

Chapter 11

RNA Folding: Structure Prediction, Folding Kinetics and Ion Electrostatics

Zhijie Tan, Wenbing Zhang, Yazhou Shi and Fenghua Wang

Abstract Beyond the “traditional” functions such as gene storage, transport and protein synthesis, recent discoveries reveal that RNAs have important “new” biological functions including the RNA silence and gene regulation of riboswitch. Such functions of noncoding RNAs are strongly coupled to the RNA structures and proper structure change, which naturally leads to the RNA folding problem including structure prediction and folding kinetics. Due to the polyanionic nature of RNAs, RNA folding structure, stability and kinetics are strongly coupled to the ion condition of solution. The main focus of this chapter is to review the recent progress in the three major aspects in RNA folding problem: structure prediction, folding kinetics and ion electrostatics. This chapter will introduce both the recent experimental and theoretical progress, while emphasize the theoretical modelling on the three aspects in RNA folding.

Keywords RNA folding · Structure · Kinetics · Metal ions · Thermodynamics

11.1 Introduction

Recent discoveries reveal the new biological functions of non-coding RNAs such as RNA silence and riboswitch. The functions of non-coding RNAs are intrinsic to RNA structures and stability, and can provide potential RNA-based therapeutic strategies. This demands the quantitative understanding and prediction on RNA

Z. Tan (✉) · W. Zhang (✉) · Y. Shi · F. Wang
Department of Physics and Key Laboratory of Artificial Micro- and Nano-Structures
of Ministry of Education, School of Physics and Technology, Wuhan University,
Wuhan, China
e-mail: zjtan@whu.edu.cn

W. Zhang
e-mail: wbzhang@whu.edu.cn

structure and its stability, i.e., RNA folding problem. RNA structural folding is driven by the intra-molecular forces, such as base pairing/stacking interactions, ion-mediated interactions and the conformational entropies [1, 2]. This chapter is focusing on the three aspects in RNA folding: structure prediction, folding kinetics and ion-mediated electrostatic interactions, respectively.

First, RNA structure prediction is one of central issues in RNA folding problem, since RNA structures involve not only their biological functions such as gene regulation, but also the interactions with other molecules which can provide the potential therapeutic strategies. Generally, RNA folding is hierarchical, i.e., the secondary structure can be formed firstly driven by strong base pairing/stacking interactions, and afterwards, the tertiary structure can be folded by the aggregation of secondary segments and the formation of tertiary contacts [3]. Since the base pairing/stacking interactions are very strong, the secondary structure of RNAs can be relatively stable. Accordingly, the RNA structure prediction can be classified into the secondary structure prediction and tertiary structure prediction. From 1980s, many efforts have been made on the secondary structure prediction based on the experimentally derived parameters and great progress has been made in accurate predictions on RNA secondary structure [4–8]. However, RNAs are often biological functional in their 3-dimensional structures. The lack of experimentally derived structures and the high cost for experimentally determining structures have enabled the computational modelling for predicting RNA structures [9–38]. The tertiary (including 3-dimensional) structure predictions have attracted much attention and important progress has been achieved in recent years, which will be focused on in the first part of this chapter.

Second, in addition to static structures, RNA folding kinetics is also directly tied to RNA biological functions. Experiments suggest that alternative conformations of the same RNA sequence perform different functions [39–41]. The capability of RNA molecules to form multiple (metastable) conformations for different functions is probably used by nature to regulate versatile functions of RNA. Furthermore, it was found that the folding of the functional structures is controlled by folding kinetics rather than by equilibrium thermodynamics. The mechanisms of ribozyme [42, 43], anti-HIV RNA aptamers [44–46], gene expression regulators such as miRNA, siRNA and riboswitches [47–53] and other RNAs are often kinetically controlled. For instance, self-induced riboswitches regulate RNA functions by limiting biologically functional properties of RNA structures to certain time windows. The *hok/sok* system of plasmid R1 [54, 55] regulates the plasmid maintenance through mRNA conformation rearrangements into different functional forms. For riboswitches, it has been proposed that the transient intermediate structure of RNA can regulate transcription and translation by creating a time window that is necessary for regulatory reactions to occur [56]. After the early work of Porschke [57] and Crothers [58, 59], extensive kinetic experiments, such as temperature-jump, single molecule and time resolved NMR spectroscopy experiments, have been employed to study the RNA [60–64] and DNA [65–67] folding kinetics. The recent progresses in understanding RNA/DNA folding kinetics will become the second part of the chapter.

Third, due to the polyanionic nature of RNAs, RNA folding causes massive build-up of the negative charges [68–73] and a strong intra-chain Coulombic repulsion. However, the folding would attract metal ions in solution and cause significant ion binding to RNA, to effectively reduce the electrostatic energy barrier and stabilize a folded RNA structure. Therefore, RNA folding including folded structure, folding kinetics, and stability are strongly coupled to ion electrostatic interactions [1, 68–73]. The third part of the chapter will focus on the recent progress in qualitative/quantitative understanding on ion roles in RNA folding.

In the following, we will review the recent progress in RNA folding, in the three aspects including structure prediction, folding kinetics, and ion electrostatics.

11.2 RNA Structure Prediction

As described above, the RNA structure prediction can be divided into two levels: secondary structure prediction and tertiary structure prediction. On predicting RNA secondary structure, many computational models have been proposed, based on the experimental thermodynamic parameters [4–8], such as Mfold through free energy minimization method [4], Vienna RNA package with dynamic programming algorithm [5], Sfold through sampling structures with Boltzmann statistics [6], etc.; see Ref. [8] for a recent review on RNA secondary structure prediction. In the following, we will focus on the RNA tertiary structure prediction methods which can be classified into three types: knowledge-based structure modelling, physics-based structure modelling, and knowledge/physics-hybridized structure modelling; see Table 11.1 for a summary on the algorithms for RNA tertiary structure prediction [9–38]. Other reviews are also available [74–79].

11.2.1 Knowledge-Based Structure Prediction

With the rapid increase in the RNA structure data deposited in protein data bank (PDB) and nucleic acid data bank (NDB), the knowledge-based modelling is becoming an important method for predicting RNA structure based on available sequences. The knowledge-based modelling relies on the database of experimentally solved structures and empirically observed structural similarities between the same (similar) sequences.

11.2.1.1 Graphics-Based Method

One kind of knowledge-based modelling is Graphics-based method, which usually involves interactive (user-guided) manipulation of RNA structures based on the assembly of fragments derived from various experimental structures (motifs),

Table 11.1 The algorithms for RNA three-dimensional/tertiary structure prediction

Algorithms and references	Input	Simulation method	Classification
MANIP [9]	Alignment of sequences, 2D structure	Graphical user interface	Graphics-based
S2S/Assemble [11, 12]	Database of known fragment, 2D structure	Graphical user interface	Graphics-based
RNA2D3D [13]	2D structure	Graphical user interface	Graphics-based
RNABuilder [15]	2D structure, tertiary contacts	Scripting interface, MD	Homology-based
ModeRNA [14]	Sequence, 3D structure of template	Scripting interface	Homology-based
YUP [17]	2D structure, tertiary contacts	MD	Physics-based
NASt [18, 19]	2D structure, tertiary contacts	MD	Physics-based
Vfold [21–23]	Sequence	Statistical mechanical calculation, MD	Physics-based
iFold/DMD [24, 25]	Sequence	Replica exchange DMD	Physics-based
FARNA/FARFAR [31, 32]	Sequence, 2D structure constrains	Fragment assembly, MC	Hybrid method
MC-Fold/MC-Sym [34]	Sequence, 2D structure	Fragment assembly, Las Vegas algorithm	Hybrid method

2D secondary; 3D three-dimensional; MD molecular dynamics simulation; MC Monte Carlo simulation; DMD discrete molecular dynamics simulation

including the algorithms of MANIP [9], S2S/Assemble [11, 12], and RNA2D3D [13].

MANIP. Massire and Westhof have developed a program MANIP [9], which allows the rapid assembly of isolated motifs (each with a specified sequence) into a complex three-dimensional architecture by users. As an interactive tool, MANIP has a toolbox, where the user can find a variety of tools that help to design a three-dimensional structure model. MANIP constitutes a quick and easy way to model small- to large-size structured RNAs, and the use of multiple connections and pairing tables opens the further development perspectives and allows, for instance, the precise modelling of RNA-protein interactions.

S2S/Assemble. Jossinet et al. [11] developed a program S2S (sequence to structure), in which a user can conveniently display, manipulate and interconnect heterogeneous RNA data. Assemble [12], an algorithm complementary to S2S, is an intuitive graphical interface to analyze, manipulate and build complex 3D RNA architectures. S2S/Assemble is a system that combines various tools and web services into a powerful package to edit sequences and structures of RNA. It contains explicit annotation of base pairing and stacking interactions, multiple sequence alignments, a motif library and an automatic procedure to generate 3D models from the annotation. But all interactions have to be annotation manually, and thus it is difficult to perform a high-throughput analysis.

RNA2D3D. With the use of the primary sequence and secondary structure information of an RNA, the program RNA2D3D [13] automatically and rapidly produces a 3-dimensional conformation (the initial) consistent with the available information. At the next step, the overlaps in the initial 3D structure model are removed and conformational changes are made aiming to the targeted features. Subsequently, the refinement needs to be performed by the user through its interactive graphical editing and the special tools such as the compacting, stem-stacking and segment-positioning energy-refinement. The most important advantage of RNA2D3D is that it is applicable to structures of arbitrary branches and pseudoknots. The algorithm has been verified in the modelling of ribozymes, viral kissing loops, and viral internal ribosome entry sites.

Obviously, all these methods are not an automatic algorithm. The graphics-based modelling requires users to set up and refine the RNA structures according to the specific principles, thus requires users' expert knowledge [9–13].

11.2.1.2 Homology-Based Modelling

Another kind of knowledge-based modelling is the homology-based modelling, i.e., comparative modelling, based on the empirical observation that evolutionarily related macromolecules usually retain similar 3D structure despite the divergence on the sequence level [81]. Several algorithms have been developed based on the homology-based modelling, such as ModeRNA [14] and RNABuilder [15].

ModeRNA. As a minimal input, ModeRNA [14] requires the 3D coordinates of template structures and a pairwise sequence alignment between the sequences of

the template and the RNA to be modeled. The ModeRNA provides a flexible scripting framework that can build RNA structures with various strategies, including the fast automated modelling based on template structure and target–template alignment without additional data. The ModeRNA was tested by 99 tRNAs with known structures (experimentally solved and each of them as a target to be modelled on each of the other 98 structures as templates) with RMSD values around 5.0 Å.

RNABuilder. Recently, Flores et al. have developed RNABuilder [15] now known as MMB (a contraction of MacroMolecule Builder), for comparative modelling of RNA structures. It generates RNA structures by treating the kinematics and forces at separate. The coarse-graining force field for an alignment used in this approach consists of forces and torques which act to bring the interacting bases into the base pairing geometry specified by the user. RNABuilder has been used to predict the structure of the 200-nucleotide *Azoarcus* group I intron in the absence of any information of the solved *Azoarcus* intron crystal structure. The model accurately depicts the global topology, secondary and tertiary connections, and gives an overall RMSD value of 4.6 Å relative to the crystal structure.

Homology-based modelling can be used to predict any RNA molecules no matter how large or small, as long as the user can find a template and an effective alignment between the template and the target [14, 15, 79]. So this method is also called template-based modelling. Although the PDB/NDB database covers many important families, it may be difficult to find a proper template RNA for a particular target. In addition, creating an accurate and biologically relevant target–template sequence alignment is also a critical issue [79, 80].

11.2.2 Physics-Based Structure Prediction

Physics-based (ab initio) approach is based on the thermodynamics hypothesis [82]: the conformation with the lowest free energy corresponds to the native structure. Since a full-atomic structural model of RNA has a large number of degrees of freedom, which results in the huge computational complexity. For physical simplification, several prediction models with coarse-graining have been proposed at different resolution levels [16].

11.2.2.1 One-Bead Coarse-Grained Model

One-bead model uses one bead to represent a nucleotide, thus significantly reduce the spatial freedoms of an RNA structure. Several algorithms have been developed for predicting RNA 3D structures, such as YUP [17] and NAST [18].

YUP. Yammp Under Python (YUP) [17] is a general-purpose molecular mechanics program for multi-scaled coarse-grained modelling, in which Python is used as a programming/scripting language. It can be used to model RNA structures

as well as DNA and protein structures by extending the Python language through adding three new data types (atom maps, atom vectors and numerous energy types). In general, YUP is an extendable and useful tool for multi-scale modelling, but its potentials are required to be changed by the user according to the problem at hand. In addition, a fragment-based approach is used to add full-atomic details to the coarse-grained structure in YUP.

NAST. The Nucleic Acid Simulation Tool (NAST) [18] is a molecular dynamics simulation tool for predicting 3D structure for large RNA molecules based on secondary structures. Three types of data are also used to rank the conformational clusters produced from molecular dynamics simulations: (1) ideal small-angle X-ray scattering (SAXS) data; (2) experimental and ideal solvent accessibility (SAS) data; and (3) NAST energy (statistical information). NAST has been tested by building the structural models for two RNA molecules—the yeast tRNA^{Phe} (76-nt) and the P4-P6 domain of the *Tetrahymena thermophila* group I intron (158-nt), with the averaged RMSD 8.0 ± 0.3 and 16.3 ± 1.0 Å, respectively. Recently, the authors also developed a fully automated fragment- and knowledge-based method, called C2A (Coarse to Atomic) [19], to add full atomic details to coarse-grained models.

Both YUP and NAST are successful for large RNA molecules at nucleotide level, but they are limited by their prior need for secondary structure and the information of some tertiary contacts derived from both experimental and computational methods.

11.2.2.2 Three-Bead Coarse-Grained Model

Beyond the one-bead models [17–20], a number of coarse-grained models with higher resolution have been developed, such as three-bead [21, 24], five-bead [27] and six to seven-bead model [28].

Vfold. Cao and Chen have developed a three-vector virtual bond-based RNA folding model (Vfold) [21] for predicting RNA 3D tertiary folds from the sequence without using the experimental constraints. In Vfold, the loop conformations are produced by the self-avoiding random walks of the virtual bonds on a diamond lattice [22, 23] and the conformational entropy of RNA structures can be calculated. The Vfold model has been tested by a systematic benchmark including a wide range of RNA motifs (such as hairpin, duplex), pseudoknot, and a large RNA (a 122-nt 5S rRNA domain) with rmsd of about 3.5, 6.0 and 7.4 Å, respectively. Due to the rigid lattice constraints, Vfold is inadequate to study the folding dynamics of RNA.

iFoldRNA. iFoldRNA [25] is a web-based methodology for RNA 3D structure prediction and analysis of RNA folding thermodynamics. It is based on discrete molecular dynamics (DMD) and a force field (including base-pairing, base-stacking and loop entropy) [24]. The ifoldRNA has been tested by simulating a set of 153 RNA molecules within an average 4 Å deviation from experimental structures. Despite its rapid conformational sampling efficiency, the CPU time for DMD simulations also depends on RNA length. Ding et al. recently reported the

development of a qualitatively structure refinement approach using hydroxyl radical probing (HRP) measurements to drive DMD simulations for large RNA molecules (80 ~ 230 nt) with complex topologies [26].

