Supporting information

Catalytic Effects of Mutations of Distant Protein Residues in Human DNA Polymerase β: Theory and Experiment

Martin Klvaňa, Drew L. Murphy, Petr Jeřábek, Myron F. Goodman, Arieh Warshel, Joann B. Sweasy, Jan Florián

Equations S1-12: Cancelling $\Delta\Delta G_{solv}$ terms from thermodynamic cycle (Figure 2) for calculating the relative binding free energies of WT (*R*) and mutant (*R'*) amino acid residue; *E* – enzyme; *W* – water; *gas* – gas phase.

$$\Delta G_E - \Delta G_{bind}(R') - \Delta\Delta G_{solv} + \Delta G_{bind}(R) = 0 \quad (S1)$$

$$\Delta G_E = \Delta G_{bind}(R') + \Delta\Delta G_{solv} - \Delta G_{bind}(R) \quad (S2)$$

$$\Delta G_{bind}(R') = G(R')_E - G(R')_W \quad (S3)$$

$$\Delta G_{bind}(R) = G(R)_E - G(R')_W \quad (S4)$$

$$\Delta\Delta G_{solv} = \Delta G_{solv}(R') - \Delta G_{solv}(R) \quad (S5)$$

$$\Delta G_{solv}(R') = G(R')_W - G(R')_{gas} \quad (S6)$$

$$\Delta G_{solv}(R) = G(R)_W - G(R)_{gas} \quad (S7)$$

$$\Delta G_E = G(R')_E - G(R')_W + G(R')_W - G(R')_{gas} - G(R)_W + G(R)_{gas} - G(R)_E + G(R')_W \quad (S8)$$

$$G(R')_{gas} = 0 \quad (S9)$$

$$G(R)_{gas} = 0 \quad (S10)$$

$$\Delta G_E = G(R')_E \cancel{- G(R')_W + G(R')_W} - 0 \cancel{- G(R)_W} + 0 - G(R)_E \cancel{+ G(R)_W} \quad (S11)$$

$$\Delta G_E = G(R')_E - G(R)_E \quad (S12)$$

Table S1: Atomic charges (a.u.) of electroneutral phosphate group and ionizable amino acid residues[a].

Atom	Charge	Atom	Charge
Phosphate group		*Arginine*	
P	1.1659	N	-0.4157
OP	-0.2761	H	0.2719
		CA	-0.0597
Aspartate		HA	0.0869
N	-0.4157	CB	0.1303
H	0.2719	HB	0.0187
CA	0.0341	CG	-0.0430
HA	0.0864	HG	0.0236
CB	0.1316	CD	-0.0660
HB	0.0488	HD	0.0186
CG	0.7755	NE	-0.8000
OD	-0.5054	HE	0.3456
C	0.5973	CZ	0.3327
O	-0.5679	NH	-0.8627
		HH	0.4478
Glutamate		C	0.5973
N	-0.4157	O	-0.5679
H	0.2719		
CA	0.0341	*Lysine*	
HA	0.0864	N	-0.4157
CB	0.0771	H	0.2719
HB	0.0256	CA	-0.2400
CG	0.0149	HA	0.1426
HG	0.0430	CB	-0.0094
CD	0.7755	HB	0.0362
OE	-0.5054	CG	-0.0907
C	0.5973	HG	0.0103
O	-0.5679	CD	-0.1200
		HD	0.0621
		CE	-0.0723
		HE	0.0335
		NZ	-0.8000
		HZ	0.3400
		C	0.5973
		O	-0.5679

[a] Our experience gained from countless calculations (long before most research groups) with studies using the SCAAS +LRF spherical boundary conditions is that the far ionized residues should be neutralized and then, if needed, treated with Coulomb law and a dielectric of 40 or more.

Table S2: FEP alchemistic mutagenesis.

Mutations[a]	Number of windows (N)	Set of λ values[b]
L-Hil282→M/L[c]	101	{1.0000, 0.9900, 0.9800, 0.9700, . . ., 0.0400, 0.0300, 0.0200, 0.0100, 0.0000}
I174→S, Q260→Q, H285→D, K289→M	104	{1.0000, 0.9900, 0.9800, 0.9700, . . ., 0.0100, 0.0050, 0.0010, 0.0001, 0.0000}
K288→E, R283→A, R283→L	203	{1.0000, 0.9950, 0.9000, 0.9850, . . ., 0.0100, 0.0050, 0.0010, 0.0001, 0.0000}

[a] Arrows indicate direction of FEP alchemistic mutagenesis; *State 1* → *State 2*;

[b] Each FEP simulation was subdivided into N separate MD simulations (N windows) that differed in the value of the coupling parameter λ. The potential energy surface used in the i^{th} window was defined as $E_i = (1 - \lambda_i) E_{State\ 1} + \lambda_i E_{State\ 2}$, i = 1, 2, ...N; $\lambda_1 = 0$ and $\lambda_N = 1$.

[c] See Supplementary Figure S2.

