## 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 $4(4 \mathrm{C} 11)$ and round $5(5 \mathrm{~B} 1,5 \mathrm{~B} 4,5 \mathrm{D} 3,5 \mathrm{D} 4)$ 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 $\mathrm{N}=\mathrm{n}-1$ (for $\mathrm{n}>58$ ), respectively $\mathrm{N}=\mathrm{n}-2$ (for $\mathrm{n}>228$ ). 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

Translation of 4C1 Translation of 5B1 Translation of 5B4 Translation of 5B4 Translation of 5D3
Translation of 5D4 Translation of 5D4
Translation of Taq po Translation of Taq pol
Translation of Tth pol Translation of Tth pol
Translation of Tfl pol Translation of Tfl pol
 $-M A M L P L F E P K G R V L L V D G H H L A Y R T F F A L K G L T T S R G E P V Q A V Y G F A K S L L K A L K E D G-D V V V V V F D A K A P S F R H E A Y E A Y K A G R A P T P E D F P R Q L A L I$
$M M A M L P L F E P K G V L L V D G H H L A Y R T F F A L K G L T T S R G E P V Q A V Y G F A K S L L K A L K E D G Y K A V F V V F D A K A P S F R H E A Y E A Y K A G R A P T P E D F P R Q L A L I ~$

Translation of 4C11 Translation of 5B1 Translation of 5B4 Translation of 5B4 Translation of 5D3
Translation of 5D4 Translation of 5D4 Translation of Taq pol Translation of Tth pol Translation of Tfl pol
$120 \quad 130$
$130 \quad 140$
$140 \quad 150$
$150 \quad 160$
$160 \quad 170$
170180
180190
190
${ }_{2} 200$ KELVDLLGFTRLEVGGYEADDVLATLAKKAEKEGYEVRILTADRDLYQLVSDRVAVLHPEGHLITPEWLWEKYGLRPEOWVDFRALVGDPSDNLPGVKGI ELVDLLGFTRLEVGYEADDVLATLAKKAEKEGYEVRILTADRDLYQLVSDRVAVLHPEGHLITPEWLWEKYGLRPEQWVDFRALVGDPSDNLPGVKGI ELVDLLLGFTRLEVPGYEADDVLATLAKKAEKEGYEVRILTADRDLYQLVSDRVAVLHPEGHLITPEWLWEKYGLRPEQWVDFRALVGDPSDNLPGVKGI ELVDLLGFTRLEVGGYEADDVLATLAKKAEKEGYEVRILTADRDLYOLVSDRVAVLHPEGHLITPEWLWEKYGLRPEOWVDFRALVGDPSDNLPGVKGI


 GEKTALKLLKEWGSLENLLKNLDRLKP-AIREKILAHMDDLKLSWDLAKVRTDLPLEVDFAKRREPDRERLRAFLERLEFGSLLHEFGLLESSKALEEAP GEKTALKLLKEWGSLENLLKNLDRLKP-AIREKILAHMDDLKLSWDLARVRTDLPLEVDFAGRREPDRERLRAFLERLEFGSLLHEFGLLESPKALEEAP GEKTALKLLKEWGSLENLLKNLDRLKP-AIREKILAHMDDLKLSWDLAKVRTDLPLEVDFAKRREPDRERLRAFLERLEFGSLLHEFGLLESPKALEEAP
 GEKTARKLLEEWGSLEALLKNLDRLKPGAIREKILAHMDDLKLSWDLAKVRTDLPLEVDFAKRREPDRERLRAFLERLEFGSLLHEFGLLESPKALEEAP

 GEKTALKLLKEWGSLENLLKNLDRLKPEAIREKILAHMDDLKLSWDLA.VRTDLPLEVDFAKRREPDRERLRAFLERLEFGSLLHEFGLLESPKALEEAP

Translation of 4C11 Translation of 5B1 Translation of 5B4 Translation of 5D3 Translation of 5D4 Translation of Taq po Translation of Taq pol Translation of Tth pol

