**MANUSCRIPT TITLE**

# RAMA: an Integrative Approach to Identify Hot Spots and Guide the Directed Evolution of High Molecular Weight Enzyme

Yihao Chen[a], Zhou Gong[a,b], Xiaoling Zhao[c,d], Mingjun Zhu[a,b], Ali Raza[a,b], Zhiqing Tao[a,b], Wenhui Li[a,b], Xu Zhang[a,b], Maili Liu[a,b,e], Lichun He[*a,b]

a. State Key Laboratory of Magnetic Resonance and Atomic Molecular Physics, National Center for Magnetic Resonance in Wuhan, Innovation Academy for Precision Measurement Science and Technology, Chinese Academy of Sciences, Wuhan 430071, China.
b. University of Chinese Academy of Sciences, Beijing, 100049, China.
c. Department of Reproductive Medicine, General Hospital of Central Theater Command of the People's Liberation Army, Wuhan, Hubei 430061, China.
d. Qinhe Life Science Ltd. Wuhan 430000, China.
e. Optics Valley Laboratory, Hubei 430074, China.

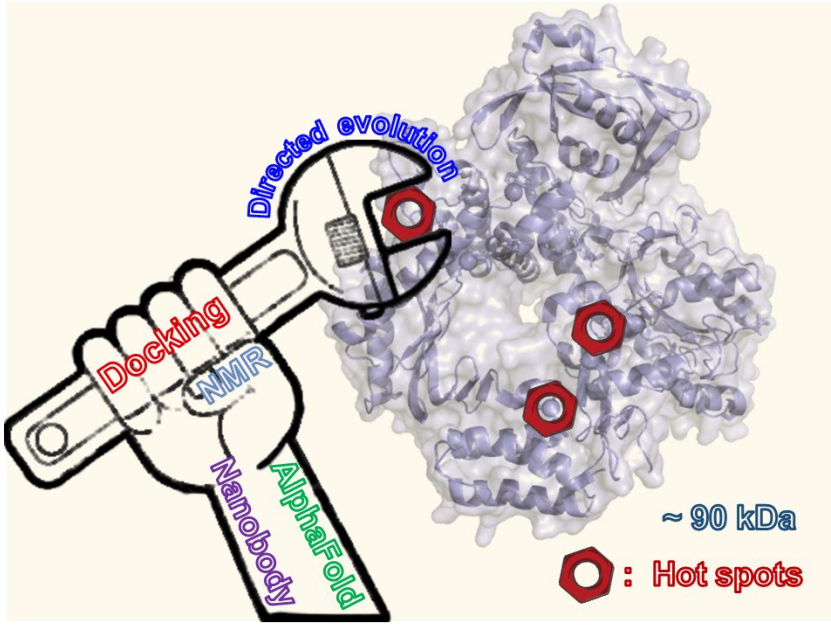
* Corresponding author:

Lichun He

**E-mail**: helichun@apm.ac.cn

**ABSTRACT**

Directed evolution is a robust method for engineering enzymes by iteratively creating and screening variant libraries. However, the challenge of developing assays for high-throughput screening persists, especially for high molecular weight enzymes, which are constrained by the vast number of possible amino acid sequences. In this study, we introduce an integrative approach named RAMA to reduce the search space in the directed evolution of enzymes. $^{1}H$-$^{13}C$ NMR spectra were applied to record the titration of a 90 kDa B-type DNA polymerase *Pyrococcus furiosus* DNA polymerase (*Pfu* pol) with neutralization nanobodies and substrates to recognize 'hot spots' involved in its activity. Communal peaks with chemical shift perturbation were analyzed to determine their amino acid types. Following by the molecular docking of *Pfu* pol-nanobody and *Pfu* pol-substrate complexes, these hot spots were assigned and used for the directed evolution of the *Pfu* pol. Saturation mutagenesis of four identified hot spots of *Pfu* pol were performed to screen polymerase with enhanced catalytical activity. Biochemical and structural analysis of three preferred *Pfu* pol variants reveal distinct molecular mechanism for their improved performances, indicating a powerful capability of RAMA in the directed evolution of *Pfu* pol. This proposed approach is also applicable to other high molecular weight enzymes, provided appropriate antibodies are available. Thus, its use will speed up the design and engineer enzymes for many bio-chemical transformations.




