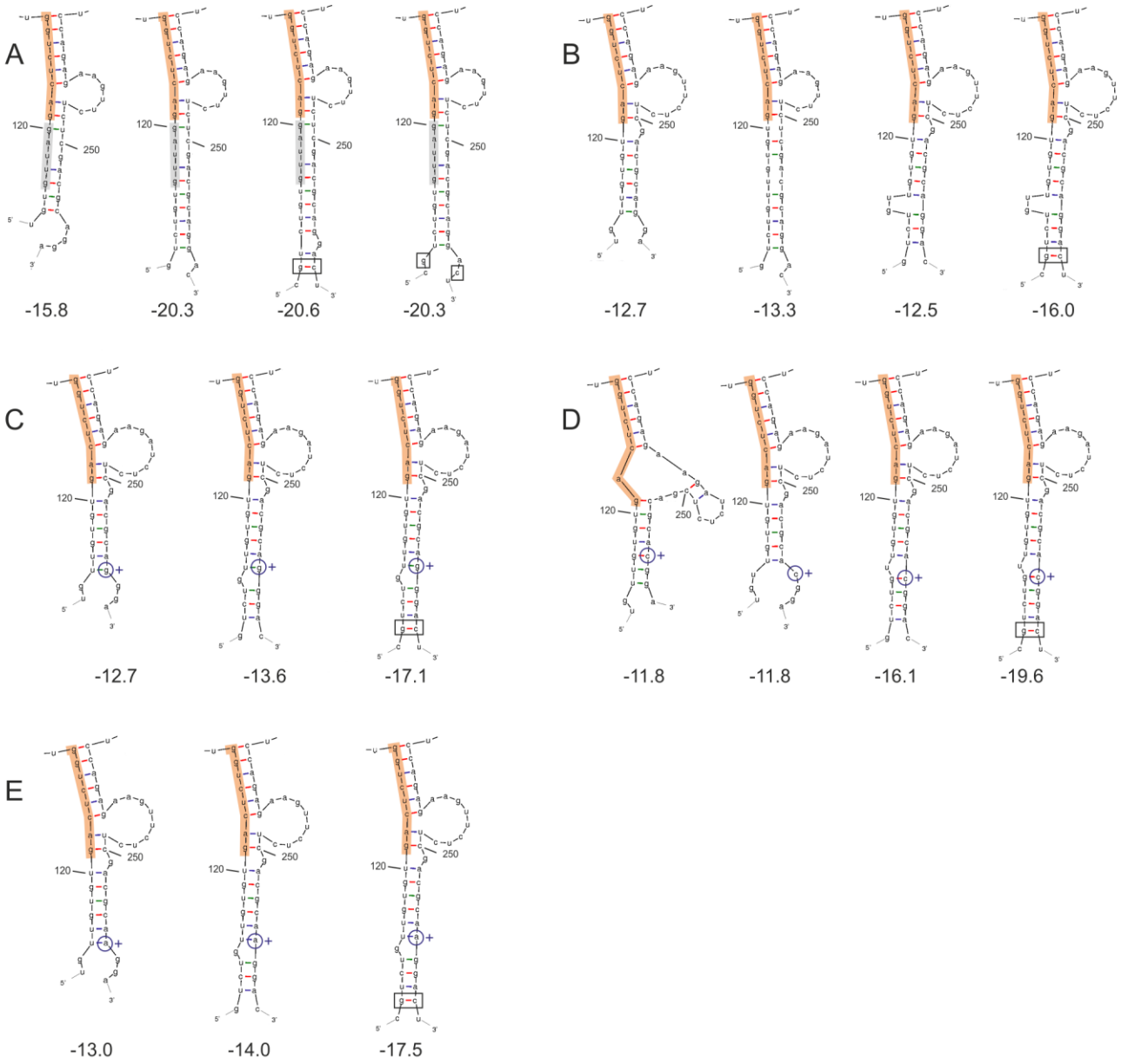

Figure S2. Mfold-calculated structures (ΔG in kcal/mol) of **PBS stem extension** in **CRF01_AE** isolates exemplified by 90CF402 (GenBank accession number U51188) (A), **subtype G** isolates - by DRCBL (AF084936) (B), **CRF02_AG** isolates – by DJ263 (AB485634) containing 256_257insG (C), 99CMBD6 (AY271690) containing 256_257insC (D) and 97GH-AG1 (AB049811) containing 256_257insA (E). The closing base pair 110G:260C is boxed. PAS motif is marked by orange. Mutated stretch ^{116}GUUAG^{120} in CRF01_AE isolates is shadowed (A), 1-nt insertion between positions 256 and 257 in CRF02_AG isolates is marked by plus (C, D, E).

Comment on Figure S2.

