

Supplement 3

Table S1. Mfold-calculated free energies (in kcal/mol) of optimal conformation $\Delta G(O)$, D-duplex conformation $\Delta G(D)$, I-duplex conformation $\Delta G(I)$ and conformations with an open PAS structure $\Delta G(Pn)$ and their differences from free energy of D-duplex conformation of U5-PBS region **fragment 1** (nt 110-243) for **52 CRF01_AE templates**. Conformations with an open PAS structure obtained by prohibiting PAS motif from base pairing are marked by asterisk.

Accession number	Suboptimality, constraints (besides PBS+CAG and U-duplex)	Optimal	D-duplex 1GAAA(3+3), 2GAAA(3)		I-duplex 1GAAA(3), 2GAAA(3+3)		PAS1a d(3+5), 1GAAA(2), 2GAAA(3+3)		PAS1a' d(5), 1GAAA(3), 2GAAA(3+3)		PAS1b d(3), 1GAAA(3), AAAG(3+5)	
		ΔG(O)	ΔG(D)	ΔG(O)- ΔG(D)	ΔG(I)	ΔG(D)- ΔG(I)	ΔG(P1)	ΔG(D)- ΔG(P1)	ΔG(P2)	ΔG(D)- ΔG(P2)	ΔG(P3)	ΔG(D)- ΔG(P3)
U51188	s10%	-25.72	-25.72	0	-23.66	-2.06	-23.44	-2.28	-23.23	-2.49	-20.57*	-5.15
AB220946	s10%	-25.72	-25.72	0	-23.66	-2.06	-23.44	-2.28	-23.23	-2.49	-20.57*	-5.15
AB070352	s10%	-25.72	-25.72	0	-23.66	-2.06	-23.44	-2.28	-23.23	-2.49	-20.57*	-5.15
JX447638	s10%	-25.72	-25.72	0	-23.66	-2.06	-23.44	-2.28	-23.23	-2.49	-20.57*	-5.15
AB032740	s10%	-25.72	-25.72	0	-23.66	-2.06	-23.44	-2.28	-23.23	-2.49	-20.27*	-5.45
JX447349	s10%	-25.72	-25.72	0	-23.66	-2.06	-23.44	-2.28	-23.23	-2.49	-20.57*	-5.15
AB565503	s10%	-25.12	-25.12	0	-23.06	-2.06	-22.84	-2.28	-22.63	-2.49	-19.97*	-5.15
EF036529	s10%	-25.72	-25.72	0	-23.66	-2.06	-23.44	-2.28	-23.33	-2.39	-20.67*	-5.05
EF036530	s10%	-25.72	-25.72	0	-23.66	-2.06	-23.44	-2.28	-23.33	-2.39	-20.67*	-5.05
EF036531	s10%	-25.72	-25.72	0	-23.66	-2.06	-23.44	-2.28	-23.33	-2.39	-20.67*	-5.05
U51189	s10%	-28.12	-28.12	0	-26.06	-2.06	-25.84	-2.28	-25.73	-2.39	-23.07*	-5.05
EF036536	s10%	-25.72	-25.72	0	-23.66	-2.06	-23.39	-2.33	-23.28	-2.44	-20.62*	-5.1
EF036527	s10%	-25.69	-25.69	0	-23.62	-2.07	-23.41	-2.28	-23.31	-2.38	-20.65*	-5.04
AB220948	s10%	-26.42	-26.42	0	-24.36	-2.06	-24.14	-2.28	-23.93	-2.49	-21.27*	-5.15
AB485652	s10%	-26.42	-26.42	0	-24.36	-2.06	-24.14	-2.28	-23.93	-2.49	-21.27*	-5.15
JX447539	s10%	-26.42	-26.42	0	-24.36	-2.06	-24.14	-2.28	-23.93	-2.49	-21.27*	-5.15
AB220945	s10%	-25.92	-25.92	0	-23.86	-2.06	-23.64	-2.28	-23.43	-2.49	-20.57*	-5.35
DQ859180	s10%	-24.72	-24.72	0	-22.66	-2.06	-22.44	-2.28	-22.23	-2.49	-19.27*	-5.45
DQ859179	s10%	-22.92	-22.92	0	-20.86	-2.06	-20.64	-2.28	-20.43	-2.49	-17.77*	-5.15
AB052995	s10%	-22.72	-22.72	0	-20.66	-2.06	-20.44	-2.28	-20.23	-2.49	-17.57*	-5.15
AB220947	s10%	-22.52	-22.52	0	-20.46	-2.06	-20.24	-2.28	-20.03	-2.49	-17.37*	-5.15
AF164485	s15%, D-d(3)*	-17.12	-17.12	0	-15.06	-2.06	-14.84	-2.28	-14.63	-2.49	-11.97*	-5.15
AF259954	s20%	-26.42	-26.42	0	-24.36	-2.06	-23.84	-2.58	-23.93	-2.49	-21.27*	-5.15
AF197339	s10%	-26.29	-26.29	0	-24.36	-1.93	-24.14	-2.15	-23.93	-2.36	-21.27*	-5.02

U54771	s15%	-23.53	-23.06	-0.47	-20.56	-2.5	-20.36	-2.7	-20.16	-2.9	-17.5*	-5.56
EF036532	s15%	-20.23	-19.96	-0.27	-17.46	-2.5	-17.24	-2.72	-17.03	-2.93	-14.37*	-5.59
AB253717	s10%, 1GAAA(2)*, 2GGAAG(3)*	-16.63	-16.36	-0.27	-13.86	-2.5	-13.64	-2.72	-13.53*	-2.83	-12.27*	-4.09
AB253703	s10%, 1GAAA(2)*, 2GGAAG(3)*	-15.92	-15.92	0	-13.86	-2.06	-13.64	-2.28	-13.43*	-2.49	-13.07*	-2.85
AB220944	s15%, 1GAAA(2)*	-19.52	-19.52	0	-17.46	-2.06	-17.24	-2.28	-17.03	-2.49	-14.37*	-5.15
AB253692	s15%, 1GAAA(2)*	-19.52	-19.52	0	-17.46	-2.06	-17.24	-2.28	-17.03	-2.49	-14.37*	-5.15
AB253423	s15%, 1GAAA(2)*	-19.52	-19.52	0	-17.46	-2.06	-17.24	-2.28	-17.03	-2.49	-14.37*	-5.15
AB032741	s15%, 1GAAA(2)*	-19.52	-19.52	0	-17.46	-2.06	-17.24	-2.28	-17.03	-2.49	-14.07*	-5.45
JX447127	s15%, 1GAAA(2)*	-20.22	-20.22	0	-18.16	-2.06	-17.94	-2.28	-17.73	-2.49	-14.77*	-5.45
AB253426	s15%, 1GAAA(2)*	-19.52	-19.52	0	-17.46	-2.06	-17.24	-2.28	-17.13	-2.39	-14.47*	-5.05
AB485655	s15%, 1GAAA(2)*	-19.52	-19.52	0	-17.46	-2.06	-17.24	-2.28	-17.13	-2.39	-14.47*	-5.05
DQ859178	s15%, 1GAAA(2)*	-19.52	-19.52	0	-17.46	-2.06	-17.24	-2.28	-17.13	-2.39	-14.17*	-5.35
EF036535	s15%, 1GACA(2)*	-17.49	-17.49	0	-15.63	-1.86	-15.3	-2.19	-15.19	-2.3	-12.61*	-4.88
DQ314731	s10%	-25.72	-25.72	0	-23.66	-2.06	-22.94	-2.78	-23.23	-2.49	-20.57*	-5.15
DQ314732	s10%	-25.72	-25.72	0	-23.66	-2.06	-22.94	-2.78	-23.23	-2.49	-20.57*	-5.15
EF036528	s15%	-19.52	-19.52	0	-17.46	-2.06	-16.74	-2.78	-17.03	-2.49	-14.37*	-5.15
AB485654	s15%, 1GAAA(2)*	-19.52	-19.52	0	-17.46	-2.06	-16.74	-2.78	-17.03	-2.49	-14.37*	-5.15
AB485653	s15%	-26.42	-26.42	0	-24.36	-2.06	-22.04	-4.38	-23.93	-2.49	-21.17*	-5.25
AF197340	s15%	-25.72	-25.72	0	-23.66	-2.06	-21.03	-4.69	-23.23	-2.49	-20.57*	-5.15
EF036533	s15%	-17.76	-13.1	-4.66	-17.76	4.66	-17.24	4.14	-17.03	3.93	-14.37*	1.27
AF197341	s15%	-27.52	-27.52	0	-25.56	-1.96	-22.43	-5.09	-25.13	-2.39	-20.57*	-6.95
JX446666	s10%, 2GAAAG(3)*	-21.31	-20.66	-0.65	-18.66	-2	-15.27	-5.39	-15.11*	-5.55	-19.55*	-1.11
AB253659	s10%	-26.1	-25.72	-0.38	-23.66	-2.06	-19.69	-6.03	-20.83*	-4.89	-19.11*	-6.61
AB253635	s10%	-26.1	-25.72	-0.38	-23.66	-2.06	-19.69	-6.03	-20.83*	-4.89	-19.11*	-6.61
AB253669	s10%	-26.1	-25.72	-0.38	-23.66	-2.06	-19.69	-6.03	-20.83*	-4.89	-19.11*	-6.61
EF036534	s15%	-28.49	-28.06	-0.43	-23.69	-4.37	-23.47	-4.59	-23.26	-4.8	-20.59*	-7.47
AB253424	s15%	-26.82	-26.82	0	-21.56	-5.26	-21.34	-5.48	-21.13	-5.69	-18.47*	-8.35
AB253647	s10%	-28.92	-28.92	0	-23.66	-5.26	-19.69	-9.23	-20.83*	-8.09	-19.11*	-9.81
	Average:			-0.15		-2.11		-2.88		-2.78		-5.23