**Keywords**

Directed enzyme evolution; hot spots; NMR Spectroscopy, residue assignment

## 1. INTRODUCTION

Directed enzyme evolution is a powerful technique for enhancing various properties of biocatalysts, such as stability, catalytic activity, and selectivity, among others [1-3]. Many methods are available to generate an astronomical number of enzyme variants with relative ease [4, 5]. However, since each amino acid can be substituted with 19 alternatives, constructing a comprehensive library to cover the entire mutational space of a high molecular weight enzyme is practically unfeasible [6]. While state-of-the-art machine learning (ML) approaches have transformed numerous fields [7], yet the availability of adequate data remains a significant challenge for bioinformatics and machine learning-assisted directed evolution. The performance of these advanced computational techniques is highly dependent on access to large, high-quality datasets. In scenarios where only limited data are available, the efficacy and reliability of ML models can be substantially compromised. Additionally, high-throughput phenotypic screening also poses a significant challenge due to the diverse substrates and products associated with biocatalysts. Given the challenges of increasing screening efficiency and throughput, it was inevitable that, from the early days of directed evolution, significant effort was devoted to narrowing the sequence search space to make experimental evaluation more practical. Reshaping the active site to modify specificity and broaden substrate compatibility, structure-guided evolutionary approaches are particularly effective for directed evolution of relatively rigid proteins [8]. However, enzymes are inherently multifunctional systems where catalytic behavior depends not only on static architecture but also on functional interactions that extend beyond the active site [9-11]. Such functional features, including transient chemical states or residue-specific interactions, are often undetectable by conventional structural analysis. Structure-based methods alone face challenges in identifying mutation sites that indirectly influence catalysis through mechanisms unrelated to spatial proximity, limiting their utility in comprehensive enzyme engineering [12].

In this context, NMR spectroscopy offers a complementary strategy by enabling direct analysis of catalytic parameters such as hydrogen bonding networks, protonation states, or substrate-induced conformational shifts without relying on prior structural data. This technique can resolve residue-specific chemical environments in soluble proteins of up to mega-Dalton sizes under catalytically relevant conditions [13], thereby bypassing the need for structural models to pinpoint functionally critical regions. In particular, 2D [$^{13}$C,$^{1}$H]-Heteronuclear Multi-Quantum Coherence (HMQC) experiments offer a powerful tool to probe the chemical environment of hydrophobic cores and functional sites, making them particularly suitable for capturing key functional hot spots in directed evolution studies. This is because the methyl groups of hydrophobic residues are highly sensitive to local conformational changes and can reveal subtle perturbations in protein dynamics or allosteric networks, even in large molecular systems. Specifically, the 2D [$^{13}$C,$^{1}$H]-HMQC experiment provides a powerful tool for probing the chemical environment of hydrophobic cores and functional sites, making it particularly suitable for capturing key functional hot spots in directed evolution studies. This is because the methyl groups of hydrophobic residues are highly sensitive to local conformational changes, allowing for the detection of subtle perturbations in protein dynamics or allosteric networks, even in macromolecular systems. Nevertheless, the assignment of these hot spots is challenging, especially for high molecular weight enzymes, where fast transverse relaxation rates make scalar magnetization transfer inefficient. Additionally, resonance overlap in large proteins complicates the interpretation of multiple states with similar chemical shifts [14]. Consequently, there is an urgent need for more effective assignment methods to facilitate NMR analysis of hot spots of high molecular weight enzymes.

Here, we presented an integrative approach (RAMA) to **R**ecognize and **A**ssign hot spots of enzymes following by sited directed **M**utagenesis and **A**ctivity-based screening for the directed evolution of the

enzyme. The high-fidelity DNA polymerase from the thermophilic archaeon *Pyrococcus furiosus* with molecular weight ~90 kDa was applied [15]. *Pyrococcus furiosus* DNA polymerase (*Pfu* pol) is widely used in molecular biology, sequencing, and synthetic biology applications based on PCR technology [16-18]. The NMR spectroscopy and computational methods were combined to identify residues (hot spots) involved in the catalytical activity of *Pfu* pol, either around or far away from the active sites to reduce the sequence search space to improve screening efficiency. As a result, saturation mutations of three out of four identified hot spots were showing improved performance of *Pfu* pol in distinct biocatalytic properties. These results demonstrated the feasibility and effectiveness of RAMA approach in directed evolution of enzymes, paving the way for new fundamental studies on enzyme functions and evolution

## MATERIAL AND METHODS

### 2.1 Cloning, expression, and purification of proteins

Genes of *Pyrococcus furiosus* DNA polymerase (*Pfu* pol), *Pfu* pol-Sso7d polymerase (Pfu-S) were codon-optimized and synthesized by Sangon Biotech Co., Ltd. and cloned in pET16b vector (Novagen) containing an N-terminal His-tag. The *Pfu* pol mutants were prepared by site-directed mutagenesis PCR (see 2.5 Library construction). The genes of nanobodies NB2 and NB4 were cloned in a modified pET22b vector containing N-terminal His-tag. The respective plasmid was transformed into BL21 (DE3) competent cells (Novagen), grown in Luria Bertani (LB) broth or M9 medium supplemented with 100 μg/ml ampicillin and induced for expression with a final concentration of 0.5 mM isopropyl-D-1-thiogalactopyranoside (IPTG) at an optical density (600 nm) of 0.6 for 6 hr at 37 °C. Uniformly [$^{13}$C]-labeled protein was prepared by growing cells in M9 medium containing [U-$^{13}$C] glucose (3 g/liter) [19]. Cells were harvested by centrifugation at 5000g for 25 min at 4 °C.

