

Supporting information:

Evolving a polymerase for hydrophobic base analogues

D. Loakes, J. Gallego, V.B. Pinheiro, E.T. Kool and P. Holliger

1. Figure legends

Supporting information Figure 1 (see pages S2-S3):

Alignment of polymerase sequences: Protein sequences of selected polymerases from round 4 (4C11) and round 5 (5B1, 5B4, 5D3, 5D4) are aligned with parent polymerases Taq, Tth and Tfl. Numbering here refers to the Tth gene, which contains two insertions (59G, 228E) compared to the Taq and Tfl genes, i.e. in Taq (and Tfl), which do not share the Tth insertions G59 and E228, actual residue numbers (N) will be N=n-1 (for n>58), respectively N=n-2 (for n>228). Identities are outlined and background is shaded dark grey. Substitutions are outlined with a white background (for non-conservative substitutions (e.g. R -> E)) or light grey background (for conservative substitutions (e.g. R -> K)).

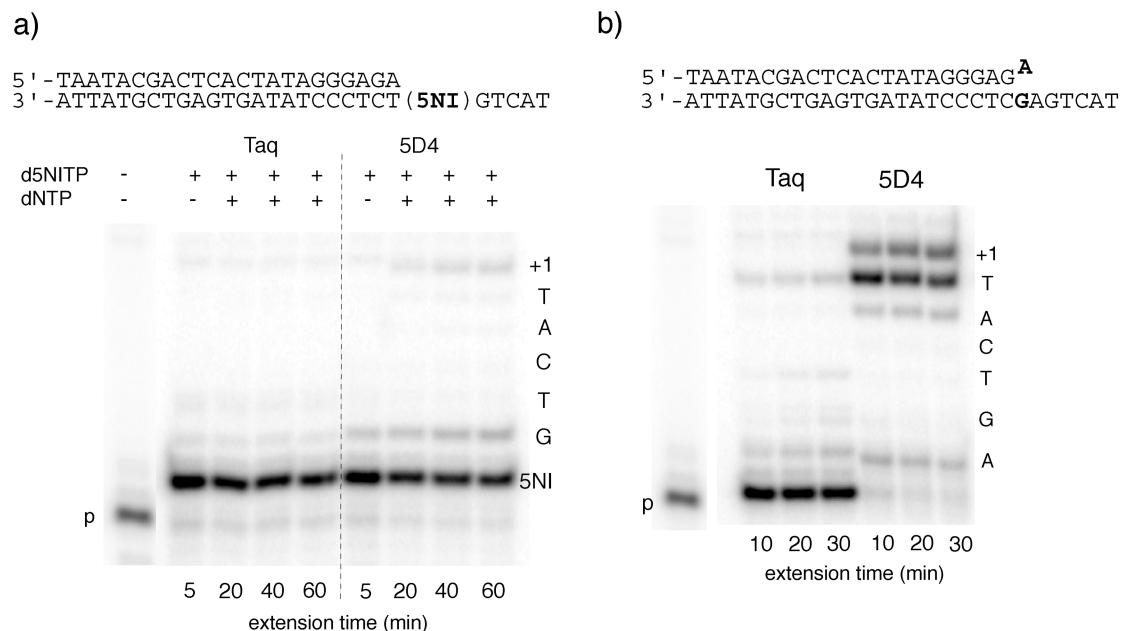
A number of mutations that are not present in the parental genes but are shared by several selected polymerases are shown in Supporting information Table 1