Translation of 4C11 Translation of 5B1 Translation of 5B4 Translation of 5D3 Translation of 5D4 Translation of Taq pol Translation of Taq pol
Translation of Tth pol Translation of Tth pol
Translation of Tfl pol

| $M R$ |  |
| :---: | :---: |
| $M$ | 1 |
| $M$ |  |

KELVDLLGFTRLE 120 K ELVDLLGLARLEV KELVDLLGFTRLEV
KELVDLLGLV KELVDLLGFTRLEV

10 MLPLF MAMLPLFEPKGRVLLVDGHHLAYRTFFALKGLTTSRGEPVQAVYGFAKSLLKALKEDGYKAVFVVFDAKAPSFRHEAYEAYKAGRAPTPEDFPRQLALI
 MAMLPLFEPKGRVLLVDGHHLAYRTVFALKGLTTSRGEPVQAVYGFAKSLLKALKEDGYKAVFVVFDAKAPSFRHEAYEAYKAGRAPAPEDFSRQLALI MAMLPLFEPKGRVLLVDGHHLAYRTFFALKGLTTSRGEPVQAVYGFAKSLLKALKEDGYKAIFVVFDAKAPSFRHEABEAYKAGRAPSPEDFPRQLALI GMLPLFEPKGRVLLVDGHHLAYRTFHALKGLTTSRGEPVQAVYGFAKSLLKALKEDGUDAVUVVFDAKAPSFRHEAYGGYKAGRAPTPEDFPRQLALI
$\qquad$
 WPPPEGAFVGFVLSRKEPMWADLLALAAARGGRVHRAPEPYKALRDLKEARGLLAKDLSVLALREGLGLPPGDDPMLLAYLLDPSNTTPEGVARRYGGEV WPPPEGAFVGFVLSRKEPMWADLLALAAARGGRVHRAPEPYKALRDLKEARGLLAKDLSVLALREGLGLPPGDDPMLLAYLLDPSNTTPEGVARRYGGKGU WPPPEGAFVGFVLSRKEPMWADLLALAAARGGRVHRAPEPYKALRDLKEARGLLAKDLSVLALREGLGLPPGDDPMLLAYLLDPSNTTPEVVARRYGGEV WPPPVGAFVGFVLSRKEPMWADLLALAAARGGRVHRAPEPYKALRDLKEARGLLAKDLSVLALREGLGLPPGDDPMLLAYLLDPSNTTPEVVARRYGGEU WPPPEGAFVGFVLSRKEPMWADLLALAAARGGRVHRAPEPYKALRDLKEARGLLAKDLSVLALREGLGLPPGDDPMLLAYLLDPSNTTPEGVARRYGGEV
 WPPPEGAFVGFVLSRKEPMWADLLALAGARGGRVHRAPEPYKALRDLKEARGLLAKDLSVLALREGLGLPPGDDPMLLAYLLDPSNTTPEGVARRYGGEW
 LEAGERAALSERLFANLWGRLEGEERLLWLYREVERPLSAVLAHMEATGVRLDVAYLRALSLEVAEEIARLEAEVFRLAGHPFNLNSRDQLERVLFDEI TEEAEERAALSERLFANLWGRLEGEERLLWLYREVERPLSAVLVHMEATGVRLDVAYLRALSLEVAEEIARLEAEVFRLAGHPFNLNSRDQLERVLFDEL TEEAGERAALSERLFANLWGRLEGEGRLPWLYREVERPLSAVLAHMEATGVRLDVAYLRALSLEVAEEIARLEAEVFRLAGHPFNLNSRDQLERVLFDEL TEEAGERAALSERLFANLWGRLEGEGRLLWLYRGVERPLSAVLAHMEATGVRLDVAYLRALSLEVAEEIARLEAEVFRLAGHPFNLNSRDQLERVLFDEL TEEAGERAALSERLFANLWGRLEGEERLLWLYREVERPLSAVLAHMEATGVRLDVAYLRALSLEVAEEIARLEAEVFRLAGHPFNLNSRDQLERVLFDEL TEDAA HRALLSERLHRNLLKRLEGEEKLLWLYHEVEKPLSRVLAHMEATGVRQDVAYLQALSLELAEEIRRLEEEVFRLAGHPFNLNSRDQLERVLFDEL TEDAGERALLAERLFQTLKERLKGEERLLWLYEEVEKPLSRVLARMEATGVRLDVAYL QALSLEVEAEVRQLEEEVFRLAGHPFNLNSRDOLERVLFDEL