The physical-based approaches not only emphasize the necessity of an accurate understanding of RNA tertiary structure but also illustrate the importance of native state dynamics. Although there are many models have been applied [16–28], how to build and choose proper force-fields is still a challenge.

11.2.3 Knowledge/Physics-Hybridized Structure Prediction

In protein structure prediction, the most successful approach is hybrid (*de novo*) modelling which combines the features of physics-based folding with the use of previously solved structures [83–85]. This hybrid (*de novo*) modelling strongly relies on the structural information from databases [79], and based on the principle, there are some existed programs for RNA structure prediction [29–34].

FARNA/FARFAR. Fragment assembly of RNA (FARNA) [31] is developed to predict RNA 3D structure from a sequence, while fragment assembly of RNA with full-atom refinement (FARFAR) [32] adds a refinement with atomic-level interactions to optimize RNA structures generated by FARNA. Based on knowledge-based energy function, FARNA can assemble three-nucleotide all-atom fragments with Monte Carlo algorithm. In a benchmark test of 20 RNA molecules (≤ 46 nt), FARNA reproduces better than 90 % of Watson-Crick base pairs. Smaller RNAs in the test are accurately reproduced with a resolution of better than 4.00 Å, but the probability of a FARNA prediction within a backbone rmsd of 4.00 Å decreases sharply as a function of RNA length. Nevertheless, combined with secondary structure and multiplexed hydroxyl radical cleavage analysis (MOHCA), FARNA can predict the structure for an RNA as large as 158 nt with the rmsd of 13 Å [32].

MC-Fold/MC-Sym. MC-Fold/MC-Sym pipeline [34] is another full-atomic RNA 3D structure prediction algorithm, which assembles RNA structures from a library of the nucleotide cyclic motif (NCM) [35]. MC-Fold predicts RNA secondary structure using a free energy minimization function, and MC-Sym builds full-atom 3D models of RNA structures based on the scripts generated by MC-Fold and 3D version of the NCM fragments. The predictive power of the pipeline has been confirmed by building 3D structures of precursor microRNA (pre-miRNA), and proposing a new 3D structure of the human immunodeficiency virus (HIV-1) cis-acting 21 frame-shifting element.

Knowledge/physics-hybridized modelling including FARNA/FARFAR and MC-Fold/MC-Sym pipeline is powerful for modelling small RNA molecules [29–32], but larger structures remain a challenge because of the computational requirements for full-atomic modelling. A coarse-grained approach would decrease computational requirements for modelling large structures.

11.3 RNA Folding Kinetics

11.3.1 Kinetic Model

Most approaches to kinetic RNA folding are based on the description of folding in terms of a stochastic process. Each model consists of three key ingredients: (1) The state space, i.e., the set of structures or conformations, (2) a move-set, i.e., the elementary transitions that can occur between such conformations, and (3) transition rates for each of these allowed transitions. The folding process can now be described as a continuous time Markov process, governed by a master equation for the state probabilities.

Consider an ensemble of conformational states. The population $p_i(t)$ for each state i at time t can be described by the following equation (master equation):

$$dp_i/dt = \sum_{\Omega} [k_{j \rightarrow i} p_j(t) - k_{i \rightarrow j} p_i(t)],$$

where Ω is the total number of conformations. $k_{j \rightarrow i}$ and $k_{i \rightarrow j}$ are the rate of the respective transitions, and they should satisfy the detailed balance condition: $p_i k_{i \rightarrow j} = p_j k_{j \rightarrow i}$, where p_i and p_j are the Boltzmann distribution of state i and j , $p_i = \frac{1}{Z} \exp(-\frac{\Delta G_i}{k_B T})$, and Z is the partition function $Z = \sum_i \exp(-\frac{\Delta G_i}{k_B T})$.

When Ω is not very large, the above rate equation can be written as the matrix form [86]: $d\mathbf{p}/dt = \mathbf{M} \cdot \mathbf{p}$, where \mathbf{p} is the vector for the population distribution, \mathbf{M} is the $\Omega \times \Omega$ rate matrix with the matrix elements defined by $M_{ij} = k_{j \rightarrow i}$ ($i \neq j$) and $M_{ii} = -\sum_{j \neq i} k_{i \rightarrow j}$. The equation can be solved with analytical form and the population kinetics is given by the eigenvalue spectrum for long times:

$$p(t) = \sum_{m=1}^{\Omega} C_m n_m e^{-\lambda_m t}, \quad (11.1)$$

where $-\lambda_m$ and n_m are the m -th eigenvalue and eigenvector of the rate matrix and C_m is the coefficient as determined by the initial condition.

Because the passage of a rate-limiting step is intrinsically related to the folding speed, it is possible to probe and to identify the rate-limiting steps through the folding from different unfolded initial conformations. In a master equation approach, slow and fast folding speeds are directly correlated to the large and small contributions of the rate-limiting slow kinetic modes. Because the contributions from the slow modes can be computed from the corresponding eigenvectors, Zhang and Chen proposed a general transition state searching method to identify the rate-limiting steps from the eigenvectors of the slow modes [87].

When Ω is large, if there exist discrete rate-limiting steps for the kinetic process, it would be possible to “renormalize” the conformational space into a number of conformational clusters. The large ensemble of chain conformations can thus be drastically reduced into a much smaller number of conformational

clusters [88, 89]. Different clusters are separated by the rate-limiting steps. If the rate-limiting steps involve sufficiently high kinetic barrier, the microstates within each cluster would have sufficient time to equilibrate and form a macrostate (in local equilibrium) before crossing the intercluster barriers to enter other kinetically neighboring clusters. The transitions between different clusters (macrostates) determine the overall folding kinetics of the molecule. Otherwise, it needs to simulate the process with Monte Carlo methods [90–93]. Due to the drawbacks such as limited sampling and slow calculation, a few methods have been applied to amend these. For examples, rejectionless Monte-Carlo approach was used to conserve the detailed balance condition [94], the simulated annealing techniques was used in order to accelerate folding [95], optimization techniques such as genetic algorithms rather than Monte-Carlo simulation was also used [96].

11.3.2 Conformation Space

The base pair is the basic subunit for RNA secondary structure, so a base-pair forming/melting corresponds to the smallest possible steps in conformation space [90]. Considering that RNA secondary structure is stabilized mainly by the base stacking interactions, and a single (unstacked) base pair is not stable and can quickly unfold, Zhang and Chen [86] defined an elementary kinetic step for RNA secondary structural change to be the formation/disruption of a stack or a stacked base pair. While this allows the most detailed description of folding pathways, the conformation space is so large that it leads to extremely long simulation runs or restricted to short sequences. To reduce the conformation space, many approaches therefore define the formation or destruction of an entire helix as the basic step [93, 97–100]. Another approach is that several uncorrelated base pairs are changed in a single time step [101]. Folding simulations in this scenario are similar to that of single base pair moves. But the relationship of the helices is different from that of base pairs [102]: the two helices can be compatible, partial compatible, and incompatible. A conformation state should consist of compatible or partial compatible helices [97–100]. Recently, thermodynamics-based RNA folding in a kinetic folding context. Coarse-grained landscapes in conjunction with stochastic sampling algorithms have been used to study the RNA folding kinetics [103]. By using the barrier trees and assuming that the basins of individual local minima are in quasi-equilibrium, the folding kinetics under transcription was studied [104]. Another approach for cotranscription folding combined the thermodynamic computations with coarse-grained local kinetics [105]. Flamm et al. developed a flooding algorithm that decomposes the landscape into basins surrounding local minima connected by saddle points [106]. Wolfinger et al. [107] use a partitioning of the landscape into macrostates, where a macrostate is defined as the set of all starting conformations for which a gradient walk ends in the same local minimum. The effective transition rates between any two macrostates were calculated from the barrier tree. Tang et al. [103, 108] adopt the probabilistic roadmaps to build an

approximated representation of the RNA folding landscape. In the roadmap graph, the vertex set represents valid sampled conformations of the folding landscape and edges are the possible transition path, and the time evolution of the population of different conformations can be calculated through the probabilistic roadmap.

11.3.3 Move Set and Kinetic Rate Models

11.3.3.1 Kinetic Rate Models

The kinetic rate for an elementary kinetic step is usually defined as:

$$k_{i \rightarrow j} = k_0 \exp\left(-\frac{\Delta G^+ - \Delta G_i}{k_B T}\right),$$

where ΔG^+ is the free energy of the transition state, ΔG_i is the free energy of the state i , k_0 is a constant. The actual models for the base pair kinetic move use:

$$k = k_0 \exp\left(-\frac{\Delta G}{2k_B T}\right),$$

where ΔG is the free energy difference between the two states. Schmitz and Steger [95] treat the stacking energy as the barrier when opening a base pair, the loop energy change as the barrier when closing a base pair. Zhang and Chen [86] define the transition rate for the formation (k_+) and the disruption (k_-) of a base stack as the following:

$$k_+ = k_0 \exp\left(-\frac{\Delta G_+}{k_B T}\right), \quad k_- = k_0 \exp\left(-\frac{\Delta G_-}{k_B T}\right),$$

where the prefactor k_0 is fitted from the experimental data and is equal is $6.6 \times 10^{12} \text{ s}^{-1}$ and $6.6 \times 10^{13} \text{ s}^{-1}$ for an AU and GC stack [109], k_B is the Boltzmann constant, T is the temperature, and ΔG_{\pm} is the free energy barrier for the respective transition. Assuming that the barrier for the formation of a stack is caused by the reduction in entropy, $\Delta G_+ = T\Delta S$. If the stack closes a loop, the formation of the stack is accompanied by concomitant entropic decrease for loop closure, thus, the kinetic barrier for loop closure is $\Delta G_+ = T\Delta S = T(\Delta S_{loop} + \Delta S_{stack})$ where ΔS_{loop} is the entropy of the loop and ΔS_{stack} is the entropy of the stack. Assuming that the barrier for the disruption of a base pair is caused by the energetic (enthalpic) cost ΔH to break the hydrogen bonding and the base stacking interactions: $\Delta G_- = \Delta H_{stack}$, where ΔH_{stack} is the enthalpy of the stack. Then the rates for the formation and disruption of a stack are:

$$k_+ = k_0 e^{-\Delta S_{stack}/k_B T};$$

$$k_- = k_0 e^{-\Delta H/k_B T}$$

respectively, and the rates for formation and disruption of a loop-closing stack (and the loop) are:

$$k_+^{loop} = k_0 \exp(-(\Delta S_{loop} + \Delta S_{stack})/k_B T), \quad k_-^{loop} = k_0 \exp(-\Delta H/k_B T).$$

11.3.3.2 Model to Calculate the Transition Rate for Helix Based Moves

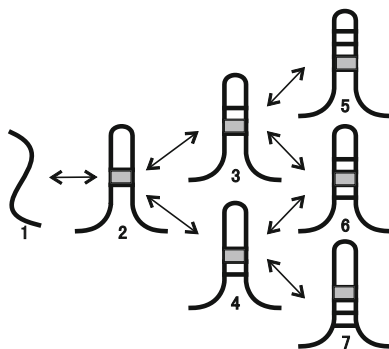
Tacker et al. [110] propose a rate model similar to that described for single base pairs: the activation energy is the change in loop energies when forming a helix, while opening a helix it is the stacking free energies. The same approach was adopted by in Refs. [111, 112]. Isambert [113] proposed that the free energy barrier for helix formation is the entropic penalty incurred by inserting the nucleus and the rate of is then given by an Arrhenius law using for nucleation. Zhao et al. [102] calculate the rate of a helix move set from that of the stack move set. If two conformations differ only in one helix, the transition between them would be the formation and disruption of the helix. Assuming that after the first stack is closed (with the concurrent formation of a loop), the helix will form along the zipping pathway. The rate of helix formation can be estimated along this zipping pathway. From the empirical thermodynamic parameters [114, 115], it can be found that for most RNA helices, the free energy landscape for a zipping pathway shows a downhill profile after the formation of the third base stack. Therefore, the rate k_f of the helix formation (along a specific pathway) is equal to the rate for the formation of the three-stack state. Considering the (slow) breaking of the stacks, for zipping along the $1 \rightarrow 2 \rightarrow 3$ pathway in Fig. 11.1:

$$k_f = k_{12} K_1 (1 - K'_2 K'_1 \sum_0^{\infty} K'_2 K_1) = k_{12} K_1 (1 - K'_2 K'_1 \frac{1}{1 - K'_2 K_1}), \quad (11.2)$$

where k_{ij} denotes the rate for the transition from state i to state j , K_1 and K'_1 are the forward (state $2 \rightarrow 3$) and reverse (state $2 \rightarrow 1$) probability of state 2, K_2 and K'_2 are the forward (state $3 \rightarrow 5$ and $3 \rightarrow 6$) and reverse (state $3 \rightarrow 2$) probability of state 3,

$$K_1 = \frac{k_{23}}{k_{23} + k_{21} + k_{24}}, K'_1 = \frac{k_{21}}{k_{23} + k_{21} + k_{24}}, K_2 = \frac{k_{35} + k_{36}}{k_{32} + k_{35} + k_{36}}, K'_2 = \frac{k_{32}}{k_{35} + k_{36} + k_{32}}. \quad (11.3)$$

Fig. 11.1 Multiple pathways for the formation of a helix after the first (nucleation) stack formed



For a given RNA molecule, the first base stack can be formed anywhere inside the helix. Therefore, the net rate k_F for the formation of a helix is the sum of the rates (Eq. 11.2) along the two pathways (Fig. 11.1) with the different first (nucleation) base stacks. The rate for the disruption of the helix can be estimated from the equilibrium constant of the helix: $k_U = k_F e^{\Delta G/k_B T}$, where ΔG is the folding free energy of the helix.

If two helices A and B overlap with each other, they cannot coexist in the same structure. The conversion of helix A to helix B through complete unfolding of helix A followed by refolding to B is extremely slow due to the high energy barrier to disrupt all the base stacks in helix A. Zhao et al. [102] proposed that there is a much faster tunneling pathway, which is classified as three process: (1) at first helix A partially disrupted, (2) exchanging, disruption of a base stack in A is accompanied by a concurrent formation of a base stack in B, (3) zipping, helix B grows through a zipping process. The pathway is fast because the formation of the base stacks in B tends to cause an overall downhill shape of the free energy landscape. This is similar to the Morgan-Higgs saddle point approach [116], in which the saddle point height is estimated as the highest point along the path. However, the free energy landscape suggests that there exist multiple high free energy points along the path. This (tunneling) pathway involves a much lower energy barrier to unwinding the helix than the complete unfolding pathway (Fig. 11.2). Based on the tunneling pathway, the rate for helix exchange is estimated as:

$$k_{A \rightarrow B} = \frac{\prod_i^n k_i}{\sum_{j=0}^{n-1} \left(\prod_{i=1}^j k'_i \prod_{m=j+2}^n k_m \right)}, \quad k_{B \rightarrow A} = k_{A \rightarrow B} e^{-\frac{\Delta G_{AB}}{k_B T}}. \quad (11.4)$$

In the above formula, k_n and k'_n are the rate constants for the process to formation (disruption) and disruption (formation) of a base stack in A (B), respectively.

When the conformation space is consisted of local minima, the transition rate is often calculated by searching the saddle point or from the free energy barrier tree [103–105].

