doi:10.3969/j.issn.1672-5565.2014.03.08

用于寡核苷酸二级结构预测的热力学数据库研究进展

刘哲言,屈武斌*,张成岗*
(军事医学科学院放射与辐射医学研究所,蛋白质组学国家重点实验室,
全军军事认知与心理卫生研究中心,北京 100850)

摘　要:基于核酸分子杂交的生物技术(如PCR)在病原微生物检测、临床诊断等诸多领域中应用广泛,此类技术的可靠性在于寡核苷酸分子与其靶点结合的高稳定性与特异性,而精确预测寡核苷酸与靶分子结合的二级结构是分析其稳定性与特异性的关键。其中,基于热力学的最近邻模型是寡核苷酸二级结构预测最为可靠的计算方法,但其精确性强烈依赖于精确的热力学参数。由于寡核苷酸分子二级结构的复杂性,除了完美匹配外,还需要错配、内环、膨胀环、末端摇摆、CNG重复、GU摆动等特殊结构的热力学数据。本文综述了近年来用于寡核苷酸二级结构预测的有效热力学数据库及相关计算方法,并指出当前热力学数据库的局限及未来发展方向。

关键词:寡核苷酸二级结构;热力学数据库;热力学计算
中图分类号:Q522　　**文献标志码**:A　　**文章编号**:1672-5565(2014)-03-196-10

Research progress of the thermodynamic database for oligonucleotide secondary structure prediction

LIU Zheyan, QU Wubin*, ZHANG Chenggang*
(*Beijing Institute of Radiation Medicine*, *State Key Laboratory of Proteomics*,
Cognitive and Mental Health Research Center, *Beijing* 100850, *China*)

Abstract: The nucleotide hybridization based molecular biological technologies like PCR have been widely used in many fields, such as pathogenic microorganism detection, clinical diagnosis. And the accurate prediction of secondary structures between oligonucleotide and its binding sites is the key to these technologies. The Nearest-Neighbor Model based on thermodynamics is the most accurate method to predict oligonucleotide secondary structure, and the precision mainly depends on the thermodynamic parameters. Meanwhile, the diversity of secondary structure requires different thermodynamic parameters for different motifs, including perfect matches, mismatches, internal loops, bulge loops, dangling ends, CNG repeats, and GU wobble base pairs. Therefore, this review summarized the current parameter sets available for oligonucleotide secondary structure prediction. We also pointed out the limitations and future development directions of the thermodynamic database.

Keywords: Oligonucleotide secondary structure; Thermodynamic database; Thermodynamic calculation

近年来,以核酸分子杂交为基础的生物技术如聚合酶链反应、DNA印迹、RNA印迹、芯片杂交等在病原微生物检测、临床诊断中应用广泛,其可靠性依赖于寡核苷酸分子与其靶点结合的高稳定性与特异性,而分析这种结合特性的关键在于寡核苷酸与靶分子结合的二级结构的精确预测,否则会导致假阴性或假阳性的检测结果[1-4]。

已有研究显示最近邻模型(Nearest-Neighbor

收稿日期:2014-05-13;**修回日期**:2014-05-20.
基金项目:国家自然科学基金课题(30900862, 30973107, 81070741, 81172770, 31371345);国家重点基础研究发展计划(973计划)课题(2012CB518200);蛋白质组学国家重点实验室自主研究课题(SKLP-O201104, SKLP-K201004, SKLP-O201002)资助。
作者简介:刘哲言,女,硕士研究生,研究方向:生物信息学;E-mail: lzy_zenger@163.com.
*通信作者:屈武斌,男,助理研究员,研究方向:生物信息学;E-mail: quwubin@gmail.com.
张成岗,男,研究员,博士生导师,研究方向:生物信息学;E-mail: zhangcg@bmi.ac.cn.

Model，简称 NN model）是预测寡核苷酸二级结构最可靠的热力学计算方法[5]，该模型指出一个给定碱基对的稳定性依赖于其临近碱基对的稳定性。其基本思想是将核酸分子结合过程中的标准焓变和熵变计算转化为由 A、T、G、C 所形成的 10 个完美匹配二聚体以及非完美匹配结构的标准焓变和熵变的累加和，再加上起始和结束单独匹配碱基对 GC 或 AT 以及序列对称性等因素的影响[6-7]。然而，由于寡核苷酸单分子自身折叠或者双分子杂交所形成的二级结构具有多样性与复杂性，除完美匹配外，还包含单独错配、连续错配、内环、膨胀环、末端摇摆、CNG 重复、GU 摆动等多种模式。因此，寡核苷酸二级结构热力学计算的精确性还需要依赖上述多种结构的热力学参数。

鉴于目前的热力学数据分散于不同的文献中[8-11]，且不同来源数据可靠性不一，因此本文根据学术界对不同来源数据的认可程度，系统性的综述了近年来广泛用于寡核苷酸二级结构预测的热力学参数及相关计算，并指出当前数据库的局限及未来发展方向，从而为相关人员进行研究提供整合数据库资源，促进寡核苷酸二级结构的精确预测。

1　DNA 与 DNA 相互作用的热力学参数

1.1　完美匹配

针对 DNA/DNA 相互作用双链中 Watson-Crick 的 AT、GC 配对，SantaLucia 于 1998 年提出的在 1mol/L NaCl 环境下的寡核苷酸最近邻热力学参数是与实验数据误差最小的参数表[8]，受到广泛应用（见表 1）。其中 ΔH、ΔS 与 ΔG 分别表示标准焓变、标准熵变和标准自由能，可直接从热力学参数表获取。完美匹配结构的热力学计算如下：

$$\Delta H\left(\begin{smallmatrix}\mathrm{AGCGA}\\ \mathrm{TCGCT}\end{smallmatrix}\right) = \Delta H\left(\begin{smallmatrix}\mathrm{AG}\\ \mathrm{TC}\end{smallmatrix}\right) + \Delta H\left(\begin{smallmatrix}\mathrm{GC}\\ \mathrm{CG}\end{smallmatrix}\right) + \Delta H\left(\begin{smallmatrix}\mathrm{CG}\\ \mathrm{GC}\end{smallmatrix}\right) + \Delta H\left(\begin{smallmatrix}\mathrm{GA}\\ \mathrm{CT}\end{smallmatrix}\right);$$
$$\Delta S\left(\begin{smallmatrix}\mathrm{AGCGA}\\ \mathrm{TCGCT}\end{smallmatrix}\right) = \Delta S\left(\begin{smallmatrix}\mathrm{AG}\\ \mathrm{TC}\end{smallmatrix}\right) + \Delta S\left(\begin{smallmatrix}\mathrm{GC}\\ \mathrm{CG}\end{smallmatrix}\right) + \Delta S\left(\begin{smallmatrix}\mathrm{CG}\\ \mathrm{GC}\end{smallmatrix}\right) + \Delta S\left(\begin{smallmatrix}\mathrm{GA}\\ \mathrm{CT}\end{smallmatrix}\right)。 \quad (1)$$

表 1　DNA/DNA 结合完美匹配最近邻热力学参数（1 mol/L，37 ℃）

Table 1　Nearest-neighbor thermodynamic parameters for DNA/DNA perfect matches（1 mol/L，37 ℃）

Sequence	ΔH（kcal/mol）	ΔS（cal/mol）	ΔG（kcal/mol）	Sequence	ΔH（kcal/mol）	ΔS（cal/mol）	ΔG（kcal/mol）
AA/TT	−7.9	−22.2	−1.00	GA/CT	−8.2	−22.2	−1.30
AT/TA	−7.2	−20.4	−0.88	CG/GC	−10.6	−27.2	−2.17
TA/AT	−7.2	−21.3	−0.58	GC/CG	−9.8	−24.2	−2.24
CA/GT	−8.5	−22.7	−1.45	GG/CC	−8.0	−19.9	−1.84
GT/CA	−8.4	−22.4	−1.44	Init/Term GC	0.1	−2.8	0.98
CT/GA	−7.8	−21.0	−1.28	Init/Term AT	2.3	4.1	1.03
Symmetry	0.0	−1.4	0.00				

