

微生物多重PCR引物设计流程_引物设计与筛选

一. 背景简介

多重PCR：在一个PCR反应管中扩增多个目标片段。

主要难点：

- 引物相互作用，有些反应无法扩增
- 非特异结合和扩增
- 不同反应扩增效率有差异，导致均一性差

微生物多重PCR难点：

- 序列差异大
- 没有SNP数据
- 样本浓度可能有较大差异

解决方法：

- 使用简并引物
- 排除引物之间可能存在的相互作用，引物与Amplicon可能存在的相互作用
- 使用Universal Tag等方法提高扩增均一性

一. 使用JCVI Primer Designer生成候选简并引物

该软件用于生成微生物扩增用高通量测序引物，可生成简并引物，J. Craig Venter Institute在2008年发布的开源软件^[1]。

要求序列简并度最好<13%，使用过程中显示序列简并度<20%也可使用。

需在Linux操作系统下使用。按照说明书按照软件后仍可能无法直接使用，需修改一下代码。

主要特点：

- 将简并引物还原为ATGC序列再设计引物，并检测非特异扩增，发卡结构，引物退火温度，引物相互作用等。
- 使用Primer3生成候选引物，检测引物相互作用
- 调用EMBOSS: palindrome，检测回文序列，避免发卡结构影响
- 调用NCBI BLAST toolkit: dust，排除低复杂度序列
- 对于较大的基因组（简并度较低），调用NCBI BLAST: blastall, bl2seq, formatdb，检测引物非特异结合
- 对于较小的基因组（简并度较高），调用EMBOSS：Dreg检测引物非特异结合

GAGTACTTCCTGTAGCATCCTTA and GAGAAGTTCTACAGGAAGAT violate rule No.3 the score is 10

```

GAGTACTTCCTGTAGCATCCTTA
| | | | | | | | | |
TAGAAGGACATCTTGAAGAG

```

4. 两引物任意位置 (with one mismatch) 互补碱基<12

ACGCCATCTTCATTACACCAT and GCTCCCAATTGTGTGAATGAAG violate rule No.4 the score is 14

```

ACGCCATCTTCATTACACCAT
| | | | | | | | | |
GAAGTAAGTGTGTTAACCCTCG

```

5. 两引物任意区域 (min = 14bp) 互补碱基比例<75%

GGAGTCAAACATGATGGACACCT and CCAACATGCTTTACCCCATACC violate rule No.5 the score is 0.76

```

GGAGTCAAACATGATGGACACCT
| | | | | | | |
CCATACCCCATTTCTGACAACC

```

根据第二篇参考文献加入了检测alignment score的方法，排除alignment score >= 8的引物组合。

6. 两引物alignment score < 8

TACCAACCCAACCATTATACAT and ATTGGGTTGACAGGTGCAGT violate rule No.6 the score is 8

```

TACCAACCCAACCATTATACAT
| | | | | | | |
TGACGTGGACAGTTGGGTTA

```

alignment score计算方法：

- match : score +1
- mismatch open : score -2
- mismatch extension : score -1

2.2 检测引物与Amplicon之间的相互作用

引物3'端与amplicon间的互补碱基数<13.(实际应用中，因为引物和amplicon均为简并序列，因此，设置为引物3'末端14bp序列与amplicon比对分数<12)

Score反映Primer与Amplicon结合的概率

Score = (Primer and Amplicon in common) / Amplicon

Score <=1

Primer	Amplicon	Score
N	A	1
A	N	1/4
B	D	2/3

B: [CGT], D: [AGT]

Found one match between the primer and amplicon above threshold, the match score is 12.0.

The primer 3 end seq is ATCHGGNTTYCCAT it is on forward strand.

The amplicon seq is AYCAGGGTTTYYAY

Find 1 matches of primer Norovirus_GI#F#0 in amplicon Adenovirus_F-G

三. 引物筛选流程及主要操作步骤

3.1 主要用途：

- 设计多重PCR引物，用于同时扩增多个靶点，每个靶点设置为一个组别
- 将简并引物还原为ATGC, 计算退火温度范围，排除退火温度不合适的引物
- 检测引物之间的相互作用；
- 检测引物-Amplicon间的相互作用；
- 设置Amplicon Size，以及Amplicon间的最小差值，排除扩增大小不合适的引物组合；
- 将经检测符合以上要求的引物进行组合，输出可行组合。