For *Pfu* pol or its mutants, the pellet was resuspended in 50 ml of Lysis Buffer A (pH 9.0, 50 mM Tris-HCl, 500 mM KCl, 0.1 mM PMSF and 0.01 mg/ml DNase I). For nanobody NB2 or NB4, the cell pellet was resuspended in 50 ml of Lysis Buffer C (pH 7.4, 20 mM NaP, 150 mM NaCl, 0.1 mM DTT, 0.1 mM PMSF and 0.01 mg/ml DNase I). Afterwards, cells were lysed by a high-pressure homogenizer and the pellet was separated by centrifugation at 27,000×g for 30 min at 4 °C. For the purification of *Pfu* pol, the collected supernatant was incubated for 30 minutes at 70 °C to precipitate cellular proteins of *E. coli*, followed by one more round of centrifugation at 27,000×g for 30 min. The supernatant containing each target protein was loaded onto a Ni-NTA (nitrilotriacetic acid) column (Qiagen) which was pre-equilibrated with the corresponding lysis buffer. The column was washed with 100 ml of wash buffer (each respective Lysis Buffer with 30 mM imidazole). *Pfu* pol or its mutants was eluted with 25 ml Elution Buffer A (pH 8.0, 50 mM Tris-HCl, 50 mM KCl, 300 mM imidazole). NB2 or NB4 was eluted with 25 ml Elution Buffer C (pH 7.4, 20 mM NaP, 150 mM NaCl, 0.1 mM DTT, 300 mM imidazole). For *Pfu* pol or its mutants, the eluted fraction was then loaded onto a 5 ml HiTrap Heparin column (GE Bioscience) that was pre-equilibrated with Heparin Buffer (pH 8.0, 50 mM Tris-HCl, 50 mM KCl). *Pfu* pol or its mutants were eluted with a 0-500 mM KCl gradient (pH 8.0, 50 mM Tris-HCl).

The eluted protein was immediately buffer exchanged *via* a PD-10 column (GE Bioscience) to the Storage Buffer (pH 7.4, 25mM Tris-HCl, 50 mM KCl, 0.1 mM EDTA, 1 mM DTT, 50% glycerol) and stored at -20 °C. For nanobody NB2 or NB4, the eluted fraction was then fractionated on a size exclusion Superdex-75 column (GE Bioscience) equilibrated in SEC Buffer B (pH 7.4, 20 mM NaP, 150 mM NaCl, 0.1 mM DTT, 0.1 mM EDTA) to further purify the proteins. The proteins (TkHSP20, NB2 and NB4) were concentrated and store at -80 °C for further usage.

**2.2 NMR spectroscopy**

All NMR experiments of the *Pfu* pol were performed in the NMR buffer (20 mM NaP, 35 mM NaCl, 0.02% $NaN_3$, 8% $D_2O$, pH 7.4). If not specified, all spectra were collected at 298K on Bruker Avance 600 MHz NMR spectrometer equipped with the cryogenically cooled probe. 2D [$^{13}$C,$^{1}$H]-HMQC spectra of 50 μM $^{13}$C labeled *Pfu* pol were recorded in the absence and presence of 50 μM hairpin DNA or 75 μM NB2 or 75 μM NB4 to monitor the chemical shift change for identifying the binding sites (hot spots). Furthermore, 2D [$^{13}$C,$^{1}$H]-HMQC spectra of 50 μM $^{13}$C labeled *Pfu* pol in the presence of 1 M betaine or 0.05% Tween-20 were recorded to measure whether and where these additives interacted with the *Pfu* pol. For all NMR spectra, 1024 complex points were recorded in the direct dimension and 154 complex points were recorded in the indirect dimension. Shifted sine bell and Polynomial baseline correction were applied in all dimensions. NMR data were processed and analyzed in CCPNMR Analysis [20].

