A new entropy model for RNA: part II, persistence-related entropic contributions to RNA secondary structure free energy calculation

Appendix: Renormalization theory as used in this work

Authors
Wayne Dawson[1*], Kenji Yamamoto[2], Kentaro Shimizu[1] and Gota Kawai[3]

Institutions
[1] Bioinformation Engineering Laboratory, Department of Biotechnology, Graduate School of Agriculture and Life Sciences, The University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113-8657, Japan

[2] Research Institute, National Center for Global Health and Medicine, 1-21-1 Toyama, Shinjuku-ku, Tokyo 162-8655 Japan

[3] Chiba Institute of Technology, 2-17-1 Tsudanuma, Narashino-shi, Chiba 275-0016 Japan.

*Current affiliation and address
Dept of Comp Bio, Fac Frontier Sci, University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa-shi, Chiba-ken 277-8561, Japan
CBRC, AIST, Tokyo Waterfront Bio-IT Research Bld, 2-4-7 Aomi, Koto-ku, Tokyo 135-0064, Japan

Corresponding Author
Correspondence should be addressed to:
Wayne Dawson: wayne-dawson@aist.go.jp

In general, renormalization is a method that is used to simplify the computation of a very large system with a vast number of interactions into a more manageable problem. Presumably, if the system is rather repetitive (like a crystal lattice) and the coupling between diverse parts of the system tapers off as a function of distance, we can search for a way to group (or effectively cluster) these local interactions in a way that encompasses the overall behavior, and then look at how these new groups couple over a much larger distance. One might see it as looking at nearest neighbor (NN) interactions, grouping them, then looking at next nearest neighbor (NNN) interactions, etc. If the coupling on the NNN scale is smaller than NN, there is a chance that this will converge. Therefore, it is a process of constructing coarse-grained structure from the fine details while presumably preserving the pertinent information of interest. A polymer chain is essentially a 1D system (at least along the chain); therefore, we will focus on explaining the renormalization process for a 1D system.

Ma [1] writes a fairly clear introduction to renormalization group theory in condensed matter physics. Rather dense studies can be found in the works of Wilson and Kagut [2,3]. Kadanoff first introduced these ideas into a condensed matter framework [4]. A readable report on renormalization and scaling theory in polymer physics is almost surely the one by McKenzie [5]. A good introduction can be found in deGennes famous work [6] and Fernandez has applied this kind of idea to RNA folding [7,8].

In renormalization, we would start with a proposed free energy (FE) function that is calculated at some initial coarse-grained length scale for our system of interest, say $\Delta G(\xi_o, \{\mathbf{x}_o\}, \{\mathbf{k}_o\})$, where ξ_o is the length scale of the polymer, $\{\mathbf{x}_o\}$ would be a set of locations on the polymer chain, and $\{\mathbf{k}_o\}$ would describe some set of coupling

constants. Then we would rescale these interactions of length scale ξ_o to a larger length scale ξ, for example $\xi = 2\xi_o$ and derive a new free energy function $\Delta G(\xi, \{\mathbf{x}\}, \{\mathbf{k}\})$ that still resembles the original system in FE but is now represented at a more coarse-grained level of approximation. The renormalization method requires that we account for the free energy within the rescaled part as a constant, and rescale the free energy associated with the long range coupling

$$\Delta G(\xi_o, \{\mathbf{x}_o\}, \{\mathbf{k}_o\}) \rightarrow \Delta G(\xi_o, \xi) + \frac{1}{\xi} \; G(\xi, \{\mathbf{x}\}, \{\mathbf{k}\}) \qquad \text{(A1)}$$

where ξ_o and ξ reflect the initial and final Kuhn length, $\Delta G(\xi_o, \xi)$ is a constant reflecting the local coupling between neighboring monomers on the chain, $\Delta G(\xi, \{\mathbf{x}\}, \{\mathbf{k}\})$ is the new function that resembles the original function $\Delta G(\xi_o, \{\mathbf{x}_o\}, \{\mathbf{k}_o\})$ and retains the long range coupling interactions but are now thinned out over length scales ξ. The function $\Delta G(\xi, \{\mathbf{x}\}, \{\mathbf{k}\})$ is weighted by $1/\xi$ because in the process of rescaling the long range coupling over a distance ξ, the influence of that coupling is reduced. In a sense, the initial NN interactions are subsumed into $\Delta G(\xi_o, \xi)$ and the quantity of interest now becomes the NNN interactions, which are presumably smaller. This is the process that builds *effective mers* in this model.