Alignment of polymerase sequences

	10	20	30	40	50	60	70	80	90	100		
Translation of 4C11	M	GMLPLFEPKGRVLLVDGHHLAYRTFFALKGLTTSRGEPVQAVYGFAKSLLKALKEDGYKAVFVVFDAKAPSFRHEAYEAYKAGRAPTPEDFPRQLALI										
Translation of 5B1	-MAMLP	LPLFEPKGRVLLVDGHHLAYRTFFALKGLTTSRGEPVQAVYGFAKSLLKALKEDGYKAVFVVFDAKAPSFRHEAYEAYKAGRAPTPEDFPRQLALI										
Translation of 5B4	-MAMLP	LPLFEPKGRVLLVDGHHLAYRTFFALKGLTTSRGEPVQAVYGFAKSLLKALKEDGYKAVFVVFDAKAPSFRHEAYEAYKAGRAPTPEDFPRQHALI										
Translation of 5D3	-MAMLP	LPLFEPKGRVLLVDGHHLAYRTVFAALKGLTTSRGEPVQAVYGFAKSLLKALKEDGYKAVFVVFDAKAPSFRHEAYEAYKAGRAPAPEDFSRQLALI										
Translation of 5D4	-MAMLP	LPLFEPKGRVLLVDGHHLAYRTFFALKGLTTSRGEPVQAVYGFAKSLLKALKEDGYKAVFVVFDAKAPSFRHEAHEAYKAGRAPSPEEDFPRQLALI										
Translation of Taq pol	M	RGMPLPLFEPKGRVLLVDGHHLAYRTFHALKGLTTSRGEPVQAVYGFAKSLLKALKEDG-DAVIVVFDAKAPSFRHEAYGGYKAGRAPTPEDFPRQLALI										
Translation of Tth pol	M	EAMLP	LPLFEPKGRVLLVDGHHLAYRTFFFALKGLTTSRGEPVQAVYGFAKSLLKALKEDGYKAVFVVFDAKAPSFRHEAYEAYKAGRAPTPEDFPRQLALI									
Translation of Tfl pol	-MAMLP	LPLFEPKGRVLLVDGHHLAYRTFFFALKGLTTSRGEPVQAVYGFAKSLLKALKEDG-DVVVVVVFDAKAPSFRHEAYEAYKAGRAPTPEDFPRQLALI										
	MMA	MLPLFEPKGRVLLVDGHHLAYRTFFFALKGLTTSRGEPVQAVYGFAKSLLKALKEDGYKAVFVVFDAKAPSFRHEAYEAYKAGRAPTPEDFPRQLALI										
	110	120	130	140	150	160	170	180	190	200		
Translation of 4C11	KELV	DLLGFTRLEVQGYEADDVLATLAKKAEEKEGYEVRLTADRDLYQLVSDRVAVLHPEGHLITPEEWLWEKYGLRPEQWVDFRALVGDPSDNLPGIKG										
Translation of 5B1	KELV	DLLGFTRLEVQGYEADDVLATLAKKAEEKEGYEVRLTADRDLYQLVSDRVAVLHPEGHLITPEEWLWEKYGLRPEQWVDFRALVGDPSDNLPGVKG										
Translation of 5B4	KELV	DLLGFTRLEVQGYEADDVLATLAKKAEEKEGYEVRLTADRDLYQLVSDRVAVLHPEGHLITPEEWLWEKYGLRPEQWVDFRALVGDPSDNLPGVKG										
Translation of 5D3	KELV	DLLGFTRLEVQGYEADDVLATLAKKAEEKEGYEVRLTADRDLYQLVSDRVAVLHPEGHLITPEEWLWEKYGLRPEQWVDFRALVGDPSDNLPGVKG										
Translation of 5D4	KELV	DLLGFTRLEVQGYEADDVLATLAKKAEK L ARLEVPGYEADDVLASLAKKAEKEGYEVRLTADRDLYQLLSDRIHVLHPEGYLI	T	A	P	W	A	T	G	E		
Translation of Taq pol	KELV	DLLLG	PGYEADDVLATLAKKAEEKEGYEVRLTADRDLYQLLSDRIHVLHPEGYLI	T	A	P	W	A	T	G		
Translation of Tth pol	KELV	DLLLG	PGYEADDVLATLAKKAEEKEGYEVRLTADRDLYQLLSDRIHVLHPEGYLI	T	A	P	W	A	T	G		
Translation of Tfl pol	KELV	DLLLG	PGYEADDVLATLAKRAAEKEGYEVRLTADRDLYQLLSDRIHVLHPEGYLI	T	A	P	W	A	T	G		
	KELV	DLLLG	PGYEADDVLATLAKRAAEKEGYEVRLTADRDLYQLLSDRIHVLHPEGYLI	T	A	P	W	A	T	G		
	KELV	DLLLG	PGYEADDVLATLAKRAAEKEGYEVRLTADRDLYQLLSDRIHVLHPEGYLI	T	A	P	W	A	T	G		
	210	220	230	240	250	260	270	280	290	300		
Translation of 4C11	GEK	TALKLKEWGSLENLLKNLDRVKPENVREKIKAHLLEDLRLSLELS	S	RVR	TDLPLEVDFAKRREPDRERLRAFLERLEFGSLLHEFGLL	E	SPK	AL	EE	AP		
Translation of 5B1	GEK	TALKLKEWGSLENLLKNLDRLKPK-AIREKI	I	LAHMDDLKL	SWDLAKVRTDLP	LEVDFAKRREPDRERLRAFLERLEFGSLLHEFGLL	S	KA	LE	EE		
Translation of 5B4	GEK	TALKLKEWGSLENLLKNLDRLKPK-AIREKI	I	LAHMDDLKL	SWDLAKVRTDLP	LEVDFAKRREPDRERLRAFLERLEFGSLLHEFGLL	S	KA	LE	EE		
Translation of 5D3	GEK	TALKLKEWGSLENLLKNLDRLKPK-AIREKI	I	LAHMDDLKL	SWDLAKVRTDLP	LEVDFAKRREPDRERLRAFLERLEFGSLLHEFGLL	S	KA	LE	EE		
Translation of 5D4	GEK	TALKLKEWGSLENLLKNLDRLKPK-AIREKI	I	LAHMDDLKL	SWDLAKVRTDLP	LEVDFAKRREPDRERLRAFLERLEFGSLLHEFGLL	S	KA	LE	EE		
Translation of Taq pol	GEK	TAKLKEWGSLENLLKNLDRLKPK-AIREKI	I	LAHMDDLKL	SWDLAKVRTDLP	LEVDFAKRREPDRERLRAFLERLEFGSLLHEFGLL	S	KA	LE	EE		
Translation of Tth pol	GEK	TALKLKEWGSLENLLKNLDRVKPENVREKIKAHLLEDLRLSLELS	S	RVR	TDLPLEVDFAKRREPDRERLRAFLERLEFGSLLHEFGLL	E	AP	AL	EE			
Translation of Tfl pol	GEK	TALKLKEWGSLENLFQHLDQVKP-SLREKLQAGMEALALSRKLSQVHTDLP	H	LEVDFAKRREPDRERLRAFLERLEFGSLLHEFGLL	TPNLEG	LR	FL	GP	AA	EE		
	GEK	TALKLKEWGSLENLLKNLDRLKPEAI	R	REKI	LAHMDDLKL	SWDLA.	VRTDLP	LEVDFAKRREPDRERLRAFLERLEFGSLLHEFGLL	E	AP	AL	EE
	310	320	330	340	350	360	370	380	390	400		
Translation of 4C11	WPP	PEGA	FVGFLSRKEPMWADLLALA	AKGGRVHRAPEPYKALRDLKEARG	LLAKDL	SVLALREG	GLLPPGDDPM	LL	DP	SNTTPEGVARRYGG		
Translation of 5B1	WPP	PEGA	FVGFLSRKEPMWADLLALA	ARGGRVHRAPEPYKALRDLKEARG	LLAKDL	SVLALREG	GLLPPGDDPM	LL	DP	SNTTPEGVARRYGG		
Translation of 5B4	WPP	PEGA	FVGFLSRKEPMWADLLALA	ARGGRVHRAPEPYKALRDLKEARG	LLAKDL	SVLALREG	GLLPPGDDPM	LL	DP	SNTTPEGVARRYGG		
Translation of 5D3	WPP	PEGA	FVGFLSRKEPMWADLLALA	ARGGRVHRAPEPYKALRDLKEARG	LLAKDL	SVLALREG	GLLPPGDDPM	LL	DP	SNTTPEGVARRYGG		
Translation of 5D4	WPP	PEGA	FVGFLSRKEPMWADLLALA	ARGGRVHRAPEPYKALRDLKEARG	LLAKDL	SVLALREG	GLLPPGDDPM	LL	DP	SNTTPEGVARRYGG		
Translation of Taq pol	WPP	PEGA	FVGFLSRKEPMWADLLALA	ARGGRVHRAPEPYKALRDLKEARG	LLAKDL	SVLALREG	GLLPPGDDPM	LL	DP	SNTTPEGVARRYGG		
Translation of Tth pol	WPP	PEGA	FVGFLSRKEPMWADLLALA	ARGGRVHRAPEPYKALRDLKEARG	LLAKDL	SVLALREG	GLLPPGDDPM	LL	DP	SNTTPEGVARRYGG		
Translation of Tfl pol	WPP	PEGA	FVGFLSRKEPMWADLLALA	ARGGRVHRAPEPYKALRDLKEARG	LLAKDL	SVLALREG	GLLPPGDDPM	LL	DP	SNTTPEGVARRYGG		
	WPP	PEGA	FVGFLSRKEPMWADLLALA	ARGGRVHRAPEPYKALRDLKEARG	LLAKDL	SVLALREG	GLLPPGDDPM	LL	DP	SNTTPEGVARRYGG		
	410	420	430	440	450	460	470	480	490	500		
Translation of 4C11	TEE	AGERAALSERLFANLWGRLEGEER	LLWL	YREVERPLSAVL	AHMEA	TGVR	LDV	AYL	RALS	LEVAEEIR		
Translation of 5B1	TEE	AGERAALSERLFANLWGRLEGEER	LLWL	YREVERPLSAVL	AHMEA	TGVR	LDV	AYL	RALS	LEVAEEIR		
Translation of 5B4	TEE	AGERAALSERLFANLWGRLEGEER	LLWL	YREVERPLSAVL	AHMEA	TGVR	LDV	AYL	RALS	LEVAEEIR		
Translation of 5D3	TEE	AGERAALSERLFANLWGRLEGEER	LLWL	YREVERPLSAVL	AHMEA	TGVR	LDV	AYL	RALS	LEVAEEIR		
Translation of 5D4	TEE	AGERAALSERLFANLWGRLEGEER	LLWL	YREVERPLSAVL	AHMEA	TGVR	LDV	AYL	RALS	LEVAEEIR		
Translation of Taq pol	TEE	AGERAALSERLFANLWGRLEGEER	LLWL	YREVERPLSAVL	AHMEA	TGVR	LDV	AYL	RALS	LEVAEEIR		
Translation of Tth pol	TEE	AGERAALSERLFANLWGRLEGEER	LLWL	YREVERPLSAVL	AHMEA	TGVR	LDV	AYL	RALS	LEVAEEIR		
Translation of Tfl pol	TEE	AGERAALSERLFANLWGRLEGEER	LLWL	YREVERPLSAVL	AHMEA	TGVR	LDV	AYL	RALS	LEVAEEIR		