1.2　单独错配

当出现单独错配的模式时，应用 Allawi、SantaLucia 以及 Peyret 等人于 1997～1999 年提供的热力学数据[9-13]（见表 2）。单独错配结构的热力学计算如下：

$$\Delta H\left(\begin{smallmatrix}ATC\\ TCG\end{smallmatrix}\right) = \Delta H\left(\begin{smallmatrix}AT\\ TC\end{smallmatrix}\right) + \Delta H\left(\begin{smallmatrix}TC\\ CG\end{smallmatrix}\right);$$
$$\Delta S\left(\begin{smallmatrix}ATC\\ TCG\end{smallmatrix}\right) = \Delta S\left(\begin{smallmatrix}AT\\ TC\end{smallmatrix}\right) + \Delta S\left(\begin{smallmatrix}TC\\ CG\end{smallmatrix}\right)。 \quad (2)$$

表 2　单独错配结构热力学参数（1 mol/L，37 ℃）

Table 2　Thermodynamic parameters for DNA/DNA single mismatch（1 mol/L，37 ℃）

Sequence	ΔH（kcal/mol）	ΔS（cal/mol）	Sequence	ΔH（kcal/mol）	ΔS（cal/mol）	Sequence	ΔH（kcal/mol）	ΔS（cal/mol）	Sequence	ΔH（kcal/mol）	ΔS（cal/mol）
AA/TA	1.2	1.7	CC/GC	−1.5	−7.2	GG/CG	−6.0	−15.8	GT/CG	−4.4	−12.3
CA/GA	−0.9	−4.2	GC/CC	3.6	8.9	TG/AG	1.6	3.6	TG/AT	−0.1	−1.7
GA/CT	−2.9	−9.8	TC/AC	6.1	16.4	AT/TT	−2.7	−10.8	TT/AG	−1.3	−5.3
TA/AA	4.7	12.9	AG/TG	−3.1	−9.5	CT/GT	−5.0	−15.8	AA/TG	−0.6	−2.3
AC/TC	0.0	−4.4	CG/GG	−4.9	−15.3	GT/CT	−2.2	−8.4	AG/TA	−0.7	−2.3
TT/AT	0.2	−1.5	CA/GG	−0.7	−2.3	GT/CC	5.2	13.5	AC/TT	0.7	0.2
AG/TT	1.0	0.9	CG/GA	−4.0	−13.2	TC/AT	1.2	0.7	AT/TC	−1.2	−6.2
AT/TG	−2.5	−8.3	GA/CG	−0.6	−1.0	TT/AC	1.0	0.7	CC/GT	−0.8	−4.5
CG/GT	−4.1	−11.7	GG/CA	0.5	3.2	AA/TC	2.3	4.6	CT/GC	−1.5	−6.1
CT/GG	−2.8	−8.0	TA/AG	0.7	0.7	AC/TA	5.3	14.6	GC/CT	2.3	5.4
GG/CT	3.3	10.4	TG/AA	3.0	7.4	CA/GC	1.9	3.7	GA/CC	5.2	14.2
GC/CA	−0.7	−3.8	TA/AC	3.4	8.0	CC/GA	0.6	−0.6	TC/AA	7.6	20.2

1.3 连续错配

当出现连续错配即两个毗连错配的模式时,应用 Allawi、SantaLucia 以及 Peyret 等人于 1997~1999 年提供的热力学数据[9-13](见表 3)。连续错配结构的热力学计算如下:

$$\Delta H\left(\begin{smallmatrix}\mathrm{ATGC}\\ \mathrm{TCAG}\end{smallmatrix}\right) = \Delta H\left(\begin{smallmatrix}\mathrm{AT}\\ \mathrm{TC}\end{smallmatrix}\right) + \Delta H\left(\begin{smallmatrix}\mathrm{TG}\\ \mathrm{CA}\end{smallmatrix}\right) + \Delta H\left(\begin{smallmatrix}\mathrm{GC}\\ \mathrm{AG}\end{smallmatrix}\right);$$
$$\Delta S\left(\begin{smallmatrix}\mathrm{ATGC}\\ \mathrm{TCAG}\end{smallmatrix}\right) = \Delta S\left(\begin{smallmatrix}\mathrm{AT}\\ \mathrm{TC}\end{smallmatrix}\right) + \Delta S\left(\begin{smallmatrix}\mathrm{TG}\\ \mathrm{CA}\end{smallmatrix}\right) + \Delta S\left(\begin{smallmatrix}\mathrm{GC}\\ \mathrm{AG}\end{smallmatrix}\right)。 \quad (3)$$

表 3 连续错配结构热力学参数(1 mol/L,37 ℃)

Table 3 Thermodynamic parameters for DNA/DNA tandem mismatches (1 mol/L,37 ℃)

Sequence	ΔH(kcal/mol)	ΔS(cal/mol)
GG/TT	5.8	16.3
GT/TG	4.1	9.5
TG/GT	-1.4	-6.2
GT/TT	5.8	16.3
GT/AT	-0.1	-1.7

1.4 内环

当出现内环即三个及其以上毗连错配的模式时,应用 Santalucia 和 Hicks 于 2004 年提供的热力学数据[14](见表 4)。内环结构的热力学计算如下:

$$\Delta H\left(\begin{smallmatrix}\mathrm{GACCGC}\\ \mathrm{CCATAG}\end{smallmatrix}\right) = \Delta H\left(\begin{smallmatrix}\mathrm{GA}\\ \mathrm{CC}\end{smallmatrix}\right) + \Delta H\left(\begin{smallmatrix}\mathrm{GC}\\ \mathrm{AG}\end{smallmatrix}\right) + \Delta H(\mathrm{loop-of-8}) = \Delta H\left(\begin{smallmatrix}\mathrm{GA}\\ \mathrm{CC}\end{smallmatrix}\right) + \Delta H\left(\begin{smallmatrix}\mathrm{GC}\\ \mathrm{AG}\end{smallmatrix}\right);$$
$$\Delta S\left(\begin{smallmatrix}\mathrm{GACCGC}\\ \mathrm{CCATAG}\end{smallmatrix}\right) = \Delta S\left(\begin{smallmatrix}\mathrm{GA}\\ \mathrm{CC}\end{smallmatrix}\right) + \Delta S\left(\begin{smallmatrix}\mathrm{GC}\\ \mathrm{AG}\end{smallmatrix}\right) + \Delta S(\mathrm{loop-of-8})。 \quad (4)$$

表 4 内环结构热力学参数(1 mol/L,37 ℃)

Table 4 Thermodynamic parameters for DNA/DNA internal loop (1 mol/L,37 ℃)

n (loop of n)	ΔS (cal/mol)	n (loop of n)	ΔS (cal/mol)	n (loop of n)	ΔS (cal/mol)
3	10.3	4	11.6	5	12.9
6	14.2	7	14.8	8	15.5
9	15.8	10	15.8	12	16.7
14	17.4	16	18.1	18	18.7
20	19.0	25	20.3	30	21.3

1.5 单独摇摆末端

所谓单独摇摆末端(Single dangling-end),指杂交序列 5'或 3'末端出现一个未匹配的核酸即空位 gap 结构(用"-"表示)。当出现这种模式时,应用 Bommarito 等人于 2000 年提供的热力学数据[15](见表 5)。Single dangling-end 结构的热力学计算如下:

$$\Delta H\left(\begin{smallmatrix}\mathrm{GCTAG-}\\ \mathrm{CGATCA}\end{smallmatrix}\right) = \Delta H\left(\begin{smallmatrix}\mathrm{GCTAG}\\ \mathrm{CGATC}\end{smallmatrix}\right) + \Delta H\left(\begin{smallmatrix}\mathrm{G-}\\ \mathrm{CA}\end{smallmatrix}\right);$$
$$\Delta S\left(\begin{smallmatrix}\mathrm{GCTAG-}\\ \mathrm{CGATCA}\end{smallmatrix}\right) = \Delta S\left(\begin{smallmatrix}\mathrm{GCTAG}\\ \mathrm{CGATC}\end{smallmatrix}\right) + \Delta S\left(\begin{smallmatrix}\mathrm{G-}\\ \mathrm{CA}\end{smallmatrix}\right)。 \quad (5)$$

表 5 Single dangling-end 结构热力学参数(1 mol/L,37 ℃)

Table 5 Thermodynamic parameters for DNA/DNA single dangling-end (1 mol/L,37 ℃)