3.2 输入文件：

候选引物文件，fasta格式

```
1. >组别#F/R#序号
2. 引物序列
3. 例如：
4. >Norovirus_GI#F#3
5. GARAARTTYTACAGRAAGAT
6. >Norovirus_GI#R#4
7. RACCCARCCATTTRTACATY
8. >Rotavirus_A#F#0
9. GGCWTTTAATGCTTTTCAGT
10. >Rotavirus_A#R#0
11. TTWACRCCWGARTCATCCAT
12. ....
```

Amplicon文件，fasta格式，包括候选引物所在区域

```
1. >Amplicon名称
2. Amplicon序列
3. 例如：
4. >SapoVirus_GI_consensus_seq
5. RGWYGTGACCYKCCTGGCRCRYDGGYCCGRCCACATCCMAYGTTGTTGTTKKCTAATCCRGARCAACCCAATGGGS
6. ....
```

3.3 运行脚本

如果候选引物文件名为test_primers.fasta，Amplicon文件名为test_input_region.fasta
将上述文件拷贝至脚本程序所在文件夹，打开cmd窗口，输出下列命令运行脚本

1. python 脚本名 候选引物文件名 (Amplicon文件名)
2. 例如:
3. python PrimerCheck.py test_primers.fasta
4. PrimerSelector_Group.py. test_primers.fasta
5. python PrimerSelector_Amplicon.py test_primers.fasta test_input_region.fasta

3.4 输出文件

将候选引物文件从简并引物转变为ATGC之后的文件

1. 文件名为:
2. 候选引物文件名_atgc.fasta
3. 例如:
4. test_primers_atgc.fasta
5. >Norovirus_GI#F#0#0
6. CTGCCATCTGGGTTTCCAT
7. >Norovirus_GI#F#0#1
8. CTGCCATCTGGGTTCCCAT
9. >Norovirus_GI#F#0#2
10. CTGCCATCTGGTTTTCAT
11. >Norovirus_GI#F#0#3
12. CTGCCATCTGGTTTCCCAT
13. >Norovirus_GI#F#0#4
14. CTGCCATCTGGCTTTCCAT
15.

引物相关信息，包括引物长度、简并度、Tm值范围，Flag标记等。如果简并度或Tm范围，不符合预先设定的值，则Flag标记变为0，被排除掉，不进行后面的运算。

1. 文件名为:
2. 候选引物文件名_information.txt
3. 例如:
4. test_primers_primer_information.txt

Name	Sequence	Len	DegNumber	Max Tm	Min Tm	Flag
Adenovirus_fg#F#8	CTACTCYGGCTCCRTYC CMTAC	22	16	66.13	58.32	0
Astrovirus_1-8#F#1	GGGRARCTCCTRTGCT AYCAGT	22	16	66.33	57.63	0
Astrovirus_1-8#R#2	GRCTTRCTAGCCATCRC ACTYCT	23	16	67.72	58.26	0
SapoVirus_GI#F#4	GGYCCGRCCACATCCM AY	18	16	66.45	57.19	0
Norovirus_GI#F#0	CTNCCATCHGGNTTYCC AT	19	96	63.02	51.79	1
Norovirus_GI#R#0	GTKGAVACWATYTCATC ATCACC	23	24	59.96	54.62	1
Norovirus_GI#R#1	ATCATCATTTACRWAWT CGG	20	8	52.02	48.87	1
Norovirus_GI#F#1	WGGDTGGCARGCCATG TT	18	12	62.51	57.38	1

.....