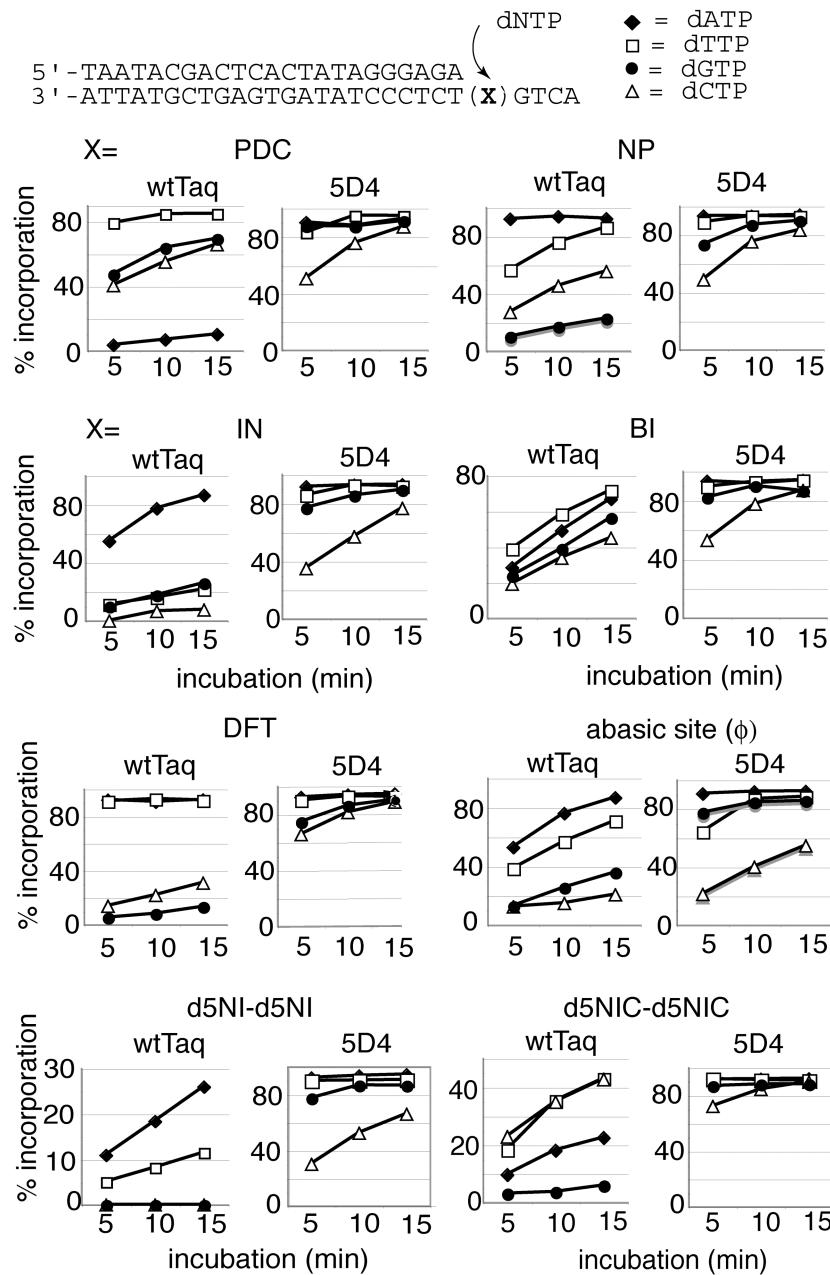
	510	520	530	540	550	560	570	580	590	600
Translation of 4C11	<i>GLPAIGKTEKTGKRSTSAAVLEALREAHPIVEKILQYRELTKLKSTYIDPLPDLIHPRTGRLHTRFNQTAATATGRLSSSDPNLQNI</i>	<i>PVRTPPLGQRIRRAF</i>								
Translation of 5B1	<i>GLPAIGKTEKTGKRSTSAAVLEALREAHPIVEKILQYRELTKLKSTYIDPLPDLIHPRTGRLHTRFNQTAATATGRLSSSDPNLQNI</i>	<i>PVRTPPLGQRIRRAF</i>								
Translation of 5B4	<i>GLPAIGKTEKTGKRSTSAAVLEALREAHPIVEKILQYRELTKLKSTYIDPLPDLIHPRTGRLHTRFNQTAATATGRLSSSDPNLQNI</i>	<i>PVRTPPLGQRIRRAF</i>								
Translation of 5D3	<i>GLPAIGKTEKTGKRSTSAAVLEALREAHPIVEKILQYRELTKLKSTYIDPLPDLIHPRTGRLHTRFNQTAATATGRLSSSDPNLQNI</i>	<i>PVRTPPLGQRIRRAF</i>								
Translation of 5D4	<i>GLPAIGKTEKTGKRSTSAAVLEALREAHPIVEKILQYRELTKLKSTYIDPLPDLIHPRTGRLHTRFNQTAATATGRLSSSDPNLQNI</i>	<i>PVRTPPLGQRIRRAF</i>								
Translation of Taq pol	<i>GLPAIGKTEKTGKRSTSAAVLEALREAHPIVEKILQYRELTKLKSTYIDPLPDLIHPRTGRLHTRFNQTAATATGRLSSSDPNLQNI</i>	<i>PVRTPPLGQRIRRAF</i>								
Translation of Tth pol	<i>RLPALGKTQ</i>	<i>TGKRSTSAAVLEALREAHPIVEKILQHRELTKLN</i>	<i>TYVDPLPSLVHPR</i>	<i>TGRLHTRFNQTAATATGRLSSSDPNLQNI</i>	<i>PVRTPPLGQRIRRAF</i>					
Translation of Tfl pol	<i>GLPAIGKTEKTGKRSTSAAVLEALREAHPIVDRILQYRELTKLN</i>	<i>TYIDPLPA</i>	<i>LVHPKTGRLHTRFNQTAATATGRLSSSDPNLQNI</i>	<i>PVRTPPLGQRIRRAF</i>						
	<i>GLPAIGKTEKTGKRSTSAAVLEALREAHPIVEKILQYRELTKLKSTYIDPLPDLIHPRTGRLHTRFNQTAATATGRLSSSDPNLQNI</i>	<i>PVRTPPLGQRIRRAF</i>								
	610	620	630	640	650	660	670	680	690	700
Translation of 4C11	<i>IAEGGWLLVVLDYSQMELRVLAHLSGDENLIRVFQEGRDIHTE</i>	<i>TASWMFGVPREAVDPLMRRRAAKTINF</i>	<i>GVLYGMMSAHLRSQELAI</i>	<i>PYEEAQAFIERYFQ</i>						
Translation of 5B1	<i>IAEGGWLLVVLDYSQMELRVLAHLSGDENLIRVFQEGRDIHTE</i>	<i>TASWMFGVPREAVDPLMRRRAAKTINF</i>	<i>GVLYGMMSAHLRSQELAI</i>	<i>PYEEAQAFIERYFQ</i>						
Translation of 5B4	<i>IAEGGWLLVVLDYSQMELRVLAHLSGDENLIRVFQEGRDIHTE</i>	<i>TASWMFGVPREAVDPLMRRRAAKTINF</i>	<i>GVLYGMMSAHLRSQELAI</i>	<i>PYEEAQAFIERYFQ</i>						
Translation of 5D3	<i>IAEGGWLLVVLDYSQMELRVLAHLSGDENLIRVFQEGRDIHTE</i>	<i>TASWMFGVPREAVDPLMRRRAAKTINF</i>	<i>GVLYGMMSAHLRSQELAI</i>	<i>PYEEAQAFIERYFQ</i>						
Translation of 5D4	<i>IAEGGWLLVVLDYSQMELRVLAHLSGDENLIRVFQEGRDIHTE</i>	<i>TASWMFGVPREAVDPLMRRRAAKTINF</i>	<i>GVLYGMMSAHLRSQELAI</i>	<i>PYEEAQAFIERYFQ</i>						
Translation of Taq pol	<i>IAE[EGWLLVVLDYSQIELRVLAHLSGDENLIRVFQEGRDIHTE</i>	<i>[QTAASWMFGVPPEAVDPLMRRRAAKTINF</i>	<i>GVLYGMMSAHLRSQELAI</i>	<i>PYEEAQAFIERYFQ</i>						
Translation of Tth pol	<i>VAAEAGWALVALDYSQIELRVLAHLSGDENLIRVFQEGRDIHTE</i>	<i>[QTAASWMFGVPSPEGVDP</i>	<i>LMRRRAAKTINF</i>	<i>GVLYGMMSAHLRSQELAI</i>	<i>PYEEAQAFIERYFQ</i>					
Translation of Tfl pol	<i>VAAE[EGWVLLVVLDYSQIELRVLAHLSGDENLIRVFQEGRDIHTE</i>	<i>[QTAASWMFGVPSPEGVDP</i>	<i>LMRRRAAKTINF</i>	<i>GVLYGMMSAHLRSQELAI</i>	<i>PYEEAQAFIERYFQ</i>					
	<i>IAEGGWLLVVLDYSQMELRVLAHLSGDENLIRVFQEGRDIHTE</i>	<i>TASWMFGVPREAVDPLMRRRAAKTINF</i>	<i>GVLYGMMSAHLRSQELAI</i>	<i>PYEEAQAFIERYFQ</i>						
	710	720	730	740	750	760	770	780	790	800
Translation of 4C11	<i>SFPKVRAWIYEKTLEEGRRRGYVETLFGRRRYVPDL</i>	<i>EARVKSVREAAERMAFNMPVQGTAADLMK</i>	<i>LAMVKLFPRL</i>	<i>ETGARM</i>	<i>LLQVHDEL</i>	<i>VLEAPKERA</i>				
Translation of 5B1	<i>SFPKVRVWIYEKTLEEGRRRGYVETLFGRRRYVPDL</i>	<i>EARVKSVREAAERMAFNMPVQGTAADLMK</i>	<i>LAMVKLFPRL</i>	<i>ETGARM</i>	<i>LLQVHDEL</i>	<i>VLEAPKERA</i>				
Translation of 5B4	<i>SFPKVRAWIYEKTLEEGRRRGYVETLFGRRRYVPDL</i>	<i>NARVKSVREAAERMAFNMPVQGTAADLMK</i>	<i>LAMVKLFPRL</i>	<i>ETGARM</i>	<i>LLQVHDEL</i>	<i>VLEAPKERA</i>				
Translation of 5D3	<i>SFPKVRAWIYEKTLEEGRRRGYVETLFGRRRYVPDL</i>	<i>EARVKSAREAAERMAFNMPVQGTAADLMK</i>	<i>LA</i>	<i>VVKLFPRL</i>	<i>ETGARM</i>	<i>LLQVHDEL</i>	<i>VLEAPKERA</i>			
Translation of 5D4	<i>SFPKVRAWIYEKTLEEGRRRGYVETLFGRRRYVPDL</i>	<i>NARVKSVREAAERMAFNMPVQGTAADLMK</i>	<i>LA</i>	<i>VVKLFPRL</i>	<i>ETGARM</i>	<i>LLQVHDEL</i>	<i>VLEAPKERA</i>			
Translation of Taq pol	<i>SFPKVRAWIYEKTLEEGRRRGYVETLFGRRRYVPDL</i>	<i>NARVKSVREAAERMAFNMPVQGTAADLMK</i>	<i>LAMVKLFPRL</i>	<i>EM</i>	<i>GARM</i>	<i>LLQVHDEL</i>	<i>VLEAPKERA</i>			
Translation of Tth pol	<i>SFPKVRAWIYEKTLEEGRRKGYVETLFGRRRYVPDL</i>	<i>NARVKSVREAAERMAFNMPVQGTAADLMK</i>	<i>LAMVRLFPRL</i>	<i>REM</i>	<i>GARM</i>	<i>LLQVHDEL</i>	<i>VLEAPQARAEE</i>			
Translation of Tfl pol	<i>SYPKVRAWIYEKTLEEGRRRGYVETLFGRRRYVPDL</i>	<i>NARVKSVREAAERMAFNMPVQGTAADLMK</i>	<i>LAMVRLFPRL</i>	<i>QEL</i>	<i>GARM</i>	<i>LLQVHDEL</i>	<i>VLEAPKDRAER</i>			
	<i>SFPKVRAWIYEKTLEEGRRRGYVETLFGRRRYVPDL</i>	<i>ARVKSVREAAERMAFNMPVQGTAADLMK</i>	<i>LAMVKLFPRL</i>	<i>ETGARM</i>	<i>LLQVHDEL</i>	<i>VLEAPKERA</i>				
	810	820	830	840	850	860	870	880	890	900
Translation of 4C11	<i>VARLAKEVMEGVYPLAVPLEVEVGIGEDWL</i>	<i>SAKE</i>								
Translation of 5B1	<i>VARLAKEVMEGVYPLAVPLEV</i>	<i>GVGIGEDWL</i>	<i>SAKE</i>							
Translation of 5B4	<i>VARLAKEVMEGVYPLAVPLEV</i>	<i>GVGIGEDWL</i>	<i>SAKE</i>							
Translation of 5D3	<i>VARLAKEVMEGVYPLAVPLEV</i>	<i>GVGIGEDWL</i>	<i>SAKE</i>							
Translation of 5D4	<i>VARLAKEVMEGVYPLAVPLEV</i>	<i>GVGIGEDWL</i>	<i>SAKE</i>							
Translation of Taq pol	<i>VARLAKEVMEGVYPLAVPLEV</i>	<i>GVGIGEDWL</i>	<i>SAKE</i>							
Translation of Tth pol	<i>VAALAKEAMEKAYPLAVPLEVEVG</i>	<i>MGEDWL</i>	<i>SAKE</i>							
Translation of Tfl pol	<i>VAALAKEAMEKAYPLAVPLEVEVG</i>	<i>MGEDWL</i>	<i>SAKE</i>							
	<i>VARLAKEVMEGVYPLAVPLEVEVGIGEDWL</i>	<i>SAKE</i>								