The involvement of G110 into the closing base pair G110:C260 into the PBS stem extension in CRF01_AE seems arguable, since its contribution into the stem stabilization is very low (0.3 kcal/mol) as compared to that in G subtype isolates (4.5 kcal/mol) and the frequent base change G110A (29% of CRF01_AE isolates) prevents formation of this base pair. Rare base changes occurring at both ends of extended fragment 2 are well tolerated by this structure, in particular, U118C (4 isolates) and G119U (4 isolates).

In CRF02_AG isolates, PBS stem extension is stabilized by 4.4 kcal/mol when 256_257insG (52%), 7.8 kcal/mol - 256_257insC (19%), 4.5 kcal/mol - 256_257insA (19%) and 5.2 kcal/mol - 257_258insA (10%).

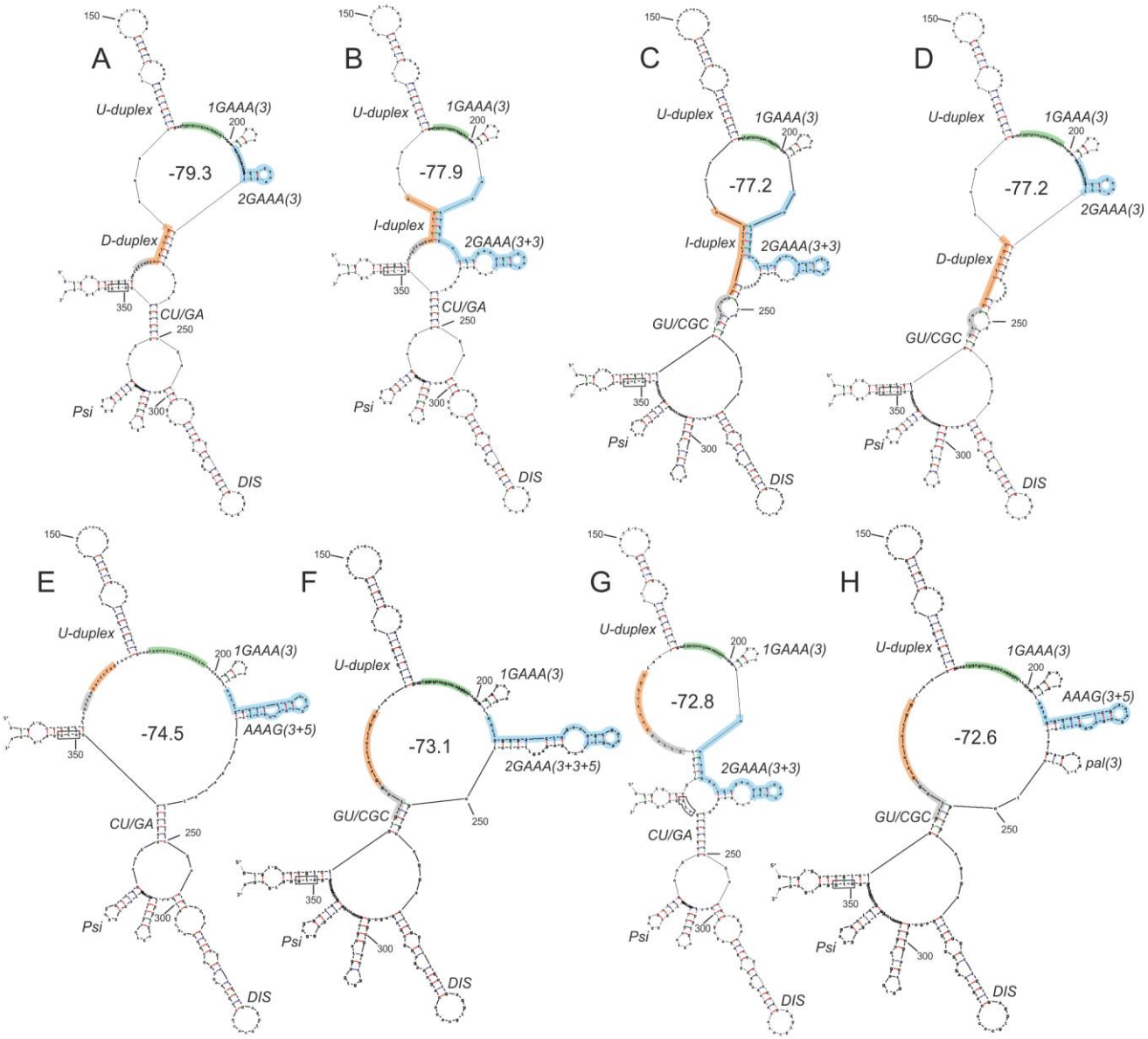

Figure S3. UNAfold-calculated structures (ΔG in kcal/mol) of **U5/AUG domain** (nt 102-359**,** MAL coordinates) of the **CRF01_AE isolate 90CF402** (GenBank accession number U51188). The most favorable structures with CU/GA-duplex and PBS region in D-duplex (A), I-duplex (B), PAS1b-like (E) and PAS1a (G) conformations; with purine-rich ring and PBS region in I-duplex (C), D-duplex (D), PAS2(gaaa) (F) and PAS2(pal) (H) conformations. Mutated stretch ^{116}GUUAG^{120} is shadowed. PBS motif, PAS motif and duplicate insertion are marked by green, orange and blue, respectively. The Gag initiation codon is boxed.

