SUPPLEMENTARY MATERIALS
to the article A.A. Berdugin, V.M. Golyshev, A.A. Lomzov
"Structural basis of the phosphoramidate N-benzimidazole group's
influence on modified primer extension efficiency by Taq DNA polymerase"

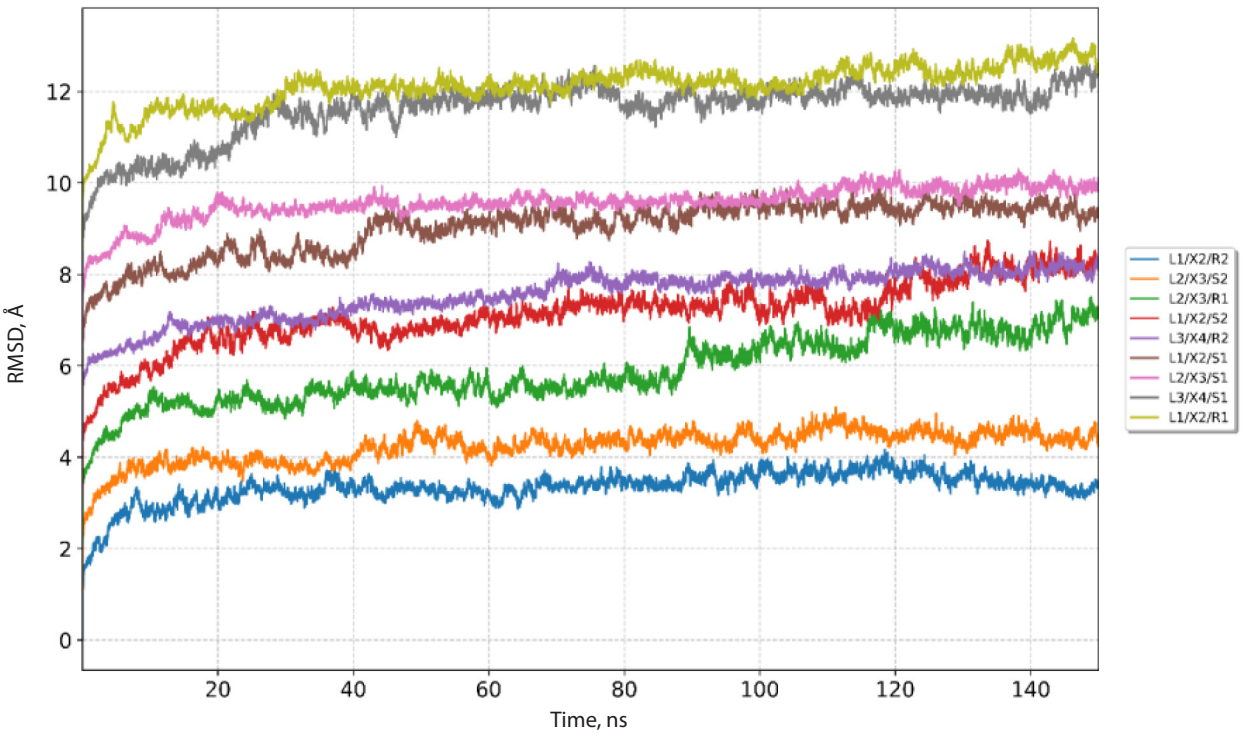

Fig. S1. RMSD as a function of time along MD trajectories of complexes that underwent an additional 50 ns of simulation.

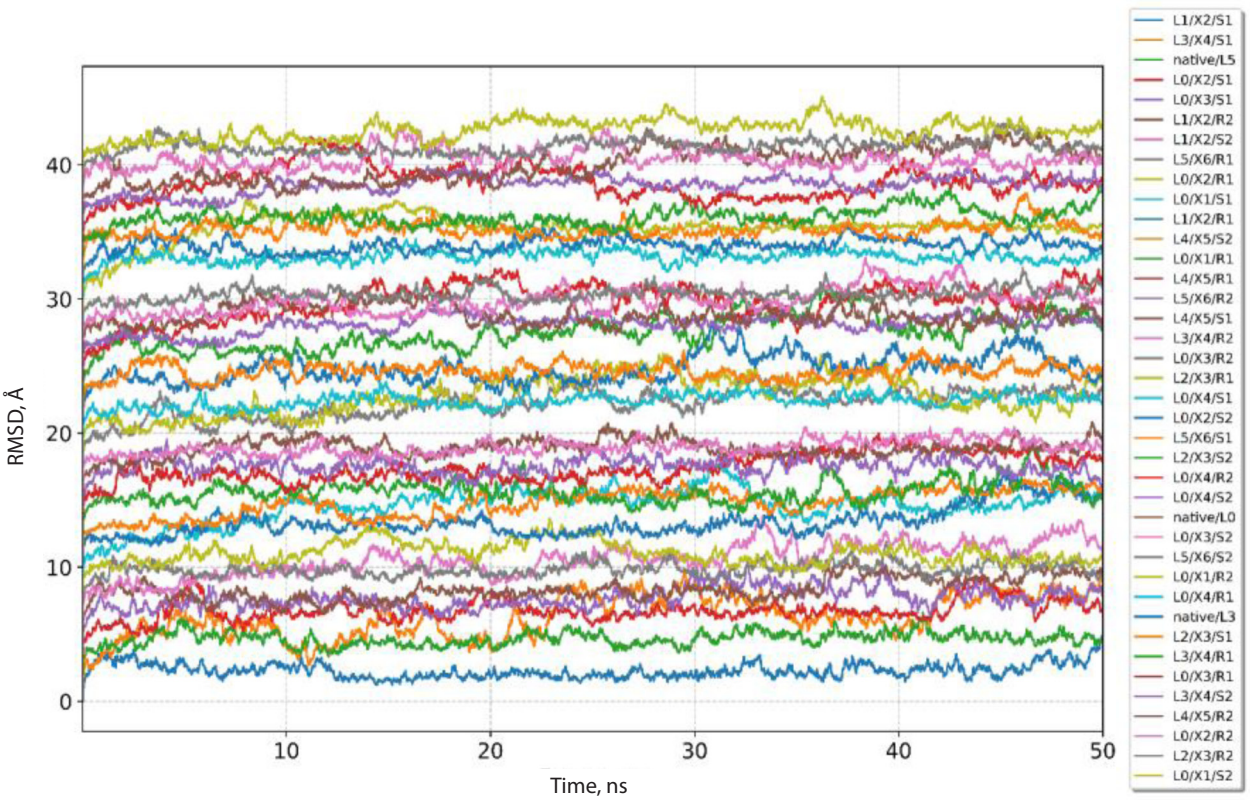

Fig. S2. 3RMSD of DNA–Taq DNA polymerase complexes (with nucleoside triphosphate and two magnesium ions) as a function of time along MD trajectories. Shown are the last 50 ns of the 100-ns simulations, used for subsequent analysis. For clarity, trajectories are vertically offset relative to one another.