20
30
40
50 TEEAGERAALSERLFANLWGRLEGEERLLWLYREVERPLSAVLAHMEATGVRLDVAYLRALSLEVAEEIARLEAEVFRLAGHPGNLISRDQLERVLFDEL
ranslation of 4C11 Translation of 5B1 Translation of 5B4 Translation of 5D3 Translation of 5D4 Translation of Taq po Translation of Tth pol Translation of Tfl pol

## 

GLPAIGKTEKTGKRSTSAAVLEALREAHPIVEKILQYRELTKLKSTYIDPLPDLIHPRTGRLHTRFNQTATATGRLSSSDPNLQNIPVRTPLGQRIRRAF GLPAIGKTEKTGKRSTSAAVLEALREAHPIVEKILQYRELTKLKSTYIDPLPDLIHPRTGRLHTRFNQTATATGRLSSSDPNLQNIPVRTPLGQRIRRAF GLPAIGKTEKTGKRSTSAAVLEALREAHPIVEKILQYRELTKLKSTYIDPLPDLIHPRTGRLHTRFNQTATATGRLSSSDPNLQNIPVRTPLGQRIRRAF GLPAIGKTEKTGKRSTSAAVLEALREAHPIVEKILQYRELTKLKSTYIDPLPDLIHPRTGRLHTRFNOTATATGRLSSSDPNLONIPVRTPLGORIRRAN GLPAIGKTEKTGKRSTSAAVLEALREAHPIVEKILQYRELTKLKSTYIDPLPDLIHPRTGRLHTRFNQTATATGRLSSSDPNLQNIPVRTPLGQRIRRA GLPAIGKTEKTGKRSTSAAVLEALREAHPIVEKILQYRELTKLKSTYIDPLPDLIHPRTGRLHTRFNQTATATGRLSSSDPNLQNIPVRTPLGQRIRRAF
 GLPAIGKTEKTGKRSTSAAVLEALREAHPIVDRILOYRELTKLKNTYIDPLPALVHPKTGRLHTRFNOTATATGRLSSSDPNLONIPVRTPLGORIRRAI GLPAIGKTEKTGKRSTSAAVLEALREAHPIVEKILQYRELTKLKSTYIDPLPDLIHPRTGRLHTRFNQTATATGRLSSSDPLQNIPVRTPLGQRIRRAF

Translation of 4C11 Translation of 5B1 Translation of 5B4 Translation of 5D3 ranslation of 5D4 Translation of Taq pol Translation of Tth pol Translation of Tfl pol

| $I I$ |
| :--- |
| $I$ |
| $I$ |
| $I$ |
| $I$ |
| $I$ |
| $I$ |
| $I$ |
| $V$ |
| $V$ |
| $V$ | 610 620 630 650 660 670 680



