Supplementary Information

# Zheng et al., Identification of drivers for the metamorphic transition of HIV-1 reverse transcriptase

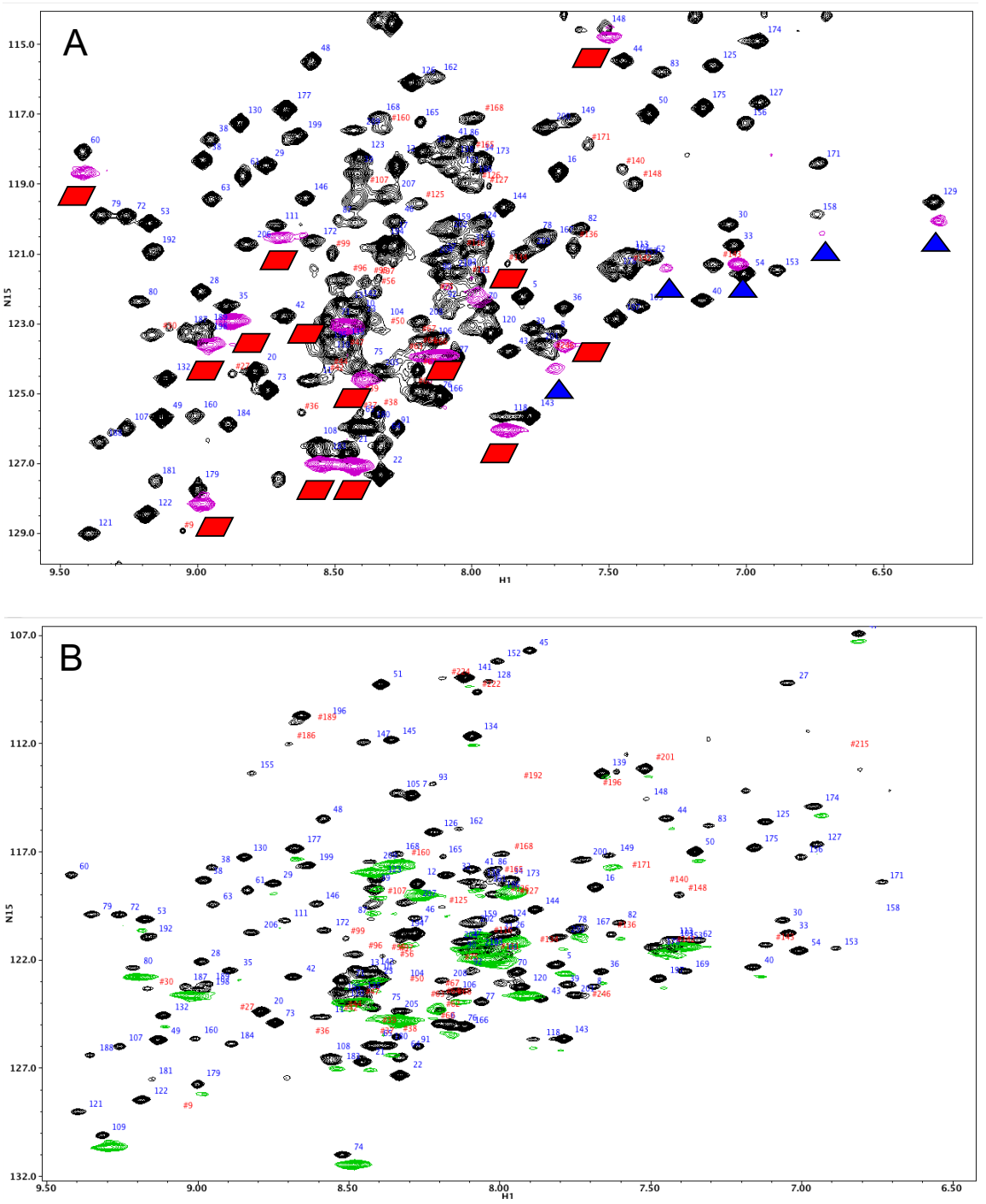


**Figure S1. Residue-specific labeling of RT216.** A) Assigned $^{1}$H-$^{15}$N TROSY spectrum of U-[$^{2}$H,$^{13}$C,$^{15}$N]RT216 overlaid with spectra of U-[$^{2}$H,$^{15}$N-Val]RT216 and U-[$^{2}$H,$^{15}$N-Ala]RT216 (magenta). The resonances identified in the valine labeling study are indicated by red parallelograms, and those identified using the $^{15}$N-alanine are identified by blue triangles. The spectra obtained with the residue-specific labels are offset in the $^{15}$N dimension. B) Assigned $^{1}$H-$^{15}$N TROSY spectrum of U-[$^{2}$H,$^{13}$C,$^{15}$N]RT216 (black) overlaid with spectra of U-[$^{2}$H,$^{15}$N-Leu]RT216 (green). The spectra of the Leu-labeled sample are offset in the $^{15}$N dimension.