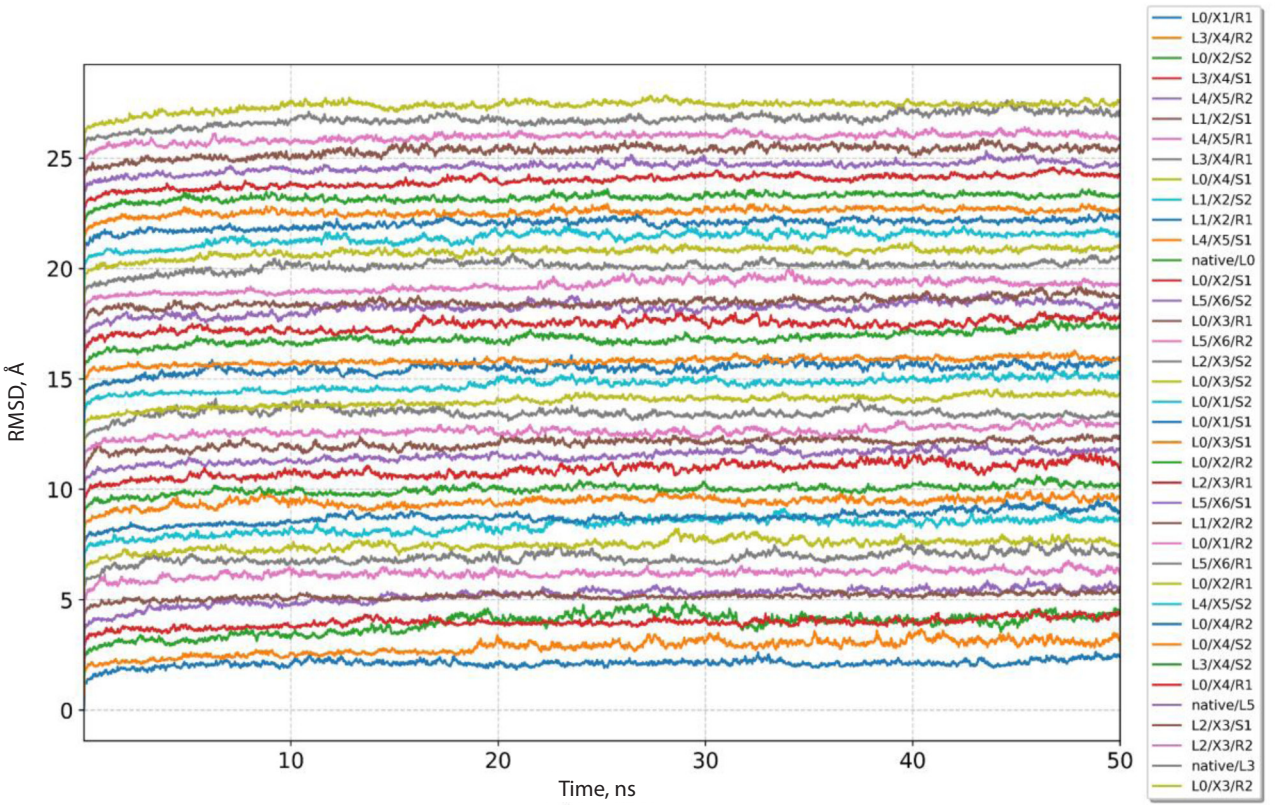

Fig. S3. RMSD of Taq DNA polymerase within the complexes as a function of time along MD trajectories. Shown are the last 50 ns of the 100-ns simulations, used for subsequent analysis. For clarity, trajectories are vertically offset relative to one another.

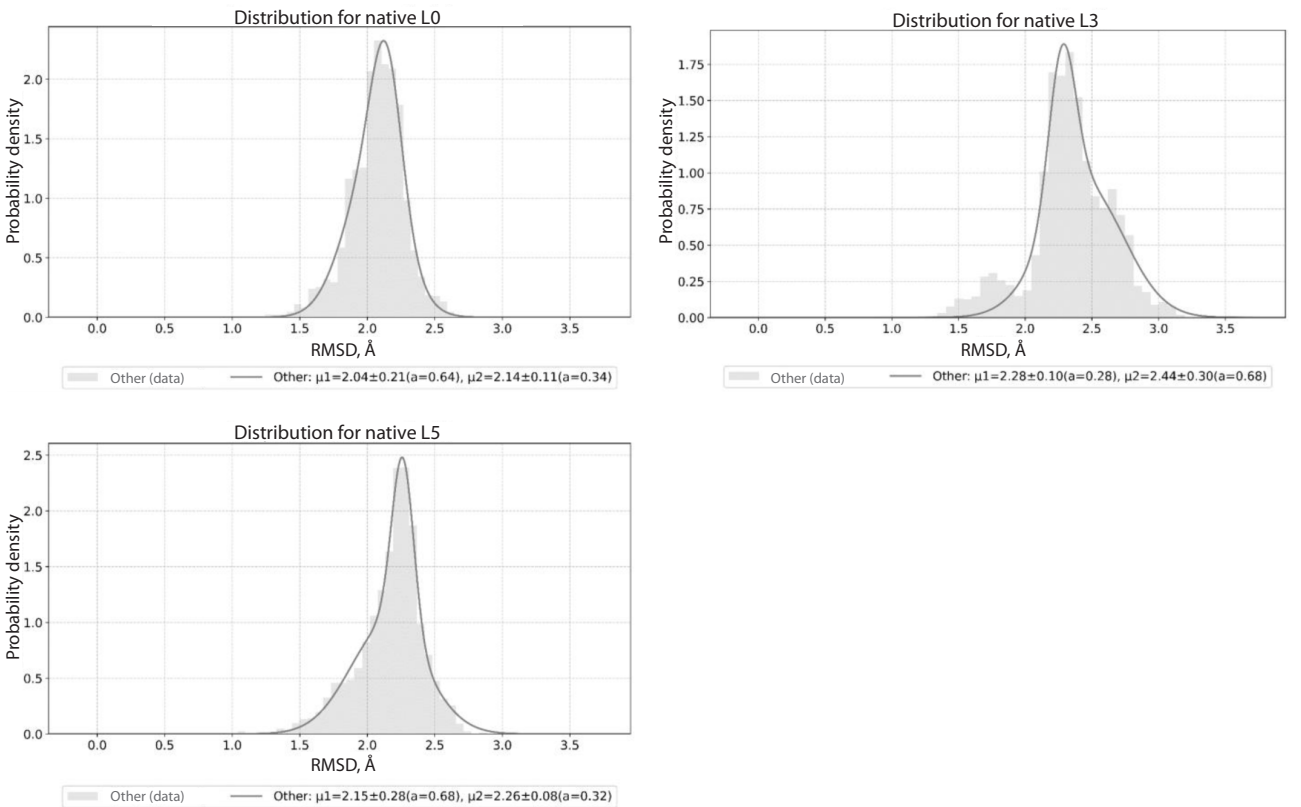

Fig. S4. Comparison of RMSD distributions for the studied complexes for Taq.