**2.3 Endpoint PCR**

20 μl reactions were prepared by mixing 2× buffer (1× Pfu DNA polymerase reaction buffer: pH 8.8, 50 mM Tris-HCl, 10 mM KCl, 6 mM ammonium sulfate, 2 mM $MgCl_2$, 0.05% Triton X-100, 0.001% BSA; 1×Pfu DNA polymerase reaction buffer: pH 8.8, 50 mM Tris-HCl, 100 mM KCl, 2 mM $MgCl_2$, 0.05% Triton X-100, 0.001% BSA) with 50 nM *Pfu* pol or its mutants, 4 ng (plasmid DNA) or 50 ng (Alpaca cDNA) or 100 ng (Mouse tail gDNA) or 40 ng (cotton gDNA) or 45 ng (rice cDNA) templates, 2 μM of each primer (see Table S1), 0.2 mM dNTP and varying amounts of additives (see figure legends). Amplification reactions were performed using a thermos cycler (C1000 Touch, Bio-rad): First, incubation at 95 ℃ for 3 min; For *Pfu* pol and its mutants, followed by 33 thermal cycles of 95 ℃ for 30 s, 55 ℃ for 30 s, and 72℃ for 45 sec to 10 min depending on the target DNA length(1 kb/min); Afterwards, 4 μl of the PCR product was mixed with the loading dye and applied onto a 1% or 0.6% agarose-TAE gel containing 1× GenRed nucleic acid gel stain (Genview). After electrophoresis, the gel was photographed under ultraviolet light. The gels were analyzed by ImageJ to qualify the intensity of PCR product bands. Plots of the gel band intensity were generated with the software GraphPad Prism 8.

**2.4 Assignments of hot spots**

The peaks from the 2D [$^{13}$C,$^{1}$H]-HMQC spectra of *Pfu* pol, showing common chemical shift perturbation upon the titration of hairpin DNA, nanobody NB2 and NB4 individually were identified as the hot spots. To assign the hot spots, we first compared the chemical shifts of hot spots with the statistics values of amino acids from the Biological Magnetic Resonance Bank (BMRB) database to calculate the possibility of these hot spots belonging to a certain amino acid type. The probability density function *f* (a,b) for a hot spots residue's belongs to a certain amino acid types is given by:

$$f(\mathrm{a},\mathrm{b}) = \frac{1}{2\pi\sigma_H\sigma_C} e^{-\frac{1}{2}\left[\left(\frac{a-\mu_H}{\sigma_H}\right)^2 + \left(\frac{b-\mu_C}{\sigma_C}\right)^2\right]} \tag{1}$$

where a is the chemical shift of the hot spot residue in the H dimension, b is the chemical shift of the hot spot residue in the C dimension, $\mu_H$ is the mean chemical shift in the H dimension for the amino acid type, $\sigma_H$ is the standard deviation of the chemical shift in the H dimension for the amino acid type, $\mu_C$ is the mean chemical shift in the C dimension for the amino acid type, $\sigma_C$ is the standard deviation of the chemical shift in the C dimension for the amino acid type.

Using Bayes' theorem, the probabilities that the hot spot belongs to certain amino acid type are:

$$P_x(\mathrm{a},\mathrm{b}) = \frac{f_x(\mathrm{a},\mathrm{b})}{\sum f(\mathrm{a},\mathrm{b})} \tag{2}$$

Where $f_x$(a, b) is the probability density value for the hot spot residue belonging to amino acid type *x*, $\sum f(\mathrm{a},\mathrm{b})$ is the total probability density values for all possible amino acid types of the hot spot residue. With the possibility of the hot spot belonging to which amino acid, we then applied the integrated methods with AlphaFold3 and HADDOCK2.4 to assign the hot spot to the exact residue. The structure of *Pfu* pol-hairpin DNA complex was generated by AlphaFold3 [21]. AlphaFold3 was executed *via* the public web server (https://www.alphafoldserver.com). Based on AlphaFold's residue-level confidence scores (pLDDT) and interface pLDDT (I-pLDDT) scores, the top 5 models were selected for later usage. The neutralization nanobodies NB2 and NB4 [24] were docked with *Pfu* pol *via* HADDOCK with a previously described protocol [22]. The DNA molecule in the crystal structure of *Pfu* pol [43] was removed prior of the docking. 100 docking models for each *Pfu* pol-nanobody complex were generated. According to haddock-score, the ensemble of 10 conformations with the lowest energy were selected. The distances between atoms of the *Pfu* pol with atoms of the nanobody NB2 or NB4 were measured. Contact residues were defined as those with any atoms within 4-5 Å of the binding partner [45-46]. Residues in the vicinity of the contact interface may also cause chemical shift perturbations (CSPs) in the NMR spectroscopy. Therefore, hot spots are the residues in the vicinity with all hairpin DNA, nanobody NB2 and NB4. The cutoff of a distance 6 Å was applied as the vicinity for *Pfu* pol with its interaction partner, which was calculated based on the averaged distance of hairpin DNA with 2 sequentially forward and backward of the contact residue of *Pfu* pol. With this constraint, we achieved

the unambiguous assignment of three hot spots out of four. Exception could occur while the chemical shift perturbation was caused by the conformational change of *Pfu* pol.

### 2.5 Library construction

Saturation mutagenesis library of *Pfu* pol was created via site-directed PCR with wild-type *P.furiosus* DNA polymerase plasmid cloned in pET16b vector as the template. Degenerate primer pairs (containing an NNK codon at the position to be mutated) were designed to mutate residues at positions 385, 680 and 684. Primer sequences are provided in Supplementary Table S1. Site-directed mutagenesis was done with FasHifi Super DNA polymerase (Swiss Affinibody Lifescience Co., Ltd.) using standard protocol. After the PCR step, the template was digested with Dpn I restriction enzyme and then inactivated at 80 °C, followed by transformation into *E.coli* BL21 (DE3) competent cells (Novagen).

