Supplementary Information

The 5-Me of thyminyl (T) interaction with the Neighboring Nucleobases dictate the relative stability of isosequential DNA/RNA hybrid duplexes

Subhrangsu Chatterjee, Wimal Pathmasiri,& Jyoti Chattopadhyaya*

Department of Bioorganic Chemistry, Box 581, Biomedical Center,Uppsala University, S-751 23 Uppsala, Sweden, E-mail: jyoti@boc.uu.se

Table S1. Calculation of ΔpK_a and $\Delta\Delta G^o_{p\text{Ka of each mid bp}}$ (kJ mol^{-1})† values from the model monomeric donors and acceptors (*i.e.* Monomeric base-pair) representing the H-bonding base pairing contribution of the middle base-pairs (mid base-pair *i.e.* excluding the terminal base pairs).The net hydrogen bonding stabilization in different base pairs are different. The $\Delta\Delta\Delta G^o_{pKa}$ values are predicting the relative hydrogen bonding stabilization of one base pair over another.

Monomeric base-pair	dC-rG	rC-dG	dG-rC	rG-dC	dA-U	rA-T	T-rA	U-dA
$\Delta pK_a{}^\ddagger$	4.94	5.35	5.35	4.94	5.43	6.43	6.43	5.43
$\Delta\Delta G^o_{p\text{Ka of each mid bp}}$	28.20	30.60	30.60	28.20	31.1	36.4	36.4	31.1
$[\Delta\Delta pK_a]_{\text{(DR)-(RD)}}$	-0.41		+0.41		-1.00		+1.00	
$\Delta\Delta\Delta G^o_{p\text{Ka of each mid bp}}$	-2.40		+2.40		-5.30		+5.30	

†See Table S2 for calculation of total free energies for base-pairing using free energies for individual middle base-pairs . ‡ For the free energy determination of the middle bp residues in duplex the ΔG^o_{pKa} values of the Etp(d/rN)pEt (**6a** – **10a**) and (**6b** – **10b**) (ref 14) have been used.

$^\ddagger\Delta pK_a$ for (rG-dC): $[pK_a]_{7b}$ - $[pK_a]_{10a}$, for (dG-rC): $[pK_a]_{7a}$ - $[pK_a]_{10b}$, for (rC-dG): $[pK_a]_{7a}$ - $[pK_a]_{10b}$, for (dC-rG): $[pK_a]_{7b}$ - $[pK_a]_{10a}$ for (U-dA): $[pK_a]_{8b}$ - $[pK_a]_{6a}$ and for (T-rA): $[pK_a]_{9a}$ - $[pK_a]_{6b}$; (dA-U): $[pK_a]_{8b}$ - $[pK_a]_{6a}$ and for (rA-T): $[pK_a]_{9a}$ - $[pK_a]_{6b}$; $[\Delta\Delta pK_a]_{\text{(DR)-(RD)}} = [\Delta pK_a]_{\textbf{DR}}$ - $[\Delta pK_a]_{\textbf{RD}}$

Table S2. Calculation of the $\Delta\Delta\Delta G^{o}_{\text{pKa of each terminal bp}}$ (kJ mol^{-1})† values from the model monomeric donors and acceptors to represent the H-bonding contribution of the base pairing of terminal base-pairs (3´→5´) in DNA-RNA and RNA-DNA hybrid duplexes.

Terminal base-pair	dC$^{3'}$-rG$^{5'}$	rC$^{3'}$-dG$^{5'}$	dG$^{3'}$-rC$^{5'}$	rG$^{3'}$-dC$^{5'}$	dA$^{3'}$-U$^{5'}$	rA$^{3'}$-T$^{5'}$	T$^{3'}$-rA$^{5'}$	U$^{3'}$-dA$^{5'}$
ΔpK_a	4.92	5.16	5.75	5.15	5.38	6.24	7.01	5.91
$\Delta\Delta$G $_{\text{p}K\text{a for terminal base-pairs}}$	28.1	29.4	32.9	29.5	30.8	35.3	40.0	33.80
$\Delta\Delta$ pK_a [**DR**-**RD**]	-0.24		0.60		-0.86		1.10	
$\Delta\Delta\Delta$G $_{\text{p}K\text{a}}$ [**DR**-**RD**]	-1.3		3.4		-4.5		6.2	

† See Table S2 for calculation of total free energies for base-pairing using free energies for individual terminal base-pairs. For the 3'- or 5'-terminal base-pairing in duplex the $\Delta G^{o}_{\text{pK}_a}$ values of the (d/rN)pEt (**1a** – **5a**) and (**1b** – **5b**) and Etp(d/rN)pEt (**6a** – **10a**) and (**6b** – **10b**) (ref 14) have been used. We have used Etp(d/rN)pEt (**6a** – **10a**) and (**6b** – **10b**) as the model for the 5'-phosphate since 3'-phosphate group in the bis-phosphate do not have any influence on the pK_a of the constituent nucleobase.