Supporting information Fig. 1: Alignment of polymerase sequences



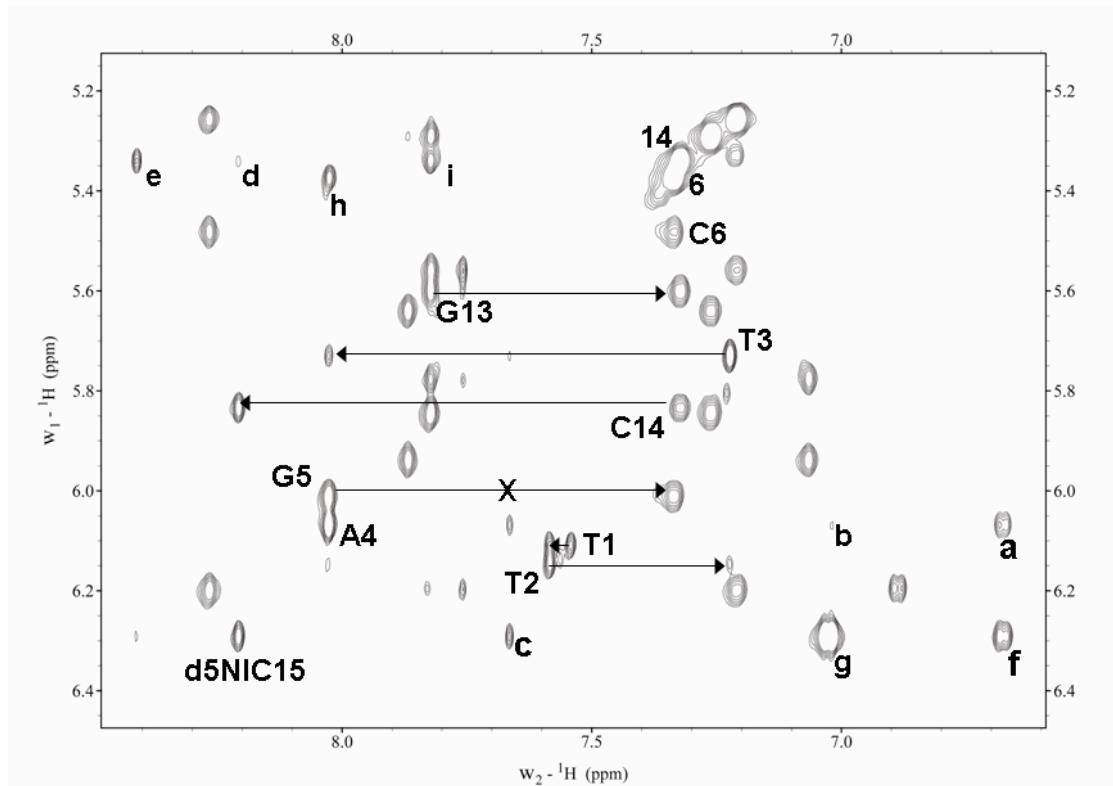
Supporting information Figure 2: Formation and extension of the d5NI self-pair and transversion mismatch extension.