Comment on Table S1.

The **D-duplex** has the same 6-bp structure (CUCUGG/CCAGAG) in **98%** of CRF01_AE isolates, it is shortened by 2 upper base pairs in the isolate EF036533 (2%). The **I-duplex** has the same 5-bp structure (CUCUG/UAGGG) in **94%** of CRF01_AE isolates, the upper G:U base pair is replaced by A:U in the isolate EF036533, by G:C in the isolate AF197341 (**4%**) and disrupted in the isolate JX446666. The **U-hairpin** bottom duplex has the same 8-bp

structure (CUAGAGAU/AUCUCUAG) in **94%** of CRF01_AE isolates, it is shortened by the bottom base pair in the isolate U54771, by two upper base pairs in the AB220947 isolate and by five upper base pairs in the AF164485 isolate.

A **1GAAA(3+3) hairpin** is the most variable structural motif in U5-PBS region of CRF01_AE and subtype G isolates. A **1GAAA(3+3) hairpin** has the same structure (UUG/UAG bottom duplex, AAxAA internal loop, AGC/GUU upper duplex, GAAA apical loop) in **60%** of CRF01_AE isolates. In **8%** of isolates, rare base changes do not disturb a whole structure: the middle G:U base pair of the upper duplex is replaced by A:U in two isolates AB565503 and JX447349, the bottom G:U base pair of the bottom duplex is replaced by G:C in two isolates AB253424 and AB253647. In **8%** of isolates, this structure is slightly altered: a single base change U198C results in 1GAAA(3+2) in two isolates U54771 and EF036532, a combination of C214A with 211del results in 1GAAA(2+2) in the JX446666 isolate, C197 in combination with G199U, A200G and A206G results in 1GGAA(3+4) in the EF036534 isolate. Totally, **76%** of isolates have a similar structure of 1GAAA(3+3) hairpin.

The frequent mutation C204U (**23%**) destabilizes the upper base pair of 1GAAA(3+3) hairpin and folding yielded conformations lacking either D-duplex or I-duplex, conformations with a short hairpin involving the PAS sequence or three conformations of interest with various unstable stem-loop structures downstream of the PBS, which were difficult to compare. For the isolates with this base change, we applied a folding constraint forcing two upper base pairs of the 1GAAA(3+3) hairpin (these isolates are marked by 1GAAA(2)* in the column “Suboptimality, constraints”) to model 1GAAA(3+3) structure observed in most of CRF01_AE isolates and all three conformations (D-duplex, I-duplex and PAS1a conformations) were obtained with common energy increments, but ΔG values of D-duplex conformation in these isolates were considerably higher than those in most CRF01_AE isolates (~ -19 kcal/mol vs. ~ -25 kcal/mol).

A **2GAAA(3) hairpin** has the same structure (AGC/GUU stem, GAAA apical loop) in **82%** of CRF01_AE isolates. In **12%** of isolates, rare base changes do not disturb a whole structure: a middle G:U is replaced by A:U in the AB220945 isolate, G229 results in GGAA apical loop instead of GAAA in five isolates (10%). Totally, **94%** of isolates have the very similar structure of 2GAAA(3) hairpin.

A combination of G229 with 232_233insG in two isolates AB253703 and AB253717 or a single mutation 232_233insG in the isolate JX446666 (totally **6%**) resulted in great disturbance of this hairpin and adjacent 2GAAA(3+3) hairpin structures. For these three isolates, we applied a constraint to force AGC/GUU stem (these isolates are marked by 2GGAAG(3)* or 2GAAAG(3)* in the column “Suboptimality, constraints”) and all three conformations were obtained with common energy increments. Two of these isolates (AB253703 and AB253717) have also the base change C204U resulting in destabilization of 1GAAA(3+3) hairpin structure (see above).

An elongation of 2GAAA(3) hairpin by CUC/GAG bottom duplex upon the I-duplex formation, **2GAAA(3+3) hairpin**, is absolutely conserved, none of mutations has been found in CUC and GAG tracts.

Several (from 3 to 6) PAS1a conformations differ in a middle duplex and 1GAAA(3) hairpin stem. Two of them, one with a regular duplex d5 and 1GAAA(3) and another with an irregular duplex d(3+5) and 1GAAA(2), are listed in Table S1. The 3' strand of this middle duplex encompasses a rather conservative 3' portion of the duplicate insertion, while the 5' strand – a very conservative portion of U5/AUG-duplex and a portion of the mutated motif ¹¹⁶GUUAG¹²⁰. The mutated motif ¹¹⁶GUUAG¹²⁰ was observed in 75% of CRF01_AE isolates, in the rest of isolates (25%), rare base changes (G117A – 2 isolates, U118C – 4 isolates, 119_120insAA – 1 isolate, G119U – 4 isolates, G119A – 1 isolate, U120A – 1 isolate) are predominantly located at positions 118 and 119 within this motif, while the extreme positions are not changed. So, structure of a middle duplex varies in these isolates.

A self-contained motif of the duplicate insertion, **AAAG(3+5) hairpin**, has the same structure (GGGAC/GUUCC, AA bulge, UCG/CGA, AAAG apical loop) in **82%** of isolates. In **14%** of isolates, rare base changes do not disturb a whole structure: U:A base pair in the upper duplex is replaced by U:G in five isolates AB032740, AB032741, DQ859178, DQ859180, JX447127 (G229, **10%**), the AB220945 isolate contains AAAA apical loop instead of

AAAG (G226A, **2%**) and the JX446666 isolate contains the AAG bulge instead of AA (232_233insG, **2%**). Two isolates AB253703 and AB253717 with G229 and 232_233insG base changes have a moderately altered structure of this hairpin - AAAG(2+6) with GAA bulge.