Table S3. Calculation for $\Sigma\Delta pK_a$ for all fourteen sequences (**1** - **14**)[4] in **RR**, **DD**, **DR**, and **RD** combination (shown in the right four columns) based on the pK_as of the middle and terminal base pairs The H-bonding in the **RR** duplexes are generally more stable (with some rare exception, *e.g.* **8$_{RR}$/8$_{DD}$/8$_{DR}$/8$_{RD}$**) than those of the **DD**, **DR** or **RD** because their $\Sigma\Delta pK_a$ are relatively smaller.

Sequence [a]	Duplex Sequences [a]	$\Sigma\Delta pK_a$ RR [c]	$\Sigma\Delta pK_a$ DD [c]	$\Sigma\Delta pK_a$ DR [c]	$\Sigma\Delta pK_a$ RD [c]
(1$_{RR}$/1$_{DD}$/1$_{DR}$/1$_{RD}$)	5'-TCCCTCCTCTCC 3'-AGGGAGGAGAGG	62.38	66.70	65.03	64.28
(2$_{RR}$/2$_{DD}$/2$_{DR}$/2$_{RD}$)	5'-CCTTCCCTT 3'-GGAAGGGAA	48.35	52.07	51.23	49.35
(3$_{RR}$/3$_{DD}$/3$_{DR}$/3$_{RD}$)	5'-TTCCCTTCC 3'-AAGGGAAGG	47.26	50.98	50.21	48.23
(4$_{RR}$/4$_{DD}$/4$_{DR}$/4$_{RD}$)	5'-GCTCTCTGGC 3'-CGAGAGACCG	51.85	55.17	54.89	52.30
(5$_{RR}$/5$_{DD}$/5$_{DR}$/5$_{RD}$)	5'-CTCGTAC CTTCCGGTCC 3'-GAGCATGGAAGGCCAGG	89.02	95.42	93.36	91.41
(6$_{RR}$/6$_{DD}$/6$_{DR}$/6$_{RD}$)	5'-CTCGTACCTTTCCGGTCCC 3'-GAGCATGG AAAGGCCAGG	94.55	101.71	99.79	96.84
(7$_{RR}$/7$_{DD}$/7$_{DR}$/7$_{RD}$)	5'-TAGTTATCTCTATCT 3'-ATCAATAGAGATAGA	81.56	90.44	88.29	84.15
(8$_{RR}$/8$_{DD}$/8$_{DR}$/8$_{RD}$)	5'-GCACAGCC 3'-CGTGTCGG	41.28	43.64	41.11	43.93
(9$_{RR}$/1$_{DD}$/1$_{DR}$/1$_{RD}$)	5'-GAGCTCCCAGGC 3'-CTCGAGGGTCCG	61.93	65.65	63.18	64.59
(10$_{RR}$/10$_{DD}$/10$_{DR}$/10$_{RD}$)	5'GCCGAGGTCCATGTCGTACGC 3'-CGGCTCCAGGTACAGCATGCG	109.25	117.01	113.83	112.83
(11$_{RR}$/11$_{DD}$/11$_{DR}$/11$_{RD}$)	5'-TGTACGTCACAACTA 3'-ACATGCAGTGTTGAT	79.91	87.67	83.09	84.91
(12$_{RR}$/12$_{DD}$/12$_{DR}$/12$_{RD}$)	5'-TATACAAGTTATCTA 3'-ATATGTTCAATAGAT	81.38	90.82	86.15	86.56
(13$_{RR}$/13$_{DD}$/13$_{DR}$/13$_{RD}$)	5'-CGACTATGCAAAAAC 3'-GCTGATACGTTT TTG	80.41	88.09	81.54	87.36
(14$_{RR}$/14$_{DD}$/14$_{DR}$/14$_{RD}$)	5'-CGCAAAAAAAAAACGC 3'-GCGTTTTTTTTTTGCG	85.94	94.38	84.97	95.79

[a] Sequences (**1** - **14**) are taken from ref 4. The type of duplexes (*i.e.,* whether **RR**, **DD**, **RD** or **DR**) are referred to as subscript to the sequence number in the text, for example, as **1$_{RR}$**, **1$_{DD}$**, **1$_{RD}$**, and **1$_{DR}$**. and so on. [c] $\Sigma\Delta pK_a$ includes the sum of pK_a differences of the middle (Table S4) as well as the terminal base-pairs (Table S5) for **RR/DD/DR** and **RD** duplexes.

