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

MOLECULAR MODELING AND THERMODYNAMICS SIMULATION OF NUCLEIC ACIDS

You Xu

Karolinska Institutet

Stockholm 2016

Published by Karolinska Institutet.
Printed by Eprint AB 2016

ISBN 978-91-7676-488-6

Molecular modeling and thermodynamics simulation of nucleic acids

THESIS FOR DOCTORAL DEGREE (Ph.D.)

By

You Xu

Principal Supervisor:
Prof. Lennart Nilsson
Karolinska Institutet
Department of Biosciences and Nutrition

Co-supervisor(s):
Prof. Roger Strömberg
Karolinska Institutet
Department of Biosciences and Nutrition

Opponent:
Prof. Richard Lavery
Université de Lyon 1, France
Institut de Biologie et Chimie des Protéines

Examination Board:
Prof. Göran Widmalm
Stockholms universitet
Department of Organic Chemistry

Prof. Alexander Lyubartsev
Stockholms universitet
Department of Materials and Environmental Chemistry
Division of Physical Chemistry

Prof. Erik Lindahl
Kungliga Tekniska högskolan
Skolan för teknikvetenskap

Through all the lying days of my youth
I swayed my leaves and flowers in the sun…
W. B. Yeats

To our time

ABSTRACT

Nucleic acids participate in many cellular processes. DNA is responsible for gene heredity and its structure is mainly in double helix, whereas RNA has wide functions in gene transcription and regulation so its structures are varied among species. RNA modifications which are known for their abundance and chemical diversity further increase the conformational variability. Functions of some RNAs closely tie to modifications. For example, modified nucleotides maintain correct tRNA structure so that enzyme and ribosome can recognize the tRNA in protein translation. Few epigenetic modifications are also found in DNA, such as 5-methyl cytidine. More often artificially modified DNA, like locked nucleic acid (LNA), is applied to alter the binding affinity of DNA duplex and triplex. Starting from the structures solved by experiments or modeled by programs, molecular dynamics (MD) simulations are employed to mimic the dynamic process and compute the thermodynamic properties, so that the structure and function of nucleic acids can be better understood. This thesis covers computational studies of both RNA and DNA structures.

In paper I, the naturally modified ribonucleotides are parameterized in an additive CHARMM force field. The parameters are targeted on quantum chemistry data. The charge and dihedral parameters are fine-tuned for some molecules to reproduce the experimental conformation. This force field allows wider computational studies on modifications involved RNA molecules.

In paper II, the new force field is used in the simulations of four tRNAs. The results show with modifications the structural stability, nucleotide conformation and base pair maintenance are almost better than those without modifications, especially in dihydrouridine loop and anticodon loop. The enhanced stability by magnesium ions is also observed.

In paper III, MD simulations combined with electrophoretic mobility shift assay illustrate the LNA effects in DNA helical structures. The results show LNA substitutions in duplex strand or the third strand improve the triplex formation, because LNA pre-organizes the DNA strands to reduce their structural adaption required upon triplex forming.

In paper IV, a method is developed to calculate free energy for LNA. The angle energies are transformed to convert the locked ribose to deoxyribose. The protocol can be in one-step or three-step by transforming bonded and nonbonded energies separately. Both protocols solve the reasonable solvation free energy and are expected to be applied in larger systems.

LIST OF SCIENTIFIC PAPERS

I. Y. Xu, K. Vanommeslaeghe, A. Aleksandrov, A. D. MacKerell Jr., L. Nilsson. Additive CHARMM force field for naturally occurring modified ribonucleotides. *J Comput Chem*. 2016, 37, 896-912.

II. Y. Xu, A. D. MacKerell Jr., L. Nilsson. Structural effects of modified ribonucleotides and magnesium in transfer RNAs. *Bioorg Med Chem*. 2016, 24, 4826–4834.

III. Y. V. Pabon[§], Y. Xu[§], A. Villa, K. E. Lundin, S. Geny, C. Nguyen, E. B. Pedersen, P. T. Jørgensen, J. Wengel, L. Nilsson, C. I. E. Smith, R. Zain. LNA effects on DNA binding and conformation: from single strand to duplex and triplex structures. manuscript.

IV. Y. Xu, A.Villa, L. Nilsson. The free energy of locking a ring:changing a deoxyribonucleoside to a locked nucleic acid. *J Comput Chem*. In revision.

[§]Joint first authors

CONTENTS

LIST OF ABBREVIATIONS

AAS	Amino acid stem (tRNA)
aa-tRNA	Aminoacyl-tRNA
ASL	Anticodon stem loop (tRNA)
BAR	Bennett Acceptance Ratio
bp	Base pair
DNA	Deoxyribonucleic acids
DSL	Dihydrouridine stem loop (tRNA)
EF-Tu	Elongation factor Tu
ESP	Electrostatic potential
FEP	Free energy perturbation
GTP	Guanosine triphosphate
H-bond	Hydrogen bond
HG	Hoogsteen (base pair)
HS	Hirsh suppressor
L-J	Lennard-Jones (expression)
LNA	Locked nucleic acid
MD	Molecular dynamics
MM	Molecular mechanics
mRNA	Messenger RNA
ncRNA	Non-coding RNA
PES	Potential energy surface
PNA	Peptide nucleic acid
QM	Quantum mechanics
RNA	Ribonucleic acids
rRNA	Ribosomal RNA
TFO	Triplex forming oligonucleotide
TI	Thermodynamic integration
TP	Thermodynamic perturbation
tRNA	Transfer RNA
$tRNA_f^{Met}$	Initiator tRNA (in bacteria)

TSL	TΨC stem loop (tRNA)
vdW	van der Waals (interaction)
WC	Watson-Crick (base pair)

1 INTRODUCTION

Nucleic acids are a family of macromolecules. They consist of deoxyribonucleic acids (DNA) and ribonucleic acids (RNA) and they are essential for life.

DNA stores genetic information. It has four building blocks (adenosine, cytidine, guanosine and thymidine) which only differ in bases (adenine, cytosine, guanine and thymine). DNA mainly exists as a double helix (duplex) where the two strands with complementary sequences bind to each other. This allows an identical daughter DNA to be replicated by taking either strand as template. RNA has almost the same bases: adenine, cytosine, guanine and uracil, but except for virus, RNA does not store genetic information. Instead they have two functional categories: one to transcribe the genetic codes from DNA strand (coding RNA) and one to regulate the gene expression (noncoding RNA). Depending on the species, they have more complicated structures than DNA, which often makes challenge to their structural determinations.

The ribosome is a complex of protein and RNA, which is responsible for translating genes to proteins. One of its delicate functions is tRNA recognition on which the correct peptides are synthesized. The ribosome selects tRNA by not only matching the mRNA codon, but also the conformational adaption of tRNA. One example is the Hirsh suppressor,[1] in which a distant mutation from anticodon in $tRNA^{Trp}$ suppresses a stop codon by miscoding. This is probably because the altered local conformation allows the mutant tRNA to be tolerated by the ribosome even though there is a mismatch between codon and anticodon. A large number of modified nucleotides are found in tRNA, and although their functions in many cases are not known, they carry a significant genetic and energetic cost for the organism and therefore are important. They influence decoding not only by altering the base pairing with the codon triplet but also by organizing the tRNA architecture for particular ribosomal interactions.

In contrast, neither abundant modifications nor varied structural features are found in DNA, since DNA exists only for carrying genes. DNA duplexes are binding targets for proteins and other molecules. One case is the DNA duplex can accommodate oligonucleotides to form a stable triplex. This feature is fascinating because people can design a nucleic acid “ligand” to selectively bind to a sequence and later regulate gene expression by disturbing the target duplex. Artificial modifications, *e.g.* peptide nucleic acids (PNA)[2] and locked nucleic acids (LNA)[3,4] are often used in oligonucleotides to improve the binding affinity of triplex formation.

Three experimental methods are used to solve nucleic acid structures. X-ray diffraction can solve very big complexes (*e.g.* ribosome) in good resolution, as long as the sample can be purified and crystalized. Nuclear magnetic resonance (NMR) spectroscopy only needs the purified sample in solution, but it becomes more difficult as the system size increases. Cryo-electron microscopy (cryo-EM) collects structural information from specimens without extra constraints, and it is getting more popular as its resolution improves steadily. Nowadays the experimental techniques can solve the structure of nucleic acids as complicated as ribosome.

Molecular dynamics (MD) simulations can be employed to mimic the structural motion and compute the related properties of a molecular model, based on which the function can be further understood. It generates conformations that are connected in time, and follow the Boltzmann distribution. From a (large) set of such conformations it is possible to compute thermodynamics properties. It uses classical mechanics and the energy is expressed in force field. A force field has formulated energy terms, and the parameters for different energy terms between two atom types are defined beforehand. There are several programs capable of MD simulation, such as AMBER[5], CHARMM[6,7] and GROMACS[8]. In this thesis CHARMM and its force fields are used.

To perform MD simulation on RNA molecules containing modifications, an additive force field of modified nucleotides is developed in paper I. Compared with the previous AMBER one,[9] this force field optimized both atomic charges and molecular conformation, and experimental data are also used to refine the parameters. In paper II the newly developed force field is employed to investigate the modification effects on tRNA structures. Those studies provide the foundations for future work, *e.g.* investigating the thermodynamic features of ribosomal mismatch tRNA such as Hirsh suppressor. In paper III the enhanced triplex formation by LNA substitutions in DNA single strand and duplex are investigated in collaboration with clinical chemistry colleagues. Further addressing the LNA effects on binding affinity requires free energy perturbation (FEP). To overcome the technical limitation of transformation for bridged molecules like LNA,[10] a method is proposed in paper IV.

In Chapter 2, a short review about the field of nucleic acids is given. This includes the structural features of nucleic acids, the ribosomal tRNA recognition, RNA modifications and the application of synthesized nucleotides in DNA helical structures. In Chapter 3, an introduction of MD simulation is given. This includes the major physical principles and techniques of this method, and details on force field and free energy calculation. In Chapter 4, the four papers in the thesis are summarized.

2 NUCLEIC ACIDS

Nucleic acids are one of the most abundant biomolecules in lives. Their functions cover wide bioprocesses related to genes, from genetic heredity to protein production and regulation. This chapter first introduces the general structural features of nucleic acids including building blocks and the secondary and tertiary structures, and then introduces some specific topics of the ribosomal tRNA recognition, the RNA modifications and the artificial nucleic acids applied in DNA structures.

2.1 STRUCTURAL FEATURES

Nucleic acids are folded from one or several strands of oligonucleotide, while the latter is linearly polymerized from nucleotides through phosphodiester bonds. They are negatively charged molecules and varied in structure and size.

2.1.1 Nucleotide

Nucleotides are the building blocks of nucleic acids. A nucleotide consists of a base, a sugar (ribose in RNA and deoxyribose in DNA) and a phosphate. The bases are adenine (A), cytosine (C), guanine (G) and thymine (T) in DNA, and with T replaced by uracil (U) in RNA.

2.1.1.1 Backbone

The backbone of RNA or DNA nucleotide is counted from P to O3'. In polymerization one nucleotide's O3' forms a phosphodiester bond with the other's P by losing a H_2O, where the first nucleotide has a free 5'end and the second has a free 3'end, both are extensible by new nucleotides. In an oligonucleotide chain, the direction is presented from 5'end to 3'end (Figure 1).

A nucleotide has six backbone torsions, denoted as: α: O3'-P-O5'-C5', β: P-O5'-C5'-C4', γ: O5'-C5'-C4'-C3', δ: C5'-C4'-C3'-O3', ε: C4'-C3'-O3'-P and ζ: C3'-O3'-P-O5'. Except for δ, which is restricted due to the ring tension, the other five torsions are free to rotate. However in ordered nucleic acids, each torsion stays in a specific range in DNA and RNA,[11] and one backbone torsion is usually correlated to another and sugar pucker, which is kind of structural fingerprint of nucleic acid tertiary structures for force field parametrisation.[11-13]

2.1.1.2 Nucleobase

Bases which include purine (A and G) and pyrimidine (C, T and U) are aromatic hetero rings with planar geometries. In a polynucleotide the sequence of base is the primary structure of nucleic acids. Due to the specific arrangement of hydrogen bond (H-bond) donors and acceptors, base pairs can be formed in several ways. For instance, A and T can form a base pair which has the same shape as a G:C base pair (Figure 1), and these Watson-Crick (WC) pairs are fundamental for the formation of DNA double helix. WC base pairs guarantee the fidelity for gene replication and transcription from DNA, where one sequence can only match

the other specifically. Other types of base pairs widely exist in RNA and nucleic acid complexes, but they are not able to extend a double helix. The primary structure folding through the formation of WC base pairs defines the secondary structure of nucleic acids.

In polynucleotide the stacking due to π-π interaction between neighbor bases is more significant in stabilizing a helical structure than H-bonds.[14] The systematic *ab initio* calculations of base-base interactions has been reviewed.[15] Not as H-bond, stacking does not have matching preference, but its strength depends on base types where purine provides stronger interaction than pyrimidine. Bases can be modified endogenically, especially in RNA so that interaction patterns with other bases are altered. This is an important strategy of life to evolve the new macromolecular structures and functions.

The base covalently connects to sugar through the glycosidic bond (χ) which is the only degree of freedom of base orientation. Although χ is free to rotate over 360°, only two low energy regions, *syn* (30°−90°) and *anti* (170°−330°), are of interest, for in other regions the repulsive interactions between ribose and base are too big.[16] Base orientation plays an important role in nucleic acid structures.

2.1.1.3 Sugar

The ribose is a five-membered sugar, where C1' connects to the base and C5' connects to the phosphate. All natural nucleic acids have β-ribose with O2' and O3' on one side of the sugar and C5' on the other side (Figure 1). Due to ring tensions, one or two atom(s) are outside the plane of the other ring atoms, causing sugar pucker. The most stable pucker is *C2'endo* or *C3'endo*, identified by if C2' or C3' is on the same side as C5' respectively. The geometry of pucker is described by the pseudorotational phase angle (P) and puckering amplitude.[17] The angle P is ranged in 0−360°, and each quarter is also denoted as *north*, *west*, *south* and *east*. In most literature the puckers *north* and *south* are equivalent to *C3'endo* and *C2'endo* respectively. Pseudorotation is the main descriptor of sugar conformation. The correlation between glycosidic torsion and pucker has been calculated using quantum chemistry and was parameterized in force field.[13,18,19]

The 2'-OH is the hallmark of RNA which makes essential difference from DNA, including: 1) 2'-OH is a chemically active site of modification whereas almost no modification happens in deoxyribose; 2) 2'-OH introduces different electrostatic interactions for nucleotides so RNA usually has *north* pucker whereas DNA has *south*; 3) 2'-OH has attractive intra-nucleotide interactions with O3', O4' and the base, or to O4' in 3'-nucleotide, which are related to RNA structural variety.[20]

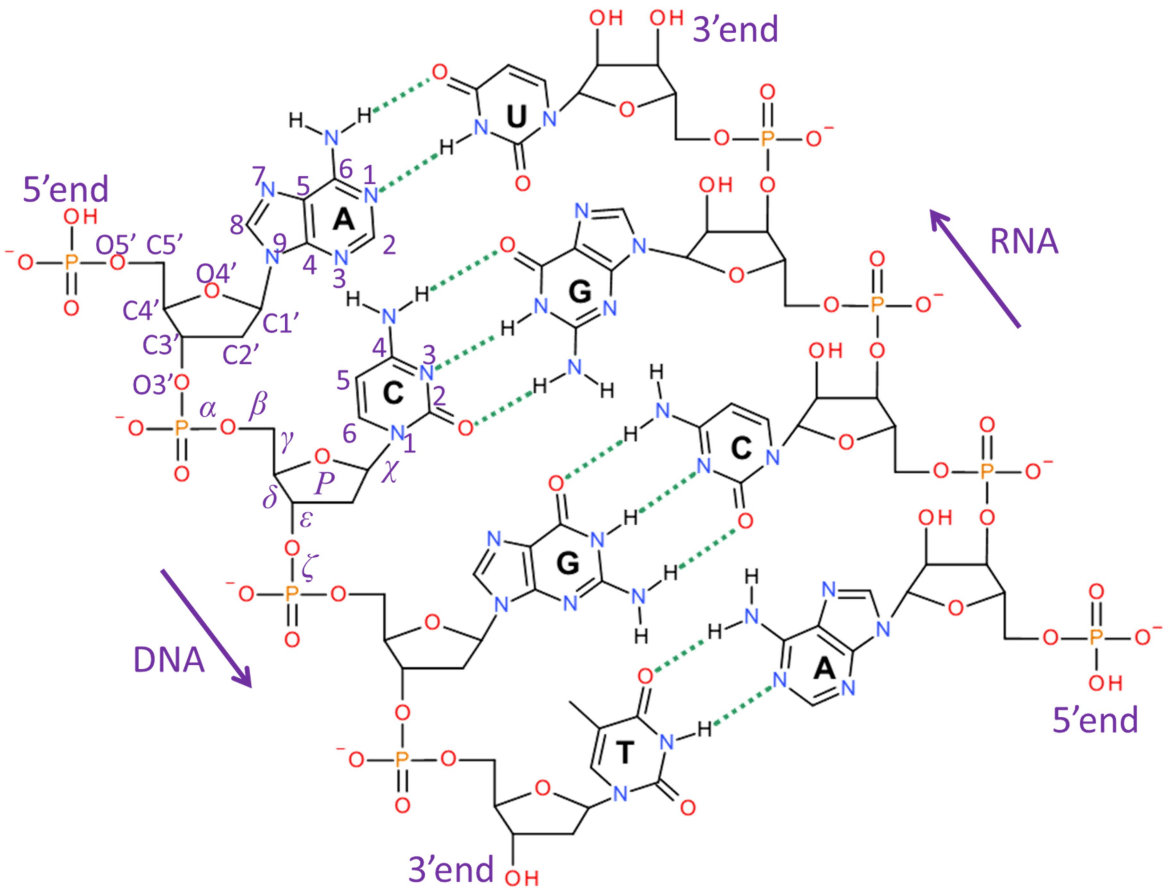

Figure 1. The structural basics of nucleic acids. The primary structures of four nucleotides are shown for a DNA strand and an RNA strand. The two strands with complementary sequences form the secondary structure through four WC base pairs, where the H-bonds are shown as green dashed lines. The arrows of the sequence direction show two strands are anti-parallel. The atom numbering of nucleoside is labeled on first two nucleotides in DNA strand and the torsions are labeled on the second nucleotide.

2.1.2 Helical structures

2.1.2.1 Duplex conformation

The double helix, or duplex, is the ubiquitous tertiary structure of DNA. As a key to understanding genes, the discovery of DNA double helix is the milestone of life science (Double helix: 50 years of DNA, http://www.nature.com/nature/dna50/archive.html).

A duplex is formed by two anti-parallel strands with complementary sequence (Figure 1), and stabilized by H-bonds within WC base pairs, stacking between base pairs, and cations in solvent as well. Due to the torsion flexibility, duplexes exist in different conformations such as A-type, B- type and Z-type. DNA is predominant in B-type in physiological condition, with a wide major groove and narrow minor groove. Its complete turn has 10.5 base pairs, with a 33-Å pitch and a 20-Å diameter. All bases are in high *anti* and puckers are in *south* conformation. Each base pair is perpendicular to the helical axis hence visually B-DNA is straight (Figure 2).

B-DNA can convert to A-DNA in dehydrated samples or upon binding of some proteins[21]. The A-type helix has 11 base pairs per turn, and the base pairs are not perpendicular to the

helical axis but with an inclination and a roll.[22,23] As a consequence the pitch and diameter of a turn are 28 Å and 23 Å respectively, so visually it is shorter but wider than B-type. All bases are in low *anti* and puckers are in *north* conformation. Compared with B-type, it has narrow but deep major groove, and wide but shallow minor groove.

Both B- and A-type are right-handed helix, whereas Z-type is left handed. It exists in high salt concentration and the sequence of alternating CG is preferred. It has a special conformation where the cytidines are in *anti* and *south* and guanosines are in *syn* and *north*. Z-DNA needs two base pairs as a repeating unit and its complete turn needs total 12 base pairs, and the pitch and diameter are 46 Å and 18 Å respectively. It therefore looks long and slim.

2.1.2.2 Base pair parameters

The parameters of base pair (bp) and base step were defined to characterize the geometry of double helix.[23,24] Conventionally 16 descriptors in three categories are adopted. The first group defines bp-axis geometries using 4 descriptors: x-displacement, y-displacement, inclination and tip. The second group defines intra-bp geometries using 6 descriptors: shear, stretch, stagger, buckle, propeller and opening. The third group defines inter-bp geometries using the rest 6 descriptors: shift, slide, rise, tilt, roll and twist. In each group the first half is distance scalar in Ångström and the second half is angle scalar in degree. The bp parameters are able to characterize all kinds of helical conformations including A-, B-, and Z-helix.

2.1.2.3 Grooves

Grooves exist between two backbones in a duplex. Either major or minor groove surface consists of phosphate, ribose and one edge of base. Two grooves are different in both size and surface potential, and their dimensions are considerably varied among the conversion of helical types (Figure 2). Because the edges toward major groove (Hoogsteen edge) of four bases are all different whereas in minor groove the pyrimidine edges are same, the major groove therefore is more sequence specific to be a target of selective binding. The DNA-binding proteins are known to embed in major groove of a particular sequence and be stabilized by the interaction with minor groove.[25] In addition to bp parameters, the groove dimension is another important feature to characterize the helical conformation,[26] for it is more intuitive for evaluation of complex binding and ligand design.

Two software, 3DNA[27,28] and Curves+[29,30], are normally used to analyze the helical structure of nucleic acids.

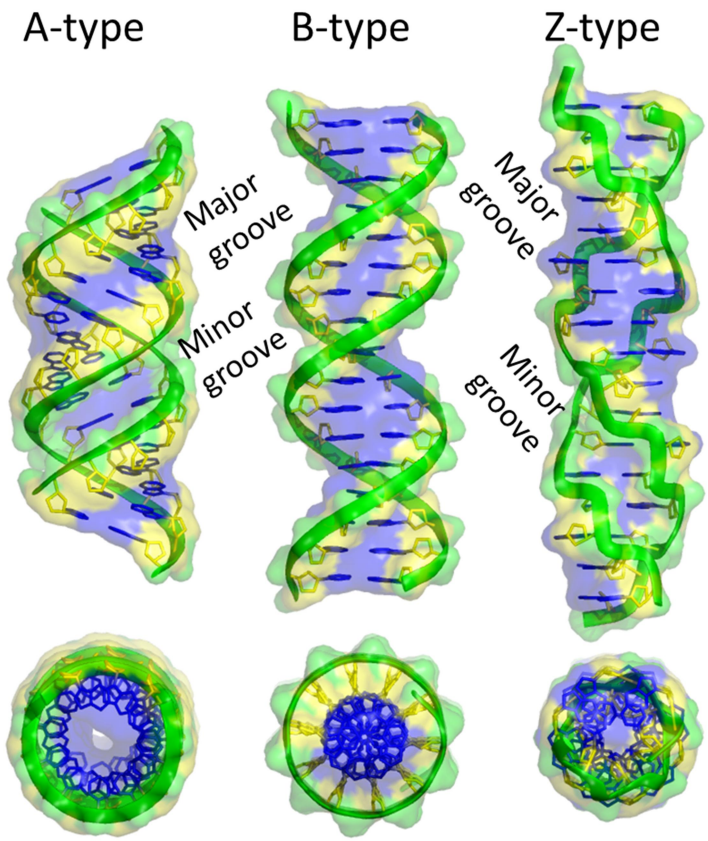

Figure 2. The duplex 3D structures. The A-, B- and Z-type helixes with 17mer sequence are shown in front and top views. The structures were generated using w3DNA[31]. The backbones are in green ribbon, and the sugars and bases are in yellow and blue sticks respectively. The solvent accessible surface is shown in the same color scheme and the grooves are labeled.

2.1.2.4 Triple helix

The duplex major groove is able to accommodate an oligonucleotide thereby forming a triple helix (triplex).[32] This third strand with a specific sequence that selectively targets the duplex is called triplex forming oligonucleotide (TFO).[33] The binding region of the duplex must have a homo-purine strand and a homo-pyrimidine strand, so that the TFO can bind to the purine strand (central strand) by forming Hoogsteen (HG) base pairs. As the rule of HG, central purine A matches T or A, and G matches $^{+}$C and G. The binding of TFO sequence is not unique. Homo-purine TFO is antiparallel with the central purine strand by forming reverse HG base pairs and homo-pyrimidine TFO is parallel by forming HG base pairs. Mixed purine-pyridine TFO is allowed the sequence containing only G and T, and can bind in either direction.[34,35] Once B-DNA accommodates a TFO, the conformation will shift toward A-type upon triplex formation.[36] Finally since cytosines have to be protonated, the triplex formation for C rich TFO requires pH < 6; the triplex formation therefore is weaker than duplex in physiological condition.