The D-duplex conformation with 1GAAA(3+3) and 2GAAA(3) hairpins or rather similar hairpins (see above) was observed in optimal folding of 62% and near-optimal folding ($\Delta\Delta G$ of about $-(0.2 \div 0.5)$ kcal/mol) of 12%, totally 74% of CRF01_AE isolates. The most frequent mutation G110A (29%) has no effect on the conformations of interest, since it is located at the 5' extreme position of fragment 1. The energy increment between the D-duplex and I-duplex conformations appeared to be about -2 kcal/mol in 86% of isolates, in the rest of isolates: -2.5 (6%) and about $-(4 \div 5)$ kcal/mole (8%). The energy increment between the D-duplex and PAS1a conformations appeared to be about $-(2.2 \div 3.0)$ kcal/mol in 79% and $-(2.4 \div 6.0)$ kcal/mol in 17%.

The second scheme of PAS exposure (PAS1b conformation with self-contained motif of the duplicate insertion) was obtained by applying a constraint on PAS motif base pairing for all CRF01_AE isolates, since $\Delta\Delta G$ value of this conformation is lower than those of the PAS1a and PAS1a' conformations (-5.2 kcal/mol vs. -2.8 and -2.9 kcal/mol) and therefore it was not observed in the initial set of calculated structures under the upper bound of 50 and suboptimality range of 10-15%.

Table S2. Mfold-calculated free energies (ΔG in kcal/mol) of optimal conformation $\Delta G(O)$, D-duplex conformation $\Delta G(D)$, I-duplex conformation $\Delta G(I)$ and conformations with an open PAS structure $\Delta G(Pn)$ and their differences from free energy of D-duplex conformation of U5-PBS region **fragment 1** (nt 110-243) for **18 subtype G templates**. Conformations with an open PAS structure obtained by prohibiting PAS motif from base pairing are marked by asterisk.

Accession number	Suboptimality, constraints (besides PBS+CAG and U-duplex)	Optimal	D-duplex 1GAAA(3+3), 2GAAA(3)		I-duplex 1GAAA(3), 2GAAA(3+3)		PAS1a d(7), 1GAAA(2), 2GAAA(3+3)		PAS1a' d(6), 1GAAA(3), 2GAAA(3+3)		PAS1b d(3), 1GAAA(3), AAAG(3+5)	
		$\Delta G(O)$	$\Delta G(D)$	$\Delta G(O)-\Delta G(D)$	$\Delta G(I)$	$\Delta G(D)-\Delta G(I)$	$\Delta G(P1)$	$\Delta G(D)-\Delta G(P1)$	$\Delta G(P2)$	$\Delta G(D)-\Delta G(P2)$	$\Delta G(P3)$	$\Delta G(D)-\Delta G(P3)$
AF084936	s10%	-22.32	-22.32	0	-20.26	-2.06	-18.69	-3.63	-18.59*	-3.73	-17.17*	-5.15
GU362884	s10%	-22.32	-22.32	0	-20.26	-2.06	-18.69	-3.63	-18.59*	-3.73	-16.97*	-5.35
FJ389367	s10%	-22.42	-22.42	0	-20.36	-2.06	-18.69	-3.73	-18.69*	-3.73	-17.27*	-5.15
AY586549	s10%	-22.42	-22.42	0	-20.36	-2.06	-18.69	-3.73	-18.69*	-3.73	-17.27*	-5.15
FJ389364	s15%	-22.42	-22.42	0	-20.36	-2.06	-18.69	-3.73	-17.87*	-4.55	-17.27*	-5.15
FJ389366	s10%	-22.52	-22.52	0	-20.46	-2.06	-18.89	-3.63	-18.79*	-3.73	-17.37*	-5.15
FJ389363	s15%	-24.12	-24.12	0	-22.16	-1.96	-20.59	-3.53	-20.49*	-3.63	-16.87*	-7.25
AY586547	s10%	-25.62	-25.62	0	-23.66	-1.96	-21.99	-3.63	-21.99*	-3.63	-18.67*	-6.95
GU362882	s10%	-28.32	-28.32	0	-26.36	-1.96	-24.69	-3.63	-24.69*	-3.63	-21.37*	-6.95
AY772535	s15%	-22.23	-21.92	-0.31	-19.59	-2.33			-17.93	-3.99	-17.27*	-4.65
FJ670520	s15%, 1GAAA(2)*	-17.32	-17.32	0	-15.26	-2.06	-13.59	-3.73	-13.59*	-3.73	-12.17*	-5.15
AB485662	s10%, 1UGAAG(2)*	-16.82	-16.82	0	-14.86	-1.96	-12.31*	-4.51	-13.19	-3.63	-9.87*	-6.95
FJ670530	s15%, 1GAAAG(2)*	-18.41	-17.39	-1.02	-15.76	-1.63	-14.11*	-4.3	-12.93*	-5.48	-13.45*	-4.96
EU786670	s10%	-32.4	-32.07	-0.33	-28.86	-3.21	-27.29	-4.78	-27.19*	-4.88	-22.01*	-10.06
FR846409	s10%	-26.22	-26.22	0	-22.26	-3.96	-20.59	-5.63	-20.59*	-5.63	-18.86*	-7.36
AY612637	s10%	-22.46	-22.46	0	-18.12*	-4.34	-17.44*	-5.02	-21.96	-0.5	-17.17*	-5.29
FR846408	s15%	-25.62	-25.62	0	-20.36	-5.26	-18.69	-6.93	-18.69*	-6.93	-17.27*	-8.35
AB231893	s20%, 1GAAA(2)*	-21.06	-20.16	-0.9	-14.16*	-6	-12.49	-7.67	-12.49*	-7.67	-11.07*	-9.09
	Average:			-0.14		-2.72		-4.44		-4.25		-6.34

Comment on Table S2.

The **D-duplex** has the same 6-bp structure (CUCUGG/CCAGAG) in **100%** of subtype G isolates. The **I-duplex** has the same 5-bp structure (CUCUG/UAGGG) in **50%** of subtype G isolates; in **33%** of isolates, the upper G:U base pair is replaced by G:C base pair (CUCUG/CAGGG) similar to

the MAL isolate. I-duplex is shortened by an upper base pair in two isolates AY772535 and FJ670530 and by two upper base pairs in the AY612637 isolate. The **U-hairpin** bottom duplex has the same 8-bp structure (CUAGAGAU/AUCUCUAG) in **all** subtype G isolates.

A 1GAAA(3+3) hairpin is the most variable structural motif in U5-PBS region of CRF01_AE and subtype G isolates. A **1GAAA(3+3) hairpin** has the same structure (UUG/UAG bottom duplex, AAxAA internal loop, AGC/GUU upper duplex, GAAA apical loop) in 6 subtype G isolates, a middle G:U of the upper duplex is replaced by A:U in 6 subtype G isolates, a AAxAA internal loop is replaced by GAxAA in the FR846409 isolate and the bottom base pair U:G is replaced by C:G in the FR846408 isolate. Totally, **78%** of subtype G isolates have a similar structure of 1GAAA(3+3) hairpin.

The C204U mutation destabilizing the upper base pair of the 1GAAA(3+3) hairpin was found in three isolates AB231893, AB485662 and FJ670520 and the C204A – in the isolate FJ670530. In the FR846408 isolate, C197 in combination with U198C results in breaking of the bottom base pair. For these isolates (marked by 1GAAA(2)*, 1UGAAG(2)*, 1GAAA(2)*, 1GAAAG(2)* in the column “Suboptimality, constraints”), we applied a folding constraint forcing two upper base pairs of the 1GAAA(3+3) to model 1GAAA(3+3) structure observed in most of subtype G isolates and all three conformations (D-duplex, I-duplex and PAS1a conformations) were obtained with common energy increments, but stability of D-duplex conformation in these isolates was lower by 5-6 kcal/mol than that in most of subtype G isolates.