### 2.6 Library Screening

*E.coli* BL21 cells containing *Pfu* pol mutants were plated on LB agar with 100 mg/ml ampicillin. Single colony was picked and inoculated into LB (10 mL) containing ampicillin in 50 ml tubes. Cultures were incubated at 37 °C induced for expression with a final concentration of 0.3 mM IPTG at an optical density (600 nm) of 0.8 for 3 hours at 220 rpm. Cells were harvested by centrifugation at 5000g for 10 min at 4 °C. Pellets were resuspended in $ddH_2O$, centrifuged again, and the supernatant was discarded. The cells were resuspended in $ddH_2O$, and their $OD_{600}$ values were measured. The cells were then diluted stepwise into $ddH_2O$, and different DNA substrates (with target length 250 bp and 10 kbp) were amplified to determine the appropriate amounts of cells expressing Pfu polymerase for subsequent experiments (Figure. S9). We found that cells with an $OD_{600}$ value of 0.33 showed the highest enzyme activity and used this for subsequent screenings. Specifically, first, incubation at 98 ℃ for 6 min to lyse cells for releasing the *Pfu* pol proteins. Then, followed by 34 thermal cycles of 98 ℃ for 15 s, 60 ℃ for 15 s, and 72 ℃ for 20 sec to 5 min depending on the target DNA length (2 kb/min). Afterwards, 4 μl of the PCR product was mixed with the loading dye and applied onto a 1% agarose-TAE gel containing 1× GenRed nucleic acid gel stain (Genview). After electrophoresis, the gel was photographed under ultraviolet light. The gels were analyzed by ImageJ to qualify the intensity of PCR product bands. The band intensities of the wild-type were used for normalization, and the relative activities of the mutants were calculated. The heatmap of relative activities was generated using the software GraphPad Prism 8. Plasmids extracted from the colonies demonstrating improved activity were sequenced by Sanger sequencing (Sangon Biotech Co., Ltd.) to determine the identities of beneficial mutations.

### 2.7 Steady-state kinetic analyses

Steady-state kinetic data were collected and analyzed using the EvaGreen-based fluorometric polymerase activity assay [23]. A hairpin DNA (sequence in Table S1, 100 μM in water) was prepared

in advance by heating at 98℃ for 5 min and annealing on the ice for 30 min. Different amounts of hairpin DNA from 0.1-2 μM were then used as the template to mixed with dNTPs, 10 nM *Pfu* pol or its mutants respectively. $MgCl_2$ was added to initiate the DNA synthesis at 72℃. The reaction buffer contained 25 mM Tris-HCl pH 8.8, 10 mM KCl, 2.5 mM $MgCl_2$, 200 μM dNTPs, and 0.05% Triton X-100. The fluorescence curve was plotted against the time, where the value of the fluorescence was derived by subtracting the background fluorescence value of each reaction. The initial rate of each reaction with different amounts of template was obtained by calculating the first derivative of the fluorescence curve. The kinetics parameters of $k_{cat}$ and $K_m$ were generated by fitting the Michaelis-Menten kinetic equation with the GraphPad Prism 8.0 software. The reported values are the average of a biological duplicate each with three technical replicates.

## 2. RESULTS

### 2.1 Recognition of hot spots of Enzymes

Here we combine four independent steps into a consensus method called RAMA, which encompasses **R**ecognition of hot spots, **A**ssignment of hot spots, site directed saturation **M**utagenesis and **A**ctivity based screening (Figure 1). A multi-domain *Pfu* polymerase, with molecular weight ~90 kDa, was applied for directed evolution with the RAMA protocol.