2.1.2.5 G-quadruplex

G-quadruplex is folded from one or more G rich sequences. It is stabilized by the intra-strand G tetrad, where four guanines interact with ether other through four sets of HG H-bonds in a

square and planar geometry. A potassium ion is often present in the center coordinating with four guanine O7 atoms to stabilize it. Multi tetrads are required so each other is stabilized through stacking. Depending on the length and guanine position of the strands, both parallel and antiparallel topologies can be formed between neighbor strands. The intra-strand G-quadruplexes are frequently found in human telomeres, which protect the telomeres from degrading by telomerase.

2.1.3 RNA structures

In addition to the information flow indicated by the central dogma, RNAs are discovered in many cellular processes including decoding, regulation and expression. One type like messenger RNA (mRNA) whose sequences are transcribed from gene and translated to protein is coding RNAs. Correspondingly the other type whose sequences have no relationship with gene is non-coding RNAs (ncRNA). The ncRNA is a big family, in which some trigger RNA modifications, like small nucleolar RNA; some regulate gene expression by slicing mRNA, like microRNA; and some are building blocks for a complicated enzyme, like ribosomal RNA (rRNA). Till now there are still a lot of RNAs whose functions and encoding genes are unknown.

To be functionalized in many fields, RNA structures are required to be diverse. Such diversity is not only given by the active 2'-OH group, but also because RNA tertiary structure is almost folded from a single strand rather than two complementary strands. As the single sequence is usually not self-complementary, only parts of local sequences can form WC base pairs in the secondary structure, and the unpaired bases have indefinite conformational contribution in RNA architecture.

2.1.3.1 Stem and loop

Helical structures are also ubiquitous for RNA. Complimentary RNA sequences form A-type duplex and rarely B-type. RNA duplexes are abundant in small interfering RNAs and in RNA virus. In other cases RNAs duplex is shown as a short stem in single-stranded RNAs. Bulge which is formed by a few unpaired bases in one strand is tolerable for a long stem.

If 3−8 unpaired nucleotides are left at one end of stem, they will form a loop. The stem-loop or hairpin is a common RNA secondary structure. Some loop motifs have predominant population, and because of the particular intra-loop interactions the conformation of those loops are conserved. An example is U-turn with a motif UNR (N is any nucleotide and R is purine), where U forms H-bonds with atoms 3'-P and N7 of R. U-turn looks sharp but it is well organized and stable, and it shows up twice in a transfer RNA (tRNA). Tetraloop is also common in hairpins where some motifs are extremely abundant, like GNRA.[37] In this tetraloop G formed three H-bonds with atoms 5'-P, N7 and N6 of A, and one H-bond with N7 of R, which is quite stable and conformationally equivalent to a U-turn.[38]

2.1.3.2 Transfer RNA

tRNA exists in all life domains and is considered as the most ancient molecule in the evolution of life. A specific tRNA has a unique three-base anticodon which matches the complementary codon on mRNA in ribosome, where the peptide chain is synthesized according to the mRNA sequence.

A tRNA consists of 76–96 nucleotides. Its secondary structure is cloverleaf shaped and has five subdomains: the acceptor stem (AS) has a 7-bp stem including 5' and 3' termini and a conserved tail CCA for amino acid attaching; the dihydrouridine stem loop (DSL) has a 4-bp stem and a 6–11-nt loop where the conserved dihydrouridine (D) is present; the anticodon stem loop (ASL) has a 5-bp stem and a 7-nt loop where the anticodon triplet is located; the variable loop (VL) in abundant type I tRNA is a short 4–5-nt loop whereas in type II tRNA is a 24-nt hairpin;[39] the thymidine stem loop (TSL) has a 5-bp stem and 7-nt loop where nearly conserved TΨC motif is present. A junction is formed among those five subdomains (Figure 3).

All tRNAs fold in L-shaped tertiary structures, in which DSL contacts TSL and VL as the elbow and ASL is left along as the arm. The important elbow interaction is formed by seven base-pair or base-triplet layers. Base variants in the seven-layer exist, but this interaction pattern is conserved in all tRNA.[39,40] A species of tRNA has specific anticodon and conformation, both are essential to correct tRNA recognition in ribosome. One significant feature of tRNA is the frequent modifications which alter the secondary structure and increase the chemical diversities for tRNAs. Furthermore Mg^{2+} ions are important to maintain the conformation for functionalized tRNA. In single molecule or bound into ribosome, several backbone turns (*e.g.* in DSL) and base pairs (*e.g.* in ASL) of tRNA are stabilized by Mg^{2+}.[41-43]

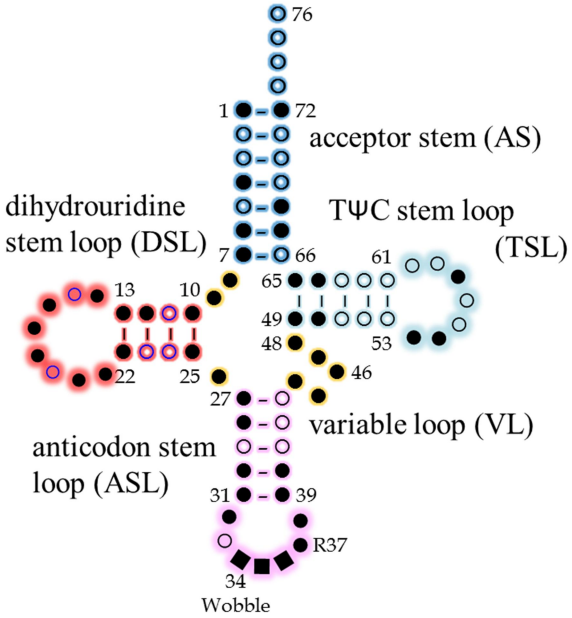

Figure 3. The cloverleaf secondary structure of tRNA (shorter VL). The circle represents a nucleotide, while square is particularly for anticodon nucleotides. The filled symbol means modifications have been found in this position. Dashes in the stems indicate WC base pairs. Each subdomain is labeled.

2.2 TRNA IN RIBOSOMAL DECODING

The gene transcribed from DNA is translated into proteins in ribosome. Ribosome is an accurate machine where each amino acyl tRNA (aa-tRNA) is identified according to the codon in mRNA.

2.2.1 The genetic code

A codon triplet consists of three adjacent bases in mRNA sequence which encodes an amino acid. This implies 20 amino acids are encoded from 64 codons, resulting in codon redundancy. Except for methionine and tryptophan which only have one codon each, codon degeneracy exists in the rest 18 amino acids, from two- to six-fold, but excluding five-fold. For the 15 from the 18 amino acids which have two- to four-fold codon degeneracy, the codons of each amino acid only differ at the third position; particularly for all two-fold codons, the third bases are either both purines or both pyrimidines.

As proposed in the wobble hypothesis,[44] the first two positions must form strict WC base pairs with the anticodon, and wobble base pair (G:U or I:U) is tolerable for the third one. That original wobble base pair illustrated how two-fold codon is read by one anticodon: a guanine or inosine (I) can match a uracil by forming still two H-bonds as A:U, but a shifted geometry from WC base pair is required. Later it was extended as the delicate modifications in first position of anticodon enable this modified base match any four bases, so the three- or four-fold codon can be recognized by one anticodon.[45] As the first two positions do not allow wobble pair, the six-fold codon will require at least two anticodons.

Codon degeneracy improves the fault toleration for point mutation of gene, *e.g.* the mutation in the third position of four-fold codon will not cause fault protein synthesis; and the transition mutation will not cause fault for two-fold codon either.

2.2.2 tRNA recognition

The tRNA recognition is understood through the comprehensive knowledge of ribosomal translation which was accumulated from kinetic and structural data in the last decades.[46-49]

2.2.2.1 Proofreading

In bacteria, the translation is initiated by $tRNA_f^{Met}$ and elongated by aa-tRNAs. After the initiator tRNA, the process and the involved enzymes for each elongator aa-tRNA are the same. Every time an aa-tRNA combines with an elongation factor EF-Tu and an energy supplier GTP, forming a tertiary complex. The complex temporarily contacts the ribosomal A site, staying in A/T state. Here the tRNA has a 30°-bent conformation, where its ASL is in the A site but the elbow and acceptor which bound to EF-Tu and GTP are outside the A site. The anticodon and codon triplet form a mini duplex, where the first two base pairs are monitored by A1492, A1493 and G530 from 16S rRNA. The cognate match (strict WC base pair for first two codon positions) forms eight H-bonds between tRNA and ribosome whereas the non-cognate match (first two are not all in WC bp) has fewer than eight. This H-bond network is very sensitive to tRNA conformation. The cognate interaction makes a unique ASL conformation that just fits the H-bond network, whereas other ASL conformation in non-cognate interaction leads to large dissociation constant for tRNA.[45-49] The first codon:anticodon base pair is also stabilized by Mg^{2+} ions and tRNA modified purine 37.[50]

As G:U pair, which has same H-bond number but shifted geometry as A:U, causes non-cognate interaction at first two codon positions, the mechanism of proofreading is thus working by 'shape' recognition rather than counting H-bonds between anticodon and codon. Indeed, if only one species of non-cognate tRNAs is provided, the ribosome forces the G:U pair in WC geometry to fit the cognate pocket shape, even though the normal H-bonds between G and U cannot form.[51]

After proofreading the cognate tRNA induces a 'closed' conformation of the decoding center which triggers the GTP hydrolysis and EF-Tu dissociation, so that the tRNA relaxes the bend and its elbow and acceptor part move into the A site. On the other hand the non-cognate tRNA is rejected and replaced by another aa-tRNA. This proofreading is precise with an error rate of 10^{-3}–10^{-4}.

2.2.2.2 Hirsh suppressor

The Hirsh suppressor is a mutant $tRNA^{Trp}$ which is the first evidence that H-bonds in codon:anticodon base pair is not the only criterion in tRNA recognition. Wide type $tRNA^{Trp}_{CCA}$ reads only one cognate codon UGG. A distant mutation G24A in DSL however makes $tRNA^{Trp}_{CCA}$ decode UGA[52] or UGU[53] thus suppressing the stop or cysteine codon. A recent experiment measured the dissociation constants of Hirsh suppressor and supposed that the mutation allows more flexibility for $tRNA^{Trp}$ to adapt the A/T state, so that accelerates the GTP hydrolysis rate which helps tRNA accommodation in ribosome.[54] This mutation also enhances the interaction between tRNA and 50S subunit, preventing the tRNA away from the

ribosome.[54,55] In crystal structure, the bent conformation of A/T state was stabilized by an additional internal interaction between A24 and G44, thereby improving surviving time in A site for the miscoding tRNA[56]. As $tRNA^{Trp}$ reads only one cognate codon without the modification in anticodon, Hirsh suppressor is a good model to further investigate the mechanism of tRNA recognition.

2.3 RNA MODIFICATIONS

At least 112 naturally modified nucleotides have their structures known and they are very common in RNA (The RNA Modification Database: http://mods.rna.albany.edu/).[57,58] RNA modifications are involved in energy metabolism related diseases, and also tumors and neurodegenerative diseases.[58] The biological effects of these RNA modifications on the mechanisms of disease are not well known, but for several cases of mitochondrial pathogeneses the hypo-modification in tRNA leads insufficient protein synthesis.[59]

Generally modifications enlarge the RNA structural repertoire and contribute to specific interactions in well-defined 3D conformations. Such structures often tie to special functions and are involved in regulation of cellular processes while sensing the cell's metabolic state.[57,58] The modified nucleotides participate in interactions between tRNA and other partners in protein translation, such as synthetases, ribosomes, mRNA, initiation and elongation factors.[40,57,59-61]

2.3.1 Diversity and nomenclature

The current modifications are composed by more than 70 chemical diversities on base and 2 on ribose. Some modifications are chemically same but exist in different nucleotides, such as 2'O-methylation. The overall number of RNA modification is shown in Table 1. These modifications are introduced by a variety of enzymes and RNAs, with a significant energetic and genetic cost.[62] Although many modifications are shared inter-domains, clear differences are still among archaea, prokaryotes and eukaryotes, which is interesting to understand the life evolution[58,62,63]

Table 1. The number of known natural modification and the substituting positions.

Type	Quantity	Modification sites	
		Abundant	Others
Adenosines	26	N6	C2, N1, C8, & O2'
Cytidines	15	N4	C5, C2, N3 & O2'
Guanosines	27	N7/C7[a]	N1, N2, C2, N3 & O2'
Uridines	44	C5/N1[b]	C2, C4, N3 & O2'

a. Shown for guanosine/7-deazaguanosine
b. Shown for uridine/pseudouridine

The chemical names of modification are adopted from IUPAC, but there is no standard one- or three-letter code as canonical nucleotides. The convention of the extended one-letter code is used in literature:[58] the base type is in capital letter and the modifications on base and

ribose are shown in the left and right respectively; the chemical group is in one or two lower letters, following which the superscript means the position of the modified group and the subscript is the number of this modified group. There is no consistence for three-letter code which is used in the PDB (Protein Data Bank: http://www.rcsb.org) coordinate files and force field topology files. AMBER force field has suggested[9] that the final letter is for base type and first two letters are picked from the chemical name of the modification. The full modifications are shown in Figure 3 of paper I and list of modification names is in Table S1.

2.3.2 tRNA modifications

Various modifications maintain the physiological functions of tRNA. Although the direct contribution of a specific modified nucleotide is hardly attributed to a biological process, by modulating the enzymes' expression to suppress the modification levels, the abnormal cellular regulations are observed.[58,59,64] The known human pathologies that tie to the modification defects in tRNA include neurological, cardiac, respiratory, cancer, metabolic and mitochondrial-linked diseases.[63]

At least 93 modifications are found in tRNA, which covers 83% of the total number. More than 40 positions of tRNA sequence have been reported to be modified and one tRNA on average has 8 positions modified (Figure 3).[39,58,61] Some modifications are conserved while some are species dependent;[45,60,65-68] some in ASL facilitate codon degeneracy thereby contributing to efficient decoding,[45,57,59,60,69-71] and some others are crucial to alter or keep the particular local structure.[40,41,72] The modifications in tRNA are attributed into three function classes:[58] 1) alter/adjust anticodon matching rules; 2) control tRNA folding; and 3) as an identification of tRNA. The structural effects of some conserved modifications are especially important and more investigated.

2.3.2.1 Wobble position

The wobble position #34 is the first position of the anticodon, *i.e.* it matches the third codon position. In all tRNA species uridine 34 is most abundant and rarely unmodified. In opposite, adenosine 34 is least present. The interesting consequence is that codon U3 is often read by modified U34 rather than A34.

The wobble modification s^2U34 is often in aa-tRNA of two-fold codon, *e.g.* $tRNA^{Glu}$, $tRNA^{Gln}$ and $tRNA^{Lys}$, where it reads both A3 and G3.[45,60] It is also in pyrimidine abundant anticodon, *e.g.* $tRNA^{Lys}_{UUU}$, where it enhances the stacking between U34:A3 and U35:A2.[68] More complicated mcm^5s^2U34 (eukaryotic) and mnm^5s^2U34 (prokaryotic) might have the same functions.[66] NMR shows the charged side chain of mnm^5s^2U likely enables the uracil ionized to be 4-enol form, thereby U34:G3 is in the WC geometry as U34:A3.[73] A recent crystal structure shows a new geometry where mnm^5s^2U moves towards minor groove, by forming either 2-thiol or 4-enol form.[74] From those observations, mnm^5s^2U enables the U:G pair in various geometries, which probably reduces the conformational penalty for near-cognate interaction.

Modification cmo^{5}U34 is often shown in four-fold degenerate codon, for its ability of reading all bases in codon #3. In crystal structure the modification facilitates the enol form for U34 in U34:G3 with WC geometry as the mnm^{5}s^{2}U case, and also enables the special geometry that keeps the C1'-C1' distance of pyrimidine pair U34:U3 or U34:C3 as similar as the purine codon cases.[75] Computational study of ribosomal ASL^{Val} attributes the versatile cmo^{5}U34 ability to the lower mismatching free energy and favorable interaction with ribosomal nucleotides by modified side chains, but not the pre-structured effect on ASL.[76]

The pre-structured effect on ASL might be another contribution of wobble modifications. One example is again s^{2}U and its derivatives. The 2-thiolation enhances the *north* pucker and rigidifies low-*anti* χ compared to unmodified U,[77-80] thereby stabilizing A-type conformation for anticodon loop. The 5-substitutions of U34 can interact with its 5'-phosphate or U33 so it reduces the flexibility of A loop.[77,81-83] In summary the effects of some modified U34 are both enthalpic (codon:anticodon interaction) and entropic (pre-organization of ASL).

The similar effect is also demonstrated for modified guanosines, such as Gm[84] and Queuosine (Q)[85]. The 2'-OMe mildly enhances the *north* pucker, but it only contributes to the loop stability not the base pair pattern.[86,87] Q is a family of hypermodified 7-deazaguanosines discovered only in wobble positions. Its aldohexose- and amino acid- attached derivatives are found in bacterial and eukaryotic tRNAs respectively.[58] Although the function of those extremely complicated side chains are still unknown, the presence of Q34 in tRNA^{His} and tRNA^{Tyr} enhances the interaction with C3 and U3 against purines, so it promotes the recognition of the two-fold histidine and tyrosine codons.[58,88]

Agmatidine (C+) and lysidine (k^{2}C) in tRNA^{Ile} of archaea and bacteria respectively enable the modified C to pair with A rather than G so that the isoleucine codon AUA is read correctly against the start codon AUG.[89,90] This is a rare case that modification differentiates one purine from the other, which is necessary for tRNA^{Ile} to read the only three-fold codon. Interestingly k^{2}C34 is also necessary to be enzymatically active for isoleucine acceptance, while the unmodified tRNA^{Ile} is prone to accept methionine.[91]

2.3.2.2 Purine #37

Position 37, the 3'-neighbor of anticodon, is conserved as a purine and mostly modified. It stabilizes the codon:anticodon mini helix and maintains the reading frame.[45,57,59,60] The effect of R37 is often synergized with U34. For example, the previously mentioned cmo^{5}U34 facilitates to read four codons for tRNA^{Val}, but this ability is lost if m^{6}A37 is unmodified.[82]

Hyper-modifications (complicated chemical groups) are common on R37, which supply extra stacking interaction for the anticodon:codon N36:N1 base pair, especially for A:U/U:A rich anticodon:codon contexts. For instance, t^{6}A37 and its derivatives after U36 stabilize the base pair U36:A1 by their extended conjugate side chains;[68,73,92] i^{6}A37 and its derivatives after A36 provide strong stacking for A36:U1.[56] Furthermore the stabilization of ms^{2}i^{6}A37 in tRNA^{Phe} is observed not only in A site, but also P and E sites.[50]

Similar as the modification 34, R37 also orders the ASL conformation. These bulky modifications are often steric hindrances to prevent C32:A38 forming base pair, such opening is supposed to be necessary for an active A loop.[45,83,92,93] The modified guanosines like wyosines are not studied popularly as modified A37 yet, but some simple modifications, *e.g.* m^1G37 in $tRNA^{Phe}$, are found to reduce the mobility and organize the ASL conformation.[71] Interestingly modified A37 in free tRNA might not stabilize the base stacking as in ribosome. For example, $(ms^2)t^6A37$, stacking with A38 or pairing with U33, either bulges out U36 or bulges out U35 from the base step in A loop.[81] The conformations of t^6A37 in single tRNA are different, where the side chain is either in a tri-cyclic conformation through intra H-bonds[81,92,94] or in a linear conformation[95]. Furthermore in free *E. coli* initiator[96] $tRNA_f^{Met}$ or transformylase bound[97] $tRNA_f^{Met}$, t^6A37 is bulged outward the A loop and A36 stacks with A38 instead. This suggests the conformation of t^6A37 derivatives is environment dependent, and only when it is contacting in ribosome the stabilization is shown.

2.3.2.3 Dihydrouridine

Dihydrouridine (D) is highly conserved in one or more positions of DSL. Among all modifications, dihydrogenation is the only one that breaks the aromaticity of the base, which makes different properties from other nucleotides. Because the inability of stacking, D always exists in loop and is either solvent exposed or non-WC base-paired.[95,96,98,99] Hence the context of D is flexible. Another special feature is that D induces *south* pucker for ribose rather than *north* of common RNA.[100] The reason is likely the loss of the π-electron interaction between C5-C6 and O4'.[101,102] In oligonucleotides this conformation propagates to its 3'-neighbor, and synergized by the stacking loss, the A-type helical conformation is not kept in D present domain.[103] The reason why D exists is not conclusive. A recent NMR shows D rather than U is required to keep the DSL-like hairpin, where the *south* pucker in specific positions is important to stabilizes the loop.[104]

2.3.2.4 Pseudouridine

Pseudouridine is the first discovered modification and is probably most ubiquitous.[40] It has opposite effect to dihydrouridine, *i.e.* enhances the *north* pucker and stacking interaction and stabilizes the A-type helix.[105-108] Mono Ψ nucleoside has more rigid conformation than U even the featured interaction between N1 and 5'-phosphate is hindered.[105,109-111] The stabilization effect is seen in single oligonucleotides by NMR,[106] where Ψ significantly increases the *north* for 5'-adenosine. Although mono Ψ nucleoside can be in *syn*/*south*, in oligonucleotide it is only in *anti*/*north* conformation. The enhanced base stacking of Ψ:A is also observed on 5'-neighbors in double helix context.[112] In tRNA Ψ is mainly present in ASL and T loop, *e.g.* #27, #39 and more popular #55. They are considered to be important to tRNA folding and local conformational maintenance.[72,107,108,111,113]

2.3.2.5 Methylation

Methylation is a common and big class of post-transcriptional modifications. They are simple but important, and over 2/3 modifications are methylation or related.[114] Evolutionarily it was

hypothesized to bifurcate the gene storage and function/catalysis from RNA to DNA and protein respectively, because methyl groups lower the activation energy of biomolecules.[115] Chemically the methyl group changes the molecular size and hydrophobicity, as well as the base aromaticity. The molecular interactions, such as base pairing/stacking and enzyme/ligand recognition, hence are affected. All the methylations are post-transcribed, and the methylated positions are recognized by enzyme according to the specific base pairs in 3D conformation.[116-119] The N1-methyl in adenosine prohibits the base pair of A9:U64 for some tRNA species, which is responsible for stabilization because those tRNAs cannot fold correctly *in vitro* in presence of this base pair.[120,121] A similar effect is supposed for 2-methylation of guanosine in G10:U25 and G26:A44, where the specific base pair geometries are kept.[116] T54 is responsible for local structural folding and flexibility for an individual tRNA.[40,41] Besides stabilizing the local structure, methylation also affects the interaction between tRNA and other molecules. For example, m^5C40 is essential to Mg^{2+} bound to the ASL.[122] Two charged methylation m^7G46 and m^1A58, and m^5C48 stabilize the tRNA structure by preventing the rapid degradation of premature cellar tRNA.[123-125] Without the methylations the tRNA still folds correctly, but the enzyme of aminoacylation cannot recognize correctly.[126]

2.3.2.6 Others

2'O-ribosylation is another natural ribose modification besides 2'OMe. It is only found in purines. The eukaryotic unique rA64 discriminates initiator tRNA from and elongator tRNAs, as the removal of rA64 leads $tRNA_i^{Met}$ to be recognized by elongation factors.[127]

2.3.3 Modifications in other RNAs

2.3.3.1 mRNA

Currently 13 modifications are found in mRNA. One example is m^7G, the simplest but most common N7-methyl guanosines. It is almost conserved in tRNA but better known as the eukaryotic mRNA cap. It is added on 5'-cap along the transcription and is essential to intron splicing, translation and poly(A)-tail addition for a mature mRNA.[128] Another well-known modification in eukaryotic mRNA is the adenosine deamination product inosine (I). Since inosine paired with C, it changes the codon and causes a substitution for the synthesized peptide chain. This modification supplies the protein diversity while increasing misguided gene editing. Although different conditions corresponding to modification levels have been observed, the whole mechanism behind the deamination regulation is still unknown.[129]

2.3.3.2 rRNA

Most of 31 rRNA modifications are methylation. The study of rRNA modification has not been as detailed as tRNA, however some methylations have been claimed to be important to hairpin formation and RNA folding, *e.g.* Um2552 in A loop[130], and m^2G and $m^6{}_2A$ in Helix 45[131]. Denser methylation also observed in A-site finger and Helix 69, which contact the A and P-site tRNA and are influential to the translation efficiency in yeast ribosome.[132] 8-

methylation is recently found in A2503 of prokaryotic 23S RNA, which interferes with the binding of Linezolid and Clindamycin, *etc.* and is one of bacterial ribosomal mutation responsible for antibiotic resistance. [133,134]

2.4 ARTIFICIAL DNA MODIFICATIONS

Abnormal gene expression is usually associated with human disease, silencing of specific gene expression thus can be applied in the treatment of these diseases. Selectively targeting the genomic DNA by TFOs is a strategy to modulate gene expression.[135] Triplex formation is straightforward to control *in vitro*, but there are substantial obstacles in living cells. The TFO must be nuclease resistant and form a stable triplex at physiological pH to interfere with the biological processes that act on DNA. This provokes the development of DNA modifications

2.4.1 Peptide nucleic acid

Peptide nucleic acid (PNA) is a synthetic compound which has the same base type as the common nucleic acids, but with a peptide backbone replacing the ribosyl phosphate.[2] It is not negatively charged and has no repulsion between strands in helix. Therefore PNA duplex is more stable than DNA or RNA duplex. Furthermore PNA backbone is more flexible so that less structural restraint is present when complex is formed.