A **2GAAA(3) hairpin** has the same structure (AGC/GUU stem, GAAA apical loop) in **83%** of subtype G isolates. In the rest of isolates, rare base changes do not disturb a whole structure: a middle G:U is replaced by A:U in the GU362884 isolate, G229 results in GGAA apical loop instead of GAAA in two isolates EU786670 and FJ389363. Totally, **100%** of isolates have the same or very similar structure of 2GAAA(3) hairpin. An elongation of 2GAAA(3) hairpin by 3-bp bottom duplex (CUC/GAG) upon the I-duplex formation, **2GAAA(3+3) hairpin**, is absolutely conserved, none of mutations has been found in CUC and GAG tracts.

Two PAS1a conformations differing in a middle duplex and 1GAAA(3) hairpin stem, d6 and 1GAAA(3) or d7 and 1GAAA(2), are listed in Table S2. The 3' strand of this middle duplex encompasses a rather conservative 3' portion of the duplicate insertion and the 5' strand – a very conservative portion of U5/AUG-duplex and 3-4 downstream nucleotides. Infrequent base changes affect a middle duplex structure in four isolates FJ389364, AY612637, AY772535 and FJ670530 (**22%**).

A self-contained motif of the duplicate insertion, **AAAG(3+5) hairpin**, has the same structure (GGGAC/GUUCC, AA bulge, UCG/CGA, AAAG apical loop) in **83%** of subtype G isolates. In **17%** of isolates, rare base changes do not disturb a whole structure: a U:A base pair in the upper duplex is replaced by U:G in two isolates EU786670 and FJ389363, the GU362884 isolate contains AAAA apical loop instead of AAAG. Totally, the AAAG(3+5) hairpin structure is conserved in all subtype G isolates.

The D-duplex conformation with 1GAAA(3+3) and 2GAAA(3) hairpins was observed in optimal folding and near-optimal folding in 78% of subtype G isolates. The energy increment between the D-duplex and I-duplex conformations appeared to be about -2 kcal/mol in 61% of isolates, within the range of – (1.6÷2.3) kcal/mol in 11% and –(3.2÷6.0) kcal/mole in 28%. The energy increment between the D-duplex and PAS1a conformation appeared to be within the range of –(3.5÷3.7) kcal/mol in 56% and –(2.4÷7.7) kcal/mol in 39%.

Similar to CRF01_AE isolates, fragment 1 adopts D-duplex, I-duplex and PAS1a conformations without restriction on D-duplex, I-duplex and PAS exposure in 72% of subtype G isolates. The PAS1a' and PAS1b conformations were obtained by applying a constraint on PAS motif base pairing for all subtype G isolates, since $\Delta\Delta G$ value of PAS1b conformation is much lower than that of the PAS1a conformation (-6.3 kcal/mol vs. -4.4 kcal/mol) and $\Delta\Delta G$ value of PAS1a' conformation is very close to that of PAS1a conformation.

Table S3. Mfold-calculated free energies (in kcal/mol) of optimal conformation $\Delta G(O)$, D-duplex conformation $\Delta G(D)$, I-duplex conformation $\Delta G(I)$ and conformations with an open PAS structure $\Delta G(P_n)$ and their differences from free energy of D-duplex conformation of U5-PBS region **fragment 1** (nt 110-243) for **21 CRF02_AG templates**. D-duplex conformation with two short hairpins obtained by forcing 1AAGU(3) formation, I-duplex conformation obtained by forcing I-duplex formation and conformations of an open PAS structure obtained by prohibiting PAS motif from base pairing are marked by asterisk.

Accession number	Suboptimality, constraints (besides PBS + CAG and U-duplex)	Optimal	D-duplex 1AAUAGG(3+4), 2GAAA(3)		D-duplex 1AAGU(3), 2GAAA(3)		I-duplex 1AAGU(2), 2GAAA(3+3)		PAS1spec 1AAUAGG(3+4), 2GAAA(3)		PAS1b d(3), AAAG(3+5)		PAS1a 1UGA(1/2), 2GAAA(3+3)	
		ΔG(O)	ΔG(D)	ΔG(O)- ΔG(D)	ΔG(D1)	ΔG(D)- ΔG(D1)	ΔG(I)	ΔG(D)- ΔG(I)	ΔG(P1)	ΔG(D)- ΔG(P1)	ΔG(P2)	ΔG(D)- ΔG(P2)	ΔG(P3)	ΔG(D)- ΔG(P3)
AY271690	s10%	-21.83	-21.83	0	-15.73*	-6.1	-14.45*	-7.38	-14.11*	-7.72	-13.3*	-8.53	-12.1*	-9.73
AB049811	s10%	-23.23	-23.23	0	-17.13*	-6.1	-15.85*	-7.38	-15.51*	-7.72	-14.7*	-8.53	-13.5*	-9.73
AB286859	s10%	-23.13	-23.13	0	-17.03*	-6.1	-15.75*	-7.38	-15.41*	-7.72	-14.6*	-8.53	-13.4*	-9.73
JQ316137	s10%	-24.23	-24.23	0	-18.13*	-6.1	-16.85*	-7.38	-16.51*	-7.72	-15.7*	-8.53	-14.5*	-9.73
JQ316136	s10%	-23.43	-23.43	0	-17.33*	-6.1	-16.05*	-7.38	-15.71*	-7.72	-14.9*	-8.53	-13.7*	-9.73
JQ316138	s10%	-23.43	-23.43	0	-17.33*	-6.1	-16.05*	-7.38	-15.71*	-7.72	-14.9*	-8.53	-13.7*	-9.73
AB231895	s10%	-22.49	-22.49	0	-17.2*	-5.29	-15.25*	-7.24	-14.23*	-8.26	-10.8*	-11.69	-12.9*	-9.59
AB485634	s10%	-24.53	-24.53	0	-19.13*	-5.4	-17.85*	-6.68	-16.81*	-7.72	-16.5*	-8.03	-15.3*	-9.23
L39106	s10%	-18.97	-18.97	0	-15.76*	-3.21	-14.49*	-4.48	-11.78*	-7.19	-13.3*	-5.67	-10.41*	-8.56
AB485636	s10%	-17.13	-17.13	0	-13.83*	-3.3	-12.55*	-4.58	-9.41*	-7.72	-13.12*	-4.01	-12.56*	-4.57
AB485633	s10%	-20.7	-19.29	-1.41	-14.44*	-4.85	-12.61*	-6.68	-9.77*	-9.52	-14.3*	-4.99	-11.3*	-7.99
EU884501	s10%	-21.29	-21.29	0	-14.04*	-7.25	-12.21*	-9.08	-13.59*	-7.7	-14*	-7.29	-10.9*	-10.39
EU786671	s10%	-19.69	-19.69	0	-12.44*	-7.25	-10.61*	-9.08	-11.99*	-7.7	-12.4*	-7.29	-9.3*	-10.39
AB286857	s10%	-20.99	-20.99	0	-13.74*	-7.25	-11.91*	-9.08	-13.29*	-7.7	-13.7*	-7.29	-10.6*	-10.39
AB286861	s10%	-22.99	-22.99	0	-15.74*	-7.25	-13.91*	-9.08	-15.29*	-7.7	-15.7*	-7.29	-12.6*	-10.39
AB231896	s10%	-23.19	-23.19	0	-15.94*	-7.25	-14.11*	-9.08	-15.49*	-7.7	-15.9*	-7.29	-12.8*	-10.39
AB485635	s10%	-23.79	-23.79	0	-16.54*	-7.25	-14.71*	-9.08	-16.09*	-7.7	-16.5*	-7.29	-13.4*	-10.39
AB286855	s10%	-19.02	-16.15	-2.87	-14.77*	-1.38	-14.25*	-1.9	-8.46*	-7.69	-13.1*	-3.05	-9.87*	-6.28
AB231898	s10%	-24.26	-22.69	-1.57	-16.74*	-5.95	-13.33*	-9.36	-15.09*	-7.6	-18.9*	-3.79	-11.31*	-11.38
AB286863	s10%	-27.38	-27.38	0	-17.93*	-9.45	-16.35*	-11.03	-20.89*	-6.49	-13.5*	-13.88	-14.3*	-13.08
AB286864	s10%	-27.18	-27.18	0	-17.73*	-9.45	-16.15*	-11.03	-20.69*	-6.49			-11.77*	-15.41
	Average:			-0.28		-6.11		-7.70		-7.68		-7.50		-9.85

Comment on Table S3.

