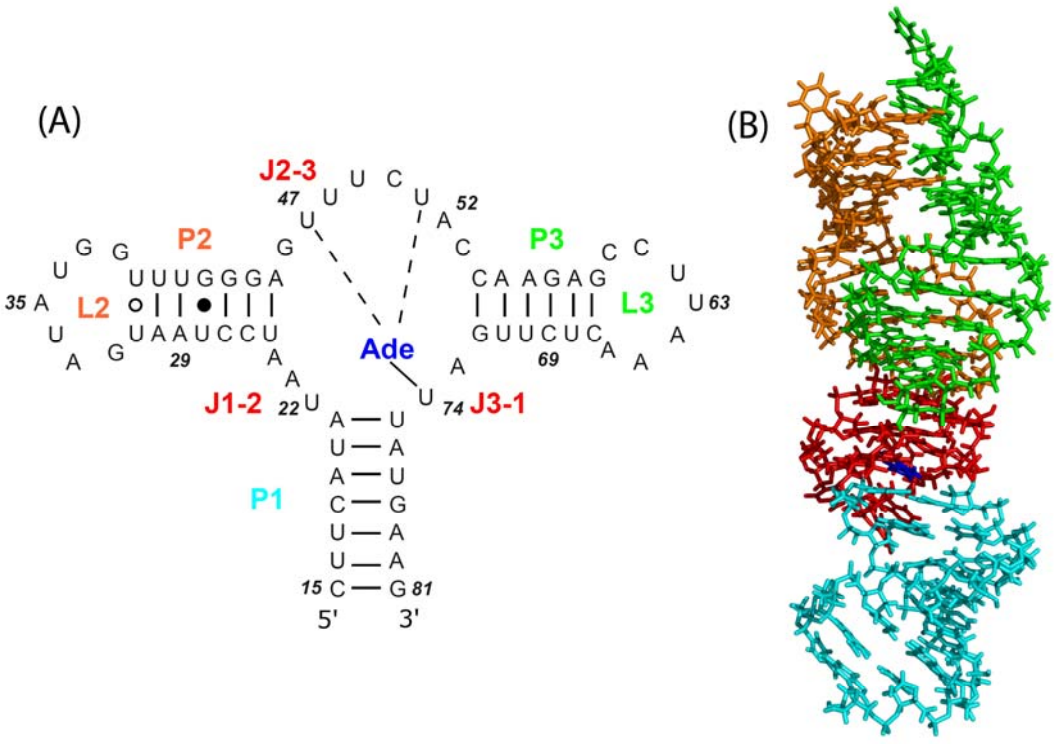

Figure 3. Aptamer domain of the *add* A-riboswitch. (A) 2-D secondary structure including stem base pair hydrogen bonds and residue numbering. Bonds to the ligand are shown as full lines for Watson-Crick interactions and dashed lines for non Watson-Crick interactions. (B) 3-D structure colored according to labels in A from the X-ray structure by Serganov *et al.*[10].

2.2.2.1 Ligand Binding and Regulatory Mechanism

The ligand-binding pocket of a purine riboswitch is situated in the three-way helical junction where P1, P2, and P3 meet. This region of the RNA is defined by a series of non-canonical base interactions, such as a base triplet between residues far apart in the RNA sequence (A21-U75•C50). At the center of the junction, a pyrimidine at position 74 forms a Watson-Crick base pair with the ligand, which is bound further by other conserved residues. The identity of this pyrimidine residue (cytosine or uridine) is the basis for specificity between the guanine and adenine classes.[11] When the ligand is bound in its pocket, it is completely encapsulated by the riboswitch making it inaccessible to the aqueous environment, a feature common to many riboswitches.[6]

This encapsulation of the purine ligand is accompanied by both local and global conformational changes. On the local level, the J3-1 strand of the junction, acts as a preformed docking station for the incoming purine ligand by making Purine74 available for Watson-Crick pairing. Once the ligand is bound, nucleotides comprising J2/3 fold over the ligand and encapsulate it. The ligand binding (together with Mg^{2+} ions) is also closely associated to the overall folding of the entire riboswitch structure.[6, 12] In the *add* A-riboswitch specifically, the binding of the adenine ligand strongly promotes the kissing loop interactions between P2 and P3 and the base pairing of P1.[13]

The coupling of ligand binding to a conformational change is central to the regulatory mechanism of purine riboswitches. As J2/3 covers the ligand, it also forms additional tertiary interactions with the switching sequence at 3' end of the P1 stem (Figure 2, green). These additional ligand-induced interactions with the P1 helix stabilize its incorporation into the aptamer domain. At the same time, this prevents the P1 helix from being able to form alternative structures with the expression platform. In this way, the expression platform forms one of two "on and off structures" that interface with the expression machinery: either RNA polymerase or the ribosome.

In the *add* A-riboswitch the ligand binding stabilizes the folded aptamer domain, which leaves the expression domain in the on structure able to interact with the Shine Delgarno sequence of the ribosome and initiate translation. If the ligand is not bound after the transcription of the riboswitch, the expression platform rapidly equilibrates into a lower-energy off structure, unable to bind to the ribosome and thus repressing gene expression. This is an example of positive feedback to the presence of ligand but riboswitches that provide negative feedback (by repressing gene expression when ligand is bound) are also known.[6]

2.3 RIBOSOMAL TRANSLATION

At the heart of translation of mRNA into the peptide chains of proteins lies the genetic code (Figure 4). This code is the link between the ribonucleotides of mRNA and the amino acid sequence of the final proteins. It is nearly completely universal with the same "dictionary" being used in close to all known organism across all domains, including viruses. It is written in a linear form consisting of 61 triplets of nucleotides called codons each specifying one amino acid. Among the codons are also the "start" and "stop" triplets that specify where the translation should be initiated and terminated. The code is unambiguous but degenerate, meaning that each triplet specifies only one single amino acid but that a given amino acid can be specified by more than one triplet codon. In fact, almost all amino acids are specified by two, three, four or even six

different codons. Only tryptophan and the methionine start codon are encoded by single codons (see color coding in Figure 4)

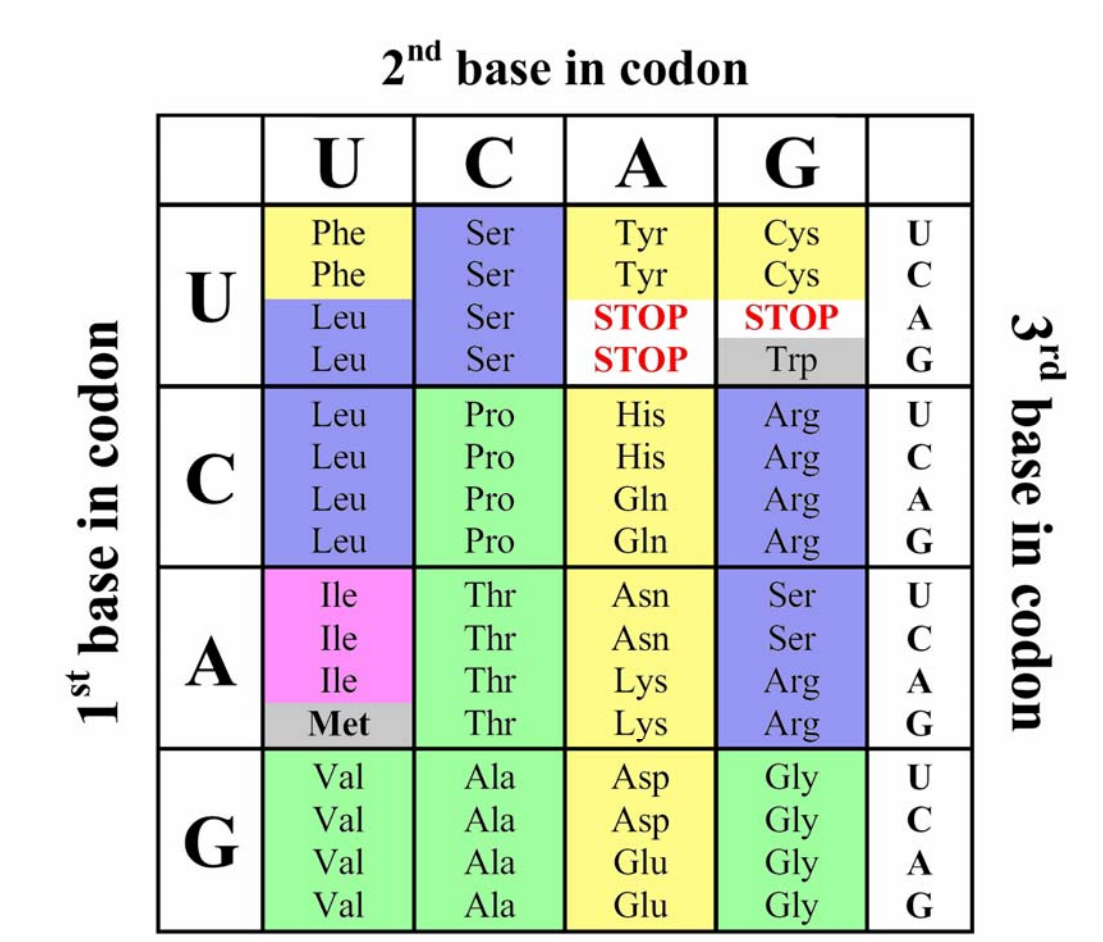

Figure 4. The genetic code. The codons (three-base code words) are translated into the 20 amino acids denoted by in their three letter codes. The codons are colored according to their degeneracy, grey for unique, yellow for twofold, magenta for threefold, green for fourfold, and blue for six fold.

When studying the degeneracy of the genetic code a pattern soon becomes apparent. Most often in a set of codons specifying the same amino acid, the first and second letters are the same with only the third one differing. When observing this Francis Crick formulated the Wobble hypothesis in 1966.[2] The hypothesis states that the third letter of the codon is read less accurately than the two first, allowing mismatches in the third so called wobble position. This relaxed base-pairing requirement allows the anti-codon of single tRNA to pair with more than one codon of mRNA. This means that a U at the first position (the 5'-end) of the tRNA anticodon may pair with A or G at the third position (the 3'-end) of the mRNA codon and G may likewise pair with U or C at the wobble position. In addition to this near-cognate wobbling, later research has found that certain tRNAs are able to read, not only near cognate bases, but all four bases at the wobble position.[14] This ability requires post transcriptional modifications of the tRNA anti-codon and is described in more detail in section 2.3.1 below. When these wobble rules are applied a theoretical minimum of about 30 different tRNA species is necessary to account for the 61 triplets coding for an amino acid. The evolutionary reason for the wobble capability is considered to be an economy measure. Current estimates are that 30-40 tRNA species are present in bacteria and up to 50 are present in animal and plant cells.

2.3.1 tRNA

The decoding process where the nucleotide sequence of mRNA is translated into the amino acid sequence of the protein is performed by tRNA molecules, carrying one amino acid each, on the ribosome. When a codon in the mRNA calls for a particular amino acid, the tRNA carrying that amino acid will recognize the specific codon and deliver its amino acid to the growing polypeptide chain. tRNAs consist of a single stranded RNA molecule, about 75 nucleotides long, folded into a L-shaped hairpin arrangement (Figure 5). The structure (first revealed over 35 years ago[15, 16]) contains four stems that are stabilized by Watson-Crick base pairing, three of which have loops (inset Figure 5). The middle hairpin (blue) harbors the anti-codon (grey), the nucleotide triplet that base-pair to the corresponding codon in mRNA. The other two hairpins (red and green) make up the L-shaped center of the tRNA molecule. The fourth stem (purple) is composed of the 3'- and 5'-ends of the chain and this unlooped stem is referred to as the acceptor or aminoacyl (aa) stem. All tRNAs have a conserved CCA sequence at their 3'-end that is not base-paired (yellow). The free 2'-hydroxyl or 3'-hydroxyl of the terminal adenosine is covalently attached to the amino acid, forming an aa-tRNA.

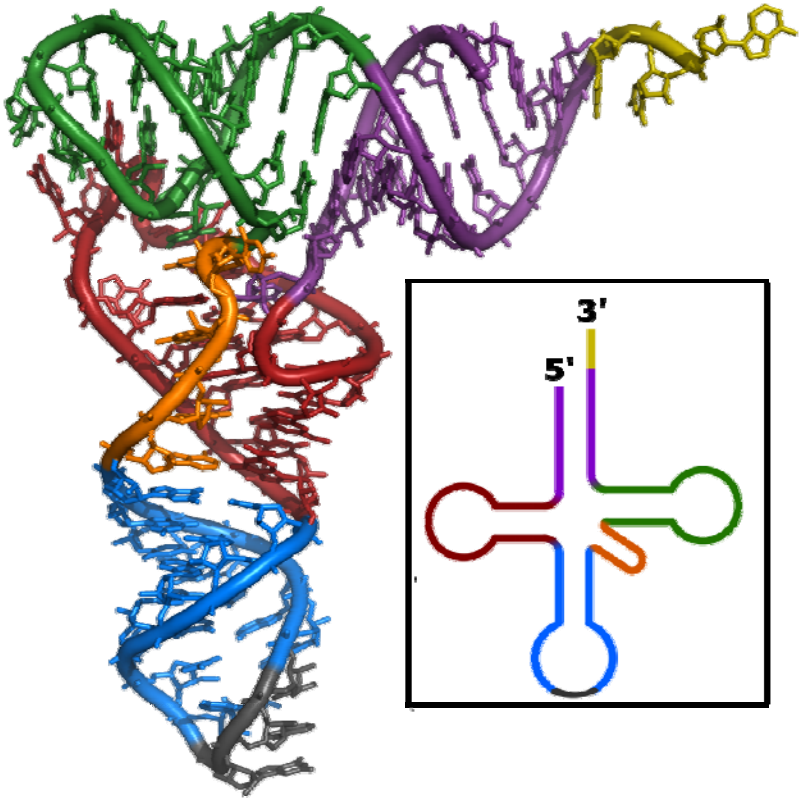

Figure 5. Structure of $tRNA^{PHE}$ from yeast. Inset shows 2-D secondary structure. Purple: acceptor stem, red: D-loop, blue: anticodon loop, orange: variable loop, green: TPsiC-loop, yellow: CCA-3' of the acceptor stem and grey: anticodon. X-ray structure from PDB: 1ehz.[17] Adapted from the Wikimedia Commons file: "TRNA-Phe yeast 1ehz.png", http://commons.wikimedia.org/wiki/File:TRNA-Phe_yeast_1ehz.png.

tRNA molecules have one of the highest frequencies of post transcriptional base modifications of all known biomolecules. Base modifications are extra atomic groups that are attached to any of the standard nucleotides after the transcription from DNA to RNA. Although modification of nucleosides comes at a considerable genetic and

energetic cost (most modifications demand their own enzyme), more than 70 distinct modifications have been identified on the known tRNAs. The modifications can be situated in all domains of the tRNA and are conserved in many organisms.[18, 19] However, the most universally conserved modified bases are situated a close to the anticodon, especially in position 34 and the purine 37 on the 3' side of the anticodon. The modifications in these places vary greatly in size, ranging from simple methyl groups to complex structures with multiple functional groups, but are all with little doubt involved in the recognition of the mRNA codon[20]. It is these modifications that let certain tRNA accept all four mRNA bases at the wobble position in the codon. The mechanism of this expanded decoding is studied in detail in Paper I in this thesis.