Sequence	ΔH (kcal/mol)	ΔS (cal/mol)	Sequence	ΔH (kcal/mol)	ΔS (cal/mol)	Sequence	ΔH (kcal/mol)	ΔS (cal/mol)	Sequence	ΔH (kcal/mol)	ΔS (cal/mol)
AA/-T	0.2	2.3	CG/-C	-4.0	-11.9	TG/A-	-1.6	-3.6	CC/-G	-4.4	-12.6
TA/-T	-6.9	-20.0	AT/-A	-2.9	-7.6	TT/A-	2.9	10.4	AG/-C	-3.7	-10
GA/-T	-1.1	-1.6	TT/-A	-0.2	-0.5	GA/C-	-2.1	-3.9	TG/-C	-4.9	-13.8
CA/-T	0.6	3.3	GT/-A	-4.2	-15.0	GC/C-	-0.2	-0.1	GG/-C	-3.9	-10.9
AC/-G	-6.3	-17.1	CT/-A	-4.1	-13.0	GG/C-	-3.9	-11.2	TA/A-	-0.7	-0.8
TC/-G	-4.0	-10.9	AA/T-	-0.5	-1.1	GT/C-	-4.4	-13.1	TC/A-	4.4	14.9
GC/-G	-5.1	-14.0	AC/T-	4.7	14.2	CA/G-	-5.9	-16.5	CT/G-	-5.2	-15.0
AT/T-	-3.8	-12.6	AG/T-	-4.1	-13.1	CC/G-	-2.6	-7.4	CG/G-	-3.2	-10.4

1.6 长摇摆末端

所谓长摇摆末端(Long dangling-end),指杂交序列 5'或 3'末端出现连续两个及其以上未匹配的核酸即空位 gap 结构(用"-"表示)。当出现这种模式时,应用 Sugimoto 等人于 2002 年提供的热力学数据[16](见表 6)。Long dangling-end 结构的热力学计算如下:

$$\Delta H\left(\begin{smallmatrix}\mathrm{GCTAG---}\\ \mathrm{CGATCAAA}\end{smallmatrix}\right) = \Delta H\left(\begin{smallmatrix}\mathrm{GCTAG}\\ \mathrm{CGATC}\end{smallmatrix}\right) + \Delta H\left(\begin{smallmatrix}\mathrm{G---}\\ \mathrm{CAAA}\end{smallmatrix}\right);$$
$$\Delta S\left(\begin{smallmatrix}\mathrm{GCTAG---}\\ \mathrm{CGATCAAA}\end{smallmatrix}\right) = \Delta S\left(\begin{smallmatrix}\mathrm{GCTAG}\\ \mathrm{CGATC}\end{smallmatrix}\right) + \Delta S\left(\begin{smallmatrix}\mathrm{G---}\\ \mathrm{CAAA}\end{smallmatrix}\right)。 \quad (6)$$

表 6　Long dangling-end 结构热力学参数(1 mol/L,37 ℃)

Table 6　Thermodynamic parameters for DNA/DNA long dangling-end (1 mol/L,37 ℃)

Sequence	ΔH (kcal/mol)	ΔS (cal/mol)	Sequence	ΔH (kcal/mol)	ΔS (cal/mol)	Sequence	ΔH (kcal/mol)	ΔS (cal/mol)
CAA/G--	-2.15	-5.5	CAAA/G---	-3.3	-8.5	CAAAA/G----	-4.85	-13.5
TAA/A--	-1.0	-1.0	TAAA/A---	-1.95	-4.5	TAAAA/A----	-2.35	-6.0
AAA/--T	-1.5	-2.5	AAAA/---T	-1.75	-3.5	AAAAA/----T	-3.95	-10.5
AAG/--C	-0.25	0.5	AAAG/---C	-0.15	1.5	AAAAG/----C	-0.75	-0.5

1.7　单独膨胀环

所谓单独膨胀环(Single bulge loop),指杂交序列内部出现一个未匹配的核酸即空位 gap 结构(用“-”表示)。当出现这种模式时,应用 Tanaka 等人于 2004 年提供的热力学数据[17](见表 7)。Single bulge loop 结构的热力学计算如下:

$$\Delta H\left(\begin{smallmatrix}\text{GCTTAGGC}\\ \text{CGA-TCCG}\end{smallmatrix}\right) = \Delta H\left(\begin{smallmatrix}\text{GCT}\\ \text{CGA}\end{smallmatrix}\right) + \Delta H\left(\begin{smallmatrix}\text{TTA}\\ \text{A-T}\end{smallmatrix}\right) + \Delta H\left(\begin{smallmatrix}\text{AGGC}\\ \text{TCCG}\end{smallmatrix}\right);$$
$$\Delta S\left(\begin{smallmatrix}\text{GCTTAGGC}\\ \text{CGA-TCCG}\end{smallmatrix}\right) = \Delta S\left(\begin{smallmatrix}\text{GCT}\\ \text{CGA}\end{smallmatrix}\right) + \Delta S\left(\begin{smallmatrix}\text{TTA}\\ \text{A-T}\end{smallmatrix}\right) + \Delta S\left(\begin{smallmatrix}\text{AGGC}\\ \text{TCCG}\end{smallmatrix}\right)。 \quad (7)$$

表 7　Single bulge loop 结构热力学参数(1 mol/L,37 ℃)

Table 7　Thermodynamic parameters for DNA/DNA single bulge loop (1 mol/L,37 ℃)

Sequence	ΔH (kcal/mol)	ΔS (cal/mol)	Sequence	ΔH (kcal/mol)	ΔS (cal/mol)	Sequence	ΔH (kcal/mol)	ΔS (cal/mol)	Sequence	ΔH (kcal/mol)	ΔS (cal/mol)
AAA/T-T	-4.0	-17.5	AAT/T-A	-13.5	-44.6	AAC/T-G	10.2	25.8	AAG/T-C	-0.2	-4.6
ATA/T-T	9.8	24.7	ATT/T-A	-19.5	-61.8	ATC/T-G	-4.8	-20.0	ATG/T-C	5.6	12.9
ACA/T-T	6.1	14.0	ACT/T-A	-3.4	-15.3	ACC/T-G	-29.1	-91.3	ACG/T-C	-1.2	-7.4
AGA/T-T	15.5	43.1	AGT/T-A	5.3	10.6	AGC/T-G	-7.2	-28.4	AGG/T-C	5.7	14.4
TAA/A-T	15.3	38.5	TAT/A-A	19.8	56.7	TAC/A-G	-2.3	-12.7	TAG/A-C	15	40.4
TTA/A-T	-2.7	-16.0	TTT/A-A	-8.2	-30.1	TTC/A-G	8.8	23.2	TTG/A-C	7.0	17.5
TCA/A-T	1.6	-3.9	TCT/A-A	9.9	30.5	TCC/A-G	-15.2	-48.7	TCG/A-C	-0.7	-7.0
TGA/A-T	-12.3	-47.0	TGT/A-A	20.9	59.8	TGC/A-G	2.6	1.6	TGG/A-C	-9.2	-31.4
CAA/G-T	-14.4	-46.6	CAT/G-A	-7.0	-25.8	CAC/G-G	4.3	8.8	CAG/G-C	-2.9	-13.0
CTA/G-T	-18.7	-61.7	CTT/G-A	-4.6	-17.0	CTC/G-G	-14.5	-48.1	CTG/G-C	-4.7	-18.8
CCA/G-T	5.8	13.4	CCT/G-A	-5.3	-24.8	CCC/G-G	-2.6	-9.4	CCG/G-C	9.1	24.4
CGA/G-T	-14.7	-49.6	CGT/G-A	2.1	6.5	CGC/G-G	-4.4	-17.8	CGG/G-C	-16.4	-51.6
GAA/C-T	-6.8	-25.2	GAT/C-A	-9.8	-35.6	GAC/C-G	-4.8	-18.3	GAG/C-C	-6.5	-21.0
GTA/C-T	-7.4	-27.0	GTT/C-A	1.8	-0.6	GTC/C-G	-12.3	-40.0	GTG/C-C	-2.3	-10.0
GCA/C-T	-2.1	-12.5	GCT/C-A	-3.5	-16.1	GCC/C-G	0.4	-1.0	GCG/C-C	13.8	42.9
GGA/C-T	-0.7	3.6	GGT/C-A	-1.2	-9.3	GGC/C-G	-1.7	-7.2	GGG/C-C	3.5	8.5