The **D-duplex** has the same 6-bp structure (CUCUGG/CCAGAG) in **81%** of CRF02_AG isolates. The second A:U base pair is replaced by G:U in the AB485636 isolate, D-duplex is elongated by additional U:A base pair in the AB231895 isolate or U:G base pair in two isolates AB286863 and AB286864, totally **19%**. The **I-duplex** has the same 5-bp structure (CUCUG/UAGGG) in **all** CRF02_AG isolates. The **U-hairpin** bottom duplex has the same 8-bp structure (CUAGAGAU/AUCUCUAG) in **all** CRF02_AG isolates.

As distinct from CRF01_AE and subtype G isolates, the mutations located immediately downstream of the PBS (a 7-nt deletion and base changes at positions 197 and 198 resulting in UU or CG) affect a structure of the hairpin formed in this region: the isolates with C197U in combination with U198 (**57%**) contain mostly a **1AAUAGG(3+4)** hairpin (UUGA/UCGA(G), A bulge, GUU/GAC, AAUAGG apical loop, **33%**) or very similar structures **1AAUAGG(4+3)** (AB286863, AB286864) and **1AAUAGGG(2+4)** (L39106); the isolates with C197 in combination with U198G (38%) – mostly **1AAUAGG(4+2)** hairpin (**29%**); the AB286855 isolate with C197, U198 and G209A – **1AAUAG(2+3)** hairpin. Totally, **47%** of CRF02_AG contain **1AAUAGG(3+4)** or very similar structures and **29%** - **1AAUAGG(4+2)**. For easy reference, the most common **1AAUAGG(3+4)** hairpin is listed in Table S3. Similar to **1GAAA(3+3)** hairpin of CRF01_AE and subtype G isolates, **1AAUAGG(3+4)** or **1AAUAGG(4+2)** hairpin appeared to be the most variable structural motif in U5-PBS region of CRF02_AG isolates.

A **2GAAA(3) hairpin** has the same structure (AGC/GUU stem, GAAA apical loop) in **85%** of CRF02_AG isolates, but it is shortened by 1 bp in **15%** of isolates, **2GAAA(2)** (3 isolates with elongated D-duplex: AB231895, AB286863, AB286864). An elongation of **2GAAA(3)** hairpin by 3-bp bottom duplex (CUC/GAG) upon the I-duplex formation, **2GAAA(3+3) hairpin**, is conserved in **95%**; none of mutations has been found in GAG tract, the rare base change A239G results in U:A to U:G substitution in the AB485636 isolate, while the rare base change U220G isolate prevents **2GAAA(3)** elongation in the AB231898 isolate.

A self-contained motif of the duplicate insertion, **AAAG(3+5) hairpin**, has the same structure (GGGAC/GUUGC, AA bulge, UCG/CGA, AAAG apical loop) in **78%** of CRF02_AG isolates. In **22%** of isolates, rare base changes significantly affect its structure: two isolates AB286863 and AB286864 have **AAAG(3+2+2)**, in the latter isolate this hairpin forms in fragment 2 only, but not in fragment 1; in the AB231895 isolate containing a double duplicate insertion, a similar hairpin **AAAG(3+2+2)** formed by the second duplicate insertion; in the AB231898 isolate – **GAAA(2+5)** with GAA bulge.

The D-duplex conformation with **1AAUAGG(3+4)** or similar hairpin (see above) and **2GAAA(3)** was observed in optimal foldings in 81% of CRF02_AG isolates. In the isolates with elongated D-duplex, switches from D-duplex to I-duplex and PAS1a or PAS2 conformations proceed with a higher energy inputs than those in other CRF02_AG isolates.

All conformation with I-duplex and open PAS structure were obtained by forcing I-duplex formation and prohibiting PAS motif from base pairing, since ΔG values of these conformations are much higher than D-duplex conformation ($\sim(10\div17)$ kcal/mol vs. $\sim(17\div24)$ kcal/mol) and therefore they were not observed in the initial set of calculated structures under the upper bound of 50 and suboptimality of 10%. Among PAS1 conformations, we frequently observed a conformation that contained U-hairpin and **2GAAA(3+3)** hairpin only, but lacked any supporting duplex (PAS1(gaaa) conformation), this conformation is not listed in Table S3.

Table S4. Mfold-calculated free energies (in kcal/mol) of optimal conformation $\Delta G(O)$, D-duplex conformation $\Delta G(D)$, I-duplex conformation $\Delta G(I)$ and conformations with an open PAS structure $\Delta G(Pn)$ and their differences from free energy of D-duplex conformation of U5-PBS region **fragment 2** (nt 113-259) for **52 CRF01_AE templates**. Conformations with an open PAS structure obtained by prohibiting PAS motif from base pairing are marked by asterisk.