As an illustrative example, Fig A1 shows a polymer of length 15 mers and an initial coarse-grained scale $\xi_o = 1$ [mer]. The long range coupling in this polymer occurs between the following pair of mers (2,13) and (3,12) and is illustrated by the dotted red lines. This structure is then rescaled to $\xi = 2$ [mer], the blue ovals in Fig A1. Because of the way the sequence was divided up, the coupling is shared between two effective mers, illustrated by the green dashed lines. Let the free energy of $\xi_o = 1$ and

$\xi = 2$ be represented by ΔG_1 and ΔG_2 respectively. Then Fig A1 would be expressed

$$\Delta G_1 = 15\Delta G(\xi = 1) + \Delta G(\mathbf{x}_{13} - \mathbf{x}_2) + \Delta G(\mathbf{x}_{12} - \mathbf{x}_3) \tag{A2a}$$

where $\Delta G(\xi = 1)$ contains whatever internal and NN interactions must be accounted for on length scale $\xi_o = 1$ and $\Delta G(\mathbf{x}_j - \mathbf{x}_i)$ reflects the long range coupling due to base pairing interactions (in the case of RNA and DNA). Upon renormalization,

$$\Delta G_2 = \frac{15}{2}\Delta G(\xi = 2) + \frac{1}{2}\Delta G(\mathbf{x}_{\tilde{7}} - \mathbf{x}_{\tilde{1}}) + \frac{1}{2}\Delta G(\mathbf{x}_{\tilde{8}} - \mathbf{x}_{\tilde{2}}) \tag{A2b}$$

where $\Delta G(\xi = 2)$ is a new constant that subsumes interactions on length scale $\xi_o = 1$ [mer], the tilde notation is used to index the rescaled effective mers and $\Delta G(\mathbf{x}_{\tilde{j}} - \mathbf{x}_{\tilde{i}})$ is the rescale FE function using these effective mers. Since we expect that the total FE is left unchanged by our general rescaling, it follows that

$$\Delta G_1 \approx \Delta G_2. \tag{A3}$$

The coupling between two independent pairs of mers is now subsumed into a single pair of effective mers. The weight of the coupling is cut in half. However, because the coupling is distributed over two links in Fig A1 and rescaled accordingly, the discernible average still resembles one bond between two effective mers comprising the original mers 2,3 and mers 12,13. It is, in essence, an effective base pair. It is also expected from Eqn (A3) that there are local coupling effects contained in $\Delta G(\xi = 2)$

that compensate for the change in the magnitude of the long range coupling. Between ΔG_1 and ΔG_2, the overall FE is largely balanced to yield the same FE.

In a second illustrative example, in Fig A2, a 20 mer sequence is renormalized from $\xi_o = 1$ [mer] to $\xi = 3$ [mer]. The other notation is the same. The polymer is expressed as

$$\Delta G_1 = 20\Delta G(\xi = 1) + \Delta G(\mathbf{x}_{16} - \mathbf{x}_5) + \Delta G(\mathbf{x}_{15} - \mathbf{x}_6) + \Delta G(\mathbf{x}_{14} - \mathbf{x}_7) \tag{A4a}$$

and upon renormalization,

$$\Delta G_3 = \frac{20}{3}\Delta G(\xi = 3) + \frac{1}{3}\Delta G(\mathbf{x}_{\%} - \mathbf{x}_{\%}) + \frac{1}{3}\Delta G(\mathbf{x}_{\%} - \mathbf{x}_{\%}) + \frac{1}{3}\Delta G(\mathbf{x}_{\%} - \mathbf{x}_{\%}) \tag{A4b}$$

Again, since we expect the FE to remain constant in the rescaling process, it follows that $\Delta G_1 \approx \Delta G_3$. Now the coupling is spread over 3 mers with a corresponding reduction in the FE. Nevertheless, the overall interaction and FE remain constant due to compensation from the constant term $\Delta G(\xi = 3)$. The 3 base pairs become one effective base pair.

A consequence of the long range coupling in Fig A1 and A2 is that the 1D Ising spin model, which assumes only strong coupling between NN, is not an adequate approximation of the strongly long range coupling diagrammed in Figs A1 and A2. In these diagrams, the long range coupling exists whether we incorporate it into the dinucleotide base pair (bp) interaction or not.