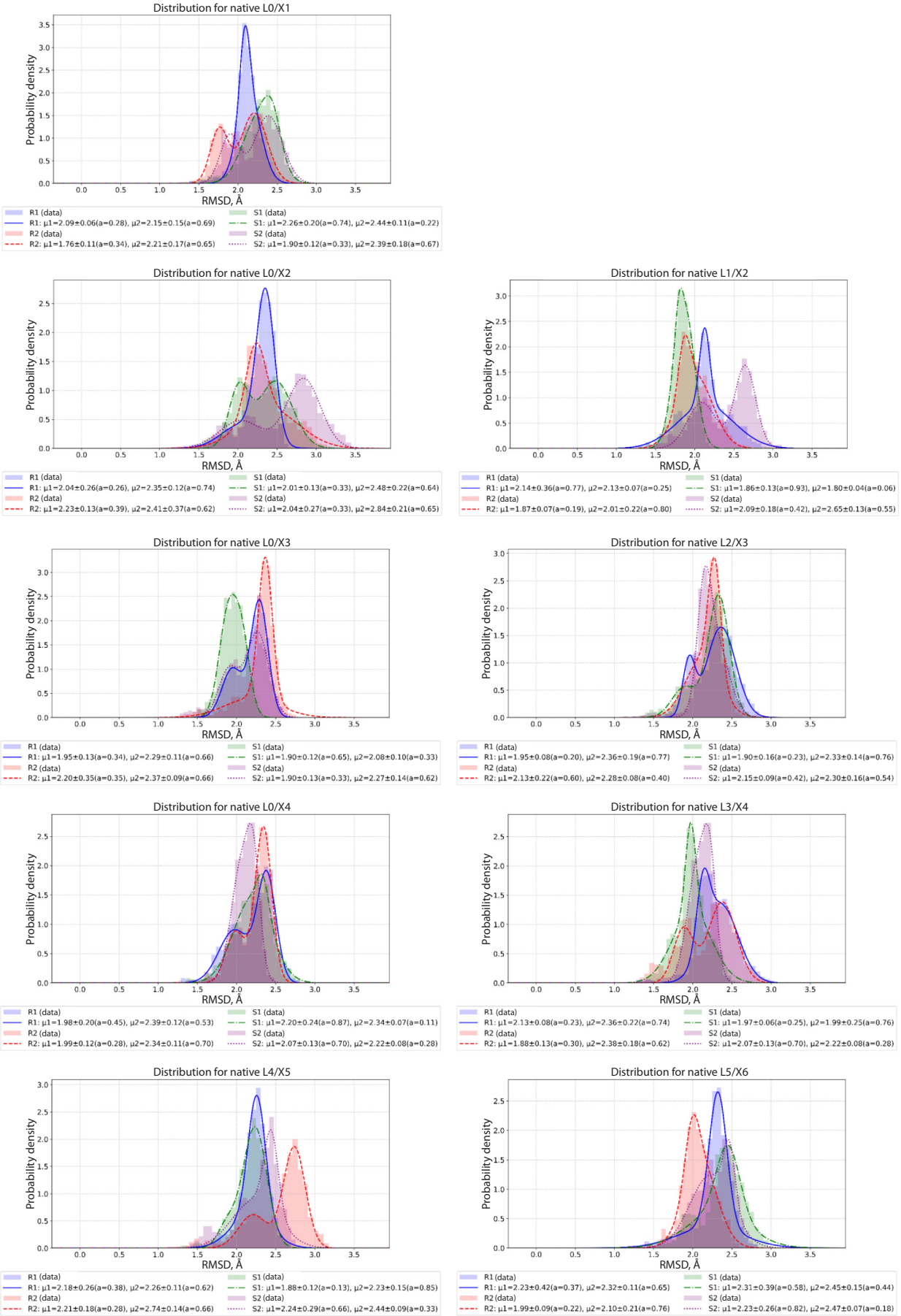
Distribution for native L0/X1
Probability density
RMSD, Å
R1 (data)
R1: μ1=2.09±0.06(a=0.28), μ2=2.15±0.15(a=0.69)
R2 (data)
R2: μ1=1.76±0.11(a=0.34), μ2=2.21±0.17(a=0.65)
S1 (data)
S1: μ1=2.26±0.20(a=0.74), μ2=2.44±0.11(a=0.22)
S2 (data)
S2: μ1=1.90±0.12(a=0.33), μ2=2.39±0.18(a=0.67)
Distribution for native L0/X2
Probability density
RMSD, Å
R1 (data)
R1: μ1=2.04±0.26(a=0.26), μ2=2.35±0.12(a=0.74)
R2 (data)
R2: μ1=2.23±0.13(a=0.39), μ2=2.41±0.37(a=0.62)
S1 (data)
S1: μ1=2.01±0.13(a=0.33), μ2=2.48±0.22(a=0.64)
S2 (data)
S2: μ1=2.04±0.27(a=0.33), μ2=2.84±0.21(a=0.65)
Distribution for native L1/X2
Probability density
RMSD, Å
R1 (data)
R1: μ1=2.14±0.36(a=0.77), μ2=2.13±0.07(a=0.25)
R2 (data)
R2: μ1=1.87±0.07(a=0.19), μ2=2.01±0.22(a=0.80)
S1 (data)
S1: μ1=1.86±0.13(a=0.93), μ2=1.80±0.04(a=0.06)
S2 (data)
S2: μ1=2.09±0.18(a=0.42), μ2=2.65±0.13(a=0.55)
Distribution for native L0/X3
Probability density
RMSD, Å
R1 (data)
R1: μ1=1.95±0.13(a=0.34), μ2=2.29±0.11(a=0.66)
R2 (data)
R2: μ1=2.20±0.35(a=0.35), μ2=2.37±0.09(a=0.66)
S1 (data)
S1: μ1=1.90±0.12(a=0.65), μ2=2.08±0.10(a=0.33)
S2 (data)
S2: μ1=1.90±0.13(a=0.33), μ2=2.27±0.14(a=0.62)
Distribution for native L2/X3
Probability density
RMSD, Å
R1 (data)
R1: μ1=1.95±0.08(a=0.20), μ2=2.36±0.19(a=0.77)
R2 (data)
R2: μ1=2.13±0.22(a=0.60), μ2=2.28±0.08(a=0.40)
S1 (data)
S1: μ1=1.90±0.16(a=0.23), μ2=2.33±0.14(a=0.76)
S2 (data)
S2: μ1=2.15±0.09(a=0.42), μ2=2.30±0.16(a=0.54)
Distribution for native L0/X4
Probability density
RMSD, Å
R1 (data)
R1: μ1=1.98±0.20(a=0.45), μ2=2.39±0.12(a=0.53)
R2 (data)
R2: μ1=1.99±0.12(a=0.28), μ2=2.34±0.11(a=0.70)
S1 (data)
S1: μ1=2.20±0.24(a=0.87), μ2=2.34±0.07(a=0.11)
S2 (data)
S2: μ1=2.07±0.13(a=0.70), μ2=2.22±0.08(a=0.28)
Distribution for native L3/X4
Probability density
RMSD, Å
R1 (data)
R1: μ1=2.13±0.08(a=0.23), μ2=2.36±0.22(a=0.74)
R2 (data)
R2: μ1=1.88±0.13(a=0.30), μ2=2.38±0.18(a=0.62)
S1 (data)
S1: μ1=1.97±0.06(a=0.25), μ2=1.99±0.25(a=0.76)
S2 (data)
S2: μ1=2.07±0.13(a=0.70), μ2=2.22±0.08(a=0.28)
Distribution for native L4/X5
Probability density
RMSD, Å
R1 (data)
R1: μ1=2.18±0.26(a=0.38), μ2=2.26±0.11(a=0.62)
R2 (data)
R2: μ1=2.21±0.18(a=0.28), μ2=2.74±0.14(a=0.66)
S1 (data)
S1: μ1=1.88±0.12(a=0.13), μ2=2.23±0.15(a=0.85)
S2 (data)
S2: μ1=2.24±0.29(a=0.66), μ2=2.44±0.09(a=0.33)
Distribution for native L5/X6
Probability density
RMSD, Å
R1 (data)
R1: μ1=2.23±0.42(a=0.37), μ2=2.32±0.11(a=0.65)
R2 (data)
R2: μ1=1.99±0.09(a=0.22), μ2=2.10±0.21(a=0.76)
S1 (data)
S1: μ1=2.31±0.39(a=0.58), μ2=2.45±0.15(a=0.44)
S2 (data)
S2: μ1=2.23±0.26(a=0.82), μ2=2.47±0.07(a=0.18)