(a) Formation and extension of a d5NI-d5NI self-pair is shown for both Taq and 5D4. While d5NI-TP is incorporated efficiently opposite template d5NI by both polymerases, further extension of the self-pair is stalled in both cases, with only 5D4 showing some weak extension (< 5%). (b) Primer extension (%) is plotted against time for extension of an A'G mismatch by both Taq and 5D4. While Taq is unable to extend the mismatch, it is efficiently extended by 5D4.



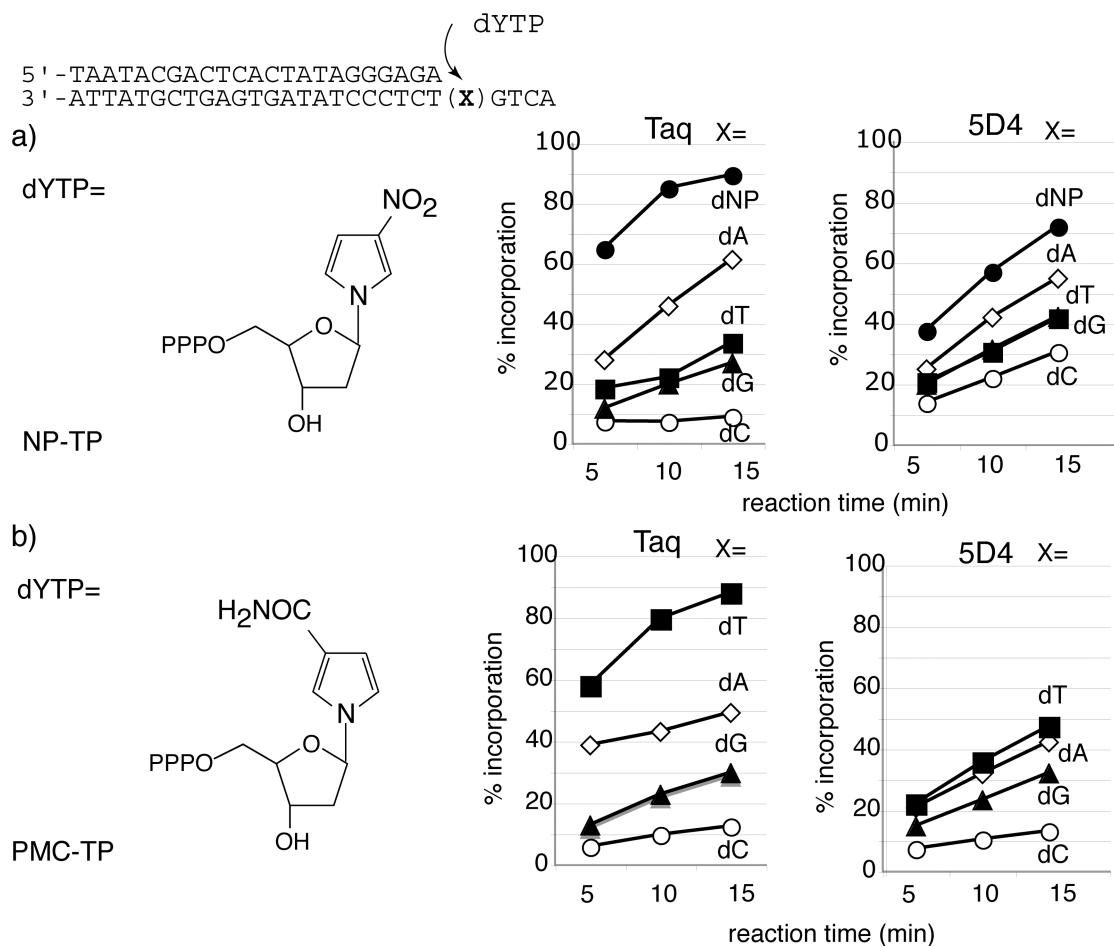
Supporting information Figure 3: Incorporation specificity opposite HBAs.

Primer extension (%) is plotted against time for incorporation of dATP (black diamonds), dTTP (open squares), dGTP (filled circles), dCTP (open triangles) for unnatural template bases NP, PDC, DFT, I, BI, $[d5NI]_2$ (2 consecutive d5NI bases), $[d5NIC]_2$ (2 consecutive d5NIC bases) and an abasic site (tetrahydrofuran (ϕ)) by polymerases Taq and 5D4. 5D4 displays greatly reduced incorporation bias for several HBA template bases with some (NP, PDC, DFT) approaching universal base behaviour.