Comment on Figure S3.

According to the descending order of UNAfold-calculated structures of U5/AUG domain of the CRF01_AE isolate 90CF402, a hypothetical (speculative) scheme of switching between alternative conformations with D-duplex, I-duplex and an open PAS structure may include the following steps. The lowest free energy structure containing the CU/GA-duplex and PBS region in the D-duplex conformation (Fig. S3A) may switch to the most favorable structure containing the same CU/GA-duplex and PBS region in the I-duplex conformation (Fig. S3B) with the energy input of 1.4 kcal/mol. The next switch results in formation of the GU/CGC-duplex and a purine-rich ring structure and requires the energy input of 0.7 kcal/mol, PBS region being in the same I-duplex conformation (Fig. S3C). The third switch to the PAS2(gaaa) conformation may proceed in the purine-rich ring structure and the GU/CGC-duplex and requires the energy input of 4.1 kcal/mol (Fig. S3F). ΔΔG values of all four structures of U5/AUG domain with the resulting PAS2(gaaa) conformation of PBS region (Fig. S3A, B, C, F) are within the range of 0÷6.2 kcal/mol.

By an alternative scheme with resulting PAS2(pal) conformation, a switch from the I-duplex conformation to the PAS1b-like conformation with a self-contained motif of the duplicate insertion may proceed in the U5/AUG domain structure with the CU/GA-duplex (Fig. S3B, E) with the energy input of 3.4 kcal/mol, bypassing the U5/AUG domain structure with PBS region in the I-duplex conformation and the GU/CGC-duplex (Fig. S3C). The PAS1b-like conformation may switch to the PAS2(pal) conformation by formation of the GU/CGC-duplex and the pal(3) hairpin instead of the CU/GA-duplex, which requires the energy input of 1.9 kcal/mol (Fig. S3H). ΔΔG values of all four structures of U5/AUG domain with the resulting PAS2(pal) conformation of PBS region (Fig. S3A, B, E, H) are within the range of 0÷6.7 kcal/mol.

All schemes are fully speculative and presented just to illustrate possible structural rearrangements upon I-duplex formation and PAS exposure in a broader context than four fragments of U5-PBS region selected for this study.