Fig. S4 (end)

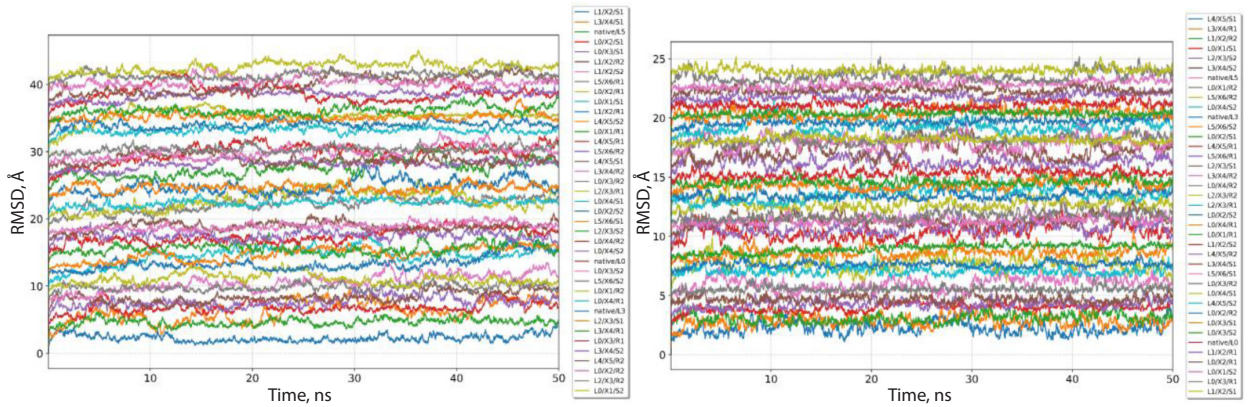

Fig. S5. RMSD of DNA in complex with Taq polymerase as a function of time along MD trajectories, including the single-stranded template overhang (left) and excluding it (right). Shown are the last 50 ns of the 100-ns simulations, used for subsequent analysis. For clarity, trajectories are vertically offset relative to one another.

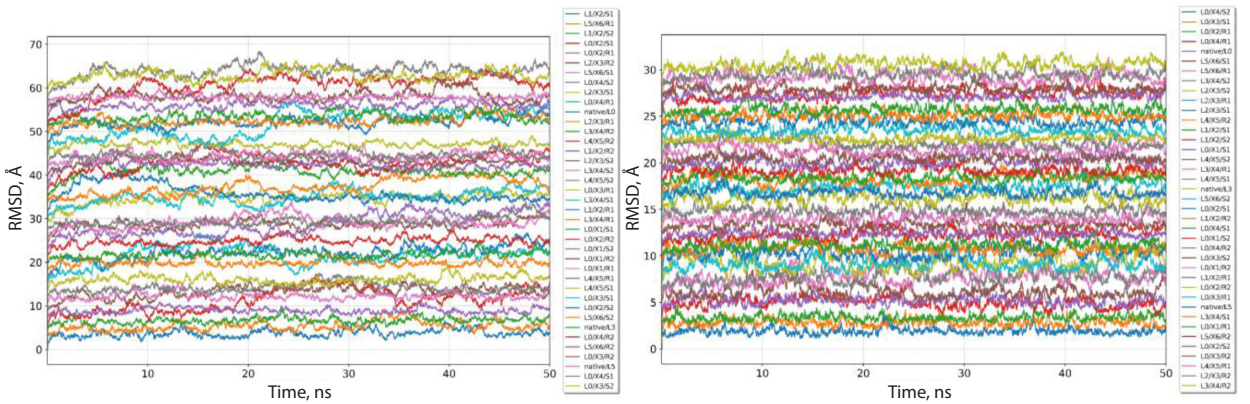

Fig. S6. RMSD of DNA with the single-stranded template overhang (left) and without it (right) as a function of time along MD trajectories. Shown are the last 50 ns of the 100-ns simulations, used for subsequent analysis. For clarity, trajectories are vertically offset relative to one another.