2.3.2 The Ribosome - Structure

The ribosome is the molecular machine responsible for all protein synthesis in the cell. It is a large particle, consisting of two subunits made up from a mixture of ribosomal RNA (rRNA) and proteins. The size of the two subunits as well as some of the steps in the translation process varies between eukaryotic and prokaryotic organisms. In this thesis only the bacterial ribosome has been studied and therefore only the prokaryotic ribosomal structure and mechanisms will be covered here.

The structure in atomic resolution of the ribosome remained unknown for long but was solved during the last decade by Thomas A. Steitz,[21] Venkatraman Ramakrishnan[22] and Ada E. Yonath[23] and was awarded the Nobel Prize in chemistry 2009. In bacterial ribosomes the two subunits are called the 30S (small) and the 50S (large) subunit. These ribosomal subunits have irregular shape, but together they form a fairly globular complex with a cleft between them through which the mRNA and tRNA molecules binds and transits (Figure 6). The simplest bacterial and archeal ribosomes contain three rRNA chains, forming the core of the ribosomal subunits and comprising about 66 % of the total mass, and more than 50 proteins representing the remaining 33 %. The 30S subunit has one large rRNA molecule (16S RNA) while the 50S subunit has two rRNA molecules, one large (23S RNA) and one smaller (5S RNA). In both subunits it is the large rRNA molecule that provides the binding sites for the ribosomal proteins. Most of the ribosomal proteins are located at the solvent exposed surface of the subunits.

There are three binding sites for tRNA molecules on the ribosome and all three are situated on the interface between the 30S and 50S subunit and are thus surrounded by both. The first binding site is the A (aminoacyl) site where the incoming aa-tRNA first binds and selection between cognate and non-cognate tRNA molecules takes place at the decoding center. If the tRNA is accepted it is transferred to the P (peptidyl) site where the carried amino acid is added to the nascent peptide chain at the peptidyl transfer center (PTC). tRNA molecules are able to concurrently contact the decoding site on the 30S subunit and the PTC on the 50S subunit because of their L-shaped structure where the anticodon is at one side and the aminoacyl tail is at the opposite side, about 75 Å apart. Finally the deacylated tRNA is moved to the E (exit) site before leaving the ribosome. The A site is located at the side of the L7/L12- stalk, the E site is at the L1-side of the ribosome and P site is situated between them at the center of the ribosome.

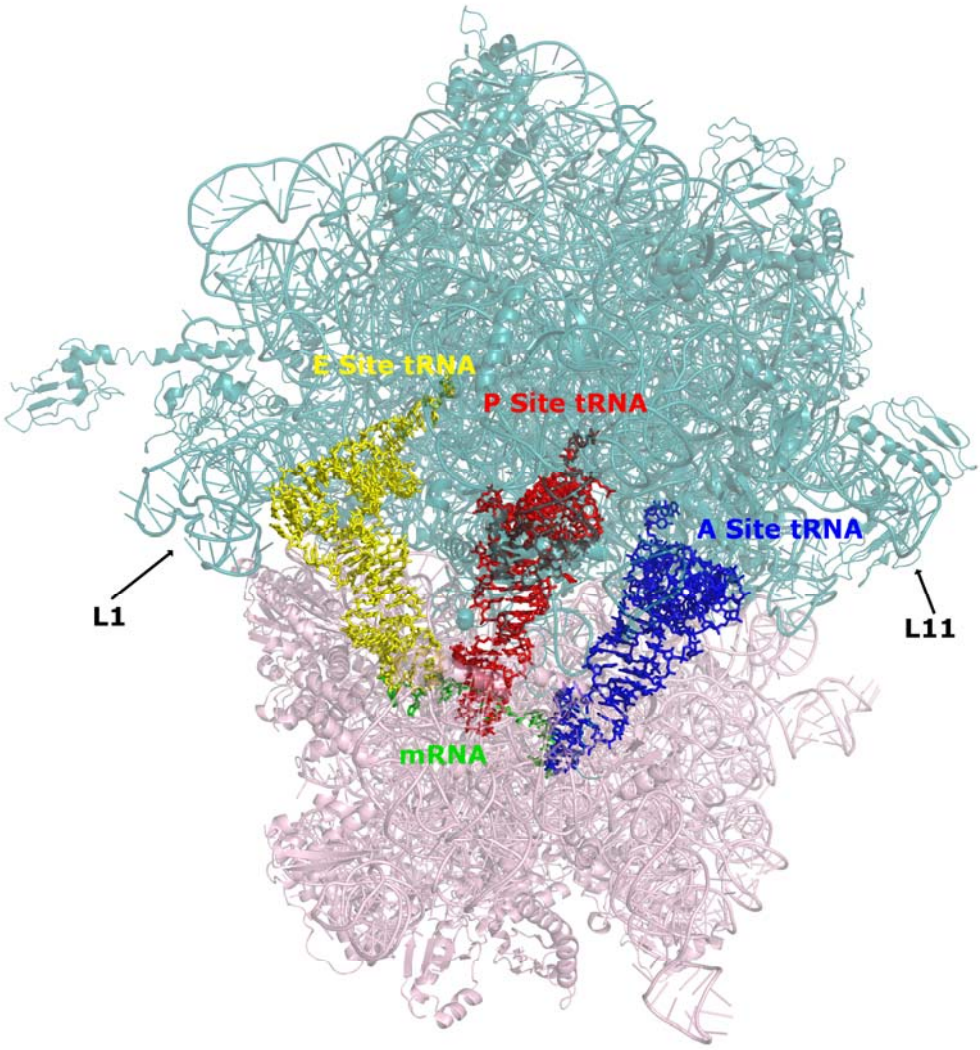

Figure 6. The prokaryotic 70S ribosome with three bound tRNAs and mRNA. The 50S subunit is colored pale blue and the 30S subunit is colored pale pink. Ribosome coordinates from he X-ray structure by Selmer *et al*..[22]

The movements of the tRNA molecule through the ribosome hint at the dynamic nature of the ribosome which undergoes large scale conformational changes as part of its function.[24] In the 30S subunit, the large-scale dynamics mainly takes place around the decoding center, which undergoes a conformational change from an ”open” to a ”closed” form after a cognate codon:anticodon interaction is registered.[25] In the 50S subunit, it is mainly the protuberances of the L1 and L11 stalk (situated on the opposite side of the ribosome from L1) that display large scale dynamics. The L1 stalk is involved in the release of the deacylated tRNA from the E site. The biological functions of L11 is not completely known but is thought to be associated with the hydrolysis of the energy bearer GTP by its hydrolization enzyme GTPase.[26]

2.3.3 Ribosomal Translation – The Three Steps

The translation process can be divided into three distinct steps: *initiation*; *elongation* and *termination*. The process is powered by the hydrolisation of GTP to GDP and is executed with the help of several cofactors, of which all will not be mentioned in the short description below. The translation process has been reviewed in more detail by Ramakrishnan[27] and Steitz[28].

2.3.3.1 Initiation

Protein synthesis is initiated when an mRNA molecule binds to the free 30S subunit (without the 50s unit). Correct initiation requires that the start codon (generally AUG) is positioned in the ribosomal P site. In bacteria, the 30S subunit identifies the start codon by forming interactions between its 16S RNA molecule and the so called Shine-Dalgarno sequence located upstream of the AUG start codon in the mRNA chain.[29] After the correct placement of the start codon in the P site of the 30S subunit, the "start" $tRNA^{Met}$ bind to the P site with the help of initiation factors (IF). With the help of energy from GTP and other initiation factors the 50S ribosomal subunit binds to 30S subunit. After release of all IFs, the result is the 70S ribosome with $tRNA^{Met}$ in the P site that is ready for elongation.

2.3.3.2 Elongation

The elongation step of the translation can be divided into three separate sub-steps:

1. Elongation starts with the binding of an aa-tRNA in complex with an elongation factor (EF-Tu) and GTP to the A site of the ribosome. Here the mRNA codon is matched with the tRNA anti-codon by the formation of three base pairs (see below for more details on the decoding process). If accepted, EF-Tu will hydrolyze its bound GTP and undergo a conformational change. The EFTu ·GDP complex has a low affinity for the aa-tRNA and the ribosome resulting in its dissociation from the aa-tRNA. This will free the aminoacyl end of tRNA and allow it to move into the P site and reach the PTC on the 50S subunit, where the peptide is synthesized.
2. The formation of the peptide bond between the amino acid brought by the tRNA and the nascent peptide chain takes place at the PTC with the help of peptidyl transferase, which is an enzymatic function of rRNA that catalyzes the formation of peptide bonds.
3. The final step of elongation is the translocation of tRNA. The now deacylated tRNA leaves the P site and exits via the E site. Meanwhile, a new aa-tRNA that is accommodated in the A site moves into the P-site together with its corresponding mRNA codon, exposing the downstream codon at the A-site. This process is catalyzed by elongation factor G (EF-G).

This three step process is repeated until a stop codon in the mRNA strand is reached at the A site.

The *decoding* of mRNA and selection of the correct tRNA for participation in the addition of a new amino acid to the growing polypeptide chain is entirely based on the base pairing between the codon on mRNA and the anticodon on tRNA. However, the energy difference in base pairing of cognate tRNA, and that of near-cognate tRNA (generally consisting of only a single mismatch) is too small to account for the accuracy of selection, which has been found to have an error rate of 10^{-3} to 10^{-4}.[30]

How can this high accuracy achieved? It turns out that the ribosome has several ways of augmenting the differences between base pairs to increase the accuracy.[25] Firstly, the decoding site of the 30S subunits works much like an enzyme by recognizing and stabilizing the geometry of correct codon-anticodon base pairing and sterically discriminates against mismatches. This is achieved by the formation of hydrogen bonds between the ribosomal residues A1492, A1493 and G530 and the tRNA and mRNA when Watson-Crick base pairs between the anti-codon and codon are present (see Figure 2 in Paper I of this thesis for more details). Secondly, the ribosome asserts a

"kinetic proofreading" by splitting the selection step into an initial selection and a proofreading step. The two steps are separated by an irreversible reaction, such as the hydrolysis of GTP by EF-Tu preventing once discarded tRNA to be evaluated again. In this scheme, an incorrect tRNA has two chances to dissociate, once during initial selection and once after GTP hydrolysis and release of EF-Tu. Theoretically, this process results in a selectivity that is the product of the selectivity at each step. A third way of increasing (or decreasing) the selectivity is by the addition of modified bases discussed in section 2.3.1 and studied in Paper I.

2.3.3.3 Termination

Protein synthesis is terminated when one of the three stop codons UAA, UAG or UGA is exposed at the A site. Three release factors (RF) work in concert to terminate the translation process. Two of them recognize the stop codons and then hydrolyze the polypeptide chain from the last tRNA in the P site. The third release factor is a GTPase, and its function is to remove two other release factors from the ribosome.[31] After the termination step and release of the produced peptide chain, the ribosome is prepared for a new round of translation by the separation of the two subunits and dissociation of the mRNA and the deacylated tRNA. This process is promoted by the ribosome recycling factor together with EF-G.

2.4 IONS AND RNA

Each nucleic acid residue in DNA and RNA carries a -1 charge at the backbone phosphate group in aqueous solution and physiological conditions. In DNA, with its ordered double helical structure, the phosphate groups are kept far apart and the repulsion of the negative charges does not create problems for the effective folding to DNAs helical structure. RNA with its additional hydroxyl group, folds in much more intricate ways, often packing the negative phosphate groups close together in space. To overcome the electrostatic repulsion hindering the folding of RNA to its functional forms, it has been found that positive metal ions are integrated in the structure and dynamics of RNA.[32, 33]

2.4.1 Types of Ions in RNA

Of the ions found in the cell, Na^+, K^+, Mg^{2+} and Ca^{2+} are present at the highest concentration. While they all have been showed to be involved with RNA structure, the monovalent cations, sodium (Na^+) and potassium (K^+), and the divalent cations, magnesium (Mg^{2+}) and calcium (Ca^{2+}) influence the structure and folding of RNA more and less effectively. Mainly two factors determine the effectiveness of an ion to stabilize RNA structures: charge and size. The divalent Mg^{2+} has been found to have a large stabilizing effect on RNA structure even in the presence of a 100-fold excess of monovalent cations.[34, 35] The ionic size is relevant to ion-RNA interactions in two ways: aside from the obvious steric problem of fitting a large ion into an RNA binding site, the radius also determines the strength of the electrostatic field. The smaller the radius of an ion is, the greater the charge density and the stronger its interactions with water and RNA atoms.