Accession number	Suboptimality, constraints (besides PBS+CAG and U-duplex)	Optimal	D-duplex GU/CGC-d, d(3), 1GAAA(3+3), 2GAAA(3)		I-duplex GU/CGC-d, d(3), 1GAAA(3), 2GAAA(3+3)		PAS2(gaaa) GU/CGC-d, 1GAAA(3), 2GAAA(3+3+5)		PAS2(pal) GU/CGC-d, 1GAAA(3), AAAG(3+5), pal(4)	
		$\Delta G(O)$	$\Delta G(D)$	$\Delta G(O)$ - $\Delta G(D)$	$\Delta G(I)$	$\Delta G(D)$ - $\Delta G(I)$	$\Delta G(P1)$	$\Delta G(D)$ - $\Delta G(P1)$	$\Delta G(P2)$	$\Delta G(D)$ - $\Delta G(P2)$
U51189	s20%	-34.42	-34.42	0	-31.39	-3.03	-31.13	-3.29	-30.61*	-3.81
U51188	s10%	-32.02	-32.02	0	-28.99	-3.03	-28.73*	-3.29	-28.21*	-3.81
AB220946	s10%	-32.02	-32.02	0	-28.99	-3.03	-28.73*	-3.29	-28.21*	-3.81
AB070352	s10%	-32.02	-32.02	0	-28.99	-3.03	-28.73*	-3.29	-28.21*	-3.81
EF036530	s10%	-32.02	-32.02	0	-28.99	-3.03	-28.73*	-3.29	-28.21*	-3.81
JX447638	s10%	-32.02	-32.02	0	-28.99	-3.03	-28.73*	-3.29	-28.21*	-3.81
JX447349	s10%	-32.02	-32.02	0	-28.99	-3.03	-28.73*	-3.29	-28.21*	-3.81
EF036529	s10%	-32.02	-32.02	0	-28.99	-3.03	-28.73*	-3.29	-28.21*	-3.81
AB485652	s10%	-32.72	-32.72	0	-29.69	-3.03	-29.43*	-3.29	-28.91*	-3.81
AB220948	s10%	-32.72	-32.72	0	-29.69	-3.03	-29.43*	-3.29	-28.91*	-3.81
AB565503	s10%	-31.42	-31.42	0	-28.39	-3.03	-28.13*	-3.29	-27.61*	-3.81
EF036527	s10%	-31.99	-31.99	0	-28.95	-3.04	-28.71*	-3.28	-28.18*	-3.81
DQ859179	s10%	-29.22	-29.22	0	-26.19	-3.03	-25.93*	-3.29	-25.41*	-3.81
AB052995	s10%	-29.02	-29.02	0	-25.99	-3.03	-25.73*	-3.29	-25.21*	-3.81
AB220947	s10%	-28.82	-28.82	0	-25.79	-3.03	-25.53*	-3.29	-25.01*	-3.81
AF164485	s10%	-23.42	-23.42	0	-20.39	-3.03	-20.13*	-3.29	-19.61*	-3.81
AB032740	s10%	-32.02	-32.02	0	-28.99	-3.03	-28.73*	-3.29	-27.91*	-4.11
DQ859180	s10%	-31.02	-31.02	0	-27.99	-3.03	-27.73*	-3.29	-26.91*	-4.11
AB220945	s10%	-32.22	-32.22	0	-29.19	-3.03	-28.93*	-3.29	-28.21*	-4.01
AF197339	s10%	-32.59	-32.59	0	-29.69	-2.9	-29.43*	-3.16	-28.91*	-3.68
AB220944	s20%, 1GAAA(2)*	-25.82	-25.82	0	-22.79	-3.03	-22.53	-3.29	-22.01*	-3.81
AB253692	s20%, 1GAAA(2)*	-25.82	-25.82	0	-22.79	-3.03	-22.53	-3.29	-22.01*	-3.81
AB253423	s20%, 1GAAA(2)*	-25.82	-25.82	0	-22.79	-3.03	-22.53	-3.29	-22.01*	-3.81
AB253426	s20%, 1GAAA(2)*	-25.82	-25.82	0	-22.79	-3.03	-22.53	-3.29	-22.01*	-3.81
AB485655	s20%, 1GAAA(2)*	-25.82	-25.82	0	-22.79	-3.03	-22.53	-3.29	-22.01*	-3.81
AB032741	s20%, 1GAAA(2)*	-25.82	-25.82	0	-22.79	-3.03	-22.53	-3.29	-21.71*	-4.11

DQ859178	s20%, 1GAAA(2)*	-25.82	-25.82	0	-22.79	-3.03	-22.53	-3.29	-21.71*	-4.11
JX447127	s20%, 1GAAA(2)*	-26.52	-26.52	0	-23.49	-3.03	-23.23	-3.29	-22.41*	-4.11
EF036535	s15%, 1GACA(2)*	-23.79	-23.79	0	-20.96	-2.83	-20.69*	-3.1	-20.16*	-3.63
AB485654	s15%, 1GAAA(2)*	-24.92	-24.92	0	-21.89	-3.03	-20.43*	-4.49	-21.31*	-3.61
AB253717	s15%, 1GAAA(2)*, 2GGAAG(3)*	-22.93	-22.66	-0.27	-19.19	-3.47	-18.93	-3.73	-19.81*	-2.85
AB253703	s10%, 1GAAA(2)*, 2GGAAG(3)*	-22.22	-22.22	0	-19.19	-3.03	-18.93*	-3.29	-21.31*	-0.91
AB253659	s10%	-32.92	-32.92	0	-29.89	-3.03	-29.53	-3.39	-29.71*	-3.21
AB253669	s10%	-32.92	-32.92	0	-29.89	-3.03	-29.53	-3.39	-29.71*	-3.21
AB253635	s10%	-32.92	-32.92	0	-29.89	-3.03	-29.53	-3.39	-29.71*	-3.21
AB485653	s10%	-28.82	-28.82	0	-25.79	-3.03	-25.53*	-3.29		
AF197341	s20%	-29.82	-29.82	0	-27.06	-2.76	-26.64*	-3.18		
DQ314732	s10%	-31.12	-31.12	0	-28.09	-3.03	-26.63*	-4.49	-28.06*	-3.06
DQ314731	s10%	-31.12	-31.12	0	-28.09	-3.03	-26.63*	-4.49	-28.06*	-3.06
EF036528	s10%	-24.92	-24.92	0	-21.89	-3.03	-20.43*	-4.49	-21.86*	-3.06
AF259954	s10%	-36.42	-36.42	0	-33.39	-3.03	-30.09*	-6.33	-30.71*	-5.71
JX447539	s10%	-35.12	-35.12	0	-32.09	-3.03	-28.29	-6.83	-30.31*	-4.81
JX446666	s10%, 2GAAAG(3)*	-29.67	-27.61	-2.06	-23.99	-3.62	-24.51*	-3.1	-27.18*	-0.43
AF197340	s10%	-28.7	-28.22	-0.48	-23.56	-4.66	-23.31*	-4.91	-24.91*	-3.31
EF036536	s10%	-31.2	-30.72	-0.48	-26.06	-4.66	-28.68*	-2.04	-28.16*	-2.56
U54771	s20%	-29.83	-29.36	-0.47	-25.89	-3.47	-25.66*	-3.7	-25.14*	-4.22
EF036532	s10%	-25.93	-25.66	-0.27	-22.79	-2.87	-22.53*	-3.13	-21.06*	-4.6
AB253424	s15%	-33.12	-33.12	0	-26.89	-6.23	-26.63*	-6.49	-26.11*	-7.01
EF036534	s15%	-34.79	-34.36	-0.43	-29.02	-5.34	-28.76*	-5.6	-28.23*	-6.13
AB253647	s10%	-36.12	-36.12	0	-29.89	-6.23	-29.53*	-6.59	-29.71*	-6.41
EF036531	s10%	-31.73	-31.42	-0.31	-28.14	-3.28	-31.73	0.31		
EF036533	s10%, D-d(4)*	-23.9	-18.74	-5.16	-22.59	3.85	-22.03*	3.29	-21.51*	2.77
	Average:			-0.19		-3.15		-3.50		-3.68

Comment on Table S4.

An elongation of 2GAAA(3+3) stem through 5-bp bottom duplex (AGGGA/UCUCU) and AAGU bulge, **2GAAA(3+3+5) hairpin**, observed in the PAS2(gaaa) conformation of CRF01_AE isolates is **highly conserved**, since none of mutations has been found in the AGGGA tract. Two isolates AF197341 and EF036531 have the base change U247C in the UCUCU tract resulting in G:U to G:C base pair substitution in the middle part of the AGGGA/UCUCU-duplex. In the DQ859178 isolate, the base change C244A results in AAGA bulge instead of AAGU.

A **pal(4) hairpin** observed in the PAS2(pal) conformation has the same structure (AGAG/CUCU stem and AAGUU apical loop) in **94%** of CRF01_AE isolates, the DQ859178 isolate has AAGAU apical loop (2%). In two isolates EF036531 and AF197341 (4%), U247C prevents a formation of pal(4) hairpin. As compared to subtype G and CRF02_AG isolates, pal(4) hairpin in optimal PAS2(pal) conformation is upstreamly shifted by 1 nt in CRF01_AE isolates (AGAG/CUCU stem), however in near-optimal foldings it has the GAGA/UCUC stem similar to subtype G and CRF02_AG isolates.