Table S3: Average RMSD (Å) of the thumb subdomain in LIE simulations of binary (E) and transition state (TS) complexes of Pol β containing right (*R*) or wrong (*W*) dNTP substrate with respect to the crystal structures with open (1BPX), closed (2FMP) and partially open (3C2M) conformation of the thumb sub-domain.[a]

Variant	1BPX			2FMP			3C2M		
	E	TS (*R*)	TS (*W*)	E	TS (*R*)	TS (*W*)	E	TS (*R*)	TS (*W*)
WT[b]	1.3 ± 1.4	5.0 ± 1.4	4.2 ± 1.4	3.7 ± 2.7	0.7 ± 0.8	0.8 ± 1.1	2.5 ± 2.3	1.3 ± 1.2	0.8 ± 1.0
I174S	1.9 ± 0.7	5.0 ± 0.6	4.0 ± 0.4	2.3 ± 0.8	0.6 ± 0.3	0.9 ± 0.3	1.3 ± 0.6	1.3 ± 0.5	0.8 ± 0.3
I260Q	3.7 ± 0.7	5.2 ± 0.5	2.8 ± 0.5	2.5 ± 0.4	0.7 ± 0.3	1.5 ± 0.4	2.2 ± 0.4	1.3 ± 0.4	0.7 ± 0.2
M282L	2.1 ± 0.5	5.2 ± 0.4	4.6 ± 0.7	2.0 ± 0.6	0.5 ± 0.2	1.0 ± 0.4	1.2 ± 0.5	1.4 ± 0.4	1.3 ± 0.5
H285D	0.9 ± 0.4	5.0 ± 0.6	3.8 ± 0.5	3.3 ± 0.9	0.8 ± 0.3	0.8 ± 0.3	2.0 ± 0.8	1.4 ± 0.4	0.8 ± 0.3
E288K	1.5 ± 0.6	5.4 ± 0.4	3.4 ± 0.4	3.0 ± 0.7	0.7 ± 0.3	1.2 ± 0.3	1.7 ± 0.6	1.6 ± 0.3	0.5 ± 0.2
K289M	0.8 ± 0.4	4.8 ± 0.6	4.0 ± 0.6	3.4 ± 0.8	0.5 ± 0.2	1.6 ± 0.3	2.2 ± 0.7	1.1 ± 0.4	1.2 ± 0.4
R283A	0.9 ± 0.5	4.6 ± 0.7	3.8 ± 0.4	3.4 ± 0.6	0.6 ± 0.3	1.0 ± 0.3	2.2 ± 0.5	1.0 ± 0.5	0.8 ± 0.3
R283L	0.9 ± 0.3	4.3 ± 0.6	4.5 ± 0.5	4.6 ± 1.1	0.6 ± 0.2	0.6 ± 0.3	3.3 ± 1.0	0.8 ± 0.4	0.9 ± 0.4

[a] LIE simulations were initiated from 1BPX (E), 2FMP (TS, *R*) and 3C2M (TS, *W*).

[b] Average of eight independent MD simulations.

Table S4. Average RMSD of thumb sub-domain in LIE simulations of binary (E) and transition state (TS) complexes with respect to their initial X-ray crystal structures.

Variant	2FMP[a]		3C2M[a]	
	E	TS	E	TS
I174 (WT)[b]	1.2 ± 0.4	0.5 ± 0.3	1.7 ± 0.8	1.0 ± 0.4
S174	0.9 ± 0.3	0.6 ± 0.3	1.0 ± 0.3	0.8 ± 0.3
I260 (WT)[b]	1.1 ± 0.3	0.9 ± 0.4	1.2 ± 0.5	1.1 ± 0.4
Q260	0.8 ± 0.3	0.7 ± 0.3	1.3 ± 0.6	0.7 ± 0.2
R283 (WT)[b]	0.7 ± 0.3	0.7 ± 0.3	0.9 ± 0.3	0.9 ± 0.4
A283	1.1 ± 0.6	0.6 ± 0.3	1.3 ± 0.6	0.8 ± 0.3
L283	1.2 ± 0.6	0.6 ± 0.2	0.8 ± 0.3	0.9 ± 0.4

[a] Initial X-ray crystal structure used in LIE simulations.
[b] Individual WT simulations differ in their definition of the probe-region of the simulated system; the probe region included the side-chain of the amino acid listed in the left column.

Table S5. Relative free energies (kcal/mol) of Pol β point mutations calculated using the FEP method.

Mutant	ΔG_E		ΔG_{TS} (GC)		ΔG_{TS} (GA)	
	FEP[a]	FEP/LIE[b]	FEP[a]	FEP/LIE[b]	FEP[a]	FEP/LIE[b]
I174S	1.8	5.9	0.4	5.6	0.8	5.7
I260Q	-32.8	-31.3	-32.8	-30.6	-31.5	-30.5
M282L	-8.7	-8.6	-9.0	-8.1	-8.8	-8.5
H285D	-16.6	-6.0	-18.2	-6.2	-17.7	-5.3
E288K	12.7	5.2	11.1	4.0	11.9	4.8
K289M	25.3	24.0	26.3	24.6	26.5	25.0
R283A	118.5	120.9	120.3	122.4	118.5	121.0
R283L	88.5	88.7	90.9	90.8	86.4	87.1

[a] Free energies scaled by an empirical factor of 0.5.
[b] hybrid FEP/LIE method (eq 10).