1.8　长膨胀环

所谓长膨胀环(Long bulge loop),指杂交序列内部出现连续两个及其以上未匹配的核酸即空位 gap 结构(用“-”表示)。当出现这种模式时,应用 Santalucia 和 Hicks 于 2004 年提供的热力学数据[14](见表 8)。Long bulge loop 结构的热力学计算如下:

$$\Delta H\left(\begin{smallmatrix}\text{GACCC}\\ \text{C---G}\end{smallmatrix}\right) = \Delta H(\text{Term - AT - penalty}) + \Delta H(\text{bulge - loop - of - 3}) = 0;$$
$$\Delta S\left(\begin{smallmatrix}\text{GACCC}\\ \text{C---G}\end{smallmatrix}\right) = \Delta S(\text{Term - AT - penalty}) + \Delta S(\text{bulge - loop - of - 3}) = \Delta S(\text{bulge - loop - of - 3})。 \quad (8)$$

表 8　Long bulge loop 结构热力学参数(1 mol/L,37 ℃)

Table 8　Thermodynamic parameters for DNA/DNA long bulge loop (1 mol/L,37 ℃)

n (loop of n)	ΔS (cal/mol)	n (loop of n)	ΔS (cal/mol)	n (loop of n)	ΔS (cal/mol)
2	-9.35	3	-10.0	4	-10.32
5	-10.64	6	-11.28	7	-11.93
8	-12.57	9	-13.22	10	-13.86
11	-14.15	12	-14.51	13	-15.00
14	-15.48	15	-15.65	16	-16.12
17	-16.26	18	-16.77	19	-16.80
20	-17.09	25	-18.06	30	-19.02

2 DNA 与 RNA 相互作用的热力学参数

当 DNA 与 RNA 相互作用时,完美匹配的最近邻热力学参数表由 Sugimoto 等人于 1995 年提出[18](见表 9),完美匹配结构的热力学计算同 1.1。

表 9 DNA/RNA 结合完美匹配最近邻热力学参数(1 mol/L,37 ℃)

Table 9 Thermodynamic parameters for DNA/RNA perfect matches (1 mol/L,37 ℃)

Sequence	ΔH (kcal/mol)	ΔS (cal/mol)	Sequence	ΔH (kcal/mol)	ΔS (cal/mol)	Sequence	ΔH (kcal/mol)	ΔS (cal/mol)
dAA/rUU	−11.5	−36.4	dCG/rGC	−16.3	−47.1	dTA/rAU	−7.8	−23.2
dAC/rUG	−7.8	−21.6	dCT/rGA	−9.1	−23.5	dTC/rAG	−5.5	−13.5
dAG/rUC	−7.0	−19.7	dGA/rCU	−8.6	−22.9	dTG/rAC	−9.0	−26.1
dAT/rUA	−8.3	−23.9	dGC/rCG	−8.0	−17.1	dTT/rAA	−7.8	−21.9
dCA/rGU	−10.4	−28.4	dGG/rCC	−9.3	−23.2	Initiation	1.9	−3.9
dCC/rGG	−12.8	−31.9	dGT/rCA	−5.9	−12.3			

3 RNA 与 RNA 相互作用的热力学参数

3.1 完美匹配

当 RNA 与 RNA 相互作用时,完美匹配的最近邻热力学参数表[19]由 Xia 等人于 1998 年提出(见表 10)。当 mRNA 与 RNA 相互作用时,完美匹配的最近邻热力学参数表[20]由 Turner 等人于 2006 年提出(见表 11)。完美匹配结构的热力学计算同 1.1。

表 10 RNA/RNA 结合完美匹配最近邻热力学参数表(1 mol/L,37 ℃)

Table 10 Thermodynamic parameters for RNA/RNA perfect matches (1 mol/L,37 ℃)

Sequence	ΔH (kcal/mol)	ΔS (cal/mol)	Sequence	ΔH (kcal/mol)	ΔS (cal/mol)	Sequence	ΔH (kcal/mol)	ΔS (cal/mol)
AA/UU	−6.82	−19.0	CC/GG	−13.39	−32.7	Initiation	3.61	−1.5
AC/UG	−11.4	−29.5	CG/GC	−10.64	−26.7	Ter per-A/U	3.72	10.5
AG/UC	−10.48	−27.1	GA/CU	−12.44	−32.5	symmetry	0.0	−1.4
AU/UA	−9.38	−26.7	GC/CG	−14.88	−36.9			
CA/GU	−10.44	−26.9	UA/AU	−7.69	−20.5			

表 11 mRNA/RNA 结合完美匹配最近邻热力学参数表(1 mol/L,37 ℃)

Table 11 Thermodynamic parameters for mRNA/RNA perfect matches (1 mol/L,37 ℃)

Sequence	ΔH (kcal/mol)	ΔS (cal/mol)	Sequence	ΔH (kcal/mol)	ΔS (cal/mol)	Sequence	ΔH (kcal/mol)	ΔS (cal/mol)
mAA/UU	−7.48	−22.3	mCG/rGC	−9.47	−23.0	mUG/rAC	−12.14	−32.9
mAC/rUG	−6.32	−15.2	mCU/rGA	−9.59	−23.89	mUU/rAA	−5.43	−14.5
mAG/rUC	−13.94	−39.1	mGA/rCU	−5.77	−11.9	Initiation	−12.8	−52
mAU/rUA	−6.33	−17.7	mGC/rCG	−11.9	−26.3	Ter per-A/U	3.14	9.1
mCA/rGU	−5.21	−10.7	mGG/rCC	−9.66	−22.1	Symmetry	0.0	−1.4
mCC/rGG	−8.88	−19.7	mGU/rCA	−6.62	−14.3			
mUA/rAU	−6.47	−17.0	mUC/rAG	−9.65	−25.0			

3.2 错配

当出现单独错配的模式时,应用 Znosko 等人于 2008年提供的热力学数据[21](见表12),其中R表示嘌呤碱基,Y 表示嘧啶碱基;单独错配结构的热力学计算如下:

$$\Delta H\left(\begin{smallmatrix}\mathrm{AUC}\\ \mathrm{UUG}\end{smallmatrix}\right) = \Delta H\left(\begin{smallmatrix}\mathrm{U}\\ \mathrm{U}\end{smallmatrix}\right) + \Delta H(\mathrm{per\_AU/GU}) + \Delta H\left(\begin{smallmatrix}\mathrm{RYY}\\ \mathrm{YYR}\end{smallmatrix}\right) = \Delta H\left(\begin{smallmatrix}\mathrm{U}\\ \mathrm{U}\end{smallmatrix}\right) + 1 \times \Delta H\left(\begin{smallmatrix}\mathrm{A}\\ \mathrm{U}\end{smallmatrix}\right) + \Delta H\left(\begin{smallmatrix}\mathrm{RYY}\\ \mathrm{YYR}\end{smallmatrix}\right);$$

$$\Delta S\left(\begin{smallmatrix}\mathrm{AUC}\\ \mathrm{UUG}\end{smallmatrix}\right) = \Delta S\left(\begin{smallmatrix}\mathrm{U}\\ \mathrm{U}\end{smallmatrix}\right) + \Delta S(\mathrm{per\_AU/GU}) + \Delta S\left(\begin{smallmatrix}\mathrm{RYY}\\ \mathrm{YYR}\end{smallmatrix}\right) = \Delta S\left(\begin{smallmatrix}\mathrm{U}\\ \mathrm{U}\end{smallmatrix}\right) + 1 \times \Delta S\left(\begin{smallmatrix}\mathrm{A}\\ \mathrm{U}\end{smallmatrix}\right) + \Delta S\left(\begin{smallmatrix}\mathrm{RYY}\\ \mathrm{YYR}\end{smallmatrix}\right)。 \quad (9)$$

当出现连续错配的模式时,应用 Turner 等人于 1999 和 2006 年提供的热力学数据[22-23](见表 13)。连续错配结构的热力学计算如下:

$$\Delta H\left(\begin{smallmatrix}\mathrm{GACC}\\ \mathrm{CCAG}\end{smallmatrix}\right) = \mathrm{adjacent\_to\_GC\_}\Delta H\left(\begin{smallmatrix}\mathrm{AC}\\ \mathrm{CA}\end{smallmatrix}\right);$$

$$\Delta S\left(\begin{smallmatrix}\mathrm{GACC}\\ \mathrm{CCAG}\end{smallmatrix}\right) = \mathrm{adjacent\_to\_GC\_}\Delta S\left(\begin{smallmatrix}\mathrm{AC}\\ \mathrm{CA}\end{smallmatrix}\right)。 \quad (10)$$