Table S4. The total number of the middle (Mid) and the terminal (Term, both $^{3'}_{5'}$ and $^{5'}_{3'}$) base pairing [dG-rC, dC-rG, T-rA, dA-U in DR and rG-dC, rC-dG, rA-T, U-dA in RD] in duplexes $(1 – 14)^4$ with their ΣΔpK_a values (DR and RD) of middle plus terminal base-pairs.

Seq No.	Duplex Sequence [a]	Total (T/U-A) bp [b]	Total (A-T/U) bp [c]	Mid (T/U-A) bp [d]	Mid (A-T/U) bp [e]	Mid (G-C) bp [f]	Mid (C-G) bp [g]	Term 5'/3' bp	Term 3'/5' bp	ΣΔpK_a DR [h]	ΣΔpK_a RD [i]
(1DR/1RD)	5'-TCCCTCCCTCTCC-3' 3'-AGGGAGGGAGAGG-5'	4	0	3	0	0	7	T/A	C/G	65.03	64.28
(2DR/2RD)	5'-CCTTCCCTT-3' 3'-GGAAGGGAA-5'	4	0	3	0	0	4	C/G	T/A	51.23	49.35
(3DR/3RD)	5'-TTCCCTTCC-3' 3'-AAGGGAAGG-5	4	0	3	0	0	4	T/A	C/G	50.21	48.23
(4DR/4RD)	5'-GCTCTCTGGC-3' 3'-CGAGAGACCG-5'	3	0	3	0	2	3	G/C	C/G	54.89	52.30
(5DR/5RD)	5'-CTCGTAC CTTCCGGTCC-3' 3'-GAGCATGGAAGGCCAGG-5'	5	1	5	1	3	6	C/G	C/G	93.36	91.41
(6DR/6RD)	5'-CTCGTACCTTTCCGGTCC-3' 3'-GAGCATGG AAAGGCCAGG-5'	6	1	6	1	3	6	C/G	C/G	99.79	96.84
(7DR/7RD)	5'-TAGTTATCTCTATCT- 3' 3'-ATCAATAGAGATAGA-5'	8	3	6	3	1	3	T/A	T/A	88.29	84.15
(8DR/8RD)	5'-GCACAGCC-3' 3'-CGTGTCGG-5'	0	2	0	2	1	3	G/C	C/G	41.11	43.93
(9DR/9RD)	5'-GAGCTCCCAGGC-3' 3'-CTCGAGGGTCCG-5'	1	2	1	2	3	4	G/C	C/G	63.18	64.59
(10DR/10RD)	5'-GCCGAGGTCCATGTCGTACGC-3' 3'-CGGCTCCAGGTACAGCATGCG-5'	4	3	4	3	6	6	G/C	C/G	113.83	112.83
(11DR/11RD)	5'-TGTACGTCACAACTA-3' 3'-ACATGCAGTGTTGAT-5'	4	5	3	4	2	4	T/A	A/T	83.09	84.91
(12DR/12RD)	5'-TATACAAGTTATCTA-3' 3'-ATATGTTCAATAGAT-5'	6	6	5	5	1	2	T/A	A/T	86.15	86.56
(13DR/13RD)	5'-CGACTATGCAAAAC-3' 3'-GCTGATACGTTT TTG-5	2	7	2	7	2	2	C/G	C/G	81.54	87.36
(14DR/14RD)	5'-CGCAAAAAAAACGC-3' 3'-GCGTTTTTTTTGCG-5'	0	10	0	10	2	2	C/G	C/G	84.97	95.79

[a]The DNA-RNA (**DR**) and RNA-DNA (**RD**) duplexes, **1** – **14**, are taken from ref. 4 and $\Sigma \Delta pK_a$ for **DR** and **RD** are used in the plot for Figure 3. [b] The total number of middle T/U-A basepairs (bp) for **DR** or **RD** . [c]A-T/U basepairs (bp) for **DR** or **DR** duplexes in each sequence. [a] total no of middle T/U-A base-pairs. [e] total no. of A-T/U base pairs. [f] total G-C bp (mid) in **DR** and **RD** (**1-14**)[4], [g] total C-G base pairing (mid) in **DR** or **RD**. [h] $\Sigma \Delta pK_a$ (**DR**) = [{ΔpK_a (T-A)}*total no of mid T-A bp + {ΔpK_a (A-U)}*total no of mid A-U bp} + {ΔpK_a (dC-G)}*total no of mid dC-G bp + {ΔpK_a (dG-C)}*total no of mid dG-C bp + {ΔpK_a terminal base-pair}(5´3)+ {ΔpK_a terminal base-pair}´(3´5)]+. [i] $\Sigma \Delta pK_a$ (**RD**) = [{ΔpK_a (U-A)}*total no of mid U-A bp + {ΔpK_a (A-T)}*total no of mid A-T bp} + {ΔpK_a (C-dG)}*total no of mid C-dG bp + {ΔpK_a (G-dC)}*total no of mid G-dC bp + {ΔpK_a terminal base-pair}(5´3)+ {ΔpK_a terminal base-pair}(3´5´)]

Table S5 : Calculation of total Gibb's free energy for base pairing of all **DR** and **RD** (1-14)[4] reference duplexes.