Table S6: Electrostatic (*ES*) and van der Waals (*vdw*) contributions to ΔG (in kcal/mol) calculated by the LIE method.[a]

Mutation	ΔG_E		ΔG_{TS}(GC)		ΔG_{TS}(GA)	
	ES	*vdw*	*ES*	*vdw*	*ES*	*vdw*
I174S	-6.6	4.1	-6.7	3.4	-6.7	3.6
I260Q	-6.8	0.2	-9.3	1.0	-8.5	0.0
M282L	-3.8	0.9	-3.6	1.5	-4.2	1.2
H285D	-50.2	6.2	-51.1	6.4	-52.0	6.9
E288K	6.5	-3.0	3.5	-2.9	5.4	-3.5
K289M	59.2	-2.4	59.2	-2.8	60.5	-2.6
R283A	60.5	4.0	54.8	6.8	64.0	4.3
R283L	53.7	2.0	48.3	2.9	57.1	1.8

[a] $\alpha = 0.45$; $\beta = 0.43$ [see eq. 7 and 8];

Table S7: Electrostatic (*ES*) and van der Waals (*vdw*) contributions to ΔG (in kcal/mol) calculated by FEP.[a]

Mutation	ΔG_E			ΔG_{TS}(GC)			ΔG_{TS}(GA)		
	QQ[b]	*ES*	*vdw*	*QQ*	*ES*	*vdw*	*QQ*	*ES*	*vdw*
I174S	3.7	-2.0	0.2	4.0	-1.8	-1.8	4.0	-1.9	-1.3
I260Q	-24.2	-7.4	-1.2	-24.2	-7.4	-1.2	-23.8	-6.7	-0.9
M282L	-10.1	0.6	0.8	-10.2	0.6	0.6	-10.5	0.8	0.9
H285D	16.9	-29.2	-4.0	17.4	-29.9	-5.1	17.1	-29.2	-4.2
E288K	4.0	4.2	3.3	3.7	3.2	3.8	2.4	5.9	3.8
K289M	-14.4	40.9	-1.2	-14.7	42.1	-1.2	-14.7	42.3	-1.1
R283A	74.5	42.4	1.4	74.2	41.4	5.3	73.9	42.8	3.2
R283L	57.6	29.1	1.7	59.3	27.2	4.3	55.8	30.0	0.6

[a] Both *ES* and *vdw* contributions scaled by a factor of 0.5;

[b] QQ = *ES*qq + *vdW*qq

Table S8. ΔGsolv (kcal/mol) calculated using the LIE method.[a]

Residue	ΔG_{solv}(E)	ΔG_{solv}(TS,GC)	ΔG_{solv}(TS,GA)
I174	-7.5	-6.6	-6.7
S174	-10.0	-9.9	-9.9
I260	-8.0	-8.1	-8.6
Q260	-14.5	-16.5	-17.0
M282	-5.3	-5.8	-5.3
L282	-8.2	-8.0	-8.3
H285	-14.7	-14.2	-14.0
D285	-58.7	-59.0	-59.2
E288	-53.1	-51.6	-52.8
K288	-49.6	-51.0	-50.9
K289	-58.7	-58.5	-59.8
M289	-2.0	-2.0	-1.9
R283	-66.3	-65.5	-70.6
A283	0.1	-0.1	0.2
L283	-8.6	-9.1	-9.6

[a] $\alpha = 0.45$; $\beta = 0.43$ (see eq. 7 and 8)

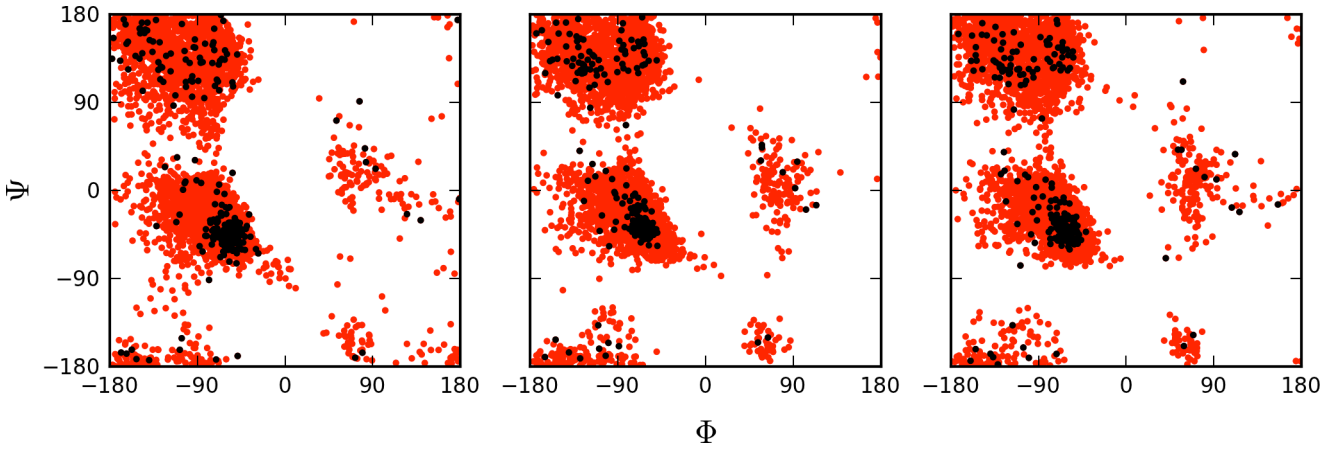

Figure S1: Ramachandran plots of distribution of ϕ and ψ dihedral angles in X-ray crystal structures (black) and MD simulations of WT Pol β (red). *Left,* 1BPX X-ray crystal structure *vs.* MD simulation of open binary complex initiated from 1BPX; *center,* 2FMP X-ray crystal structure *vs.* MD simulation of $\mathcal{R}$ TS initiated from 2FMP; *right,* 3C2M X-ray crystal structure *vs.* MD simulation of $\mathcal{W}$ TS initiated from 3C2M. 10-ns MD trajectories were generated using ff94 force field and sampled for the dihedral angles every 500 ps; only residues included in the 33 Å simulation sphere are shown.