Since no intrinsic RNA restriction enzyme has been reported, one application of PNA is the synthesized RNA nucleases (PNAzyme).[136-139] The PNA oligonucleotide carrying a catalyst selectively binds to target RNA and cleaves the target mRNA into fragments. The most effective sequence binds to target RNA by WC base pairs and pinches four unpaired nucleotides into a bulge. Then the carrying Cu^{2+}-phenanthroline complex binds to the target site on bulge backbone and cleaves the RNA from the phosphate ester bond.[136] However, the *in vitro* catalytic rate, 0.5 h, is still not enough regarding a cellular process.

2.4.2 Locked nucleic acid

Locked nucleic acid (LNA, Figure 1 in paper IV) is another artificially modified nucleotide with an oxymethylene group bridging between C2' and C4' which locks the pucker in *north* conformation.[3,4] The modification does not change the WC base pairing rule and the solubility of LNA is likely between DNA and RNA. Thus LNA keeps the similar physical properties as the original DNA/RNA helical structures. The common LNA refers to β-D-LNA which constantly constrains the ribose to be *north*, and its stereoisomer which is in *south* is specified as α-L-LNA.[140] Both isomers introduced in one strand will increase the stability of DNA-DNA or DNA-RNA duplexes,[140,141] but β-D-LNA is dominantly common in bioassays because β-ribose is more relevant to natural sugar. In the follow texts only β-D-LNA (referred as LNA) is described.

The LNA containing oligonucleotides shows enhanced binding affinity in complementary DNA sequence compared to the counterpart DNA sequences by detecting the significantly increased melting temperature; and this enhancement is additive when more DNAs are mutated to LNA nucleotides.[4,142-144] The LNA propagates the *north* conformation to

especially the 3'-neighbor whereas the sugar puckers in the complementary strand are less affected. The LNA containing duplex conformations, such as backbone torsions and groove dimension, are shifted toward A-type.[141,145,146] The enhancement of binding affinity by LNA is also present in triplex, where the TFOs with LNA substitution in triplex have significantly increased melting temperature.[147] But different from the duplex formation, a TFO with full LNA nucleotides is not possible to form a triplex while a TFO with LNA flanked by one or two DNA nucleotides has the best performance of the triplex formation.[146,147] Nowadays 5-methylcytidine combined LNA (m^5C_{LNA}) is often used in triplex forming oligonucleotide (TFO) instead of cytidine,[143,144,146,148,149] because m^5C has better base stacking in physiological pH.[150,151] Therefore the stabilization of C contained TFO can be also contributed by base methylation.

Kinetically LNA contributes to decrease the dissociation constant in duplex and triplex formation.[146,147,152] The binding enhancement brought by LNA is attributed to the pre-organization of backbone conformation and enhancement of base stacking for duplex formation, which corresponds to entropy and enthalpy effect respectively.[142,145,147,153] Furthermore LNA stabilizes the stacking of matched base pair rather than the mismatched base pair which enables LNA to be a probe that discriminates the base pair mismatch at specific position of DNA duplexes in gene assay.[144,154,155]

The effect of LNA in DNA duplex was investigated using experimental and designed sequences in molecular modeling. Those studies checked the properties of LNA containing duplexes. The LNA containing double helixes undergo an obvious unwinding and the base-pair geometries turned to be more A-like,[156-158] where slide and twist are very discriminable between LNA-modified and unmodified DNA duplex. The hydration properties of groove are different because of the changed groove dimensions,[156,159] where LNA duplexes have less hydration number but network of water is more regular. The binding free energies of duplexes that calculated using Generalized Born Surface Area demonstrated LNA containing duplexes have lower free energy than canonical DNA or RNA duplex, which is qualitatively consistent with experiments.[156,157]

3 MOLECULAR MODELING AND SIMULATION

Molecular modeling and simulation employ mathematical, physical and chemical principles to predict the behavior of molecules, in particular of biomolecules. The molecular system can be described in molecular mechanics (MM) or quantum mechanics (QM) based model. In MM the mass and coordinate are at the center of an atom, and electrons are invisible but mimicked by atomic charges and covalent bonds. During the regular simulations no charge is transferring and no bond is creating or breaking. This chapter first introduces force field, and then the molecular dynamics (MD) simulation, and finally the free energy calculation.

3.1 FORCE FIELD

Force field is a set of energy functions combining with the parameters. It defines the potential energy for a given system so the force on each atom can be calculated in simulation. It consists of non-bonded terms and bonded terms. The former includes the electrostatic and van der Waals (vdW) interactions between two atoms; and the latter includes the bond, angle, and dihedral angles which are defined by one, two and three covalently bonded atoms respectively. The potential energy of a system is the sum of them:

$$\begin{aligned} V = & \sum_{\substack{\text{nonbonded}\\ \text{pairs}}} \left(\frac{q_i q_j}{4\pi\varepsilon\varepsilon_0 r_{ij}} + \epsilon_{ij} \left[\left(\frac{R_{\min,ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{R_{\min,ij}}{r_{ij}} \right)^{6} \right] \right) + \sum_{\text{bonds}} K_b (b - b_0)^2 \\ & + \sum_{\text{angles}} K_\theta (\theta - \theta_0)^2 + \sum_{\text{UBs}} K_{UB} \left(r_{1,3} - r_{1,3;0} \right)^2 \\ & + \sum_{\text{impropers}} K_\varphi (\varphi - \varphi_0)^2 + \sum_{\text{dihedrals}} K_\Phi \left(1 + cos(n\Phi - \delta) \right) \end{aligned} \tag{1}$$

The electrostatic term is formulated by Coulomb potential, where q_i and q_j are partial charges of distance r_{ij}, ε_0 is the vacuum permittivity, and ε is the relative permittivity which is generally assigned a value of 1 for explicit solvent simulations. The vdW term uses the Lennard-Jones (L-J) expression, where $R_{\min,ij}$ is the distance of minimal potential energy between the two particles, and ϵ_{ij} is the depth of the potential well. The bonds, angles, Urey-Bradley (UB) and improper dihedrals are described by harmonic expressions, where K_b, K_θ, K_{UB} and K_φ are force constants and b_0, θ_0, $r_{1,3;0}$ and φ_0 are equilibrium values. The dihedral angle (torsion) is described by cosine functions, where K_Φ, n and δ are the amplitude, periodicity and phase angle, respectively. A given torsion may be treated as a Fourier series over one or a sum of periodicities.

3.1.1 Nonbonded interactions

All the nonbonded interactions physically are because of electronic distribution. Although the empirical atomic charges can evaluate the electrostatic interactions, the interactions of multipoles *e.g.* between neutral atoms, need another term.

3.1.1.1 van der Waals interaction

The L-J expression in eq 1 is a simple but widely used approximation. It includes two physically different interactions: the repulsive exchange force due to overlap of electron orbitals and the attractive dispersion force due to the instantaneous dipole.

The molecular electrons are dynamically redistributing by the induction of environmental molecular dipoles. Thus there is an instantaneous dipole interaction (dispersion force, or loosely vdW force) between two nonbonded atoms in any moment. Such attractive interaction is usually small, but its accumulation for a whole system is considerable and it becomes more dominant in nonpolar molecules. It is related to the polarizability of atoms. According to Drude model, it decays as r^{-6}, which is the basis of the attractive term in L-J expression (eq 1).

On the other hand, if two atoms are so close to each other that their electron orbitals are overlapping, a strong repulsion will be present between electrons of same state according to the Pauli Exclusion Principle. For the reason of computational efficiency, the repulsive r^{-12} term is simply taken as the square of vdW term, which is not physically based. Other forms of the potential for vdW interaction have been suggested, *e.g.* using the QM based r^{-9} repulsive term instead of r^{-12} term. Very accurate vdW parameters are more required for nonpolar systems, like rare gases,[160] however, for biomolecular simulations the current L-J expression is adequate.[161]

In most force fields the L-J parameters for interactions between two different atom types are obtained using combining rules for the intrinsic parameters of each atom in the pair:

$$\epsilon_{ij} = \sqrt{\epsilon_{ii} * \epsilon_{jj}}\,, \qquad R_{\min,ij} = (R_{\min,ii} + R_{\min,jj})/2 \tag{2}$$

3.1.1.2 Additive electrostatic interaction

The electrostatic interaction expressed using Coulomb's law, but the partial atomic charge is a simplified treatment. In additive force field, the partial atomic charges are static and the electrostatic interactions can be summed from energies between any charges in given coordinates. Such treatment currently is most popular for its computational efficiency. To remedy the omitted induced dipole interactions, the atomic charges in additive force field are artificially increased to overestimate the interaction energy and molecular dipole at least 20% against QM calculated data. By this enhanced charges the electrostatic interaction calculated in aqueous solution is close to the experimental measurement. Such empirical charge is capable of reaching considerable accuracy for most biomolecular simulations. However since the charge is unable to redistribute with the change of the environment dielectric, the interaction might be less well described when a molecule moves from a solvent accessible surface to being buried in a cavity.

3.1.1.3 Water model

The interaction energy involving water molecules is often the biggest contribution in explicit solvent systems. All the physical properties of water models therefore need to be well defined. The simplest and popular additive water models are TIP3P[162] and SPC[163] which have small differences in geometric parameter and atomic charge, but both treat a water molecule as a rigid three-particle model. The positive charges are equally assigned on two hydrogen atoms and the negative charge is on oxygen. In early version only oxygen has L-J radius but later some variants also suggest the L-J radius for hydrogen.[164] The four-particle model TIP4P[162] has a charge particle shifted a little from the oxygen atom, but the geometry is almost same as TIP3P. The five-particle model has two charge particles for oxygen as the lone pair electrons. One set of force fields is usually parameterized based on one kind of water model, for example, CHARMM, OPLS and traditional versions of AMBER use TIP3P, and GROMOS uses SPC. It is important to use the same water model in simulations as was used in parametrisation.

3.1.1.4 Polarizable electrostatic interaction

Alternatively in non-additive force field, the atomic charges are allowed to change and correlate to each other. The molecule is polarizable and the dipole moment can be explicitly described. So far there are three methods introduced in the polarizable force field.[165] The first is point dipole model which introduces a parameter polarizability to calculate the dipole moment. The second is Drude model which introduces massless and negatively charged particles that harmonically attach to the positively charged atom cores, to mimic the induced dipole. The third is the electronegativity model which does not introduce new interactions but allows atomic charges to fluctuate and redistribute.

CHARMM polarizable force field employs the second method,[166] in which Drude particles are attached on all non-hydrogen atoms. The total electrostatic energy is:

$$V_{\mathrm{elec}} = V_{\mathrm{c,c}} + V_{\mathrm{D,D}} + V_{\mathrm{c,D}} + V_{\mathrm{pol}} \tag{3}$$

where $V_{\mathrm{c,c}}$, $V_{\mathrm{D,Dc}}$, and $V_{\mathrm{c,D}}$ mean interactions between core and Drude charges which are evaluated as Coulomb term in eq 1, and $V_{\mathrm{pol}} = \frac{1}{2}\sum k_i d_i^2$ that mimics the polarizability is the energy of harmonic spring of Drude particles. Obviously the number of electrostatic interactions is more than doubled compared to the additive force field; a small time step is also required due to the high frequency of the Drude spring. Overall this increases the computational cost approximately 5-fold.

3.1.2 Bonded interactions

3.1.2.1 Covalent bond and angle

An empirical approximation for bond stretching is expressed by the Morse potential:

$$V = \varepsilon[1 - e^{-a(b-b_0)}]^2 \quad (4)$$

where ε is the depth of the potential minimum and $a = \sqrt{k/2\varepsilon}$ with k being the force constant of the bond. Most force fields employ a harmonic potential to mimic the Morse potential because of the computational efficiency. Although a deviation is present, the harmonic energy surface approximately fits the Morse potential around the minimum. Considering that the bond force constant is usually large and MD simulation rarely samples the high energy conformation, this approximation is acceptable.

The bending of covalent angle also uses the harmonic expression but the force constant is smaller. To reproduce the coupling between bond stretching and angle bending the Urey-Bradley potential of cross-term is needed. It adds an harmonic potential on 1,3-interaction. Although the cross-term coupling is everywhere, the UB correction is only needed for some sensitive angles according to the vibrational analysis. Finally all the nonbonded interactions are excluded for atoms involved in the same bond or angle.

3.1.2.2 Dihedral angle

For four consecutive atoms dihedral angle describes the rotation of the middle bond. A Fourier series is used to express the periodic energy change. Dihedral potential is defined as a correction term addition to 1,4-nonbonded interaction (between two end atoms), so that correct energy surface can be obtained related to the rotation of a bond. In CHARMM 1,4-interaction keeps the original electrostatic and vdW interaction, whereas AMBER and later OPLS employ the interaction scaled by 0.83.[161] This different 1,4-factors show the varied philosophy behind different force fields, especially the treatment of nonbonded interaction.

3.1.2.3 Improper dihedral

Improper dihedral is an artificial term to reproduce motion of out-of-plane in a planar molecule. It is usually applied on sp2-atom centered chemical groups. In the case that the three angle parameters are transferred from other molecules and the sum of the equilibrium angles is not 360°, the improper dihedral is used to keep those three angles on the same plane.

3.1.3 Parametrisation

The quality of a force field is judged by the performance of reproducing experimental properties. However since experimental data are not always available, QM calculations are provided as target data for the systematic parameterization and the NMR and X-ray data are often used for validation and refinement. Although the QM calculation supplies all the parameters that describe a molecule, such as charge, geometry and energy, they are not applicable directly for a force field due to the difference between QM and MM energy functions. So protocols are developed to adapt *ab initio* data for molecular mechanics.

The philosophy of empirical force field is transferability, *i.e.* the necessity that the parameters determined in a single residue are also valid in polymer. Allowing the same parameters for

repeating blocks in polymer systems is the basis of molecular mechanics. The selection of model compounds hence is important, as too simple or too sophisticated model may decrease the accuracy or transferability respectively. A complicated molecule is usually broken into several fragment models and each is parameterized separately. A general idea of building a model is that, 1) keep conjugate atoms as a moiety and split from aliphatic bond; 2) do not break down the H-bond donor or acceptor and other polar groups; 3) avoid target atoms involving in the intra-molecular interaction, such as H-bond and steric hindrance. The order of parametrisation usually is first charge, and then harmonic bonded terms and finally torsions. The L-J parameters are determined before or after charge, but whenever it is done all the other parameters need to be rechecked. Finally whatever is altered, the concerning torsion should always be finally checked.

3.1.3.1 Atomic charges

The electrostatic interaction is often the dominant energy term in aqueous solution. The QM optimization of the model compound is usually performed using MP2/6-31G*.The MM partial charges are then obtained using one of two different strategies. The one is the electrostatic potential (ESP)[167] based method which is employed in AMBER and polarizable CHARM force fields. The other one used in additive OPLS and CHARMM force fields calculates the water-complex to get the interaction energy. In the latter method of CHARMM, each solvent accessible atom of the model is probed by a water molecule in the poses of hydrogen bond. The HF/6-31G* energy of polar compound is scaled by 1.16,[162,168] and the dipole moment magnitude of neutral molecule is scaled by 20%~50%[169] to fulfill the overestimation in empirical fitting. Polarizable force field does not require such overestimation since the induced dipole moment can be explicitly evaluated. The QM H-bond distance between of water and probed atom is also scaled by −0.2 Å before the empirical parameters are fitting[169]. Nonpolar water interaction distances are usually not scaled.

The CHARMM charges have been optimized to be an integer value for a group of atoms since early stage. This is to keep the transferability of basic functional groups so their parameters can be immediately applied to a new molecule with the same substructure. For instance, in methyl or methylene group that is not next to heteroatoms, the carbons is −0.27 or −0.18 respectively, and the aliphatic hydrogen is constantly +0.09.

3.1.3.2 L-J parameters

The L-J parameters are only concerning with the atom types that were designed to represent different hybrid orbitals. For example, all force fields differentiate sp3, sp2 and sp carbons. The L-J parameters are optimized only for the atom types in new chemical space; otherwise they can be taken from existed force fields. The vdW interaction is difficult to calculate because it is weak and easily shed by other interactions. Thus the target atom is probed by a rare gas atom with zero charge. The QM calculation also should be in higher level, at least MP3/6-311++G, to get accurate energy. The potential well and minimum distance are adjusted to reproduce QM results.[170,171] The supplemented methods as well as the validation

to L-J parameters include the liquid simulation and the reproductions of the experimental molecular volume and vaporization heat.

3.1.3.3 Bonds, angles and improper torsions

Almost all force fields optimize these parameters in a same way. The equilibrium points of bond, angle and improper dihedral are adjusted directly toward the value of QM data, which are calculated in MP2/6-31G*. The force constants are determined to reproduce the vibrational analysis that is solved from the Hessian matrix in the same QM level. In CHARMM, the infrared spectra of the molecular vibration are analyzed by potential energy distribution (PED) using MOLVIB module.[7] The assignment with frequency higher than 1500 cm^{-1} belongs to the stretch of hydrogen or double/triple bond. The movements of those bonds are either constrained or limited in simulation, so the most efforts are usually made on the terms with frequencies < 1500 cm^{-1}. The empirical parameters target on MP2/6-31G* results that are scaled by 0.94 to reproduce the best experimental data.[172]

3.1.3.4 Flexible dihedrals

The stiff dihedral, *e.g.* torsion of aromatic or conjugate bond, can be approximated using harmonic function so is determined in the previous step. The flexible is optimized using potential energy surface (PES) scan. The target potential energy is calculated in MP2/6-31G* level. This energy surface represents the whole molecule with respect to the conformational change of the rotatable bond, but it is dominated by dihedral parameters. It is never trivial to be fine-tuned to reproduce the correct molecular conformation, and the most effort is made on the potential surface below 12 kcal/mol. There are several dihedral terms associated with one bond, each term has the different 1,4-atoms, so there can be multiple terms with different periodicities. In principle, n can be any integer and δ any value between 0° and 360°, but in practice n is usually taken to be 1, 2, 3, 4 and/or 6, and δ is either 0° or 180° because more terms are not only trivial to improve accuracy but also sacrifice the transferability.

3.1.4 Biomolecular force fields

There is no universal force field to cover the entire chemical space; rather each force field is developed for a specific set of molecules. Force fields have been developed for all biological macromolecules, *i.e.* proteins[173-178], nucleic acids[174,178-181], carbohydrates[178,182-184] and lipids[178,185-188], as well as drug-like organic molecules[169,178,189-191].

3.1.4.1 Proteins

The side chain and backbone of amino acids are parameterized separately. Except for proline and glycine, backbone parameters of the rest residues are parameterized from alanine dipeptide. AMBER and CHARMM include the vibrational analysis for the force constants of bonds and angles, targeting on QM calculations and PDB survey. Different force fields have own charge and parameter assignment, but they can reproduce the consistent behavior for proteins.

Extra efforts have been made on optimization of two backbone torsions φ and ψ, because they are determinant to protein conformation. Ramachandran plot[192] provides an informative reference to represent the preferred distribution for φ and ψ. Based on the survey and QM calculation, a 2D dihedral energy correction map of φ and ψ is introduced in CHARMM. An energy term $\sum V_{\mathrm{cmap}}$ is then added to eq 1 for CHARMM protein force field. The updated potential function significantly improves the reproduction of protein structures.[176,193] The refinements on φ and ψ were also made in AMBER.

3.1.4.2 Nucleic acids

Nucleic acid force fields usually take the transferred L-J parameters from proteins, however there are still two challenges: the intensive negative charge and up to six flexible backbone torsions. All AMBER, CHARMM and GROMOS have succeeded in simulating stable nucleic acid structures since early stage.[194,195] Later refinements of CHARMM and AMBER focused on the re-optimization of charge and conformational parameters. CHARMM27 got outstanding improvements for its reproduction of A-, B- and Z-helix referring to PDB survey.[11,179] The investigations were done on series model blocks and nucleosides in both QM and MM levels, which thoroughly provided detail for the parametrisation.[12,13,16,18,19,196] The correlation between the 2'-OH torsion and RNA conformation was later optimized to obtain an improved agreement with experimental structure.[20,197] The DNA backbone torsions were also updated to allow better sampling on B_I/B_{II} conformations.[176] These two features are included in CHARMM36.

With respect to the modification on nucleic acids, analogous parameters transferred from models of other force fields are often used. The first specialized RNA modification force field was developed for AMBER,[9] but that optimization was limited to charges without further validating the conformations.

3.1.4.3 Other macromolecules

The carbohydrate force field aims to reproduce good conformation for poly saccharide. Its biggest challenge is the sensitive conformational interconverting of hexose puckers and the diversities of linkage glycosidic torsions. CHARMM carbohydrate force field first aimed to optimize the charge, L-J parameters and 1,4-interaction for monosaccharide, so that the reasonable ring conformation and intra hydrogen bonds were reproduced.[182] The glycosidic torsions φ and ψ (not confusing with peptide torsions) in disaccharides and glycans were then optimized using 2D correction map.[198] As the model blocks of lipid, the parametrisation of alkanes began at early stage.[170,171] However the difficulty of lipid force field is always from the hydrophobicity which makes lipid behave differently from other macromolecules in solvent, and require the high quality of L-J parameters. The current lipid force field enables the reproduction of different lipid layer systems.[185] Alternatively the united and coarse-grained force fields are also the good choice for lipid simulation, which gains the computational efficiency for big sized lipid system.

3.1.4.4 Small molecules

Small organic molecules have much wider chemical diversity than macromolecules. The parameters of small molecule are partly applied in the building blocks of each macromolecular force field. MMFF is one of the early professional force fields of drug-like organic molecules,[191] and still earns good reputation till now. It covered an extensive chemical space and optimized both bonded and nonbonded interaction terms. MMFF has good performance in protein-ligand interaction and is popular in computer-aided drug design. The general AMBER force field (GAFF) came later, which first time developed an automated procedure based on the restrained ESP charge assignment.[199] This brought the advantage that users can get parameters for new molecules without force field knowledge.

The CHARMM general force field (CGenFF)[169,200] covers sufficient chemical space base on the predecessors and later-added sulfonyl compounds. CGenFF focused on both charge and bonded parameters using newly developed automation method, which assigns L-J parameters based on atom type and assigns other parameters based on analogs in current CGenFF.[201,202] Furthermore the CGenFF program generates a penalty score on each initial guess based on the similarity from the analogs, so that users know which parameters are reliable and which likely need validation. Its web server ParamChem is available: http://cgenff.paramchem.org/.

3.2 MOLECULAR DYNAMICS SIMULATION

MD simulation computes the dynamic motion and thermodynamic properties for molecular systems in multiple scales.[7,203,204] This section introduces the basic principles and techniques of modern MD simulation.

3.2.1 Newtonian mechanics

In MD simulation the system configuration is generated by applying Newton's law of motion, *i.e.* the position of a particle i with mass m has the relationship with its force F along time t, and in any moment the force along x component in Cartesian coordinates is obtained from the differential of the potential energy:

$$\frac{d^2x_i}{dt^2} = \frac{F_{x_i}}{m_i} = -\frac{1}{m_i}\frac{\partial V}{\partial x} \tag{5}$$

Given an initial set of coordinates and velocities for a particle, the potential and kinetic energies are defined. In the next moment the new position and velocity are solved by the force acting on the particle (eq 5). Thus every next moment is predictable and the trajectory of a motion is determined. However for many-body systems, the motions of particles are coupled, thus the equation has to be solved numerically by several integration algorithms.

3.2.1.1 Integration of Motion

If assuming the forces are constant within a time step δt, having the current position $r(t)$, acceleration $a(t)$ and the position of the last step $r(t-\delta t)$, the position of the next step can be calculated by Verlet integration:[205]

$$r(t+\delta t) = 2r(t) - r(t-\delta t) + \delta t^2 a(t) \tag{6}$$

It has limitations: the velocity is not explicitly included but can only be calculated in the next time step, and a small term δt^2 is added to two big terms which may lose precision. The leap-frog integration calculates the velocity and improves the precision by first calculating the velocity in halt time step:

$$\begin{aligned} v\left(t+\frac{1}{2}\delta t\right) &= v\left(t-\frac{1}{2}\delta t\right) + \delta t a(t) \\ r(t+\delta t) &= r(t) + \delta t v(t+\frac{1}{2}\delta t) \end{aligned} \tag{7}$$

As the name indicates, the velocity is always half time step ahead the position, hence they are not synchronized. Thus the kinetic and potential energies are not at the same moment.