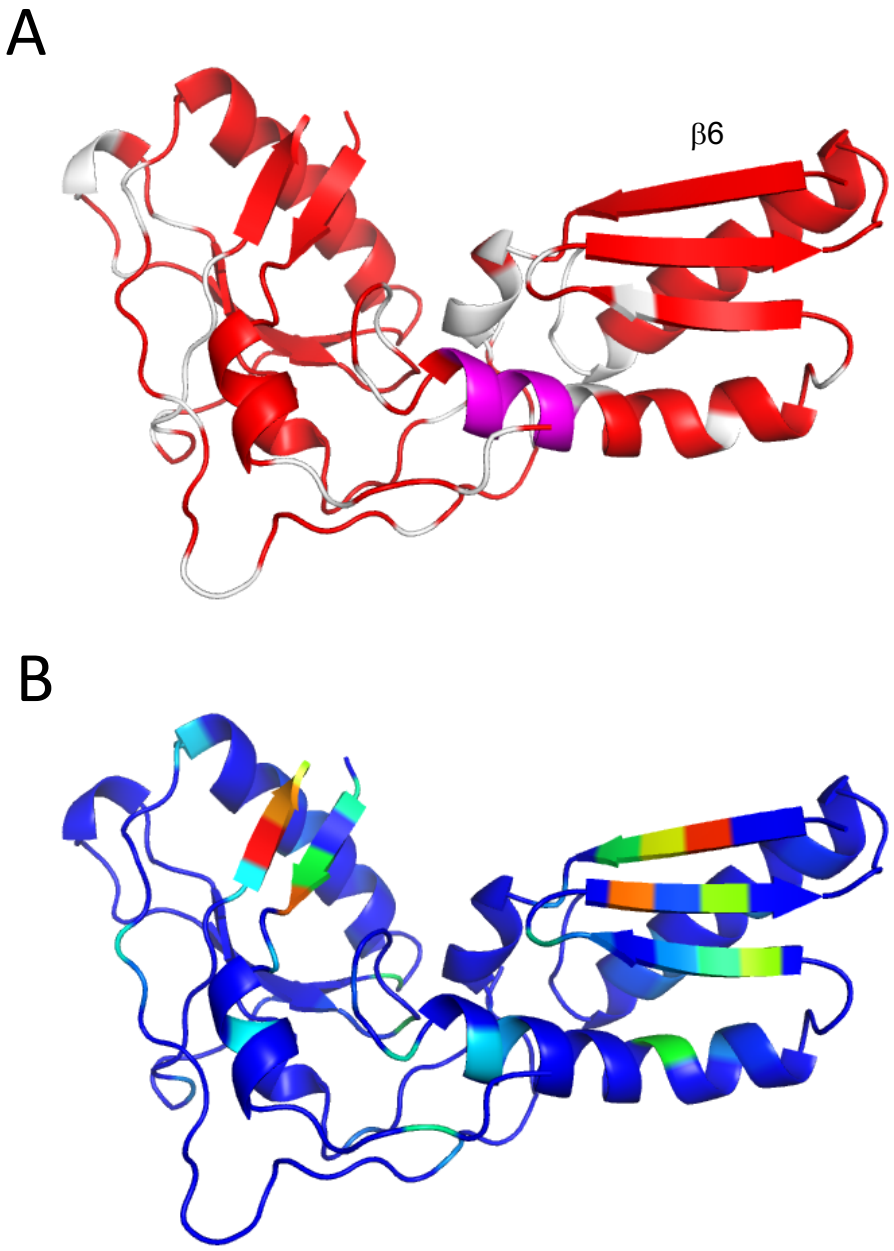


**Figure S2. Extent of RT216 assignments and location of residues exhibiting concentration-dependent amide shifts.** A) assigned residues – red or, for residues Helix E residues 157-163, magenta. Unassigned residues are white. B) Concentration dependent shifts – blue – unassigned or no significant shift; other ecolored residues correspond to shifted residues, following a rainbow ordering with red corresponding to the largets shifts.