表 12　单独错配结构热力学参数(1 mol/L,37 ℃)

Table 12　Thermodynamic parameters for RNA/RNA single mismatch (1 mol/L,37 ℃)

Sequence	ΔH (kcal/mol)	ΔS (cal/mol)	Sequence	ΔH (kcal/mol)	ΔS (cal/mol)
A/G	−0.8	−10.4	YRR/RRY	0.8	6.2
U/U	−14.9	−45.7	RYY/YYR	−0.3	−1.2
G/G	−17.9	−52.2	YYR/RYY	2.6	6.5
per_A/U	−4	−15.0	YRY/RYR	−7.0	−20.3
per_G/U	−4	−19.3	RRY/YYR	−17.3	−54.5

表 13　连续错配结构热力学参数表(1 mol/L,37 ℃)

Table 13　Thermodynamic parameters for RNA/RNA tandem mismatches (1 mol/L,37 ℃)

Sequence (closing="G/C")	ΔH (kcal/mol)	ΔS (cal/mol)	Sequence (closing="C/G")	ΔH (kcal/mol)	ΔS (cal/mol)	Sequence (closing="U/A")	ΔH (kcal/mol)	ΔS (cal/mol)	Sequence (closing="A/U")	ΔH (kcal/mol)	ΔS (cal/mol)
GA/AG	−25.1	−72.5	GA/AG	−14.6	−44.8	GA/AG	−13.4	−45.5	GA/AG	−14.4	−47.4
AG/GA	−15.6	−46.1	AG/GA	−12.7	−38.7	AG/GA	−19.4	−64.8	AG/GA	−10.8	−35.8
UU/UU	−14.4	−44.8	UU/UU	−17.5	−55.1	UU/UU	−6.7	−25.1	UU/UU	−12.2	−41.3
GG/GG	−22.8	−76.1	GG/GG	−22.8	−76.1	GG/GG	2.7	3.9	GG/GG	−1.0	−7.7
CA/AC	−10.3	−36.4	CA/AC	−10.8	−38.4	CA/AC	9.1	23.2	CA/AC	7.2	15.8
CU/UC	−20.9	−70.9	CU/UC	−0.6	−6.4	CU/UC	3.3	3.5	CU/UC	7.4	−16.8
UC/CU	−14.7	−50.6	UC/CU	−2.8	−13.5	UC/CU	9.5	21.6	UC/CU	7.4	−16.8
CC/CC	−14.7	−50.6	CC/CC	−1.8	−11.3	CC/CC	12.1	30.0	CC/CC	7.4	−16.8
AC/CA	−8.6	−30.6	AC/CA	−1.7	−11.9	AC/CA	12.1	30.1	AC/CA	7.5	16.1
AA/AA	−1.3	−9.0	AA/AA	−4.6	−19.0	AA/AA	14.7	38.4	AA/AA	13.4	34.2

3.3　内环

当出现内环结构,应用 Turner 等人于 1999 年和 2006 年提供的热力学数据[22-23](见表 14)。内环结构的热力学计算如下:

$$\Delta H\left(\begin{smallmatrix}\mathrm{AACCGC}\\ \mathrm{UC-UAG}\end{smallmatrix}\right) = \Delta H(\mathrm{loop-of-}n) + \Delta H(\mathrm{per\_AU/GU}) + (n1-n2)\Delta H(\mathrm{asymmetry}) = \Delta H\,\mathrm{loop}(7) + \Delta H\left(\begin{smallmatrix}\mathrm{A}\\ \mathrm{U}\end{smallmatrix}\right) + (4-3)\Delta H(\mathrm{asymmetry});$$

$$\Delta S\left(\begin{smallmatrix}\mathrm{AACCGC}\\ \mathrm{UC-UAG}\end{smallmatrix}\right) = \Delta S(\mathrm{loop-of-}n) + \Delta S(\mathrm{per\_AU/GU}) + (n1-n2)\Delta S(\mathrm{asymmetry}) = \Delta H\,\mathrm{loop}(7) + \Delta S\left(\begin{smallmatrix}\mathrm{A}\\ \mathrm{U}\end{smallmatrix}\right) + (4-3)\Delta S(\mathrm{asymmetry})。 \quad (11)$$

3.4　CNG 重复序列

所谓 CNG 重复序列,是指一条序列(5'至 3'方向)由多个 G(CNG)xC 的子片段组成的序列,这里 x 是指 CNG 重复的数目;N 代表一种单独错配类型即 N/N。当出现这种模式时,应用 Broda 等人于 2005 年提供的热力学数据[24](见表 15)。CNG 重复结构的热力学计算如下:

$$\Delta H\left(\begin{smallmatrix}\mathrm{GCAGCAGCAGC}\\ \mathrm{CGACGACGACG}\end{smallmatrix}\right) = \Delta H(3-\mathrm{CAG}-\mathrm{repeats});$$

$$\Delta S\left(\begin{smallmatrix}\mathrm{GCAGCAGCAGC}\\ \mathrm{CGACGACGACG}\end{smallmatrix}\right) = \Delta S(3-\mathrm{CAG}-\mathrm{repeats})。 \quad (12)$$

表 14　内环结构热力学参数(1 mol/L,37 ℃)

Table 14　Thermodynamic parameters for RNA/RNA internal loop (1 mol/L,37 ℃)

n(loop of n)	ΔH (kcal/mol)	ΔS (cal/mol)	n(loop of n)	ΔH (kcal/mol)	ΔS (cal/mol)	n(loop of n)	ΔH (kcal/mol)	ΔS (cal/mol)
2	−10.5	−35.4	5	−6.8	−28.4	per_A/U	5	13.9
3	0.3	−4.5	6	−1.3	−10.6	per_G/U	5	13.9
4	−7.2	−26.8	>6	−1.3	−10.6	asymmetry	3.2	8.4

表 15 CNG 重复结构热力学参数(1 mol/L,37 ℃)

Table 15 Thermodynamic parameters for CNG repeats (1 mol/L,37 ℃)

Sequence(x)	ΔH (kcal/mol)	ΔS (cal/mol)	Sequence(x)	ΔH (kcal/mol)	ΔS (cal/mol)	Sequence(x)	ΔH(kcal/mol)	ΔS (cal/mol)
CAG(2)	−40.1	−110.1	CAG(4)	−32.0	−80.2	CAG(6)	−35.5	−104.0
CGG(2)	−50.1	−139.5	CGG(4)	−58.0	−153.0	CGG(6)	−23.9	−69.8
CCG(2)	−29.3	−74.9	CCG(4)	−45.8	−66.7	CCG(6)	−24.4	−72.0
CUG(2)	−49.6	−141.9	CUG(4)	−43.9	−116.7	CUG(6)	−39.8	−118.6
CAG(3)	−30.1	−75.4	CAG(5)	−29.3	−86.6	CAG(7)	−30.6	−89.4
CGG(3)	−50.7	−134.3	CGG(5)	−19.9	−58.2	CGG(7)	−23.4	−67.5
CCG(3)	−42.2	−114.4	CCG(5)	−32.8	−121.5	CCG(7)	−27.3	−80.1
CUG(3)	−54.9	−147.9	CUG(5)	−42.8	−128.3	CUG(7)	−49.86	−143.5

3.5 GU 摆动配对

当 RNA/RNA 相互作用呈现多个连续的 GU 配对(非 Watson-Crick 配对)结构时,应用 Turner 等人于 1999 年提供的热力学数据[22](见表 16)。GU 摆动配对结构的热力学计算如下:

One GU base pair: $\Delta H\left(\begin{smallmatrix}\mathrm{GUC}\\ \mathrm{CGG}\end{smallmatrix}\right)=\Delta H\left(\begin{smallmatrix}\mathrm{GU}\\ \mathrm{CG}\end{smallmatrix}\right)+\Delta H\left(\begin{smallmatrix}\mathrm{UC}\\ \mathrm{GG}\end{smallmatrix}\right)$,

$\Delta S\left(\begin{smallmatrix}\mathrm{GUC}\\ \mathrm{CGG}\end{smallmatrix}\right)=\Delta S\left(\begin{smallmatrix}\mathrm{GU}\\ \mathrm{CG}\end{smallmatrix}\right)+\Delta S\left(\begin{smallmatrix}\mathrm{UC}\\ \mathrm{GG}\end{smallmatrix}\right)$;