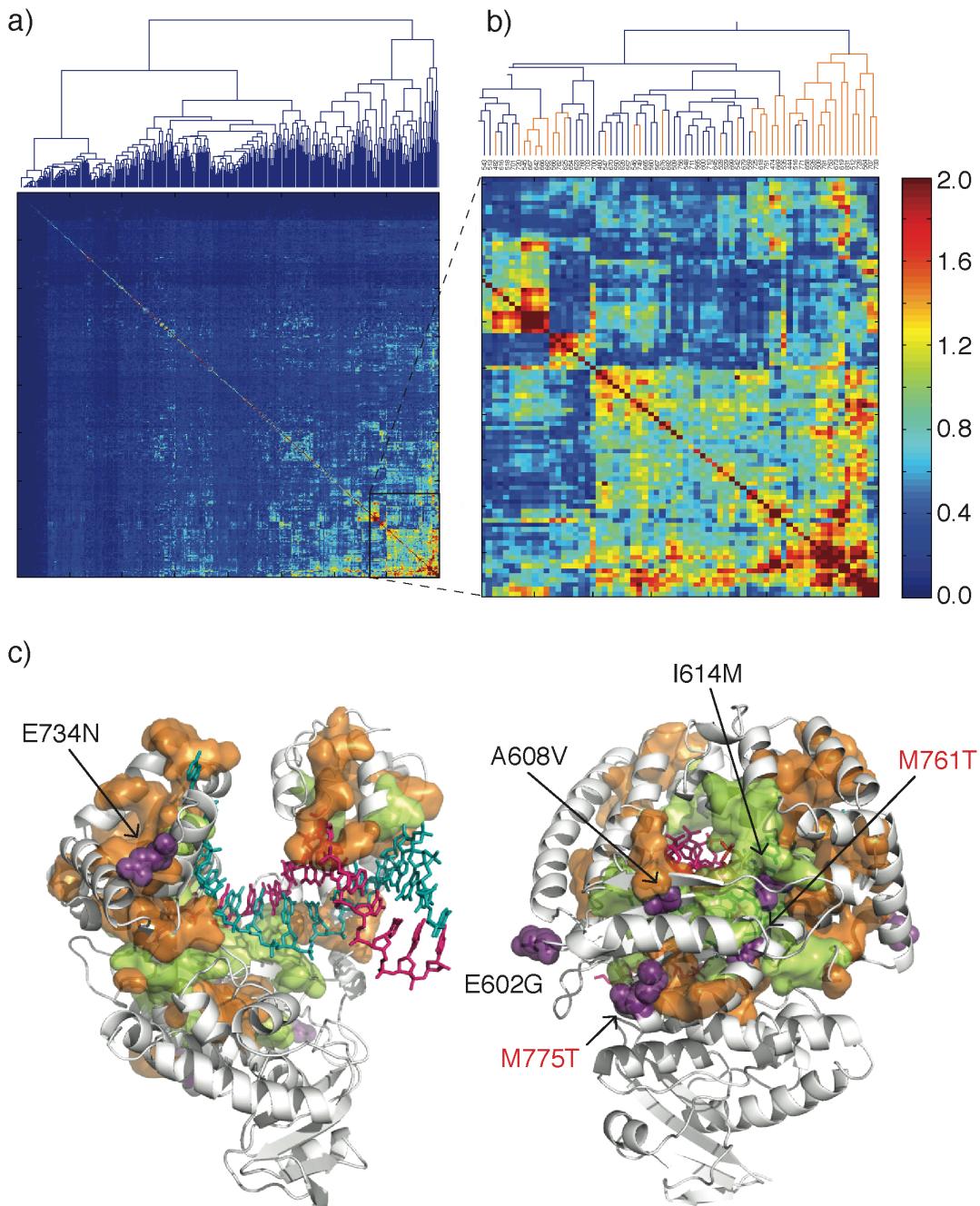
Supporting information Figure 4: NMR.

Aromatic–H1' region of the D₂O NOESY spectra (250 ms mixing time) of *tnic*, recorded at 25°C. Intra-residue pyrimidine H6-H1' and purine/d5NIC H8-H1' cross-peaks are labelled with residue name and number, intra-residue pyrimidine H5-H6 cross-peaks are labelled with residue number, and sequential NOE connectivities are indicated with arrows. For clarity, only T1 to C6 and G13 to d5NIC15 assignments are shown. Cross-peaks (a) to (i) are assigned as follows: (a) d5NIC15 H2-A4 H1'; (b) d5NIC15 H3-A4 H1'; (c) A4 H2-d5NIC15 H1'; (d) d5NIC15 H8-C14 H5; (e) d5NI15 H6-C14 H5; (f) d5NIC15 intra-residue H2-H1'; (g) d5NIC15 intra-residue H3-H1'; (h) G5 H8-C6 H5; (i) G13 H8-C14 H5. The absence of the sequential A4 H2-G5 H1' NOE is marked with an “X”.

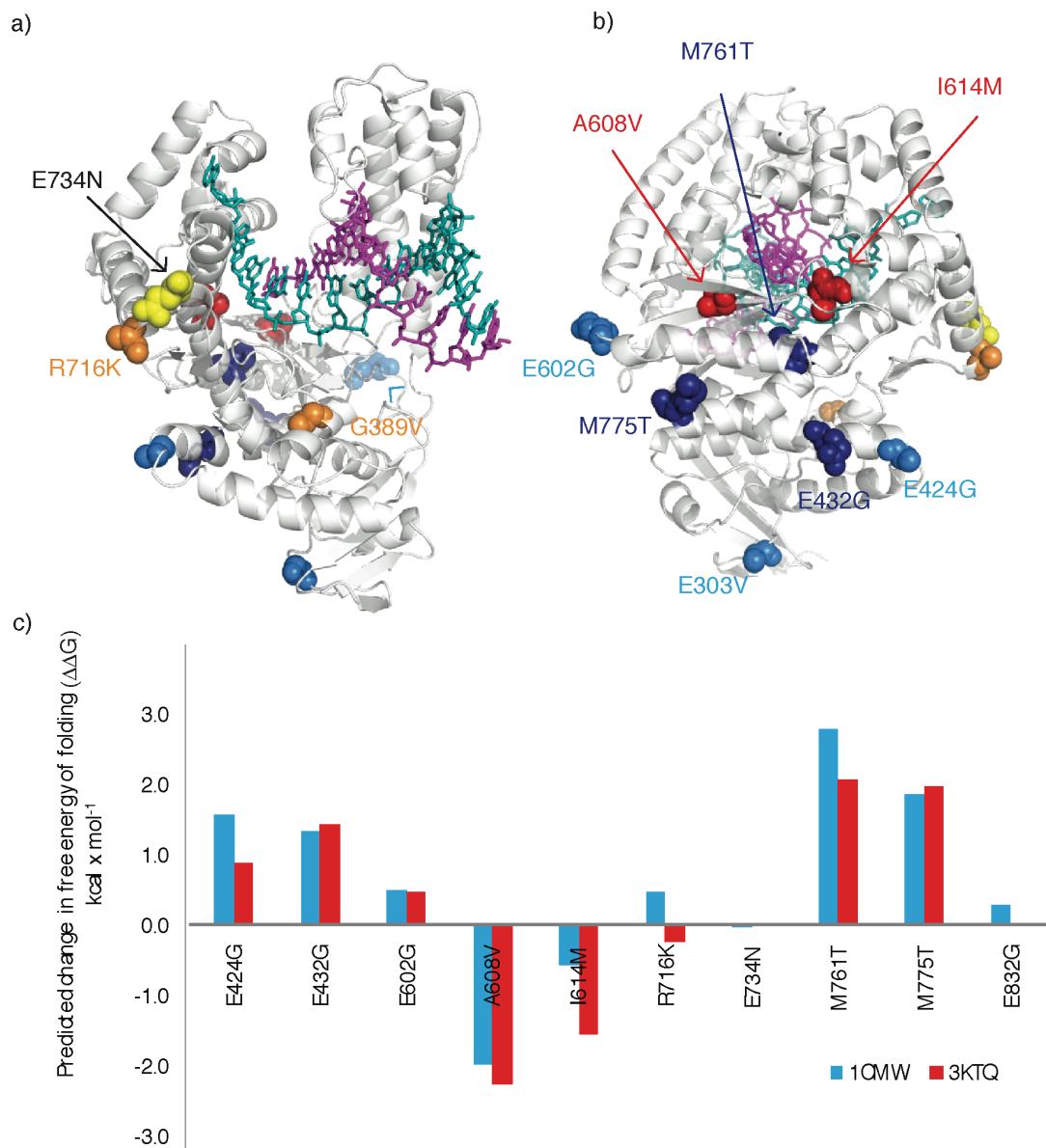


Supporting information Figure 5: Incorporation of HBA triphosphates.

Nucleotide triphosphate incorporation (%) is plotted against time for incorporation of dYTP (a) Nitro-pyrrol-triphosphate (NP-TP) and (b) Pyrrol-monocarboxamide-triphosphate (PMC-TP) for template X=dA (open diamonds), dT (filled squares), dG (filled triangles), dC (open circles) and dNP (filled circles (a)) for both Taq and 5D4. Incorporation efficiencies are comparable with Taq slightly superior.