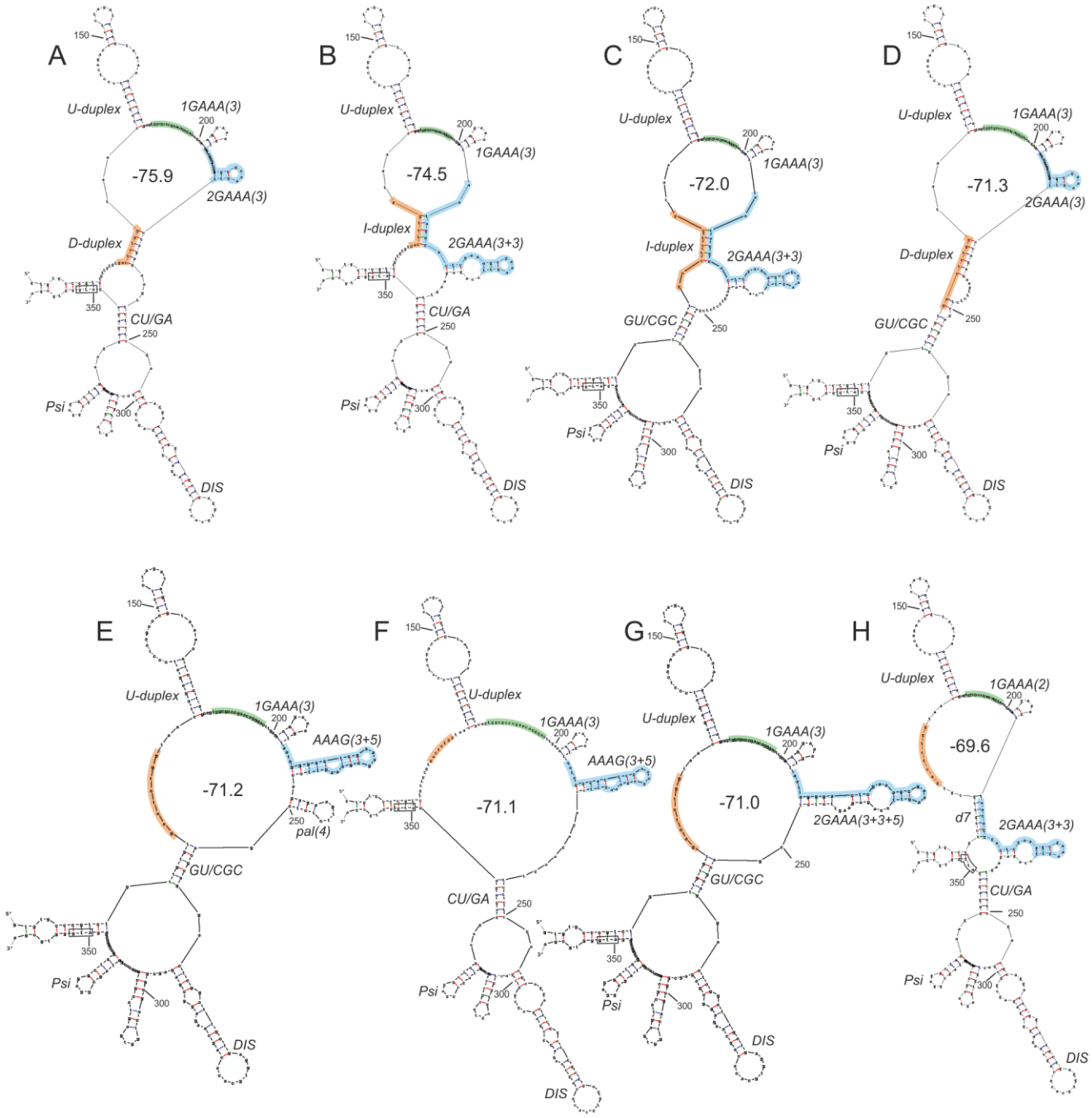

Figure S4. UNAfold-calculated structures (ΔG in kcal/mol) of **U5/AUG domain** (nt 102-359, MAL coordinates) of the **subtype G isolate DRCBL** (GenBank accession number AF084936). The most favorable structures with CU/GA-duplex and PBS region in D-duplex (A), I-duplex (B), PAS1b-like (F) and PAS1a (H) conformations; with purine-rich ring and PBS region in I-duplex (C), D-duplex (D), PAS2(pal) (E) and PAS2(gaaa) (G) conformations. PBS motif, PAS motif and duplicate insertion are marked by green, orange and blue, respectively. The Gag initiation codon is boxed.

Comment on Figure S4.

The lowest free energy structure containing the CU/GA-duplex and PBS region in the D-duplex conformation (Fig. S4A) may switch to the most favorable structure containing the same CU/GA-duplex and PBS region in the I-duplex conformation (Fig. S4B) with the energy input of 1.4 kcal/mol. The next switch results in formation of the GU/CGC-duplex and a purine-rich ring structure and requires the energy input of 2.5 kcal/mol, PBS region being in the same I-duplex conformation (Fig. S4C). The third switch to the PAS2(gaaa) conformation may proceed in the purine-rich ring structure and the GU/CGC-duplex and requires the energy input of 1.0 kcal/mol (Fig. S4G). ΔΔG values of all four structures of U5/AUG domain with the resulting PAS2(gaaa) conformation of PBS region (Fig. S4A, B, C, G) are within the range of 0÷4.9 kcal/mol.

By an alternative scheme with resulting PAS2(pal) conformation, the third switch from the I-duplex conformation to the PAS2(pal) conformation (Fig. S4C, E) requires the energy input of 0.8 kcal/mol. ΔΔG values of all four structures of U5/AUG domain with the resulting PAS2(pal) conformation of PBS region (Fig. S4A, B, C, E) are within the range of 0÷4.7 kcal/mol.