候选引物相互作用报告

```

1. 文件名为:
2. 候选引物文件名_confliction_report.fasta
3. 例如:
4. test_primers_confliction_report.fasta
5. 主要包括:
6. 1. 引物在Amplicon上的位置
7. SapoVirus_GI#R#3
8. ('SapoVirus_GI_consensus_seq', 'reverse', 816)
9. Norovirus_GI#F#3
10. ('Norovirus_GI_consensus_seq', 'forward', 478)
11. Adenovirus_f-g#R#0
12. ('Adenovirus_F-G', 'reverse', 329)
13. Astrovirus_1-8#F#4
14. ('Astrovirus_1-8_consensus_seq', 'forward', 281)
15. ....
16.
17. 2. 候选引物间可能存在的相互作用
18. The confliction of primers are :
19.
20. GGAAACTCCTATGCTATCAGTTG and TTGATACCATAGATTGGAA violate rule No.6 the score is
21. GGAAACTCCTATGCTATCAGTTG
22.      ||||| |||||
23.      AAGGTTAGATACCATAGTT
24. The penalty_score is 1
25. Rotavirus_A#F#2#5 and Astrovirus_1-8#F#3#14 may have interaction.
26. ATGTTGGGATGAAGCGATATAGA and TCAGTCATCCAATTTCTAT violate rule No.1 the score is
27. ATGTTGGGATGAAGCGATATAGA
28.      |||||
29.      TATCTTTAACCTACTGACT
30. The penalty_score is 1
31. Rotavirus_A#R#1#19 and SapoVirus_GI#R#0#0 may have interaction.
32. ....
33.
34. 3. 各引物的罚分
35. The penalty score of degenerate primers are :
36. SapoVirus_GI#R#1                10
37. Adenovirus_f-g#F#0              65
38. Astrovirus_1-8#F#7              32
39. Astrovirus_1-8#F#6              276
40. Adenovirus_f-g#F#4               4
41. Rotavirus_A#F#2                 84
42. Astrovirus_1-8#R#1              40
43. Norovirus_GII#R#0               194
44. Astrovirus_1-8#F#1              144
45. ....

```

引物相互作用的表格

1. 文件名为:
2. 候选引物名_confliction_table.csv
3. 例如:
4. test_primers_confliction_table.csv

Confl iction _Tabl _e	Sapo Virus _GI#R #3	Noro virus _GI#F #3	Adeno virus_ f-g#R# 0	Astro virus_ 1-8#F# 4	Norov irus_ GI#F# 1	Adeno virus_ f-g#R# 4	SapoV irus_ GI#R# 2
SapoV irus_G I#R#3	2	0	1	0	0	0	0
Norov irus_G I#F#3	0	2	1	0	0	0	0
Adeno virus_ f-g#R# 0	1	1	0	0	0	0	0
Astro virus_ 1-8#F# 4	0	0	0	2	0	0	0
Norov irus_G I#F#1	0	0	0	0	0	0	0

.....

候选引物组合


```

1. 文件名为:
2. 候选引物名_combinations.txt
3. 例如:
4. test_primers_combinations.txt
5. 主要包括:
6. 1. 可选引物组合:
7. Primer Combinations Without Interaction:
8. Amplicon Length are : [115, 179, 238, 320, 398, 532]
9. >Norovirus_GI#F#1
10. WGGDTGGCARGCCATGTT
11. >SapoVirus_GI#R#1
12. GGTGGTAYGACAGAAGCAATGAT
13. >Adenovirus_f-g#R#7
14. CACGTTGTAYCCTTCMCCRTC
15. ....
16.
17. 2. 自身存在相互作用的引物:
18. The self interaction primers are:
19. >Norovirus_GI#F#0
20. CTNCCATCHGGNTTYCCAT
21. >Norovirus_GI#F#3
22. GARAARTTYTACAGRAAGAT
23. >SapoVirus_GI#R#3
24. ATTGGRTRTRACNGGTGCVGT
25. >Adenovirus_f-g#R#8
26. AAACCCTGRTAKCCRATRRTTGTA
27. >SapoVirus_GI#R#2
28. AGTGATRCGAGCATCAACGAAAG
29. ....
30.
31. 3. 尝试引物组合数以及可用引物组合数:
32. The total combination number is 846720, the combination number without intera
33. The primers can be used are:
34. >Adenovirus_f-g#F#0
35. TACATGAACGGKCGKGTKRC
36. >Norovirus_GII#F#3
37. GCTCCCRRTKKTGTGAATGAAG
38. >Adenovirus_f-g#F#6
39. CMAACATGCTKTAYCCCATHCC
40. >Adenovirus_f-g#F#2
41. GCATYAACTGTAYGCYAACTTTTT
42. >Norovirus_GI#F#1
43. WGGDTGGCARGCCATGTT
44. ....
45.
46. 4. 可用引物出现次数:
47. The primers can be used and there appearance value are:
48. Adenovirus_f-g#F#0 6
49. Norovirus_GII#F#3 56
50. Adenovirus_f-g#F#6 39
51. Adenovirus_f-g#F#2 55
52. Norovirus_GI#F#1 311
53. Adenovirus_f-g#R#4 77