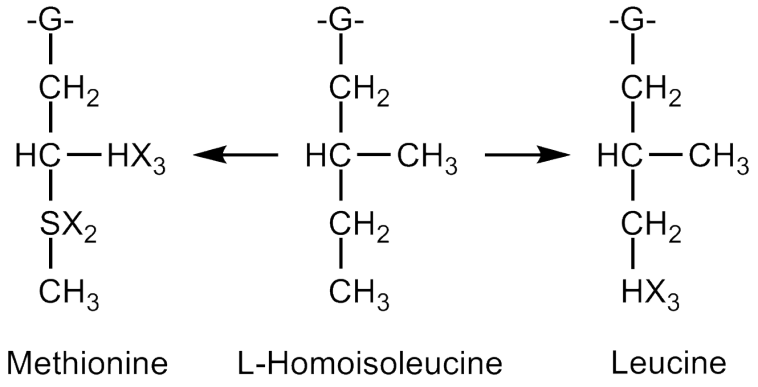

Figure S2: ΔG_{FEP} values for M282L mutation were calculated by combining two FEP simulations: FEP mutation of L-homoisoleucine (Hil) to methionine and FEP mutation of L-homoisoleucine to leucine; ΔG_{FEP} = $-\Delta G$(Hil → Met) + ΔG(Hil → Leu). -G-, glycine fragment of the amino acid residues; X, dummy atom.

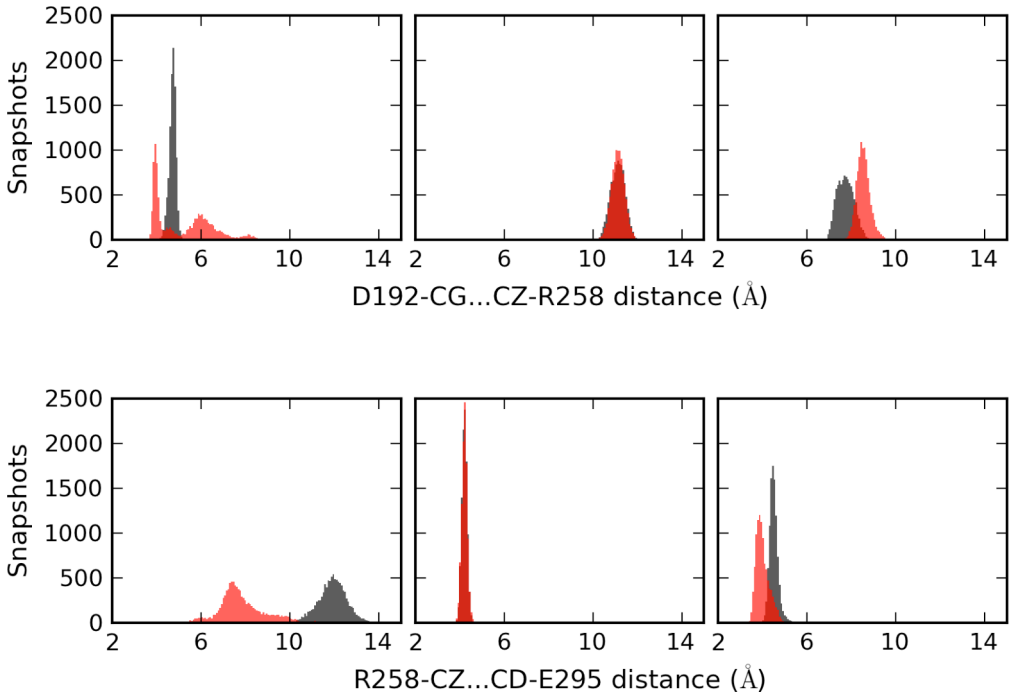

Figure S3: Histograms of distances between active site residues in MD simulations of WT (grey) and I174S mutant (red) of Pol β. *Left,* MD simulations of binary complexes initiated from 1BPX (open) X-ray crystal structure; *center,* MD simulations of $\mathcal{R}$ TS complexes initiated from 2FMP (closed) X-ray crystal structure; *right,* MD simulations of $\mathcal{W}$ TS complexes initiated from 3C2M (partially open) X-ray crystal structure. Note the breaking of D192-R258 interaction in the binary complex of I174S mutant and concomitant shortening the distance between R258 and E295, two hallmarks of catalytically competent state formation.