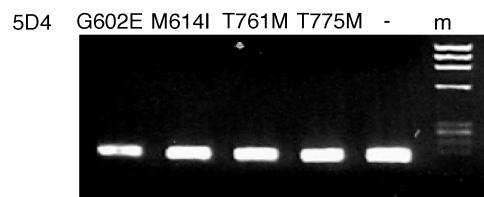
Supporting information Figure 6: SCA. Statistical coupling analysis (SCA)¹ of bacterial polymerases. The hierarchical clustering of the output coupling matrix (a) and a detail of the region containing the highest coupling values (b) are shown. Residues with couplings above $\Delta\Delta G = 1.8 \text{ } kT^*$ were selected to represent the SCA network, mapped to the structure of the Taq polymerase domain (3KTQ²) shown coloured orange in surface representation (c). Conserved residues (> 97% identity) are shown as the green surface. The template strand is shown in teal and primer strand in magenta. 5D4 mutations are shown in purple. Mutations M761T and M775T with connections to the SCA network are labelled in red.



Supporting information Figure 7: FoldX analysis of 5D4 mutations. FoldX was used to predict the change in free energy of folding due to individual point mutations in the available Taq structures (see supporting information Table 5). Predictions were ranked and mutations are shown superimposed on the Taq structure (3KTQ) (a, b) and coloured according to their effect on $\Delta\Delta G$: red = stabilising mutations ($\Delta\Delta G < -1.0$ kcal mol⁻¹ in most structures, e.g. A608V, I614M), orange = weakly stabilising mutations ($-1.0 < \Delta\Delta G < 0.0$ kcal mol⁻¹, e.g. R715K, G389V), yellow = neutral mutations (e.g. E734N), light blue = potentially destabilising ($0.0 < \Delta\Delta G < 1.0$ kcal mol⁻¹, e.g. E602G, E424G) and dark blue = destabilising mutations ($\Delta\Delta G > 1.0$ kcal mol⁻¹, e.g. M761T, M775T). (c) Predicted change in free energy of folding ($\Delta\Delta G$) in kcal mol⁻¹ is plotted for individual mutations for two different Taq structures 1CMV³ (Apoform) and 3KTQ² (closed ternary complex).

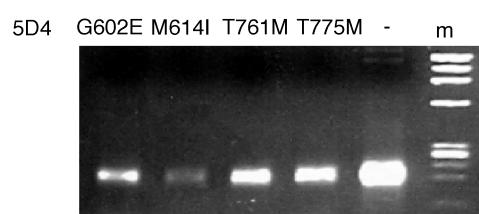
a)

1 : 5' - CAGGAAACAGCTATGACAAAAATCTAGATAACGAGGGCAA
4 : 5' - GTAAAACGACGCCAGTACCACCGAACTGCAGGTGACGCCAAGCG



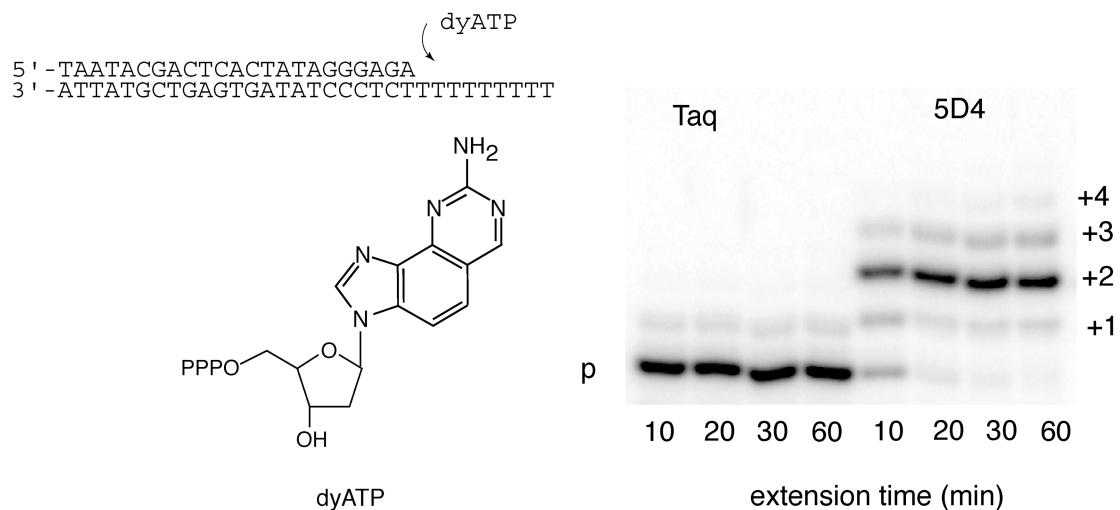
b)

2 : 5' - CAGGAAACAGCTATGACAAAAATCTAGATAACGAGGGCA (x)
6 : 5' - GTAAAACGACGCCAGTACCAC (x) GAACTGCAGGTGACGCCAAGC (x)
x = d5NIC



Supporting information Figure 8: d5NIC PCR of reversion mutants.

PCR amplification using (a) unmodified standard primers (1,4), or primers in which either only the 3' terminal base (2) or the 3' terminal base as well as an internal base are substituted by d5NIC (6) for reversion mutants (from left to right) (5D4:G602E, 5D4:M614I, 5D4:T761M, 5D4:T775M) as well as 5D4. While all polymerases show comparable activity with unmodified primers only 5D4:M614I shows a significant relative reduction in PCR activity with d5NIC-modified primers. m, ϕ X174 HaeIII digest marker.



Supporting information Figure 9: Incorporation of dyATP

Incorporation of the size-expanded dATP analogue dyATP⁴ is shown for both Taq and 5D4. While incorporation by Taq is poor and stalls at +1, 5D4 can incorporate up to +4.

2. Supporting Tables

Supporting Table 1: Shared point mutations in the polymerase domain
 (numbering of mutations refers to the Taq gene).

Mutations ^a	4C11	5B1	5B4	5D3	5D4
E602G	+	+	+	+	+
A608V	+	+	+	+	+
I614M	+	+	+	+	+
M762T			+		+
M775T	+	+	+	+	+

Supporting Table 2: K_m (μ M) and k_{cat}/K_m values (% min⁻¹ μ M⁻¹) for Taq and 5D4 (Fig. 2a)

dXTP	Template base	K _m wtTaq	k _{cat} /K _m wtTaq	K _m 5D4	k _{cat} /K _m 5D4	f
5NI	dA	111	2.8	9	227.6	81
	dT	109	1.5	8	316.1	211
	dC	26	2.0	3	1364.4	682
	dG	34	2.1	5	683.9	326
5NIC	dA	13	29.6	3	1475.9	50
	dT	10	42.4	5	529.9	13
	dC	9	15.6	2	796.6	51
	dG	15	19.3	1	2103.4	109
A	d5NI	76	0.3	28	21.4	71
T		51	0.5	38	11.8	24
C		11	0.10	14	5.3	53
G		20	0.07	28	5.3	76
NI		8	9.2	9	314.9	34
A	d5NIC	27	0.2	41	13.4	67
T		55	0.1	33	13.6	136
C		47	0.1	93	2.2	22
G		99	0.04	145	4.4	109
NIC		9	3.2	1	979.3	306
T	dA	3	187.6	3	498.9	2.7
G	dC	3	763.0	2	1069.0	1.4

Relative catalytic efficiency: $f = k_{cat}/K_m \text{ 5D4} / k_{cat}/K_m \text{ wtTaq}$

Supporting Table 3: Steady-state extension constants k_{cat} of 3' d5NI by 5D4 (% min⁻¹ μM⁻¹)

3' base	Template N	template N+1	k_{cat}/K_m
d5NI	A	A	3.0
	A	T	2.6
	C	A	4.6
	C	T	3.7
	C	C	4.4
d5NIC	T	A	40.3
	T	T	10.3

Supporting Table 4 NMR.