Two adjacent GU base pairs: $\Delta H\left(\begin{smallmatrix}\mathrm{GUGC}\\ \mathrm{CGUG}\end{smallmatrix}\right)=\Delta H\left(\begin{smallmatrix}\mathrm{GU}\\ \mathrm{CG}\end{smallmatrix}\right)+\Delta H\left(\begin{smallmatrix}\mathrm{UG}\\ \mathrm{GU}\end{smallmatrix}\right)+\Delta H\left(\begin{smallmatrix}\mathrm{UC}\\ \mathrm{GG}\end{smallmatrix}\right)$, $\Delta S\left(\begin{smallmatrix}\mathrm{GUGC}\\ \mathrm{CGUG}\end{smallmatrix}\right)=\Delta S\left(\begin{smallmatrix}\mathrm{GU}\\ \mathrm{CG}\end{smallmatrix}\right)+\Delta S\left(\begin{smallmatrix}\mathrm{UG}\\ \mathrm{GU}\end{smallmatrix}\right)+\Delta S\left(\begin{smallmatrix}\mathrm{UC}\\ \mathrm{GG}\end{smallmatrix}\right)$。 (13)

表 16 GU 摆动配对结构热力学参数(1 mol/L,37 ℃)

Table 16 Thermodynamic parameters for GU wobble base pairs (1 mol/L,37 ℃)

n(loop of n)	ΔH (kcal/mol)	ΔS (cal/mol)	n(loop of n)	ΔH (kcal/mol)	ΔS (cal/mol)	n(loop of n)	ΔH (kcal/mol)	ΔS (cal/mol)
AG/UU	−3.2	−8.6	GG/CU	−8.33	−21.9	GG/UU	−13.47	−44.9
AU/UG	−8.81	24.0	GU/CG	−12.59	−32.5	UG/GU	−9.26	−30.8
CG/GU	−5.61	−13.5	UG/AU	−6.99	−19.3	per_G/U	3.72	10.5
CU/GG	−12.11	−32.2	UU/AG	−12.83	−37.3			

3.6 摇摆

对于 Sigle dangling-end 模式,应用 Serra 等人于 2006 和 2008 年提供的热力学数据[25-26](见表 17);对于 Long dangling-end 模式,应用 Sugimoto 与 Serra 等人于 2002 和 2006 年提供的热力学数据[16, 25](见表 18),其中 R 表示嘌呤碱基,Y 表示嘧啶碱基。Dangling-end 结构的热力学计算同 1.5 与 1.6。

表 17 Single dangling-end 结构热力学参数(1 mol/L,37 ℃)

Table 17 Thermodynamic parameters forRNA/RNA single dangling-end (1 mol/L,37 ℃)

Sequence	ΔH (kcal/mol)	ΔS (cal/mol)	Sequence	ΔH (kcal/mol)	ΔS (cal/mol)	Sequence	ΔH (kcal/mol)	ΔS (cal/mol)	Sequence	ΔH (kcal/mol)	ΔS (cal/mol)
UA/A−	−5.2	−14.3	UC/A−	−3	−8.7	UG/A−	−6.6	−18.8	UU/A−	−3.1	−8.9
UA/G−	−2.9	−6.5	UC/G−	−2.8	−6.7	UG/G−	−6.4	−17.5	UU/G−	−1.5	−3.1
GA/U−	−2.6	−6.8	GC/U−	−4.0	−10.2	GG/U−	−1.7	−4.4	GU/U−	−2.2	−5.4
AU/−G	2.8	9.7	CU/−G	0.4	2.3	GU/−G	1.4	5.3	UU/−G	4.3	14
AG/−U	1.6	5.3	CG/−U	0.5	1.6	GG/−U	−0.6	−1.7	UG/−U	2.2	6.8

表 18 Long dangling-end 结构热力学参数(1 mol/L,37 ℃)

Table 18 Thermodynamic parameters forRNA/RNA long dangling-end (1 mol/L,37 ℃)

Sequence	ΔH (kcal/mol)	ΔS (cal/mol)	Sequence	ΔH (kcal/mol)	ΔS (cal/mol)	Sequence	ΔH (kcal/mol)	ΔS (cal/mol)	Sequence	ΔH (kcal/mol)	ΔS (cal/mol)
CA/G−	−1.05	1.0	CAA/G−−	−6.75	−16.0	CAAA/G−−−	−10.45	−26.5	CAAA/G−−−	−11.95	−31
AG/−C	−0.4	−0.5	AAG/−−C	−5.2	−20.5	AAAG/−−−C	−8.25	−24.5	AAAG/−−−C	−13.1	−39.5
UA/A−	−0.95	−1.0	UAA/A−−	−3.2	−7.0	UAAA/A−−−	−4.2	−9.5	UAAA/A−−−	−5.25	−13.0
AA/−U	−1.1	−2.5	AAA/−−U	−3.35	−9.0	AAAA/−−−U	−3.8	−10.0	AAAA/−−−U	−3.85	−9.0
RY/Y−	0.0	0.0	RRR/Y−−	0.0	0.0	RRY/Y−−	−2.5	−6.7	YY/R−	0.0	0.0
YR/R−	−2.5	−6.7									

3.7　膨胀环

对于 Single bulge loop 模式,应用 Serra 等人于 2007 年提供的热力学数据[27](见表 19),Single bulge loop 结构的热力学计算同 1.7;对于 Long bulge loop 结构应用 Turner 等人于 1999 和 2006 年提供的热力学数据[22-23](见表 20),Long bulge loop 结构的热力学计算如下:

$$\Delta H\left(\begin{smallmatrix}\mathrm{AACGC}\\ \mathrm{U--G}\end{smallmatrix}\right) = \Delta H(\mathrm{per\_AU/GU}) + \Delta H(\mathrm{bulge-loop-of-3});$$

$$\Delta S\left(\begin{smallmatrix}\mathrm{AACGC}\\ \mathrm{U--G}\end{smallmatrix}\right) = \Delta S(\mathrm{per\_AU/GU}) + \Delta S(\mathrm{bulge-loop-of-3})。\quad (14)$$

表 19　Single bulge-end 结构热力学参数(1 mol/L,37 ℃)

Table 19　Thermodynamic parameters for RNA/RNA single bulge-end (1 mol/L,37 ℃)

Sequence	ΔH (kcal/mol)	ΔS (cal/mol)	Sequence	ΔH (kcal/mol)	ΔS (cal/mol)	Sequence	ΔH (kcal/mol)	ΔS (cal/mol)	Sequence	ΔH (kcal/mol)	ΔS (cal/mol)
CAC/G-G	4.7	2.6	CAG/G-C	14.7	32.9	CAU/G-A	11.9	24.2	CAU/G-G	5.3	13.2
CGC/G-G	8.8	15.5	CGU/G-A	21.8	55.1	CGU/G-G	19.7	58.0	GAG/C-C	8.3	14.5
GAC/C-G	8.2	17.1	GAU/C-A	9.8	17.1	GAU/C-G	21.3	62.2	GCG/C-C	-4.5	-24.2
GCA/C-U	9.4	16.4	GCU/C-A	4.2	9.4	GCG/C-U	1.61	35.8	GUG/C-C	1.3	29.3
GUA/C-U	-0.3	-12.9	GUG/C-U	15.6	35.1	UAG/A-C	16.4	45.8	UAC/A-G	26	66.4
UAU/A-A	20.8	54.2	UAU/A-G	36.8	100.0	UGC/A-G	26.6	69.6	UGU/A-A	12.2	28.4
UGU/A-G	0.056	163.8	UCA/A-U	8.5	14.8	UCU/A-A	4,2	1.0	UCG/A-C	20.1	49.3
UCG/A-U	34.6	97.0	ACA/U-U	-3.7	-25.8	ACG/U-C	22.6	60.1	ACU/U-A	32.4	89.0
ACG/U-U	39.7	114.5	AUG/U-C	15.1	35.8	AUA/U-U	37	0.0	AUG/U-U	42	119.9
GCA/U-U	25.8	70.9	GCG/U-C	11.4	23.9	GCU/U-A	25.2	68.7	GCG/U-U	23.8	66.7
GUG/U-C	4.2	11.9	GUA/U-U	10	19.3	GUG/U-U	24.6	68.4	UAC/G-G	6.2	10.6
UAG/G-C	0.1	-4.8	UAU/G-A	20.7	56.1	UAU/G-G	18.4	50.9	UGC/G-G	21.8	61.6
UGU/G-A	13.8	31.0	UGU/G-G	17.5	46.4						