With these factors is mind, Mg^{2+} stands out as being the most relevant for RNA stability and folding.[36, 37] Mg^{2+} ions are characterized by a high charge density and strong electrostatic field due to the +2 charge and small radius (~0.65 Å). This gives the

ion the ability to transfer a large amount of charge into sterically confined spaces and effectively mitigate the negative charge. The high charge density also results in extremely strong interactions with water molecules. Mg^{2+} forms a complex of six water molecules ($[Mg(H_2O)_6]^{+2}$) packed in a octahedral arrangement and surrounded by a second solvation shell of twelve less strongly bound water molecules. In comparison; K^+, the most common metal in physiological conditions, has with its 1.3 Å radius and +1 charge considerably weaker interactions, both with RNA and surrounding water molecules. Another reason for the more efficient stabilization of RNA by Mg^{2+} (and other divalent ions) is a smaller entropic cost for the association of ions to an RNA molecule. Due to their higher charge density, fewer ion have to by restricted to the surface of the RNA molecule resulting in a lower entropy cost.[38]

2.4.2 Types of Ion Binding to RNA

Metal ions can bind to RNA in two ways: directly by replacing 1-3 of the waters in the first solvation shell with RNA atoms (inner sphere contact) or indirectly with one of the first shell water molecules bridging between the ion and the RNA acceptor atom (outer sphere contact) (Figure 7). The type of binding has significant effects on the properties of the interaction, especially for Mg^{2+} which binds several orders of magnitudes stronger than Na^+, K^+ and Ca^{2+}. While each direct interaction is stronger than an indirect, the view that the main stabilizing contribution comes from direct binding has received less support lately and it is now believed that indirect contacts are responsible for most of the stabilizing effect that Mg^{2+} ions have on RNA structures.[39, 40] The importance of indirect interactions to the overall stability of RNA can be attributed to their significantly greater number. The relative scarcity of direct interactions can in turn be explained by the high free energy cost associated with the partial dehydration of the $[Mg(H_2O)_6]^{+2}$ complex[32] and RNA systems[41] that is required for direct binding. The cost of preparing the ion for direct binding is larger than the gain from the stronger interactions.

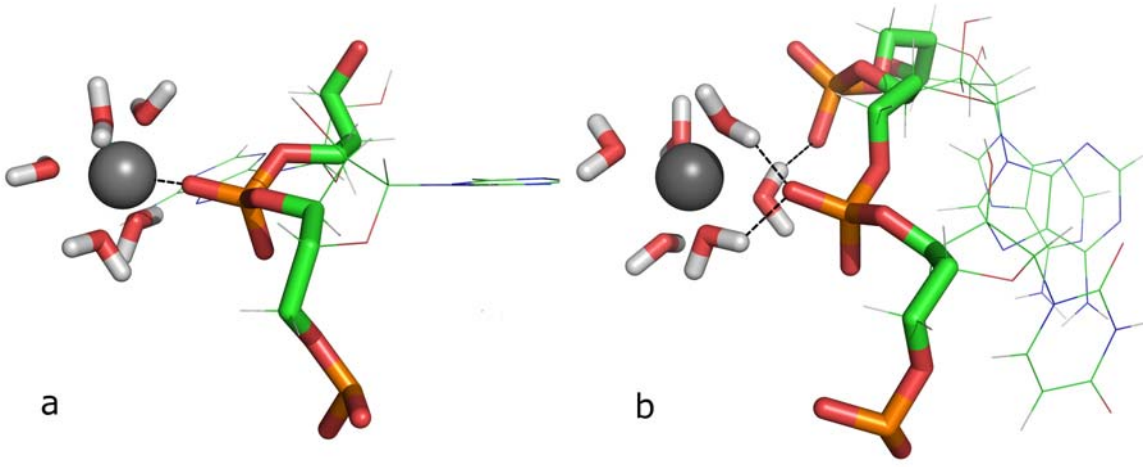

Figure 7. Direct (a) and indirect (b) binding of an Mg^{2+} ion to an RNA phosphate group. Magnesium atoms are in grey, phosphor in orange, oxygen in red, carbon in green and hydrogen in white.

2.4.3 Ions in Riboswitches and the Ribosome

The ribosome, the largest natural RNA structure, unsurprisingly requires many cations to form its folded, functional form. High resolution X-ray structures of the ribosome

have revealed 60 directly bound Mg^{2+} ions and 88 monovalent ions in the 50S subunit alone[42] but many more of the "hard to detect", indirectly bound ions are expected. Although the functional centers of the ribosome are densely packed with divalent metal ions, the ions have not been found to directly participate in the ribosomal function. The main function of the ribosomal ions is however believed to be to promote the folding and stability of the ribosome's functional architecture.[33]

Recent investigations have also elucidated the role of Mg^{2+} ions in the folding of riboswitches. It has been shown that a high Mg^{2+} concentration can compensate for the absence of a bound ligand and promote the structure formed with the correct ligand bound.[43, 44] Two studies of adenine riboswitches have also found that the presence of Mg^{2+} strongly stabilizes the folded state and that Mg^{2+} enhances the stability of the kissing loop structure and the binding pocket, which leads to an increased affinity for the adenine ligand.[31, 40]

2.5 ANTIBIOTICS AND RNA

With the surge in antibiotic resistance seen in clinics all over the world, the need for novel drug targets in bacteria becomes more and more acute. While RNA was believed to lack the chemical diversity of proteins and thus not be good targets for drugs, we now know that it has many attractive characteristics. Like proteins, RNA can fold into intricate 3-D structures with pockets and cavities that have the potential to bind ligands specifically. RNAs, such as the ribswitches, undergo extensive structural rearrangement upon associating with small molecules, and therefore exist in multiple distinct conformations that can be targeted. In addition, many cellular RNAs are subjected to extensive post-transcriptional modifications, which increases their chemical versatility further.[45] The characteristics together with key differences between prokaryotic and eukaryotic cells and RNA's close connection with gene expression make RNA molecules enticing antibiotics targets. The majority of antibiotics in clinical use target the rRNA of the ribosome,[46] but novel and promising drug candidates targeting riboswitches are currently in development.[47]

2.5.1 Antibiotics Targeting the Ribosome

Antibiotics that target the ribosome almost exclusively bind to one of three functional sites: the decoding (A-site) on the 30S, the PTC (P- site) on the 50S, and the peptide exit tunnel on the 50S subunit.[48] Antibiotics that bind the A-site, such as aminoglycosides, interfere mainly with codon recognition. At the decoding center, the binding of these aminoglycosides induces a conformational change of ribosomal residues A1492, A1493 and G530 (see section 2.3.3.2) that mimics the structure corresponding to Watson-Crick base pairing between the codon and anti-codon. This results in the acceptance of non-cognate tRNAs and malfunctioning proteins. Drugs targeting the P site, like oxazolidinones, inhibit the actual protein synthesis by either hindering tRNA substrates from binding or disrupting peptide bond formation by binding to the PTC itself. Finally, antibiotics that target the peptide exit tunnel function by blocking the growth of the nascent amino acid chain at the peptide exit tunnel. This prevents the elongation process and results in incomplete proteins.

2.5.2 Antibiotics Targeting Riboswitches

Riboswitches are attractive targets for new drugs for two additional reasons compared to other RNA. First, riboswitches evolved to recognize small molecules. Although numerous RNA domains have been experimentally targeted for drug discovery, small molecule binding to these RNAs is often circumstantial, not related to their physiological function.[49] Ligands targeting such RNAs can exhibit poor selectivity.[50] Second, almost all known riboswitches occur predominantly in bacteria, not eukaryotes. If eukaryotes do employ riboswitches, it is likely that these will be distinct from those of bacteria, minimizing cross-reactivity of bacterial riboswitch targets.

Several antibiotic compounds (e.g. roseoflavin and pyrithiamine) with previously unknown mechanism have recently serendipitously been found to target riboswitches.[49] These molecules are highly similar to the cognate ligands of the FMN and TPP riboswitches respectively and inhibit gene expression by inducing the off state of the riboswitch. Active development of drugs targeting the guanine (Purine) riboswitch is currently also being performed.[51, 52] Guanine analogues differing in the C2 and C6 positions have been shown to bind and inhibit growth significantly, also in bacteria known for antibiotics resistance. Compound stability and bioavailability are however problems yet to be solved.

3 METHODOLOGY

"It is unworthy of excellent men to lose hours like slaves in the labor of calculation, which could be relegated to anyone else if machines were used."
- Gottfried Wilhelm von Leibnitz (1646-1716),
German philosopher and mathematician.

Molecular modeling is the science of studying molecular structure and function through structure building and computation. The structures usually consist of atomic coordinates of large biomolecules, metals and small, bound organic molecules obtained from experimental methods like X-ray crystallography or NMR-spectroscopy. The computational methods stretch from detailed *ab initio* and semi-empirical quantum mechanics, through empirical molecular dynamics and Monte Carlo simulations to the large scale, less detailed coarse-graining and homology modeling.

The questions that can be addressed by computational approaches today are as complex as the biological systems themselves. Possible inquires range from understanding the equilibrium structure of a small biomolecule, the energetics of hydrogen-bond formation in proteins and nucleic acids, binding affinities of ligands/drugs to their target to the complex kinetics of macromolecular folding and functioning of supramolecular aggregates. Molecular modeling provides a way to systematically explore structural/dynamical/thermodynamic patterns, and test and develop hypotheses to help understand and make practical use of the structure, flexibility and function of biomolecules.

Simulations of systems of the size of macromolecules like proteins and folded nucleic acid molecules are generally carried out with methods based on classical physics like Molecular Dynamics (MD) and Monte Carlo (MC). While structural configurations are generated randomly in MC simulations, MD simulation generates a trajectory of structures connected in time. The time scales studied with MD simulations typically range from picoseconds (10^{-12} s) to microseconds (10^{-6} s) but simulations on the millisecond (10^{-3} s) scale have been reported recently.[53] In this thesis only MD simulations have been used and they will be described in more detail below.

3.1 MOLECULAR DYNAMICS SIMULATIONS

Many software packages that perform MD simulations exists, but the for most widely spread include: CHARMM[35, 54], AMBER[55], GROMACS[56] and NAMD[57]. The work presented in this thesis has been carried out exclusively using the CHARMM package. While these packages are subjected to constant development, the proper and

effective usage of them is far from generally applicable, automated procedures and their application relies on user expertise as well as biological intuition.

Many successful MD studies have been published since the first simulation of a protein was reported in 1977[58] thanks to the development of more effective algorithms and the explosive increase in computer power. Numerous applications have surfaced, including the refinement of low-resolution experimental structures[59], interpretation of various experimental data such as single-molecule force-extension curves[39] or NMR spin-relaxation in proteins[60-62] and improvement of structure-based function predictions, for example, by predicting calcium-binding sites[63]. MD simulations are also an important tool in calculating free energy differences and have been able to estimate the importance of quantum effects in lowering free-energy barriers of biomolecular reactions[64] and to propose free-energy pathways and associated mechanisms.[65, 66] MD simulations have also been an important tool in developing an understanding of enzyme binding-pockets[67, 68] and the interactions of drug compounds[69].

3.1.1 Force Fields

At the core of a classical MD-simulation lies the force field, which is responsible for the description of the potential energy of a system of particles. The force field consists of two main components: the potential energy function and the parameters used in the function. The parameters are derived empirically to reproduce available experimental data or detailed quantum mechanics calculations on small model compounds. The successful parameterization of biological systems is heavily dependent on highly detailed experimental data on structure, dynamics and energetic. This have until recently limited the development of accurate force fields, especially for nucleic acids, for which less data is available compared to proteins.

The potential energy function used in CHARMM consists of bonded energy terms (bonds, angles and torsions) to the non-bonded terms (van der Waals and electrostatic):

$$V_{TOTAL} = \sum_{BONDS} k_B \left(b - b_0\right)^2 + \sum_{ANGLES} k_\theta \left(\theta - \theta_0\right)^2 + \sum_{DIHEDRALS} k_\chi \left[1 + \cos\left(n\chi - \sigma\right)\right] + \\ + \sum_{\substack{NON-BONDED \\ PAIRS,ij}} \left(\frac{q_i q_j}{4\pi\varepsilon_0 \varepsilon_1 r_{ij}} + \varepsilon_{ij} \left[\left(\frac{R_{\min,ij}}{r_{ij}} \right)^{12} - 2\left(\frac{R_{\min,ij}}{r_{ij}} \right)^{6} \right] \right) \quad (1)$$

The three bonded terms represent the stretching of bonds, the bending of angles and the rotation of dihedrals. k_b, k_θ and k_χ are the force constants for the bond length, bond angle and the dihedral angle respectively. The values of these variables in the current configuration are denoted b, θ and χ, while the values representing the potential energy minima are denoted with the subscript zero. The rotation of a dihedral is periodic with periodicity n and phase σ. In the non-bonded term *i, j* are all atom pair combinations within a cutoff distance, q is the charge of atoms *i or j*, ε_0 is the permittivity of vacuum and ε_1 is set to 1 for explicit solvents and r_{ij} is distance between atom *i* and *j*. ε_{ij} and $R_{\min,ij}$ are the combined atomic Lennard-Jones parameters, specifying the depth (ε_{ij}) and position ($R_{\min,ij}$) of energy minimum. ε_{ij} is obtained by the geometric mean of the two atomic parameters, $\varepsilon_{ij} = \sqrt{\varepsilon_i \varepsilon_j}$, found in the parameter set while, $R_{\min,ij}$, is

obtained by the arithmetic mean of the two atomic parameters, $R_{\min,ij} = \frac{R_{\min,i} + R_{\min,j}}{2}$.

To avoid an insurmountable number non-bonded pairs that should be evaluated at each dynamics step, a distance cutoff is introduced to limit pairs to surrounding atoms. While this cutoff provides a considerable decrease in computational cost, it also introduces a significant approximation by neglecting all non-bonded interactions outside the cutoff. Several methods that account for long-range electrostatics exist, but one of the most widely used is the particle mesh Ewald (PME) algorithm which provide an accurate but computational affordable representation of long-range electrostatics. [70, 71]

A number of empirical force fields, used for the simulation of biomolecular systems, have evolved during the recent decades. The most widely used are currently CHARMM[54, 72], AMBER[57, 73], GROMOS[74] and OPLS[75]. The form of the potential energy function, used to describe specific interactions, is very similar between these force fields but there are differences in the actual values of the specific parameters and how they are implemented in the potential energy function. The largest differences lies in the parameters describing the non-bonded interactions, i.e. the partial charges used to represent the electrostatic interactions between molecules and the Lennard-Jones terms used to represent the van der Waals interactions between atoms. But it is also important to note that each parameter exists in a delicate context together with other parameters in each of the force fields. Modifying a parameter or inserting parameters from one force filed into another might cause severe and unexpected problems in other parts of the parameter sets. The quality of MD simulations is ultimately heavily dependent on the accuracy of the empirical force fields and an improvement of the Mg^{2+} ion parameters is presented in Paper III.