3.2.1.2 Constraints

The choice of time step δt depends on how the degree of freedom is described in MD simulation, *i.e.* δt needs to be small enough so that the momentum from each motion is conserved. For a biomolecule people are often interested to know the conformational change. To describe the atom translation and bond rotation for flexible molecules the time step has to be 2×10^{-15} s. However the bond vibration which is unlikely interesting needs δt to be 1 or 0.5×10^{-15} s. Since the bond stretching will increase the computational time, constraints are usually applied on covalent bonds in MD simulation. For example TIP3P water has two O-H bonds and one fake H-H bond constrained so that TIP3P has no vibrational degree of freedom. Biomolecules usually have all hydrogen covalent bonds constrained, so a time step of 2 fs is feasible. The constraint algorithm used in projects of this thesis is SHAKE.[206]

3.2.2 Sampling and properties

If the initial coordinates are defined, beginning with a random velocity the system will reach the thermodynamics equilibrium after some time steps. Although a particle can be described by exact momentum and position, for a biomolecular system from thousands to millions of particles the properties must be considered in a statistical way. The total energy of a system is described by the Hamiltonian $\mathcal{H}$ which includes kinetic energy $\mathcal{K}$ and potential energy $\mathcal{V}$:

$$\mathcal{H}(\mathbf{p},\mathbf{r}) = \mathcal{K}(\mathbf{p}) + \mathcal{V}(\mathbf{r}) \tag{8}$$

where $\mathcal{K}$ and $\mathcal{V}$ are only functions of momenta **p** and coordinates **r** respectively. The kinetic energy is calculated from the particle velocities ($\sum \frac{1}{2} m_i v_i^2$) while the potential energy is calculated from the force field (eq 1).

3.2.2.1 Phase space

The time evolution of a system is governed by the Hamiltonian as:

$$\frac{d\mathbf{p}_i}{dt} = -\frac{\partial \mathcal{H}(\mathbf{p}, \mathbf{r}, t)}{\partial \mathbf{r}_i}$$
$$\frac{d\mathbf{r}_i}{dt} = \frac{\partial \mathcal{H}(\mathbf{p}, \mathbf{r}, t)}{\partial \mathbf{p}_i} \tag{9}$$

The momenta and coordinates have three components (*x*, *y*, *z*) respectively. An *N*-particle system thus has 6*N* degree of freedom, in which 3*N* are momenta and 3*N* are coordinates. The whole state of all the possible momenta and positions that the system can visit is phase space. In thermodynamic equilibrium the possible microstates of the system in the phase space are collected in an ensemble which obeys the Boltzmann distribution: $\exp(-E_i(\mathbf{p}, \mathbf{r})/k_B T)$, where E_i is the total energy of state *i*, *T* is temperature and k_B is Boltzmann constant. The probability of a certain microstate *i* is given by:

$$\rho_i(\mathbf{p}, \mathbf{r}) = \frac{\exp(-E_i(\mathbf{p}, \mathbf{r})/k_B T)}{\frac{1}{N!\, h^{3N}} \iint d\mathbf{p}\, d\mathbf{r}\ \exp(-\mathcal{H}(\mathbf{p}, \mathbf{r})/k_B T)} \tag{10}$$

In this equation the denominator is the normalized partition function Q_{NVT} which describes the phase space. Here the subscript *NVT* means the phase space is visited through canonical ensemble, i.e. in the conditions of constant particle number (*N*), volume (*V*) and temperature (*T*). Eq 10 suggests the lower energy level gets more popular distribution.

A property of the system is an observable averaged over an ensemble. According to the ergodic hypothesis: in an equilibrated system, the measurement of a property from different microstates averaged over the ensemble is equivalent to the measurement from one microstate averaged over the time series. So connection can be built between real system observables (many molecules, one moment) and simulated system observables (one molecule, many times):

$$A = \iint d\mathbf{p}\, d\mathbf{r}\ A(\mathbf{p}, \mathbf{r})\, \rho(\mathbf{p}, \mathbf{r}) = \frac{1}{M} \sum_{i=1}^{M} A(\mathbf{p}, \mathbf{r}) = \langle A \rangle \tag{11}$$

where the brackets mean time average, and *M* is the number of time steps. Obviously in eq 11 the equality holds if the ensemble is very large and time is very long. So the simulation time is required to be sufficient.

3.2.2.2 System properties

Some properties can be calculated directly along the trajectory production in MD simulation. For example the internal energy of a system can be expressed from eq 11:

$$U = \iint d\mathbf{p}\, d\mathbf{r}\, \mathcal{H}(\mathbf{p},\mathbf{r})\, \rho(\mathbf{p},\mathbf{r}) \tag{12}$$

As implied in eq 10, the MD simulation walking in the phase space is to seek the configurations of lower internal energy until the equilibrium is reached. Some other basic properties are directly related to the energy terms. The first one is the temperature:

$$T = \frac{2\mathcal{K}(\mathbf{p})}{k_B N_{df}} \tag{13}$$

where N_{df} is the number of degrees of freedom of the system. Usually $N_{df} = 3N - N_c$, where N_c is the number of constraints used in MD simulation. The second is the pressure:

$$p = \frac{1}{V}\left[Nk_BT - \frac{1}{3}\sum_{i<j}^{N} \mathbf{r}_{ij}\frac{d\mathcal{v}(\mathbf{r}_{ij})}{d\mathbf{r}_{ij}}\right] \tag{14}$$

where *V* is the volume. The first term is for ideal gas and the second term is for real gas, where $\mathbf{r}_{ij}$ is the vector between two particles, 3 denotes the dimension, and $\mathcal{v}$ is the pairwise potential energy with $\mathcal{V} = \sum \mathcal{v}$

3.2.2.3 Thermodynamic properties

According to Boltzmann's definition, thermodynamic properties are related to the partition function itself, for example the Helmholtz free energy *A* is:

$$A = -k_BT \ln Q_{NVT} = k_BT \ln\left(\iint d\mathbf{p}\, d\mathbf{r}\ \exp\left(\frac{\mathcal{H}(\mathbf{p},\mathbf{r})}{k_BT}\right)\rho(\mathbf{p},\mathbf{r})\right) \tag{15}$$

Different from internal energy, the contribution from high-energy states in this equation are more significant. But since high-energy states cannot be sufficiently sampled in standard MD simulations, the free energy is hence difficult to evaluate correctly along the trajectory. The solution to get free energy will be introduced in the section 3.3.

3.2.3 Long-range force

This subsection describes the techniques for dealing with the most computationally expensive part of MD simulation, the long-range interactions. Calculating force of nonbonded interaction between atoms is time consuming, because its complexity related to the particle number is $O(N^2)$. To count these interactions the system firstly has to be placed in a proper dimension.

3.2.3.1 Boundary conditions

A simulation sized system (1k−1000k particles) is much smaller than real system. If the lattice has a rigid boundary (like a container), the collision between the hard wall and solvent molecules will cause considerable interactions on solute. If there is no boundary, the surface solvent molecules will feel in vacuum which still causes strange behavior. To overcome this, the image arrays centered on the system lattice are extending along each direction, which is in periodic boundary condition (PBC). All coordinates of images are copied from the central cell and the particles in central cell are surrounded by particles out to infinite distance. As a consequence a particle which moves outside the central cell will be compensated by its image on the opposite side entering the central cell. PBC is reasonable as the situation is similar to the middle cell in a crystal lattice. The cell shape needs to fill all the space by translation of the cell in three dimensions, thus orthorhombic, hexagonal, octahedral and rhombic dodecahedral cells are commonly used.

The spherical cell with a non-periodic boundary can be also applied for truncated systems. In this case, the residues on the surface need to be restrained and only the center part is allowed to move freely. To reduce the boundary effect, the sphere is usually built to be larger than the active site to allow a buffer region between the surface and inner sphere. Mild restraints can be used in buffer region to keep the truncated residues stable meanwhile the central part is not affected by the restraints.

3.2.3.2 Pairwise cutoff

An idea to reduce computational cost is to use cutoff, outside which the nonbonded interaction is omitted. As the vdW interaction decays to be < 1% when distance is > 6 Å, a cutoff greater than 8 Å is sufficient. For most systems the cutoffs greater than 12 Å is also reliable for electrostatic interaction. However the energy discontinuity will break the energy conservation for whole system. Two methods are used to smoothen the energy gap after the cutoff. The first uses a shifted potential:

$$U_{\text{shift}}(r) = U(r) - U_c - \left(\frac{dU(r)}{dr}\right)_{r=r_c}(r - r_c), \qquad r \le r_c$$
$$U_{\text{shift}}(r) = 0, \qquad r > r_c \tag{16}$$

where r_c is the cutoff distance and U_c is the potential energy at cutoff. The last term in the upper equation is to make the first derivative, *i.e.* force, continuous at $r = r_c$. The shifted potential always has an offset U_c from real potential and it vanishes after r_c. The second method needs an inner cutoff r_{in} and uses a switching function between two cutoffs:

$$U_{\text{switch}}(r) = U(r), \qquad r < r_{in}$$

$$U_{\text{switch}}(r) = U(r)\left(c_0 + c_1\frac{r - r_{in}}{r_c - r_{in}} + c_2\left(\frac{r - r_{in}}{r_c - r_{in}}\right)^2 + \cdots\right), r_{in} \le r \le r_c \qquad (17)$$

$$U_{\text{switch}}(r) = 0, \qquad r > r_c$$

The polynomial (switching function) in the middle equation is to switch off the energy and force gradually while keeping them continuous at r_{in} and r_c. The switching potential is identical to real potential before r_{in} until it decays fast between at r_{in} and r_c. Alternatively the shifted and switching functions can be also applied on force which is often more preferable for MD simulations.

3.2.3.3 Ewald summation

Charge-charge interaction decays as r^{-1}, so unless a large cutoff is used the shifted or switching function may not be sufficiently accurate. In periodic boundary conditions, a point charge in central cell interacts with the charges in the same cell and the image cells. The total energy is summed up over the image arrays for each point charge. To get this summation converged, Ewald summation provides a solution. One can assume that for a point charge $+q_i$, a diffuse charge cloud $-q_i$ is surrounding by. As compensation a second charge cloud $+q_i$ is added to counteract the $-q_i$. The charge clouds obey Gaussian distribution. In the outer space if the Poisson equation of compensating charge is solved in reciprocal space using Fourier transformation, the distribution will become periodic and the energy is converged fast. In the real space (inner) the electrostatic energy is screened by Gaussian charge cloud, thus decays rapidly as $\text{erfc}(\mathbf{r}_{ij})/\mathbf{r}_{ij}$, where $\text{erfc}(x) = 2/\sqrt{\pi}\int_x^\infty \exp(-t^2)\,dt$. The total electrostatic energy is summed from two parts; for conciseness the formula is written by excluding numerator and coefficients:

$$\frac{1}{r_{ij}} = \frac{1 - f(r)}{r_{ij}} + \frac{f(r)}{r_{ij}} = \frac{\text{erfc}(\kappa r_{ij})}{r_{ij}} + \frac{\text{erf}(\kappa r_{ij})}{r_{ij}} \qquad (18)$$

This idea is to convert the interactions into short range (erfc(x) term, real space) and long-range (erf(x) term, to be transformed in Fourier space) and both converge in each space. The parameter κ is chosen to control the convergence efficiency.

The Ewald summation solved the accuracy problem because of cutoff, but the computational complexity of the most expensive long-range term is still $O(N^2)$. The variant method, particle

mesh Ewald (PME) treats the charge distribution of the long-range term in Fourier space as a density field by subdividing the space into small grids.[207] This approximation loses very little accuracy, but reduces the complexity to $O(N\log N)$.

3.2.4 Control of conditions

MD simulation can be performed in the ensemble of constant volume and energy (*NVE*), also constant temperature and volume (*NVT*) or constant temperature and pressure (*NPT*). In most normal simulation the difference of derived properties between ensembles might not be significant, but the ensemble control is still important in case the desired condition has to be chosen.

3.2.4.1 Constant temperature

Since temperature is determined by particle velocities (eq 13), a possible control is scaling the velocity. Suppose that there is a heat bath with the desired temperature, in each time interval the velocity is changed. The changing speed depends on how tightly the heat bath is coupled. The scaling factor is given by:

$$\lambda^2 = 1 + \frac{\delta t}{\tau}\left(\frac{T_{\text{bath}}}{T(t)} - 1\right) \tag{19}$$

where τ is the coupling factor and $T(t)$ is the current temperature. This is the Berendsen thermostat.[208] It has computational efficiency to change a system toward a target temperature. However the coupling between system and heat bath is weak, which often causes problem of inhomogeneous temperature of the system. Thus unless being employed to quickly heat the system in equilibration, it is superseded by Nosé-Hoover thermostat in canonical ensembles.

Nosé-Hoover method[209,210] considers the heat bath as an integral part of the system. It includes an extra coordinate s for the heat bath with a fictitious mass Q, and its conjugate momentum is p_s. The time step of the extended system is stretched by s: $\delta t' = s\delta t$, and the coordinate and momentum are transformed as $\mathbf{r}' = \mathbf{r}$ and $\mathbf{p}' = \mathbf{p}/s$. All the properties related to time are affected by s. The Hamiltonian (eq 8) is extended as:

$$\mathcal{H}_{\text{Nosé}}(\mathbf{p}', \mathbf{r}', p_s, s) = \mathcal{K}(\mathbf{p}') + \mathcal{V}(\mathbf{r}') + \frac{p_s^2}{2Q} + (N_{df}+2)k_B T_{\text{bath}} \ln s \tag{20}$$

The first two terms are the kinetic and potential energies of real system with velocity being stretched; and the last two terms are the kinetic and potential energies of the heat bath. This Hamiltonian is conserved as the time evolution ($\frac{\partial \mathcal{H}_{\text{Nosé}}}{\partial t} = 0$), hence the extended system is sampled in microcanonical ensemble (*NVE*). However the partition function of the extended system can be projected on the real system, being a form of canonical ensemble (*NVT*):

$$Q_{\text{ext}} = \int \delta(\mathcal{H}_{\text{Nosé}} - E) d\mathbf{p}' d\mathbf{r}' dp_s ds = C \int \mathcal{H}(\mathbf{p}', \mathbf{r}') d\mathbf{p} d\mathbf{r} = C Q_{\text{real}} \qquad (21)$$

This indicates the energies are exchanged between the real system and the virtual bath so that the ensemble of constant temperature is fulfilled for target system. The coupling between real and bath is controlled by the virtual mass Q, in which big number leads slow energy exchange and small number causes instability.

3.2.4.2 Constant pressure

It is popular to run MD simulation in *NPT* ensemble because chemical reactions take place in constant pressure. Pressure is controlled by changing volume. The compressibility of a condense system like aqueous solution is small, with the volume fluctuation only 1.5%. The methods to pressure control are analogously taken from those to temperature control. The weak-coupling Berendsen barostat introduces a pressure "bath" using the similar way shown in eq 19. The new coordinates of particles are coupled with pressure bath so the volume can be changed. Alternatively the fictitious mass Q introduced by Andersen[211] for the extended system acts as a piston whose energy is coupled with target pressure and the system volume. The Nosé method stretches time step to express the new coordinates in a function of volume.

3.2.4.3 Stochastic dynamics

In most cases the solvent-solvent interaction especially the part far from solute is less interesting, so it would be good to mimic the solvent effect using a simplified model. Langevin equation introduces the terms of frictional drag and random collision to represent the solvent effects on solute particles while original potential within solute particles is unchanged. This idea assumes the solvent particles are smaller so that the motion of bigger solute particles is damped. The force of stochastic dynamics is briefly written as:

$$m_i \frac{d^2 \mathbf{r}_i}{dt^2} = \frac{\partial U_i}{\partial \mathbf{r}_i} - \gamma_i \frac{d\mathbf{r}_i}{dt} m_i + \mathbf{F}_i \qquad (22)$$

where γ_i is the friction coefficient and $\mathbf{F}_i$ is the Gaussian distributed random-collision force which implicitly contains γ_i. Based on the ratio between $1/\gamma_i$ and time step δt, the adapted integration methods have to be used. When γ_i is zero the Langevin equation reduces to Newton's motion. The stochastic dynamics required less computational effort and often can be applied with bigger time step, which improves the efficiency a lot. More importantly, the coefficient in $\mathbf{F}_i$ is related to γ_i through the fluctuation-dissipation theorem, and both are proportional to temperature. So in a given friction coefficient, the update of momenta at each time step is damped by the particular thermal collision that is specialized for canonical ensemble. Therefore the stochastic dynamics has good performance on temperature control.

3.3 FREE ENERGY CALCULATIONS

Free energy is the amount of work a system can perform. Its difference is often used to describe the work required for changing one thermodynamic state to another. In biomolecular systems, it is useful to evaluate the structural effects of point mutation, the binding affinity between molecules and the energy cost of reactions. The free energy is difficult to evaluate in normal MD simulation because of the insufficient sampling. Techniques are therefore developed to "enforce" the system to visit the adequate phase space. Two types of free energy difference are most interesting: the one between two molecules and the one along a reaction coordinate. In this section the ensemble is notated as *NVT* and the Helmholtz free energy *A* is derived. Alternatively Gibbs free energy *G* is derived from *NPT* ensemble.

3.3.1 Free energy difference

Free energy is defined in eq 15, but directly solving the partition function Q is difficult. Instead the difference (ΔA) is often calculated. The popular methods include thermodynamic perturbation (TP), thermodynamic integration (TI) and Bennett acceptance ratio (BAR). They are physically rigorous and the precision can be controlled as 1 kcal/mol, corresponding to one magnitude of binding constants measured in experiment.

3.3.1.1 Thermodynamic perturbation

The free energy difference between two molecules can be written as:

$$\Delta A = A_B - A_A = -k_B T \ln \frac{Q_B}{Q_A} \tag{23}$$

To eliminate the partition functions one has to assume the evaluation of Hamiltonians is taken in only one ensemble (either A or B). Thus eq 23 can be written as the ensemble average:

$$\Delta A = -k_B T \ln \left\langle \exp\left(-\frac{\mathcal{H}_B - \mathcal{H}_A}{k_B T}\right)\right\rangle_A = -k_B T \ln \left\langle \exp\left(-\frac{\mathcal{H}_A - \mathcal{H}_B}{k_B T}\right)\right\rangle_B \tag{24}$$

This is the solution using TP, which was illustrated from Zwangzig's exponential formula method.[212] However if the phase spaces of A and B do not overlap, $\langle \mathcal{H}_B \rangle_A$ will not be evaluated correctly and *vice versa*, hence eq 24 may encounter the convergence problem. Intermediate states therefore need to be inserted between A and B until adequate overlap is obtained between two adjacent states. In each intermediate state *i*, the Hamiltonian is scaled by a factor λ_i ($0 \leq \lambda_i \leq 1$), where $\lambda_i = 0$ or 1 is the initial (A) or final (B) state. Thus the free energy is obtained from the summations of two end states and $N-1$ intermediates:

$$\Delta A = -k_B T \sum_{i=1}^{N-1} \ln \left\langle \exp\left(-\frac{\mathcal{H}_{\lambda_{i+1}} - \mathcal{H}_{\lambda_i}}{k_B T}\right)\right\rangle_{\lambda_i} \tag{25}$$

In this equation the perturbation proceeds from A ($\lambda_1 = 0$) to B ($\lambda_{N-1} = 1$), *i.e.* forward sampling. It is equivalent to do so in backward sampling. Actually it is beneficial to do double-wide sampling. In this case the half point of forward and backward λ-space is evaluated on current ensemble, *i.e.* $\langle\frac{\lambda_i-\lambda_{i-1}}{2}\rangle_{\lambda_i}$ and $\langle\frac{\lambda_{i+1}-\lambda_i}{2}\rangle_{\lambda_i}$. This has shown to get much better results than only one direction sampling.[213]

3.3.1.2 Thermodynamic integration

TI is an alternative method which originated from Kirkwood's coupling parameter method.[214] It assumes that the free energy A is a continuous function and it can be written as an integral over the parameter λ:

$$\Delta A = \int_0^1 \left\langle \frac{\partial \mathcal{H}}{\partial \lambda} \right\rangle_\lambda d\lambda \tag{26}$$

In practice, the number of $\left\langle \frac{\partial \mathcal{H}}{\partial \lambda} \right\rangle_\lambda$ is finite so the integral can be approximately replaced by summation using numerical methods.[215]

3.3.1.3 Bennett Acceptance Ratio

BAR is the third option which samples both directions on current ensemble:[216]

$$\Delta A = k_B T \sum_{i=1}^{N-1} \left(\ln \frac{\langle f(U_{\lambda_i} - U_{\lambda_{i+1}} + C)\rangle_{\lambda_{i+1}}}{\langle f(U_{\lambda_{i+1}} - U_{\lambda_i} - C)\rangle_{\lambda_i}} + C \right) \tag{27}$$

where $f(x)$ is the Fermi function:

$$f(x) = \frac{1}{1 + e^{x/k_B T}} \tag{28}$$

and C is an energy offset related to the partition functions of states λ_i and λ_{i+1}. The free energy difference is obtained when the following condition is fulfilled:

$$\langle f(U_{\lambda_i} - U_{\lambda_{i+1}} + C)\rangle_{\lambda_{i+1}} = \langle f(U_{\lambda_{i+1}} - U_{\lambda_i} - C)\rangle_{\lambda_i} \tag{29}$$

It is an efficient method as it is converged without requiring very big overlap between two intermediates.[213]

3.3.2 Implementations

This subsection introduces implementation detail. As the kinetic energy changes always cancel out in free energy difference, only the potential energy V is of interest.

3.3.2.1 Intermediates

The coupling parameter λs used for intermediates are between 0 and 1 and are not necessarily equally spaced. In transformation the factor scales the energy of one state from full to zero, and from zero to full for the other state:

$$V(\mathbf{r}, \lambda) = (1-\lambda)^n V_A(\mathbf{r}) + \lambda^n V_B(\mathbf{r}) \tag{30}$$

If $n = 1$, the energy scaling is linear, which is applied in paper IV of this thesis. As shown in the equation when λ is not 0 or 1 the system is in unphysical hybrid states, so this process is referring to alchemical transformation. Those intermediates (or windows) enable the system to visit the complete phase space between initial and final states. Obviously the choice of path is arbitrary, but for efficiency some high energy regions need denser windows and low energy regions do not need. More windows can be inserted if phase space coverage is not adequate, without rerunning other windows.

3.3.2.2 Hybrid topologies

The transformation from molecule A to B may include changes of atom types, charges and topologies. To have the correct sampling, the difference between A and B should not be very large. As MD simulation never creates or annihilates atoms, the hybrid topology needs to be built. There are two types of hybrid topology, which have system-dependent performance but solve the final results identically.

In dual topology the atoms that are different in the two states have separate coordinate sets (dual position). Both sets have full interactions with the environment, but never interact with each other. Since the coordinates in the two states are independent, the transformation is equal to turning off A-atoms meanwhile turning on B-atoms. At the end, A-atoms do not disappear but have no charges and L-J interactions (dummy atoms), so do the B-atoms at the beginning. In this process, the bonded energies of either state do not need to transform. The system potential energy in dual topology is:

$$V(\lambda) = (1-\lambda)V_A^{\mathrm{nb}} + \lambda V_B^{\mathrm{nb}} + V_A^{\mathrm{bd}} + V_B^{\mathrm{bd}} + V_0^{\mathrm{env}} \tag{31}$$

where the superscripts nb and bd denote nonbonded energy (charge and L-J) and bonded energy (bond, angle and dihedral) respectively, and env denotes the energy of environment which includes the unchanged atoms in target molecule and atoms in all other molecules.

In single topology, the atoms from different states have the identical coordinates (single position). Thus at the end A-atoms are replaced by B-atoms, and so are B-atoms at the beginning. Since the connectivity (bonds, angles and dihedrals) has different parameters between both states, during the transformation the effective force constants mixed by λ are acting on. As consequence both nonbonded and bonded energies are changed. If the topologies are different, *i.e.* atoms in one state have no companions in the other state, the

dummy atoms are added for them. The transformation of those atoms becomes the same way as dual topology. The system potential energy in single topology is:

$$V(\lambda) = (1-\lambda)\left(V_A^{\mathrm{nb}} + V_A^{\mathrm{bd}\prime}\right) + \lambda\left(V_B^{\mathrm{nb}} + V_B^{\mathrm{bd}\prime}\right) + V_0^{\mathrm{env}} \tag{32}$$

where the prime in $V^{\mathrm{bd}\prime}$ implies bonded energy for dummy atoms are excluding (belong to U_0^{env}). In eq 31 and 32, the potential energies of single and dual topology are different so the resulting ΔA are different. Thus the "reference" system is needed to subtract the redundant hybrid energy. This will be introduced in thermodynamic cycle in section 3.3.3.