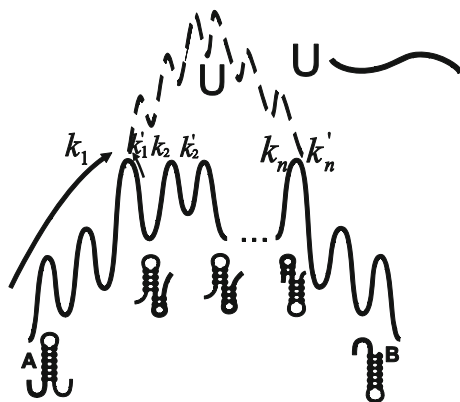


Fig. 11.2 The free energy landscape of the tunneling pathway that connects two overlapping helices *A* and *B*. *U* is the open state. The unfolding of *A* is accompanied by the folding of *B*. k_1 denotes the transition rate for the unfolding of helix *A* to form the first stack of helix *B*. $k'_1, k_2, k'_2, \dots, k_n, k'_n$ denote the transition rates between the neighboring intermediates along the tunneling pathways

11.3.3.3 Folding Kinetics During Transcription

The folding of functional RNA structures are often coupled with the transcription process [117–119]: since transcription is slow compared to local folding processes, the partially synthesized RNA will start folding while the molecule is still being synthesized. For instance, in the auto-catalyzed splicing reaction of tetrahymena group intron, the functional native structure may form within the timescale of transcription, which is much faster than the refolding of the complete chain in vitro [120–125]. Investigations about the RNA component of *Bacillus subtilis* RNase P folding indicate site-specific pausing could greatly influence the folding result of RNA molecule [126]. Addition of NusA which causes pausing in the process of transcription provides longer duration of temporary RNA chain to undergo the conformational search. Recently, several RNA folding kinetics algorithms were developed in connection with the thermodynamic energetics of the folding system. For instance, by using the barrier trees and assuming that the basins of individual local minima are in quasi-equilibrium, the folding kinetics under transcription was studied [104]. Combining the thermodynamic properties with coarse-grained local folding kinetics, a heuristic approach was also developed to successfully predict cotranscriptional folding for large RNAs [105]. Zhao et al. [127] treat the transcription of a single nucleotide as an elementary time step. The real time for each step is a constant or variable if the nucleotides are synthesized at a constant or variable speed, respectively. The transcriptional pausing at a specific site can be simulated by assigning a large number of effective time steps for the corresponding (paused) step. If the transcription speed of an RNA sequence is v nucleotides per second, the (real) time window for each step would be $1/v$ seconds, i.e., the polymerase spends

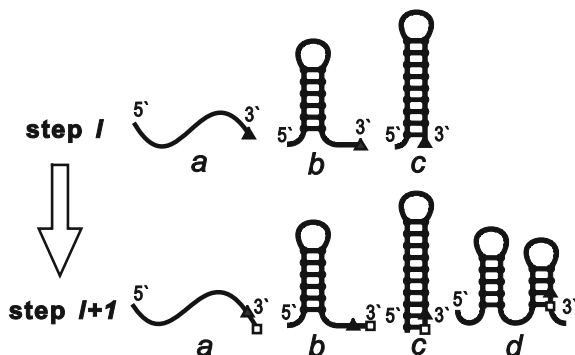


Fig. 11.3 The relationships between l -nt and $(l + 1)$ -nt structures: elongation of an open chain (*a*), a dangling tail (*b*), a helix (*c*), and the formation of a new structure (*d*). The filled triangle denotes the last transcribed nucleotide in step l , and the square denotes the last transcribed nucleotide in step $l + 1$

$1/\nu$ seconds to synthesis a nucleotide. Assuming that at time t the l -nt chain is (newly) transcribed, the population distribution of the l -nt chain conformational space is relaxed from $[P_1(l)_{begin}, P_2(l)_{begin}, \dots, P_\Omega(l)_{begin}]$ to $[P_1(l)_{end}, P_2(l)_{end}, \dots, P_\Omega(l)_{end}]$ at time t to time $t + 1/\nu$, when the $(l + 1)$ -th nucleotide is transcribed, here Ω is the number of conformations for an l -nt chain (Fig. 11.3). The beginning population of the $(l + 1)$ -th step is inherited from the ending population of the l -th step. However, the RNA chain in the $(l + 1)$ -th step is one nucleotide longer than in the l -th step. According to the possible changes of the structures upon the elongation of the chain by one nucleotide, the structures are classified as four types. The population distribution at the beginning of step $l + 1$ can be derived from that of the step at the end: $P(l + 1)_{begin} = P(l)_{end}$ for *a*, *b*, & *c*; $P(l + 1)_{begin} = 0$ for *d*. Applying this method from the first step to the end of transcription, we compute the folding kinetics for the RNA chain during transcription.

11.4 Metal Ions in RNA Folding

11.4.1 Ions Stabilize RNA/DNA Folded Structure

11.4.1.1 Ion Binding to RNAs/DNAs

Metal ions would like to bind to negatively charged nucleic acids to neutralize the negatively charged RNAs/DNAs [128–139]. The number of binding ions is important to DNA/RNA structure and stability, and can be measured via several experimental methods such as the small angle X-ray scattering (SAXS) [128–139], the ion-counting method [132], and the thermodynamic method [133–138].

The experimental methods have been applied to various RNAs/DNAs, including yeast 58-nt ribosomal RNA fragment [136], tRNA [137, 138], poly(A.U) [133], beet western yellow virus pseudoknot fragment [135], polymeric calf thymus DNA [134], oligomeric DNA/RNA duplexes [132], and DNA triplex [132]; see Ref [72] for a collection on the experimental data for ion binding to RNAs/DNAs. The extensive experimental data have yielded the following major conclusions:

- (1) The detailed distributions of binding ions near molecular surface are sensitive to the specific atomistic structure of the RNAs/DNAs [131], and the ion-binding's of different (monovalent and divalent) species of ions appear anti-cooperative [132];
- (2) Metal ions can give more efficient charge neutralization for RNA than for DNA. Such difference possibly comes from the higher charge density on backbone of A-form helix has than B-form helix [131].
- (3) Multivalent ions (e.g., Mg^{2+}) are much more efficient in charge neutralization than monovalent ions (e.g., Na^+). Such unusually higher efficiency of multivalent ions is beyond the mean-field description such as ionic strength, and is more pronounced for larger RNAs with more compact structures.

In addition to the diffusive ion binding, the specific ion binding may make significant contribution to stabilizing specific RNA folded structure and the function of RNA. The specific ion binding may be related to the RNA sequence, the local geometry, and the property of ion and water, and is a challenge in both experiments and modelling.

11.4.1.2 Ion Contribution to Flexibility of Single-Stranded RNA

Single-strand (ss) RNA is a fundamental segment in RNA structure and the flexibility of ssRNA is important to the global stability of RNA. The ion contribution to the flexibility (e.g., persistence length l_p) of single-strand RNAs have been quantified by a variety of experimental approaches, such as force-extension, single molecule fluorescence resonance energy transfer (smFRET), small angle X-ray scattering (SAXS), and fluorescence recovery after photobleaching, over different kinds of ssRNAs/DNAs [140–147]. The major conclusions are in the following:

- (1) Mg^{2+} is approximately 60-120 times more efficient than Na^+ in neutralizing ss RNAs/DNAs, which is beyond the mean-field concept (e.g., ionic strength) [145];
- (2) The persistence length of ss RNA/DNA decreases with the increase of $[Na^+]$ or $[Mg^{2+}]$, and the ion-concentration dependence is stronger for $[Na^+]$ than for $[Mg^{2+}]$ [145, 146];
- (3) The dependence of persistence length of ss RNA/DNA is stronger for longer sequence. For long ss generic sequence, there is a crude empirical formula for ion-dependent persistence length: $l_p = 5 + 1.5/\sqrt{I}$, where I is the ionic strength [147].

- (4) Poly(A)/poly(U) are more stiff than poly(T) at high salt while the ion-dependence of l_p are similar. Such stronger intrinsic stiffness may result from the stronger self-stacking of Poly(A)/poly(U) [144, 145].

However, there are also questions remained: (1) How do the flexibility of a ss RNA/DNA and its $\text{Na}^+/\text{Mg}^{2+}$ dependence rely on the surrounding space? (2) How is the $\text{Na}^+/\text{Mg}^{2+}$ -dependent l_p on the sequence length quantified? To answer the questions remains a challenge due to the high conformational fluctuation of ss chain and possible stronger correlations between Mg^{2+} .

11.4.1.3 Ions Stabilize Helices and Hairpins

Helix is the most fundamental segment of RNA structure (ranging from several to about ten base pairs), and hairpin is the simplest secondary structural motif. The thermodynamic experiments have revealed that the stability of helices and hairpins is sensitive to ionic environments. Most of the experiments were performed in a Na^+ solution or a mixed $\text{Na}^+/\text{Mg}^{2+}$ solution; see Ref. [72] for a brief summary on the experimental data [148–159]. These thermodynamic data lead to the following major features for ion effects in helix and hairpin stability:

- (1) In Na^+ or K^+ solution, the stabilities of DNA/RNA helices/hairpins depend on ion concentration with the approximately linear dependence on the logarithm of Na^+ or K^+ concentration and such dependence is strong at low salt (<0.1 M Na^+ or K^+), and relatively weak at a high Na^+ or K^+ concentration (≥ 0.1 M Na^+);
- (2) Compared with Na^+ or K^+ , divalent ions (e.g., Mg^{2+}) are more efficient in stabilizing helices/hairpins. For an example, the stability for short DNA/RNA duplexes/hairpins in a 10 mM Mg^{2+} solution is approximately equivalent to the stability in a 1 M Na^+ solution [139, 147–149].

The thermodynamic parameters for the formation of helix and loop have been measured extensively at 1 M Na^+ (i.e., the standard ion condition). These parameters have enabled the accurate predictions on RNA (DNA) secondary structure, stability and kinetics [160–164]. For ion condition other than 1 M NaCl , RNA/DNA thermodynamic data and theoretical modelling for various ionic conditions yields a set of fitted formulas for the thermodynamic parameters of RNA/DNA helices versus $\text{Na}^+/\text{Mg}^{2+}$ concentrations. In contrast to Na^+ solutions, experimental data on Mg^{2+} -dependent helix/hairpin stability has been relatively limited, and the $[\text{Mg}^{2+}]$ -dependent thermodynamic parameters [154, 155] may need to be validated through more extensive experimental data; see the Sect. 11.4.3. For ion-dependent loop formation thermodynamics, the hairpin loop stability has been derived as functions of $[\text{Na}^+]$ and $[\text{Mg}^{2+}]$, based on the statistical mechanical modelling [156]; see the Sect. 11.4.3.

11.4.1.4 Ions Stabilize Tertiary Structures

Generally, RNA tertiary structure is folded by the aggregation of secondary segments, the minor rearrangements of secondary segments and the formation of tertiary contact. Since RNA tertiary folding generally involves massive charge build-up, ion-RNA interaction is stronger for tertiary structures and consequently the quantitative understanding on ion effects in RNA tertiary folding becomes more challenging. Extensive experiments have investigated how metal ions assist RNA tertiary folding and stabilize tertiary structures for various RNAs, such as tRNA, 58-nt ribosomal RNA fragment, beet western yellow virus pseudoknot fragment, Tetrahymena ribozyme, kissing complex etc.; see Ref. [72] for a summary for the experimental data of ion effects in RNA tertiary folding [165–179]. These experiments have revealed the following important major features on the effects of metal ions, especially Mg^{2+} :

- (1) Metal ions of higher charge density (i.e. higher valence and smaller size) are more efficient in stabilizing RNA tertiary folds [166, 169]. For the Tar–tar RNA complex, smaller ions can enhance the folding stability [170].
- (2) Mg^{2+} can make a significant contribution to RNA tertiary structure stability even at high Na^+ (or other monovalent ions) concentration, and Mg^{2+} can induce more compact folded structures than Na^+ .
- (3) For HIV-1 dimerization initiation signal (DIS) type kissing loop-loop complexes, the melting temperature shows much stronger ion-dependence than for the corresponding duplex of the same sequence at the kissing interface [174, 175].
- (4) The higher efficiency of Mg^{2+} over Na^+ is much more pronounced for the kissing loop complex than for the duplex [174]. Such phenomena may result from the significantly higher massive built-up when loop-loop kissing.

In addition to the non-specific effects of metal ions shown above, some experiments also suggest that, depending on the sequence and geometry, specific interactions of binding ions with the RNA could make critical contribution to RNA tertiary structure [72, 135, 136]. The unclear understanding on roles of specific binding ions suggest the demand for the further more careful and extensive investigations, especially theoretical investigations, on the role of specific binding ions.

11.4.1.5 Ion-Mediated Structural Collapse

RNA structural collapse during tertiary folding often involves the helix-helix packing, and therefore, the helix-helix interaction is important for RNA tertiary folding. Rau and Parsegian have performed osmotic pressure measurements to quantify the ion-mediated interactions between *long* DNA helices [159, 160], leading to the following general conclusions:

- (1) Multivalent ions, such as Co^{3+} , can induce effective attraction between DNA helices, while monovalent ions (e.g., Na^+) can only screen the helix-helix electrostatic repulsion [180];
- (2) Certain types of divalent ions such Mn^{2+} can induce effective helix-helix attractive force [181], while other divalent ions such as Ca^{2+} cannot. Mg^{2+} in the presence of methanol could induce the effective helix-helix attraction [181]. The different roles of divalent ions might be attributed to the different ion binding affinities to the different groups [1]. For example, Mn^{2+} likes to binding into grooves, while Ca^{2+} likes to binding to phosphate groups [1].

However, in realistic RNA structures, helices are generally very short (ranging from several to around ten base pairs). Ions may have different effects in the effective interactions between short helices from long helices due to the greater rotational freedom and stronger end-effects of short helices. The recent experiments for short helices [68, 182–185] indicate the following conclusions:

- (1) For a system of dispersed short DNA helices, the SAXS experiments suggest that Mg^{2+} of high concentration can induce effective helix-helix attraction through end-end stacking [183].
- (2) For a system of loop-tethered short helices, the SAXS experiments suggested a possible weak side-side helix-helix attraction in a Mg^{2+} solution of high concentration (~ 0.6 M) [68].
- (3) The experiments showed that the PB theory underestimates the efficiency of Mg^{2+} in RNA structural collapse by over 10 times [184].
- (4) In trivalent ion solution, short DNA duplexes can become condensed while RNA duplexes keep soluble [185].
- (5) Mg^{2+} cannot condense long DNA duplexes while could condense short DNA triplex in aqueous solution [181, 184].

However, for the system of short helices, further investigations are still required to make clear: (1) How do the different ions (with different valences and sizes) cause the different effective helix-helix interactions? (2) Is the relaxation state a randomly disordered state or a state with certain order or ion-specific?

11.4.2 Theoretical Modelling for Ion Electrostatics

To quantitatively explore the ion effects in RNA folding, some theoretical approaches have been developed. The application of these theories on the ion-RNA (DNA) system has significantly enhanced the quantitative understanding on the ion role in RNA folding, which will be introduced in the following; see Ref. [186] for a review on the theoretical models.