3.1.2 Molecular Dynamics Calculations

In an MD simulation the movements of the simulated atoms are described by Newton's second law of motion, $F = ma$, where F is the force acting on the particle, m is its mass and a is its acceleration. Newton's equation of motion can be applied if the positions and velocities are known or can be assumed and if all forces acting on each atom as a function of the atom position can be calculated. By knowing the 3-dimensional coordinates of the system, applying the atomic force field discussed above and assigning initial, randomly generated velocities to each atom, these conditions are generally fulfilled.

The propagation of the coordinates and the creation of a time dependent trajectory are achieved by integration of the second law of motion over time. Time integration algorithms are based on finite difference methods, where time is discretized in short time steps, Δt. If the positions, velocities and forces are known at time t, the integration scheme then gives the same quantities at time $t + \Delta t$. These methods are completely deterministic, meaning that once the positions and velocities of each atom are known; the state of the system can be predicted at any time in the future or the past.
Mainly two types off errors are associated with these types of methods: *Truncation errors*, related to the accuracy of the finite difference method with respect to the true solution and *Round-off errors*, related to the finite number of digits used in computer arithmetics. Both errors can be reduced by decreasing the size of Δt.

One of the most basic and widely used methods for integration is the Verlet algorithm. The basic idea is to write two second-order Taylor expansions for the positions $r(t)$, one forward and one backward in time:

$$\begin{aligned} r(t+\Delta t) &= r(t) + v(t)\Delta t + \frac{1}{2}a(t)\Delta t^2 + \frac{1}{6}a'(t)\Delta t^3 \\ r(t-\Delta t) &= r(t) - v(t)\Delta t + \frac{1}{2}a(t)\Delta t^2 - \frac{1}{6}a'(t)\Delta t^3 \end{aligned} \tag{2}$$

where r is the position, v is the velocity and a is the acceleration. Adding the two expression gives:

$$r(t+\Delta t) = 2r(t) - r(t-\Delta t) + a(t)\Delta t^2 \tag{3}$$

Since we are integrating Newton's equations, $a(t)$ is just the force divided by the mass, and the force is in turn a function of the positions $r(t)$:

$$a(t) = -\frac{F(r(t))}{m} \tag{4}$$

It evident from the above equations that the truncation error of the algorithm is of the order Δt^4, since this is the largest term not accounted for in the Taylor expansion. The Verlet algorithm is at the same time simple to implement, accurate and has been proven to be stable, explaining its large popularity among molecular dynamics simulators. However, a problem with the Verlet algorithm is that velocities are not directly generated. While they are not needed for the time evolution, they are necessary on some occasions. For example, they are required to compute the kinetic energy K, which evaluation is necessary to test the conservation of the total energy $E=K+V$. This is one of the most important tests to verify that an MD simulation is proceeding correctly. In the Verlet algorithm, the velocities can be computed from the positions by using:

$$v(t) = \frac{r(t+\Delta t) - r(t-\Delta t)}{2\Delta t} \tag{5}$$

But the error associated with this expression is of the order Δt^2 rather than Δt^4.

To overcome this deficiency, variants of the Verlet algorithm have been developed. One of them is the *leapfrog* algorithm. To obtain more accurate velocities, the leapfrog algorithm calculates velocities at half time steps:

$$v\left(t+\frac{\Delta t}{2}\right) = v\left(t-\frac{\Delta t}{2}\right) + a(t)\Delta t \tag{6}$$

After this equation has been evaluated, the velocities of the current time step can be calculated using:

$$v(t) = \frac{v\left(t + \frac{\Delta t}{2}\right) + v\left(t - \frac{\Delta t}{2}\right)}{2} \tag{7}$$

This equation is necessary when the kinetic energy is needed at time *t*, for example when evaluating the conservation of energy. The atomic positions are then obtained from:

$$r(t + \Delta t) = r(t) + v\left(t + \frac{\Delta t}{2}\right)\Delta t \tag{8}$$

The leapfrog scheme, which involves only addition of terms compared to the subtraction of terms in the Verlet algorithm, results in greater numerical stability and it is the algorithm that has been used throughout the work of this thesis.

3.1.3 Running a Molecular Dynamics Simulation

When setting up a molecular dynamics simulation there are some necessary decisions regarding simulations conditions to be made apart from choice of biological system and software package. Here I will briefly describe what they are and how I have performed the simulations presented in this thesis.

3.1.3.1 Solvation

Most biological macromolecules exist in a water environment and to properly reproduce the behavior of in vivo conditions, the biological molecule must be solvated. The representation of solvent can be either explicit or implicit. The explicit model, where each water molecule is represented and simulated explicitly, is the most rigorous and commonly used method. This representation is required if the important interactions between nucleic- or amino acids and water are to be studied. A problem with explicit representation of water is that after a proper solvation (about 10 Å of water around all sides of the solute), around 90 % of the system consists of water atoms, imposing large computational costs compared to the solute molecule alone. Several different water models of different complexities exist, but the simple TIP3P[75] and SPC/E[56] models has proven to give acceptable results at minimal costs and are the most widely used.

To circumvent the considerable increase in computational cost introduced by the addition of explicit solvent, implicit representations of water have been developed. In these solvation methods the water molecules are replaced with a continuum potential that tries to reproduce the solvent effect. This decreases the computational cost dramatically but comes at the expense of loss of atomic details of the system. The method is mainly used when studying large systems on long time scales.

In this thesis, explicit solvation with the TIP3P water models have been used in all papers except in Paper III, where the interaction of TIP3P, SPC/E and TIP5P[76] models with Mg^{2+} ions was compared.

3.1.3.2 System Boundary

In an explicit system, the studied solute molecule and the surrounding water molecules make up a finite system with the water molecules at the boundary having solvent on one side and vacuum at the other. A common solution to this problem is to duplicate the system periodically in all directions to mimic an infinite system. By applying these so-called periodic boundary conditions it is ensured that all simulated atoms are surrounded by neighboring atoms. Periodic boundary conditions can be used with geometries other than the cubic box such as the rhombic dodecahedron and the truncated octahedron. This is often done in order to reduce the number of water molecules in the system and make it more computationally efficient. Another way of solving the boundary problem and at the same time minimize the number of atoms in the calculations can be to apply a spherical (i.e. non-periodical) boundary potential. This is useful when studying a particular part of a larger system where the region of interest is enclosed within a shell. The atoms outside the shell are cut out and deleted and the atoms within the shell simulated. The boundary solute atoms in the shell are restrained to their starting positions, while waters and ions are prevented from leaving by an energy potential shaped as a sphere.

The work presented in Paper II-IV was carried out using periodic boundary conditions while the ribosomal A-site, studied in Paper I, was simulated with a spherical boundary potential.

3.1.3.3 Ensembles

Molecular dynamics simulations are usually performed as close to experimental conditions as possible and physical properties such as pressure and temperature are chosen to achieve this. MD-simulations are performed under different conditions to produce a group of microscopic states corresponding to the same macroscopic or thermodynamical state, called ensembles. Several ensembles can be used and the most widely used in MD-simulations are:

- Microcanonical ensemble (*NVE*): The thermodynamic state characterized by a fixed number of atoms, *N*, a fixed volume, *V*, and a fixed energy, *E*. This corresponds to an isolated system which strives towards maximizing entropy at equilibrium.

- Canonical Ensemble (*NVT*): The thermodynamic state characterized by a fixed number of atoms, fixed volume, and a fixed temperature, *T*. Will result in a minimum of Helmholz free energy at equilibrium.

- Isobaric-Isothermal Ensemble (*NPT*): This ensemble is characterized by a fixed number of atoms a fixed pressure, *P*, and a fixed temperature. Will result in a minimum of Gibbs free energy at equilibrium.

In the work presented in this thesis, simulations have been carried out using the NVT ensemble in Paper I and the NPT ensemble in Paper II-IV. In order to keep the pressure and temperature constant in an NPT ensemble, the Berendsen thermostat and barostat[56] can be used. In an canonical ensemble the Nosé-Hoover thermostat[77, 78] is often used to control the temperature.

3.1.4 Sampling in Biomolecular Simulations

The main purpose of a biomolecular simulation is to sample as many configurations of the system as possible. Despite a wealth of algorithmic innovations over the past decades, MD largely remains the most popular tool for achieving maximum sampling

of biomolecular simulations systems. This comes from the simple fact that no other method can routinely and reliably outperform MD in sampling efficiency by a significant amount.[79] The most straightforward way of obtaining more sampling is of course to run longer simulations. But, although explicit-solvent MD simulations are now routine performed on the 100 ns – 1 μs scale , modern MD studies still often fall short of what is needed for statistically valid equilibrium simulation. Ideally, one would like to run a simulation at least 10 times longer than the slowest important timescale in a system. Unfortunately, many biomolecular timescales exceed 1 ms, far away from the achievable lengths.

Here I will briefly go through the reasons why many configurations are so hard to sample adequately and how the sampling can be assessed quantitatively. Methods for improving sampling will not be discussed in detail since they have not been extensively employed in this thesis. The discussion will be made from the perspective of all-atom, explicit solvent MD-simulations.

3.1.4.1 The Sampling Problem

The configurations of a biomolecular system can be described as an energy landscape and the probability of finding the system at a specific configuration $\boldsymbol{x}$ can be expressed as:

$$\rho(x) \propto e^{-E(x)/k_b T} \tag{9}$$

where ρ is the probability density function, E is the free energy function (depending on the ensemble), k_B is Boltzmann's constant, and T is the temperature. An ideal, *uncorrelated* sampling of the generated configurations would follow the distribution in equation (9). In an ensemble of N configurations any region of configurational space with probability $\rho > 1/N$ are likely to be represented and this would be true regardless of kinetic barriers among states. In a real life simulations however, the energy landscape is riddled with local minima and kinetic barriers preventing an ideal sampling (Figure 8). This results in that the sampling is *correlated* with previous configurations and with the initial condition. In a typical simulation of a folded biomolecule the problem of reaching the free energy minima of the system is usually not limited by the position of the initial structure. The bottleneck of obtaining a complete sampling of all relevant configurations will instead be to pass over the multiple barriers connecting the substates representing the folded structure.

These transitions over barriers allowing the communication between substates are not only essential for the assessment of relative population (thermodynamical stability) of substates, the average time spent in each minima also tells gives us the rate constant of the actual transitions. The error sizes in determining these properties are directly proportional to the number of transitions, making extensive sampling one of the most important factors for a successful MD-simulation.

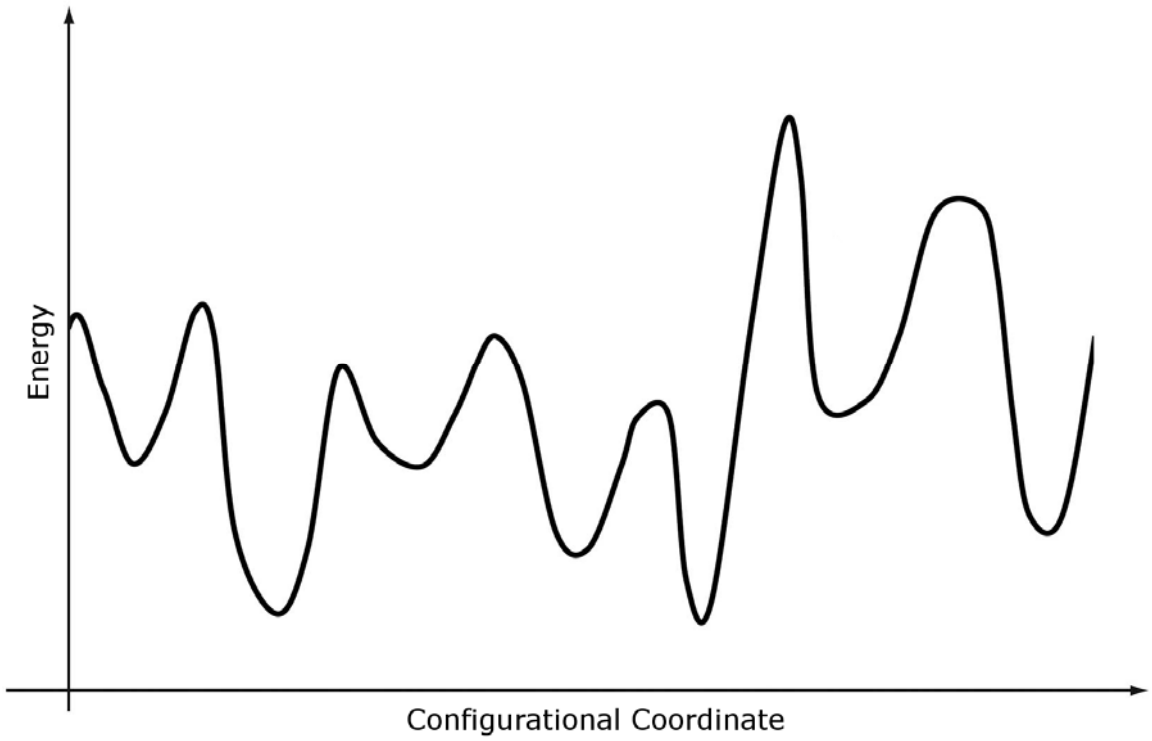

Figure 8. Schematic representation of the energy landscape of a biomolecular system. In a simulation, the system easily gets trapped in a minima and the crossing of the barriers is crucial for successful simulations.

3.1.4.2 Quantitative Assessment of Sampling

Despite the central role of sampling in simulations and awareness of our limitations in obtaining sufficient amounts of it, quantitative assessment of sampling has often received relatively little attention. No widely accepted and used methods for quantifying the effectiveness of sampling procedures have surfaced during the last decades.