表 20　Long bulge-end 结构热力学参数(1 mol/L,37 ℃)

Table 20　Thermodynamic parameters forRNA/RNA long bulge-end (1 mol/L,37 ℃)

n(loop of n)	ΔH (kcal/mol)	ΔS (cal/mol)	n(loop of n)	ΔH (kcal/mol)	ΔS (cal/mol)	n(loop of n)	ΔH (kcal/mol)	ΔS (cal/mol)
2	7.1	13.9	5	7.1	10.0	per_A/U	5	13.9
3	7.1	12.6	6	7.1	8.7	per_G/U	5	13.9
4	7.1	11.3	>6	7.1	8.7			

4　特殊碱基的热力学参数

除上述 ATCGU 之外,生物体中同样存在一些非正常的碱基如肌苷(Inosine base,I)、羟基腺嘌呤(2_hydroxyadenine pairs,A*)等。目前发现肌苷在 RNA/RNA 相互作用中以 IU 匹配的形式出现,相关的热力学数据由 Znosko 等人于 2007 年提供[28](见表 21);肌苷在 DNA/DNA 相互作用中的热力学数据由 Santalucia 等人于 2005 年提供(见表 22)[29],热力学计算如下:

One inosine base:

$$\Delta H\left(\begin{smallmatrix}\mathrm{AIC}\\ \mathrm{TAG}\end{smallmatrix}\right) = \Delta H\left(\begin{smallmatrix}\mathrm{AI}\\ \mathrm{TA}\end{smallmatrix}\right) + \Delta H\left(\begin{smallmatrix}\mathrm{IC}\\ \mathrm{AG}\end{smallmatrix}\right), \Delta S\left(\begin{smallmatrix}\mathrm{AIC}\\ \mathrm{TAG}\end{smallmatrix}\right) = \Delta S\left(\begin{smallmatrix}\mathrm{AI}\\ \mathrm{TA}\end{smallmatrix}\right) + \Delta S\left(\begin{smallmatrix}\mathrm{IC}\\ \mathrm{AG}\end{smallmatrix}\right);$$

Two adjacentbase pairs containing inosine :

$$\Delta H\left(\begin{smallmatrix}\mathrm{GIAC}\\ \mathrm{CAIG}\end{smallmatrix}\right) = \Delta H\left(\begin{smallmatrix}\mathrm{GI}\\ \mathrm{CA}\end{smallmatrix}\right) + \Delta H\left(\begin{smallmatrix}\mathrm{IA}\\ \mathrm{AI}\end{smallmatrix}\right) + \Delta H\left(\begin{smallmatrix}\mathrm{AC}\\ \mathrm{IG}\end{smallmatrix}\right), \Delta S\left(\begin{smallmatrix}\mathrm{GIAC}\\ \mathrm{CAIG}\end{smallmatrix}\right) = \Delta S\left(\begin{smallmatrix}\mathrm{GI}\\ \mathrm{CA}\end{smallmatrix}\right) + \Delta S\left(\begin{smallmatrix}\mathrm{IA}\\ \mathrm{AI}\end{smallmatrix}\right) + \Delta S\left(\begin{smallmatrix}\mathrm{AC}\\ \mathrm{IG}\end{smallmatrix}\right)。\quad (15)$$

羟基腺嘌呤(A*)的热力学参数由 Sugimoto 等人于 2001 年提供[30],通常 A* 只在 5'-GA*C-3' 以及 5'-TA*A-3' 序列中出现(见表 23)。热力学计算如下:

$$\Delta H\left(\begin{smallmatrix}\mathrm{GA^*C}\\ \mathrm{CCG}\end{smallmatrix}\right) = \Delta H\left(\begin{smallmatrix}\mathrm{GA}\\ \mathrm{CT}\end{smallmatrix}\right) + \Delta H\left(\begin{smallmatrix}\mathrm{AC}\\ \mathrm{CG}\end{smallmatrix}\right) + \Delta H(\mathrm{GA^*C/CCG});$$

$$\Delta S\left(\begin{smallmatrix}\mathrm{GA^*C}\\ \mathrm{CCG}\end{smallmatrix}\right) = \Delta S\left(\begin{smallmatrix}\mathrm{GA}\\ \mathrm{CT}\end{smallmatrix}\right) + \Delta S\left(\begin{smallmatrix}\mathrm{AC}\\ \mathrm{CG}\end{smallmatrix}\right) + \Delta S(\mathrm{GA^*C/CCG})。\quad (16)$$

表 21 RNA/RNA 结合含有 I 碱基的热力学参数(1 mol/L,37 ℃)

Table 21 Thermodynamic parameters for RNA/RNA inosine base (1 mol/L,37 ℃)

Sequence	ΔH (kcal/mol)	ΔS (cal/mol)	Sequence	ΔH (kcal/mol)	ΔS (cal/mol)	Sequence	ΔH (kcal/mol)	ΔS (cal/mol)	Sequence	ΔH (kcal/mol)	ΔS (cal/mol)
UA/IU	−10.08	−33.8	UU/IA	−11.68	−36.3	UG/IC	−11.99	−36.2	UC/IG	−9.81	27.4
UU/II	9.53	8.8	UI/IU	8.41	19.9	IA/UU	−8.22	−27.9	IU/UA	−15.83	−49.4
IG/UC	−13.38	−39.3	IC/UG−	−11.56	−34.0	IU/UI	17	43.3	per_I/U	−0.08	4.0

表 22 DNA/DNA 结合含有 I 碱基的热力学参数(1 mol/L,37 ℃)

Table 22 Thermodynamic parameters for DNA/DNA inosine base (1 mol/L,37 ℃)

Sequence	ΔH (kcal/mol)	ΔS (cal/mol)	Sequence	ΔH (kcal/mol)	ΔS (cal/mol)	Sequence	ΔH (kcal/mol)	ΔS (cal/mol)	Sequence	ΔH (kcal/mol)	ΔS (cal/mol)
AA/IT	−3.4	−11.2	GC/IG	−1.1	−3.2	IG/AC	−7	−20.0	IT/GA	0.1	−1.8
AC/IG	−7.8	−21.1	GG/IC	7.1	21.3	IT/AA	−0.7	−2.6	IG/GI	3.2	6.1
AG/IC	2.6	8.9	GT/IA	−4.9	−15.8	IA/AI	−13.9	−40.5	IA/TT	−0.8	−4.3
AT/IA	−8.3	−25	GI/IG	−4.2	−39.8	IA/CT	−4.9	−13.9	IC/TG	−4.3	−12.1
AI/IA	−9.5	−29.1	TA/IT	−6.5	−22.0	IC/CG	−5	−12.6	IG/TC	0.1	−1.0
CA/IT	−5.9	−17.4	TC/IG	−3.5	−10.6	IG/CC	−8.3	−23.8	IT/TA	−5.6	−18.7
CC/IG	−6.8	−19.1	TG/IC	−1	−2.4	IT/CA	−8.8	−25.4	IT/TI	−7.6	−31.1
CG/IC	−5.4	−13.7	TT/IA	0.49	−0.7	IC/CI	−12.1	−36.3	IC/IG	−0.5	−1.3
CT/IA	−8.9	−25.5	TI/IT	−1.47	−49.4	IA/GT	1.0	1.0	IG/IC	1.3	3.0
CI/IC	−1.8	−6.4	IA/AT	−1.3	−4.6	IC/GG	−7.6	−22	IT/IA	−3.3	−11.9
GA/IT	−1.9	−8.5	IC/AG	−7.6	−20.2	IG/GC	5.8	16.9	IA/IT	0.1	−2.3

表 23 DNA/DNA 结合含有 A* 结构的热力学参数(1 mol/L,37 ℃)

Table 23 Thermodynamic parameters for DNA/DNA A* (1 mol/L,37 ℃)