Translation of 4C11 Translation of 5B1 Translation of 5B4 Translation of 5D3 Translation of 5D4 ranslation of Taq pol Translation of Tth pol Translation of Tfl pol

| 710 | 720 | 730 |
| :---: | :---: | :---: |
| FPKVRAWIEKTLEEGRRRGY |  |  |
| SFPKVRVWIEKTLEEGRRRGYVETLFGRRRYVPDL |  |  |
| SFPKVRAWIEKTLEEGRKRGYVETLFGRRRYVPDL |  |  |
| SFPKVRAWIEKTLEEGRRRGYVETLFGRRRYVPDL |  |  |
| SFPKVRAWIEKTLEEGRKRGYVETLFGRRRYVPDLN |  |  |
| SFPKVRAWIEKTLEEGRRRGYVETLFGRRRYVPDLE |  |  |
| SFPKVRAWIEKTLEEGRKRGYVETLFGRRRYVPDL |  |  |
| SYPKVRAWI |  |  |

740
750
760
770
780
790 SFPKVRVWIEKTLEEGRRRGYVETLFGRRRYVPDLEA

RVKREAAERMAFNMPVQGTAADLMKLAMVKLFPRLEETGARMLLQVHDELVLEAPKERAEA
 ARVKSVREAAERMAFNMPVQGTAADLTKLAMVKLFPRLEETGARMLLQUHDELVLEAPKERAEA ARVKSAREAAERMAFNMPVQGTAADLMKLAVVKLFPRLEETGARMLLQVHDELVLEAPKERAEA ARVKSVREAAERMAFNMPVQGTAADLTKLAMVKLFPRLEETGARMLLQVHDELVLEAPKERAEA ARVKSVREAAERMAFNMPVQGTAADLMKLAMVKLFPRLEEMGARMLLQVHDELVLEAPKERAE ARVKSVREAAERMAFNMPVOGTAADLMKLAMVKLFPRLREMGARMLLOVHDELLLEAP OARAE
 ARVKSVREAAERMAFNMPVQGTAADLMKLAMVKLFPRLEETGARMLLQVHDELVLEAPKERAEA

Translation of 4C11
Translation of 5B1 Translation of 5B4 Translation of 5D3 Translation of 5D4 Translation of Taq pol Translation of Tth pol Translation of Tfl pol


840
850


## 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), $\left[\mathrm{d} 5 \mathrm{NIC}_{2}(2\right.$ consecutive d5NIC bases) and an abasic site (tetrahydrofuran ( $\phi$ ) ) 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 $\mathrm{D}_{2} \mathrm{O}$ NOESY spectra ( 250 ms mixing time) of tnic, recorded at $25^{\circ} \mathrm{C}$. Intra-residue pyrimidine $\mathrm{H} 6-\mathrm{H} 1$ ' and purine/d5NIC H8-H1' crosspeaks are labelled with residue name and number, intra-residue pyrimidine $\mathrm{H} 5-\mathrm{H} 6$ 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".
5'-TAATACGACTCACTATAGGGAGA ( ${ }^{\text {dYTP }}$
a)
dYTP=
NP-TP


b)