The duplex closing all conformations of fragment 2 (the **GU/CGC-duplex of PBS stem extension**) has the same structure (GUGUU/GACGC) in **86%** of CRF01_AE isolates, in four isolates AB253635, AB253647, AB253659 and AB253669 (**8%**) the upper base pair U:G is replaced by C:G. In two isolates AF197340 and AF197341, the rare base change G117A results in the shortened GU/CGC-duplex by 2 upper base pairs (GUG/CGC). In the AB485653 isolate, the rare base change G114A results in the shortened GU/CGC-duplex by bottom base pair (UGUU/GACG).

The D-duplex conformation with 1GAAA(3+3) and 2GAAA(3) hairpins or similar hairpins (see Comment on Table S1) was observed in optimal folding of 62% and near-optimal folding ($\Delta\Delta G$ of about $-(0.3\div0.5)$ kcal/mol) of 12%, totally 74% of CRF01_AE isolates. For the isolates with the C204U base change (23%), we applied a folding constraint forcing two upper base pairs of the 1GAAA(3+3) hairpin (these isolates are marked by 1GAAA(2)* in the column “Suboptimality, constraints”) and D-duplex and I-duplex conformations were obtained with common energy increment, but ΔG values of D-duplex conformation in these isolates were considerably higher than those in most of CRF01_AE isolates ($-(23\div26)$ kcal/mol vs. $-(29\div32)$ kcal/mol).

An energy increment between the D-duplex and I-duplex conformations appeared to be about -3 kcal/mol in 83% of isolates. The PAS2(pal) and PAS2(gaaa) conformations were obtained by applying a constraint prohibiting PAS motif from base pairing for all CRF01_AE isolates. The energy increments between the D-duplex and PAS2(gaaa) or PAS2(pal) conformations appeared to be within the range of $-(3.1\div3.7)$ kcal/mol in 75% or $-(3.6\div4.2)$ kcal/mol in 62%. For three isolates AB485653, AF197341 and EF036531, we did not manage to obtain the PAS2(pal) conformation.

Table S5. Mfold-calculated free energies (in kcal/mol) of optimal conformation $\Delta G(O)$, D-duplex conformation $\Delta G(D)$, I-duplex conformation $\Delta G(I)$ and conformations with an open PAS structure $\Delta G(Pn)$ and their differences from free energy of D-duplex conformation of U5-PBS region **fragment 2** (nt 113-259) for **18 subtype G templates**. I-duplex conformation obtained by forcing I-duplex base pairing and conformations of an open PAS structure obtained by prohibiting PAS motif from base pairing are marked by asterisk.

Accession number	Suboptimality, constraints (besides PBS+CAG and U-duplex)	Optimal	D-duplex GU/CGC-d, d(2), 1GAAA(3+3) 2GAAA(3)		I-duplex GU/CGC-d, d(2), 1GAAA(3), 2GAAA(3), pal(4)		PAS2(pal) GU/CGC-d, 1GAAA(3) AAAG(3+5), pal(4)		PAS2(gaaa) GU/CGC-d, 1GAAA(3) 2GAAA(3+3+5)	
		$\Delta G(O)$	$\Delta G(D)$	$\Delta G(O)-\Delta G(D)$	$\Delta G(I)$	$\Delta G(D)-\Delta G(I)$	$\Delta G(P1)$	$\Delta G(D)-\Delta G(P1)$	$\Delta G(P2)$	$\Delta G(D)-\Delta G(P2)$
GU362882	s15%	-31.52	-31.52	0	-28.79	-2.73	-30.43	-1.09	-28.36	-3.16
FR846409	s15%	-29.42	-29.42	0	-24.69	-4.73	-27.96	-1.46	-24.26	-5.16
AY586547	s15%	-28.82	-28.82	0	-26.09	-2.73	-27.73	-1.09	-25.66	-3.16
FJ389363	s15%	-28.82	-28.82	0	-25.91	-2.91	-22.78*	-6.04	-20.51*	-8.31
FR846408	s10%	-28.82	-28.82	0	-22.79	-6.03	-26.33	-2.49	-24.16	-4.66
EU786670	s10%	-35.6	-35.27	-0.33	-32.01	-3.26	-31.89*	-3.38	-29.06*	-6.21
AB485662	s10%, 1UGAAG(2)*	-20.02	-20.02	0	-17.29*	-2.73	-18.93*	-1.09	-16.86*	-3.16
GU362884	s15%	-25.09	-23.12	-1.97	-20.11	-3.01	-20.94	-2.18	-20.39	-2.73
AB231893	s20%, 1GAAA(2)*	-23.36	-23.36	0	-16.59	-6.77	-20.13	-3.23	-17.96	-5.4
AF084936	s10%	-26.23	-25.52	-0.71	-22.69	-2.83	-26.23	0.71	-24.06	-1.46
FJ389367	s10%	-26.33	-25.62	-0.71	-22.79	-2.83	-26.33	0.71	-24.16	-1.46
AY586549	s10%	-26.33	-25.62	-0.71	-22.79	-2.83	-26.33	0.71	-24.16	-1.46
FJ389366	s10%	-26.43	-25.72	-0.71	-22.89	-2.83	-26.43	0.71	-24.26	-1.46
FJ670520	s15%, 1GAAA(2)*	-21.66	-20.52	-1.14	-17.69	-2.83	-21.23	0.71	-19.06	-1.46
FJ389364	s10%	-25.39	-21.72	-3.67	-20.79	-0.93	-23.73	2.01	-22.36	0.64
FJ670530	s10%, 1GAAAG(2)*	-22.93	-21.61	-1.32	-17.11	-4.5	-22.51	0.9	-20.53	-1.08
AY772535	s10%	-26.76	-25.12	-1.64	-22.02*	-3.1	-26.33	1.21	-24.36	-0.76
AY612637	s15%	-26.23	-25.66	-0.57	-21.56	-4.1	-26.23	0.57	-23.31	-2.35
	Average:			-0.75		-3.43		-0.77		-2.93

Comment on Table S5.

An elongation of 2GAAA(3+3) stem through 5-bp bottom duplex (AGGGA/UCUCU) and AAGU bulge, **2GAAA(3+3+5) hairpin**, observed in the PAS2(gaaa) conformation of subtype G isolates is **very conserved**, since none of mutations has been found in the UCUCU tract; the AY612637 isolate has

the base change A215C in the AGGGA tract resulting in the disruption of the upper base pair A:U. In three isolate AB485662, FJ389364 and FJ389367 (17%), the base change C244A results in AAGA bulge instead of AAGU.

A **pal(4) hairpin** observed in the PAS2(pal) conformation has the same stem (GAGA/UCUC) in **all** subtype G isolates, its apical part is AGUUC (83%) or AGAUC (17%).

The duplex closing all conformations of fragment 2 (the **GU/CGC-duplex of PBS stem extension**) has the same structure (UUGUGU/ACGCAG) in **83%** of subtype G isolates. In the FJ389363 isolate, the rare base change U120C results in a shortened GU/CGC-duplex by the upper base pair (UUGUG/CGCAG), but a 2-bp duplex between D- and GU/CGC-duplexes is elongated by 1 base pair instead. In the GU362884 isolate, the rare base change G119U results in a shortened GU/CGC-duplex by 2 bps (UUGU/GCAG), but a 2-bp duplex between D- and GU/CGC-duplexes is elongated by 2 bps instead. In the FJ389364 isolate, the rare base change G117U results in a 6-bp duplex UUUGUG/CGCAGG, but an absence of a 2-bp duplex between D- and GU/CGC-duplexes.

The D-duplex conformation with 1GAAA(3+3) and 2GAAA(3) hairpins or similar hairpins (see Comment on Table S2) was observed in optimal or near-optimal foldings of 50% of subtype G isolates, the PAS2(pal) conformation was observed in optimal foldings in the rest of isolates. The PAS2(pal) conformation appeared to be more favorable than the I-duplex conformation in 89% of subtype G isolates.