11.4.2.1 Counterion Condensation Theory

The counterion condensation (CC) theory was developed for describing the interaction between ion and long DNA [187, 188]. In the theory, a DNA helix is approximated as a line-charge, and metal ions around DNA are either in the condensed state (near the DNA surface) or in the free state (away from the vicinity of DNA). The binding of an ion would decrease the electrostatic energy and simultaneously increase the ion entropic free energy. The competition between the two components (electrostatic energy and ion entropy) determines the thermodynamically equilibrium state.

The application of the CC theory on the effect of monovalent ions (e.g., Na^+ , K^+) in DNA helix thermodynamics [160, 187] shows a linear dependence of melting temperature T_m of DNA helix on the logarithm of monovalent ion concentration, which is in accordance with the experimental measurements. For the system of multi-body helices, the CC theory predicts that two DNA helices attract each other in both monovalent and multivalent salts. For lower salt concentration, the predicted attractive force becomes stronger while two helices are equilibrated at a larger separation [189]. The predictions are somewhat inconsistent with the experiment measurements [180, 181, 183] and computer simulations [190, 191] on nucleic acid helix-helix interactions.

Although the CC theory has gained the great success in the analysis of DNA thermodynamics, the theory still has the serious shortcomings: (1) The CC theory cannot be strictly employed to the RNA with complex structure in finite salt solutions since the theory is derived based on the assumptions of infinite-length DNA line-charge structural model and a infinite-dilute ion concentration; (2) The CC theory ignores the fluctuation and correlation of condensed ions, by assuming a uniform distribution of condensed ions along DNA. Consequently, the theory may become invalid for the multivalent ion solution where the correlations can be strong.

11.4.2.2 Poisson-Boltzmann Theory

The Poisson-Boltzmann (PB) theory has its early and simplified versions known as Gouy-Chapmann theory and Debye-Huckel theory. These two theories are the simplified versions of the PB theory for different specific systems [186]. The PB equation can be derived based on the Poisson equation for mean electric potential ψ and a Boltzmann distribution for diffusive ions in solutions

$$\nabla \cdot [\varepsilon(\mathbf{r})\varepsilon_0 \nabla \psi(\mathbf{r})] = -4\pi \left[\rho_f + \sum_{\alpha} ec_{\alpha}^0 N_{AV} z_{\alpha} e^{-z_{\alpha} e \psi(\mathbf{r}) / k_B T} \right], \quad (11.5)$$

where $z_{\alpha} e \psi$ is approximated to be the electrostatic energy of a diffusive ion with ionic charge $z_{\alpha} e$. ε is the dielectric constant; ρ_f is the charge density of fixed

charges in biomolecules; and c_{α}^0 is the bulk concentration of ion species α . In the recent two decades, some algorithms have been developed to numerically solve the PB equation, and the PB theory has been widely used in the electrostatics of biomolecules in solutions [192–196]. For electrostatics of biomolecules in aqueous/monovalent ion solutions, the experimental comparisons show that the PB theory makes rather accurate predictions [e.g., 197].

However, a mean-field approximation is employed in deriving the PB equation, i.e., diffusive charges (ions) obey a mean Boltzmann distribution based on the mean electric potential instead of the potential of mean force. As the result, (1) the PB theory ignores the fluctuation of ions in solution by assuming a mean ion distribution; (2) the PB theory ignores the ion-ion correlations by assuming the mean electric potential for diffusive ions rather than the potential of mean force, and (3) the PB theory ignores the ion finite size by the point-charge approximation. Therefore, the PB could not make reliable predictions on the electrostatic interactions for nucleic acid in multivalent ion solution where ion-ion interactions can be strong [186, 198]. An important example is the (multivalent) ion-mediated like-charge interaction, the PB always predicts repulsive force between two like-charged polyelectrolytes (DNA helices), while the experiments have shown the attractive force in multivalent salts [180, 181]. For ion-mediated RNA structural collapse, the PB theory underestimates the efficiency of Mg^{2+} by over 10 times [179, 182].

11.4.2.3 Modified Models Beyond Mean-Field Approximation

Aiming to improve the prediction for polyelectrolyte-multivalent ions, many efforts have been made in the recent years to overcome the shortcomings of the PB theory. Here, we will introduce several major modified models beyond the mean-field approximation [198–207].

Size-modified Poisson-Boltzmann model. The simplest modification for the PB model is to incorporate discrete ion size into the model. Recently, a size-modified PB model was proposed based on the lattice gas formalism, where the ion solution is discretized into a lattice with grid cells which can be occupied by ions with finite size [199, 200]. The application of the model for 3-dimensional complex nucleic acids shows that the modification can improve the prediction on monovalent ion-binding profiles at high salt concentration by capturing the binding saturation effect at high ion concentration. But for RNA solution with multivalent ions, the model still cannot give reliable predictions since it ignores the ion-ion electrostatic correlations [199, 200].

Modified Poisson-Boltzmann theory based on Kirkwood/BBGY hierarchy. A modified PB model has been developed based on Kirkwood/BBGY hierarchy through taking into account the fluctuation potential and ion-exclusion term in the potential of mean force for diffusive ions [e.g., 201]. The comparisons with the computer simulations show that, the theory gives the improved predictions for

multivalent ion distributions near polyelectrolytes with ideal 3D shapes such as cylinder, sphere and plane. But for realistic nucleic acids (or proteins) with arbitrary 3D shape, the numerical solution requires huge computational cost and is computationally inapplicable for realistic nucleic acids/proteins with complex 3D shape because the equation for the fluctuation potential is coupled to the equation for the mean electrostatic potential [201, 202].

Correlation-corrected Poisson-Boltzmann model. Recently, a so-called correlation-corrected Poisson-Boltzmann model was developed to account for the ion-ion correlations, by introducing an effective potential between like-charge ions [203]. Such effective potential is the same as the Coulomb potential at large ion-ion separation, while becomes a reduced repulsive Coulomb potential for a close ion-ion separation. For the electric double layers, the comparisons with the computer simulations showed that the model makes improved predictions on multivalent ion distributions and predicts an attractive force between the two planes in the presence of multivalent ions [203]. However, the model is computationally expensive for RNAs with complex structures. Moreover, such effective potential is somewhat ad hoc and the model is still lack of the validation on thermodynamics of nucleic acids.

Other theories beyond the mean-field approximation. Other theories beyond the mean-field approximation such as the integration theory [204], the density function theory [205] and the local molecular field theory [206] have been developed to account for the ion-ion correlation effects around nucleic acids/polyelectrolytes. But due to the huge computation cost for realistic nucleic acid system, these theories are also practically inapplicable. Recently, a tightly bound ion (TBI) theory is developed by explicitly treating the strongly correlated ions which reside in the vicinity of nucleic acid surface [186, 198, 207–211]. The extensive experimental comparisons showed that this theory has been shown to make improved predictions on the ion effects for various nucleic acid structures in the presence of Mg^{2+} . In the following, we will focus on the TBI theory and its applications on modelling ion effects in DNA/RNA structure stabilities, including helices, hairpins, tertiary folds, and assembly.

11.4.3 Tightly Bound Ion Theory

As described above, extensive experiments have shown that Mg^{2+} plays a special role in RNA folding: Mg^{2+} is much more efficient than Na^+ in RNA folding and Mg^{2+} can induce more compact structure than Na^+ . For example, Mg^{2+} is generally $\sim 1,000$ times more efficient than Na^+ in RNA tertiary folding [169]. Aiming to quantitatively understand the role of multivalent ions in RNA folding, Tan and Chen have developed a TBI theory, by accounting for ion-correlations and fluctuations for realistic RNAs in ion solutions [186, 198, 207–211]. In the following, we will introduce the TBI theory with theoretical framework and applications on modelling ion effects in RNA/DNA structure stabilities.

11.4.3.1 Framework of the Tightly Bound Ion Theory

Since RNAs/DNAs are highly charged polyanionic molecules, the positively charged metal ions in solutions would aggregate on nucleic acid surface, causing high ion concentration in the vicinity of RNA/DNA surface. These condensed ions of high concentration would interact (correlate) strongly with each other. The correlation strength between ions can be characterized by the coupling parameter Γ [186, 198]

$$\Gamma(\mathbf{r}) = \frac{(z_z e)^2}{\varepsilon a_{wz}(\mathbf{r}) k_B T} \geq \Gamma_c. \quad (11.6)$$

Previous studies have shown that for ionic system, the change of coupling parameter Γ can induce the gas-liquid transition, and the critical value Γ_c was shown to reside in the range of [2.3, 2.9] [186, 198]. In the TBI theory, according to the critical inter-ion correlation strength Γ_c (2.6, a mean value over [2.3–2.9]), the ions around RNAs/DNAs are divided into two types: (strongly correlated) tightly bound ions in the vicinity of RNA and (weakly correlated) diffusive ions in the outer space. Correspondingly, the space around RNAs is also divided into the tightly bound region and diffusive region. Due to the weak inter-ion correlations, the diffusive ions can be treated by the mean-field PB approach. While for the (strongly correlated) tightly bound ion, the tightly bound region is discretized into different tightly bound cells, each of them around a (negatively) phosphate group. Every tightly bound cell can keep empty or be occupied by an ion, and all possible states of tightly bound cells (either empty or occupied by an ion) give the ensemble of ion-binding modes. The different ion-binding modes (M) in different cells are explicitly considered to account for the effects of ion correlations and fluctuations.

For a nucleic acid-ion system, the partition function Z is given by the summation of the partition function Z_M for all possible modes M

$$Z = \sum_M Z_M, \quad (11.7)$$

where Z_M is given by [177, 186, 198, 208]

$$Z_M = Z^{id}(c_z)^{N_b} \left(\int \prod_{i=1}^{N_b} d\mathbf{R}_i \right) e^{-(\Delta G_b + \Delta G_d + \Delta G_b^{pol})/k_B T}. \quad (11.8)$$

Here, Z^{id} is the partition function for the uniform ion solution (without RNAs). N_b is the number of the tightly bound ions for model M . c_z is the bulk concentration of the z -valent ions, and \mathbf{R}_i denotes the position of tightly bound ion i . $\int \prod_{i=1}^{N_b} d\mathbf{R}_i$ is the volume integral over the tightly bound region for the N_b tightly bound ions. ΔG_b is the free energy for the discrete charges in the tightly bound region

(including the tightly bound ions and phosphate charges); ΔG_d is the free energy for the diffusive ions, including the electrostatic interactions between the diffusive ions, and between the diffusive ions and the charges in the tightly bound region as well as the entropic free energy of the diffusive ions; ΔG_b^{pol} is the (Born) self-polarization energy for the discrete charges within the tightly bound region [177, 186, 198, 208].

The TBI theory has been widely applied to quantitatively understanding the ion contributions to RNA secondary and tertiary structure stability, which will be described in the following.

11.4.3.2 Modelling Ion Effects Stability of DNA/RNA Helices

The stability of helices is essential to the global stability and the functions of RNAs (DNAs). Due to the polyanionic nature, metal ions can be important to the stability of DNA/RNA helices. Based on a polyelectrolyte theory, the melting of a helix can be modelled as a two-state model. The free energy change due to the melting can be decoupled into a non-electrostatic contribution ΔG^{NE} and an electrostatic contribution ΔG^E , and ΔG^{NE} can be evaluated by combining the experimental data at a reference state with a polyelectrolyte theory (e.g. the TBI theory) [154, 155]

$$\begin{aligned}\Delta G &= \Delta G^E + \Delta G^{NE} \\ &= \Delta G^E + (\Delta G_{1MNa^+} - \Delta G_{1MNa^+}^E).\end{aligned}\quad (11.9)$$

With the use of the TBI theory for treating ion-DNA(RNA) interactions, the Na^+/Mg^{2+} dependence of helix stability can be quantitatively evaluated.

The comparisons with the extensive experimental data showed that the TBI theory makes reliable predictions on the stability of DNA and RNA helices in Na^+/Mg^{2+} solutions [154, 155]. Furthermore, the comprehensive calculations with the TBI theory give a series of empirical formulas for describing the thermodynamics of DNA (RNA) helices in Na^+/Mg^{2+} solutions.

Thermodynamic parameters for DNA helix in Na^+/Mg^{2+} solution.

For a DNA helix in Na^+ solution, the following formulas of Na^+ -dependent thermodynamics can be obtained from the TBI theory [154]

$$\begin{aligned}\Delta G[Na^+] &= \Delta G[1\text{ M } Na^+] + (N - 1)\Delta g_1^{DNA}; \\ \Delta S[Na^+] &= \Delta S[1\text{ M } Na^+] - 3.22(N - 1)\Delta g_1^{DNA}; \\ 1/T_m[Na^+] &= 1/T_m[1\text{ M } Na^+] - 0.00322(N - 1)\Delta g_1^{DNA}/\Delta H[1\text{ M } Na^+],\end{aligned}\quad (11.10)$$

where ΔG , ΔS , T_m , ΔH are the free energy change, entropy change, melting temperature, enthalpy change for helix formation at $[Na^+]$ (in molar), or 1 M $[Na^+]$ (standard ion condition). Δg_1^{DNA} is a function associated with electrostatic folding

free energy per base stack of DNA, and is a function of helix length and $[\text{Na}^+]$ [154]

$$\begin{aligned}\Delta g_1^{DNA} &= a_1^{DNA} + b_1^{DNA}/N; \\ a_1^{DNA} &= -0.07 \ln[\text{Na}^+] + 0.012 \ln^2[\text{Na}^+]; \\ b_1^{DNA} &= 0.013 \ln^2[\text{Na}^+].\end{aligned}\quad (11.11)$$

The thermodynamics for DNA helix at any given $[\text{Na}^+]$ can be calculated through the above empirical formulas, since those at 1 M $[\text{Na}^+]$ (standard ion condition) can be obtained from the nearest neighbor model with the experimental parameters of SantaLucia et al. [160]. The quantitative comparisons with extensive experimental data show that the empirical formulas can give rather accurate estimates for thermodynamics of DNA helices in Na^+ solutions [154].

For the thermodynamics of a DNA helix in Mg^{2+} solution, the TBI model gives the following similar empirical formulas [154]

$$\begin{aligned}\Delta G[\text{Mg}^{2+}] &= \Delta G[1 \text{ M Mg}^{2+}] + (N-1)\Delta g_2^{DNA}; \\ \Delta S[\text{Mg}^{2+}] &= \Delta S[1 \text{ M Mg}^{2+}] - 3.22(N-1)\Delta g_2^{DNA}; \\ 1/T_m[\text{Mg}^{2+}] &= 1/T_m[1 \text{ M Mg}^{2+}] - 0.00322(N-1)\Delta g_2^{DNA}/\Delta H[1 \text{ M Mg}^{2+}],\end{aligned}\quad (11.12)$$

where Δg_2^{DNA} is given by

$$\begin{aligned}\Delta g_2^{DNA} &= a_2^{DNA} + b_2^{DNA}/N^2; \\ a_2^{DNA} &= 0.02 \ln[\text{Mg}^{2+}] + 0.0068 \ln^2[\text{Mg}^{2+}]; \\ b_2^{DNA} &= 1.18 \ln[\text{Mg}^{2+}] + 0.344 \ln^2[\text{Mg}^{2+}].\end{aligned}\quad (11.13)$$

Through the above empirical formulas, the thermodynamics of a DNA helix at a given $[\text{Mg}^{2+}]$ can be calculated easily. The experimental comparisons show that the empirical formulas can make reliable estimates for the stability of a short DNA helix ranging from 6-bp to 30-bp at an arbitrary $[\text{Mg}^{2+}]$ [154].