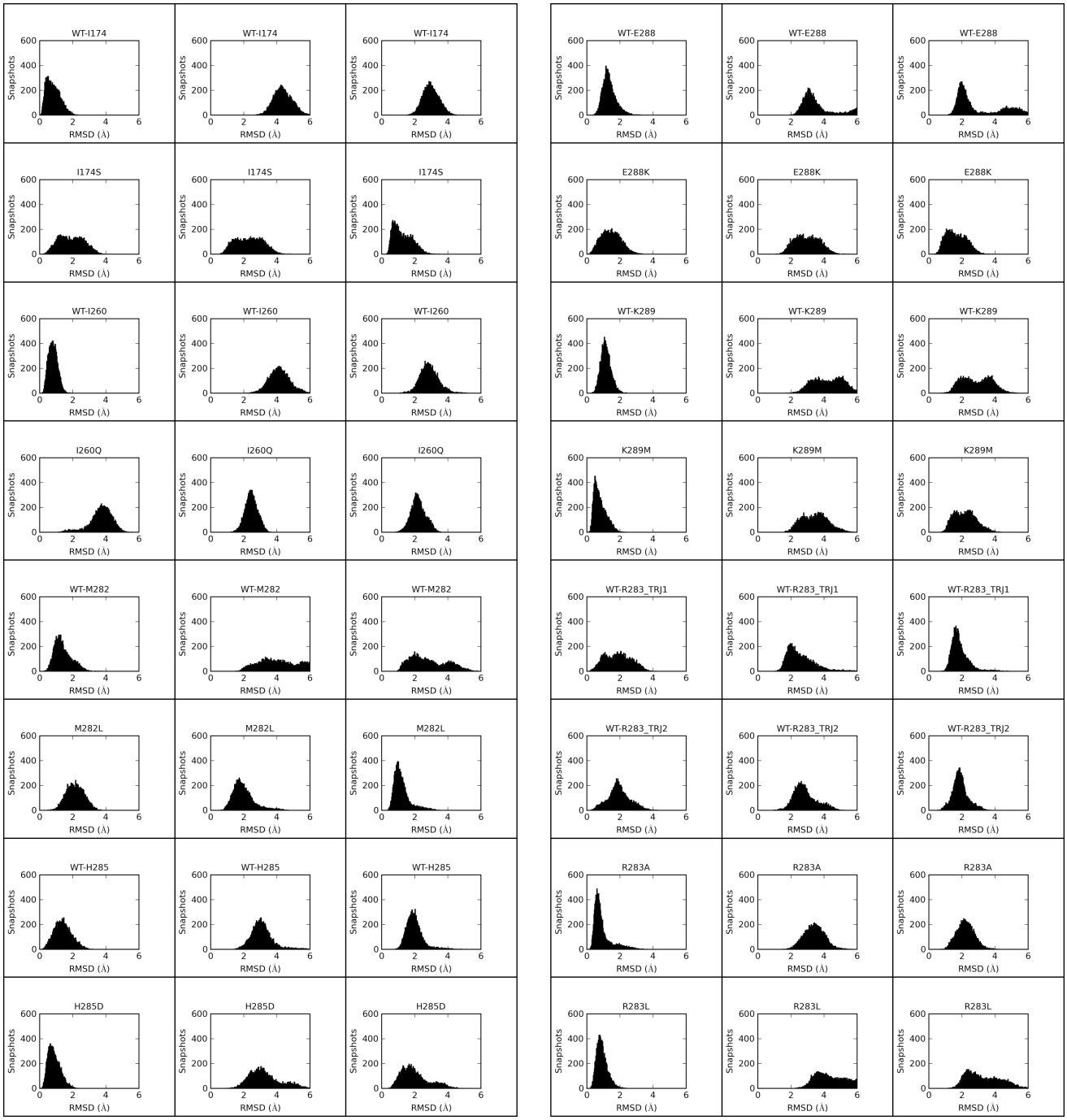

Figure S4: Histograms of the calculated RMSD of the thumb sub-domain of the binary complexes of WT Pol β and its mutants. The RMSD values were measured from the open (left column), closed (middle column) and partially open (right column) protein conformations that were defined by the coordinates of the corresponding 1BPX, 2FMP and 3C2M crystal structures. All RMSD ensembles were generated from 10 ns MD trajectories that were initiated from the open (pdb code 1BPX) X-ray crystal structure. Note that several histograms for the WT Pol β are shown. These histograms correspond to independent simulations of the WT Pol β that differ in the definition of the probe region of the protein (Figure 3).