```

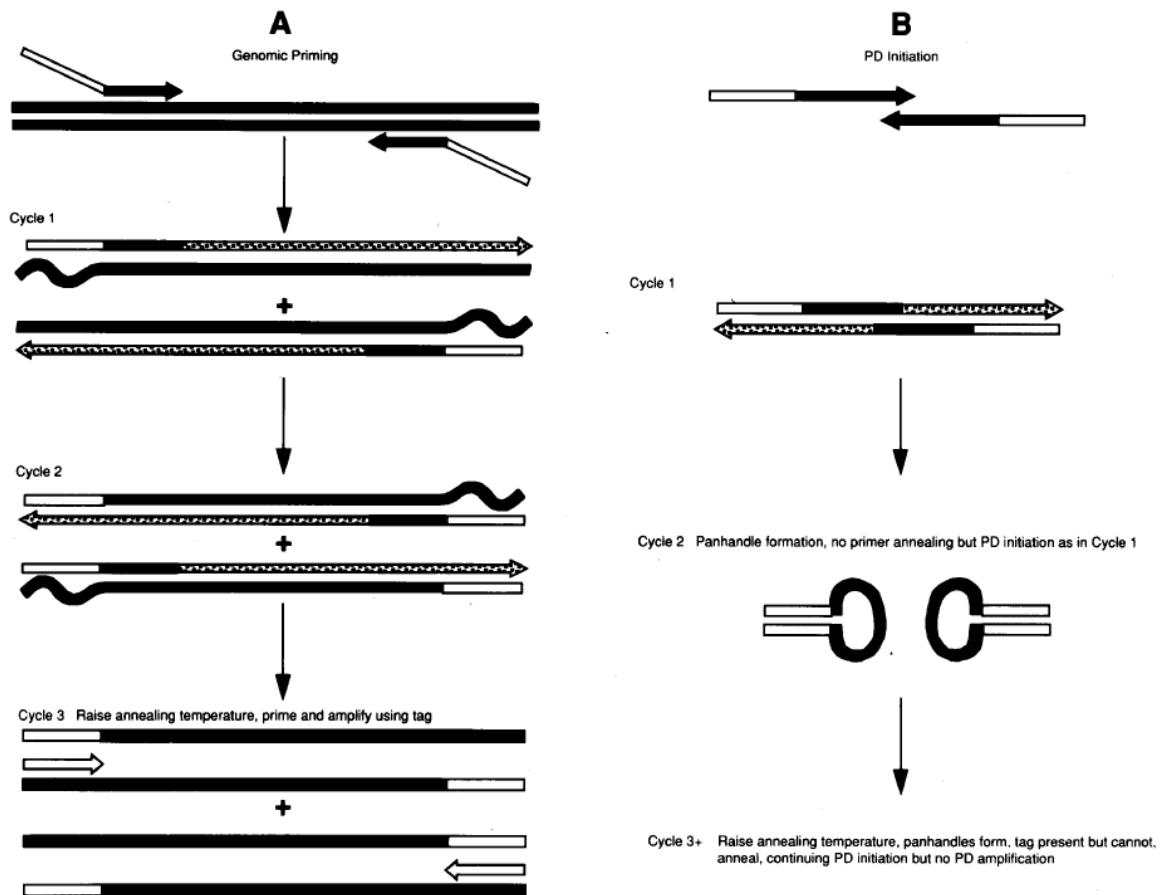
```

54. Norovirus_GII#F#2      80
55. Rotavirus_A#R#0      311
56. Norovirus_GII#R#0    172
57. Adenovirus_f-g#R#0   56
58. .....
59.
60. 5. 引物-Amplicon相互作用, >1则可能有引物与其他Amplicon有非特异结合
61. Primer_Amplcion interaction number are:
62. Sapovirus_GI#R#3      1
63. Norovirus_GI#F#3      1
64. Adenovirus_f-g#R#0    1
65. Norovirus_GII#R#2     2
66. Adenovirus_f-g#R#5    1
67. .....
68.
69. 6. 没有出现非特异结合的引物
70. Primers with unspecific interactions are:
71. >Norovirus_GI#R#4
72. >RACCCARCCATTTRTACATY
73. >Astrovirus_1-8#R#5
74. >GAGTRCTYCCWGTAGCRTCCTTA
75. >Norovirus_GI#R#0
76. >GTKGAVACWATYTCATCATCACC
77. >Astrovirus_1-8#R#3
78. >TTACGGACACGYTGAKWAGAYTG
79. >Adenovirus_f-g#R#8
80. >AAACCCTGRTAKCCRATRTTGTA
81. .....