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-monocarboxamidetriphosphate (PMC-TP) for template $\mathrm{X}=\mathrm{dA}$ (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 $\Delta \Delta \mathrm{G}=1.8 k T^{*}$ were selected to represent the SCA network, mapped to the structure of the Taq polymerase domain $\left(3 \mathrm{KTQ}^{2}\right)$ 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 ( 3 KTQ ) ( $\mathrm{a}, \mathrm{b}$ ) and coloured according to their effect on $\Delta \Delta \mathrm{G}$ : red $=$ stabilising mutations $(\Delta \Delta \mathrm{G}<-1.0$ $\mathrm{kcal} \mathrm{mol}^{-1}$ in most structures, e.g. A608V, I 614 M ), orange $=$ weakly stabilising mutations $\left(-1.0<\Delta \Delta \mathrm{G}<0.0 \mathrm{kcal} \mathrm{mol}^{-1}\right.$, e.g. R715K, G389V), yellow $=$ neutral mutations (e.g. E734N), light blue $=$ potentially destabilising $(0.0<\Delta \Delta \mathrm{G}<1.0 \mathrm{kcal}$ $\mathrm{mol}^{-1}$, e.g. E602G, E424G) and dark blue $=$ destabilising mutations $(\Delta \Delta \mathrm{G}>1.0 \mathrm{kcal}$ $\mathrm{mol}^{-1}$, e.g. M761T, M775T). (c) Predicted change in free energy of folding $(\Delta \Delta \mathrm{G})$ in kcal $\mathrm{mol}^{-1}$ is plotted for individual mutations for two different Taq structures $1 \mathrm{CMV}^{3}$ (Apoform) and $3 \mathrm{KTQ}^{2}$ (closed ternary complex).
1: $5^{\prime}$ - CAGGAAACAGCTATGACAAAAATCTAGATAACGAGGGCAA
$4: 5^{\prime}-$ GTAAAACGACGGCCAGTACCACGGAACTGCGGGTGACGCCAAGCG
5D4 G602E M614I T761M T775M - m
$\underline{\underline{\underline{Z}}}$
$-\rightarrow \sim \rightarrow \infty$,
b)
2:5'-CAGGAAACAGCTATGACAAAAATCTAGATAACGAGGGCA (X)
6:5'-GTAAAACGACGGCCAGTACCAC (X) GAACTGCGGGTGACGCCAAGC ( $\mathbf{x}$ )
$\mathrm{X}=\mathrm{d} 5 \mathrm{NIC}$
5D4 G602E M614I T761M T775M -


## 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, $\phi$ X174 HaeIII digest marker.
$5^{\prime}$ - TAATACGACTCACTATAGGGAGA
3'-ATTATGCTGAGTGATATCCCTCTTTTTTTTTT

dyATP

Taq 5D4


## Supporting information Figure 9: Incorporation of dyATP

Incorporation of the size-expanded dATP analogue dyATP ${ }^{4}$ is shown for both Taq and 5D4. While incorporation by Taq is poor and stalls at $+1,5 \mathrm{D} 4$ 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 $^{\mathrm{a}}$ | 4C11 | 5B1 | 5B4 | 5D3 | 5D4 |
| :--- | :---: | :---: | :---: | :---: | :---: |
| E602G | + | + | + | + | + |
| A608V | + | + | + | + | + |
| I614M | + | + | + | + | + |
| M762T |  |  | + |  | + |
| M775T | + | + | + | + | + |

Supporting Table 2: $K_{m}(\mu \mathbf{M})$ and $k_{\text {cat }} / K_{m}$ values $\left(\% \min ^{-1} \mu M^{-1}\right)$ for Taq and 5D4 (Fig. 2a)

| dXTP | Template base | $\begin{aligned} & \hline \mathrm{K}_{\mathrm{m}} \\ & \mathrm{wtTaq} \end{aligned}$ | $\begin{aligned} & \hline \mathrm{k}_{\mathrm{cat}} / \mathrm{K}_{\mathrm{m}} \\ & \mathrm{wTaq} \\ & \hline \end{aligned}$ | $\begin{array}{\|l\|} \hline \mathrm{K}_{\mathrm{m}} \\ \text { 5D4 } \\ \hline \end{array}$ | $\begin{aligned} & \mathrm{k}_{\mathrm{cal}} / \mathrm{K}_{\mathrm{m}} \\ & \text { 5D4 } \end{aligned}$ | $f$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| 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: $f=\mathrm{k}_{\mathrm{cal}} / \mathrm{K}_{\mathrm{m}} 5 \mathrm{D} 4 / \mathrm{k}_{\mathrm{ca}} / \mathrm{K}_{\mathrm{m}} \mathrm{wtTaq}$