Moreover, the folding is path independent: in essence, which bond forms first is not important in the statistical mechanics. The only means of removing the

independence of this coupling is by subsuming it in the renormalization length.

To illustrate the impact of path independence on the folding of the polymer, consider the folding diagram in Fig A3. Let stem A and stem B represent unique and mutually exclusive interactions, i.e., there can be no mixing of any part of the polymer chain comprising stem A with that of stem B. (This is admittedly unrealistically idea, but this is just for the purpose of illustration and more realistic examples just complicate the math without changing the general observations.)

In principle, all structures are in thermodynamic equilibrium, and there is a small probability at some time in sampling data of the polymer's structure, it will be found in the unfolded state (U) on the left or one of the two folding intermediate states (I_1 and I_2). It also describes the denaturing/refolding process of a polymer. Due to the path independence, the folding through I_1 (forming stem A first) is

$$p(A)p(B) = p(B \mid A)p(A), \text{ stem A forms first} \qquad \text{(A5a)}$$

likewise, the folding through I_2 (forming stem B first) is

$$p(A)p(B) = p(A \mid B)p(B), \text{ stem B forms first} \qquad \text{(A5b)}$$

where by definition, the path independence requires that the total probability for the folded structure is $p(A)p(B)$. The consequence of Eqns (A5ab) is $p(A) = p(A \mid B)$ and $p(B) = p(B \mid A)$. All models that calculate the structure of RNA from dinucleotide base pair interactions use additive terms in the free energy, indicating that they are assumed to be based on path independent interactions. The mere assumption of that these thermodynamic parameters can be added implicitly implies the assumption that these interactions are in fact *not* conditional probabilities in any of these models.

When considering base pairing, the spreading out of the interaction over an effective mer allows us to think about the folding event involving the formation of stem A or stem B in Fig A3. This effectively transforms the response of individual bps in a stem into a collective dependent interaction. Nevertheless, it remains true for bps in the stem that, regardless of which bp forms first, it is still a path independent process. The folding takes on a cooperative character when we include the Kuhn length in the evaluation of RNA folding. Hence, due to the renormalization approach that is used, the CLE model incorporates cooperativity as a natural consequence of the renormalization process in the case of folded RNA structures. This cooperativity can be seen in the unzipping of RNA in force-extension experiments [9,10] and in melting experiments where different stems melt at different temperatures [11-15].

In the case of double strand DNA (dsDNA) and dsRNA, there is no significant long-range coupling, Fig A4. In Fig A4, the coupling (red dashed lines) happens over independent chains and is highly local. On the other hand, the coupling described in Figs A1 and A2 is quite long range. Based on the diagram in Fig A4, it is clear that there are no long-range interactions except between the two RNA chains. Therefore, the main interactions will all be local in nature.

The long-range coupling, shown in Figs A1 and A2, results from an interaction force that pushes the polymer chain apart when two points on the polymer chain i and j are squeezed closer than their equilibrium position or stretched beyond this position. In the non-interacting state, the root-mean-squared distance between mers i and j ($r_{rms.ij}$) is a function of $N_{ij} = j - i + 1$ and thus $r_{rms.ij} = (\xi N_{ij})^{1/2} b$, where ξ is the Kuhn length and b is the distance between consecutive monomers in the RNA chain.

The state equation for a Gaussian polymer chain (GPC) is

$$f_{\text{int}}(r_{ij}) = k_{\text{B}}T\left(\frac{1}{r_{ij}} - \frac{3}{\xi N_{ij} b^2} r_{ij}\right) \tag{A6}$$

Hence, if $r_{ij}^2 > \xi N_{ij} b^2 / 3$, the force between mers i and j is repulsive, and if $r_{ij}^2 > \xi N_{ij} b^2 / 3$, the force is attractive. From the relationship $f_{\text{int}}(r) = T(\partial S / \partial r)_T$, we can derive the entropy of the polymer, particularly since, in the expression, the entropy is independent of T and only depends on N_{ij}, ξ and r_{ij}. From here, we can generalize the concepts to what has been previously published [16-19].

There are several points that need to be discussed.

First, traditional renormalization theory in condensed matter physics and polymers is applied to infinite systems of homogeneous objects (e.g., atoms in a unit cell or identical monomers). Here, although the monomers are of similar size and chemical behavior, they are heterogeneous.