Generally, a buffer contains both Na^+ (or K^+) and Mg^{2+} ions. The TBI theory also gives the empirical formulas for DNA helix thermodynamics in a mixed $\text{Na}^+/\text{Mg}^{2+}$ solution [155]

$$\begin{aligned}\Delta G &= \Delta G[1 \text{ M Na}^+] + (N-1)(x_{\text{duplex}}\Delta g_1^{DNA} + (1-x_{\text{duplex}})\Delta g_2^{DNA}) + \Delta g_{12}; \\ \Delta S &= \Delta S[1 \text{ M Na}^+] - 3.22((N-1)(x_{\text{duplex}}\Delta g_1^{DNA} + (1-x_{\text{duplex}})\Delta g_2^{DNA}) + \Delta g_{12}); \\ 1/T_m &= 1/T_m[1 \text{ M Na}^+] - 0.00322((N-1)(x_{\text{duplex}}\Delta g_1^{DNA} + (1-x_{\text{duplex}})\Delta g_2^{DNA}) + \Delta g_{12})/\Delta H[1 \text{ M Na}^+],\end{aligned}\quad (11.14)$$

where x_{duplex} stands for the contribution fraction from Na^+ , and Δg_{12} is a crossing term. x_{duplex} and Δg_{12} are given by

$$x_{duplex} = \frac{[\text{Na}^+]}{([\text{Na}^+] + (8.1 - 32.4/N)(5.2 - \ln[\text{Na}^+])[\text{Mg}^{2+}])}; \quad (11.15)$$

$$\Delta g_{12} = -0.6x_{duplex}(1 - x_{duplex}) \ln[\text{Na}^+] \ln((1/x_{duplex} - 1)[\text{Na}^+])/N,$$

where $[\text{Na}^+]$ and $[\text{Mg}^{2+}]$ are both in molar. The comparisons with experimental data show that the above formulas give good estimate for the thermodynamics of a DNA helix in mixed $\text{Na}^+/\text{Mg}^{2+}$ solutions [155].

Thermodynamic parameters for RNA helix in $\text{Na}^+/\text{Mg}^{2+}$ solution.

For an RNA helix in a Na^+ solution, the thermodynamics can also be formulated by Eq. (11.10), except that Δg_1^{DNA} needs to be replaced by Δg_1^{RNA} . Δg_1^{RNA} can be given by [155]

$$\begin{aligned} \Delta g_1^{RNA} &= a_1^{RNA} + b_1^{RNA}/N; \\ a_1^{RNA} &= -0.075 \ln[\text{Na}^+] + 0.012 \ln^2[\text{Na}^+]; \\ b_1^{RNA} &= 0.018 \ln^2[\text{Na}^+]. \end{aligned} \quad (11.16)$$

The combination of Eq. (11.10) and Eq. (11.16) can give good estimate for an RNA helix in a Na^+ solution, as shown in Ref [155].

Similarly, for an RNA helix in a Mg^{2+} solution, the thermodynamics can be described by Eq. (11.12), except that Δg_2^{DNA} needs to be changed into Δg_2^{RNA}

$$\begin{aligned} \Delta g_2^{RNA} &= a_2^{RNA} + b_2^{RNA}/N^2; \\ a_2^{RNA} &= -0.6/N + 0.025 \ln[\text{Mg}^{2+}] + 0.0068 \ln^2[\text{Mg}^{2+}]; \\ b_2^{RNA} &= \ln[\text{Mg}^{2+}] + 0.38 \ln^2[\text{Mg}^{2+}]. \end{aligned} \quad (11.17)$$

For an RNA helix in a mixed $\text{Na}^+/\text{Mg}^{2+}$ solution, the thermodynamics can be calculated by the following empirical formulas

$$\begin{aligned} \Delta G &= \Delta G[1 \text{ M Na}^+] + (N - 1)(x_{duplex}\Delta g_1^{RNA} + (1 - x_{duplex})\Delta g_2^{RNA}) + \Delta g_{12}; \\ \Delta S &= \Delta S[1 \text{ M Na}^+] - 3.22((N - 1)(x_{duplex}\Delta g_1^{RNA} + (1 - x_{duplex})\Delta g_2^{RNA}) + \Delta g_{12}); \\ 1/T_m &= 1/T_m[1 \text{ M Na}^+] - 0.00322((N - 1)(x_{duplex}\Delta g_1^{RNA} + (1 - x_{duplex})\Delta g_2^{RNA}) + \Delta g_{12})/\Delta H[1 \text{ M Na}^+], \end{aligned} \quad (11.18)$$

where x_{duplex} and Δg_{12} are given by Eq. (11.15). Δg_1^{RNA} and Δg_2^{RNA} are given by Eqs. (11.16 and 11.17), respectively. As shown in Ref. [155]. The predictions from Eq. (11.18) are quite reliable for the thermodynamics of an RNA helix in a mixed $\text{Na}^+/\text{Mg}^{2+}$ solution (Fig. 11.4).

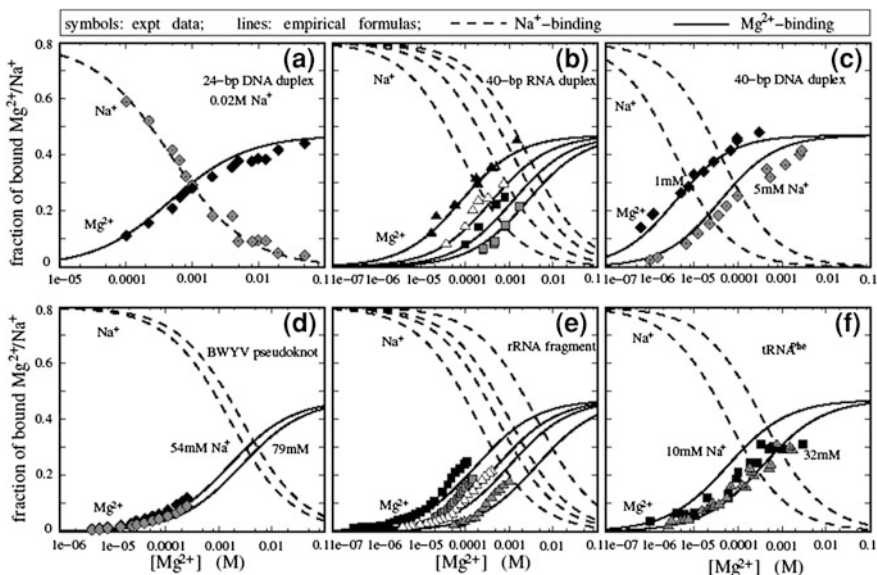


Fig. 11.4 The Mg^{2+} and Na^+ binding fractions per nucleotide for various RNA/DNA molecules. The *solid lines* are from the empirical formulas (Eqs. 11.26 and 11.28); and the symbols are experimental data. **a** 24-bp DNA duplex in $[Mg^{2+}]$ with fixed $[Na^+] = 20$ mM [132]; **b** 40-bp RNA duplex. The experimental data are for poly(A.U) [133]; From the *left to right*, $[Na^+] = 10, 29, 60,$ and 100 mM, respectively; **c** 40-bp DNA duplex. The experimental data are for the calf thymus DNA [134]; **d** BWYV pseudoknot RNA [135]; **e** 58-nt rRNA fragment [136]. Please note that the experimental data are for mixed Mg^{2+}/K^+ (not Mg^{2+}/Na^+) solution: from *left to right*, $[K^+] = 20, 40, 60,$ and 150 mM, respectively. Here we show the experimental data for semi-quantitative comparisons. **f** Yeast $tRNA^{Phe}$ [137, 138]

11.4.3.3 Modelling Ion Effects in Stability of DNA/RNA Hairpins

An RNA/DNA hairpin consists of a helix stem and a hairpin loop. The Na^+/Mg^{2+} dependence of a helix can be quantified by the empirical formulas described above. The TBI model can also give the empirically analytical Na^+/Mg^{2+} -dependent thermodynamics for a single-stranded loop, with the combination with the virtual bond model for the single-stranded loop conformation [156].

For a loop formation in Na^+ solutions, the systematic calculations of the TBI model give the following empirical relation for the folding free energy of an N -nt loop with end-to-end distance l [156]:

$$\Delta G[Na^+] = -k_B T \left(a_1 \ln(N - l/d + 1) + b_1 (N - l/d + 1)^2 - b_1 - (c_1 N - d_1) \right), \tag{11.19}$$

where $d = 6.4 \text{ \AA}$. The coefficients a_1 , b_1 , c_1 , and d_1 are given by

$$\begin{aligned} a_1 &= (0.02N - 0.026) \ln[\text{Na}^+] + 0.54N + 0.78; \\ b_1 &= (-0.01/(N + 1) + 0.006) \ln[\text{Na}^+] - 7/(N + 1)^2 - 0.01; \\ c_1 &= 0.07 \ln[\text{Na}^+] + 1.8; \\ d_1 &= 0.21 \ln[\text{Na}^+] + 1.5. \end{aligned} \quad (11.20)$$

For a loop in Mg^{2+} solutions, the empirical formulas from the TBI theory for the folding free energy can be written as

$$\Delta G[\text{Mg}^{2+}] = -k_B T \left(a_2 \ln(N - l/d + 1) + b_2 (N - l/d + 1)^2 - b_2 - (c_2 N - d_2) \right), \quad (11.21)$$

where a_2 , b_2 , c_2 , and d_2 are given by

$$\begin{aligned} a_2 &= (-1/(N + 1) + 0.32) \ln[\text{Mg}^{2+}] + 0.7N + 0.43; \\ b_2 &= 0.0002(N + 1) \ln[\text{Mg}^{2+}] - 5.9/(N + 1)^2 - 0.003; \\ c_2 &= 0.067 \ln[\text{Mg}^{2+}] + 2.2; \\ d_2 &= 0.163 \ln[\text{Mg}^{2+}] + 2.53. \end{aligned} \quad (11.22)$$

For a loop in mixed $\text{Na}^+/\text{Mg}^{2+}$ solutions, the folding free energy is represented by

$$\Delta G[\text{Na}^+/\text{Mg}^{2+}] = x_{loop} \Delta G[\text{Na}^+] + (1 - x_{loop}) \Delta G[\text{Mg}^{2+}], \quad (11.23)$$

where x_{loop} stands for the contribution fraction from Na^+ and is given by

$$x_{loop} = \frac{[\text{Na}^+]}{[\text{Na}^+] + (7.2 - 20/N)(40 - \ln[\text{Na}^+])[\text{Mg}^{2+}]}. \quad (11.24)$$

With the use of the above formulas for loop formation, the folding thermodynamics of hairpin loop, bulge loop, internal loop in an arbitrary $\text{Na}^+/\text{Mg}^{2+}$ solution can be easily calculated [156].

For a hairpin loop, the $\text{Na}^+/\text{Mg}^{2+}$ -dependent thermodynamics can be obtain by fixing the loop end-to-end distance at $\sim 17 \text{ \AA}$. Then the thermodynamics of an RNA (or DNA) hairpin can be calculated by the following formula [156]

$$\begin{aligned} \Delta G_{\text{hairpin}} &= \Delta H_{\text{stem}} - T \Delta S_{\text{stem}} + \Delta G_{\text{hairpin loop}}[\text{Na}^+/\text{Mg}^{2+}]; \\ \Delta H_{\text{stem}} &= \Delta H_{\text{stem}}[1 \text{ M Na}^+] + \Delta H_{\text{terminal mismatch}}; \\ \Delta S_{\text{stem}} &= \Delta S[\text{Na}^+/\text{Mg}^{2+}] + \Delta S_{\text{terminal mismatch}}, \end{aligned} \quad (11.25)$$

where ΔH_{stem} , $\Delta H_{\text{terminal mismatch}}$, and $\Delta S_{\text{terminal mismatch}}$ can be obtained from the nearest neighbour model with the measured thermodynamic parameters. $\Delta H_{\text{stem}}[\text{Na}^+/\text{Mg}^{2+}]$ can be given by the previously introduced empirical formulas (Eq. 11.14 for DNA and Eq. 11.18 for RNA). The extensive experimental comparisons show that the empirical formulas can make rather reliable predictions on the hairpin stability in a $\text{Na}^+/\text{Mg}^{2+}$ solution; see Ref. [156].

In a very recent single-molecule experiment, the quantitative comparisons with the measurements on a 20-bp RNA hairpin show that, the above empirical formulas are rather accurate in describing the $\text{Na}^+/\text{Mg}^{2+}$ -dependent thermodynamics for short RNA hairpin [146] (Fig. 11.4).

11.4.3.4 Modelling Ion Binding to RNA Tertiary Structures

As described above (Sect. 11.4.1.1), ion binding is critical for stabilizing RNA folded structure. The TBI theory has been developed for a static atomistic structure, and can quantify the ion atmosphere around an RNA (or DNA) with complex 3D structure [177]. In a mixed $\text{Na}^+/\text{Mg}^{2+}$ solution, the binding's of Na^+ and Mg^{2+} are competitive and anti-cooperative. The TBI model has given an empirical equivalence relation between Mg^{2+} and Na^+ as [177]

$$\log[\text{Na}^+]_{\text{Mg}} = A \log[\text{Mg}^{2+}] + B, \quad (11.26)$$

where $[\text{Na}^+]$ and $[\text{Mg}^{2+}]$ are both in millimolar (mM). A and B are two parameters depending on the (low-resolution) RNA (or DNA) structure

$$A = 0.65 + \frac{4.2}{N} \left(\frac{R_g}{R_g^0} \right)^2; \quad B = 1.8 - \frac{9.8}{N} \left(\frac{R_g}{R_g^0} \right)^2, \quad (11.27)$$

where N is the number of nucleotides of an RNA, and R_g is the radius of gyration of the RNA (or DNA) backbone, and R_g^0 is the radius of gyration of an N -nt RNA duplex.

Based on Eq. (11.23), the binding fractions of Na^+ and Mg^{2+} can be calculated through [177]

$$f_{\text{Na}^+} = \frac{[\text{Na}^+]}{[\text{Na}^+] + [\text{Na}^+]_{\text{Mg}}} f_{\text{Na}^+}^0; \quad f_{\text{Mg}^{2+}} = \frac{[\text{Na}^+]_{\text{Mg}}}{[\text{Na}^+] + [\text{Na}^+]_{\text{Mg}}} f_{\text{Mg}^{2+}}^0, \quad (11.28)$$

where $[\text{Na}^+]_{\text{Mg}}$ is given by Eq. (11.20). $f_{\text{Na}^+}^0$ and $f_{\text{Mg}^{2+}}^0$ are the binding fractions for pure Na^+ and pure Mg^{2+} solutions, respectively. Generally, $f_{\text{Na}^+}^0 \approx 0.8$, and $f_{\text{Mg}^{2+}}^0 \approx 0.47$ [177]. As shown in Fig. 11.4, the above empirical formulas could make reliable predictions for $\text{Na}^+/\text{Mg}^{2+}$ binding to RNAs/DNAs with complex 3D structures.