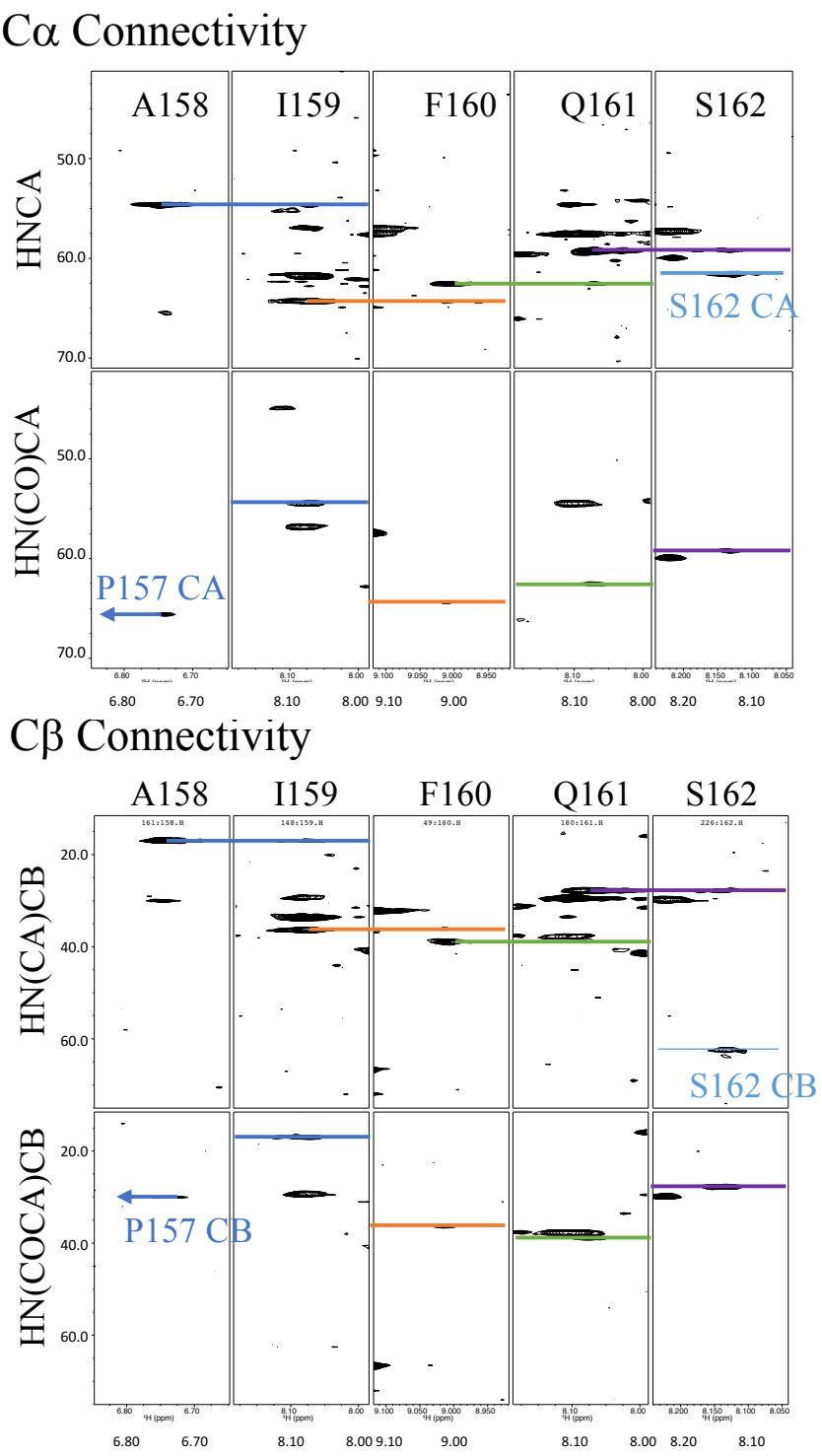


**Figure S3. Assignments of RT216 residues 158-162.** Assignments were based on HNCA, HN(CO)CA, HN(CA)CB, and HN(COCA)CB experiments. The assignment strategy is illustrated for residues 158-162 that were subjected to further TALOS analysis in order to evaluate the conformation of this segment of α-helix E. The CA and CB shifts for Ala158 are easily identifiable as they are very low in both ranges, the methyl CB being below 20 ppm. The Ser 162 CA and CB shifts are also characteristic of a serine, with both being above 60 ppm. Also the rather high CA shift of P157 is another distinguishing feature of this segment. All three of these are residue type specific features. Next the connectivity between the Ala158 and S162 is strongly supported by all $CA_i$, $CA_{i-1}$, $CB_i$, and $CB_{i-1}$ being visible in all of the appropriate spectra.