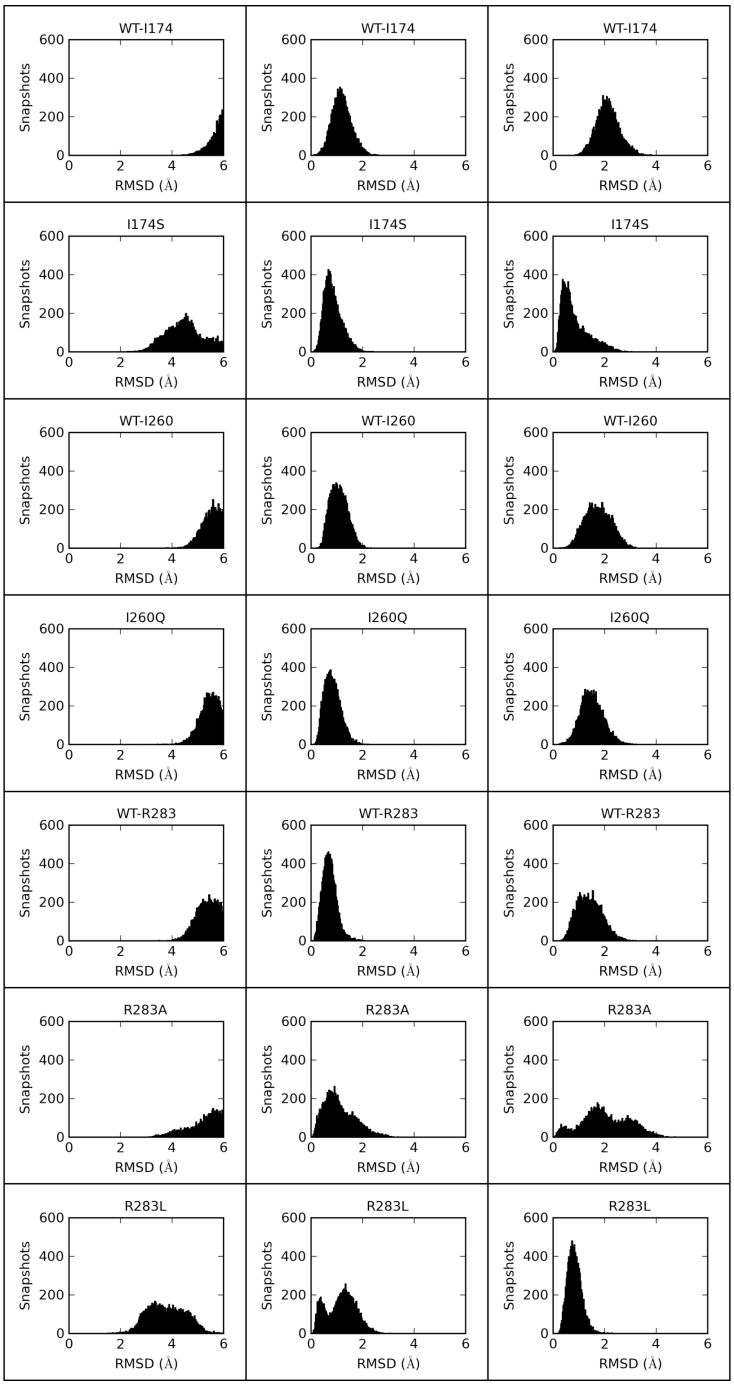

Figure S5: Histograms of the calculated RMSD of the thumb sub-domain of the binary complexes of WT Pol β and its four mutants. The RMSD values were measured from the open (left column), closed (middle column) and partially open (right column) protein conformations that were defined by the coordinates of the corresponding 1BPX, 2FMP and 3C2M crystal structures. All RMSD ensembles were generated from 10 ns MD trajectories that were initiated from the closed (pdb code 2FMP) X-ray crystal structure after the removal of the dCTP·Na^+·Mg^{2+} substrate. Note that the histograms labeled WT-I174S, WT-I260 and WT-R283 correspond to three independent simulations of the WT Pol β, that differ in the definition of the probe region of the protein (I174S, I260 or R283).