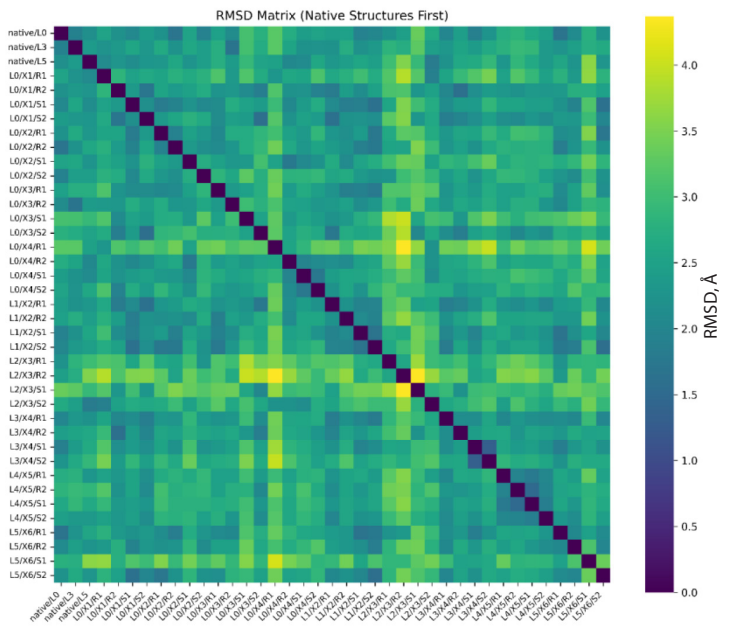

Fig. S7. Heatmap of RMSD values for the most representative protein structures from MD trajectory clusters.

Table S1. Average RMSD values between Cα atoms of the protein structures and their standard deviations for the various complexes. RMSD values were calculated for each structure relative to all other structures in this study

Complex	Average RMSD, Å	SD of RMSD, Å	Complex	Average RMSD, Å	SD of RMSD, Å
native/L0	2.38	0.39			
native/L3	2.52	0.33			
native/L5	2.50	0.37			
L0/X1/R1	2.78	0.41			
L0/X1/R2	2.36	0.36			
L0/X1/S1	2.36	0.42			
L0/X1/S2	2.36	0.41			
L0/X2/R1	2.53	0.42	L1/X2/R1	2.33	0.40
L0/X2/R2	2.46	0.41	L1/X2/R2	2.53	0.43
L0/X2/S1	2.66	0.38	L1/X2/S1	2.36	0.43
L0/X2/S2	2.61	0.35	L1/X2/S2	2.38	0.50
L0/X3/R1	2.54	0.43	L2/X3/R1	2.97	0.38
L0/X3/R2	2.53	0.32	L2/X3/R2	3.26	0.51
L0/X3/S1	2.88	0.45	L2/X3/S1	3.07	0.45
L0/X3/S2	2.57	0.45	L2/X3/S2	2.74	0.39
L0/X4/R1	3.24	0.44	L3/X4/R1	2.37	0.39
L0/X4/R2	2.55	0.40	L3/X4/R2	2.50	0.35
L0/X4/S1	2.59	0.35	L3/X4/S1	2.46	0.43
L0/X4/S2	2.61	0.35	L3/X4/S2	2.76	0.43
			L4/X5/R1	2.59	0.37
			L4/X5/R2	2.64	0.43
			L4/X5/S1	2.62	0.40
			L4/X5/S2	2.49	0.35
			L5/X6/R1	2.38	0.41
			L5/X6/R2	2.53	0.42
			L5/X6/S1	3.07	0.37
			L5/X6/S2	2.48	0.46

Table S2. Comparison of RMSD values for the duplex region of nucleic acid substrates

	R1[1]	R2[1]	S1[1]	S2[1]	Without Taq	Comparison with free DNA[3]	Comparison with DNA from MD sim. with Taq[4]
L0/X1/R1		1.66	2.32	1.47	3.00	1.80	1.91
L0/X1/R2	1.66		1.82	1.26	2.46	1.49	1.58
L0/X1/S1	2.32	1.47		1.78	2.53	1.78	1.92
L0/X1/S2	1.47	1.26	1.78		2.67	1.67	1.88
L0/X2/R1		1.82	2.17	1.85	2.34	1.61	1.71
L0/X2/R2	1.82		1.80	1.78	2.35	1.70	1.81
L0/X2/S1	2.17	1.80		1.85	2.67	1.45	1.69
L0/X2/S2	1.85	1.78	1.85		2.25	1.62	1.50
L1/X2/R1		2.44	3.88	1.94	2.12		
L1/X2/R2	2.44		3.33	2.48	1.75		
L1/X2/S1	3.88	3.33		3.32	2.39		
L1/X2/S2	1.94	2.48	3.32		2.41		
L0/X3/R1		1.86	2.35	1.50	2.66	1.74	1.88
L0/X3/R2	1.86		2.60	1.66	2.65	1.80	1.90
L0/X3/S1	2.35	2.60		2.26	2.94	1.73	1.53
L0/X3/S2	1.50	1.66	2.26		1.90	1.47	1.72
L2/X3/R1		2.36	2.13	2.52	2.40		
L2/X3/R2	2.36		2.22	2.12	1.68		
L2/X3/S1	2.13	2.22		1.74	2.57		
L2/X3/S2	2.52	2.12	1.74		2.78		
L0/X4/R1	0	1.63	1.53		2.72	1.60	1.71
L0/X4/R2	1.63	0	1.66		1.96	1.96	1.92
L0/X4/S1	1.53	1.66	0		3.33	1.81	1.86
L3/X4/R1		1.50	1.63		2.44	2.07	1.80
L3/X4/R2	1.50		1.64		2.30	2.21	1.76
L3/X4/S1	1.63	1.64			1.85	2.10	1.60
L4/X5/R1		1.77	2.45	2.30	2.61		
L4/X5/R2	1.77		2.56	2.25	2.40		
L4/X5/S1	2.45	2.56		1.53	1.89		
L4/X5/S2	2.30	2.25	1.53		2.10		
L5/X6/R1		1.99	1.99	1.93	2.66	2.49	2.20
L5/X6/R2	1.99		1.86	1.68	2.02	2.19	1.55
L5/X6/S1	1.99	1.86		1.87	1.87	2.57	1.81
L5/X6/S2	1.93	1.68	1.87		1.83	2.30	1.63

[1] Comparison of structures obtained after MD simulations of DNA in complex with the protein for different rotamers and stereoisomers. [2] Comparison of structures with the same initial modification position for a single stereoisomer, simulated with and without Taq polymerase. [3, 4] Comparison of structures after MD simulations of modified and unmodified DNA in complex with the protein.