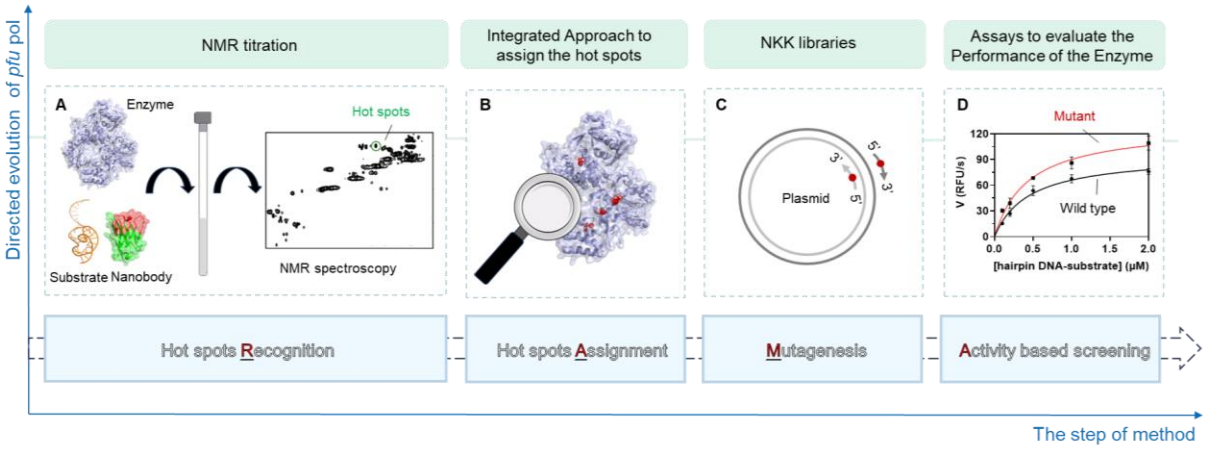



**Figure 1.** Schematic representation of the RAMA process used to identify hot spots and direct the evolution of enzymes. (**A**) Recognition of hot spots: Multiple two-dimensional NMR spectra of the target enzyme were recorded including the apo-state and holo-states of the enzyme titrated with its substrate and neutralization nanobodies respectively. Communal NMR peaks of the enzyme with chemical shift perturbation in different holo-states titrated with its substrate or neutralization nanobodies were identified as hot spots. (**B**) Assignment of hot spots: Computational and experimental data were integrated to assign detected hot spots of the enzyme. (**C**) Site directed saturation Mutagenesis:

Site directed saturation mutagenesis was created with the NNK-based randomization at the assigned hot spots respectively. (**D**) Activity-based Screening: Reaction catalyzed by the targeted enzyme in one or more different selective conditions were conducted to evaluate the performance of mutations.

To overcome the size limitation of protein NMR, the side-chain methyl groups of the *Pyrococcus furiosus* DNA polymerase (*Pfu* pol) were measured, as methyl groups offer several advantages for protein NMR spectroscopy. Firstly, the terminal end position of methyl groups of each amino acid enables the methyl groups to exhibit a local 'tumbling' motion faster than the overall tumbling of the biomolecule, resulting in sharper and more resolved NMR signals. Secondly, methyl groups possess three protons which effectively triples the initial magnetization, resulting in a stronger signal. Moreover, the methyl groups are distributed throughout the tertiary structure of enzymes and have the high sensitivity to their local chemical environment, making them possible to detect very subtle structural changes in enzymes during the catalytic process, such as domain movements, conformational changes, and side-chain rotations. Compared to 2D [$^{15}$N,$^{1}$H]-Heteronuclear Single Quantum Coherence (HSQC), which primarily provides information on backbone assignment and secondary structure, the 2D [$^{13}$C,$^{1}$H]-Heteronuclear Multi-Quantum Coherence (HMQC) spectrum offers complementary insights into the chemical environments of hydrophobic cores and functional sites of proteins, making it particularly suitable for identifying critical functional hot spots during directed evolution.

Following this, 2D [$^{13}$C,$^{1}$H]-HMQC spectra of [$^{13}$C]-labeled *Pfu* pol protein was recorded in the absence and presence of a hairpin DNA template (See Table S1) to identify the potential hot spots of the *Pfu* pol (Figure S2A). Substantial chemical shift perturbations (CSPs) were observed for five aliphatic and aromatic residues (Figure. 2, Figure. S2), named ‘ali1’, ‘aro1’, ‘aro2’, ‘aro3’, and ‘aro4’ (Figure. 2B, Figure. S2A), when comparing the HMQC spectra of [$^{13}$C]-labeled *Pfu* pol protein in the absence and presence of the hairpin DNA template. As the chemical shift perturbation may arise from the binding sites and the allosteric conformational changed regions which may not directly involved in the catalytic activity of *Pfu* pol, we then employed two in-house prepared *Pfu* pol neutralization nanobody NB2 and NB4 [24] (Figure S1) to narrow down hot spots of the *Pfu* pol. Adding both unlabeled NB2 and NB4 nanobodies into the $^{13}$C labeled *Pfu* pol resulted in significant CSPs of methyl group resonance respectively (Figure. S2B and C). The presence of the nanobody NB2 induced notable chemical shift perturbations of the peaks ali1, aro1, aro3, and aro4 (Figure. 2C), while the presence of NB4 caused pronounced chemical shift perturbations of the peaks ali1, aro2, aro3, and aro4 (Figure. 2D). Residues ali1, aro1, aro2, and aro4 revealed communal chemical shift perturbation while adding the hairpin DNA template and neutralization nanobody NB2 or NB4. Thus, these four residues were identified as hot spots to probe the activity related residues of *Pfu* pol subsequent experiments. Whereas the peak ali2 residue did not show chemical shift perturbation upon adding both nanobody NB2 and NB4, indicating it might arise from the conformation changes of the *Pfu* pol upon adding the hairpin DNA. To further