Seq.No	Duplex Sequence[a]	$[\Delta G^o{}_{bp}]$ DR kcal mol^{-1} [b]	$[\Delta G^o{}_{bp}]$ RD kcal mol^{-1} [c]	$[\Delta\Delta G^o{}_{bp}]$ DR-RD kcal mol^{-1} [d]
(1$_{DR}$/1$_{RD}$)	5'-TCCCTCCTCTCC-3' 3'-AGGGAGGAGAGG-5'	88.10	87.55	0.55
(2$_{DR}$/2$_{RD}$)	5'-CCTTCCCTT-3' 3'-GGAAGGGAA-5'	69.40	67.24	2.16
(3$_{DR}$/3$_{RD}$)	5'-TTCCCTTCC-3' 3'-AAGGGAAGG-5	67.95	65.69	2.26
(4$_{DR}$/4$_{RD}$)	5'-GCTCTCTGGC-3' 3'-CGAGAGACCG-5'	74.40	71.19	3.21
(5$_{DR}$/5$_{RD}$)	5'-CTCGTAC CTTCCGGTCC-3' 3'-GAGCATGGAAGGCCAGG-5'	126.60	124.38	2.22
(6$_{DR}$/6$_{RD}$)	5'-CTCGTACCTTTCCGGTCC-3' 3'-GAGCATGG AAAGGCCAG5'	135.26	131.79	3.47
(7$_{DR}$/7$_{RD}$)	5'-TAGTTATCTCTATCT-3' 3'-ATCAATAGAGATAGA-5'	119.57	114.38	5.19
(8$_{DR}$/8$_{RD}$)	5'-GCACAGCC-3' 3'-CGTGTCGG-5'	55.93	59.60	-3.19
(9$_{DR}$/9$_{RD}$)	5'-GAGCTCCCAGGC-3' 3'-CTCGAGGGTCCG-5'	85.88	87.71	-1.83
(10$_{DR}$/10$_{RD}$)	5'GCCGAGGTCCATGTCGTACGC-3' 3'CGGCTCCAGGTACAGCATGCG-5'	154.57	153.31	1.26
(11$_{DR}$/11$_{RD}$)	5'-TGTACGTCACAACTA-3' 3'-ACATGCAGTGTTGAT-5'	112.79	115.19	-2.40
(12$_{DR}$/12$_{RD}$)	5'-TATACAAGTTATCTA-3' 3'-ATATGTTCAATAGAT-5'	116.81	117.38	-0.57
(13$_{DR}$/13$_{RD}$)	5'-CGACTATGCAAAAAC-3' 3'-GCTGATACGTTT TTG-5	110.88	118.31	-7.43
(14$_{DR}$/14$_{RD}$)	5'-CGCAAAAAAAAACGC-3' 3'-GCGTTTTTTTTTGCG-5'	115.76	129.50	-13.74

[a] All fourteen sequences, **DR** and **RD** (**1-14**) [4] . [b] Total base pairing energy in kcal mol^{-1} in **DR** using the free energy in monomeric base pairing and following the equation *i.e.* $[\Delta G^o{}_{bp}]_{\textbf{DR}} = \{[\ \Delta\Delta G^o{}_{p\text{Ka of each mid bp}}\ ]_{\textbf{DR}}\}$ * Number of middle bp + $\{[\Delta\Delta G^o{}_{p\text{Ka of terminal basepair (3´5´)}}\ ]_{\textbf{DR}} + \{[\Delta\Delta G^o{}_{p\text{Ka of terminal basepair (5´3´)}}\ ]_{\textbf{DR}}\}$, the energy calculation is done in kcal mol^{-1} dividing the base-pairing energy (kJmol^{-1}) by 4.2. [c] Total base pairing energy in kcal mol^{-1} in **RD** using the free energy in monomeric base pairing and following the equation ie $[\Delta G^o{}_{bp}]_{\textbf{RD}} = \{[\ \Delta\Delta G^o{}_{p\text{Ka of each mid bp}}\ ]_{\textbf{RD}}\}$ * Number of middle bp + $\{[\Delta\Delta G^o{}_{p\text{Ka of terminal basepair (3´5´)}}\ ]_{\text{RD}} + \{[\Delta\Delta G^o{}_{p\text{Ka of terminal basepair (5´3´)}}\ ]_{\text{RD}}\}$. [d] Difference in bp energy (kcal mol^{-1})between **DR** and **RD** with same sequence context.(**1-14**)[4]

∞