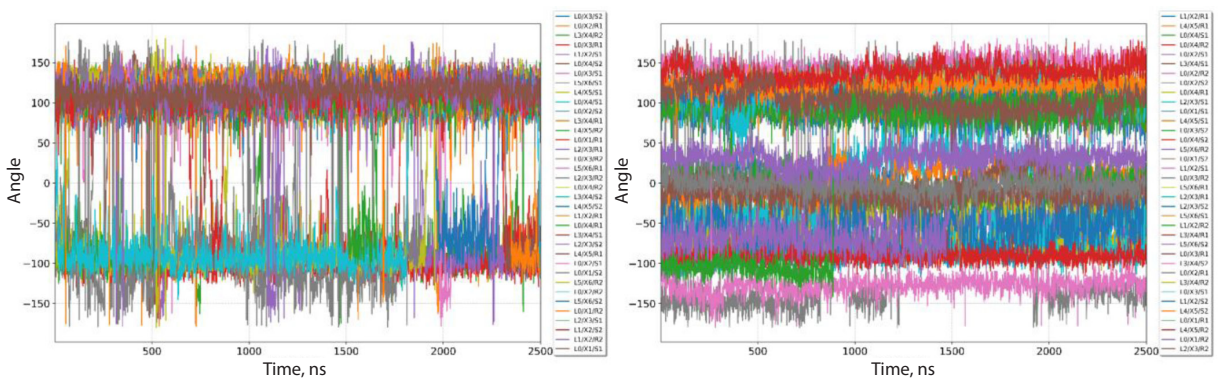

Fig. S8. Dihedral angle of the P–N bond in the phosphoramidate moiety along MD trajectories for free DNA (left) and DNA in complex with the protein (right).

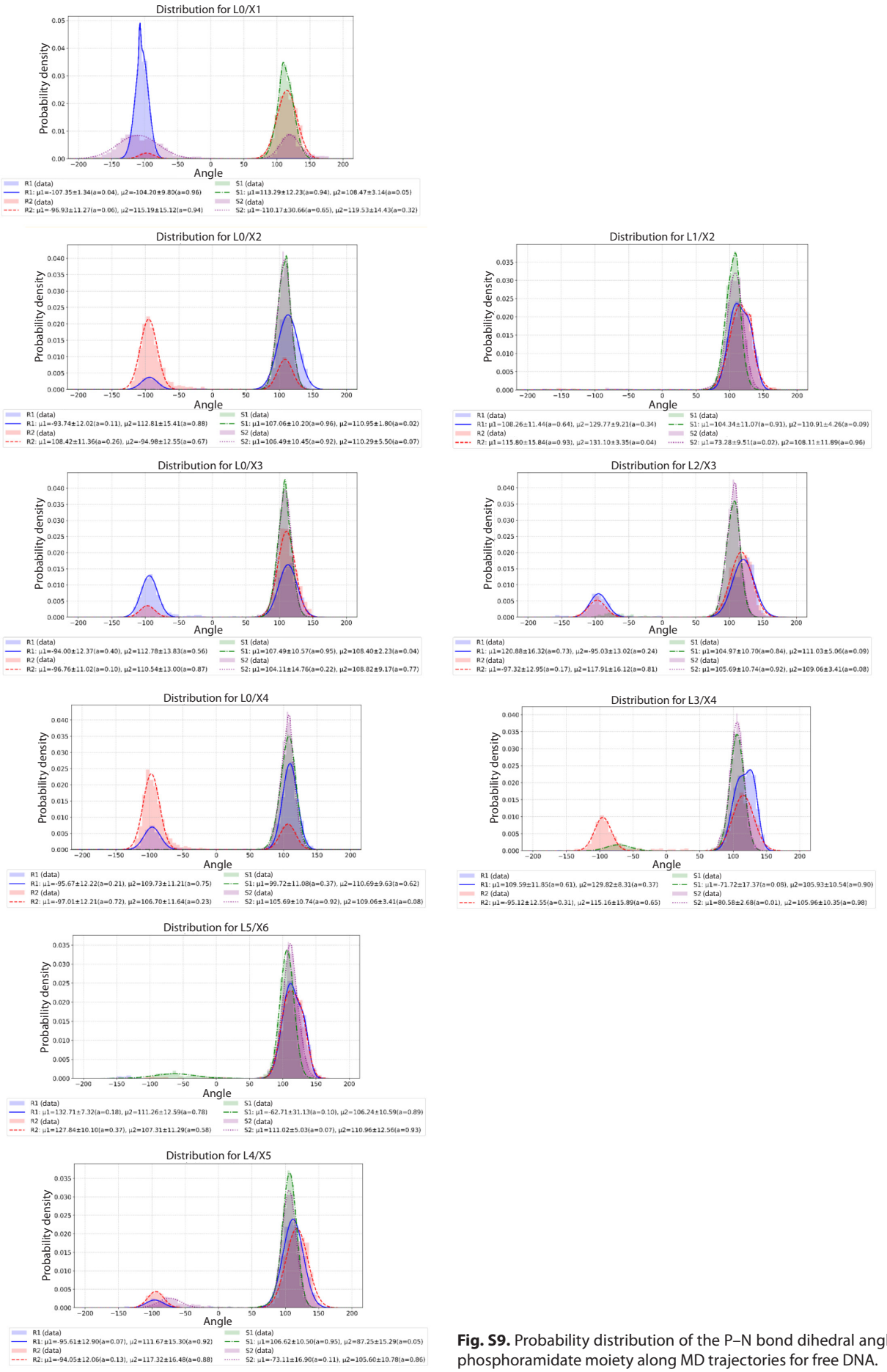

Fig. S9. Probability distribution of the P–N bond dihedral angle in the phosphoramidate moiety along MD trajectories for free DNA.