Supporting Table 3: Steady-state extension constants $\mathbf{k}_{\text {cat }}$ of 3' d5NI by 5D4 $\left(\% \min ^{-1} \mu \mathrm{M}^{-1}\right)$

| 3' base | Template <br> N | template <br> $\mathrm{N}+1$ | $\mathrm{k}_{\text {cat }} / \mathrm{K}_{\mathrm{m}}$ |
| :--- | :--- | :--- | :--- |
| 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
$\left.\begin{array}{|l|c|c|}\hline \text { Experimental constraints }^{\mathbf{1}} & \text { tni } & \text { tnic } \\ \hline \text { NOE distance constraints } & 228 & 236 \\ \hline \text { Hydrogen bonding distance constraints } & 6 & 8 \\ \hline \text { Dihedral constraints } & 50 & 48 \\ \hline \text { Refinement Analysis } & \text { tni } & \text { tnic } \\ \hline \begin{array}{l}\text { Distance constraint violation energy } \\ \text { (kcal mol }\end{array} \\ \hline \begin{array}{l}\text { Dihedral constraint violation energy } \\ \\ \text { (kcal mol }\end{array} \\ \hline \text { (1) }\end{array}\right)$
${ }^{1}$ for residues T1-C6 and G13-d5NI(C) 15 (see Methodology)
${ }^{2}$ Force constant $\mathrm{K}=50 \mathrm{Kcal} \mathrm{mol}^{-1} \AA^{-2}$
${ }^{3}$ Force constant $\mathrm{K}=0.01 \mathrm{Kcal} \mathrm{mol}^{-1} \mathrm{deg}^{-2}$

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

| Mutation | Structure <br> 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 $\mathrm{Taq}\left(1 \mathrm{CMW}^{3}\right)$ and Klentaq fragment $\left(1 \mathrm{KTQ}^{2}\right)$, the binary complex of polymerase bound to DNA (4KTQ ${ }^{2}$ ), and both open $\left(2 \mathrm{KTQ}^{2}\right)$ and closed $\left(3 \mathrm{KTQ}^{2}\right)$ ternary complexes of Taq polymerase with DNA and incoming nucleotide. Energy is given as $\mathrm{kcal} \mathrm{mol}^{-1}$. Residues not present in the structure are shown as N.S. .

## 3. List of primers ( $\mathrm{X}=\mathbf{d 5 N I C}, \mathbf{Y}=\mathbf{d 5 N I}$ )

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 GCX -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 GCX-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 GCY-3’

8-10: primers $1-3$ as described ${ }^{5}$

11-13: primer $1,2,5$ as described ${ }^{6}$

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 TTB TTC GTG GTC GCG ACG GAT GCC X-3 $\mathbf{N}=\mathbf{d A}, \mathbf{d G}, \mathbf{d C}, \mathbf{d T} ; \mathbf{B}=\mathbf{d U}$-biotin

20-23: 5'-AGC TAC CAT GCC TGC ACG CAG CNG GCA TCC GTC GCG ACC ACG TTB TTC GTG GTC GCG ACG GAT GCC Y-3' $\mathbf{N}=\mathbf{d A}, \mathbf{d G}, \mathbf{d C}, \mathbf{d T} ; \mathbf{B}=\mathbf{d U}$-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. $5^{\prime}$ - 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 $(\phi)$, IN, BI, ICS, 7AI
26. 5'-TAG CTA XTC TCC CTA TAG TGA GTC GTA TTA-3'
27. $5^{\prime}$-TAG CTA YTC TCC CTA TAG TGA GTC GTA TTA-3'
28. 5'-TAA TAC GAC TCA CTA TAG GGA GAZ-3'
$\mathbf{Z}=\mathbf{I C S}, 7 \mathrm{AI}$
29. $5^{\prime}$-NNT CTC CCT ATA GTG AGT CGT ATT A-3'
$\mathbf{N}=\mathbf{d A}, \mathbf{d T}, \mathbf{d C}, \mathbf{d G}$
30. 5'-TTT TTT TTT TZT CTC CCT ATA GTG AGT CGT ATT A-3'
$\mathbf{Z}=$ abasic site $(\phi)$

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: $5^{\prime}$ - 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.