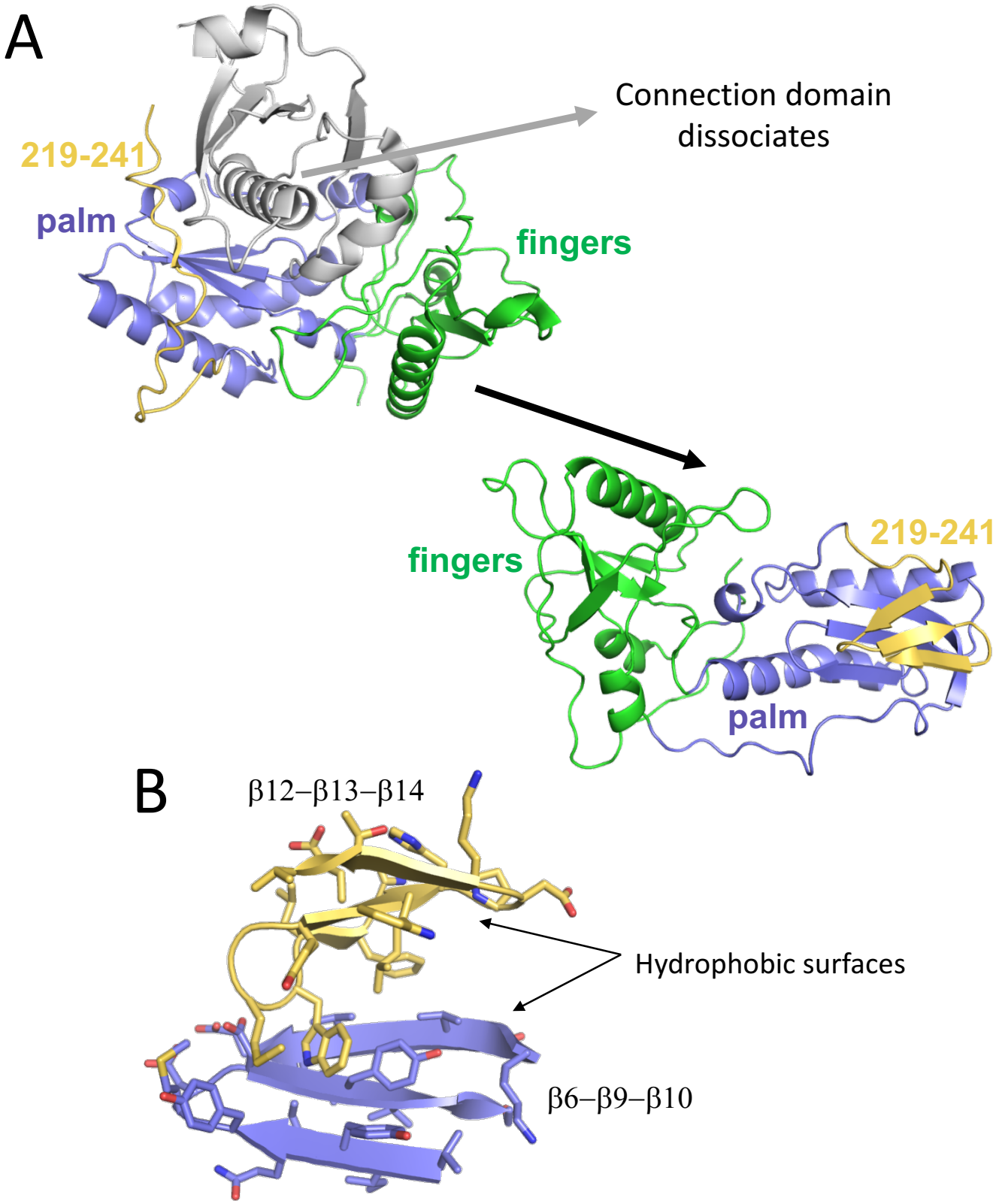


**Figure S4. Palm domain β-sheet formation. A)** Unimolecular dissociation of the fingers/palm:connection complex in the monomer is viewed as the first step of the metamorphic transition. Removal of the connection domain (gray) exposes a hydrophobic surface of the palm domain (blue) that includes the β6–β9–β10 sheet. The exposed hydrophobic surface mediates aggregation of RT216, but in p66, it interacts with the C-terminal segment of the palm domain (yellow), helping to stabilize the nascent β-sheet formed from residues 226-241. A fourth β-strand (not shown) is derived from the connection domain. B) The interacting hydrophobic surfaces of the two β-sheets are shown.