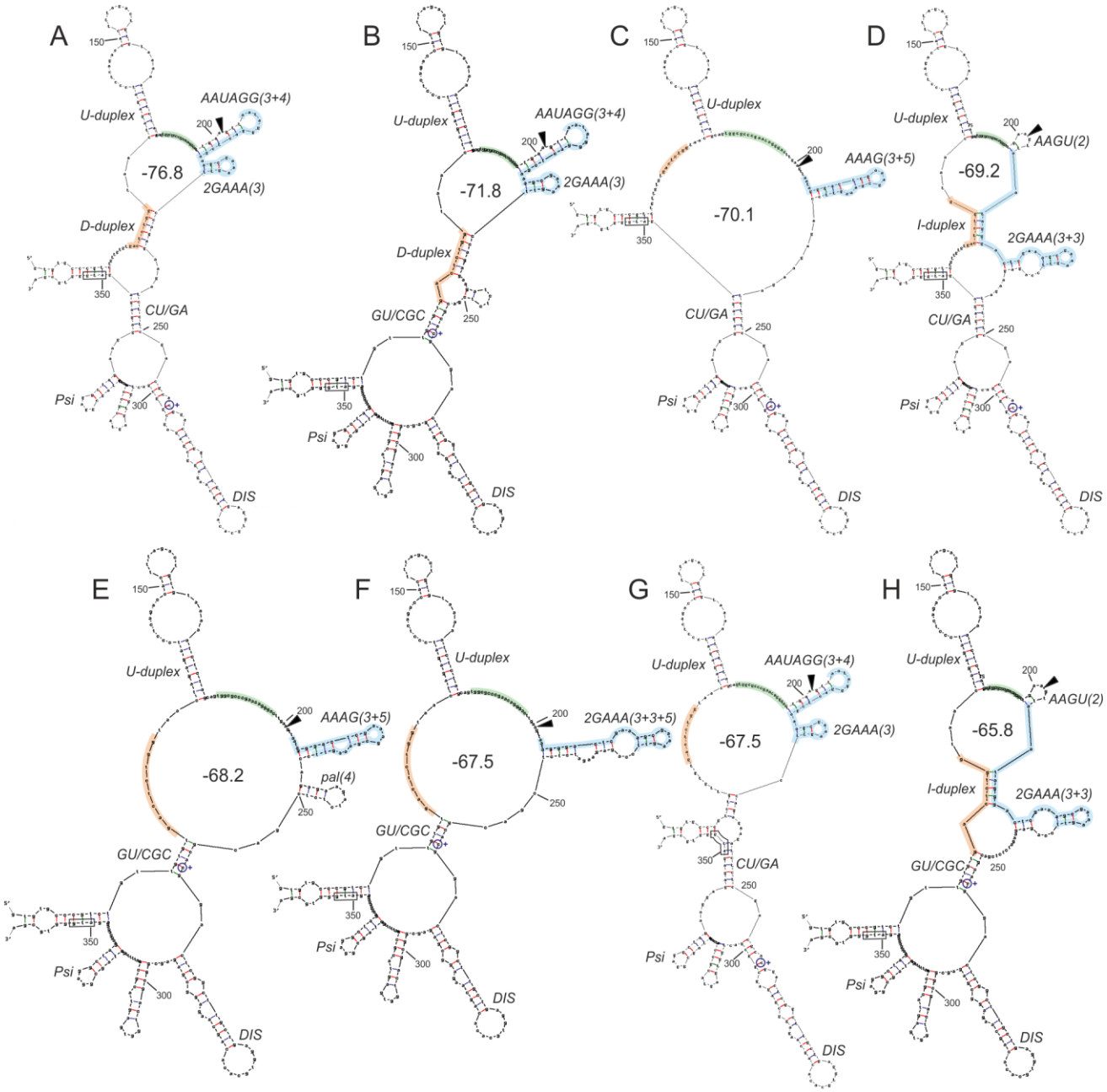

Figure S5. UNAfold-calculated structures (ΔG in kcal/mol) of **U5/AUG domain** (nt 102-359**,** MAL coordinates) of the **CRF02_AG isolate 99CMBD6** (GenBank accession number AY271690). The most favorable structures with CU/GA-duplex and PBS region in D-duplex (A), PAS1b-like (C), I-duplex (D) and PAS1*specific* (G) conformations; with purine-rich ring and PBS region in D-duplex (B), PAS2(pal) (E), PAS2(gaaa) (F) and I-duplex (H) conformations. A 7-nt deletion downstream of PBS and 1-nt insertion between positions 256 and 257 is indicated by triangle and by plus, respectively. PBS motif, PAS motif and duplicate insertion are marked by green, orange and blue, respectively. The Gag initiation codon is boxed.

Comment on Figure S5.

A switch from the D-duplex conformation to the I-duplex conformation within the structure with the CU/GA-duplex requires the energy input is 7.6 kcal/mol (Fig. S5A, D). The next switch implies a concurrent formation of the GU/CGC-duplex and PAS exposure (PAS2(gaaa) conformation) with the energy input of 1.7 kcal/mol (Fig. S5F). ΔΔG values of three structures of U5/AUG domain with resulting PAS2(gaaa) conformation (Fig. S5A, D, F) are within the range of 0÷9.3 kcal/mol. As compared with the subtype G isolate DRCBL, a switch to the I-duplex is greatly impeded in the CRF02_AG isolate 99CMBD6 (7.6 kcal/mol against 1.4 kcal/mol), while a formation of the GU/CGC-duplex and PAS exposure is notably advanced (1.7 kcal/mol against 3.5 kcal/mol).