Second, our rescaling approach is applied to a system according to its observed (or at least observable) regularity. For example, a loose, highly flexibile RNA structure (e.g., random interactions in a polyA sequence) would have a small Kuhn length (ξ) whereas a large structure of RNA with extensive scaffolding (e.g., ribosomal RNA) would tend to have a large ξ. Traditionally, the approach has been directed to scale identical spins progressively in 2,4,8 etc groups together (in an Ising spin model). Hence, traditional renormalization has been used as a mathematical tool to approximate the ground state of the system. We also use renormalization as a tool, but ξ has a

genuine meaning as the flexibility of the RNA. Therefore, there should be a minimum in the FE where ξ is optimized in the model. It is possible to apply additional renormalization steps (as with the Ising model) to higher levels of structure. For example, calculating local 3D structure of stem-loops and then considering the coupling of the stem-stem scale interactions. Nevertheless, the simple 2^n rescaling process is not possible to maintain.

Third, the application of renormalization is to find ξ in this method, not to search for a critical temperature or critical parameters per se. Renormalization methods have been used in polymers, however, not in quite so applied a manner. In short, the model is necessarily less rigorous and is used in an applied fashion to real problems in RNA structure. Therefore, it's not on the same ground of rigor.

Furthermore, Kuhn length (or sometimes Kuhn statistical length) and persistence length are generally derived quantities measured from macroscopically observed quantities like the mean square end-to-end distance $\langle r^2 \rangle$. Hence, it is usually cited as an average value. Since we are using ξ to reflect real RNA structure – an inherently regional, heterogeneous and strongly coupled system rather than a long, uniform, homogeneous and virtually non-interacting system – it is more germane to define ξ in terms of the true regional flexibility of a given RNA structure.

Since the model retains the main features of real space renormalization theory, this should not pose a significant issue. Moreover, by using this approach in an applied manner, the true power and utility of this technique emerges, both as a mathematical tool, and as a physical concept. Hence, we argue that by the results of the approach, it shows that renormalization is more than just a mere tool. It is a fundamental property of at least some physical systems and one that can manifest itself at far more mundane

theoretical levels than the developers may have expected.