Sequence	ΔH (kcal/mol)	ΔS (cal/mol)	Sequence	ΔH (kcal/mol)	ΔS (cal/mol)
TAA/AA*T	26.5	71.0	GAC/CA*G	70.8	203.0
TGA/AA*T	21	61.0	GGC/CA*G	25.6	76.0
TCA/AA*T	9	27.0	GCC/CA*G	19	54.0
TTA/AA*T	14	38.0	GTC/CA*G	9	29.0

5 结语与展望

寡核苷酸二级结构的精确预测在分子生物学应用中发挥越来越重要的作用,热力学参数的不断完善使其精确性得到逐步提高。本文针对不同二级结构,综述目前可利用的有效热力学数据库,涵盖完美匹配、错配、内环、膨胀环、末端摇摆、CNG 重复、GU 摆动配对以及包含肌苷和羟基腺嘌呤等结构。基于本文的数据,用户可以根据不同需求选择合适的数据集,开发自有或者改进现有的寡核苷酸二级结构的预测算法,从而提高以寡核苷酸杂交为基础的分子生物技术的成功率。

然而,生物学过程是复杂的,许多基于核酸杂交的生物技术实验中包含的酶,环境的 pH 值,Ca^{2+}、Mg^{2+}等离子以及一些变性剂如甲酰胺(formamide)、DMSO(二甲基亚砜)都会显著影响寡核苷酸二级结构的评估[14];并且,除 A、T、C、G、U 正常碱基以及少数报道的特殊碱基外,其他生物分子如经甲基化或者乙酰化修饰后的核苷酸也大量存在于生物体中。因此,未来有必要通过实验补充新的热力学参数,使现有的热力学数据得到充实和完善,以进一步提高热力学计算的精确性,从而促进寡核苷酸二级结构的准确预测。

参考文献(References)

[1] FURUICHI T, INOUYE S, INOUYE M. Biosynthesis and structure of stable branched RNA covalently linked to the 5′end of multicopy single-stranded DNA of Stigmatella aurantiaca[J]. Cell, 1987, 48(1):55-62.

[2] GACY A M, GOELLNER G, JURANIC N, et al. Trinucleotide repeats that expand in human disease form hairpin structures in vitro[J]. Cell, 1995, 81(4):533-540.

[3] HIGASHITANI N, HIGASHITANI A, HORIUCHI K. Nucleotide sequence of the primer RNA for DNA replication of filamentous bacteriophages[J]. J Virol, 1993, 67(4): 2175-2181.

[4] MOROZOV S, CHERNOV B K, MERITS A, et al. Computer-assisted predictions of the secondary structure in the plant virus single-stranded DNA genome[J]. J Biomol Struct Dyn, 1994, 11(4):837-847.

[5] CROTHERS D M ,ZIMM B H. Theory of the melting transition of synthetic polynucleotides:Evaluation of the stacking free energy[J]. J Mol Biol, 1964, 9:1-9.

[6] SCHILDKRAUT C. Dependence of the melting temperature of DNA on salt concentration[J]. Biopolymers, 1965, 3(2):195-208.

[7] 张艳春,屈武斌,卢一鸣,等. DNA 解链温度(Tm)不同预测方法的比较[J]. 军事医学, 2011(03):231-235.
ZHANG Yanchun, QU Wubin, LU Yiming, et al. A Comparison of different melting temperature (Tm) prediction methods for DNA[J]. Military Medical Sciences, 2011, 35(3):231-235.

[8] SANTALUCIA J, JR. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics[J]. Proc Natl Acad Sci U S A, 1998, 95(4):1460-1465.

[9] ALLAWI H T, SANTALUCIA J, JR. Thermodynamics and NMR of internal G.T mismatches in DNA[J]. Biochemistry, 1997, 36(34):10581-10594.

[10] ALLAWI H T, SANTALUCIA J, JR. Nearest neighbor thermodynamic parameters for internal G.A mismatches in DNA[J]. Biochemistry, 1998, 37(8):2170-2179.

[11] ALLAWI H T, SANTALUCIA J, JR. Thermodynamics of internal C.T mismatches in DNA[J]. Nucleic Acids Res, 1998, 26(11):2694-2701.

[12] ALLAWI H T, SANTALUCIA J, JR. Nearest-neighbor thermodynamics of internal A.C mismatches in DNA: sequence dependence and pH effects [J]. Biochemistry, 1998, 37(26):9435-9444.

[13] PEYRET N, SENEVIRATNE P A, ALLAWI H T, et al. Nearest-neighbor thermodynamics and NMR of DNA sequences with internal A.A, C.C, G.G, and T.T mismatches[J]. Biochemistry, 1999, 38(12):3468-3477.

[14] SANTALUCIA J, JR, HICKS D. The thermodynamics of DNA structural motifs [J]. Annu Rev Biophys Biomol Struct, 2004, 33:415-440.

[15] BOMMARITO S, PEYRET N, SANTALUCIA J, et al. Thermodynamic parameters for DNA sequences with dangling ends[J]. Nucleic Acids Res, 2000, 28(9):1929-1934.

[16] OHMICHI T, NAKANO S, MIYOSHI D, et al. Long RNA dangling end has large energetic contribution to duplex stability[J]. J Am Chem Soc, 2002, 124(35):10367-10372.

[17] TANAKA F, KAMEDA A, YAMAMOTO M, et al. Thermodynamic parameters based on a nearest-neighbor model for DNA sequences with a single-bulge loop[J]. Biochemistry, 2004, 43(22):7143-7150.

[18] SUGIMOTO N, NAKANO S, KATOH M, et al. Thermodynamic parameters to predict stability of RNA/DNA hybrid duplexes [J]. Biochemistry, 1995, 34(35):11211-11216.

[19] XIA T, SANTALUCIA J, JR, et al. Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson-Crick base pairs[J]. Biochemistry, 1998, 37(42):14719-14735.

[20] KIERZEK E, MATHEWS D H, CIESIELSKA A, et al. Nearest neighbor parameters for Watson-Crick complementary heteroduplexes formed between 2′-O-methyl RNA and RNA oligonucleotides[J]. Nucleic Acids Res, 2006, 34(13):3609-3614.

[21] DAVIS A R, ZNOSKO B M. Thermodynamic characterization of naturally occurring RNA single mismatches with G-U nearest neighbors[J]. Biochemistry, 2008, 47(38):10178-10187.

[22] MATHEWS D H, SABINA J, ZUKER M, et al. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure[J]. J Mol Biol, 1999, 288(5):911-940.

[23] LU Z J, TUMER D H, MATHEWS D H. A set of nearest neighbor parameters for predicting the enthalpy change of RNA secondary structure formation [J]. Nucleic Acids Res, 2006, 34(17):4912-4924.

[24] BRODA M, KIERZEK E, GDANIEC Z, et al. Thermodynamic stability of RNA structures formed by CNG trinucleotide repeats. Implication for prediction of RNA structure [J]. Biochemistry, 2005, 44(32):10873-10882.

[25] O′TOOLE A S, MILLER S, HAINES N, et al. Comprehensive thermodynamic analysis of 3′ double-nucleotide overhangs neighboring Watson-Crick terminal base pairs [J]. Nucleic Acids Res, 2006, 34(11):3338-3344.

[26] MILLER S, JONES L E, GIOVANNITTI K, et al. Thermodynamic analysis of 5′ and 3′ single-and 3′ double-nucleotide overhangs neighboring wobble terminal base pairs [J]. Nucleic Acids Res, 2008, 36(17):5652-5659.

[27] BLOSE J M, MANNI M L, KLAPEC K A, et al. Nonnearest-neighbor dependence of the stability for RNA bulge loops based on the complete set of group I single-nucleotide bulge loops[J]. Biochemistry, 2007, 46(51):15123-15135.

[28] WRIGHT D J, RICE J L, YANKER D M, et al. Nearest neighbor parameters for inosine *x* uridine pairs in RNA duplexes[J]. Biochemistry, 2007, 46(15):4625-4634.

[29] WATKINS N E, JR, SANTALUCIA J, et al. Nearest-neighbor thermodynamics of deoxyinosine pairs in DNA duplexes[J]. Nucleic Acids Res, 2005, 33(19):6258-6267.

[30] KAWAKAMI J, KAMIYA H, YASUDA K, et al. Thermodynamic stability of base pairs between 2-hydroxyadenine and incoming nucleotides as a determinant of nucleotide incorporation specificity during replication [J]. Nucleic Acids Res, 2001, 29(16):3289-3296.