confirm the recognized hot spots are associated with the activity of the *Pfu* pol, a well-known PCR additives, betaine [25, 26], was investigated if they acted on hot spots of *Pfu* pol. The addition of 0.5-2.5 M Betaine (N,N,N-trimethylglycine) into the *Pfu* pol reaction system could enhance its performance in amplifying different templates with varied length (Figure.S3) respectively. Meanwhile, significant chemical shift perturbations of *Pfu* pol were observed on the hot spot residues aro1, aro2, aro3, and aro4 upon the addition of 1 M betaine, while slight chemical shift perturbations of residues aro5 and aro6 (Figure. 2E, Figure. S4) were detected. The results showed that all four hot spots are indeed related to the performance of *Pfu* pol, paving the way for the following steps.

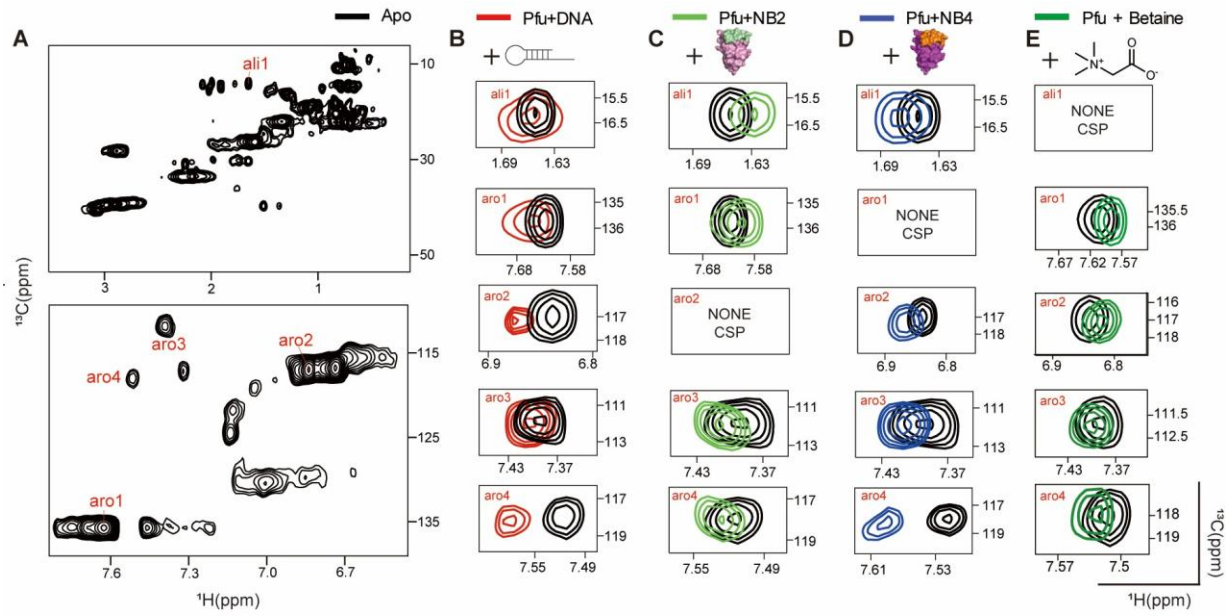



**Figure 2.** Recognition of hot spots of *Pfu* pol by NMR spectroscopy. (**A**) Two-dimensional [$^{13}$C,$^{1}$H]-HMQC spectra of 50 μM [$^{13}$C]-labeled *Pfu* pol were acquired on the 600 MHz spectrometer at 298 K, where the hot spots are labeled in red (ali1, aro1, aro2, aro3, and aro4). (**B**-**E**) Enlarged view of NMR peaks of hot spot residues, which show chemical shift perturbations upon adding 50 μM hairpin DNA (**B**, red), 75 μM NB2 (**C**, pale green), 75 μM NB4 (**D**, blue), 1M betaine (**E**, green).

**Assignments of hot spots of *Pfu* pol**

The terminal methyl groups of aliphatic amino acid side chains exhibit relatively dispersed carbon chemical shifts, except for Val and Leu. As shown in Figure 3A, based on the $^{1}$H and $^{13}$C chemical shift, the peaks of methyl groups in the 2D [$^{13}$C,$^{1}$H]- HMQC spectrum can be assigned to no more than three amino acid types, except threonine. In contrast, aromatic amino acids have a broader $^{13}$C chemical shift distribution, significantly reducing peak overlap. In the 2D [$^{13}$C,$^{1}$H]-HMQC spectrum, the methyl peaks of aromatic amino acid can be assigned to no more than two amino acid types, except histidine. Furthermore, only the $CH_2$ groups of isoleucine (Ile) exhibit minor spectral overlap with the $CH_3$ groups