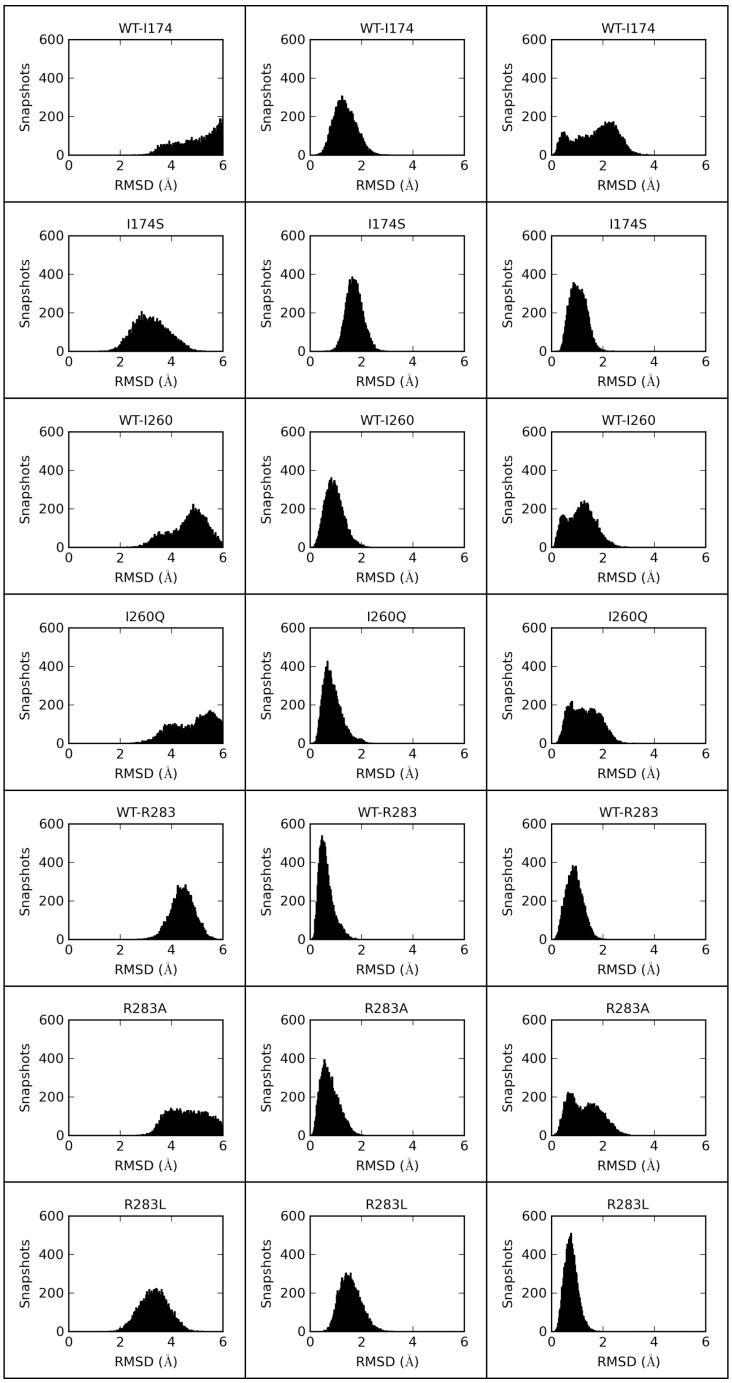

Figure S6: Histograms of the calculated RMSD of the thumb sub-domain of the binary complexes of WT Pol β and its four mutants. The RMSD values were measured from the open (left column), closed (middle column) and partially open (right column) protein conformations that were defined by the coordinates of the corresponding 1BPX, 2FMP and 3C2M crystal structures. All RMSD ensembles were generated from 10 ns MD trajectories that were initiated from the partially closed (pdb code 3C2M) X-ray crystal structure after the removal of the dGTP·2Mn^{2+} substrate. Note that the histograms labeled WT-I174, WT-I260 and WT-R283 correspond to three independent simulations of the WT Pol β that differ in the definition of the probe region of the protein (I174, I260 or R283).

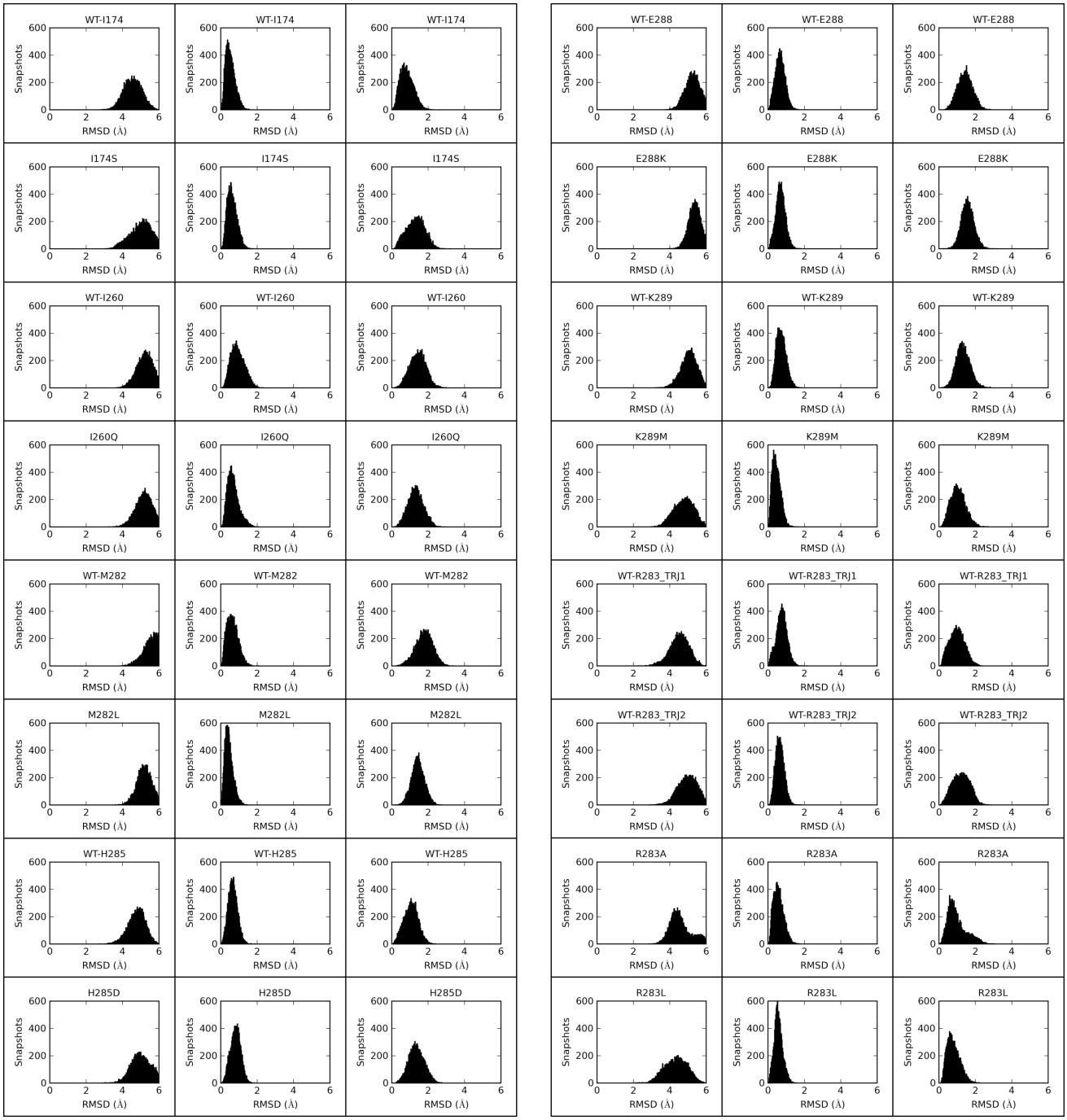

Figure S7: Histograms of the calculated RMSD of the thumb sub-domain of the $\mathcal{R}$ TS complexes of WT Pol β and its mutants. The RMSD values were measured from the open (left column), closed (middle column) and partially open (right column) protein conformations that were defined by the coordinates of the corresponding 1BPX, 2FMP and 3C2M crystal structures. All RMSD ensembles were generated from 10 ns MD trajectories that were initiated from the closed (pdb code 2FMP) X-ray crystal structure subjected to distance constraints on $PO_{3'}$ and PO_{lg} bonds. Note that several histograms for the WT Pol β are shown. These histograms correspond to independent simulations of the WT Pol β that differ in the definition of the probe region of the protein (Figure 3).