3.3.2.3 End point issue

In the case of different topologies, the atoms have to be "created" or "annihilated" (charge and L-J on or off) from environmental space at the end point (λ = 0 or 1), which may cause atom clash for overlapping radii. In MD simulation the configuration of high repulsive interaction is never sampled. However in integration (TI) or evaluation at the end ensemble, the singularity will be met and the calculation diverges. A remedy is to use soft-core potential for L-J terms so that when the distance is approaching zero the repulsive energy is finite. Two soft-core methods were developed. The one using linear coupling for L-J potential[217] is applied in PERT module[7] of CHARMM. It modified the denominator in the L-J repulsive term (eq 1) as $(r_{ij}^2 + \delta(1-\lambda))^6$, where δ is the user defined parameters to control the strength of interaction. The implementation of soft-core does not shift the energy minimum position but will change the depth of the potential well somehow, and this bias is vanishing as the distance increases.

3.3.3 Thermodynamic cycle

The way of calculating free energy difference is alchemical. So the final results with chemical significance has to be calculated using a thermodynamic cycle, taking advantage that free energy is a state function.

3.3.3.1 Alchemical vs. chemical

Assuming molecules A and B are ligands of a receptor, to get the binding free energy difference the experiment will measure $\Delta A_{\mathrm{bind}}^{A}$ and $\Delta A_{\mathrm{bind}}^{\mathrm{B}}$ individually and then subtract the difference. However this is hardly done by computational model, because directly sampling a ligand moving from infinite distance into a receptor or moving from receptor to infinite distance is difficult. Instead it is possible to calculate the free energy difference transforming from complex with ligand A to complex with ligand B ($\Delta A_{\mathrm{com}}^{\mathrm{A}\to\mathrm{B}}$) and from free ligand A to free ligand B ($\Delta A_{\mathrm{lig}}^{\mathrm{A}\to\mathrm{B}}$) (Figure 4A). Finally the computational path and experimental path can form a closed cycle, in which the subtraction from either path is equivalent:

$$\Delta\Delta A_{\mathrm{bind}}(A\to B) = \Delta A_{\mathrm{bind}}^{\mathrm{B}} - \Delta A_{\mathrm{bind}}^{\mathrm{A}} = \Delta A_{\mathrm{com}}^{\mathrm{A}\to\mathrm{B}} - \Delta A_{\mathrm{lig}}^{\mathrm{A}\to\mathrm{B}} \tag{33}$$

The thermodynamic cycle is always required for alchemical transformation. Once taking subtraction between two systems with the same topology, all the energies of self-term cancel out. The net different $\Delta\Delta A(A \to B)$ has the exact meaning, which represents the free energy difference between A and B in response to two different environments. When the environments are solvent and receptor $\Delta\Delta A$ is the relative binding free energy, and when the environments are vacuum and solvent it is the relative solvation free energy.

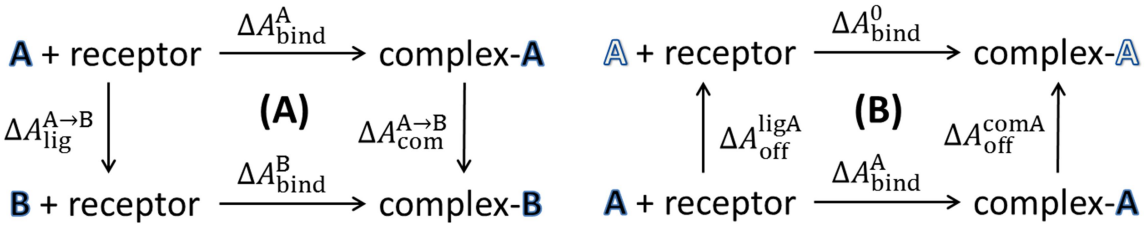

Figure 4. Thermodynamic cycles for determination of free energy. (A) The relative binding free energy between two ligands. $\Delta A_{\mathrm{bind}}^{\mathrm{A}}$ and $\Delta A_{\mathrm{bind}}^{\mathrm{B}}$ are binding free energies for ligand A and B, and $\Delta A_{\mathrm{lig}}^{\mathrm{A\to B}}$ and $\Delta A_{\mathrm{com}}^{\mathrm{A\to B}}$ are free energy differences of mutating complex that binds to A into complex that binds to B, respectively. (B) The absolute binding free energy of ligand A. The filled letter means nonbonded interaction turning on (real atoms) while hallow letter means nonbonded interaction turning off (dummy atoms). $\Delta A_{\mathrm{off}}^{\mathrm{ligA}}$ and $\Delta A_{\mathrm{off}}^{\mathrm{ligA}}$ are free energy differences of turning off nonbonded interaction on free ligand and bound ligand respectively. $\Delta A_{\mathrm{bind}}^{0}$ and $\Delta A_{\mathrm{bind}}^{\mathrm{A}}$ are the free energy differences of moving dummy and real molecule A into receptor. All the transformations take place in solution. The left vertical leg in each graph only calculates the ligand transformation, because the difference of receptor is zero (unvaried).

3.3.3.2 Absolute free energy

The straightforward results from alchemical transformation are relative free energy difference. Sometime the absolute free energy ($\Delta A_{\mathrm{bind}}^{\mathrm{A}}$ or $\Delta A_{\mathrm{bind}}^{\mathrm{B}}$) is important, as it can be compared with experimental data directly. This is solved using another thermodynamic cycle (Figure 4B). The ligand atoms are turned off in solvent and in receptor respectively. To close the cycle, the ligand in off state is moved from solvent into receptor. When nonbonded energy is off, the ligand is a dummy molecule, so $\Delta A_{\mathrm{bind}}^{0} = 0$. Thus the absolute binding free energy is given by:

$$\Delta A_{\mathrm{bind}}^{\mathrm{A}} = \Delta A_{\mathrm{bind}}^{0} + \Delta A_{\mathrm{off}}^{\mathrm{ligA}} - \Delta A_{\mathrm{off}}^{\mathrm{comA}} = \Delta A_{\mathrm{off}}^{\mathrm{ligA}} - \Delta A_{\mathrm{off}}^{\mathrm{comA}} \tag{34}$$

The trickiest calculation is $\Delta A_{\mathrm{off}}^{\mathrm{comA}}$, where turning off the ligand makes its phase space become whole system which is difficult to sample. The solution is to apply a set of restraints between ligand and receptor before turning off the ligand, so the sampling of dummy ligand is adequate. Thus the energy contribution in $\Delta A_{\mathrm{off}}^{\mathrm{comA}}$ has nonbonded term and restraint term. The latter has to be deducted analytically.[218] Alternatively the same restraints can be applied on both legs ($\Delta A_{\mathrm{off}}^{\mathrm{comA}}$ and $\Delta A_{\mathrm{off}}^{\mathrm{ligA}}$) and the effects of the restraints are expected to cancel out. But this is not physically rigorous in the case that the environments are very different (Here solvent and receptor), because the nonbonded interactions responding to the restraint are not sufficiently considered.

3.3.4 Potential of mean force

The potential of mean force (PMF) is the free energy surface along a reaction coordinate (path). Different from the alchemical path of calculating free energy difference, this surface has physical meaning which describes the work required to overcome the interaction between two configurations. It requires umbrella sampling to pass over the barriers along the path.

3.3.4.1 Umbrella sampling

Umbrella sampling[219] applies a biased potential that enables high energy regions to be visited, *i.e.* to bridge the energetically disconnected regions along the path. The reaction coordinate between two states is divided into several windows. The biasing potential w_i applied in each window is a function of the reaction coordinate ξ. It usually takes a harmonic form:

$$w_i(\xi) = \frac{k_i}{2}(\xi - \xi_{i,0})^2 \tag{35}$$

where $\xi_{i,0}$ is the reference coordinate and k_i is the force constant. The system thus is added by this biased potential and the probability distribution sampled in MD simulation is biased (ρ_i^{b}), and related to the coordinate ξ:[220]

$$\rho_i^{\mathrm{b}}(\xi) = \frac{\int \delta[\xi(\mathbf{r}) - \xi]\exp\big(-[V_i(\mathbf{r}) + w_i\big(\xi(\mathbf{r})\big)]/k_BT\big)d\mathbf{r}}{\int \exp\big(-[V_i(\mathbf{r}) + w_i\big(\xi(\mathbf{r})\big)]/k_BT\big)d\mathbf{r}} \tag{36}$$

The numerator of this equation means the integral over all coordinates except ξ. The unbiased distribution ρ_i^{u} can be written as a function of ρ_i^{b}. The free energy on one window based on its unbiased distribution is:

$$A_i(\xi) = -k_BT \ln \rho_i^{\mathrm{u}}(\xi) = -k_BT \ln \rho_i^{\mathrm{b}}(\xi) - w_i(\xi) + F_i \tag{37}$$

where F_i is the free energy offset at window i, which is a constant depending on the biased potential w_i. For many-window systems, the weighted histogram analysis method (WHAM)[219] calculates F_i and the global unbiased probability distribution ρ^{u} iteratively. The global free energy A (PMF) is solved from the finally optimized ρ^{u}.

4 SUMMARY OF THE PAPERS

4.1 A FORCE FIELD FOR RNA MODIFICATIONS

An empirical force field of 112 known modified ribonucleotides is parameterized in this study, in line with the philosophy behind the development of the CHARMM additive force field. The L-J parameters were transferred from previous force fields of nucleic acids, organic molecules and carbohydrates. The other parameters were optimized targeting QM data and further refined towards experimental data. It is compatible with other components in the CHARMM force field.

The initial guess is generated by ParamChem server using CGenFF program (Figure 5). The optimization of the partial atomic charges is based on the reproduction of water-model compound minimum interaction energies. A least squares fitting approach is used to reduce the RMS difference between QM and MM to be averagely 0.25 kcal/mol for all models. The dipole moment of 20% overestimation is the second target to be reproduced. The equilibrium values and force constants of harmonic terms are optimized by targeting the energy-minimized geometries and vibrational analysis respectively. The average RMS difference over all molecular terms were 0.015 Å, 1.46°, 2.13° and 0.99° for bond, angle, stiff and improper dihedral geometry respectively, and 7.0% for vibrational frequency terms. The dihedral angles were carefully optimized based on PES scans. The new glycosidic and 2'-OMe torsions which correlate with sugar pucker were fine-tuned based on MD simulations to achieve good agreement between simulations and experiments.

Emphasis was placed on some common nucleotides, dihydrouridines, pseudouridines, 7-methylguanosine and 2'-OMe nucleotides, whose conformations were shown to be consistent with experimental data (Figure 5). Important properties include that the A-type conformation can be stabilized by Ψ, m^7G and 2'OMe, and destabilized by D. Correlations of *anti*/*north* for Ψ and m^7G and *Base*/*north* for 2'OMe nucleotides were observed, respectively, thereby illustrating how they stabilize RNA structures.

In cases where the χ and pucker were identical to canonical nucleic acids and could not be optimized, the charges of base were adjusted based on the chemical fact that electron-withdrawing 5-substitutions increase *north* for pyrimidines whereas electron-donating 5-substitutions increase *south*.[101,102] However, compromise is required that the charges are fine-tuned toward the experimental evidence while the results of water interaction are not adversely affected too much. An example is ac^4C whose *north* population was improved by adjusting the base charges, leading to better agreement with NMR data.

Further refinement beyond targeting the QM data is limited because experimental structures are lacking for many modified nucleic acids. For many cases staying with QM data is sufficient. The example is N-methylated bases, in which the parameters targeting on QM potential surface are able to reproduce the experimental conformations. The discrepancies usually come from the solvent effect on the torsions ending with polar atoms, in which

validation is necessary. Nevertheless it is always recommended to refine the parameters based on explicit solvent simulations as long as the experimental data are available.

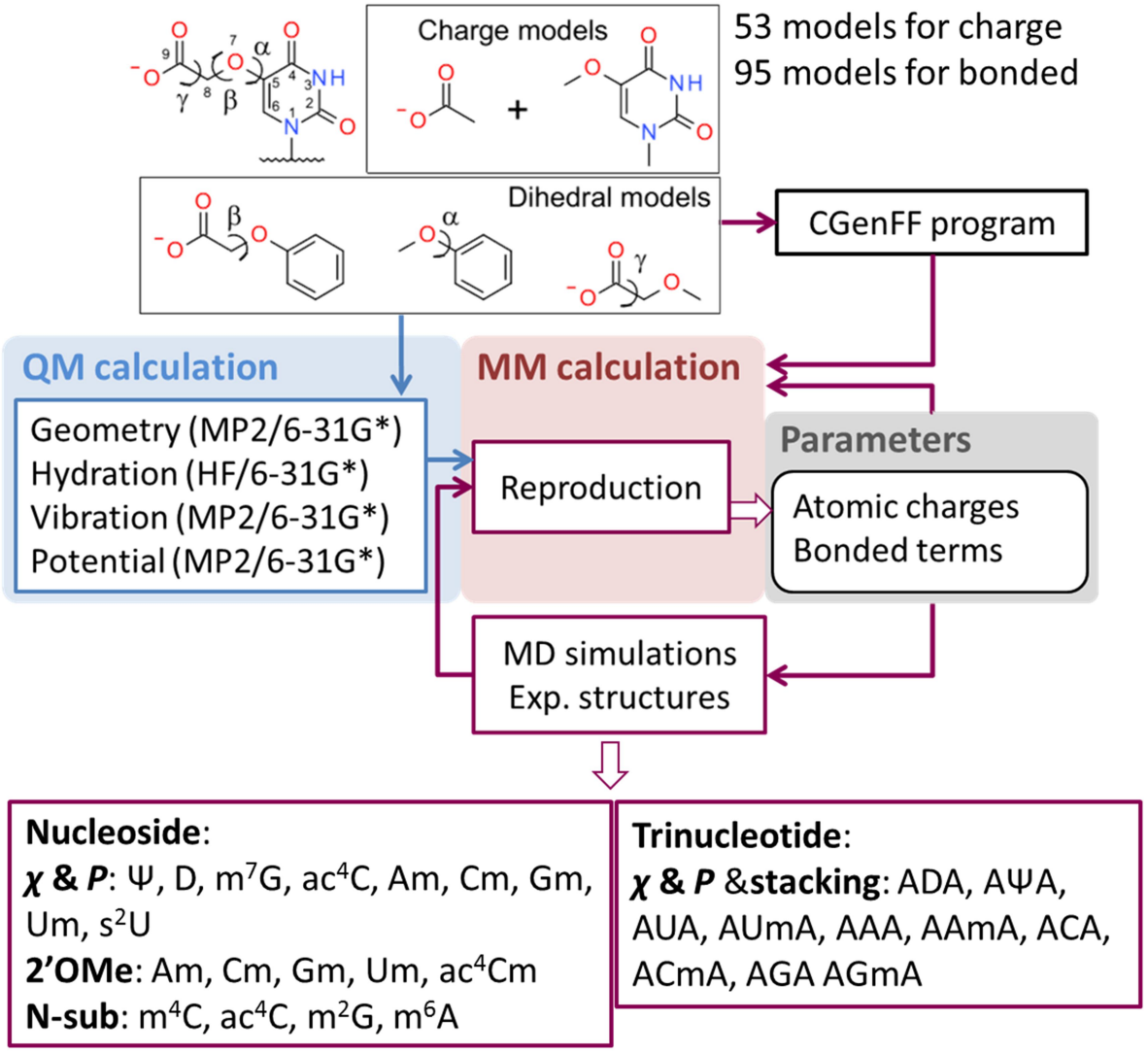

Figure 5. Flowchart of parametrisation. The upper boxes show the example of model compounds of charges and dihedral torsions for cmo^{5}U. The middle boxes briefly show the process of parametrisation. The blue frame is for QM calculations, violet for MM calculations and black for output parameters. The two bottom boxes show the modified nucleosides which were involved in validation referring to experimental structures, with respect to their conformations of interest.

4.2 EFFECTS OF MODIFICATIONS IN TRNA

Modified nucleotides are ubiquitous and important to tRNA structure and function. To understand their effects on tRNA conformation, totally 6-μs MD simulations were performed on yeast tRNA^{Phe} (1EHZ) and tRNA$_i^{Met}$ (1YFG), *E. coli* tRNA$_f^{Met}$ (3WC5) and HIV tRNA^{Lys} (1FIR) with or without nucleotide modifications and Mg^{2+} ions.

In ASL, yW37 combined with a Mg^{2+} in the anticodon loop contributes to the stabilization and rigidity of the anticodon in tRNA^{Phe} (Figure 6A), and Gm34 and Cm32 may also contribute to the stabilization. mcm5s2U34 keeps rigid conformation but unmodified U34 similarly does not show much destabilization (Figure 6C). (ms^2)t^{6}A in 1YFG and 1FIR were

not found to improve the stability of the anticodon in free tRNA. Such modifications probably contribute to the stabilization of the anticodon:codon mini-helix in ribosome. No significant impact of adenosine protonation, $^{+}$A37 and $^{+}$A38, is observed on *E. coli* $tRNA_f^{Met}$.

In D loop, D16/D17 keeps the specific pucker of G18 and G19 (Figure 6D) which participate in conserved base pair G18:U55 and G19:C56 that are necessary for interactions between the D arm and the T/V loops. Mg^{2+} ions are abundant in D loop and at least one Mg^{2+} obviously contributes to stabilization of the D loop backbone in the simulation (Figure 6B). *E. coli* $tRNA_f^{Met}$ has different conformations in D loop and ASL from other structures, which might link to its special function.

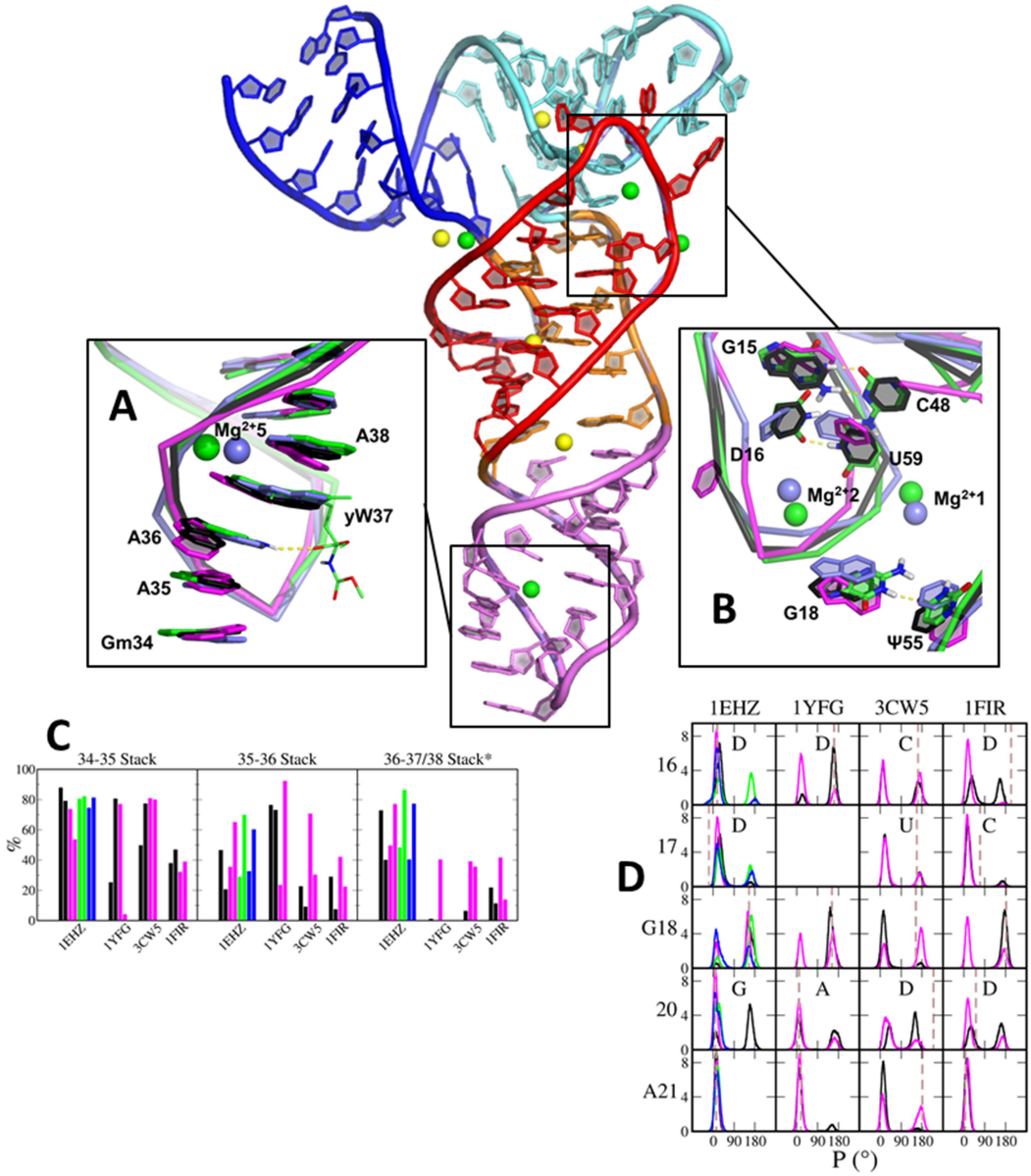

Figure 6. Graphic summary of tRNA simulations. The 3D structure of 1EHZ is shown with CCA arm truncated. The Mg^{2+} ions are shown in spheres in which green ones are stable in MD simulation and yellow ones are mobile. The DSL and ASL are shown in detail: (A) the A loop structural superimposition of 1EHZ; (B) the D loop structural superimposition of 1EHZ; (C) the percentage of anticodon base stacking for four tRNAs, in which dashed lines are the conformations of crystal structure; and (D) the sugar pucker distributions of D loop nucleotides in four tRNAs. The color scheme for the graphs A, B, C and D is black (WT), magenta (CAN), green (WTMG), blue (CNMG) and slate (crystal).

Some base-methylated nucleotides like m^5C, m^7G and $m^2{}_2G$ exhibited slightly improved hydrogen bonding compared to the corresponding unmodified nucleotides. The strong stabilization of Ψ was not observed, since the U counterparts had similar structural properties. An unexpected destabilization is present in TΨC loop, maybe because of the underestimation of T54-Ψ55 stacking and/or m1A58:T54 hydrogen bonding.

Overall, the role of many modifications in maintaining the tRNA structure and impacting their dynamic properties is obvious compared with the unmodified systems. In addition, Mg^{2+} shows the importance in the stabilization of loop region, especially the folding of backbone. Some conformational differences were observed between the simulation and experiment, which may indicate the influence of crystal packing effects and the requirement of force field updating.

4.3 LNA ENHANCES THE DNA TRIPLEX FORMATION

In anti-gene strategy of modulating gene expression, a specific DNA duplex is recognized by TFO or invading oligonucleotides. The oligonucleotides should target DNA selectively and in high binding affinity. In collaboration with biological colleagues, the LNA effects on triplex formation were investigated using molecular dynamics (MD) combing hybridization analysis and electrophoretic mobility shift assay (EMSA). A 19mer duplex sequence (DS19) of regulator c-*MYC* was chosen used as the target (Figure 7A). The modeling shows the LNA inclusion in both TFO and duplex strand is beneficial to triplex formation by affecting conformational flexibility of not only single strand but also duplex.

As shown from EMSA, no triplex was formed in presence of the TFO sequence with pure DNA (TFO1) whereas triplex was detected after 24 h in the presence of TFO sequence with alternative DNA-LNA (TFO-5'DNA). The models of single strand TFO1 and TFO-5'DNA and triplex DS19·TFO5'DNA were built. In TFO1 DNA nucleotides are mainly in loosely *south*/*anti*, whereas in TFO-5'DNA all LNA nucleotides are restricted *north*/low-*anti* and DNA nucleotides are similar as in TFO1 but with more induced *north* conformation (Figure 7B, top). When triplex was formed, all the nucleotides are in *north*/low-*anti*, which is fit by LNA nucleotides intrinsically (Figure 7B, bottom). It therefore suggests LNA contained TFOs are conformationally pre-organized for major groove binding. This finding is helpful to future design of TFO.

The bisLNA invaded DNA duplex in two steps: first one part of bisLNA sequence targets the DNA duplex by forming HG base pairs thus forming a triplex; and second the other part of sequence clamps the duplex and competitively forms WC base pairs against one prototype strand. The models of hetero-and homo-duplexes were built (Figure 7A) to mimic the triplex conformations when the target DNA is invaded or not. The LNA-substitutions in the WC pyrimidine strand were found to alter the duplex conformation. $DS19^{Hetero}$ is unwinding so has larger diameter compared with DS19; and this helix shape is visually closer to triplex (Figure 7C). $DS19^{Hetero}$ is between A- and B-type helix, with base pairs remaining almost perpendicular to the helical axis (B like) but negative slide and x-displacement, and reduced

twist (A-like). This conformation is similar to both modified and unmodified DNA duplexes when TFO is bound (Figure 7D). This suggests LNA nucleotides also pre-organize the duplex to be the triplex favorable conformation, so that hetero duplex is superior to homo helix to accommodate a TFO. This finding extends the mechanism of bisLNA invasion: the efficiency is not only because LNA contained WC strand has better duplex stability, but also because of the faster formation of the intermediate triplex which moves the reaction equilibration towards forming hetero-duplex formation.