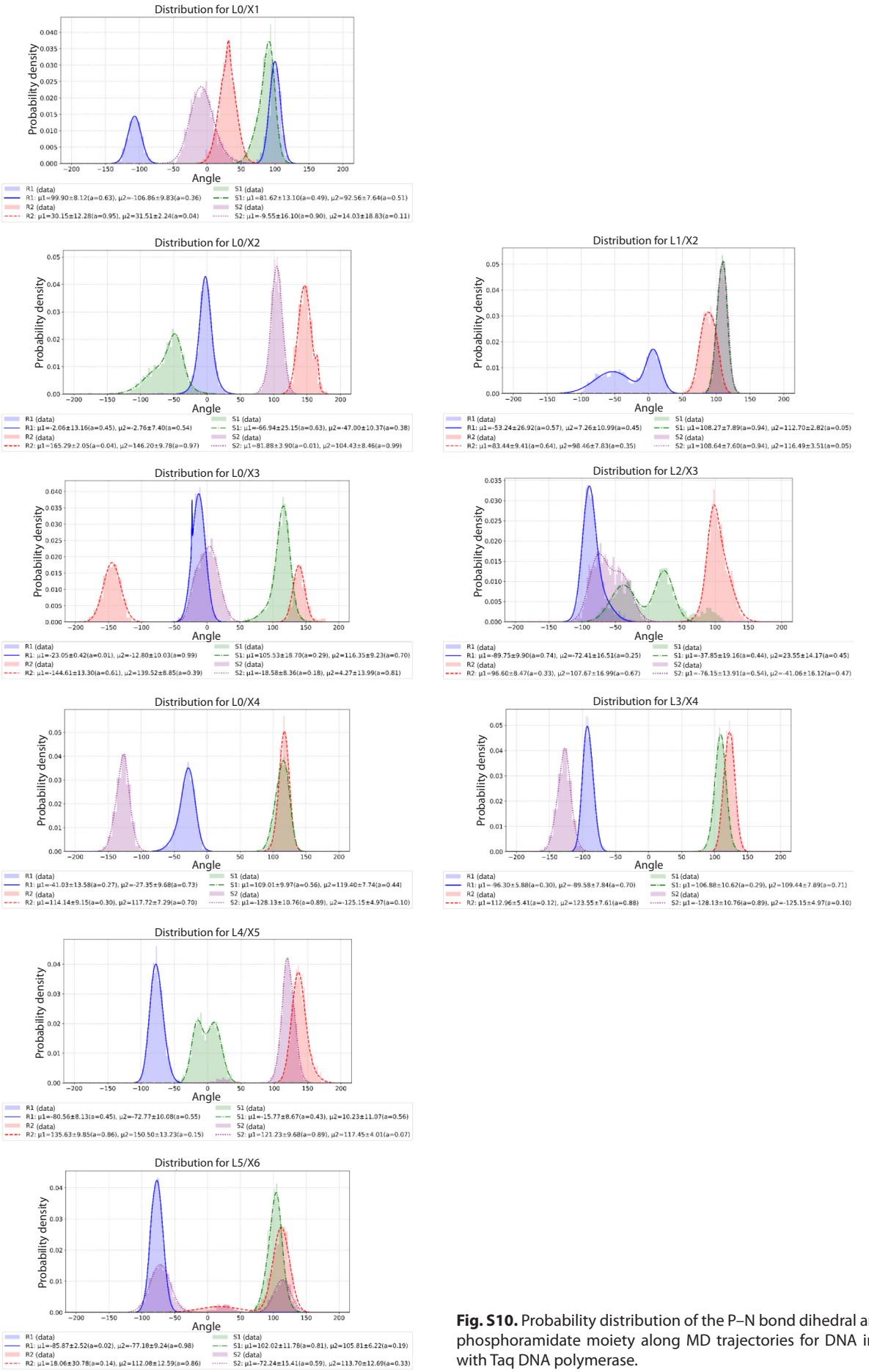

Fig. S10. Probability distribution of the P–N bond dihedral angle in the phosphoramidate moiety along MD trajectories for DNA in complex with Taq DNA polymerase.

Table S3. Positioning of the N-benzimidazole modification in the DNA–Taq polymerase complex and its impact on structure

	Location in protein pocket	Number of atoms		Impact
		within 3 Å	within 3.5 Å	
L0/X1/R1	+ +	26	82	Non-planarity of base pairs 2 and 3
L0/X1/R2	+	18	59	Non-planarity of base pairs 1 and 2
L0/X1/S1	–	12	37	Modification interacts with the incoming dNTP
L0/X1/S2	– +	18	47	Non-planarity of base pairs 1 and 3
L0/X2/R1	– +	5	21	Non-planarity of base pair 3 and disruption of Watson–Crick pairing in base pair 2
L0/X2/R2	+	27	65	Base pair 3 lacks Watson–Crick pairing
L0/X2/S1	– +	13	28	Modification shields bases toward the 3′-end of the primer
L0/X2/S2	– +	14	31	Modification shields bases toward the 3′-end of the primer
L1/X2/R1	+	18	62	
L1/X2/R2	+ +	25	68	
L1/X2/S1	+ –	19	38	
L1/X2/S2	–	11	31	Base pairs 3 and 4 lack Watson–Crick pairing
L0/X3/R1	+ –	22	61	Base pair 5 lacks Watson–Crick pairing
L0/X3/R2	+ +	25	76	Base pair 5 lacks Watson–Crick pairing
L0/X3/S1	– +	5	25	
L0/X3/S2	– +	6	30	
L2/X3/R1	+ +	28	68	
L2/X3/R2	+	14	37	
L2/X3/S1	–	19	35	
L2/X3/S2	– +	8	26	Non-planarity of base pairs 3
L0/X4/R1	+ + +	36	100	
L0/X4/R2	+ + +	28	86	
L0/X4/S1	– +	15	41	Modification shields bases toward the 3′-end of the primer
L3/X4/R1	+ +	28	72	
L3/X4/R2	+ + +	37	89	Non-planarity of base pairs 5
L3/X4/S1	+	22	49	
L4/X5/R1	+	20	58	
L4/X5/R2	+ +	28	81	
L4/X5/S1	+ –	15	51	
L4/X5/S2	+ –	17	49	
L5/X6/R1	+	25	61	
L5/X6/R2	–	4	15	
L5/X6/S1	– –	0	0	Modification shields bases toward the 3′-end of the primer
L5/X6/S2	–	9	28	

Table S4. Amino acid residues located within 3 Å of the phosphoramidate *N*-benzimidazole moiety