Figures

Figure A1

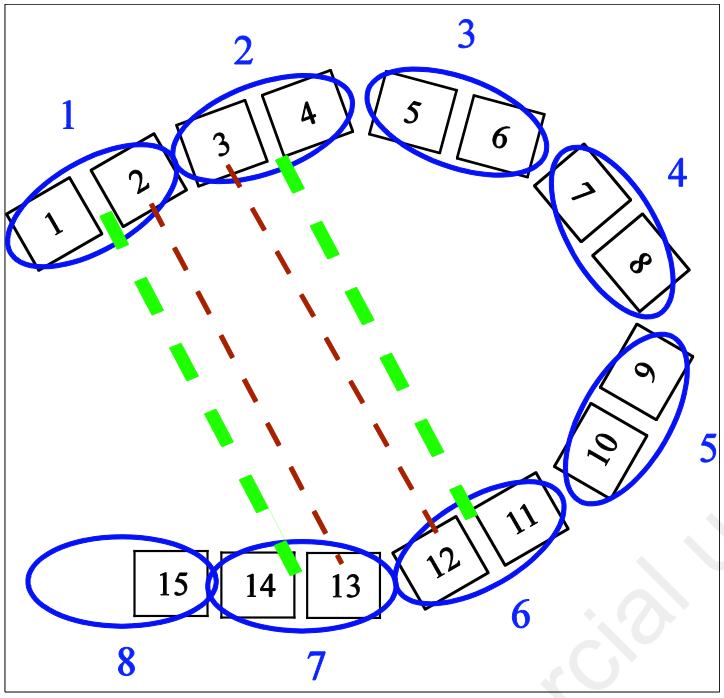

Figure A2

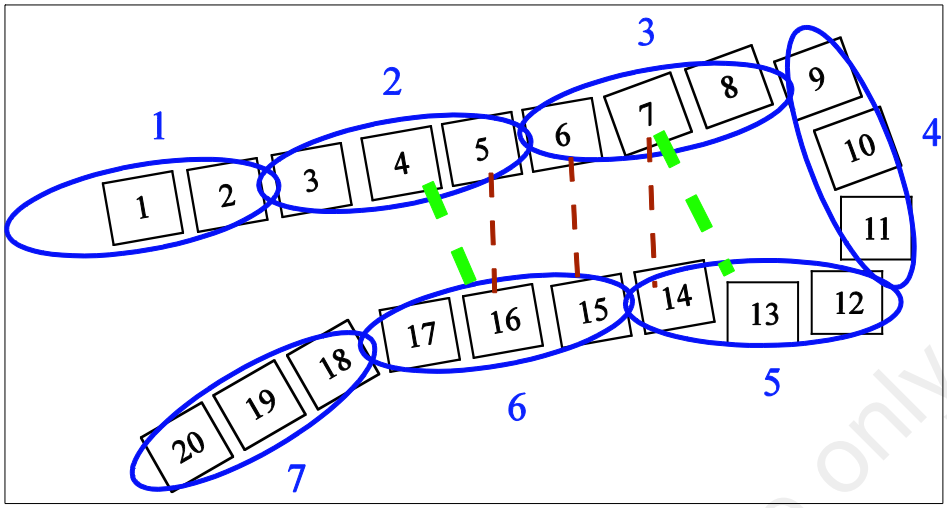

Figure A3

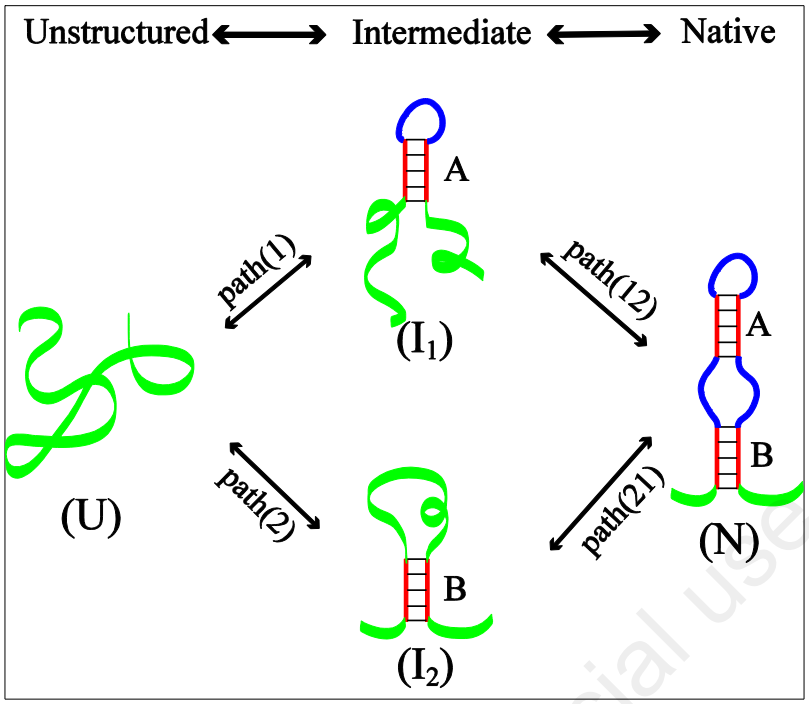

Figure A4

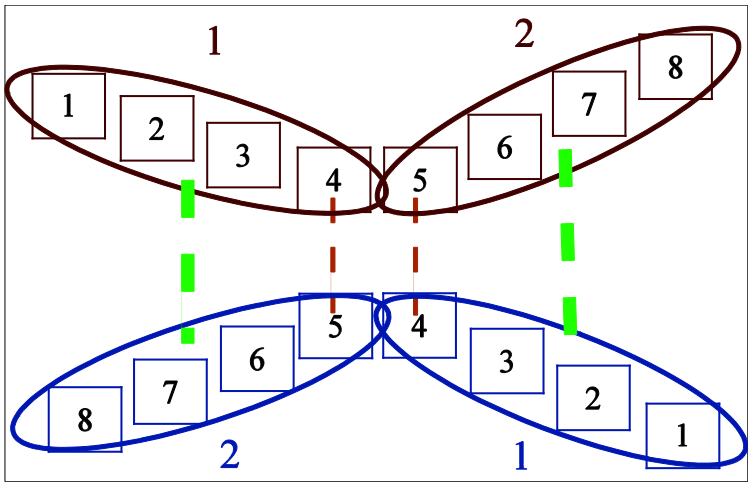

Figure Captions

Figure A1

Figure A1. Illustrative example of renormalization of a polymer chain from a single mer of length $\xi_o = 1$ mer to an effective mer of length $\xi = 2$ mers. The polymer chain also is folded such that there are strong intra-chain coupling interactions indicated by the red dashed lines (for $\xi_o = 1$) and the green dashed lines (for $\xi = 2$). The effective mers are labeled with blue numbers.