```

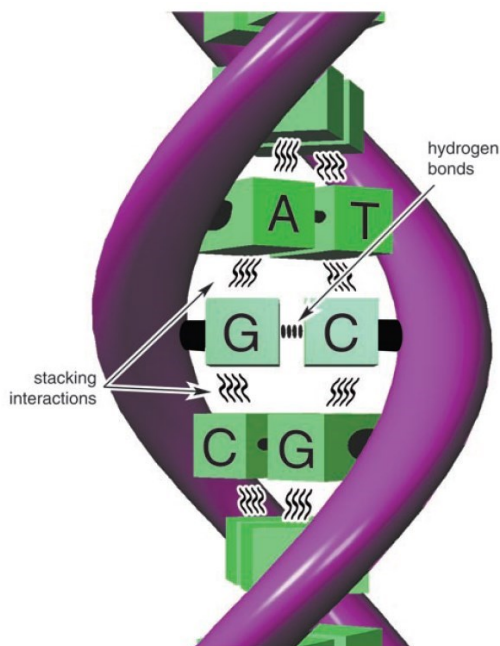
四. 实验方面的建议

1. 先进行单管扩增验证引物，从目前实验结果来看，单管扩增成功的都能实现共同扩增。
2. 引物浓度对扩增的影响并不大，建议采用0.1uM*简并度为初始引物浓度。
3. 对于AT含量较高的Amplicon，多重PCR中使用dUTP可能无法很好地扩增。在dUTP中加入少量dNTP(1/8),可以较好地实现同时扩增。
4. 提高Mg²⁺浓度，降低退火温度等有助于共同扩增，但可能出现较多引物二聚体。
5. 使用F,R相同Universal Tag (HANDS系统^[5]) 有助于减少引物二聚体，可能对提高扩增均一性和检测限有帮助。



五. 一些改进

仅根据碱基互补配对，检测引物间的相互作用。相当于仅考虑碱基之间的氢键。没有考虑碱基堆积力等作用力，也没有考虑退火温度、 Mg^{2+} 浓度、dNTP浓度等对反应的影响。



Nearest Neighbor Model：研究分子热动力学的模型，认为两条链上碱基的相互作用，除了与自身相关，与邻近碱基也相关。

- 是目前主流引物设计软件预测引物相互作用，计算引物Tm值的方法。
- 将碱基堆积力纳入计算，在某些情况下，可能具有更好的效果。
- 很多人进行过相关研究，计算得到的参数不尽相同，可能应用范围也有所不同。

将MPprimer中计算delta G值的方法补充到脚本中用于运算^[6]。MPprimer使用SantaLucia的模型进行计算，主要参数及计算方法如下^[7]：

$$\Delta G_{37}^{\circ}(\text{total}) = \Delta G_{37}^{\circ} \text{initiation} + \Delta G_{37}^{\circ} \text{symmetry} + \sum \Delta G_{37}^{\circ} \text{stack} + \Delta G_{\text{AT terminal}}^{\circ}$$

$$\begin{array}{l} 5'\text{-CGTTGA-3'} = \Delta G_{37}^{\circ} \text{initiation} + \Delta G_{37}^{\circ} \text{symmetry} \\ 3'\text{-GCAACT-5'} \quad + \text{CG} + \text{GT} + \text{TT} + \text{TG} + \text{GA} + \text{AT}_{\text{terminal}} \\ \quad \quad \quad \text{GC} \quad \text{CA} \quad \text{AA} \quad \text{AC} \quad \text{CT} \end{array}$$

$$\Delta G_{37}^{\circ} (\text{predicted}) = 1.96 + 0 - 2.17 - 1.44 - 1.00 - 1.45 - 1.30 + 0.05$$

$$\Delta G_{37}^{\circ} (\text{predicted}) = -5.35 \text{ kcal mol}^{-1}.$$

TABLE 1 Nearest-neighbor thermodynamic parameters for DNA Watson-Crick pairs in 1 M NaCl