Another scheme implies the first switch from CU/GA-duplex to the GU/CGC-duplex, PBS region being in the same D-duplex conformation, the energy input is 5.0 kcal/mol (Fig. S5A, B). The next step may include a switch from D-duplex to PAS exposure (PAS2(pal) conformation) within the structure with purine-rich ring and the GU/CGC-duplex, the energy input is 3.6 kcal/mol (Fig. S5E). ΔΔG values of three structures of U5/AUG domain with the resulting PAS2(pal) conformation (Fig. S5A, B, E) are within the range of 0÷8.6 kcal/mol. This scheme bypasses the I-duplex conformation (Fig. S5D), otherwise it should mean that a switch between the CU/GA-duplex and the GU/CGC-duplex occurs three times: the CU/GA-duplex to the GU/CGC-duplex (Fig. S5A, B), the GU/CGC-duplex to the CU/GA-duplex upon I-duplex formation (Fig. S5B, D) and again the CU/GA-duplex to the GU/CGC-duplex upon PAS exposure (Fig. S5D, F).

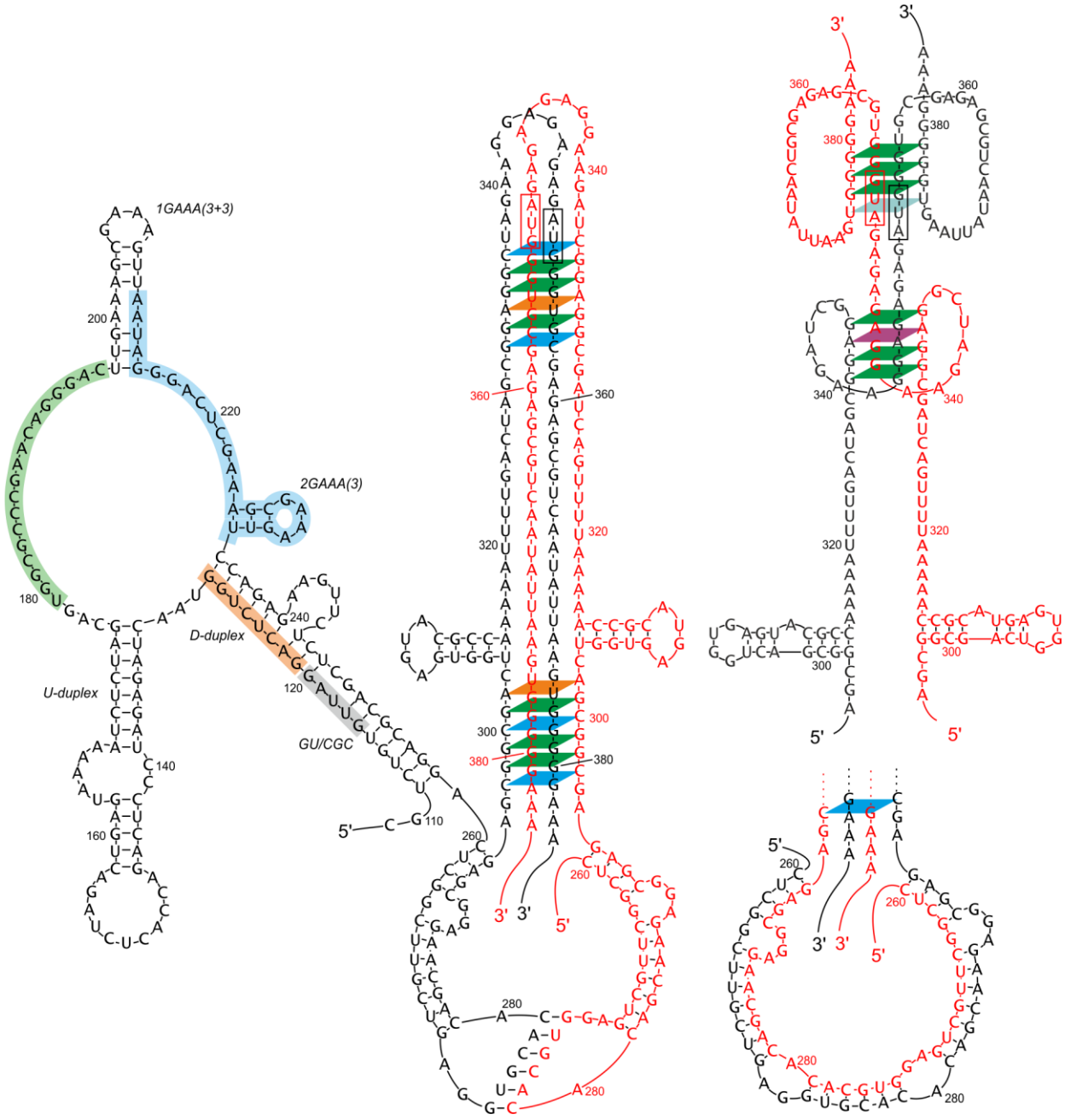

Figure S6. Structural model of the U5-PBS region extended fragment 2 and DLS region encompassing DIS kissing loop complex (or DIS extended duplex, shown in bottom insert) and intermolecular quadruplexes in antiparallel orientation (or parallel orientation, shown in upper insert) for the **CRF01_AE isolate 90CF402** (GenBank accession number U51188; nt 109-384, MAL coordinates). The extended fragment 2 is given in the optimal conformation possessing the D-duplex and PBS stem extension. Mutated stretch ^{116}GUUAG^{120} is shadowed. PBS motif, PAS motif and duplicate insertion are marked by green, orange and blue, respectively. The second copy of vRNA is given in red. G-tetrads are indicated by green, A-tetrads – by lilac, U-tetrads – by blue-gray, and mixed tetrads - by blue and orange. The Gag initiation codon is boxed.