The methods for assessing sampling that have been proposed can be divided into two groups, absolute and relative measures. Absolute measures attempt to give a binary indication of whether convergence has been achieved, while relative measures estimate how much sampling has been achieved, e.g. by counting the number of times an event occurs. The main limitation of absolute measures lie in the problem of defining an unambiguous point where sampling can be considered absolutely converged. In spite of this shortcoming, absolute methods, like ergodic measures[80] and cluster counting[81] can be valuable in detecting extremely poor sampling or when comparing sampling of the same system, obtained with different approaches. In relative measures it is assumed that there is a fundamental correlation time for the overall sampling of the system, t^*_{corr}. The number of effectively independent configurations (the number of uncorrelated times a configuration is generated) can then be defined as:

$$N^{eff} = \frac{t_{sim}}{t^*_{corr}} \tag{10}$$

where t_{sim} is the total simulation time used. In broad terms absolute lack of sampling correspond to $N^{eff} < 1$ and to make statistically relevant observations it is desirable that $N^{eff} > 10$.[79] The problem with this methodology lies of course in determining t^*_{corr} but several approximations have been proposed. Lyman & Zuckerman[82] derived their correlation time from the overall distribution in configuration space, estimating the time that must elapse between trajectory frames so that they behave as if statistically independent. Another method for determining the overall correlation time, using principal component analysis, has been proposed by Hess.[83] These approaches have the twin advantages of being based on the full configuration-space distribution (as opposed to isolated observables) and of being blindly and objectively applicable to any dynamical trajectory.

In Paper IV, cluster counting was used to compare the sampling obtained with one long or several shorter simulations of Glutaredoxin 1.

3.1.4.3 Improving Sampling

Many methods to improve the sampling of standard MD-simulations have been suggested over the years. Most of them algorithmic tools, involving e.g. increased temperature, but lately also hardware approaches, like running on GPUs and custom built hardware have been used to improve the sampling of biomolecular systems. In this thesis only the SHAKE algorithm, which reduces the degrees of freedom by constraining the high frequency vibrations of hydrogens, and umbrella sampling (see Free Energy Calculations section below) have been used. More exhaustive descriptions of methods for improving sampling can be found in reviews by Zuckerman[79] and van Gunsteren.[84]

3.1.5 Free Energy Calculations

Perhaps the most important quantity in thermodynamics is Gibbs free energy and it is defined as:

$$G(T, p) = U + pV + TS \tag{11}$$

where U is the internal energy, pV is volume work, T is the temperature and S is the entropy. The free energy of a system is hard to obtain due to the limited sampling of all regions of phase space inherent to an MD-simulation. However ways to obtain the free energy difference of an event in a molecular system, such as a conformational change, binding of a ligand, mutation of a residue or solvation of an atom have been developed. The free energy difference of a process provides a thorough description of its driving forces and is often reported in biology in investigations of relative stability such as comparisons of binding affinities and structural conformations.

The free energy difference between the initial and final state, $\Delta G_{A\to B}$, gives us the equilibration constant:

$$\ln K = -\frac{\Delta G_{A\to B}}{RT} \tag{12}$$

Using transition state theory, the free energy barrier between the two states, $\Delta G^{\dagger}$, gives us the rate at which a process occurs:

$$k = Ae^{-\Delta G^{\dagger}/RT} \tag{13}$$

where R is the molar gas constant and A is a pre-exponential factor with unit s^{-1}.

Biological macromolecules with large number of atoms have complex free energy landscapes which are practically impossible to sample with conventional simulations. Several methods to circumvent this problem are used. Two of the most commonly used methods are free energy perturbation (FEP) and potential of mean force (PMF) with umbrella sampling, each suitable for specific problems. FEP is used for chemical process such as mutations of atoms and phase transitions, while PMF is used for physical process such as conformational changes. All of the papers except Paper IV included in this thesis involve free energy calculations in one or several ways.

3.1.5.1 Free Energy Perturbation

Free energy perturbation (FEP) is often referred to as an alchemical method since it involves transforming atoms in the initial state to the final state in an unphysical way. This methodology is possible thanks to the fact that the free energy is a thermodynamic state function. This means that its value does not depend on the actual path of the process, only on the end states. This allows the use of a thermodynamics cycle, exemplified with the difference in ΔG binding of two small molecules to a macromolecule in Figure 9.

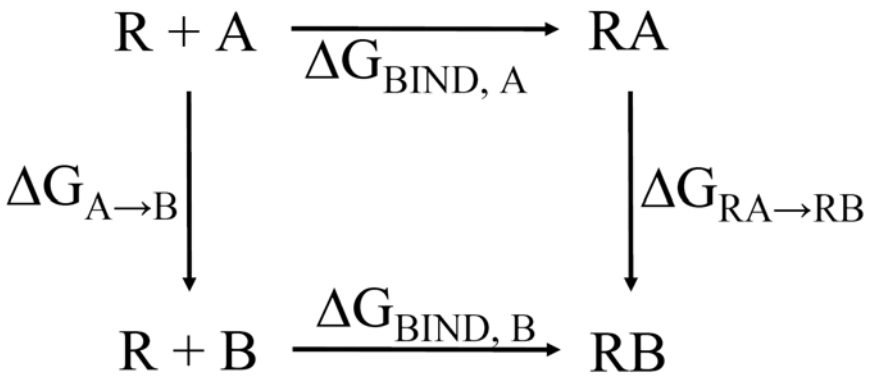

Figure 9. Thermodynamic cycle. The reactions exemplify the fictive binding of two small molecules (A and B) to a large target (R).

The horizontal reactions, which are the experimentally relevant reactions, would be computationally very demanding using a physics based potential energy since they involve moving the small molecule through the solvent and possibly also requiring large conformational changes to allow access to the binding site. Utilizing the fact that the free energy is a state function, we can more easily obtain the difference in binding affinity between A and B by instead calculating the vertical reactions. The difference in affinities is then given by $\Delta G_{Bind,B} - \Delta G_{Bind,A} = \Delta G_{RA \to RB} - \Delta G_{A \to B}$.

To be able to perform the transformation of one residue to another, a hybrid residue consisting of both the reactant and product residues is created. The energy of the transformation is described by:

$$H(\lambda) = \lambda H_B + (1-\lambda) H_A \tag{14}$$

where λ is a coupling parameter ranging from 0 (product state) to 1 (reactant state) in small steps. The free energy difference is then calculated by summing the contributions from the simulations at all intermediate λ-values:

$$\Delta G = \sum_{\lambda=0}^{1} -RT \ln \left\langle e^{-\Delta E'/RT} \right\rangle_{\lambda} \tag{15}$$

where the angular brackets denote averaging over a simulation performed at a given value of λ, E is the energy corresponding to the Hamiltonian in eq. 14 and $\Delta E' = E_{\lambda+d\lambda} - E_{\lambda}$.

In Paper I this methodology has been used to compare the binding affinities of all four RNA nucleosides in wobble position of mRNA to the modified uracil cmo^5U34 in the tRNA anti-codon. In Paper III a variant of FEP, thermodynamic integration, was used to compare the free energy of solvation between two Mg^{2+} parameter sets.

3.1.5.2 Potential of Mean Force and Umbrella Sampling

A potential of mean force (PMF) is the free energy of a system as a function of some internal coordinate or reactions coordinate (x), for example a distance between two atoms or the rotation of a dihedral. Unlike the non-physical pathways in the free energy perturbation calculations, the PMF is calculated for a physically achievable process. The free energy is related to the probability distribution of states as a function of that specific coordinate. The probability of finding a certain state can be described as a function of the PMF according to the Boltzmann distribution:

$$\rho(x) \propto e^{-g(x)/k_b T} \tag{16}$$

where $g(x)$ is the PMF, k_b is the Boltzmann constant and T the temperature. As previously discussed, MD-simulations do not sample regions that deviate largely from the equilibrium state adequately, leading to poor values for the potential of mean force of high energy regions. To sample all configurations along the reaction coordinate at reasonable computational times, the potentials of mean force are obtained from a series of simulations, each having an extra potential (w) to restrain the coordinate to specific points along desired reaction coordinate. This potential, referred to as an umbrella potential, often takes the form of a harmonic potential:

$$w_i(x) = k_i (x - x_i)^2 \tag{17}$$

where x is the value of the reaction coordinate, x_i is the constrained value and k_i the force constant. The PMF is extracted from the probability distribution in each simulation after the bias potential has been removed. A popular method to perform this post-processing and receive the final PMF is the Weighted Histogram Analysis Method[85, 86].

In Paper I, PMF was used to assess the free energy difference between two alternate binding conformations of a base pair and in Paper II it was used to investigate free

energy of unfolding of a riboswitch. In Paper III PMF was used to calculate the kinetics of Mg^{2+}-water exchange.

3.2 EXPERIMENTAL TECHNIQUES FOR STUDYING RNA AND PROTEINS

A close inter-dependence exists between molecular dynamics simulations and experimental techniques. MD-simulations rely on crystallographic and spectroscopic methods for three dimensional starting structures of proteins and nucleic acids. Large amounts of accurate structures and dynamical data are also needed for the development of force fields. On the other hand, experimentalists rely on computer simulations for refinement of structures and the detailed analysis and interpretation of results.

The two most successful techniques of obtaining atom resolution structures of proteins and nucleic acids are X-ray crystallography and Nuclear Magnetic Resonance (NMR). These two methods have been used to solve the majority of the around 80000 structures available in the Protein Data Bank (PDB)[87]. The techniques will be briefly explained below together with notes on how they relate to the work presented in this thesis.

3.2.1 X-ray Crystallography

Perhaps the most important technique for revealing the structure of proteins and nucleic acids is X-ray crystallography. In this method, the biological molecule is crystallized and struck by a high intensity beam of X-ray radiation. The X-rays will diffract into many specific directions after hitting the molecule. From the angles and intensities of the diffracted beams, a three-dimensional picture of the density of electrons can be derived. This density can then be used to determine the positions of the atoms and the type of chemical bonds between them. The limiting step in solving a structure using X-ray crystallography is often the crystallization of the studied molecule. A large uniform crystal is needed to obtain high quality data. The most important parameters to determine the quality of a published structure is the resolution (commonly between 2 Å and 3.5 Å) and the R-factor which is a measure of how well the refined structure predicts the observed data.

Historically, the method has been hugely successful and numerous Nobel prizes have been awarded for discoveries directly related to X-ray crystallography. Among them are the 1914 Physics prize to M. von Laue for the discovery of X-ray diffractions, the 1962 medicine prize to F. Crick, J. Watson and M. Wilkins for their contributions leading to the discovery of the DNA helix, published in 1953.[2] More recent prizes include the 2003 Chemistry prize to R. MacKinnon for his structures of the potassium channels and the 2009 chemistry prize to V. Ramakrishnan, T.A. Steitz and A.E. Yonath for their structures of the ribosome, some of which have been used as starting structures in Paper I of this thesis.[22] In paper II and III the X-ray structure of the *add* A-riboswitch[10] has been used.

3.2.2 Nuclear Magnetic Resonance Spectroscopy

A second method providing atomic resolution structures utilizes the Nuclear Magnetic Resonance (NMR) phenomenon. It occurs when the nuclei of certain atoms are subjected to a static magnetic field and then exposed to a second oscillating magnetic field. For nuclei to experience this phenomenon it must possess a property called spin or magnetic momentum, which comes from an odd number of protons in the nuclei. Some of the atoms with this property most commonly found in biomolecules are ^{1}H, ^{13}C, ^{15}N and ^{25}Mg, making them detectable in an NMR experiment.

When these atoms are subjected to the alternating magnetic field they give off signals called chemical shifts. The chemical shifts also depend on the shielding effect of neighboring atoms. This effect gives rise to unique signals dependent on the surrounding atoms which can be used to determine the structural environment. The spins are also sensitive to the presence of magnetic moments from other spins and are therefore "coupled". How well the spins are coupled is represented by the so called J-coupling and together with the chemical shifts, they can be used to provide high resolution structural information.

Another property that can be measured with NMR spectroscopy is spin relaxation. It is evoked by replacing the oscillating magnetic field described above with short radio frequency pulses perpendicular to the static field. The pulse causes a perturbation of the orientation of NMR-active atoms. What is measured is however what time it takes for the system to relax to its initial orientation. Three relaxation parameters can be measured in an NMR relaxation experiment, *NOE*, T_1 and T_2. *NOE*, or Nuclear Overhauser Effect, is a consequence of dipole–dipole coupling of nuclear spins and can be used to in the determination and refinement of structures. T_1, or the spin-lattice relaxation time, can be thought of as a characterization of the return of the spin orientations to their equilibrium positions. T_2, or the spin-spin relaxation time, describes the change in magnetic field and is related to the exchange of spin between nuclei. The time dependence of the relaxation rates provides unique information on the local and global dynamics of biomolecules, not available from X-ray crystallography. The dynamical data from an NMR relaxation experiment is often presented with the so called generalized order parameters (S^2), which are obtained using the model-free approach of Lipari-Szabo[88].

NMR spectroscopy requires the studied molecule to be in solution and ability to successfully solvate the biomolecule is often the limiting factor in which systems that can be studied. The high concentration needed to obtain high quality data sets a limit on the size of the systems possible to investigate at about 70 kD and also prevents the study insoluble systems such as membrane proteins.

In this thesis, the Mg^{2+} parameters developed in paper III was optimized to fit data on the kinetics of Mg^{2+}-H_2O exchange obtained with NMR spectroscopy. In paper IV, NMR relaxation parameters are calculated from an MD-trajectory and linked to corresponding dynamical events, providing valuable information in the analysis of experimental NMR data. The 3-D structures of Glutaredoxin 1 and 3 used in Paper IV were solved using solution NMR.[89, 90]

4 SUMMARY OF MAIN RESULTS

"Science increases our power in proportion as it lowers our pride."
- Claude Bernard (1813-78), French physiologist.