Propagation sequence	ΔH° (kcal mol ⁻¹)	ΔS° (e.u.)	ΔG_{37}° (kcal mol ⁻¹)
AA/TT	-7.6	-21.3	-1.00
AT/TA	-7.2	-20.4	-0.88
TA/AT	-7.2	-21.3	-0.58
CA/GT	-8.5	-22.7	-1.45
GT/CA	-8.4	-22.4	-1.44
CT/GA	-7.8	-21.0	-1.28
GA/CT	-8.2	-22.2	-1.30
CG/GC	-10.6	-27.2	-2.17
GC/CG	-9.8	-24.4	-2.24
GG/CC	-8.0	-19.9	-1.84
Initiation	+0.2	-5.7	+1.96
Terminal AT penalty	+2.2	+6.9	+0.05
Symmetry correction	0.0	-1.4	+0.43

TABLE 2 Nearest-neighbor delta G₃₇[°] increments (kcal mol⁻¹) for internal single mismatches next to Watson-Crick pairs in 1 M NaCl

Propagation sequence	X	Y			
		A	C	G	T
GX/CY	A	0.17	0.81	-0.25	WC
	C	0.47	0.79	WC	0.62
	G	-0.52	WC	-1.11	0.08
	T	WC	0.98	-0.59	0.45
CX/GY	A	0.43	0.75	0.03	WC
	C	0.79	0.70	WC	0.62
	G	0.11	WC	-0.11	-0.47
	T	WC	0.40	-0.32	-0.12
AX/TY	A	0.61	0.88	0.14	WC
	C	0.77	1.33	WC	0.64
	G	0.02	WC	-0.13	0.71
	T	WC	0.73	0.07	0.69
TX/AY	A	0.69	0.92	0.42	WC
	C	1.33	1.05	WC	0.97
	G	0.74	WC	0.44	0.43
	T	WC	0.75	0.34	0.68

采用MPprimer中默认的盐浓度等参数，设置cutoff值为-9kcal/mol⁻¹，两引物deltaG value > -9kcal，则标记为可能存在相互作用。从目前结果来看，基本与rule No.6及rule No.5一致，但也可以检测到一些之前没有的相互作用。

ATTGGATTGACCGGTGCCGT and ATTGGGTTGACGGGTGCGGT violate rule No.7 the score is -9.10

```

ATTGGATTGACCGGTGCCGT
      ||||  ||||
      TGGCGTGGGCAGTTGGGTTA

```

六. 相关资源及参考文献

1. JCVI Primer Designer:
<https://sourceforge.net/projects/primerdesigner/>
2. MPprimer:
<https://sourceforge.net/projects/mpprimer/>
<http://biocompute.bmi.ac.cn/MPprimer/>
3. Thermo Fisher Multiple Primer Analyzer:
<https://www.thermofisher.com/cn/zh/home/brands/thermo-scientific/molecular-biology/molecular-biology-learning-center/molecular-biology-resource-library/thermo-scientific-web-tools/multiple-primer-analyzer.html>
4. Primer Pooler:
<http://people.ds.cam.ac.uk/ssb22/pooler/>

[1]Li K, Brownley A, Stockwell T B, et al. Novel computational methods for increasing PCR primer design effectiveness in directed sequencing[J]. BMC bioinformatics, 2008, 9(1): 191.

[2]Wang H Y, Luo M, Tereshchenko I V, et al. A genotyping system capable of simultaneously analyzing> 1000 single nucleotide polymorphisms in a haploid genome[J]. Genome research, 2005, 15(2): 276-283.

- [3]Schoske R, Vallone P M, Ruitberg C M, et al. Multiplex PCR design strategy used for the simultaneous amplification of 10 Y chromosome short tandem repeat (STR) loci[J]. *Analytical and Bioanalytical Chemistry*, 2003, 375(3): 333-343.
- [4]Brownie J, Shawcross S, Theaker J, et al. The elimination of primer-dimer accumulation in PCR[J]. *Nucleic acids research*, 1997, 25(16): 3235-3241.
- [5]Shen Z, Qu W, Wang W, et al. MPprimer: a program for reliable multiplex PCR primer design[J]. *BMC bioinformatics*, 2010, 11(1): 143.
- [6]SantaLucia Jr J, Hicks D. The thermodynamics of DNA structural motifs[J]. *Annu. Rev. Biophys. Biomol. Struct.*, 2004, 33: 415-440.