Table S6. Mfold-calculated free energies (in kcal/mol) of optimal conformation $\Delta G(O)$, D-duplex conformation $\Delta G(D)$, I-duplex conformation $\Delta G(I)$ and conformations with an open PAS structure $\Delta G(Pn)$ and their differences from free energy of D-duplex conformation of U5-PBS region **extended fragment 2** (nt 109-261) for **21 CRF02_AG templates**. I-duplex conformation obtained by forcing I-duplex base pairing and conformations with an open PAS structure obtained by prohibiting PAS motif from base pairing are marked by asterisk.

Accession number	Suboptimality, constraints (besides PBS+CAG and U-duplex)	Optimal	D-duplex d(4/5), GU/CGC-d, d(2), 1AAUAGG(3+4), 2GAAA(3)		PAS2(pal) d(4/5), GU/CGC-d, 2AAAG(3+5), pal(4)		PAS2spec d(4/5), GU/CGC-d, 1AAUAGG(3+4), 2GAAA(3), pal(4)		PAS2(gaaa) d(4/5), GU/CGC-d, 2GAAA(3+3+5)		I-duplex d(4/5), GU/CGC-d, d(2), 1AAGU(2), 2GAAA(3), pal(4)	
			$\Delta G(D)$	$\Delta G(O)-\Delta G(D)$	$\Delta G(P1)$	$\Delta G(D)-\Delta G(P1)$	$\Delta G(P2)$	$\Delta G(D)-\Delta G(P2)$	$\Delta G(P3)$	$\Delta G(D)-\Delta G(P3)$	$\Delta G(I)$	$\Delta G(D)-\Delta G(I)$
AY271690	s10%	-31.93	-31.93	0	-29.26	-2.67	-28.1*	-3.83	-27.08	-4.85	-23.78*	-8.15
AB231896	s10%	-34.4	-33.29	-1.11	-31.86	-1.43	-29.47*	-3.82	-29.68	-3.61	-23.54*	-9.75
AB286857	s10%	-32.2	-31.09	-1.11	-29.66	-1.43	-27.27*	-3.82	-27.48	-3.61	-21.34*	-9.75
AB286861	s10%	-31.7	-30.59	-1.11	-29.16	-1.43	-26.77*	-3.82	-26.98	-3.61	-20.84*	-9.75
EU786671	s10%	-29	-27.29	-1.71	-25.86*	-1.43	-23.47*	-3.82	-23.68*	-3.61	-17.36*	-9.93
AB485634	s10%	-32.13	-32.13	0	-29.96	-2.17	-28.3*	-3.83	-27.78	-4.35	-24.68*	-7.45
AB485635	s10%	-32.5	-31.39	-1.11	-29.96	-1.43	-27.57*	-3.82	-27.78	-3.61	-21.64*	-9.75
AB286859	s10%	-30.73	-30.73	0	-28.06	-2.67	-26.9*	-3.83	-25.88	-4.85	-22.58*	-8.15
JQ316137	s10%	-30.13	-30.13	0	-27.86	-2.27	-26.9*	-3.23	-26.98	-3.15	-22.7*	-7.43
JQ316138	s10%	-31.93	-31.03	-0.9	-28.36	-2.67	-27.2*	-3.83	-27.16	-3.87	-22.88*	-8.15
AB049811	s10%	-31.23	-31.23	0	-28.56	-2.67	-27.4*	-3.83	-26.38	-4.85	-23.08*	-8.15
EU884501	s10%	-29.8	-28.69	-1.11	-27.26	-1.43	-24.87*	-3.82	-25.08	-3.61	-18.76*	-9.93
JQ316136	s10%	-30.13	-30.13	0	-27.86	-2.27	-26.9*	-3.23	-26.98	-3.15	-22.7*	-7.43
AB231895	s10%	-41.17	-41.17	0	-38.14*	-3.03	-36.71*	-4.46	-36.26*	-4.91	-35.73*	-5.44
AB286863	s10%	-34.98	-34.98	0	-26.96*	-8.02	-32.37*	-2.61	-26.78*	-8.2	-23*	-11.98
AB286864	s10%	-36.28	-36.28	0	-23.61*	-12.67	-29.01*	-7.27	-23.17*	-13.11	-24.3*	-11.98
AB286855	s10%	-27.37	-24.15	-3.22	-26.96	2.81	-20.34*	-3.81	-24.78	0.63	-21.1*	-3.05
AB485633	s10%	-28.65	-26.89	-1.76	-27.86	0.97	-23.07*	-3.82	-25.68	-1.21	-19.36*	-7.53
L39106	s10%	-27	-26.37	-0.63	-26.58	0.21	-22.59*	-3.78	-24.41	-1.96	-20.94*	-5.43
AB485636	s10%	-27.27	-25.63	-1.64	-26.86	1.23	-22.4*	-3.23	-24.98	-0.65	-20.5*	-5.13
AB231898	s10%	-37.2	-32.89	-4.31	-34.66	1.77	-31.66*	-1.23	-29.66*	-3.23	-27.36*	-5.53
	Average:			-0.94		-2.03		-3.75		-3.97		-8.09

Comment on Table S6.

An elongation of 2GAAA(3+3) stem through 5-bp bottom duplex (AGGGA/UCUCU) and AAGA/C/U bulge, **2GAAA(3+3+5) hairpin**, observed in the PAS2(gaaa) conformation of **86%** CRF02_AG isolates. As distinct from CRF01_AE and subtype G isolates, the last position of AAGU bulge is much more variable: AAGA (38%), AAGC (38%), AAGU (14%) and AAG (10%). None of mutations has been found in the AGGGA tract, while the base change C246A in the UCUCU tract in combination with the C244del found in the AB231898 isolate results in a great disturbance of the 5-bp duplex structure, 2GAAA(3+3+3) hairpin with 9x9 internal loop. In the isolate AB485636, the deletion at the first position of AAGC bulge in combination with the A239G results in the internal loop 3x5 instead of 3x3 and C bulge instead of AAGC. In the isolate AB231895, the 244Cdel in combination with the 250_251insC results in 2GAAA(2+2+6) hairpin with the internal loop 5x6.

A **pal(4) hairpin** observed in the PAS2(pal) and PAS2*specific* conformations has the same stem (GAGA/UCUC) in **86%** of CRF02_AG isolates, while its apical part depends on a base change at position 244 and adjacent deletions: AGAUC (38%), AGCUC (24%), AGUUC (14%), AGUA (5%) AGUC (5%). In the AB485636 isolate, 241Gdel in combination with the A239G results in A:U to G:U substitution in hairpin stem and GCUC apical loop. In two isolates JQ316136 and JQ316137, the C250U base change leads to a 1-nt upstream shift of pal(4) hairpin (AGAG/CUCU, AAGCU apical loop).

The structures of **PBS stem extension** (the **GU/CGC-duplex** and a bottom extension with involvement of 110G:260C) are shown in Supplement 2, Fig. S2.

The D-duplex conformation with 1AAUAGG(3+4) or rather similar hairpin and 2GAAA(3) (see Comment on Table S3) was observed in optimal or near-optimal foldings in 76% of subtype G isolates. The PAS2(pal) conformation appeared to be more favorable than the D-duplex conformation in 24% of CRF02_AG isolates, the D-duplex conformation and the PAS2(pal) conformation are close in free energy in 29%.

The PAS2*specific* conformation was obtained by prohibiting the PAS motif from base pairing in all CRF02_AG isolates, since the PAS2(gaaa) and PAS2*specific* conformations are very close in free energy. The I-duplex conformation was obtained by forcing I-duplex formation, since it adopts with very high energy input.