4.1 PAPER I

Nucleotide modifications and tRNA anticodon - mRNA codon interactions on the ribosome

In the first paper we performed MD simulations of the tRNA anticodon and mRNA codon in the decoding center of the ribosome to study the effect of the common tRNA modifications cmo^5U34 and m^6A37. Post transcriptional modifications are very common in these positions of tRNA and in $tRNA^{Val}$ these two modifications allow all four nucleotides to be successfully read at the wobble position in a codon. This extraordinary ability makes $tRNA^{Val}$ a very interesting system for studying the influence of modifications on the decoding of mRNA during protein synthesis in the ribosome. Previous data suggest that entropic effects are mainly responsible for the extended reading capabilities. However, detailed mechanisms have remained unknown and the aim of this paper was to provide an atomic explanation of how the decoding capabilities are expanded by the cmo^5U34 and m^6A37 modifications. To elucidate the details of these mechanisms and quantify their effects, we performed a wide range of simulations: Extensive free energy perturbation coupled with umbrella sampling, entropy calculations of tRNA (free and bound to the ribosome) and thorough structural analysis of the ribosomal decoding center. After analyzing the entropy of the modified and unmodified ASL we found that no prestructuring effect on the tRNA anticodon stem loop from the two modifications could be observed. We could however identify two other mechanisms that the modifications employ that may contribute to the expanded decoding capability. *1. Alternate binding conformations*: The extended reach of the cmo^5U34 allows an alternative binding conformation to be formed for the three non cognate base pairs at the wobble position. In the case of Ura-Ura and Ura-Cyt base pairing, the highly polar carboxyl group of cmo^5 bridges over the distance gap forming a hydrogen bond. In the case of the Ura-Gua pair a network of contacts is formed. These conformations lie lower in free energy than the standard mismatch binding by approximately 1-3 kcal/mol (Figure 10). *2. Increased interactions with the ribosome.* Additional contacts between the ribosomal residues of the decoding center and anti-

codon enhance the "catalyzing" effect of the ribosome are observed. One of these contacts is the bridging, by the carboxyl group of cmo^{5}U34, between ribosomal C1054 and the first base of the next mRNA codon.

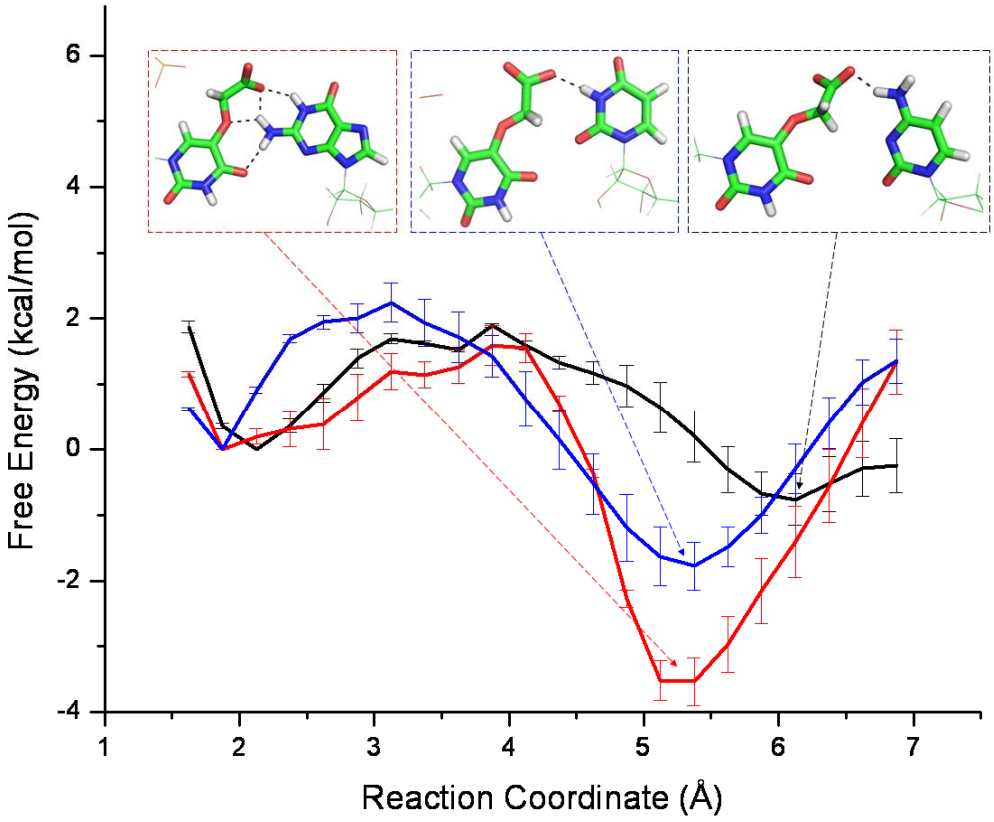

Figure 10. Relative free energy between the standard and outer (insets) conformations obtained by PMF calculations. The outer conformation, made possible with the cmo^{5}U34 modification lie lower in energy than standard mismatch binding for three base pairs. The reaction coordinate of the umbrella sampling is the distance between O4 of cmo^{5}U34 and H1, H42 and H3 for guanine (red), cytosine (black) and uracil (blue) respectively.

4.2 PAPER II

Effects of Ion Environment and Ligand on the RNA Kissing Loops in the add Adenine Riboswitch

In the second paper we study the effect of ions and ligand binding on the free energy and structure during folding of a riboswitch. Like other structured RNA molecules, riboswitches are heavily dependent on cations, especially Mg^{2+}, to function properly and can in some aspects be a substitution for ligand binding. We performed MD simulations of the adenine-sensing *add* A-riboswitch to study the effect of ion environment and ligand presence on the formation of the kissing hairpin loops, a key step in the unfolding mechanism of the aptamer domain. We investigated the aptamer domain of *add* A-Riboswitch in complex with its conjugate ligand and in ligand-free state, in Mg^{2+} and Na^{+} ion solution Only minor structural differences among the systems are observed during the simulation time, but the Mg^{2+} ions are found to provide 33 % more stabilization than the monovalent Na^{+} ions to the RNA system. The opening of the hairpins was simulated using umbrella sampling, using the distance between two kissing loops as reaction coordinate. A two-step process was observed in all the simulated systems. First, a general loss of stacking and hydrogen bond interactions was observed. Lastly, two base pairs, G37-C60 and G38-C61, broke, pointing to their essential role in the tertiary kissing loops structure. Moreover, when increasing the loop-loop distances, the L3 hairpin displayed more flexibility by changing its orientation in the structure while the other conserved its coaxial arrangement with rest of the structure (Figure 11). The calculation of free energy profiles of the hairpin opening, however, revealed large deviations between replica simulations, suggesting insufficient sampling. A fourfold increase of the production sampling of each window of some of the simulations did not significantly improve the situation. A possible remedy will be to considerably extend the computationally expensive initial generation of starting structures for the umbrella sampling.

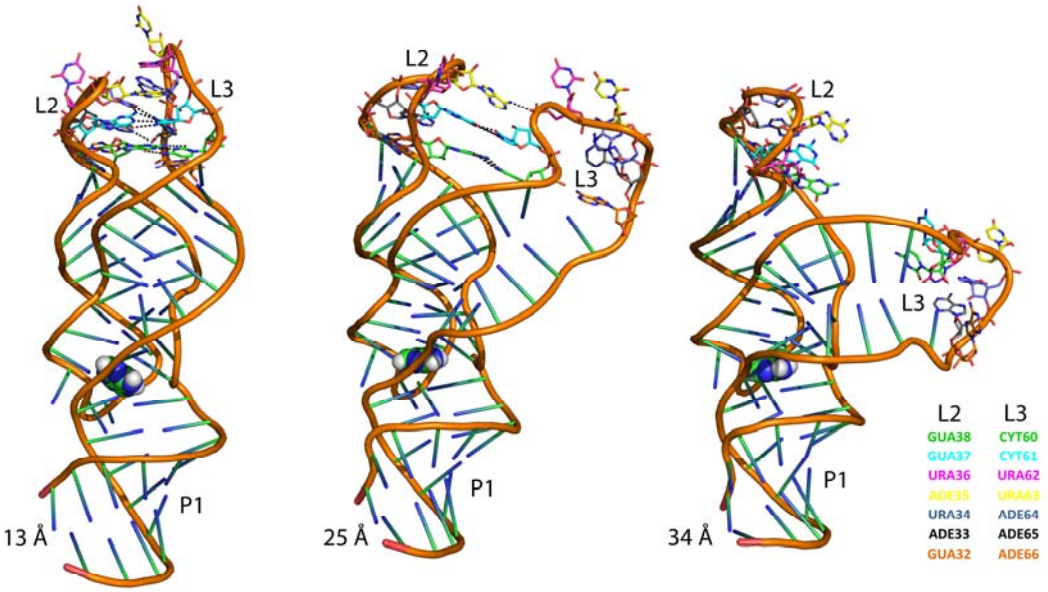

Figure 11. Conformational changes at three stages of the kissing loop unfolding of the *add* A-riboswitch. The L3 loop is observed to larger conformational changes than L2 during the opening. The ligand is shown in spheres and L2 and L3 loop residues are represented in sticks and colored according to the legend

4.3 PAPER III

Magnesium ion-water coordination and exchange in biomolecular simulations

The third paper is a methodology paper that aims to improve the representation of magnesium ions in simulation of nucleic acids. Mg^{2+} ions have an important role in the structure and folding mechanism of RNA systems. To properly simulate these biophysical processes, the applied molecular models should reproduce, among other properties, the kinetics of the ions in water solution. In this paper we studied the kinetics of the binding of magnesium ions with water molecules and nucleic acids systems in detail, using MD simulation. We began by validating the parameters used in biomolecular force fields, such as AMBER and CHARMM, for Mg^{2+} and for the other, most biological relevant ions, Na^{+}, K^{+} and Ca^{2+} together with three different water models (TIP3P, SPC/E and TIP5P). The initial results showed that Mg^{2+} ions have a slower exchange rate than Na^{+}, K^{+} and Ca^{2+} in agreement with the experimental trend, but the simulated value underestimates the experimentally observed Mg^{2+}-water exchange rate with several orders of magnitudes, irrespective of force field and water model. To remedy this situation we developed a new set of parameters for Mg^{2+} to reproduce the experimental kinetic data (Figure 12). The new set was also shown to give better reproduction of structural data than existing models. These improvements were accompanied with only slight changes (6%) in solvation free energy; the property previous parameters were optimized to. We then applied the new parameters set to the more physical process of Mg^{2+} binding with a mono-phosphate model system and with a purine riboswitch, the *add* A-riboswitch. In line with the Mg^{2+}-water results, the newly developed parameters showed a better description of the structure and kinetic of the Mg^{2+}-phosphate binding than all existing parameters. The characterization of the ion binding to the riboswitch system showed that the new parameter set does not affect the global structure of the ribonucleic acid system or the number of ions involved in direct or indirect binding. A slight decrease in the number of water-bridged contacts between A-riboswitch and Mg^{2+} ion was however observed. Combined, the results clearly show that the newly developed parameters improve the kinetic and structural description of Mg^{2+} ions and their applicability in nucleic acid simulation.

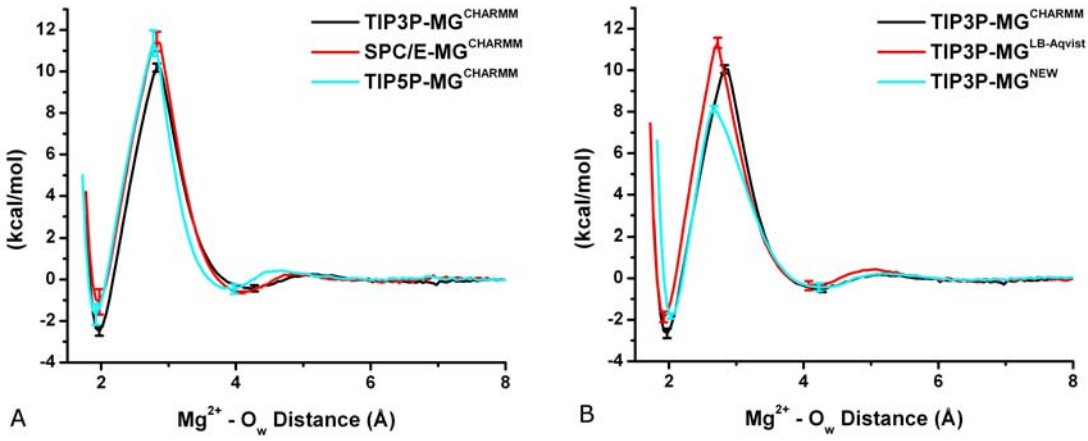

Figure 12. Potential of mean force between Mg^{2+} and a water oxygen around Mg^{2+} using different water models (A) and Mg^{2+} parameters (B). While water models have little effect on the Mg^{2+} - water exchange rate, our new parameters significantly improve the agreement with experimental data.

4.4 PAPER IV

Motions and Entropies in Proteins as Seen in NMR Relaxation and Molecular Dynamics Simulations

In another methodology study, paper IV aims to improve the interpretation of experimental data on dynamics. We performed MD simulations of the proteins glutaredoxin1 in water to relate the data of dynamics and entropy obtained in NMR relaxation experiments with data extracted from a MD trajectory. NMR relaxation is the most widely used experimental method to obtain data on dynamics of proteins but it is limited to relatively short timescales and to motions of backbone amides, or in some cases ^{13}C-H vectors. By relating the experimental data to the complete picture given by the molecular dynamics trajectory, valuable insights on the interpretation of the experiment can be gained. The internal dynamics and their time scales were estimated by calculating the generalized order parameters (S^2) for different time windows. We went on to calculate the quasiharmonic entropy (S) and compare it to the entropy calculated from the generalized order parameter of the amide vector, which is available from an NMR experiment. The focus was put on characterizing dynamics that are not expressed through the motions of the amide group. We finally evaluated the amount of sampling obtained with our molecular dynamics simulations and how it is affected by the length of individual simulations by clustering of the sampled conformations. The results show that two turns act as hinges, allowing the alpha-helix between them to undergo large, long time scale motions that can not be detected in the NMR time window. We also find that the entropy obtained from the amide vector is unable to account for the often heavily correlated motions of adjacent residues in regions of the protein with large dynamics (Figure 13A). Finally we show that the sampling in a total of 100 ns molecular dynamics simulation can be can be increased by around 50% by dividing the trajectory into 10 replicas with different starting velocities (Figure 13B).