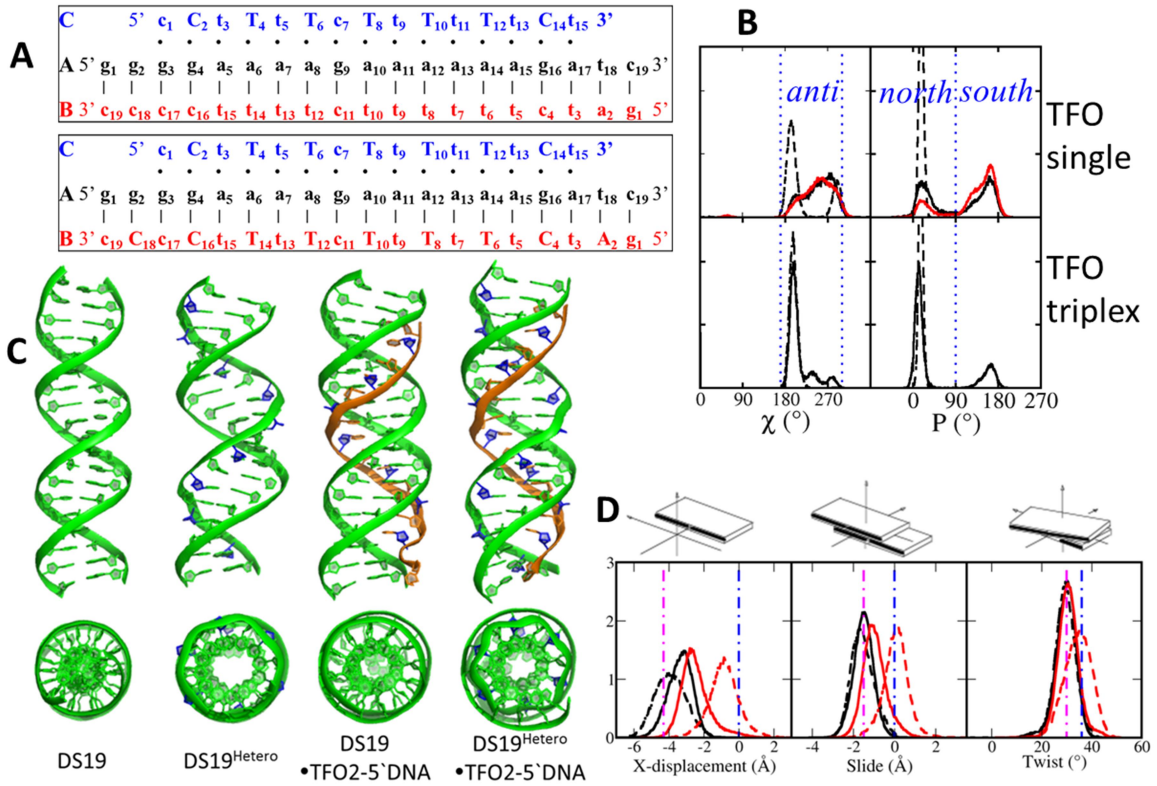

Figure 7. Graphic summary of molecular models of LNA contained duplex and triplex. (A) The triplex sequences forming by homo (upper, DS19) and hetero (lower, DS19^{Hetero}) duplexes. The TFO sequences are identical as TFO2-5'DNA. The uppercase is LNA and lowercase is DNA nucleoside. (B) Distribution of χ and P of TFO2-5'DNA (black) and TFO1 (red) as single strands free in solution (top) and as TFO bound to the DS19 (bottom). The solid lines represent the DNA nucleosides and dashed lines represent the LNA nucleosides. The regions separated by blue dotted lines show *anti* for χ (low *anti*: < 220° and high *anti*: > 270°), and *north* and *south* for P. C) The front view and top view of average structures of duplexes and the corresponding triplexes. LNA sugars are in blue and TFO strands are in orange. D) Distribution of base-pair-step parameters. DS19^{Hetero} and DS19 are in black and red respectively, and their duplexes are in dashed line while triplexes are in solid line. The vertical dash-dotted lines represent the corresponding value of ideal A-DNA (magenta) and B-DNA (blue) duplex.

4.4 FREE ENERGY PERTURBATION OF LOCKED RIBOSE

LNA improves the stability of DNA/RNA duplexes significantly, and therefore is of interest to know the free energy change between LNA and DNA nucleosides. The transformation requires the breaking of the bridged ribose ring, which leaves technical challenge of the transformation in free energy calculations.[10,221] We developed two protocols for the transformation between locked ribose and deoxyribose (LNA→DNA) using single and dual

topology of pyrimidine nucleosides. The computations were performed using CHARMM36 force field and the Bennett Acceptance Ratio method.

The first protocol is divided in three-steps: 1) the DNA angle and dihedral energies are turned on (*s0*), 2) the charges and the L-J parameters are transmuted (*s1*), 3) the LNA angle and dihedral energies are turned off (*s*2) (Figure 8A). The phase spaces are continuous between steps and the sampling in each step is fair. The conformational distributions of two end states are identical to MD simulation of canonical LNA and DNA nucleosides (Figure 8B). The free energy changes in each step are analyzed (Table 2). For the single topology *s0* can be omitted, and the main transformation is completed in *s1*. For the dual topology *s1* also involves the most significant energy change, but here *s0* and *s2* are nontrivial because they are required to reproduce correct ring conformations.

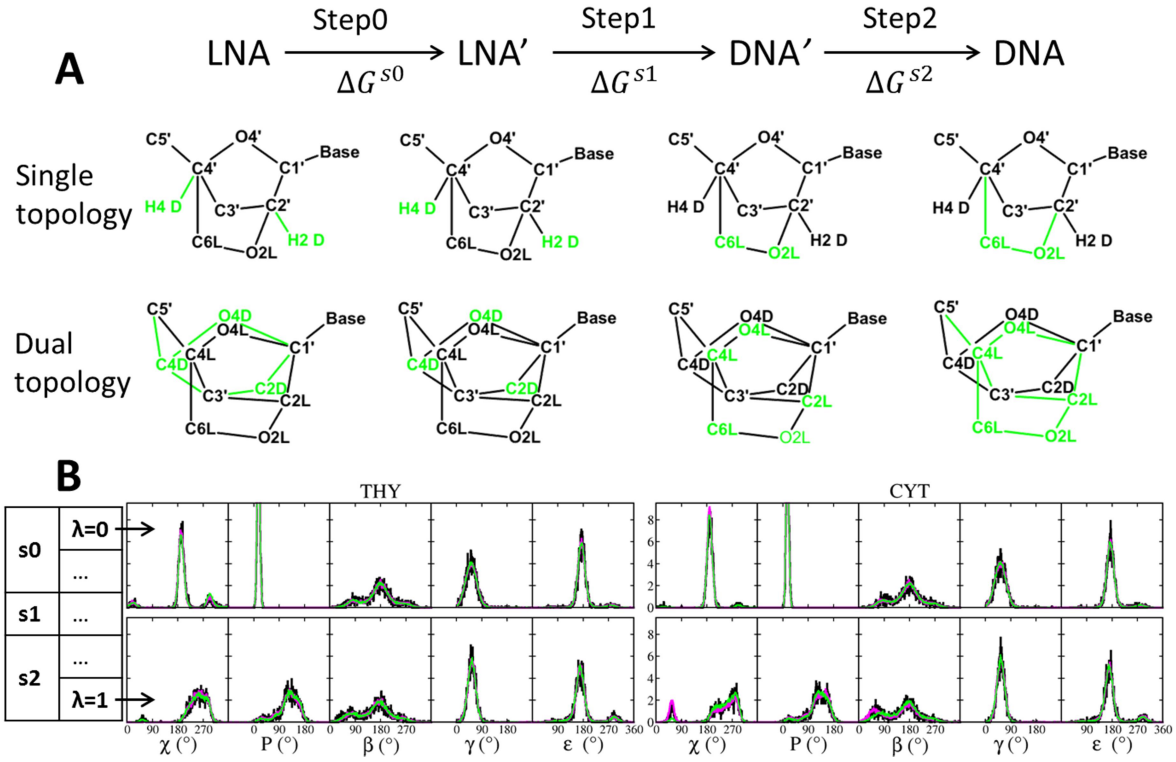

Figure 8. Scheme and conformational sampling of three-step protocol. (A) The transformation process and states are shown for LNA→DNA, blow which are the corresponding topologies. The real atoms and the bonded terms are in black, whereas the dummy atoms (with zero electrostatic and L-J interactions) and dummy bonded terms (with zero angle and dihedral but original bond force constants) are green. (B) The nucleoside conformational distributions of χ, P, β, γ and ε are shown for the initial (λ_{s0}=0, LNA) and final (λ_{s2}=1, DNA) states. Magenta and green distributions are summed from five replicas using single and dual topology, respectively, and black distribution is sampled from 200 ns standard MD simulations for regular LNA and DNA nucleosides.

Those steps can be partly or all performed together. As an example, the second protocol merges three steps into one step, *i.e.* mutating bonded and non-bonded terms at the same time. Since soft-core is not applied, this protocol requires some careful adaptation in the transformation of angle terms to make sure LNA and DNA atoms overlap on each other. This approach requires fewer windows than the three-step approach, so it is more efficient. Both approaches give similar results, *viz*. the solvation free energy for an LNA nucleoside is 1-2 kcal/mol more favorable than for a DNA (Table 2), because of the hydrophilicity of

oxymethylene. We are not aware of any direct measurement of the solvation free energy of an LNA nucleoside, but $\Delta\Delta G_{\mathrm{solv}}^{\mathrm{L}\to\mathrm{D}}$ is comparable with experimental data for cyclic and linear ether analogs.

Although most windows were run for 10 ns (20 ns for some in dual topology), 5 ns sampling time seems to be enough to get converged results for the three-step and one-step approaches in either topology, except for thymidine in the three-step approach of dual topology, where longer sampling time is suggested for the windows near λ_{s0}=1/λ_{s1}=0. Both three-step and one-step protocols can be used, but the one-step approach in dual topology requires the user to be familiar with the force field parameters of the molecule.

Table 2. The step-wise transformation free energy (kcal/mol) of converting LNA to DNA for thymidine and cytidine using single and dual topology. Numbers in parenthesis are the standard deviations (σ) from five runs.

		Single topology				Dual topology			
		ΔG_{aq}	ΔG_{vac}	$\Delta\Delta G$	$\Delta\Delta G_{\mathrm{solv}}^{\mathrm{L}\to\mathrm{D}}$	ΔG_{aq}	ΔG_{vac}	$\Delta\Delta G$	$\Delta\Delta G_{\mathrm{solv}}^{\mathrm{L}\to\mathrm{D}}$
T	*s0*	7.93 (0.03)	7.90 (0.07)	0.03 (0.08)	1.17 (0.18)	42.74 (0.09)	43.51 (0.06)	-0.77 (0.11)	1.47 (0.22)
	s1	-82.81 (0.05)	-84.06 (0.01)	1.25 (0.05)		-73.94 (0.03)	-76.55 (0.02)	2.61 (0.04)	
	s2	-52.44 (0.15)	-52.33 (0.06)	-0.11 (0.16)		-105.73 (0.06)	-105.36 (0.18)	-0.37 (0.19)	
C	*s0*	7.89 (0.04)	7.96 (0.03)	-0.07 (0.05)	1.31 (0.12)	42.09 (0.13)	42.00 (0.04)	0.09 (0.14)	1.54 (0.18)
	s1	-75.60 (0.07)	-77.14 (0.00)	1.54 (0.07)		-67.32 (0.02)	-69.74 (0.02)	2.42 (0.03)	
	s2	-52.25 (0.07)	-52.09 (0.05)	-0.16 (0.09)		-104.56 (0.09)	-103.59 (0.06)	-0.97 (0.11)	

4.5 SUMMARY AND OUTLOOK

4.5.1 Conclusive remarks

There are only two chemical differences between DNA and RNA, in which DNA has the deoxyribose (without 2'-OH) and thymine (5-methyluracil). But their functions and the corresponding structures are remarkably different. DNA stores gene and mainly exists in B-helix in cellular environment, though structural conversions between B- and A-type or BI and BII are possible. RNA regulates gene expression. It has a large repertoire of possible secondary and tertiary structures, such as A-helix, hairpin, bulge, junction and pseudoknot. Modifications further extend such diversity so that the short molecules like tRNA have well-defined conformation in ribosome.

Force field determines the potential energy between atoms, so its accuracy is important to MD simulation of nucleic acids. The current additive CHARMM force field has been refined to reproduce reasonable conformational population for nucleic acid structures.[179,181,197] In line

with the previous parametrisation strategy, the chemical space of bases is extended for RNA modifications in paper I. Besides the parametrisation on base modifications, the glycosidic torsion and sugar pucker are re-optimized for new base scaffold, and correlations between modified bases and their sugar conformations are considered. Crystal and NMR structures of some modified nucleosides and oligonucleotides are collected to validate and refine the parameters. A limitation is that known experimental data are insufficient for three quarters of the modifications. Although for some types of modification staying with QM data is enough, fine-tuning parameters targeting experimental monomer or oligomer is never trivial for biomolecular force field.

In paper II the prototype tRNA with modifications and the tRNA with only canonical nucleotides are simulated. The quality of starting crystal structure strongly affects the results. For the structure of highest resolution (PDB ID: 1EHZ), both modified and unmodified systems kept stable in simulations, but modifications maintain more stable local conformation. The stabilization of magnesium ions is also clearly shown. For low resolution structures or those with crystal packing, local disturbances are observed in simulations for both modified and unmodified systems. It is therefore hard to illustrate the effect of some modifications. A few local instabilities in modified nucleotides might suggest some revision of force field is needed. The present force field and observation on tRNA structures provide good starting point for simulations on larger and more complicated RNA systems.

LNA is one of the commonly used artificial modifications in DNA binding oligonucleotide. The experiment shows DNA strands containing LNA has enhanced triplex formation. The simulations on single strand, duplex and triplex models show LNA pre-organizes the TFO and duplex structures to be triplex preferred conformations, so that they do not need much conformational adaption upon binding. This study evaluates the modified and unmodified structures in detail and provides more theoretic basis about LNA-substitution in TFO and bisLNA for DNA duplex invasion. One limitation is the protonation states of cytosine cannot respond to the environmental change (pH and base pair opening) in simulations, which may lead biased results for C rich sequences.

Thermodynamic effect of LNA in DNA context requires free energy calculation, but dealing with the bridged ring of locked ribose introduces several technical problems. Possible solution is reported in last paper where angle and dihedral terms are transformed to achieve the conformational conversion between DNA and LNA ribose. However the transformation energies between deoxyribose and locked ribose in target and reference system are big, from which the small free energy difference is obtained. This will lead the loss of precision when the calculation is taken in larger systems, because of expensive sampling in both target and reference systems.

The increasingly developed experimental techniques enable to solve bigger biomolecular structures and clarify the puzzle of complicated function involved in nucleic acids. This has changed researchers' mind in recent decades. A lot of ncRNAs which were attributed as junk RNA are now drawing tremendous interest. The combination of experiment and simulations

becomes a more efficient way to explore complicated nucleic acid systems. The improved computational devices and algorithm also make it possible to handle the system of millions atoms and the time scale with millisecond. The force fields with new particle properties and the new ideas of sampling are in development to allow larger scale simulation with rational results.

4.5.2 Further work

4.5.2.1 tRNA recognition

The tRNA conformation plays an important role in protein translation. The first two papers supply better understanding of RNA modifications and tRNA structure, based on which the simulation on more complicated RNA models can be performed. Hirsh suppressor, where two different mutants lead to the same miscoding, is a good system to study tRNA recognition because both kinetic and structural data have been provided in literature. Investigating the roles of different tRNA structural domains is helpful to further understand ribosomal selection.

4.5.2.2 More details of LNA and triplex formation

More detail of LNA position and number affecting on DNA contexts will be studied. It is interesting to know the similarity of triplex structures of different duplex sequences, so that the conformational parameters can be the criteria to identify the triplex forming ability for duplexes. This will be helpful to future design of oligonucleotides for DNA duplex invasion.

The protocol free energy calculation supposed in paper IV will be further applied in duplex and triplex structures. Meanwhile it might be necessary to update the old LNA parameters[159] according to experimental data. This study will aim to generalize the usefulness and efficiency (GPU implementation) of free energy calculation on large systems containing bridged ring.

Involved in the design of restrict RNA enzyme,[136] MD simulation has been performed on PNA-RNA duplex which mimics the bulge domain of active site.[222] The updating of PNA parameters based on the old version[223] will be completed in collaboration with other groups. The new parameters are expected to support further study of DNA or RNA binding oligonucleotides using PNA.

5 ACKNOWLEDGEMENTS

Stockholm is my second home city now. I am grateful to Sweden which shows me a different but beautiful world. The excellent characters I learned from this country will guide me in whenever future. I spend my calmest 5 years trying almost only one thing, finishing my PhD. Although not always delighted, I survived with the help from people.

I appreciate my supervisor Lennart who not only advises me on all my projects but also introduces me into the scientific community. More importantly, I learned from Lennart how a scientist treats a data representation even a sentence rigorously.

My appreciation is to my senior colleague Alessandra who generously shares her knowledge and intelligence to help me proceed and catch my problems from the detail.

I would like to thank Alex who instructs me the force field, conducts me in communication with his lab and helps me in career plan.

I also thank Roger, my co-supervisor and also collaborator, who shares his intelligent idea from the different field.

My current colleague Yossa, for both lab work and gym work; my previous colleagues Arzu, Eva, Olof, Mika, and especially Mauricio and Alok who shared me experience and tricks when I was a fresh; and also Andrey and Nicolas following whom I learned the specialized knowledge.

My experimental collaborators, Vladimir, Rula, Ted and Karin, and Alice in Roger's lab, who provide me the wonderful experience of solving interdiscipline questions together.

My force field collaborators Alexey and especially Kenno who guided me in professional detail with their advanced insight; and Alex's lab members especially Justin and Jing, from whom I gained fruitful discussion in U.S.

My fellows Ting & Jian, Yumei & Ning, Jiyu, Rongrong, Wenbo, Yichun, Zhiqiang, Fangjie, Qie, Yimeng and Tenghao, and senior colleagues Chunyan and Hui, who make the cordial Chinese atmosphere in Novum. My Chinese friends, Guanghong & Gen, Lin, Wei, etc., for the cheerful corridor time, and Yan for supplying me the first accommodation. Other Chinese societies, who brought many joyful moments for my foreign life.

My office neighbors Carsten's group, Olle and Peter's group especially Debora, who provide a nice working environment and mind exchange of different culture.

My domestic supervisor Yun, who helped me apply for the PhD program; my domestic colleagues in Sweden, Yaozong, Guanglin and Xianqiang, for making fun in research and leisure; and definitely the China Scholarship Council, my four-year support.

My administration and IT colleagues especially Monica, Lena, Eva, Maria, Thomas, Anders and Erik, for your kind help with elegant smile.

The special gratitude is to my family outside Sweden. My mom & dad, for bringing me up and tolerating my self-will, and standing behind me wherever I am. My beloved zaizi, in whatever cloudy and dark days, you always paint our moments in bright colors.

6 REFERENCES

1. D. Hirsh. Tryptophan transfer RNA as the UGA suppressor. *J. Mol. Biol.* **1971**, *58*(2), 439-458.

2. P. E. Nielsen; M. Egholm; R. H. Berg; O. Buchardt. Sequence-Selective Recognition of DNA by Strand Displacement with a Thymine-Substituted Polyamide. *Science.* **1991**, *254*(5037), 1497-1500.

3. S. Obika; D. Nanbu; Y. Hari; K. Morio; Y. In; T. Ishida; T. Imanishi. Synthesis of 2'-O,4'-C-methyleneuridine and -cytidine. Novel bicyclic nucleosides having a fixed C-3,-endo sugar puckering. *Tetrahedron Lett.* **1997**, *38*(50), 8735-8738.

4. A. A. Koshkin; S. K. Singh; P. Nielsen; V. K. Rajwanshi; R. Kumar; M. Meldgaard; C. E. Olsen; J. Wengel. LNA (Locked Nucleic Acids): Synthesis of the adenine, cytosine, guanine, 5-methylcytosine, thymine and uracil bicyclonucleoside monomers, oligomerisation, and unprecedented nucleic acid recognition. *Tetrahedron.* **1998**, *54*(14), 3607-3630.

5. D. A. Case; T. E. Cheatham; T. Darden; H. Gohlke; R. Luo; K. M. Merz; A. Onufriev; C. Simmerling; B. Wang; R. J. Woods. The Amber biomolecular simulation programs. *J. Comput. Chem.* **2005**, *26*(16), 1668-1688.

6. B. R. Brooks; R. E. Bruccoleri; B. D. Olafson; D. J. States; S. Swaminathan; M. Karplus. Charmm - a Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *J. Comput. Chem.* **1983**, *4*(2), 187-217.

7. B. R. Brooks; C. L. Brooks, 3rd; A. D. Mackerell, Jr.; L. Nilsson; R. J. Petrella; B. Roux; Y. Won; G. Archontis; C. Bartels; S. Boresch; A. Caflisch; L. Caves; Q. Cui; A. R. Dinner; M. Feig; S. Fischer; J. Gao; M. Hodoscek; W. Im; K. Kuczera; T. Lazaridis; J. Ma; V. Ovchinnikov; E. Paci; R. W. Pastor; C. B. Post; J. Z. Pu; M. Schaefer; B. Tidor; R. M. Venable; H. L. Woodcock; X. Wu; W. Yang; D. M. York; M. Karplus. CHARMM: the biomolecular simulation program. *J. Comput. Chem.* **2009**, *30*(10), 1545-1614.

8. D. Van der Spoel; E. Lindahl; B. Hess; G. Groenhof; A. E. Mark; H. J. C. Berendsen. GROMACS: Fast, flexible, and free. *J. Comput. Chem.* **2005**, *26*(16), 1701-1718.

9. R. Aduri; B. T. Psciuk; P. Saro; H. Taniga; H. B. Schlegel; J. SantaLucia. AMBER Force Field Parameters for the Naturally Occurring Modified Nucleosides in RNA. *J. Chem. Theory Comput.* **2007**, *3*(4), 1464-1475.

10. S. Liu; L. L. Wang; D. L. Mobley. Is Ring Breaking Feasible in Relative Binding Free Energy Calculations? *J. Chem. Inf. Medel.* **2015**, *55*(4), 727-735.

11. A. D. MacKerell; N. K. Banavali. All-atom empirical force field for nucleic acids: II. Application to molecular dynamics simulations of DNA and RNA in solution. *J. Comput. Chem.* **2000**, *21*(2), 105-120.

12. N. Foloppe; L. Nilsson; A. D. MacKerell, Jr. Ab initio conformational analysis of nucleic acid components: intrinsic energetic contributions to nucleic acid structure and dynamics. *Biopolymers.* **2001**, *61*(1), 61-76.

13. N. Foloppe; A. D. MacKerell. Contribution of the phosphodiester backbone and glycosyl linkage intrinsic torsional energetics to DNA structure and dynamics. *J. Phys. Chem. B.* **1999**, *103*(49), 10955-10964.

14. P. Yakovchuk; E. Protozanova; M. D. Frank-Kamenetskii. Base-stacking and base-pairing contributions into thermal stability of the DNA double helix. *Nucleic Acids Res.* **2006**, *34*(2), 564-574.

15. P. Hobza; J. Sponer. Structure, energetics, and dynamics of the nucleic acid base pairs: Nonempirical ab initio calculations. *Chem. Rev.* **1999**, *99*(11), 3247-3276.

16. N. Foloppe; B. Hartmann; L. Nilsson; A. D. MacKerell. Intrinsic conformational energetics associated with the glycosyl torsion in DNA: A quantum mechanical study. *Biophys. J.* **2002**, *82*(3), 1554-1569.

17. C. Altona; M. Sundaralingam. Conformational analysis of the sugar ring in nucleosides and nucleotides. A new description using the concept of pseudorotation. *J. Am. Chem. Soc.* **1972**, *94*(23), 8205-8212.

18. N. Foloppe; A. D. MacKerell. Conformational properties of the deoxyribose and ribose moieties of nucleic acids: A quantum mechanical study. *J. Phys. Chem. B.* **1998**, *102*(34), 6669-6678.

19. N. Foloppe; A. D. MacKerell, Jr. Intrinsic conformational properties of deoxyribonucleosides: implicated role for cytosine in the equilibrium among the A, B, and Z forms of DNA. *Biophys. J.* **1999**, *76*(6), 3206-3218.