Experimental constraints and refinement analysis for the *tni* and *tnic* NMR model structures

Experimental constraints ¹	<i>tni</i>	<i>tnic</i>
NOE distance constraints	228	236
Hydrogen bonding distance constraints	6	8
Dihedral constraints	50	48
Refinement Analysis	<i>tni</i>	<i>tnic</i>
Distance constraint violation energy ² (kcal mol ⁻¹)	6.9	3.3
Dihedral constraint violation energy ³ (kcal mol ⁻¹)	14.2	12.4
rms deviation from ideal bond lengths (Å)	0.02	0.02
rms deviation from ideal bond angles (°)	1.9	1.8

¹for residues T1-C6 and G13-d5NI(C)15 (see Methodology)

²Force constant K = 50 Kcal mol⁻¹ Å⁻²

³Force constant K = 0.01 Kcal mol⁻¹ deg⁻²

Supporting Table 5: FoldX calculation of $\Delta\Delta G$ changes of polymerase domain mutations for different structures of Taq polymerase

Mutation	Structure				
	1CMW	1KTQ	4KTQ	2KTQ	3KTQ
P264S	0.91	N.S.	N.S.	N.S.	N.S.
E303V	0.10	0.40	0.19	0.74	0.56
G389V	0.72	-1.19	-0.38	-0.71	0.12
E424G	1.57	0.03	-0.18	0.54	0.88
E432G	1.34	1.67	0.22	1.73	1.42
E602G	0.5	0.57	-0.04	0.84	0.47
A608V	-1.98	-2.79	-2.35	0.10	-2.27
I614M	-0.56	-0.52	-0.86	-1.10	-1.54
R716K	0.48	-0.55	-0.2	-0.41	-0.23
E734N	-0.03	0.19	0.08	0.26	0.00
M761T	2.79	2.45	1.84	2.30	2.07
M775T	1.85	1.47	0.91	2.33	1.98
E832G	0.28	N.S.	N.S.	N.S.	N.S.

Changes in the predicted free energy of folding ($\Delta\Delta G$) of Taq polymerase domain point mutants. FoldX was used to predict the effect on protein stability for each individual 5D4 mutation using some of the different Taq polymerase domain structures available. Mutations are numbered as per 1CMW structure. Structures used included apo structures for the full length Taq (1CMW³) and Klentaq fragment (1KTQ²), the binary complex of polymerase bound to DNA (4KTQ²), and both open (2KTQ²) and closed (3KTQ²) ternary complexes of Taq polymerase with DNA and incoming nucleotide. Energy is given as kcal mol⁻¹. Residues not present in the structure are shown as N.S. .

3. List of primers (X=d5NIC, Y=d5NI)

- 1: 5'-CAG GAA ACA GCT ATG ACA AAA ATC TAG ATA ACG AGG GCA **X**-3'
- 2: 5'- GTA AAA CGA CGG CCA GTA CCA CCG AAC TGC GGG TGA CGC CAA GC**X** -3'
- 3: 5'-CAG GAA ACA GCT ATG ACA AAA ATC TAG ATA **XCG** AGG GCA **X**-3'
- 4: 5'-GTA AAA CGA CGG CCA GTA CCA **CXG** AAC TGC GGG TGA CGC CAA GC**X**-3'
- 5: 5'-GTA AAA CGA CGG CCA GTA **CXA** CCG AAC TGC GGG TGA CGC CAA **GX**-3'
- 6: 5'-CAG GAA ACA GCT ATG ACA AAA ATC TAG ATA **YCG** AGG GCA **Y**-3'
- 7: 5'-GTA AAA CGA CGG CCA GTA CCA **CYG** AAC TGC GGG TGA CGC CAA GC**Y**-3'
- 8-10: primers 1-3 as described⁵
- 11-13: primer 1, 2, 5 as described⁶
- 14: 5'-CAG GAA ACA GCT ATG ACA AA-3'
- 15: 5'-GTA AAA CGA CGG CCA GTA CC-3'
- 16-19: 5'-AGC TAC CAT GCC TGC ACG CAG CNG GCA TCC GTC GCG ACC ACG TT**B** TTC
GTG GTC GCG ACG GAT GCC **X**-3'
N=dA, dG, dC, dT; B=dU-biotin
- 20-23: 5'-AGC TAC CAT GCC TGC ACG CAG CNG GCA TCC GTC GCG ACC ACG TT**B** TTC
GTG GTC GCG ACG GAT GCC **Y**-3'
N=dA, dG, dC, dT; B=dU-biotin
20. 5'-TAA TAC GAC TCA CTA TAG GGA GA-3'
21. 5'-TAA TAC GAC TCA CTA TAG GGA GAX-3'
22. 5'-TAA TAC GAC TCA CTA TAG GGA GAY-3'
23. 5'- ACTGXT CTC CCT ATA GTG AGT CGT ATT A-3'
24. 5'- ACTGYT CTC CCT ATA GTG AGT CGT ATT A-3'
25. 5'- ACTGZT CTC CCT ATA GTG AGT CGT ATT A-3'
Z = DPC, NP, DFT, abasic site (φ), IN, BI, ICS, 7AI
26. 5'-TAG CTA **XTC** TCC CTA TAG TGA GTC GTA TTA-3'
27. 5'-TAG CTA **YTC** TCC CTA TAG TGA GTC GTA TTA-3'
28. 5'-TAA TAC GAC TCA CTA TAG GGA GAZ-3'
Z = ICS, 7AI
29. 5'-NNT CTC CCT ATA GTG AGT CGT ATT A-3'
N = dA, dT, dC, dG
30. 5'-TTT TTT TTT TZT CTC CCT ATA GTG AGT CGT ATT A-3'
Z = abasic site (φ)
- 31: 5'-ACC GCC GCC GAC CTC ATG AAG CTG GCT ATG GTG A-3'
- 32: 5'-TCA CCA TAG CCA GCT TCA TGA GGT CGG CGG CGG T-3'

33: 5'-GGC CTT CAT CGC CGA GGA GGG GTG GCT ATT GGT GG-3'
34: 5'-CCA CCA ATA GCC ACC CCT CCT CGG CGA TGA AGG CC-3'
35: 5'-GGT CCT GGA CTA TAG TCA GAT CGA GCT CAG GGT GCT G-3'
36: 5'-CAG CAC CCT GAG CTC GAT CTG ACT ATA GTC CAG GAC C-3'
37: 5'-CCC CAG GCT GGA GGA AAT GGG CGC CAG GAT GCT CCT TCA G-3'
38: 5'-CTG AAG GAG CAT CCT GGC GCC CAT TTC CTC CAG CCT GGG G-3'
39: 5'-ACC CGC CTC GAG GTC CCA GGC TAC GAG GCG GAC-3'
40: 5'-GTC CGC CTC GTA GCC TGG GAC CTC GAG GCG GGT-3'

References

1. Lockless, S. W.; Ranganathan, R., *Science* **1999**, 286, 295-299.
2. Li, Y.; Korolev, S.; Waksman, G., *EMBO J.* **1998**, 17, 7514-7525.
3. Urs, U. K.; Murali, R.; Krishna Murthy, H. M., *Acta Crystallogr D Biol Crystallogr* **1999**, 55, 1971-1977.
4. Lu, H.; He, K.; Kool, E. T., *Angew. Chem. Int. Ed.* **2004**, 43, 5834-5836.
5. d'Abbadie, M.; Hofreiter, M.; Vaisman, A.; Loakes, D.; Gasparutto, D.; Cadet, J.; Woodgate, R.; Paabo, S.; Holliger, P., *Nat Biotechnol* **2007**, 25, 939-943.
6. Ghadessy, F. J.; Ong, J. L.; Holliger, P., *Proc. Natl. Acad. Sci. USA* **2001**, 98, 4552-4557.