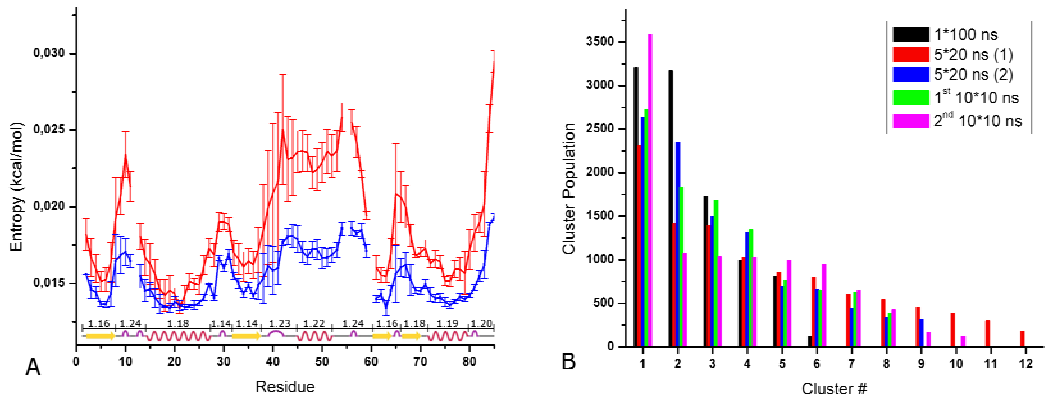

Figure 13. Comparison of NMR and quasiharmonic entropy of GRX1 (A) and amount of sampling from different simulations (B). (A): Entropy per residue of GRX1. The red curve is the entropy of the ^{15}N-^{1}H vector estimated by the quasiharmonic model and blue curve is the entropy of the same atoms estimated by the NMR model. The bottom values are the ratios of entropy calculated individually per residue in relation to the summed entropy of the segment as a whole. Significant entropy is missed with the NMR model (B): Number of conformational clusters produced by 100 ns sampling obtained by different simulation lengths and parts. Division of the trajectory into smaller parts yields more sampling.

4.5 SAMMANFATTNING PÅ SVENSKA

Processen med vilken information lagrad i DNA i gener omvandlas till funktionella RNA-molekuler och proteiner tillhör de mest fundamentalt viktiga processerna i allt känt liv. Även om den processen involverar stora makromolekyler samt rörelser som tar relativt lång tid att slutföra beror de alla till syvende och sist på interaktioner mellan enskilda atomer i nukleinsyror (DNA och RNA) och amino syror (proteiner). Endast ett fåtal experimetella metoder finns tillgängliga som kan studera dessa detaljerade interaktioner i de stora system som är involverade i biologiska processer. Datorsimuleringar av biologiska system är därför ett oumbärligt verktyg för att förstå grundläggande biologiska mekanismer. I den här avhandlingen har Molekyl Dynamik (MD) simuleringar används för att studera hur den genetiska koden i mRNA avkodas i ribosomen för att forma proteiner och hur en riboswitch fungerar när den reglerar uttrycket av en gen. Andra halvan av avhandlingen täcker arbeten som syftar till att förbättra de metoder som används för att utföra dessa simuleringar.

I Artikel I undersöks effekten av hur modifikationer (extra atomgrupper) på tRNA påverkar avkodningen av mRNA i ribosomen. MD simuleringar utfördes på en del av ribosomen, med och utan de undersökta modifikationerna för att bl.a. fastställa skillnader i fri energi. Resultaten visar att avkodningen påverkas genom två olika mekansimer: Den extra räckvidden som modifikationen ger möjliggör en alternativ bindning mellan tRNA och mRNA som är mer stabil än bindingen utan modifikation. Modifikatinen resulterar även i flera kontakter mellan tRNA och ribosomen, vilket förstärker den katalytiska funktionen.

I Artikel II studeras veckningsmekanismen hos en riboswitch i olika jonmiljöer samt med och utan bunden ligand. Utöver traditionella MD simuleringar användes också en speciell metod för att ”tvinga” RNA molekylen att öppna sig. Resultaten visar inte några signifikanta effekter av vare sig jontyp eller närvaron av bunden ligand. Dock observerade vi konsekvent en öppningsmekanism där vissa delar av molekylen är betydligt mer flexibla än andra. Mer data kan dock behövas för att kunna dra säkra, generella slutsatser.

I Artikel III presenteras en ny modell för att simulera de biologiskt mycket viktiga magnesiumjonerna. Den nya modellen har optimerats för att reproducera dynamiska egenskaper istället för energetiska som tidigare modellar optimerats för. Vi visar att den nya modellen inte bara uppvisar bättre dynamiska egenskper utan även strukturella.

I Artikel IV jämförs dynamisk data erhållet med NMR experiment (ett av få sätt att mäta molekylrörelser experimentellt) med dynamik observerad i MD simuleringar. Denna analys ger viktiga insikter i hur man ska tolka experimentell data samt hur förbättrade simuleringmetoder kan utvecklas. Resultaten visa bl.a. att väsentliga delar av entropin hos ett protein inte ses genom NMR p.g.a. ett begränsat tidsfönster samt en oförmåga av metoden att beskriva korrelerade rörelser.

4.6 CONCLUDING REMARKS AND FUTURE PERSPECTIVES

The work presented in this thesis covers both the investigations of fundamental biological questions and the improvement of computational and experimental

techniques. The ability to answer new, previously unattainable, biological questions will always be closely linked to the improvement of methodology and method development will continue to be important as long as we do not have a full understanding of biology. During my work I encountered the ever present problem of insufficient sampling and a sometimes less than intuitive usage of software packages. But new and exciting biological fields with questions and problems that can be solved with the help of computer simulations have also opened up.

In terms of method development, a few key areas of improvement can be identified. One origin for problems when running MD simulations is that most applications of software packages as black boxes are not reliable. The expertise needed for setting up and running anything but the simplest MD simulations is exclusive to a few research groups around the world. To remedy this problem, programs in computational biology or bioinformatics could offer undergraduate and graduate-level courses like bio/molecular modeling, computational chemistry, or modeling tools for experimentalists. Combined with improved user interfaces and documentation, the use of MD simulations will be made accessible to a wider range of researchers. Furthermore, in the development of the most important features of simulation packages such as new sampling techniques, better force fields, hybrid methods and parallelization of algorithms there is a lot to be gained in consolidation of efforts. Newly developed features should by made available to as many researchers as possible for testing and application by decreasing barriers between competing software packages. A consolidation of simulation software will also lower the barrier of entry for new researcher with fewer, more user friendly choices of software.

Specifically, I have during my work missed simple tools for assessing and comparing sampling and the performance of force fields, factors that are essential for any successful MD simulation. The problem with insufficient sampling became apparent during the simulations of the unfolding of the *add* A-riboswitch and a significant extension of the umbrella sampling will be needed to describe the folding energy landscape with certainty.

During my work, groundbreaking discoveries have opened up new exciting biological fields of study. In 2011 the first high resolution structures of the eukaryotic ribosome was solved.[91-93] The structures show a translational machinery that is significantly different from the prokaryotic ribosome in several key structural and mechanistic aspects. The new data opens up interesting new possibilities for designing novel antibiotics targeting the prokaryotic ribosome. Another emerging and promising field involving riboswitches and other RNA motifs is synthetic biology. Computational methods have proven to be an effective tool in designing functional RNA functional components that can be used as building blocks in larger assemblies.[94, 95] Synthetic riboswitches have important applications in biosynthesis processes, where they can be implemented as noninvasive sensors of metabolite accumulation and controllers for optimizing product yield.[96] Another very interesting application of RNA in synthetic biology is it's recently shown function as an molecular "scaffolding" that can be used to engineer the spatial organization of cellular molecules.[97] Further developments in these fields will require a detailed atomic understanding of the biology, an area where MD simulations have much to offer.

5 ACKNOWLEDGEMENTS

"In science one tries to tell people, in such a way as to be understood by everyone, something no one ever knew before. In poetry, it is the exact opposite."
- Paul Dirac (1902-1984), English theoretical physicist.

Many people have contributed to the work of this thesis and without their help and support it would never have been completed. Here are but a few of them:

First of all I would like to thank my main supervisor Lennart Nilsson for taking me on as a PhD student and guiding me through the whole process with adamant optimism and brilliant intelligence.

Secondly, I would like to thank my co supervisor Alessandra Villa, who with her openness and extraordinary attention to detail, makes science, not just correct, but also enjoyable to work with.

My beloved Annelie, for standing by me in trouble and joy. It is your time next…

Past and present members of the group: Mika, Mauricio, Alok, Christian, Vineet, Katarina, Sofia, Boel and especially Greg for all the stimulating and fruitful discussions.

All the people who helped organize Nov2k: John, Kristina, Elisa, Lars and especially Johan for all the things we experienced and learned.

All my friends from KTH (Gustaf, John, Christofer, Sara, Clinton, and Patrik) and the Stockholm Research School (Christoffer, Karl and Erik) for making the road here enjoyable.

My family for encouraging me to make the right choices.

6 REFERENCES

"If I have seen further than others, it is by standing upon the shoulders of giants."
- Sir Isaac Newton (1642-1727), English physicist and mathematician.

1. Franklin, R.E. and R.G. Gosling, *Molecular Configuration in Sodium Thymonucleate.* Nature, 1953. **171**(4356): p. 740-741.
2. Watson, J.D. and F.H.C. Crick, *The Structure of DNA.* Cold Spring Harbor Symposia on Quantitative Biology, 1953. **18**: p. 123-131.
3. Wilkins, M.H.F., A.R. Stokes, and H.R. Wilson, *Molecular Structure of Deoxypentose Nucleic Acids.* Nature, 1953. **171**(4356): p. 738-740.
4. Nudler, E. and A.S. Mironov, *The riboswitch control of bacterial metabolism.* Trends in Biochemical Sciences, 2004. **29**(1): p. 11-17.
5. Winkler, W.C. and R.R. Breaker, *Regulation of bacterial gene expression by riboswitches.* Annual Review of Microbiology, 2005. **59**: p. 487-517.
6. Garst, A.D., A.L. Edwards, and R.T. Batey, *Riboswitches: Structures and Mechanisms.* Cold Spring Harbor Perspectives in Biology, 2011. **3**(6).
7. Hammann, C. and E. Westhof, *Searching genomes for ribozymes and riboswitches.* Genome Biology, 2007. **8**(4).
8. Montange, R.K. and R.T. Batey, *Riboswitches: Emerging themes in RNA structure and function.* Annual Review of Biophysics, 2008. **37**: p. 117-133.
9. Kim, J.N. and R.R. Breaker, *Purine sensing by riboswitches.* Biology of the Cell, 2008. **100**(1): p. 1-11.
10. Serganov, A., et al., *Structural basis for discriminative regulation of gene expression by adenine- and guanine-sensing mRNAs.* Chemistry & Biology, 2004. **11**(12): p. 1729-1741.
11. Gilbert, S.D. and R.T. Batey, *Riboswitches: Fold and function.* Chemistry & Biology, 2006. **13**(8): p. 805-807.
12. Greenleaf, W.J., et al., *Direct observation of hierarchical folding in single riboswitch aptamers.* Science, 2008. **319**(5863): p. 630-633.
13. Neupane, K., et al., *Single-molecule force spectroscopy of the add adenine riboswitch relates folding to regulatory mechanism.* Nucleic Acids Research, 2011. **39**(17): p. 7677-7687.
14. Nasvall, S.J., P. Chen, and G.R. Bjork, *The modified wobble nucleoside uridine-5-oxyacetic acid in tRNA(cmo5UGG)(Pro) promotes reading of all four proline codons in vivo.* Rna-a Publication of the Rna Society, 2004. **10**(10): p. 1662-1673.
15. Kim, S.H., et al., *3-Dimensional Tertiary Structure of Yeast Phenylalanine Transfer-Rna.* Science, 1974. **185**(4149): p. 435-440.
16. Robertus, J.D., et al., *Structure of Yeast Phenylalanine Transfer-Rna at 3 a Resolution.* Nature, 1974. **250**(5467): p. 546-551.