Complex	Residues
L0/X1/R1	Leu581, Gln582, Arg595, Val783, Glu786, Trp827
L0/X1/R2	Val586, Arg587, Leu657, Arg660
L0/X1/S1	Val586, Lyn663
L0/X1/S2	Val586, Arg587, Leu657
L0/X2/R1	Asn583, Val586, Arg587
L0/X2/R2	Lys540, Leu541, Tyr545, Pro585, Arg587
L0/X2/S1	Pro585, Val586, Arg587
L0/X2/S2	Pro585, Val586, Arg587
L1/X2/R1	Arg587, Leu657, Arg660
L1/X2/R2	Glu537, Leu541, Pro585, Val586, Thr588, Leu590, Gly591
L1/X2/S1	Pro585, Val586, Arg587
L1/X2/S2	Pro585, Val586, Arg587
L0/X3/R1	Glu537, Leu538, Thr588, Leu590
L0/X3/R2	Glu537, Lys540, Leu541, Pro585, Thr588, Leu590
L0/X3/S1	Lyn540, Arg587
L0/X3/S2	Ala516, Glu537, Lyn540, Arg587^{+}
L2/X3/R1	Glu537, Lyn540, Leu541, Thr588
L2/X3/R2	Ala516, Leu533, Arg536, Glu537
L2/X3/S1	Ala516, Ala517, Lyn540, Arg587
L2/X3/S2	Arg587
L0/X4/R1	Ser515, Ala516, Leu519, Glu520, Arg523, Val529, Ile532, Leu533, Arg536
L0/X4/R2	Leu490, Thr514, Ser515, Ala516, Leu519, Ile532, Leu533, Arg536
L0/X4/S1	Ser515, Ala516, Arg536
L3/X4/R1	Ser515, Ala516, Leu533, Arg536, Glu537
L3/X4/R2	Leu490, Thr514, Ser515, Ala516, Leu519, Ile532, Leu533
L3/X4/S1	Ser515, Ala516, Ala517, Arg536
L4/X5/R1	Arg487, Leu494, Thr506, Ser513, Thr514
L4/X5/R2	Asp488, Glu491, Thr506, Lys511, Arg512, Ser513, Thr514, Arg536
L4/X5/S1	Arg487, Lys505, Thr506, Ser513, Thr514, Ser515
L4/X5/S2	Arg487, Lys505, Thr506, Ser513, Thr514, Ser515
L5/X6/R1	Thr506, Glu507, Lys508, Thr509, Lys511
L5/X6/R2	Thr506, Lys508, Thr509
L5/X6/S1	–
L5/X6/S2	Glu507, Lys508

Table S5. Energies of the entire complex, DNA duplex, and protein calculated using the MM/GBSA method with a single-trajectory approach (in kcal/mol)

	Complex	DNA	Protein
L1/X2/R1	–26388	–4885.1	–21299
L1/X2/R2	–26363	–4880.3	–21309
L1/X2/S1	–26367	–4897.1	–21303
L1/X2/S2	–26473	–4879.8	–21388
L2/X3/R1	–26149	–4757.6	–21245
L2/X3/R2	–26117	–4753.1	–21208
L2/X3/S1	–26218	–4756.9	–21244
L2/X3/S2	–26229	–4759.8	–21268
L3/X4/R1	–26337	–4816.5	–21340
L3/X4/R2	–26351	–4808.4	–21360
L3/X4/S1	–26356	–4808.3	–21357
L4/X5/R1	–26356	–4887.8	–21322
L4/X5/R2	–26328	–4885.2	–21285
L4/X5/S1	–26355	–4884.9	–21298
L4/X5/S2	–26334	–4881.3	–21275
L5/X6/R1	–26360	–4931.1	–21249
L5/X6/R2	–26381	–4929.7	–21284
L5/X6/S1	–26362	–4929.0	–21254
L5/X6/S2	–26394	–4929.6	–21293
Native	–26711	–4800.8	–21616
L0/X1/R1	–26510	–4955.9	–21363
L0/X1/R2	–26544	–4910.9	–21432
L0/X1/S1	–26516	–4895.9	–21414
L0/X1/S2	–26517	–4916.2	–21393
L0/X2/R1	–26577	–4939.7	–21432
L0/X2/R2	–26417	–4902.4	–21314
L0/X2/S1	–26486	–4920.2	–21358
L0/X2/S2	–26435	–4919.2	–21347
L0/X3/R1	–26514	–4896.2	–21404
L0/X3/R2	–26459	–4909.5	–21350
L0/X3/S1	–26495	–4922.5	–21386
L0/X3/S2	–26512	–4915.9	–21398
L0/X4/R1	–26529	–4903.4	–21399
L0/X4/R2	–26549	–4918.8	–21413
L0/X4/S1	–26461	–4917.4	–21327

Table S6. Binding energies of the nucleic acid substrate to Taq DNA polymerase, calculated using the MM/GBSA method with a single-trajectory approach (in kcal/mol)

	ΔE (MMGBSA)	ΔΔE, between conformers	ΔE_min between one type of stereoisomers	ΔΔE, between stereoisomers
native/L0	–181.0			
Native/L3	–221.0			
Native/L5	–249.7			
L0/X1/R1	–191.2	10.2	–201.4	6.9
L0/X1/R2	–201.4			
L0/X1/S1	–206.3	2.0	–208.3	
L0/X1/S2	–208.3			
L0/X2/R1	–205.7	–4.6	–205.7	2.5
L0/X2/R2	–201.2			
L0/X2/S1	–208.2	–39.7	–208.2	
L0/X2/S2	–168.6			
L1/X2/R1	–173.5	–30.9	–204.3	0.4
L1/X2/R2	–204.3			
L1/X2/S1	–204.7	–38.3	–204.7	
L1/X2/S2	–166.5			
L0/X3/R1	–213.7	–13.7	–213.7	–15.6
L0/X3/R2	–200.0			
L0/X3/S1	–187.0	11.2	–198.1	
L0/X3/S2	–198.1			
L2/X3/R1	–147.1	9.2	–156.2	59.4
L2/X3/R2	–156.2			
L2/X3/S1	–216.0	–14.5	–216.0	
L2/X3/S2	–201.5			
L0/X4/R1	–226.7	–8.8	–226.7	–9.9
L0/X4/R2	–217.9			
L0/X4/S1	–216.8		–216.8	
L3/X4/R1	–181.2	1.5	–182.7	8.7
L3/X4/R2	–182.7			
L3/X4/S1	–191.4		–191.4	
L4/X5/R1	–145.6	12.2	–157.8	20.3
L4/X5/R2	–157.8			
L4/X5/S1	–172.4	5.8	–178.1	
L4/X5/S2	–178.1			
L5/X6/R1	–180.9	–13.4	–180.9	–1.7
L5/X6/R2	–167.6			
L5/X6/S1	–179.2	–7.7	–179.2	
L5/X6/S2	–171.5			