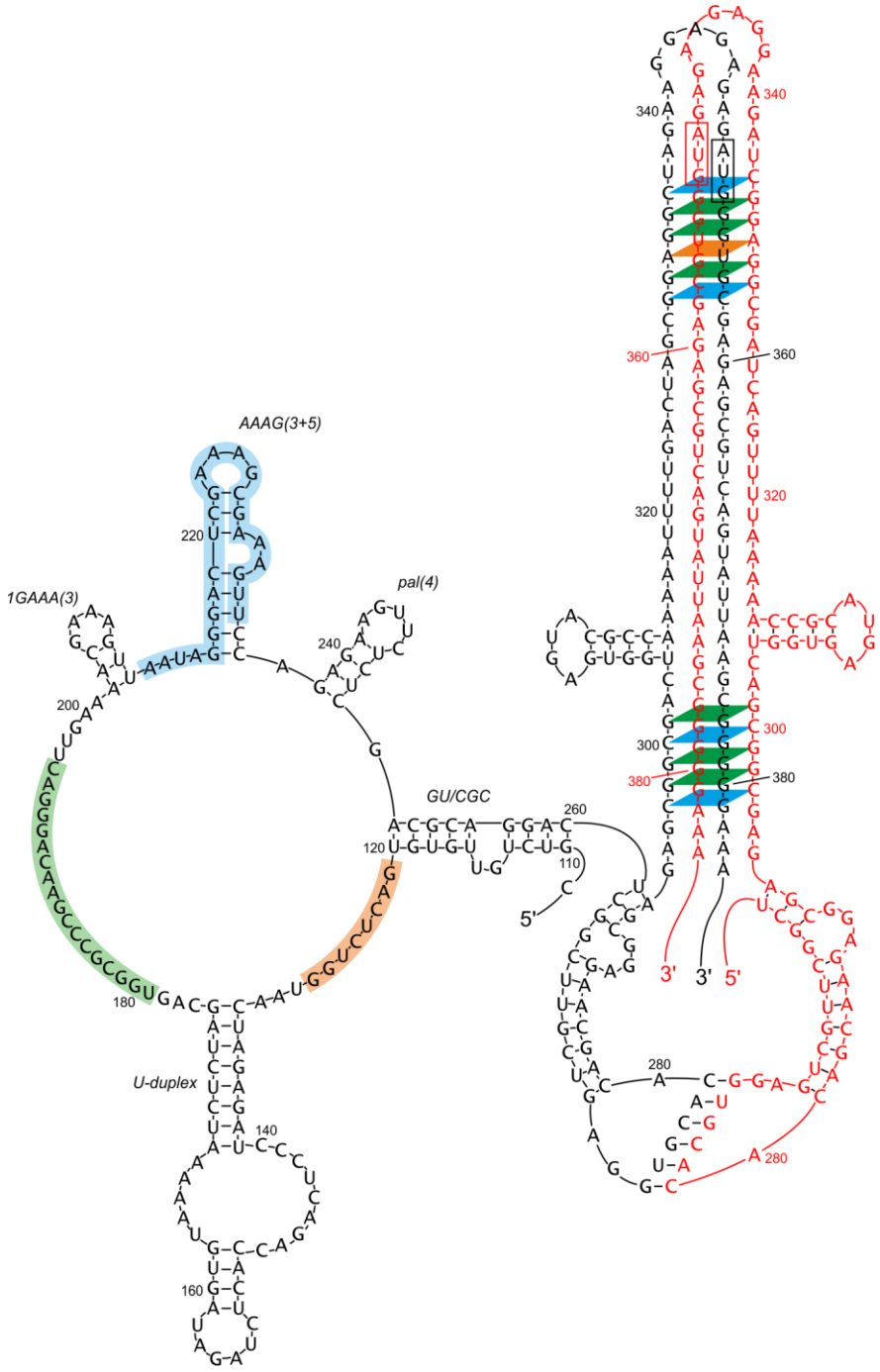

Figure S7. Structural model of the U5-PBS region extended fragment 2 and DLS region encompassing DIS kissing loop complex and intermolecular quadruplexes (nt 109-384, MAL coordinates) for the **subtype G isolate DRCBL** (GenBank accession number AF084936). The extended fragment 2 is given in the optimal conformation possessing an open PAS structure and PBS stem extension. PBS motif, PAS motif and duplicate insertion are marked by green, orange and blue, respectively. The second copy of vRNA is given in red. G-tetrads are indicated by green, mixed tetrads - by blue and orange. The Gag initiation codon is boxed.

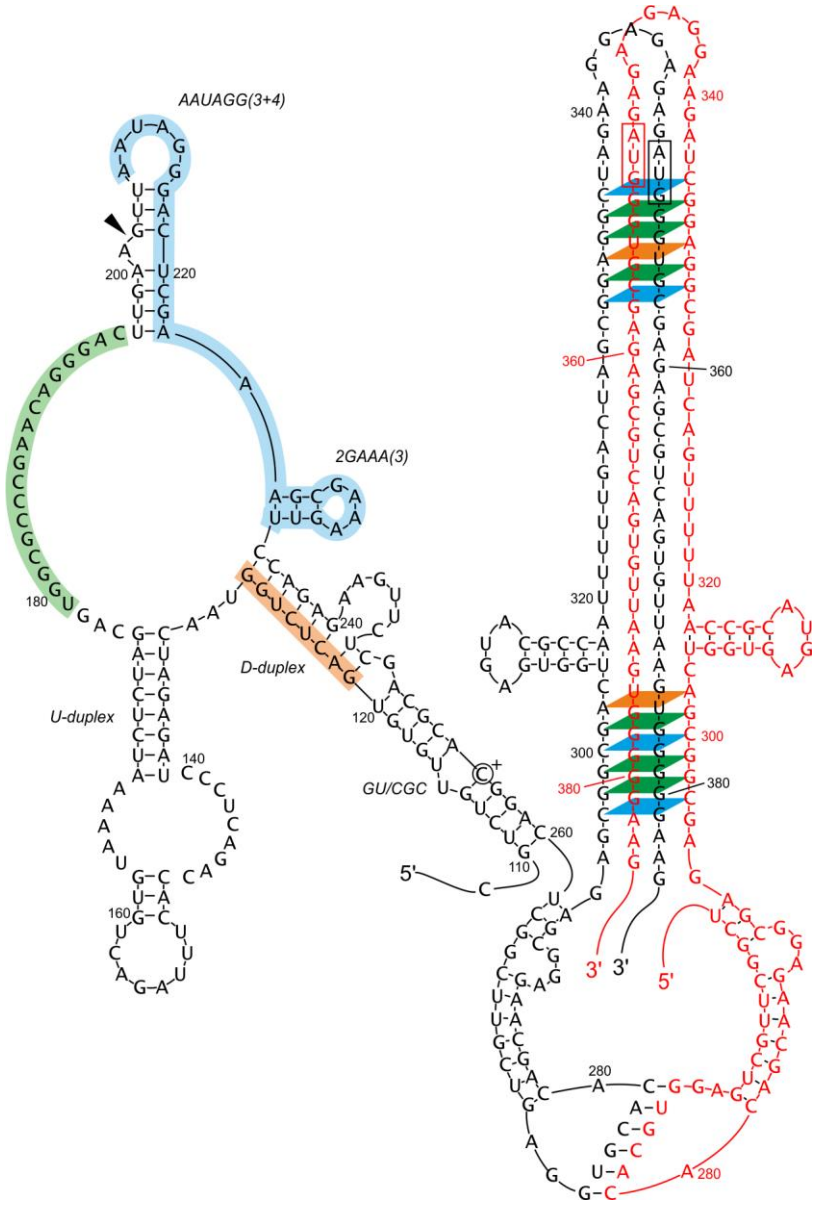

Figure S8. Structural model of the U5-PBS region extended fragment 2 and DLS region encompassing DIS kissing loop complex and intermolecular quadruplexes (nt 109-384, MAL coordinates) for the **CRF02_AG isolate 99CMBD6** (GenBank accession number AY271690). The extended fragment 2 is given in the optimal conformation possessing the D-duplex and PBS stem extension. A 7-nt deletion downstream of PBS and 1-nt insertion between positions 256 and 257 is indicated by triangle and by plus, respectively. PBS motif, PAS motif and duplicate insertion are marked by green, orange and blue, respectively. The second copy of vRNA is given in red. G-tetrads are indicated by green, mixed tetrads - by blue and orange. The Gag initiation codon is boxed.