17. Shi, H.J. and P.B. Moore, *The crystal structure of yeast phenylalanine tRNA at 1.93 angstrom resolution: A classic structure revisited.* RNA-a Publication of the RNA Society, 2000. **6**(8): p. 1091-1105.
18. Agris, P.F., *The importance of being modified: Roles of modified nucleosides and Mg2+ in RNA structure and function.* Progress in Nucleic Acid Research and Molecular Biology, Vol 53, 1996. **53**: p. 79-129.
19. Agris, P.F., F.A. Vendeix, and W.D. Graham, *tRNA's Wobble Decoding of the Genome: 40 Years of Modification.* J Mol Biol, 2007. **366**(1): p. 1-13.
20. Nishimura, S. and K. Watanabe, *The discovery of modified nucleosides from the early days to the present: A personal perspective.* Journal of Biosciences, 2006. **31**(4): p. 465-475.
21. Ban, N., et al., *The complete atomic structure of the large ribosomal subunit at 2.4 angstrom resolution.* Science, 2000. **289**(5481): p. 905-920.
22. Selmer, M., et al., *Structure of the 70S ribosome complexed with mRNA and tRNA.* Science, 2006. **313**(5795): p. 1935-1942.
23. Schluenzen, F., et al., *Structure of functionally activated small ribosomal subunit at 3.3 angstrom resolution.* Cell, 2000. **102**(5): p. 615-623.
24. Korostelev, A., D.N. Ermolenko, and H.F. Noller, *Structural dynamics of the ribosome.* Curr Opin Chem Biol, 2008. **12**(6): p. 674-83.
25. Ogle, J.M., A.P. Carter, and V. Ramakrishnan, *Insights into the decoding mechanism from recent ribosome structures.* TIBS, 2003. **28**(5): p. 259-266.
26. Rodnina, M.V., et al., *GTPases mechanisms and functions of translation factors on the ribosome.* Biol Chem, 2000. **381**(5-6): p. 377-87.
27. Ramakrishnan, V., *Ribosome structure and the mechanism of translation.* Cell, 2002. **108**(4): p. 557-572.
28. Steitz, T.A., *A structural understanding of the dynamic ribosome machine.* Nature Reviews Molecular Cell Biology, 2008. **9**(3): p. 242-253.
29. Shine, J. and L. Dalgarno, *3'-Terminal Sequence of Escherichia-Coli 16S Ribosomal-Rna - Possible Role in Initiation and Termination of Protein-Synthesis.* Proceedings of the Australian Biochemical Society, 1974. **7**(May27): p. 72-72.
30. Kurland, C. and J. Gallant, *Errors of heterologous protein expression.* Current Opinion in Biotechnology, 1996. **7**(5): p. 489-493.
31. Freistroffer, D.V., et al., *Release factor RF3 in E-coli accelerates the dissociation of release factors RF1 and RF2 from the ribosome in a GTP-dependent manner.* Embo Journal, 1997. **16**(13): p. 4126-4133.
32. Draper, D.E., *A guide to ions and RNA structure.* RNA-a Publication of the RNA Society, 2004. **10**(3): p. 335-343.
33. Draper, D.E., D. Grilley, and A.M. Soto, *Ions and RNA folding.* Annual Review of Biophysics and Biomolecular Structure, 2005. **34**: p. 221-243.
34. Leroy, J.L., et al., *ROLE OF DIVALENT IONS IN FOLDING OF TRANSFER-RNA.* European Journal of Biochemistry, 1977. **74**(3): p. 567-574.
35. Brooks, B.R., et al., *CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations.* J. Comp. Chem., 1983. **4**: p. 187-217.
36. Stein, A. and D.M. Crothers, *Conformational-Changes of Transfer-Rna - Role of Magnesium(II).* Biochemistry, 1976. **15**(1): p. 160-168.
37. Romer, R. and R. Hach, *Transfer-Rna Conformation and Magnesium Binding - Study of Yeast Phenylalanine-Specific Transfer-Rna by a Fluorescent Indicator and Differential Melting Curves.* European Journal of Biochemistry, 1975. **55**(1): p. 271-284.
38. Manning, G.S., *Molecular Theory of Polyelectrolyte Solutions with Applications to Electrostatic Properties of Polynucleotides.* Quarterly Reviews of Biophysics, 1978. **11**(2): p. 179-246.
39. Lee, E.H., et al., *Discovery Through the Computational Microscope.* Structure, 2009. **17**(10): p. 1295-1306.
40. Leipply, D. and D.E. Draper, *Effects of Mg2+ on the Free Energy Landscape for Folding a Purine Riboswitch RNA.* Biochemistry, 2011. **50**(14): p. 2790-2799.

41. Vaiana, A.C., E. Westhof, and P. Auffinger, *A molecular dynamics simulation study of an aminoglycoside/A-site RNA complex: conformational and hydration patterns.* Biochimie, 2006. **88**(8): p. 1061-1073.
42. Klein, D.J., P.B. Moore, and T.A. Steitz, *The contribution of metal ions to the structural stability of the large ribosomal subunit.* Rna-a Publication of the Rna Society, 2004. **10**(9): p. 1366-1379.
43. Lemay, J.F., et al., *Folding of the adenine riboswitch.* Chemistry & Biology, 2006. **13**(8): p. 857-868.
44. Noeske, J., H. Schwalbe, and J. Wohnert, *Metal-ion binding and metal-ion induced folding of the adenine-sensing riboswitch aptamer domain.* Nucleic Acids Research, 2007. **35**(15): p. 5262-5273.
45. Ferre-D'Amare, A.R., *RNA-modifying enzymes.* Current Opinion in Structural Biology, 2003. **13**(1): p. 49-55.
46. Sutcliffe, J.A., *Improving on nature: antibiotics that target the ribosome.* Current Opinion in Microbiology, 2005. **8**(5): p. 534-542.
47. Deigan, K.E. and A.R. Ferre-D'Amare, *Riboswitches: Discovery of Drugs That Target Bacterial Gene-Regulatory RNAs.* Accounts of Chemical Research, 2011. **44**(12): p. 1329-1338.
48. McCoy, L.S., Y. Xie, and Y. Tor, *Antibiotics that target protein synthesis.* Wiley Interdisciplinary Reviews-Rna, 2011. **2**(2): p. 209-232.
49. Blount, K.F. and R.R. Breaker, *Riboswitches as antibacterial drug targets.* Nature Biotechnology, 2006. **24**(12): p. 1558-1564.
50. Hermann, T. and Y. Tor, *RNA as a target for small-molecule therapeutics.* Expert Opinion on Therapeutic Patents, 2005. **15**(1): p. 49-62.
51. Mulhbacher, J., et al., *Novel Riboswitch Ligand Analogs as Selective Inhibitors of Guanine-Related Metabolic Pathways.* Plos Pathogens, 2010. **6**(4).
52. Kim, J.N., et al., *Design and Antimicrobial Action of Purine Analogues That Bind Guanine Riboswitches.* Acs Chemical Biology, 2009. **4**(11): p. 915-927.
53. Dror, R.O., et al., *Exploring atomic resolution physiology on a femtosecond to millisecond timescale using molecular dynamics simulations.* Journal of General Physiology, 2010. **135**(6): p. 555-562.
54. Brooks, B.R., et al., *CHARMM: The Biomolecular Simulation Program.* Journal of Computational Chemistry, 2009. **30**(10): p. 1545-1614.
55. Pearlman, D.A., et al., *Amber, a Package of Computer-Programs for Applying Molecular Mechanics, Normal-Mode Analysis, Molecular-Dynamics and Free-Energy Calculations to Simulate the Structural and Energetic Properties of Molecules.* Computer Physics Communications, 1995. **91**(1-3): p. 1-41.
56. Berendsen, H.J.C., D. van der Spoel, and R. Vandrunen, *Gromacs - a Message-Passing Parallel Molecular-Dynamics Implementation.* Computer Physics Communications, 1995. **91**(1-3): p. 43-56.
57. Phillips, J.C., et al., *Scalable molecular dynamics with NAMD.* Journal of Computational Chemistry, 2005. **26**(16): p. 1781-1802.
58. McCammon, J.A., B.R. Gelin, and M. Karplus, *Dynamics of Folded Proteins.* Nature, 1977. **267**(5612): p. 585-590.
59. Schroder, G.F., M. Levitt, and A.T. Brunger, *Super-resolution biomolecular crystallography with low-resolution data.* Nature, 2010. **464**(7292): p. 1218-U146.
60. Case, D.A., *Molecular dynamics and NMR spin relaxation in proteins.* Accounts of Chemical Research, 2002. **35**(6): p. 325-331.
61. Henzler-Wildman, K.A., et al., *Intrinsic motions along an enzymatic reaction trajectory.* Nature, 2007. **450**(7171): p. 838-U13.
62. Tsui, V., et al., *NMR and molecular dynamics studies of the hydration of a zinc finger-DNA complex.* Journal of Molecular Biology, 2000. **302**(5): p. 1101-1117.
63. Glazer, D.S., R.J. Radmer, and R.B. Altman, *Improving Structure-Based Function Prediction Using Molecular Dynamics.* Structure, 2009. **17**(7): p. 919-929.
64. Hu, H., M. Elstner, and J. Hermans, *Comparison of a QM/MM force field and molecular mechanics force fields in simulations of alanine and glycine "dipeptides" (Ace-Ala-Nme and Ace-Gly-Nme) in water in relation to the*

problem of modeling the unfolded peptide backbone in solution. Proteins-Structure Function and Genetics, 2003. **50**(3): p. 451-463.
65. Faraldo-Gomez, J.D. and B. Roux, *On the importance of a funneled energy landscape for the assembly and regulation of multidomain Src tyrosine kinases.* Proceedings of the National Academy of Sciences of the United States of America, 2007. **104**(34): p. 13643-13648.
66. Radhakrishnan, R. and T. Schlick, *Fidelity discrimination in DNA polymerase beta: Differing closing profiles for a mismatched (G : A) versus matched (G : C) base pair.* Journal of the American Chemical Society, 2005. **127**(38): p. 13245-13252.
67. Kuhlman, B., et al., *Design of a novel globular protein fold with atomic-level accuracy.* Science, 2003. **302**(5649): p. 1364-1368.
68. Jiang, L., et al., *De novo computational design of retro-aldol enzymes.* Science, 2008. **319**(5868): p. 1387-1391.
69. Read, M., et al., *Structure-based design of selective and potent G quadruplex-mediated telomerase inhibitors.* Proceedings of the National Academy of Sciences of the United States of America, 2001. **98**(9): p. 4844-4849.
70. Darden, T., D. York, and L. Pedersen, *Particle Mesh Ewald - an N.Log(N) Method for Ewald Sums in Large Systems.* Journal of Chemical Physics, 1993. **98**(12): p. 10089-10092.
71. Essmann, U., et al., *A Smooth Particle Mesh Ewald Method.* Journal of Chemical Physics, 1995. **103**(19): p. 8577-8593.
72. Foloppe, N. and A.D. MacKerell, *All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data.* Journal of Computational Chemistry, 2000. **21**(2): p. 86-104.
73. Cornell, W.D., et al., *A second generation force field for the simulation of proteins, nucleic acids, and organic molecules.* Journal of the American Chemical Society, 1995. **117**: p. 5179-5197.
74. Oostenbrink, C., et al., *A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6.* Journal of Computational Chemistry, 2004. **25**(13): p. 1656-1676.
75. Damm, W., et al., *OPLS_2002: A new version of the OPLS-AA force field.* Abstracts of Papers of the American Chemical Society, 2002. **224**: p. U471-U471.
76. Mahoney, M.W. and W.L. Jorgensen, *A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions.* Journal of Chemical Physics, 2000. **112**(20): p. 8910-8922.
77. Hoover, W.G., *Canonical Dynamics - Equilibrium Phase-Space Distributions.* Physical Review A, 1985. **31**(3): p. 1695-1697.
78. Nose, S., *A Unified Formulation of the Constant Temperature Molecular-Dynamics Methods.* Journal of Chemical Physics, 1984. **81**(1): p. 511-519.
79. Zuckerman, D.M., *Equilibrium Sampling in Biomolecular Simulations.* Annual Review of Biophysics, Vol 40, 2011. **40**: p. 41-62.
80. Mountain, R.D. and D. Thirumalai, *Measures of Effective Ergodic Convergence in Liquids.* Journal of Physical Chemistry, 1989. **93**(19): p. 6975-6979.
81. Smith, L.J., X. Daura, and W.F. van Gunsteren, *Assessing equilibration and convergence in biomolecular simulations.* Proteins-Structure Function and Genetics, 2002. **48**(3): p. 487-496.
82. Lyman, E. and D.M. Zuckerman, *On the structural convergence of biomolecular simulations by determination of the effective sample size.* Journal of Physical Chemistry B, 2007. **111**(44): p. 12876-12882.
83. Hess, B., *Convergence of sampling in protein simulations.* Physical Review E, 2002. **65**(3).
84. Christen, M. and W.F. Van Gunsteren, *On searching in, sampling of, and dynamically moving through conformational space of biomolecular systems: A review.* Journal of Computational Chemistry, 2008. **29**(2): p. 157-166.
85. Boczko, E.M. and C.L. Brooks, *Constant-Temperature Free-Energy Surfaces for Physical and Chemical Processes.* Journal of Physical Chemistry, 1993. **97**(17): p. 4509-4513.

86. Priyakumar, U.D. and A.D. MacKerell, *Role of the Adenine Ligand on the Stabilization of the Secondary and Tertiary Interactions in the Adenine Riboswitch.* Journal of Molecular Biology, 2010. **396**(5): p. 1422-1438.
87. Berman, H.M., et al., *The Protein Data Bank.* Nucleic Acids Research, 2000. **28**(1): p. 235-242.
88. Lipari, G. and A. Szabo, *Model-Free Approach to the Interpretation of Nuclear Magnetic-Resonance Relaxation in Macromolecules .1. Theory and Range of Validity.* J. Am. Chem. Soc., 1982. **104**(17): p. 4546-4559.
89. Nordstrand, K., et al., *NMR structure of Escherichia coli glutaredoxin 3-glutathione mixed disulfide complex: Implications for the enzymatic mechanism.* Journal of Molecular Biology, 1999. **286**(2): p. 541-552.
90. Sodano, P., et al., *Sequence-Specific H-1-Nmr Assignments and Determination of the 3-Dimensional Structure of Reduced Escherichia-Coli Glutaredoxin.* Journal of Molecular Biology, 1991. **221**(4): p. 1311-1324.
91. Ben-Shem, A., et al., *The Structure of the Eukaryotic Ribosome at 3.0 angstrom Resolution.* Science, 2011. **334**(6062): p. 1524-1529.
92. Klinge, S., et al., *Crystal Structure of the Eukaryotic 60S Ribosomal Subunit in Complex with Initiation Factor 6.* Science, 2011. **334**(6058): p. 941-948.
93. Rabl, J., et al., *Crystal Structure of the Eukaryotic 40S Ribosomal Subunit in Complex with Initiation Factor 1.* Science, 2011. **331**(6018): p. 730-736.
94. Salis, H.M., E.A. Mirsky, and C.A. Voigt, *Automated design of synthetic ribosome binding sites to control protein expression.* Nature Biotechnology, 2009. **27**(10): p. 946-U112.
95. Liang, J.C., R.J. Bloom, and C.D. Smolke, *Engineering Biological Systems with Synthetic RNA Molecules.* Molecular Cell, 2011. **43**(6): p. 915-926.
96. Win, M.N. and C.D. Smolke, *A modular and extensible RNA-based gene-regulatory platform for engineering cellular function.* Proceedings of the National Academy of Sciences of the United States of America, 2007. **104**(36): p. 14283-14288.
97. Delebecque, C.J., et al., *Organization of Intracellular Reactions with Rationally Designed RNA Assemblies.* Science, 2011. **333**(6041): p. 470-474.