Figure A2

Figure A2. Illustrative example of renormalization of a polymer chain from a single mer of length $\xi_o = 1$ mer to an effective mer of length $\xi = 3$ mers. The polymer chain also is folded such that there are strong intra-chain coupling interactions indicated by the red dashed lines (for $\xi_o = 1$) and the green dashed lines (for $\xi = 3$). The effective mers are labeled with blue numbers.

Figure A3

Figure A3. An illustration of path independence for folding of the polymer in which stem A and stem B represent unique and mutually exclusive interactions, i.e., there can be no mixing of any part of the polymer chain comprising stem A with that of stem B.

Figure A4

Figure A4. Illustrative example of renormalization from a single mer of length $\xi_o = 1$ mer to an effective mer of length $\xi = 4$ mers for two independent polymer chains, where the interaction forms a structure like dsRNA. Unlike Figs A1 and A2, the polymer chain does not have strong intra-chain coupling, only inter-chain coupling. The coupling is indicated by the red dashed lines (for $\xi_o = 1$) and the green dashed lines (for $\xi = 4$). The effective mers are labeled with blue numbers.

References

1. Ma S-K. Introduction to Renormalization Group. Reviews of Modern Physics 1973;45: 589-614.
2. Wilson KG. Renormalization group and critical phenomena. I. Renormalization group and Kadanoff scaling picture. Phys Rev B 1971;4:3174-3205.
3. Wilson KG, Kogut J. The renormalization group and the epsilon expansion. Phys Repts C 1974;12:75-200.
4. Kadanoff LP, Gotze W, Hamblen D, Hecht R, Lewis EAS, et al. Rev Mod Phys 1967;39:395-431.
5. McKenzie DS. Polymers and scaling. Physics Reports 1976;27C:35-88.
6. de Gennes P-G. Scaling Concepts in Polymer Physics. Ithaca: Cornell University Press; 1979, 324 p.
7. Fernandez A. Folding a protein by discretizing its backbone torsional dynamics. Phys Rev E Stat Phys Plasmas Fluids Relat Interdiscip Topics 1999;59:5928-5939.
8. Fernandez A, Salthu R, Cendra H. Discretized torsional dynamics and the folding of an RNA chain. Phys Rev E Stat Phys Plasmas Fluids Relat Interdiscip Topics 1999;60:2105-2119.
9. Liphardt J, Onoa B, Smith SB, Tinoco I, Jr., Bustamante C. Reversible unfolding of single RNA molecules by mechanical force. Science 2001;292:733-737.
10. Li PT, Bustamante C, Tinoco I, Jr. Real-time control of the energy landscape by force directs the folding of RNA molecules. Proc Natl Acad Sci U S A 2007;104:7039-7044.
11. Hinz HJ, Filimonov VV, Privalov PL. Calorimetric studies on melting of tRNA Phe (yeast). Eur J Biochem 1977;72:79-86.
12. Privalov PL, Filimonov VV. Thermodynamic analysis of transfer RNA unfolding. J Mol Biol 1978;122:447-464.
13. Gluick TC, Draper DE. Thermodynamics of folding a pseudoknotted mRNA fragment. J Mol Biol (1994;241:246-262.
14. Laing LG, Draper DE. Thermodynamics of RNA folding in a conserved ribosomal RNA domain. J Mol Biol (1994;237:560-576.
15. Laing LG, Gluick TC, Draper DE. Stabilization of RNA structure by Mg ions. Specific and non-specific effects. J Mol Biol 1994;237:577-587.
16. Dawson W, Fujiwara K, Kawai G. Prediction of RNA pseudoknots using heuristic modeling with mapping and sequential folding. PLoS One (2007;2:905.
17. Dawson W, Kawai G. Modeling the Chain Entropy of Biopolymers: Unifying Two Different Random Walk Models under One Framework. J Comput Sci Syst Biol

2009;2:001-023.

18. Dawson W, Suzuki K, Yamamoto K. A physical origin for functional domain structure in nucleic acids as evidenced by cross linking entropy: part I. J Theor Biol 2001;213:359-386.

19. Dawson W, Suzuki K, Yamamoto K. A physical origin for functional domain structure in nucleic acids as evidenced by cross linking entropy: part II. J Theor Biol (2001;213:387-412.