20. E. J. Denning; A. D. MacKerell, Jr. Intrinsic contribution of the 2'-hydroxyl to RNA conformational heterogeneity. *J. Am. Chem. Soc.* **2012**, *134*(5), 2800-2806.

21. F. DiMaio; X. Yu; E. Rensen; M. Krupovic; D. Prangishvili; E. H. Egelman. A virus that infects a hyperthermophile encapsidates A-form DNA. *Science.* **2015**, *348*(6237), 914-917.

22. R. E. Dickerson; H. L. Ng. DNA structure from A to B. *Proc Natl Acad Sci U S A.* **2001**, *98*(13), 6986-6988.

23. W. K. Olson; M. Bansal; S. K. Burley; R. E. Dickerson; M. Gerstein; S. C. Harvey; U. Heinemann; X. J. Lu; S. Neidle; Z. Shakked; H. Sklenar; M. Suzuki; C. S. Tung; E. Westhof; C. Wolberger; H. M. Berman. A standard reference frame for the description of nucleic acid base-pair geometry. *J. Mol. Biol.* **2001**, *313*(1), 229-237.

24. R. E. Dickerson. Definitions and Nomenclature of Nucleic-Acid Structure Components. *Nucleic Acids Res.* **1989**, *17*(5), 1797-1803.

25. C. O. Pabo; R. T. Sauer. Protein-DNA Recognition. *Annu. Rev. Biochem.* **1984**, *53*, 293-321.

26. E. Stofer; R. Lavery. Measuring the Geometry of DNA Grooves. *Biopolymers.* **1994**, *34*(3), 337-346.

27. X. J. Lu; W. K. Olson. 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res.* **2003**, *31*(17), 5108-5121.

28. X. J. Lu; W. K. Olson. 3DNA: a versatile, integrated software system for the analysis, rebuilding and visualization of three-dimensional nucleic-acid structures. *Nat Protoc.* **2008**, *3*(7), 1213-1227.

29. R. Lavery; H. Sklenar. Defining the Structure of Irregular Nucleic-Acids - Conventions and Principles. *Journal of Biomolecular Structure & Dynamics.* **1989**, *6*(4), 655-667.

30. R. Lavery; M. Moakher; J. H. Maddocks; D. Petkeviciute; K. Zakrzewska. Conformational analysis of nucleic acids revisited: Curves+. *Nucleic Acids Res.* **2009**, *37*(17), 5917-5929.

31. G. H. Zheng; X. J. Lu; W. K. Olson. Web 3DNA-a web server for the analysis, reconstruction, and visualization of three-dimensional nucleic-acid structures. *Nucleic Acids Res.* **2009**, *37*, W240-W246.

32. M. D. Frankkamenetskii; S. M. Mirkin. Triplex DNA Structures. *Annu. Rev. Biochem.* **1995**, *64*, 65-95.

33. M. D. Frank-Kamenetskii; S. M. Mirkin. Triplex DNA structures. *Annu. Rev. Biochem.* **1995**, *64*, 65-95.

34. K. M. Vasquez; P. M. Glazer. Triplex-forming oligonucleotides: principles and applications. *Q. Rev. Biophys.* **2002**, *35*(1), 89-107.

35. M. Duca; P. Vekhoff; K. Oussedik; L. Halby; P. B. Arimondo. The triple helix: 50 years later, the outcome. *Nucleic Acids Res.* **2008**, *36*(16), 5123-5138.

36. M. Esguerra; L. Nilsson; A. Villa. Triple helical DNA in a duplex context and base pair opening. *Nucleic Acids Res.* **2014**, *42*(18), 11329-11338.

37. L. Jaeger; F. Michel; E. Westhof. Involvement of a Gnra Tetraloop in Long-Range Tertiary Interactions. *J. Mol. Biol.* **1994**, *236*(5), 1271-1276.

38. F. M. Jucker; A. Pardi. Gnra Tetraloops Make a U-Turn. *RNA.* **1995**, *1*(2), 219-222.

39. R. Giege; F. Juhling; J. Putz; P. Stadler; C. Sauter; C. Florentz. Structure of transfer RNAs: similarity and variability. *WIREs-RNA.* **2012**, *3*(1), 37-61.

40. M. Helm. Post-transcriptional nucleotide modification and alternative folding of RNA. *Nucleic Acids Res.* **2006**, *34*(2), 721-733.

41. K. N. Nobles; C. S. Yarian; G. Liu; R. H. Guenther; P. F. Agris. Highly conserved modified nucleosides influence Mg2+-dependent tRNA folding. *Nucleic Acids Res.* **2002**, *30*(21), 4751-4760.

42. H. D. Kim; J. D. Puglisi; S. Chu. Fluctuations of transfer RNAs between classical and hybrid states. *Biophys. J.* **2007**, *93*(10), 3575-3582.

43. D. E. Draper. RNA folding: thermodynamic and molecular descriptions of the roles of ions. *Biophys. J.* **2008**, *95*(12), 5489-5495.

44. F. H. Crick. Codon--anticodon pairing: the wobble hypothesis. *J. Mol. Biol.* **1966**, *19*(2), 548-555.

45. P. F. Agris; F. A. Vendeix; W. D. Graham. tRNA's wobble decoding of the genome: 40 years of modification. *J. Mol. Biol.* **2007**, *366*(1), 1-13.

46. J. M. Ogle; D. E. Brodersen; W. M. Clemons, Jr.; M. J. Tarry; A. P. Carter; V. Ramakrishnan. Recognition of cognate transfer RNA by the 30S ribosomal subunit. *Science.* **2001**, *292*(5518), 897-902.

47. J. M. Ogle; F. V. Murphy; M. J. Tarry; V. Ramakrishnan. Selection of tRNA by the ribosome requires a transition from an open to a closed form. *Cell.* **2002**, *111*(5), 721-732.

48. M. Valle; A. Zavialov; W. Li; S. M. Stagg; J. Sengupta; R. C. Nielsen; P. Nissen; S. C. Harvey; M. Ehrenberg; J. Frank. Incorporation of aminoacyl-tRNA into the ribosome as seen by cryo-electron microscopy. *Nat. Struct. Biol.* **2003**, *10*(11), 899-906.

49. H. Stark; M. V. Rodnina; H. J. Wieden; F. Zemlin; W. Wintermeyer; M. van Heel. Ribosome interactions of aminoacyl-tRNA and elongation factor Tu in the codon-recognition complex. *Nat. Struct. Biol.* **2002**, *9*(11), 849-854.

50. L. B. Jenner; N. Demeshkina; G. Yusupova; M. Yusupov. Structural aspects of messenger RNA reading frame maintenance by the ribosome. *Nat. Struct. Mol. Biol.* **2010**, *17*(5), 555-560.

51. N. Demeshkina; L. Jenner; E. Westhof; M. Yusupov; G. Yusupova. A new understanding of the decoding principle on the ribosome. *Nature.* **2012**, *484*(7393), 256-259.

52. D. Hirsh; L. Gold. Translation of the UGA triplet in vitro by tryptophan transfer RNA's. *J. Mol. Biol.* **1971**, *58*(2), 459-468.

53. R. H. Buckingham; C. G. Kurland. Codon specificity of UGA suppressor tRNATrp from Escherichia coli. *Proc Natl Acad Sci U S A.* **1977**, *74*(12), 5496-5498.

54. L. Cochella; R. Green. An active role for tRNA in decoding beyond codon:anticodon pairing. *Science.* **2005**, *308*(5725), 1178-1180.

55. T. Daviter; F. V. t. Murphy; V. Ramakrishnan. Molecular biology. A renewed focus on transfer RNA. *Science.* **2005**, *308*(5725), 1123-1124.

56. T. M. Schmeing; R. M. Voorhees; A. C. Kelley; V. Ramakrishnan. How mutations in tRNA distant from the anticodon affect the fidelity of decoding. *Nat. Struct. Mol. Biol.* **2011**, *18*(4), 432-436.

57. M. Helm; J. D. Alfonzo. Posttranscriptional RNA Modifications: Playing Metabolic Games in a Cell's Chemical Legoland. *Chem. Biol.* **2014**, *21*(2), 174-185.

58. T. Carell; C. Brandmayr; A. Hienzsch; M. Muller; D. Pearson; V. Reiter; I. Thoma; P. Thumbs; M. Wagner. Structure and function of noncanonical nucleobases. *Angew. Chem. Int. Ed. Engl.* **2012**, *51*(29), 7110-7131.

59. B. El Yacoubi; M. Bailly; V. de Crecy-Lagard. Biosynthesis and function of posttranscriptional modifications of transfer RNAs. *Annu. Rev. Genet.* **2012**, *46*, 69-95.

60. P. F. Agris. Bringing order to translation: the contributions of transfer RNA anticodon-domain modifications. *EMBO reports.* **2008**, *9*(7), 629-635.

61. J. E. Jackman; J. D. Alfonzo. Transfer RNA modifications: nature's combinatorial chemistry playground. *WIREs-RNA.* **2013**, *4*(1), 35-48.

62. B. T. Thole. Molecular Polarizabilities Calculated with a Modified Dipole Interaction. *Chem. Phys.* **1981**, *59*(3), 341-350.

63. T. A. Halgren; W. Damm. Polarizable force fields. *Curr. Opin. Struct. Biol.* **2001**, *11*(2), 236-242.

64. A. G. Torres; E. Batlle; L. R. de Pouplana. Role of tRNA modifications in human diseases. *Trends Mol. Med.* **2014**, *20*(6), 306-314.

65. J. A. McCloskey; S. Nishimura. Modified Nucleosides in Transfer RNA. *Acc. Chem. Res.* **1977**, *10*, 403-410.

66. P. F. Agris. Decoding the genome: a modified view. *Nucleic Acids Res.* **2004**, *32*(1), 223-238.

67. S. S. Phelps; A. Malkiewicz; P. F. Agris; S. Joseph. Modified nucleotides in tRNA(Lys) and tRNA(Val) are important for translocation. *J. Mol. Biol.* **2004**, *338*(3), 439-444.

68. F. V. t. Murphy; V. Ramakrishnan; A. Malkiewicz; P. F. Agris. The role of modifications in codon discrimination by tRNA(Lys)UUU. *Nat. Struct. Mol. Biol.* **2004**, *11*(12), 1186-1191.

69. S. J. Nasvall; P. Chen; G. R. Bjork. The wobble hypothesis revisited: Uridine-5-oxyacetic acid is critical for reading of G-ending codons. *RNA.* **2007**, *13*(12), 2151-2164.

70. S. J. Nasvall; P. Chen; 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.* **2004**, *10*(10), 1662-1673.

71. J. W. Stuart; K. M. Koshlap; R. Guenther; P. F. Agris. Naturally-occurring modification restricts the anticodon domain conformational space of tRNA(Phe). *J. Mol. Biol.* **2003**, *334*(5), 901-918.

72. J. J. Dalluge; T. Hamamoto; K. Horikoshi; R. Y. Morita; K. O. Stetter; J. A. McCloskey. Posttranscriptional modification of tRNA in psychrophilic bacteria. *J. Bacteriol.* **1997**, *179*(6), 1918-1923.

73. F. A. Vendeix; F. V. t. Murphy; W. A. Cantara; G. Leszczynska; E. M. Gustilo; B. Sproat; A. Malkiewicz; P. F. Agris. Human tRNA(Lys3)(UUU) is pre-structured by natural modifications for cognate and wobble codon binding through keto-enol tautomerism. *J. Mol. Biol.* **2012**, *416*(4), 467-485.

74. A. Rozov; N. Demeshkina; I. Khusainov; E. Westhof; M. Yusupov; G. Yusupova. Novel base-pairing interactions at the tRNA wobble position crucial for accurate reading of the genetic code. *Nature Communications.* **2016**, *7*, 1-10.

75. A. Weixlbaumer; F. V. Murphy; A. Dziergowska; A. Malkiewicz; F. A. P. Vendeix; P. F. Agris; V. Ramakrishnan. Mechanism for expanding the decoding capacity of transfer RNAs by modification of uridines. *Nat. Struct. Mol. Biol.* **2007**, *14*(6), 498-502.

76. O. Allner; L. Nilsson. Nucleotide modifications and tRNA anticodon-mRNA codon interactions on the ribosome. *RNA.* **2011**, *17*(12), 2177-2188.

77. A. C. Bajji; D. R. Davis. Synthesis and biophysical characterization of tRNA(LyS,3) anticodon stem-loop RNAs containing the mcm(5)s(2)U nucleoside. *Org. Lett.* **2000**, *2*(24), 3865-3868.

78. A. Rodriguez-Hernandez; J. L. Spears; K. W. Gaston; P. A. Limbach; H. Gamper; Y. M. Hou; R. Kaiser; P. F. Agris; J. J. Perona. Structural and Mechanistic Basis for Enhanced Translational Efficiency by 2-Thiouridine at the tRNA Anticodon Wobble Position. *J. Mol. Biol.* **2013**, *425*(20), 3888-3906.

79. P. F. Agris; H. Sierzputowskagracz; W. Smith; A. Malkiewicz; E. Sochacka; B. Nawrot. Thiolation of Uridine Carbon-2 Restricts the Motional Dynamics of the Transfer-Rna Wobble Position Nucleoside. *J. Am. Chem. Soc.* **1992**, *114*(7), 2652-2656.

80. A. T. Larsen; A. C. Fahrenbach; J. Sheng; J. L. Pian; J. W. Szostak. Thermodynamic insights into 2-thiouridine-enhanced RNA hybridization. *Nucleic Acids Res.* **2015**, *43*(16), 7675-7687.

81. P. C. Durant; A. C. Bajji; M. Sundaram; R. K. Kumar; D. R. Davis. Structural effects of hypermodified nucleosides in the Escherichia coli and human tRNA(Lys) anticodon loop: The effect of nucleosides s(2)U, mcm(5)U, mcm(5)s(2)U, mnm(5)s(2)U, t(6)A, and ms(2)t(6)A. *Biochemistry.* **2005**, *44*(22), 8078-8089.

82. F. A. Vendeix; A. Dziergowska; E. M. Gustilo; W. D. Graham; B. Sproat; A. Malkiewicz; P. F. Agris. Anticodon domain modifications contribute order to tRNA for ribosome-mediated codon binding. *Biochemistry.* **2008**, *47*(23), 6117-6129.

83. C. Yarian; M. Marszalek; E. Sochacka; A. Malkiewicz; R. Guenther; A. Miskiewicz; P. F. Agris. Modified nucleoside dependent Watson-Crick and wobble codon binding by tRNA(UUU)(Lys) species. *Biochemistry.* **2000**, *39*(44), 13390-13395.

84. P. G. Schmidt; H. Sierzputowskagracz; P. F. Agris. Internal Motions in Yeast Phenylalanine Transfer-Rna from C-13 Nmr Relaxation Rates of Modified Base Methyl-Groups - a Model-Free Approach. *Biochemistry.* **1987**, *26*(26), 8529-8534.

85. R. C. Morris; K. G. Brown; M. S. Elliott. The effect of queuosine on tRNA structure and function. *J. Biomol. Struct. Dyn.* **1999**, *16*(4), 757-774.

86. S. S. Ashraf; E. Sochacka; R. Cain; R. Guenther; A. Malkiewicz; P. F. Agris. Single atom modification (O -> S) of tRNA confers ribosome binding. *RNA.* **1999**, *5*(2), 188-194.

87. S. S. Ashraf; R. H. Guenther; G. Ansari; A. Malkiewicz; E. Sochacka; P. F. Agris. Role of modified nucleosides of yeast tRNA(Phe) in ribosomal binding. *Cell Biochemistry and Biophysics.* **2000**, *33*(3), 241-252.

88. F. Meier; B. Suter; H. Grosjean; G. Keith; E. Kubli. Queuosine Modification of the Wobble Base in Transfer Rnahis Influences Invivo Decoding Properties. *EMBO J.* **1985**, *4*(3), 823-827.

89. T. Muramatsu; S. Yokoyama; N. Horie; A. Matsuda; T. Ueda; Z. Yamaizumi; Y. Kuchino; S. Nishimura; T. Miyazawa. A novel lysine-substituted nucleoside in the first position of the anticodon of minor isoleucine tRNA from Escherichia coli. *J. Biol. Chem.* **1988**, *263*(19), 9261-9267.

90. D. Mandal; C. Kohrer; D. Su; S. P. Russell; K. Krivos; C. M. Castleberry; P. Blum; P. A. Limbach; D. Soll; U. L. RajBhandary. Agmatidine, a modified cytidine in the anticodon of archaeal tRNA(Ile), base pairs with adenosine but not with guanosine. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*(7), 2872-2877.

91. T. Muramatsu; K. Nishikawa; F. Nemoto; Y. Kuchino; S. Nishimura; T. Miyazawa; S. Yokoyama. Codon and Amino-Acid Specificities of a Transfer-Rna Are Both Converted by a Single Post-Transcriptional Modification. *Nature.* **1988**, *336*(6195), 179-181.

92. J. W. Stuart; Z. Gdaniec; R. Guenther; M. Marszalek; E. Sochacka; A. Malkiewicz; P. F. Agris. Functional anticodon architecture of human tRNALys3 includes disruption of intraloop hydrogen bonding by the naturally occurring amino acid modification, t6A. *Biochemistry.* **2000**, *39*(44), 13396-13404.

93. J. Cabello-Villegas; M. E. Winkler; E. P. Nikonowicz. Solution conformations of unmodified and A(37)N(6)-dimethylallyl modified anticodon stem-loops of Escherichia coli tRNA(Phe). *J. Mol. Biol.* **2002**, *319*(5), 1015-1034.

94. E. Lescrinier; K. Nauwelaerts; K. Zanier; K. Poesen; M. Sattler; P. Herdewijn. The naturally occurring N6-threonyl adenine in anticodon loop of Schizosaccharomyces pombe tRNAi causes formation of a unique U-turn motif. *Nucleic Acids Res.* **2006**, *34*(10), 2878-2886.

95. P. Benas; G. Bec; G. Keith; R. Marquet; C. Ehresmann; B. Ehresmann; P. Dumas. The crystal structure of HIV reverse-transcription primer tRNA(Lys,3) shows a canonical anticodon loop. *RNA.* **2000**, *6*(10), 1347-1355.

96. P. Barraud; E. Schmitt; Y. Mechulam; F. Dardel; C. Tisne. A unique conformation of the anticodon stem-loop is associated with the capacity of tRNAfMet to initiate protein synthesis. *Nucleic Acids Res.* **2008**, *36*(15), 4894-4901.

97. E. Schmitt; M. Panvert; S. Blanquet; Y. Mechulam. Crystal structure of methionyl-tRNAfMet transformylase complexed with the initiator formyl-methionyl-tRNAfMet. *EMBO J.* **1998**, *17*(23), 6819-6826.

98. R. Basavappa; P. B. Sigler. The 3 A crystal structure of yeast initiator tRNA: functional implications in initiator/elongator discrimination. *EMBO J.* **1991**, *10*(10), 3105-3111.

99. H. Shi; P. B. Moore. The crystal structure of yeast phenylalanine tRNA at 1.93 A resolution: a classic structure revisited. *RNA.* **2000**, *6*(8), 1091-1105.

100. Sundaral.M; S. T. Rao; J. Abola. Stereochemistry of nucleic acids and their constituents .23. Crystal and molecular structure of dihydrouridine hemihydrate a rare nucleoside with a saturated base occurring in dihydrouridine loop of transfer ribonucleic acids. *J. Am. Chem. Soc.* **1971**, *93*(25), 7055-7062.

101. E. Egert; H. J. Lindner; W. Hillen; M. C. Bohm. Influence of Substituents at the 5 Position on the Structure of Uridine. *J. Am. Chem. Soc.* **1980**, *102*(11), 3707-3713.

102. W. Uhl; J. Reiner; H. G. Gassen. On the Conformation of 5-Substituted Uridines as Studied by Proton Magnetic-Resonance. *Nucleic Acids Res.* **1983**, *11*(4), 1167-1180.

103. J. J. Dalluge; T. Hashizume; A. E. Sopchik; J. A. McCloskey; D. R. Davis. Conformational flexibility in RNA: the role of dihydrouridine. *Nucleic Acids Res.* **1996**, *24*(6), 1073-1079.

104. N. Dyubankova; E. Sochacka; K. Kraszewska; B. Nawrot; P. Herdewijn; E. Lescrinier. Contribution of dihydrouridine in folding of the D-arm in tRNA. *Org Biomol Chem.* **2015**, *13*(17), 4960-4966.

105. M. Charette; M. W. Gray. Pseudouridine in RNA: What, where, how, and why. *IUBMB Life.* **2000**, *49*(5), 341-351.

106. D. R. Davis. Stabilization of RNA stacking by pseudouridine. *Nucleic Acids Res.* **1995**, *23*(24), 5020-5026.

107. D. R. Davis; C. A. Veltri; L. Nielsen. An RNA model system for investigation of pseudouridine stabilization of the codon-anticodon interaction in tRNALys, tRNAHis and tRNATyr. *J. Biomol. Struct. Dyn.* **1998**, *15*(6), 1121-1132.

108. P. C. Durant; D. R. Davis. Stabilization of the anticodon stem-loop of tRNA(LyS,3) by an A(+)-C base-pair and by pseudouridine. *J. Mol. Biol.* **1999**, *285*(1), 115-131.

109. M. I. Newby; N. L. Greenbaum. Investigation of Overhauser effects between pseudouridine and water protons in RNA helices. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*(20), 12697-12702.

110. C. S. Yarian; M. M. Basti; R. J. Cain; G. Ansari; R. H. Guenther; E. Sochacka; G. Czerwinska; A. Malkiewicz; P. F. Agris. Structural and functional roles of the N1- and N3-protons of psi at tRNA's position 39. *Nucleic Acids Res.* **1999**, *27*(17), 3543-3549.

111. J. G. Arnez; T. A. Steitz. Crystal-Structure of Unmodified Trna(Gln) Complexed with Glutaminyl-Transfer-Rna Synthetase and Atp Suggests a Possible Role for Pseudo-Uridines in Stabilization of Rna Structure. *Biochemistry.* **1994**, *33*(24), 7560-7567.

112. G. A. Hudson; R. J. Bloomingdale; B. M. Znosko. Thermodynamic contribution and nearest-neighbor parameters of pseudouridine-adenosine base pairs in oligoribonucleotides. *RNA.* **2013**, *19*(11), 1474-1482.

113. J. G. Arnez; T. A. Steitz. Crystal structure of unmodified tRNA(Gln) complexed with glutaminyl-tRNA synthetase and ATP suggests a possible role for pseudo-uridines in stabilization of RNA structure. *Biochemistry.* **1994**, *33*(24), 7560-7567.

114. Y. Motorin; M. Helm. RNA nucleotide methylation. *WIREs-RNA.* **2011**, *2*(5), 611-631.

115. J. D. Nickels; J. E. Curtis; H. O'Neill; A. P. Sokolov. Role of methyl groups in dynamics and evolution of biomolecules. *Journal of Biological Physics.* **2012**, *38*(3), 497-505.

116. J. Urbonavicius; J. Armengaud; H. Grosjean. Identity elements required for enzymatic formation of N-2, N-2-dimethylguanosine from N-2-monomethylated derivative and its possible role in avoiding alternative conformations in archaeal tRNA. *J. Mol. Biol.* **2006**, *357*(2), 387-399.

117. H. Hori; S. Kubota; K. Watanabe; J. M. Kim; T. Ogasawara; T. Sawasaki; Y. Endo. Aquifex aeolicus tRNA (Gm18) methyltransferase has unique substrate specificity. TRNA recognition mechanism of the enzyme. *J. Biol. Chem.* **2003**, *278*(27), 25081-25090.

118. T. Awai; S. Kimura; C. Tomikawa; A. Ochi; Ihsanawati; Y. Bessho; S. Yokoyama; S. Ohno; K. Nishikawa; T. Yokogawa; T. Suzuki; H. Hori. Aquifex aeolicus tRNA (N2,N2-guanine)-dimethyltransferase (Trm1) catalyzes transfer of methyl groups not only to guanine 26 but also to guanine 27 in tRNA. *J. Biol. Chem.* **2009**, *284*(31), 20467-20478.

119. J. Finer-Moore; N. Czudnochowski; J. D. O'Connell, 3rd; A. L. Wang; R. M. Stroud. Crystal Structure of the Human tRNA m(1)A58 Methyltransferase-tRNA(3)(Lys) Complex: Refolding of Substrate tRNA Allows Access to the Methylation Target. *J. Mol. Biol.* **2015**, *427*(24), 3862-3876.

120. M. Helm; H. Brule; F. Degoul; C. Cepanec; J. P. Leroux; R. Giege; C. Florentz. The presence of modified nucleotides is required for cloverleaf folding of a human mitochondrial tRNA. *Nucleic Acids Res.* **1998**, *26*(7), 1636-1643.