11.4.3.5 Modelling Salt Contribution to RNA Tertiary Structure Stability

Since RNA folding is hierarchical, the tertiary structure folding can be crudely modelled as a two-state transition from an intermediate (I) to the native (N) state. Similarly to the helix stability, the RNA tertiary folding free energy can be decoupled into two contributions: an electrostatic part and a nonelectrostatic part [178]

$$\begin{aligned}\Delta G &= \Delta G^E[\text{Na}^+/\text{Mg}^{2+}] + \Delta G^{NE}; \\ &= \Delta G^E[\text{Na}^+/\text{Mg}^{2+}] + (\Delta G[\text{expt Na}^+] - \Delta G^E[\text{expt Na}^+]); \\ &= \Delta G[\text{expt Na}^+] + (\Delta G^E[\text{Na}^+/\text{Mg}^{2+}] - \Delta G^E[\text{expt Na}^+]),\end{aligned}\quad (11.29)$$

where $\Delta G[\text{expt Na}^+]$ is the experimental folding free energy at a reference ion condition. ΔG^E can be given by the empirical formulas derived from the TBI model [178].

For an RNA folding in Na^+ solutions, ΔG^E can be calculated by the following empirical formula

$$\Delta G^E[\text{Na}^+] = \Delta G^E[1 \text{ M Na}^+] + a_1 N \ln[\text{Na}^+] + b_1 N \ln^2[\text{Na}^+], \quad (11.30)$$

where a_1 and b_1 are the parameters related to the RNA folded structure. a_1 and b_1 can be formulated by [178]

$$\begin{aligned}a_1 \times \varepsilon^*(T)T^* &= -0.086 + 7/(Nr_g^3 + 65); \\ b_1 \times \varepsilon^*(T)T^* &= 0.008 - 3.6/(N - 5)^2,\end{aligned}\quad (11.31)$$

where $r_g = R_g^0/R_g$. $\varepsilon^*(T) = \varepsilon(T)/\varepsilon(298.15 \text{ K})$ is the relative dielectric constant, and $T^* = T/298.15$ is the relative temperature.

For an RNA folding in Mg^{2+} solutions, the TBI model gives the following empirical formula for ΔG^E [178]

$$\Delta G^E[\text{Mg}^{2+}] = \Delta G^E[1 \text{ M Mg}^{2+}] + a_2 N \ln[\text{Mg}^{2+}] + b_2 N \ln^2[\text{Mg}^{2+}] + c_2 NT^*, \quad (11.32)$$

a_2 , b_2 , c_2 are given by

$$\begin{aligned}a_2 \times \varepsilon^*(T)T^* &= 0.012 - 1.4/(Nr_g^3 + 75); \\ b_2 \times \varepsilon^*(T)T^* &= 0.0048 - 57/(Nr_g^3 + N + 75)(N + 75); \\ c_2 \times \varepsilon^*(T)T^* &= -0.27 + 0.16/r_g^3 + 1.4/N.\end{aligned}\quad (11.33)$$

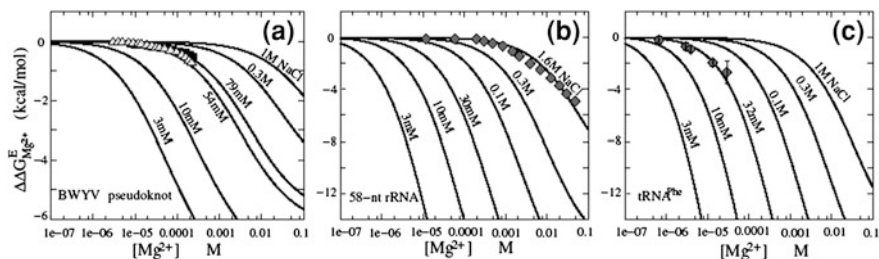


Fig. 11.5 The Mg^{2+} -contribution $\Delta\Delta G_{\text{Mg}^{2+}}^E$ to RNA tertiary structure folding free energy as a function of $[\text{Mg}^{2+}]$ for three RNA molecules: BWYV pseudoknot (a), 58-nt ribosomal RNA fragment (b), and yeast tRNA^{Phe} (c) at room temperature. *Solid lines*, empirical formulas derived from the TBI model; *symbols*, experimental data: **a** BWYV pseudoknot in 54 and 79 mM Na^+ solution [135]; **b** 58-nt rRNA fragment in solution with 1.6 M monovalent ions [178, 197]; **c** yeast tRNA^{Phe} in solution with 32 mM Na^+ [138, 179, 197]

For RNA folding in a mixed $\text{Na}^+/\text{Mg}^{2+}$ solution, ΔG^E is given by the empirical relation

$$\Delta G^E[\text{Na}^+/\text{Mg}^{2+}] = x_{3^o} \Delta G^E[\text{Na}^+] + (1 - x_{3^o}) \Delta G^E[\text{Mg}^{2+}] + N \Delta g_{12}, \quad (11.34)$$

where x_{3^o} denotes the contribution fraction from Na^+ , and Δg_{12} is a crossing term. x and Δg_{12} are given by

$$x_{3^o} = \frac{[\text{Na}^+]}{[\text{Na}^+] + \left(3.8 - 34/(N - 20)r_g^3\right)(1 + 0.2[\text{Na}^+])[\text{Mg}^{2+}]^{0.64}}; \quad (11.35)$$

$$\Delta g_{12} = -x_{3^o}(1 - x_{3^o})(0.26 - 1.2/(N - 20)).$$

With the use of the above empirical formulas for ΔG^E and Eq. (11.29), the $\text{Na}^+/\text{Mg}^{2+}$ -dependent RNA tertiary folding thermodynamics can be conveniently calculated. Figure 11.5 shows that Eq. (11.34) gives good estimates for the Mg^{2+} -contribution to the total folding stability $\Delta\Delta G_{\text{Mg}^{2+}}^E = \Delta G_{\text{Na}^+, \text{Mg}^{2+}}^E - \Delta G_{\text{Na}^+, \text{Mg}^{2+}=0}^E$, as compared with the experimental data. It is also shown that Eq. (11.29) with the empirical formulas for ΔG^E (Eqs. 11.30–11.35) can make good evaluation for the $\text{Na}^+/\text{Mg}^{2+}$ -dependent folding thermodynamics of small RNAs [178].

11.5 Perspectives

Although many RNA 3D structure modelling methods have proposed, further developments and refinements of the existing models are still required. Current algorithms have shown how the use of available experimental data can dramatically improve the structure prediction, e.g. the discrete molecular dynamics simulations

with the use of HRP measurements can predict structure of RNAs ranging in size from 80 to 230 nucleotides [26]. In addition, several algorithms have exploited the hierarchical properties of RNA folding [36–38], and consequently the prediction accuracy can be improved by adding the knowledge of secondary structure and tertiary contacts from experiments to the existed programs. However, there are still some essential problems remaining challenging, including: (1) Could the structures for larger RNA molecules be predicted reliably and efficiently? (2) Could RNA structures be predicted versus different environments (temperature, ion conditions, etc.)?

Despite the significant progress, modelling of RNA folding dynamics remains a challenging problem. The current form of the theories involves several limitations. First, the theories do not treat folding/unfolding of tertiary folds such as pseudo-knots. Second, the theories cannot treat, at the explicitly atomistic level, the effects of cofactors such as magnesium ions, ligands and proteins. In the future, we expect that the RNA folding kinetic theories can overcome these limitations and will be applicable to design RNA and DNA molecules with particular dynamic properties, which is of great importance in the emerging fields of synthetic biology and nucleic acid-based nanotechnology.

The extensive investigations have significantly enhanced the qualitative/quantitative understanding on ion effects in RNA folding. However, the quantitative understanding on ion roles is still challenging at least in the following issues: (1) How are the specific properties of ions correlated to their specific roles in RNA folding? (2) Is the efficient role of multivalent ions come from the inter-ion Coulombic correlation? (3) What are the roles of the specific-site binding of Mg^{2+} in RNA tertiary binding [212]? More issues related to RNA ion electrostatics includes: (1) What are the role of ions in RNA-ligand interaction? (2) What is the role of ions in RNA-protein interaction? To answer the questions requires the further development of theoretical modelling [209–211], combined with the progress in experiments.

Most above introduced progress in RNA folding problem were obtained for *in vitro* systems, while in cells, RNAs are surrounded by many other macromolecules. Therefore, in reality, RNA folds in a possibly interactive and dynamic confined space [213–216]. Limited existed investigations have revealed that the spatial confinement may significantly influence the folded structure and the ion role in folding [179, 214–216]. Further investigations on RNA folding should also involve the complex effects from the other macromolecules *in vivo*.

References

1. Bloomfield VA, Crothers DM, Tinoco IJ (2000) Nucleic acids: structure, properties and functions. University Science Books, Sausalito
2. Walter NG, Woodson SA, Batey RT (eds) (2009) Non-coding RNA. Non-protein coding RNAs. Springer, Berlin

3. Cruz JA, Westhof E (2009) The dynamic landscapes of RNA architecture. *Cell* 136:604–609
4. Zuker M, Stiegler P (1981) Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Res* 9:133–148
5. Hofacker IL, Fontana W, Stadler PF, Bonhoeffer S, Tacker M, Schuster P (1994) Fast folding and comparison of RNA secondary structures. *Monatsh Chem* 125:167–188
6. Ding Y, Lawrence CE (2003) A statistical sampling algorithm for RNA secondary structure prediction. *Nucleic Acids Res* 31:7280–7301
7. Do CB, Woods DA, Batzoglou S (2006) CONTRAfold: RNA secondary structure prediction without physics-based models. *Bioinformatics* 23:90–98
8. Mathews DH, Turner DH (2006) Prediction of RNA secondary structure by free energy minimization. *Curr Opin Struct Biol* 16:270–278
9. Massire C, Westhof E (1998) MANIP: an interactive tool for modelling RNA. *J Mol Graph Model* 16(197–205):255–257
10. Zwieb C, Mueller F (1997) Three-dimensional comparative modeling of RNA. *Nucleic Acids Symp Ser* 36:69–71
11. Jossinet F, Westhof E (2005) Sequence to Structure (S2S): display, manipulate and interconnect RNA data from sequence to structure. *Bioinformatics* 21:3320–3321
12. Jossinet F, Ludwig TE, Westhof E (2010) Assemble: an interactive graphical tool to analyze and build RNA architectures at the 2D and 3D levels. *Bioinformatics* 26:2057–2059
13. Martinez HM, Maizel JV, Shapiro BA (2008) RNA2D3D: a program for generating, viewing, and comparing 3-dimensional models of RNA. *J Biomol Struct Dyn* 25:669–683
14. Rother M, Rother K, Puton T, Bujnicki JM (2011) ModeRNA: a tool for comparative modeling of RNA 3D structure. *Nucleic Acids Res* 1–16
15. Flores SC, Wan YQ, Russell R, Altman RB (2010) Predicting RNA structure by multiple template homology modeling. *Pac Symp Biocomput* 15:216–227
16. Paliy M, Melnik R, Shapiro BA (2010) Coarse-graining RNA nanostructures for molecular dynamics simulations. *Phys Biol* 7:036001
17. Tan RKZ, Petrov AS, Harvey SC (2006) YUP: molecular simulation program for coarse-grained and multiscaled models. *J Chem Theory Comput* 2:529–540
18. Jonikas MA, Radmer RJ, Laederach A, Das R, Pearlman S, Herschlag D, Altman RB (2009) Coarse-grained modeling of large RNA molecules with knowledge-based potentials and structural filters. *RNA* 15:189–199
19. Jonikas MA, Radmer RJ, Altman RB (2009) Knowledge-based instantiation of full atomic detail into coarse-grain RNA 3D structural models. *Bioinformatics* 25:3259–3266
20. Taxilaga-Zetina O, Pliego-Pastrana P, Carbajal-Tinoco MD (2010) Three-dimensional structures of RNA obtained by means of knowledge-based interaction potentials. *Phys Rev E* 81:041914
21. Cao S, Chen SJ (2011) Physics-based de novo prediction of RNA 3D structures. *J Phys Chem B* 115:4216–4226
22. Cao S, Chen SJ (2006) Predicting RNA pseudoknot folding thermodynamics. *Nucleic Acids Res* 34:2634–2652
23. Cao S, Chen SJ (2005) Predicting RNA folding thermodynamics with a reduced chain representation model. *RNA* 11:1184–1897
24. Ding F, Sharma S, Chalasani P, Demidov VV, Broude NE, Dokholyan NV (2008) Ab initio RNA folding by discrete molecular dynamics: from structure prediction to folding mechanisms. *RNA* 14:1164–1173
25. Sharma S, Ding F, Dokholyan NV (2008) iFoldRNA: three-dimensional RNA structure prediction and folding. *Bioinformatics* 24:1951–1952
26. Ding F, Lavender CA, Weeks KM, Dokholyan NV (2012) Three-dimensional RNA structure refinement by hydroxyl radical probing. *Nat Methods* 9:603–608
27. Xia Z, Gardner DP, Gutell RR, Ren P (2010) Coarse-grained model for simulation of RNA three-dimensional structures. *J Phys Chem B* 114:13497–13506

28. Pasquali S, Derreumaux P (2010) HiRE-RNA: a high resolution coarse-grained model for RNA. *J Phys Chem B* 114:11957–11966
29. Zhang J, Dundas J, Lin M, Chen M, Wang W, Liang J (2009) Prediction of geometrically feasible three-dimensional structures of pseudoknotted RNA through free energy estimation. *RNA* 15:2248–2263
30. Zhang J, Bian YQ, Wang W (2012) RNA fragment modeling with a nucleobase discrete-state model. *Phys Rev E* 85:021909
31. Das R, Baker D (2007) Automated de novo prediction of native-like RNA tertiary structures. *Proc Natl Acad Sci* 104:14664–14669
32. Das R, Karanicolas J, Baker D (2010) Atomic accuracy in predicting and designing noncanonical RNA structure. *Nat Methods* 7:291–294
33. Bida JP, Maher LJ III (2012) Improved prediction of RNA tertiary structure with insights into native state dynamics. *RNA* 18:385–393
34. Parisien M, Major F (2008) The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. *Nature* 452:51–55
35. Lemieux S, Major F (2006) Automated extraction and classification of RNA tertiary structure cyclic motifs. *Nucleic Acids Res* 34:2340–2346
36. Zhao YJ, Gong Z, Xiao Y (2011) Improvement of the hierarchical approach for predicting RNA tertiary structure. *J Biomol Struct Dyn* 28:815–826
37. Gong Z, Zhao Y, Xiao Y (2010) RNA stability under different combinations of amber force fields and solvation models. *J Biomol Struct Dyn* 28(3):431–441
38. Seetin MJ, Mathews DH (2011) Automated RNA tertiary structure prediction from secondary structure and low-resolution restraints. *J Comput Chem* 32:2232–2244
39. Baumstark T, Schroder AR, Riesner D (1997) Viroid processing: switch from cleavage to ligation is driven by a change from a tetraloop to a loop E conformation. *EMBO J* 16:599–610
40. Perrotta AT, Been MD (1998) A toggle duplex in hepatitis delta virus self-cleaving RNA that stabilizes an inactive and a salt-dependent pro-active ribozyme conformation. *J Mol Biol* 279:361–373
41. Schultes EA, Bartel DP (2000) One sequence, two ribozymes: implications for the emergence of new ribozyme folds. *Science* 289:448–452
42. Kruger K, Grabowski P, Zaug AJ, Sands J, Gottschling DE, Cech TR (1982) Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of *Tetrahymena*. *Cell* 31:147–157
43. Bartel DP, Szostak JW (1993) Isolation of new ribozymes from a large pool of random sequences. *Science* 261:1411–1418
44. Joyce GF (1989) Amplification, mutation and selection of catalytic RNA. *Gene* 82:83–87
45. Ellington AE, Szostak JW (1990) In vitro selection of RNA molecules that bind specific ligands. *Nature* 346:818–822
46. Tuerk C, Gold L (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* 249:505–510
47. Mironov AS, Gusarov I, Rafikov R, Lopez LE, Shatalin K, Kreneva RA, Perumov DA, Nudler E (2002) Sensing small molecules by nascent RNA: a mechanism to control transcription in bacteria. *Cell* 111:747–756
48. Nahvi A, Sudarsan N, Ebert MS, Zou X, Brown KL, Breaker RR (2002) Genetic control by metabolite binding mRNA. *Chem Biol* 9:1043–1049
49. Winkler WC, Breaker RR (2003) Genetic control by metabolite-binding riboswitches. *Chem Bio Chem* 4:1024–1032
50. Bartel DP (2004) MicroRNAs: genomics, biogenesis, mechanism and function. *Cell* 116:281–297
51. Nudler E, Mironov AS (2004) The riboswitch control of bacterial metabolism. *Trends Biochem Sci* 29:11–17
52. He L, Hannon GJ (2004) MicroRNAs: small RNAs with a big role in gene regulation. *Nature Rev Genet* 5:522–531