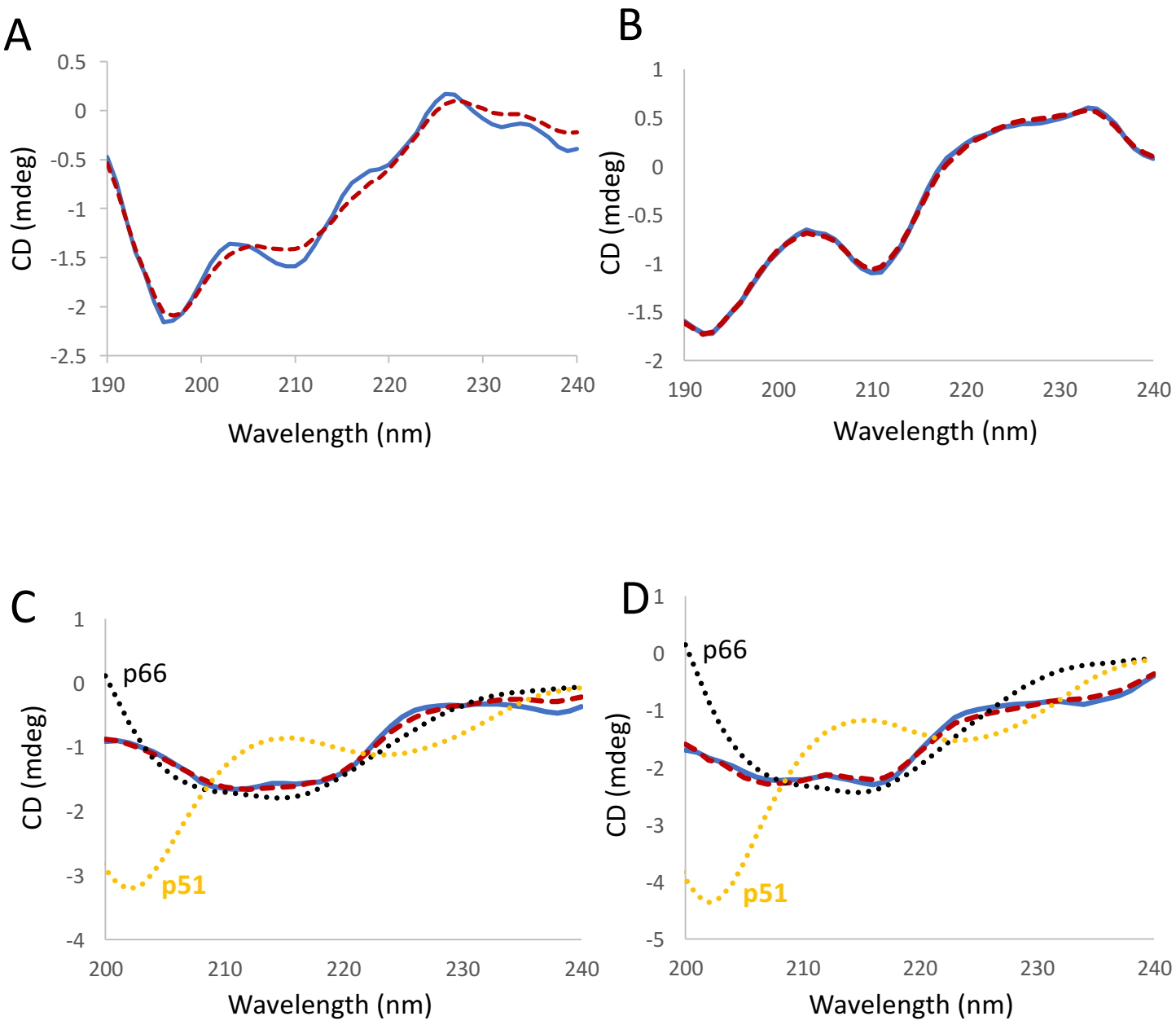


Figure S5. Illustrative CD spectra of palm C-terminal peptides. A) 20 µM PFL in 20 % EtOH; B) 10 µM PYK in buffer; C) 15 µM PFL in 40 % EtOH; D) 20 µM PFL in 40 % EtOH. The sample buffer also contained 10 mM sodium phosphate, pH 8.0. The blue curves correspond to the experimental data, the dashed red curves correspond to the fits obtained using the program DICHROWEB [1]. Figures C and D also contain theoretical curves generated for residues 226-241 in either the p66 subunit (black dotted curve) or the p51 subunit (yellow dotted curve) of RT (PDB accession number: 3DLK, [2]), generated with the PDB2CD on-line server (http://pdb2cd.cryst.bbk.ac.uk) [3]. They are scaled by 0.7 and 0.95 for the comparison with CD spectra from 15 µM and 20 µM peptides, respectively. The higher ethanol levels are expected to reduce the tendency of the hydrophobic peptide to aggregate, and the curves show much closer correspondence to the CD spectra expected for a fold similar to that in the p66 subunit.