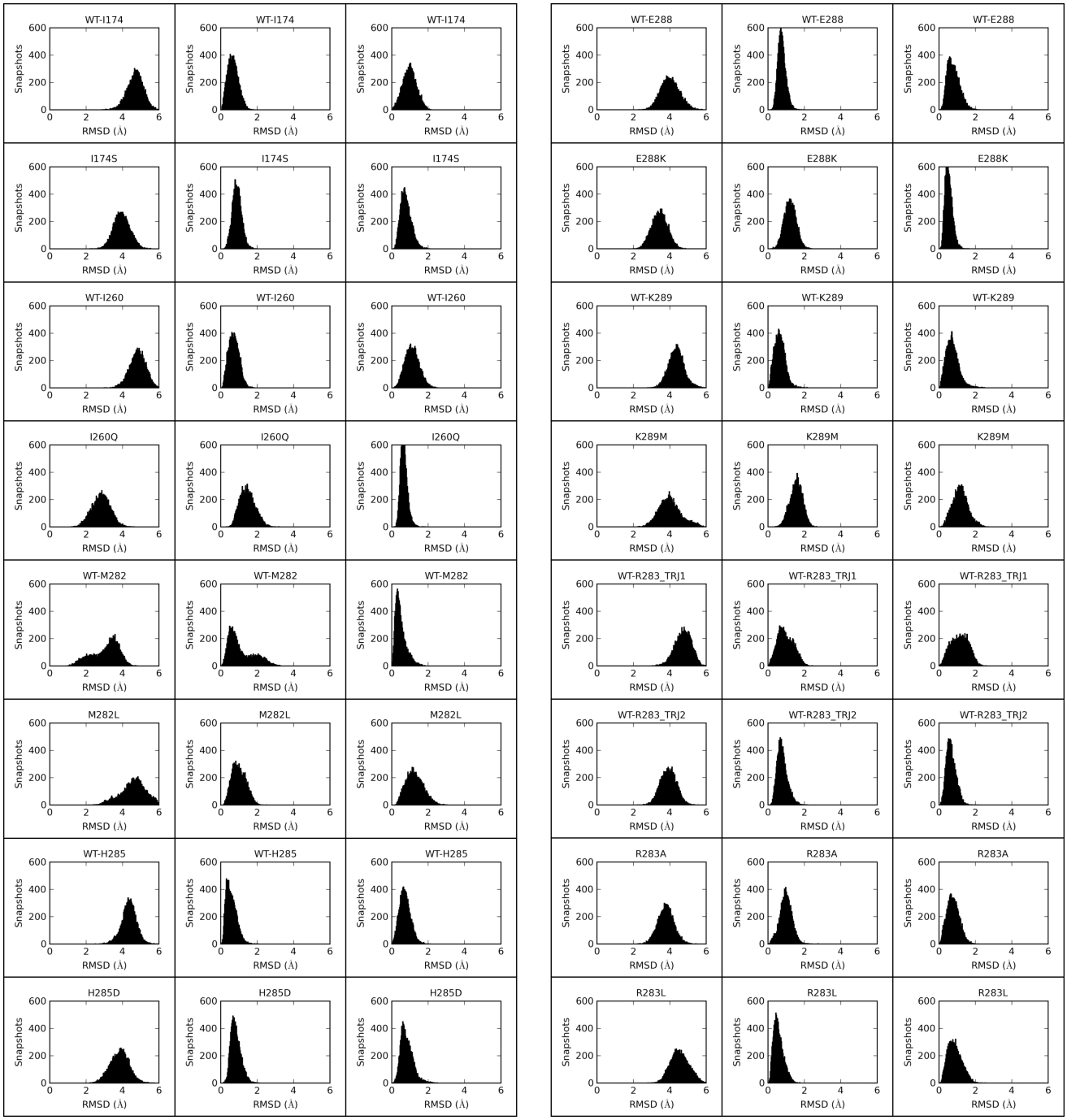

Figure S8: Histograms of the calculated RMSD of the thumb sub-domain of the $\mathcal{W}$TS complexes of WT Pol β and its mutants. The RMSD values were measured from the open (left column), closed (middle column) and partially open (right column) protein conformations that were defined by the coordinates of the corresponding 1BPX, 2FMP and 3C2M crystal structures. All RMSD ensembles were generated from 10 ns MD trajectories that were initiated from the partially closed (pdb code 3C2M) X-ray crystal structure subjected to distance constraints on the $PO_{3'}$ and PO_{lg} bonds. Note that several histograms for the WT Pol β are shown. These histograms correspond to independent simulations of the WT Pol β that differ in the definition of the probe region of the protein (Figure 3).