53. Winkler WC, Nahvi A, Roth A, Collins JA, Breaker RR (2004) Control of gene expression by a natural metabolite-responsive ribozyme. *Nature* 428:281–286
54. Gerdes K, Wagner EGH (2007) RNA antitoxins. *Curr Opin Microbiol* 10:117
55. Nagel JHA, Gultyaev AP, Gerdes K, Pleij CWA (1999) Metastable structures and refolding kinetics in hok mRNA of plasmid R1. *RNA* 5:1408–1419
56. Groeneveld H, Thimon K, Duin J (1995) Translational control of maturation-protein synthesis in phage MS2: a role for the kinetics of RNA folding? *RNA* 1:79–88
57. Porschke D (1977) Elementary steps of base recognition and helix-coil transitions in nucleic acids. *Mol Biol Biochem Biophys* 24:191–218
58. Craig ME, Crothers DM, Doty P (1971) Relaxation kinetics of dimer formation by self complementary oligonucleotides. *J Mol Biol* 62:383–401
59. Crothers DM, Cole PE, Hilbers CW, Shulman RG (1974) The molecular mechanism of the thermal unfolding of *Escherichia coli* formylmethionine transfer RNA. *J Mol Biol* 87:63–88
60. Micura R, Hobartner C (2003) On secondary structure rearrangements and equilibria of small RNAs. *Chem Biochem* 4:984–990
61. Furtig B, Buck J, Manoharan V, Bermel W, Jaschke A, Wenter P, Pitsch S, Schwalbe H (2007) Time-resolved NMR studies of RNA folding. *Biopolymers* 86:360–383
62. Harlepp S, Marchal T, Robert J, Leger J, Xayaphoummine A, Isambert H, Chatenay D (2003) Probing complex RNA structures by mechanical force. *Eur Phys J E-Soft Matter* 12:605–615
63. Jean JM, Hall KB (2001) 2-Aminopurine fluorescence quenching and lifetimes: role of base stacking. *Proc Natl Acad Sci USA* 98:37–41
64. Liphardt J, Onoa B, Smith SB, Tinoco II, Bustamante C (2001) Reversible unfolding of single RNA molecules by mechanical force. *Science* 292:733–737
65. Bonnet G, Krichevsky O, Libchaber A (1998) Kinetics of conformational fluctuations in DNA hairpin-loops. *Proc Natl Acad Sci USA* 95:8602–8606
66. Ansari A, Kunznetsov SV, Shen Y (2001) Configurational diffusion down a folding funnel describes the dynamics of DNA hairpins. *Proc Natl Acad Sci USA* 98:7771–7776
67. Wallace MI, Ying L, Balasubramanian S, Klenerman D (2001) Non-arrhenius kinetics for the loop closure of a DNA hairpin. *Proc Natl Acad Sci USA* 98:5584–5589
68. Bai Y, Das R, Millett IS, Herschlag D, Doniach S (2005) Probing counterion modulated repulsion and attraction between nucleic acid duplexes in solution. *Proc Natl Acad Sci USA* 102:1035–1040
69. Chu VB, Herschlag D (2008) Unwinding RNA's secrets: advances in the biology, physics, and modeling of complex RNAs. *Curr Opin Struct Biol* 18:305–314
70. Draper DE (2008) RNA folding: thermodynamic and molecular descriptions of the roles of ions. *Biophys J* 95:5489–5495
71. Chen SJ (2008) RNA folding: conformational statistics, folding kinetics, and ion electrostatics. *Annu Rev Biophys* 37:197–214
72. Tan ZJ, Chen SJ (2011) Importance of diffuse ions binding to RNA. *Met Ions Life Sci* 9:101–124
73. Bowman JC, Lenz TK, Hud NV, Williams LD (2012) Cations in charge: magnesium ions in RNA folding and catalysis. *Curr Opin Struct Biol* 22:262–272
74. Cruz JA et al (2012) RNA-puzzles: A CASP-like evaluation of RNA three-dimensional structure prediction. *RNA* 18:610–625
75. Shapiro BA, Yingling YG, Kasprzak W, Bindewald E (2007) Bridging the gap in RNA structure prediction. *Curr Opin Struct Biol* 17:157–165
76. Hajdin CE, Ding F, Dokholyan NV, Weeks KM (2010) On the significance of an RNA tertiary structure prediction. *RNA* 16:1340–1349
77. Laing C, Schlick T (2011) Computational approaches to RNA structure prediction, analysis, and design. *Curr Opin Struct Biol* 21:1–13
78. Laing C, Schlick T (2010) Computational approaches to 3D modeling of RNA. *J Phys Condens Matter* 22:283101

79. Rother K, Rother M, Boniecki M, Puton T, Bujnicki JM (2011) RNA and protein 3D structure modeling: similarities and differences. *J Mol Model* 17:2325–2336
80. Levitt M (1969) Detailed molecular model for transfer ribonucleic acid. *Nature* 224:759–763
81. Chothia C, Gerstein M (1997) How far can sequences diverge? *Nature* 385:579–581
82. Anfinsen CB (1973) Principles that govern the folding of protein chains. *Science* 181:223–230
83. Baker D, Sali A (2001) Protein structure prediction and structural genomics. *Science* 294:93–96
84. Zhang Y, Skolnick J (2004) Automated structure prediction of weakly homologous proteins on a genomic scale. *Proc Natl Acad Sci USA* 101:7594–7599
85. Simons KT, Kooperberg C, Huang E (1997) Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and bayesian scoring functions. *J Mol Biol* 268:209–225
86. Zhang WB, Chen SJ (2002) RNA hairpin-folding kinetics. *Proc Natl Acad Sci USA* 99:1931–1936
87. Zhang WB, Chen SJ (2003) Master equation approach to finding the rate-limiting steps in biopolymer folding. *J Chem Phys* 118:3413
88. Konishi Y, Ooi T, Scheraga HA (1982) Regeneration of ribonuclease a from the reduced protein rate-limiting steps. *Biochemistry* 21:4734–4740
89. Zhang WB, Chen SJ (2003) Analyzing the biopolymer folding rates and pathways using kinetic cluster method. *J Chem Phys* 119:8716–8729
90. Flamm C, Fontana W, Hofacker IL, Schuster P (2000) RNA folding at elementary step resolution. *RNA* 6:325–338
91. Isambert H, Siggia ED (2000) Modeling RNA folding paths with pseudoknots: application to hepatitis d-virus ribozyme. *Proc Natl Acad Sci USA* 97:6515–6520
92. Danilova LV, Pervouchine DD, Favorov AV, Mironov AA (2006) RNA kinetics: a web server that models secondary structure kinetics of an elongating RNA. *J Bioinform Comp Biol* 4:589–596
93. Martinez HM (1984) An RNA folding rule. *Nucleic Acids Res* 12:323–335
94. Gillespie DT (1976) A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J Comput Phys* 22:403–434
95. Schmitz M, Steger G (1996) Description of RNA folding by simulated annealing. *J Mol Biol* 255:254–266
96. Gulyaev AP, Batenburg FH, Pleij CW (1995) The computer-simulation of RNA folding pathways using a genetic algorithm. *J Mol Biol* 250:37–51
97. Mironov A, Kister A (1985) A kinetic approach to the prediction of RNA secondary structures. *J Biomol Struct Dyn* 2:953–962
98. Mironov AA, Lebedev VF (1993) A kinetic model of RNA folding. *Biosystems* 30:49–56
99. Isambert H, Siggia ED (2000) Modeling RNA folding paths with pseudoknots: application to hepatitis delta virus ribozyme. *Proc Natl Acad Sci USA* 97:6515–6520
100. Danilova LV, Pervoud DD, Favorov AA, Mironov AA (2006) RNAKinetics: a web server that models secondary structure kinetics of an elongating RNA. *J Bioinform Comput Biol* 4:589–596
101. Ndifon W (2005) A complex adaptive systems approach to the kinetic folding of RNA. *Biosystems* 82:257–265
102. Zhao PN, Zhang WB, Chen SJ (2010) Predicting secondary structural folding kinetics for nucleic acids. *Biophys J* 98:1617–1625
103. Tang X, Thomas S, Tapia L, Giedroc DP, Amato NM (2008) Simulating RNA folding kinetics on approximated energy landscapes. *J Mol Biol* 381:1055–1067
104. Hofacker IL, Flamm C, Heine C, Wolfinger MT, Scheuermann G, Stadler PF (2010) BarMap: RNA folding on dynamic energy landscapes. *RNA* 16:1308–1316

105. Geis M, Flamm C, Wolfinger MT, Tanzer A, Hofacker IL, Middendorf M, Mandl C, Stadler PF, Thurner C (2008) Enhancement of transactivation activity of Rta of Epstein-Barr virus by RanBPM. *J Mol Biol* 379:242–261
106. Flamm C, Hofacker IL, Stadler PF, Wolfinger MT (2002) Barrier trees of degenerate landscapes. *Z Phys Chem* 216:155–173
107. Wolfinger MT, Svrcek-Seiler WA, Flamm C, Hofacker IL, Stadler PF (2004) Efficient computation of RNA folding dynamics. *J Phys A Math Gen* 37:4731–4741
108. Tang X, Kirkpatrick B, Thomas S, Song G, Amato NM (2005) Using motion planning to study RNA folding kinetics. *J Comp Biol* 12:862–881
109. Zhang WB, Chen SJ (2006) Exploring the complex folding kinetics of RNA hairpins: I. general folding kinetics analysis. *Biophys J* 90:765–777
110. Tacker M, Fontana W, Stadler PF, Schuster P (1994) Statistics of RNA melting kinetics. *Eur Biophys J* 23:29
111. Suvernev AA, Frantsuzov PA (1995) Statistical description of nucleic acid secondary structure folding. *J Biomol Struct Dyn* 13:135–144
112. Jacob C, Breton N, Daegelen P, Peccoud J (1997) Probability distribution of the chemical states of a closed system and thermodynamic law of mass action from kinetics: the RNA example. *J Chem Phys* 107:2913
113. Isambert H, Siggia ED (2000) Modeling RNA folding paths with pseudoknots: application to hepatitis delta virus ribozyme. *Proc Natl Acad Sci USA* 97:6515–6520
114. Xia TB, SantaLucia J, Turner DH (1998) Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson–Crick base pairs. *Biochemistry* 37:14719–14735
115. Mathews DH, Sabina J, Zuker M, Turner DH (1999) Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J Mol Biol* 288:911–940
116. Morgan SR, Higgs PG (1998) Barrier heights between ground states in a model of RNA secondary structure. *J Phys A Math Gen* 31:3153–3170
117. Henkin TM, Yanofsky C (2002) Regulation by transcription attenuation in bacteria: how RNA provides instructions for transcription termination/antitermination decisions. *BioEssays* 24:700–707
118. Merino E, Yanofsky C (2005) Transcription attenuation: a highly conserved regulatory strategy used by bacteria. *Trends Genet* 21:260–264
119. Franch T, Gulyaev AP, Gerder K (1997) Programmed cell death by hok/sok of plasmid R1: processing at the hok mRNA 3'-end triggers structural rearrangements that allow translation and antisense RNA binding. *J Mol Biol* 273:38–51
120. Heilman-Miller SL, Woodson SA (2003) Effect of transcription on folding of the *Tetrahymena* ribozyme. *RNA* 9:722–733
121. Brehm SL, Cech TR (1983) The fate of an intervening sequence RNA: excision and cyclization of the *Tetrahymena* ribosomal RNA intervening sequence in vivo. *Biochemistry* 22:2390–2397
122. Zhang F, Ramsay ES, Woodson SA (1995) In vivo facilitation of *Tetrahymena* group I intron splicing in *Escherichia coli* pre-ribosomal RNA. *RNA* 1:284–292
123. Treiber DK, Williamson JR (2001) Beyond kinetic traps in RNA folding. *Curr Opin Struct Biol* 11:309–314
124. Woodson SA (2002) Folding mechanisms of group I ribozymes: role of stability and contact order. *Biochem Soc Trans* 30:1166–1169
125. Zhang LB, Bao P, Michael JL, Zhang Y (2009) Slow formation of a pseudoknot structure is rate limiting in the productive co-transcriptional folding of the self-splicing *Candida* intron. *RNA* 15:1986–1992
126. Pan T, Artsimovitch I, Fang X, Landick R, Sosnick TR (1999) Folding of a large ribozyme during transcription and the effect of the elongation factor NusA. *Proc Natl Acad Sci USA* 96:9545–9550