Table S1. Secondary structure analysis of CD data using DICHROWEB [1]

A. RT residues 226-241: PFLWMGYELHPDKWTV

| EtOH | Concentration (μM) | Total α-helix[a] | Total β-sheet[b] | Disordered | Total |
|---|---|---|---|---|---|
| 10 % | 10 | 0.15 | 0.45 | 0.39 | 0.99 |
| | 15 | 0.05 | 0.37 | 0.57 | 0.99 |
| | 20 | 0.04 | 0.28 | 0.68 | 1.00 |
| 20 % | 10 | -0.03 | 0.56 | 0.44 | 0.97 |
| | 15 | 0.00 | 0.49 | 0.48 | 0.97 |
| | 20 | 0.01 | 0.55 | 0.43 | 0.99 |
| 40 % | 10 | N/A[c] | N/A | N/A | N/A |
| | 15 | 0.05 | 0.62 | 0.33 | 1.00 |
| | 20 | 0.17 | 0.57 | 0.27 | 1.01 |

B. RT(226-241) containing three hydrophilic substitutions: PYKWMGYELHPDKWTK

| Concentration (μM) | Total α-helix[a] | Total β-sheet[b] | Disordered | Total |
|---|---|---|---|---|
| 10 | -0.02 | 0.56 | 0.42 | 0.96 |
| 15 | 0.01 | 0.46 | 0.52 | 0.99 |
| 20 | -0.01 | 0.41 | 0.58 | 0.98 |

[a]Summation of helix1 and helix2 as defined by DICHROWEB.
[b]Summation of β-strands and turns.
[c]The estimation is not available due to low reliability of raw data in UV region from190 to 200 nm)