121. M. Helm; R. Giege; C. Florentz. A Watson-Crick base-pair-disrupting methyl group (m(1)A9) is sufficient for cloverleaf folding of human mitochondrial tRNA(Lys). *Biochemistry.* **1999**, *38*(40), 13338-13346.

122. Y. Chen; H. Sierzputowskagracz; R. Guenther; K. Everett; P. F. Agris. 5-Methylcytidine Is Required for Cooperative Binding of Mg2+ and a Conformational Transition at the Anticodon Stem-Loop of Yeast Phenylalanine Transfer-Rna. *Biochemistry.* **1993**, *32*(38), 10249-10253.

123. A. Alexandrov; I. Chernyakov; W. F. Gu; S. L. Hiley; T. R. Hughes; E. J. Grayhack; E. M. Phizicky. Rapid tRNA decay can result from lack of nonessential modifications. *Mol. Cell.* **2006**, *21*(1), 87-96.

124. J. Anderson; L. Phan; A. G. Hinnebusch. The Gcd10p/Gcd14p complex is the essential two-subunit tRNA(1-methyladenosine) methyltransferase of Saccharomyces cerevisiae. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*(10), 5173-5178.

125. I. Chernyakov; J. M. Whipple; L. Kotelawala; E. J. Grayhack; E. M. Phizicky. Degradation of several hypomodified mature tRNA species in Saccharomyces cerevisiae is mediated by Met22 and the 5 '-3 ' exonucleases Rat1 and Xrn1. *Genes Dev.* **2008**, *22*(10), 1369-1380.

126. Y. Motorin; M. Helm. tRNA Stabilization by Modified Nucleotides. *Biochemistry.* **2010**, *49*(24), 4934-4944.

127. S. U. Astrom; A. S. Bystrom. Rit1, a tRNA backbone-modifying enzyme that mediates initiator and elongator tRNA discrimination. *Cell.* **1994**, *79*(3), 535-546.

128. V. H. Cowling. Regulation of mRNA cap methylation. *Biochem. J.* **2010**, *425*, 295-302.

129. S. Farajollahi; S. Maas. Molecular diversity through RNA editing: a balancing act. *Trends Genet.* **2010**, *26*(5), 221-230.

130. S. C. Blanchard; J. D. Puglisi. Solution structure of the A loop of 23S ribosomal RNA. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*(7), 3720-3725.

131. C. Hobartner; M. O. Ebert; B. Jaun; R. Micura. RNA two-state conformation equilibria and the effect of nucleobase methylation. *Angew. Chem. Int. Edit.* **2002**, *41*(4), 605-+.

132. A. Baudin-Baillieu; C. Fabret; X. H. Liang; D. Piekna-Przybylska; M. J. Fournier; J. P. Rousset. Nucleotide modifications in three functionally important regions of the Saccharomyces cerevisiae ribosome affect translation accuracy. *Nucleic Acids Res.* **2009**, *37*(22), 7665-7677.

133. K. S. Long; J. Poehlsgaard; C. Kehrenberg; S. Schwarz; B. Vester. The Cfr rRNA methyltransferase confers resistance to phenicols, lincosamides, oxazolidinones, pleuromutilins, and streptograrnin A antibiotics. *Antimicrob. Agents Chemother.* **2006**, *50*(7), 2500-2505.

134. D. N. Wilson. On the specificity of antibiotics targeting the large ribosomal subunit. *Antimicrobial Therapeutics Reviews: Antibiotics That Target the Ribosome.* **2011**, *1241*, 1-16.

135. M. Paramasivam; S. Cogoi; V. V. Filichev; N. Bomholt; E. B. Pedersen; L. E. Xodo. Purine twisted-intercalating nucleic acids: a new class of anti-gene molecules resistant to potassium-induced aggregation. *Nucleic Acids Res.* **2008**, *36*(10), 3494-3507.

136. M. Murtola; M. Wenska; R. Stromberg. PNAzymes That Are Artificial RNA Restriction Enzymes. *J. Am. Chem. Soc.* **2010**, *132*(26), 8984-8990.

137. M. Murtola; R. Stromberg. PNA based artificial nucleases displaying catalysis with turnover in the cleavage of a leukemia related RNA model. *Organic & Biomolecular Chemistry.* **2008**, *6*(20), 3837-3842.

138. M. Murtola; D. Ossipov; J. Sandbrink; R. Stromberg. RNA cleavage by 2,9-diamino-1,10-phenanthroline PNA conjugates. *Nucleosides Nucleotides & Nucleic Acids.* **2007**, *26*(10-12), 1479-1483.

139. M. Murtola; R. Stromberg. 2 '-O-Methyloligoribonucleotide based artificial nucleases (2 '-O-MeOBANs) cleaving a model of the leukemia related M-BCR/ABL m-RNA. *Arkivoc.* **2009**, 84-94.

140. M. D. Sorensen; L. Kvaerno; T. Bryld; A. E. Hakansson; B. Verbeure; G. Gaubert; P. Herdewijn; J. Wengel. alpha-L-ribo-configured locked nucleic acid (alpha-L-LNA): synthesis and properties. *J. Am. Chem. Soc.* **2002**, *124*(10), 2164-2176.

141. M. Petersen; K. Bondensgaard; J. Wengel; J. P. Jacobsen. Locked nucleic acid (LNA) recognition of RNA: NMR solution structures of LNA : RNA hybrids. *J. Am. Chem. Soc.* **2002**, *124*(21), 5974-5982.

142. D. A. Braasch; D. R. Corey. Locked nucleic acid (LNA): fine-tuning the recognition of DNA and RNA. *Chem. Biol.* **2001**, *8*(1), 1-7.

143. V. V. Filichev; U. B. Christensen; E. B. Pedersen; B. R. Babu; J. Wengel. Locked nucleic acids and intercalating nucleic acids in the design of easily denaturing nucleic acids: Thermal stability studies. *ChemBioChem.* **2004**, *5*(12), 1673-1679.

144. Y. You; B. G. Moreira; M. A. Behlke; R. Owczarzy. Design of LNA probes that improve mismatch discrimination. *Nucleic Acids Res.* **2006**, *34*(8).

145. M. Petersen; C. B. Nielsen; K. E. Nielsen; G. A. Jensen; K. Bondensgaard; S. K. Singh; V. K. Rajwanshi; A. A. Koshkin; B. M. Dahl; J. Wengel; J. P. Jacobsen. The conformations of locked nucleic acids (LNA). *Journal of molecular recognition : JMR.* **2000**, *13*(1), 44-53.

146. J. Wengel; M. Petersen; M. Frieden; T. Koch. Chemistry of locked nucleic acids (LNA): Design, synthesis, and bio-physical properties. *Lett. Pept. Sci.* **2003**, *10*(3-4), 237-253.

147. H. Torigoe; Y. Hari; M. Sekiguchi; S. Obika; T. Imanishi. 2'-O,4'-C-methylene bridged nucleic acid modification promotes pyrimidine motif triplex DNA formation at physiological pH: thermodynamic and kinetic studies. *J. Biol. Chem.* **2001**, *276*(4), 2354-2360.

148. K. E. Nielsen; J. Rasmussen; R. Kumar; J. Wengel; J. P. Jacobsen; M. Petersen. NMR studies of fully modified locked nucleic acid (LNA) hybrids: Solution structure of an LNA : RNA hybrid and characterization of an LNA : DNA hybrid. *Bioconjugate Chem.* **2004**, *15*(3), 449-457.

149. V. Marin; H. F. Hansen; T. Koch; B. A. Armitage. Effect of LNA modifications on small molecule binding to nucleic acids. *J. Biomol. Struct. Dyn.* **2004**, *21*(6), 841-850.

150. S. F. Singleton; P. B. Dervan. Influence of Ph on the Equilibrium Association Constants for Oligodeoxyribonucleotide-Directed Triple Helix Formation at Single DNA Sites. *Biochemistry.* **1992**, *31*(45), 10995-11003.

151. J. S. Lee; M. L. Woodsworth; L. J. P. Latimer; A. R. Morgan. Poly(Pyrimidine).Poly(Purine) Synthetic Dnas Containing 5-Methylcytosine Form Stable Triplexes at Neutral Ph. *Nucleic Acids Res.* **1984**, *12*(16), 6603-6614.

152. U. Christensen; N. Jacobsen; V. K. Rajwanshi; J. Wengel; T. Koch. Stopped-flow kinetics of lacked nucleic acid (LNA)-oligonucleotide duplex formation: studies of LNA-DNA and DNA-DNA interactions. *Biochem. J.* **2001**, *354*, 481-484.

153. S. Karkare; D. Bhatnagar. Promising nucleic acid analogs and mimics: characteristic features and applications of PNA, LNA, and morpholino. *Appl. Microbiol. Biotechnol.* **2006**, *71*(5), 575-586.

154. L. A. Ugozzoli; D. Latorra; R. Pucket; K. Arar; K. Hamby. Real-time genotyping with oligonucleotide probes containing locked nucleic acids. *Anal. Biochem.* **2004**, *324*(1), 143-152.

155. T. W. van Ravesteyn; M. Dekker; A. Fish; T. K. Sixma; A. Wolters; R. J. Dekker; H. P. J. te Riele. LNA modification of single-stranded DNA oligonucleotides allows subtle gene modification in mismatch-repair-proficient cells. *Proc. Natl. Acad. Sci. U.S.A.* **2016**, *113*(15), 4122-4127.

156. G. Suresh; U. D. Priyakumar. Structures, Dynamics, and Stabilities of Fully Modified Locked Nucleic Acid (beta-D-LNA and alpha-L-LNA) Duplexes in Comparison to Pure DNA and RNA Duplexes. *J. Phys. Chem. B.* **2013**, *117*(18), 5556-5564.

157. G. Suresh; U. D. Priyakumar. Atomistic investigation of the effect of incremental modification of deoxyribose sugars by locked nucleic acid (beta-D-LNA and alpha-L-LNA) moieties on the structures and thermodynamics of DNA-RNA hybrid duplexes. *J. Phys. Chem. B.* **2014**, *118*(22), 5853-5863.

158. I. Yildirim; E. Kierzek; R. Kierzek; G. C. Schatz. Interplay of LNA and 2 '-0-Methyl RNA in the Structure and Thermodynamics of RNA Hybrid Systems: A Molecular Dynamics Study Using the Revised AMBER Force Field and Comparison with Experimental Results. *J. Phys. Chem. B.* **2014**, *118*(49), 14177-14187.

159. V. Pande; L. Nilsson. Insights into structure, dynamics and hydration of locked nucleic acid (LNA) strand-based duplexes from molecular dynamics simulations. *Nucleic Acids Res.* **2008**, *36*(5), 1508-1516.

160. T. A. Halgren. Representation of Vanderwaals (Vdw) Interactions in Molecular Mechanics Force-Fields - Potential Form, Combination Rules, and Vdw Parameters. *J. Am. Chem. Soc.* **1992**, *114*(20), 7827-7843.

161. A. D. Mackerell, Jr. Empirical force fields for biological macromolecules: overview and issues. *J. Comput. Chem.* **2004**, *25*(13), 1584-1604.

162. W. L. Jorgensen; J. Chandrasekhar; J. D. Madura; R. W. Impey; M. L. Klein. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*(2), 926-935.

163. H. J. C. Berendsen; J. R. Grigera; T. P. Straatsma. The Missing Term in Effective Pair Potentials. *J. Phys. Chem.* **1987**, *91*(24), 6269-6271.

164. P. Mark; L. Nilsson. Structure and dynamics of the TIP3P, SPC, and SPC/E water models at 298 K. *J. Phys. Chem. A.* **2001**, *105*(43), 9954-9960.

165. S. W. Rick; S. J. Stuart. Potentials and algorithms for incorporating polarizability in computer simulations. *Reviews in Computational Chemistry, Vol 18.* **2002**, *18*, 89-146.

166. V. M. Anisimov; G. Lamoureux; I. V. Vorobyov; N. Huang; B. Roux; A. D. MacKerell. Determination of Electrostatic Parameters for a Polarizable Force Field Based on the Classical Drude Oscillator. *J. Chem. Theory Comput.* **2005**, *1*(1), 153-168.

167. U. C. Singh; P. A. Kollman. An Approach to Computing Electrostatic Charges for Molecules. *J. Comput. Chem.* **1984**, *5*(2), 129-145.

168. A. D. Mackerell; M. Karplus. Importance of Attractive Vanderwaals Contribution in Empirical Energy Function Models for the Heat of Vaporization of Polar Liquids. *J. Phys. Chem.* **1991**, *95*(26), 10559-10560.

169. K. Vanommeslaeghe; E. Hatcher; C. Acharya; S. Kundu; S. Zhong; J. Shim; E. Darian; O. Guvench; P. Lopes; I. Vorobyov; A. D. Mackerell, Jr. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comput. Chem.* **2010**, *31*(4), 671-690.

170. D. X. Yin; A. D. Mackerell. Combined ab initio empirical approach for optimization of Lennard-Jones parameters. *J. Comput. Chem.* **1998**, *19*(3), 334-348.

171. I. J. Chen; D. Yin; A. D. MacKerell, Jr. Combined ab initio/empirical approach for optimization of Lennard-Jones parameters for polar-neutral compounds. *J. Comput. Chem.* **2002**, *23*(2), 199-213.

172. A. P. Scott; L. Radom. Harmonic vibrational frequencies: An evaluation of Hartree-Fock, Moller-Plesset, quadratic configuration interaction, density functional theory, and semiempirical scale factors. *J. Phys. Chem.* **1996**, *100*(41), 16502-16513.

173. A. D. MacKerell; D. Bashford; M. Bellott; R. L. Dunbrack; J. D. Evanseck; M. J. Field; S. Fischer; J. Gao; H. Guo; S. Ha; D. Joseph-McCarthy; L. Kuchnir; K. Kuczera; F. T. Lau; C. Mattos; S. Michnick; T. Ngo; D. T. Nguyen; B. Prodhom; W. E. Reiher; B. Roux; M. Schlenkrich; J. C. Smith; R. Stote; J. Straub; M. Watanabe; J. Wiorkiewicz-Kuczera; D. Yin; M. Karplus. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B.* **1998**, *102*(18), 3586-3616.

174. W. D. Cornell; P. Cieplak; C. I. Bayly; I. R. Gould; K. M. Merz; D. M. Ferguson; D. C. Spellmeyer; T. Fox; J. W. Caldwell; P. A. Kollman. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules (vol 117, pg 5179, 1995). *J. Am. Chem. Soc.* **1996**, *118*(9), 2309-2309.

175. V. Hornak; R. Abel; A. Okur; B. Strockbine; A. Roitberg; C. Simmerling. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins.* **2006**, *65*(3), 712-725.

176. A. D. Mackerell, Jr.; M. Feig; C. L. Brooks, 3rd. Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.* **2004**, *25*(11), 1400-1415.

177. W. L. Jorgensen; J. Tiradorives. The Opls Potential Functions for Proteins - Energy Minimizations for Crystals of Cyclic-Peptides and Crambin. *J. Am. Chem. Soc.* **1988**, *110*(6), 1657-1666.

178. C. Oostenbrink; A. Villa; A. E. Mark; W. F. van Gunsteren. A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. *J. Comput. Chem.* **2004**, *25*(13), 1656-1676.

179. N. Foloppe; A. D. MacKerell. All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data. *J. Comput. Chem.* **2000**, *21*(2), 86-104.

180. A. Perez; I. Marchan; D. Svozil; J. Sponer; T. E. Cheatham, 3rd; C. A. Laughton; M. Orozco. Refinement of the AMBER force field for nucleic acids: improving the description of alpha/gamma conformers. *Biophys. J.* **2007**, *92*(11), 3817-3829.

181. K. Hart; N. Foloppe; C. M. Baker; E. J. Denning; L. Nilsson; A. D. MacKerell. Optimization of the CHARMM Additive Force Field for DNA: Improved Treatment of the BI/BII Conformational Equilibrium. *J. Chem. Theory Comput.* **2012**, *8*(1), 348-362.

182. O. Guvench; S. N. Greene; G. Kamath; J. W. Brady; R. M. Venable; R. W. Pastor; A. D. Mackerell, Jr. Additive empirical force field for hexopyranose monosaccharides. *J. Comput. Chem.* **2008**, *29*(15), 2543-2564.

183. K. N. Kirschner; A. B. Yongye; S. M. Tschampel; J. Gonzalez-Outeirino; C. R. Daniels; B. L. Foley; R. J. Woods. GLYCAM06: A generalizable Biomolecular force field. Carbohydrates. *J. Comput. Chem.* **2008**, *29*(4), 622-655.

184. W. Damm; A. Frontera; J. TiradoRives; W. L. Jorgensen. OPLS all-atom force field for carbohydrates. *J. Comput. Chem.* **1997**, *18*(16), 1955-1970.

185. J. B. Klauda; R. M. Venable; J. A. Freites; J. W. O'Connor; D. J. Tobias; C. Mondragon-Ramirez; I. Vorobyov; A. D. MacKerell; R. W. Pastor. Update of the CHARMM All-Atom Additive Force Field for Lipids: Validation on Six Lipid Types. *J. Phys. Chem. B.* **2010**, *114*(23), 7830-7843.

186. C. J. Dickson; B. D. Madej; A. A. Skjevik; R. M. Betz; K. Teigen; I. R. Gould; R. C. Walker. Lipid14: The Amber Lipid Force Field. *J. Chem. Theory Comput.* **2014**, *10*(2), 865-879.

187. L. D. Schuler; X. Daura; W. F. Van Gunsteren. An improved GROMOS96 force field for aliphatic hydrocarbons in the condensed phase. *J. Comput. Chem.* **2001**, *22*(11), 1205-1218.

188. X. Daura; A. E. Mark; W. F. van Gunsteren. Parametrization of aliphatic CHn united atoms of GROMOS96 force field. *J. Comput. Chem.* **1998**, *19*(5), 535-547.

189. J. Wang; R. M. Wolf; J. W. Caldwell; P. A. Kollman; D. A. Case. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25*(9), 1157-1174.

190. W. L. Jorgensen; D. S. Maxwell; J. TiradoRives. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* **1996**, *118*(45), 11225-11236.

191. T. A. Halgren. Merck molecular force field .1. Basis, form, scope, parameterization, and performance of MMFF94. *J. Comput. Chem.* **1996**, *17*(5-6), 490-519.

192. G. N. Ramachandran; C. Ramakrishnan; V. Sasisekharan. Stereochemistry of Polypeptide Chain Configurations. *J. Mol. Biol.* **1963**, *7*(1), 95-&.

193. A. D. MacKerell, Jr.; M. Feig; C. L. Brooks, 3rd. Improved treatment of the protein backbone in empirical force fields. *J. Am. Chem. Soc.* **2004**, *126*(3), 698-699.

194. P. K. Weiner; P. A. Kollman. Amber - Assisted Model-Building with Energy Refinement - a General Program for Modeling Molecules and Their Interactions. *J. Comput. Chem.* **1981**, *2*(3), 287-303.

195. L. Nilsson; M. Karplus. Empirical Energy Functions for Energy Minimization and Dynamics of Nucleic-Acids. *J. Comput. Chem.* **1986**, *7*(5), 591-616.

196. N. Foloppe; L. Nilsson. Toward a full characterization of nucleic acid components in aqueous solution: Simulations of nucleosides. *J. Phys. Chem. B.* **2005**, *109*(18), 9119-9131.

197. E. J. Denning; U. D. Priyakumar; L. Nilsson; A. D. Mackerell. Impact of 2'-Hydroxyl Sampling on the Conformational Properties of RNA: Update of the CHARMM All-Atom Additive Force Field for RNA. *J. Comput. Chem.* **2011**, *32*(9), 1929-1943.

198. O. Guvench; E. R. Hatcher; R. M. Venable; R. W. Pastor; A. D. Mackerell. CHARMM Additive All-Atom Force Field for Glycosidic Linkages between Hexopyranoses. *J. Chem. Theory Comput.* **2009**, *5*(9), 2353-2370.

199. J. M. Wang; P. Cieplak; P. A. Kollman. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.* **2000**, *21*(12), 1049-1074.

200. W. Yu; X. He; K. Vanommeslaeghe; A. D. MacKerell, Jr. Extension of the CHARMM General Force Field to sulfonyl-containing compounds and its utility in biomolecular simulations. *J. Comput. Chem.* **2012**, *33*(31), 2451-2468.

201. K. Vanommeslaeghe; A. D. MacKerell, Jr. Automation of the CHARMM General Force Field (CGenFF) I: bond perception and atom typing. *J. Chem. Inf. Model.* **2012**, *52*(12), 3144-3154.

202. K. Vanommeslaeghe; E. P. Raman; A. D. MacKerell, Jr. Automation of the CHARMM General Force Field (CGenFF) II: assignment of bonded parameters and partial atomic charges. *J. Chem. Inf. Model.* **2012**, *52*(12), 3155-3168.

203. M. Karplus; R. Lavery. Significance of Molecular Dynamics Simulations for Life Sciences. *Isr. J. Chem.* **2014**, *54*(8-9), 1042-1051.

204. M. Karplus. Development of multiscale models for complex chemical systems: from H+H(2) to biomolecules (Nobel Lecture). *Angew. Chem. Int. Ed. Engl.* **2014**, *53*(38), 9992-10005.

205. L. Verlet. Computer Experiments on Classical Fluids .I. Thermodynamical Properties of Lennard-Jones Molecules. *Phys Rev.* **1967**, *159*(1), 98-103.

206. J. P. Ryckaert; G. Ciccotti; H. J. C. Berendsen. Numerical-Integration of Cartesian Equations of Motion of a System with Constraints - Molecular-Dynamics of N-Alkanes. *J. Comput. Phys.* **1977**, *23*(3), 327-341.

207. T. Darden; D. York; L. Pedersen. Particle Mesh Ewald - an N.Log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*(12), 10089-10092.

208. H. J. C. Berendsen; J. P. M. Postma; W. F. Vangunsteren; A. Dinola; J. R. Haak. Molecular-Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984**, *81*(8), 3684-3690.

209. S. Nose. A Molecular-Dynamics Method for Simulations in the Canonical Ensemble. *Mol. Phys.* **1984**, *52*(2), 255-268.

210. W. G. Hoover. Canonical Dynamics - Equilibrium Phase-Space Distributions. *Phys Rev A.* **1985**, *31*(3), 1695-1697.

211. H. C. Andersen. Molecular-Dynamics Simulations at Constant Pressure and-or Temperature. *J. Chem. Phys.* **1980**, *72*(4), 2384-2393.

212. R. W. Zwanzig. High‐Temperature Equation of State by a Perturbation Method. I. Nonpolar Gases *J. Chem. Phys.* **1954**, *22*(8), 1420-1426.

213. S. Bruckner; S. Boresch. Efficiency of alchemical free energy simulations. I. A practical comparison of the exponential formula, thermodynamic integration, and Bennett's acceptance ratio method. *J. Comput. Chem.* **2011**, *32*(7), 1303-1319.

214. J. G. Kirkwood. Statistical Mechanics of Fluid Mixtures *J. Chem. Phys.* **1935**, *3*, 300-313.

215. S. Bruckner; S. Boresch. Efficiency of alchemical free energy simulations. II. Improvements for thermodynamic integration. *J. Comput. Chem.* **2011**, *32*(7), 1320-1333.

216. C. H. Bennett. Efficient Estimation of Free-Energy Differences from Monte-Carlo Data. *J. Comput. Phys.* **1976**, *22*(2), 245-268.

217. M. Zacharias; T. P. Straatsma; J. A. Mccammon. Separation-Shifted Scaling, a New Scaling Method for Lennard-Jones Interactions in Thermodynamic Integration. *J. Chem. Phys.* **1994**, *100*(12), 9025-9031.

218. S. Boresch; F. Tettinger; M. Leitgeb; M. Karplus. Absolute binding free energies: A quantitative approach for their calculation. *J. Phys. Chem. B.* **2003**, *107*(35), 9535-9551.

219. S. Kumar; D. Bouzida; R. H. Swendsen; P. A. Kollman; J. M. Rosenberg. The Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules .1. The Method. *J. Comput. Chem.* **1992**, *13*(8), 1011-1021.

220. J. Kastner. Umbrella sampling. *WIRES Comput Mol Sci.* **2011**, *1*(6), 932-942.

221. S. Boresch; M. Karplus. The role of bonded terms in free energy simulations: 1. Theoretical analysis. *J. Phys. Chem. A.* **1999**, *103*(1), 103-118.

222. B. N. Macchion; R. Stromberg; L. Nilsson. Analysis of the stability and flexibility of RNA complexes containing bulge loops of different sizes. *J. Biomol. Struct. Dyn.* **2008**, *26*(2), 163-173.

223. S. Sen; L. Nilsson. Molecular dynamics of duplex systems involving PNA: Structural and dynamical consequences of the nucleic acid backbone. *J. Am. Chem. Soc.* **1998**, *120*(4), 619-631.