127. Zhao PN, Zhang WB, Chen SJ (2011) Cotranscriptional folding kinetics of ribonucleic acid secondary structures. *J Chem Phys* 135:245101
128. Das R, Mills TT, Kwok LW, Maskel GS, Millett IS, Doniach S, Finkelstein KD, Herschlag D, Pollack L (2003) Counterion distribution around DNA probed by solution X-ray scattering. *Phys Rev Lett* 90:188103
129. Andresen K, Qiu X, Pabit SA, Lamb JS, Park HY, Kwok LW, Pollack L (2008) Mono- and trivalent ions around DNA: a small-angle scattering study of competition and interactions. *Biophys J* 95:287–295
130. Pabit SA, Qiu X, Lamb JS, Li L, Meisburger SP, Pollack L (2009) Both helix topology and counterion distribution contribute to the more effective charge screening in dsRNA compared with dsDNA. *Nucleic Acids Res* 37:3887–3896
131. Kirmizialtin S, Pabit SA, Meisburger SP, Pollack L, Elber R (2012) RNA and its ionic cloud: solution scattering experiments and atomically detailed simulations. *Biophys J* 102:819–828
132. Bai Y, Greenfeld M, Travers KJ, Chu VB, Lipfert J, Doniach S, Herschlag D (2007) Quantitative and comprehensive decomposition of the ion atmosphere around nucleic acids. *J Am Chem Soc* 129:14981–14988
133. Krakauer H (1971) The binding of Mg^{++} ions to polyadenylate, polyuridylylate, and their complexes. *Biopolymers* 10:2459–2490
134. Clement RM, Sturm J, Daune MP (1973) Interaction of metallic cations with DNA VI. Specific binding of Mg^{2+} and Mn^{2+} . *Biopolymers* 12:405–421
135. Soto M, Misra V, Draper DE (2007) Tertiary structure of an RNA pseudoknot is stabilized by “diffuse” Mg^{2+} ions. *Biochemistry* 46:2973–2983
136. Grilley D, Misra V, Caliskan G, Draper DE (2007) Importance of partially unfolded conformations for Mg^{2+} -induced folding of RNA tertiary structure: structural models and free energies of Mg^{2+} interactions. *Biochemistry* 46:10266–10278
137. Rialdi G, Levy J, Biltonen R (1972) Thermodynamic studies of transfer ribonucleic acids. I. Magnesium binding to yeast phenylalanine transfer ribonucleic acid. *Biochemistry* 11:2472–2479
138. Romer R, Hach R (1975) tRNA conformation and magnesium binding. A study of a yeast phenylalanine-specific tRNA by a fluorescent indicator and differential melting curves. *Eur J Biochem* 55:271–284
139. Stellwagen E, Dong Q, Stellwagen NC (2007) Quantitative analysis of monovalent counterion binding to random-sequence, double-stranded DNA using the replacement ion method. *Biochemistry* 46:2050–2058
140. Smith SB, Cui YJ, Bustamante C (1996) Overstretching B-DNA: The elastic response of individual double-stranded and single-stranded DNA molecules. *Science* 271:795–799
141. Murphy MC, Rasnik I, Cheng W, Lohman TM, Ha T (2004) Probing single-stranded DNA conformational flexibility using fluorescence spectroscopy. *Biophys J* 86:2530–2537
142. Tinland B, Pluen A, Sturm J, Weill G (1997) Persistence length of single-stranded DNA. *Macromolecules* 30:5763–5765
143. McIntosh DB, Saleh O (2011) Slat species-dependent electrostatic effects on ssDNA elasticity. *Macromolecules* 44:2328–2333
144. Sim AYL, Lipfert J, Herschlag D, Doniach S (2012) Salt dependence of the radius of gyration and flexibility of single-stranded DNA in solution probed by small-angle x-ray scattering. *Phys Rev E* 86:021901
145. Chen H, Meisburger SP, Pabit SA, Sutton JL, Webb WW, Pollack L (2012) Ionic strength-dependent persistence lengths of single-stranded RNA and DNA. *Proc Natl Acad Sci USA* 109:799–804
146. Bizarro CV, Alemany A, Ritort F (2012) Non-specific binding of Na^+ and Mg^{2+} to RNA determined by force spectroscopy methods. *Nucleic Acids Res* 40:6922–6935
147. Tan ZJ, Chen SJ (2006) Electrostatic free energy landscape for nucleic acid helix assembly. *Nucleic Acids Res* 34:6629–6639

148. Williams P, Longfellow CE, Freier SM, Kierzek R, Turner DH (1989) Laser temperature-jump, spectroscopic, and thermodynamic study of salt effects on duplex formation by dGCATGC. *Biochemistry* 28:4283–4291
149. Nakano S, Fujimoto M, Hara H, Sugimoto N (1999) Nucleic acid duplex stability: influence of base composition on cation effects. *Nucleic Acids Res* 27:2957–2965
150. Serra MJ, Baird JD, Dale T, Fey BL, Retatagos K, Westhof E (2002) Effects of magnesium ions on the stabilization of RNA oligomers of defined structures. *RNA* 8:307–323
151. Owczarzy R, Moreira BG, You Y, Behlke MA, Walder JA (2008) Predicting stability of DNA duplexes in solutions containing magnesium and monovalent cations. *Biochemistry* 47:5336–5353
152. Kuznetsov SV, Ren CC, Woodson SA, Ansari A (2008) Loop dependence of the stability and dynamics of nucleic acid hairpins. *Nucleic Acids Res* 36:1098–1112
153. Viereggs J, Cheng W, Bustamante C, Tinoco I Jr (2007) Measurement of the effect of monovalent cations on RNA hairpin stability. *J Am Chem Soc* 129:14966–14973
154. Tan ZJ, Chen SJ (2006) Nucleic acid helix stability: effects of salt concentration, cation valency and size, and chain length. *Biophys J* 90:1175–1190
155. Tan ZJ, Chen SJ (2007) RNA helix stability in mixed $\text{Na}^+/\text{Mg}^{2+}$ solution. *Biophys J* 92:3615–3632
156. Tan ZJ, Chen SJ (2008) Salt dependence of nucleic acid hairpin stability. *Biophys J* 95:738–752
157. Nixon PL, Giedroc DP (1998) Equilibrium unfolding (folding) pathway of a model H-type pseudoknotted RNA: the role of magnesium ions in stability. *Biochemistry* 37:16116–16129
158. Stellwagen E, Muse JM, Stellwagen NC (2011) Monovalent cation size and DNA conformational stability. *Biochemistry* 50:3084–3094
159. Anthony PC, Sim AY, Chu VB, Doniach S, Block SM, Herschlag D (2012) Electrostatics of nucleic acid folding under conformational constraint. *J Am Chem Soc* 134:4607–4614
160. SantaLucia JJ (1998) A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc Natl Acad Sci USA* 95:1460–1465
161. Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* 31:3406–3415
162. Chen SJ, Dill KA (2000) RNA folding energy landscapes. *Proc Natl Acad Sci USA* 97:646–651
163. SantaLucia J, Hicks D (2004) The thermodynamics of DNA structural motifs. *Annu Rev Biophys Biomol Struct* 33:415–440
164. Zhang WB, Chen SJ (2006) Exploring the complex folding kinetics of RNA hairpins: II. Effect of sequence, length, and misfolded states. *Biophys J* 90:778–787
165. Theimer A, Giedroc DP (2000) Contribution of the intercalated adenosine at the helical junction to the stability of the gag-pro frameshifting pseudoknot from mouse mammary tumor virus. *RNA* 6:409–421
166. Koculi E, Hyeon C, Thirumalai D, Woodson SA (2007) Charge density of divalent metal cations determines RNA stability. *J Am Chem Soc* 129:2676–2682
167. Takamoto K, He Q, Morris S, Chance MR, Brenowitz M (2002) Monovalent cations mediate formation of native tertiary structure of the *Tetrahymena* thermophila ribozyme. *Nature Struct Biol* 9:928–933
168. Moghaddam S, Caliskan G, Chauhan S, Hyeon C, Briber RM, Thirumalai D, Woodson SA (2009) Metal ion dependence of cooperative collapse transitions in RNA. *J Mol Biol* 393:753–764
169. Heilman-Miller SL, Thirumalai D, Woodson SA (2001) Role of counterion condensation in folding of the *Tetrahymena* ribozyme. I. Equilibrium stabilization by cations. *J Mol Biol* 306:1157–1166
170. Lambert D, Leipply D, Shiman R, Draper DE (2009) The influence of monovalent cation size on the stability of RNA tertiary structures. *J Mol Biol* 390:791–804

171. Walter NG, Burke JM, Millar DP (1999) Stability of hairpin ribozyme tertiary structure is governed by the interdomain junction. *Nature Struct Biol* 6:544–549
172. Pljevaljcic G, Millar DP, Deniz AA (2004) Freely diffusing single hairpin ribozymes provide insights into the role of secondary structure and partially folded states in RNA folding. *Biophys J* 87:457–467
173. Lepply D, Draper DE (2011) Evidence for a thermodynamically distinct Mg^{2+} ion associated with formation of an RNA tertiary structure. *J Am Chem Soc* 133:13397–13405
174. Weixlbaumer A, Werner A, Flamm C, Westhof E, Schroeder R (2004) Determination of thermodynamic parameters for HIV DIS type loop-loop kissing complexes. *Nucleic Acids Res* 32:5126–5133
175. Lorenz C, Piganeau N, Schroeder R (2006) Stabilities of HIV-1 DIS type RNA loop-loop interactions in vitro and in vivo. *Nucleic Acids Res* 34:334–342
176. Vander Meulen KA, Butcher SE (2012) Characterization of the kinetic and thermodynamic landscape of RNA folding using a novel application of isothermal titration calorimetry. *Nucleic Acids Res* 40:2140–2151
177. Tan ZJ, Chen SJ (2010) Predicting ion binding properties for RNA tertiary structures. *Biophys J* 99:1565–1576
178. Tan ZJ, Chen SJ (2011) Salt contribution to RNA tertiary structure folding stability. *Biophys J* 101:176–187
179. Tan ZJ, Chen SJ (2012) Ion-mediated RNA structural collapse: effect of spatial confinement. *Biophys J* 103:827–836
180. Rau DC, Parsegian VA (1992) Direct measurement of the intermolecular forces between counterion-condensed DNA double helices. Evidence for long range attractive hydration forces. *Biophys J* 61:246–259
181. Rau DC, Parsegian VA (1992) Direct measurement of temperature-dependent solvation forces between DNA double helices. *Biophys J* 61:260–271
182. Bai Y, Chu VB, Lipfert J, Pande VS, Herschlag D, Doniach S (2008) Critical assessment of nucleic acid electrostatics via experimental and computational investigation of an unfolded state ensemble. *J Am Chem Soc* 130:12334–12341
183. Qiu X, Andresen K, Kwok LW, Lamb JS, Park HY, Pollack L (2007) Inter-DNA attraction mediated by divalent counterions. *Phys Rev Lett* 99:038104
184. Qiu X, Parsegian VA, Rau DC (2010) Divalent counterion-induced condensation of triple-strand DNA. *Proc Natl Acad Sci USA* 107:21482–21486
185. Li L, Pabit SA, Meisburger SP, Pollack L (2011) Double-stranded RNA resists condensation. *Phys Rev Lett* 106:108101
186. Tan ZJ, Chen SJ (2009) Predicting electrostatic force in RNA folding. *Methods Enzymol* 469:465–487
187. Manning GS (1978) The molecular theory of polyelectrolyte solutions with applications to the electrostatic properties of polynucleotides. *Q Rev Biophys* 11:179–246
188. Schurr MJ (2009) Polyanion models of nucleic acid-metal ion interactions. In: Hud NV (ed) *Nucleic acid-metal ion interactions*. Royal Society of Chemistry, London, pp 307–344
189. Ray J, Manning GS (2000) Formation of loose clusters in polyelectrolyte solutions. *Macromolecules* 33:2901–2908
190. Lyubartsev P, Nordenskiöld L (1995) Monte Carlo simulation study of ion distribution and osmotic pressure in hexagonally oriented DNA. *J Phys Chem* 99:10373–10382
191. Dai L, Mu Y, Nordenskiöld L, van der Maarel JR (2008) Molecular dynamics simulation of multivalent-ion mediated attraction between DNA molecules. *Phys Rev Lett* 100:118301
192. Gilson MK, Sharp KA, Honig B (1987) Calculating the electrostatic potential of molecules in solution: method and error assessment. *J Comput Chem* 9:327–335
193. Boschitsch H, Fenley MO (2007) A new outer boundary formulation and energy corrections for the nonlinear Poisson-Boltzmann equation. *J Comput Chem* 28:909–921
194. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA (2000) Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci USA* 98:10037–10041

195. Zhou YC, Feig M, Wei GW (2008) Highly accurate biomolecular electrostatics in continuum dielectric environments. *J Comput Chem* 29:87–97
196. Lu B, Cheng X, Huang J, McCammon JA (2010) AFMPB: an adaptive fast multipole Poisson-Boltzmann solver for calculating electrostatics in biomolecular systems. *Comput Phys Commun* 181:1150–1160
197. Misra VK, Shiman R, Draper DE (2003) A thermodynamic framework for the magnesium-dependent folding of RNA. *Biopolymers* 69:118–136
198. Tan ZJ, Chen SJ (2005) Electrostatic correlations and fluctuations for ion binding to a finite length polyelectrolyte. *J Chem Phys* 122:044903
199. Chu VB, Bai Y, Lipfert J, Herschlag D, Doniach S (2007) Evaluation of ion binding to DNA duplexes using a size-modified Poisson-Boltzmann theory. *Biophys J* 93:3202–3209
200. Kirmizialtin S, Silalahi AR, Elber R, Fenley MO (2012) The ionic atmosphere around A-RNA: Poisson-Boltzmann and molecular dynamics simulations. *Biophys J* 102:829–838
201. Gavryushov S (2008) Electrostatics of B-DNA in NaCl and CaCl₂ solutions: ion size, interionic correlation, and solvent dielectric saturation effects. *J Phys Chem B* 112:8955–8965
202. Grochowski P, Trylska J (2008) Continuum molecular electrostatics, salt effects and counterion binding. A review of the Poisson-Boltzmann theory and its modifications. *Biopolymers* 89:93–113
203. Forsman J (2004) A simple correlation-corrected Poisson-Boltzmann theory. *J Phys Chem B* 108:9236–9245
204. Vlachy V (1999) Ionic effect beyond Poisson-Boltzmann theory. *Annu Rev Phys Chem* 50:145–165
205. Wang K, Yu YX, Gao GH (2008) Density functional study on the structural and thermodynamic properties of aqueous DNA-electrolyte solution in the framework of cell model. *J Chem Phys* 128:185101
206. Chen YG, Weeks JD (2006) Local molecular field theory for effective attractions between like charged objects in systems with strong Coulomb interactions. *Proc Natl Acad Sci USA* 103:7560–7565
207. Tan ZJ, Chen SJ (2006) Ion-mediated nucleic acid helix-helix interactions. *Biophys J* 91:518–536
208. Tan ZJ, Chen SJ (2008) Electrostatic free energy landscapes for DNA helix bending. *Biophys J* 94:3137–3149
209. Chen G, Tan ZJ, Chen SJ (2010) Salt-dependent folding energy landscape of RNA three-way junction. *Biophys J* 98:111–120
210. Chen G, Chen SJ (2011) Quantitative analysis of the ion-dependent folding stability of DNA triplexes. *Phys Biol* 8:066006
211. He Z, Chen SJ (2012) Predicting ion-nucleic acid interactions by energy landscape-guided sampling. *J Chem Theo Comp* 8:2095–2102
212. Cate JH, Gooding AR, Podell E, Zhou K, Golden BL, Kundrot CE, Cech TR, Doudna JA (1996) Crystal structure of a group I ribozyme domain: principles of RNA packing. *Science* 273:1678–1685
213. Zhou HX, Rivas G, Minton AP (2008) Macromolecular crowding and confinement: biochemical, biophysical, and potential physiological consequences. *Annu Rev Biophys* 37:375–397
214. Lambert D, Leipply D, Draper DE (2010) The osmolyte TMAO stabilizes native RNA tertiary structures in the absence of Mg²⁺: evidence for a large barrier to folding from phosphate dehydration. *J Mol Biol* 404:138–157
215. Pincus DL, Hyeon C, Thirumalai D (2008) Effects of trimethylamine N-oxide (TMAO) and crowding agents on the stability of RNA hairpins. *J Am Chem Soc* 130:7364–7372
216. Kilburn D, Roh JH, Guo L, Briber RM, Woodson SA (2010) Molecular crowding stabilizes folded RNA structure by the excluded volume effect. *J Am Chem Soc* 132:8690–8696