of leucine (Leu) and valine (Val). Since the signal intensity of methyl groups ($CH_3$) is usually higher than that of $CH_2$ groups, and methyl ($CH_3$), methylene ($CH_2$), and methine (CH) groups can be unambiguously distinguished *via* 2D [$^{13}$C,$^{1}$H]-HMQC integrated with the Distortionless Enhancement by Polarization Transfer (DEPT) [47]. With these advantages, further analysis of the $^{1}$H and $^{13}$C chemical shifts of these hot spots allows us to narrow down the type(s) of amino acids that hot spots belong. We first compared the chemical shifts of the hot spots with the chemical shift data in the Biological Magnetic Resonance Bank (BMRB) [27]. In Figure 3A, we plotted the $^{1}$H atom and $^{13}$C atom chemical shift range map of each amino acids in a rectangle with center on the mean value of the chemical shift of the amino acid and width and length equals to 2 times of the value of the standard deviations of $^{1}$H or $^{13}$C chemical shift in each dimension respectively. The chemical shift range map of each amino acid was plotted in different color. The 2D [$^{13}$C,$^{1}$H]- HMQC spectra of [$^{13}$C]-labeled *Pfu* pol protein was overlaid with this chemical shift range map (Figure 3A). The likelihood of each hot spot belongs to which amino acid was calculated with the formula in the method section. Ali1 may correspond to either the Cβ of alanine with a possibility of 78.8% or the Cε of methionine with a possibility of 21.2%. Aro2 has a possibility of 70.3%% to be Cε of tyrosine, a possibility of 25.3% to be Cζ of tryptophan, a possibility of 4.4% to be Cδ of histidine. While aro3 has a possibility of 95.9% to be Cζ of tryptophan, a possibility of 4.1% to be Cδ of histidine, and aro4 has a possibility of 85.1% to be Cε of tryptophan, a possibility of 14.9% to be Cδ of histidine. Aro1 could be determined to be Cε of histidine based on its unique chemical shift (Figure 3A). To further unambiguously assign these hot spots to certain residue, we introduce an integrated approach with the experimental data and computational methods. Firstly, AlphaFold3 was employed to generate the structure of the complex between the experimental hairpin DNA and *Pfu* pol. The structure with lowest energy was used to compare to the crystal structure of *Pfu* pol with double strand DNA [43], the contact residue positions of the polymerase domain (positions 369 to 775) in *Pfu* pol with hairpin DNA were 95.6 % similar with these from the crystal structure of *Pfu* pol- double strand DNA complex [43] (Figure S5), indicating a high confidence of the complex structure generated by AlphaFold3. Structures of nanobodies NB2 and NB4 were generated by AlphaFold. The software Haddock was then used to generate the complex structure of NB2 and NB4 with *Pfu* pol [22]. The top ten lowest-energy structures of the complex of NB2 and NB4 with *Pfu* pol were selected respectively. The lowest distance of each residue of *Pfu* pol with the surface of nanobody NB2 and NB4 were calculated for each conformer and plotted in Figure 3B. Residue 149, residue 380, residue 615 and residue 680 of *Pfu* pol revealed a short distance less than 6 Å with all of hairpin DNA, nanobody NB2, NB4 conformers. Therefore, we took Residue 149, residue 380, residue 615 and residue 680 of *Pfu* pol as the common contact residues between *Pfu* pol with hairpin DNA, nanobody NB2 and NB4. Following this, 5 sequentially forward and backward amino acid around each of the residue 149, residue 380, residue 615, and residue 680 were taken as individual cluster to calculated its distances to nucleotides from the hairpin DNA. The average distance of the

hairpin DNA with amino acids from all four cluster was determined to be 6.08 Å, which was used as the reference for the vicinity of interaction partner with *Pfu* pol. Taken the possibility of each hot spot belonging to a certain amino acid, distances of hot spots to hairpin DNA NB2 and NB4, all these information together. The residue of the assigned amino acid type with the lowest distance to hairpin DNA, NB2 and NB4 was determined to be the hot spot. Thus, we achieved the unambiguous assignment of hot spot ali1 as $C_\beta$ of A684 (Figure S6), hot spot aro1 as Cε of H680 and hot spot aro2 as Cε of Y385. Hot spots aro3 and aro4 could be not unambiguously assigned. As the chemical shift perturbation could arise from the common conformational change of *Pfu* pol upon binding with hairpin DNA, nanobody NB2 and NB4. The distance constrain from the complex structure could not be applied in this case. All three unambiguously assigned hot spots were plotted on the structure of *Pfu* pol (Figure 3E). Among these, Y385 is within the "YxGG/A" motif, which situated in the linker region between the N- and C-terminal domains and had been shown playing a critical role in the DNA synthesis processes. A684 is located close to the active site of *Pfu* pol, whereas the residue H680 is positioned at the periphery of the interaction cluster, distal from the active center.

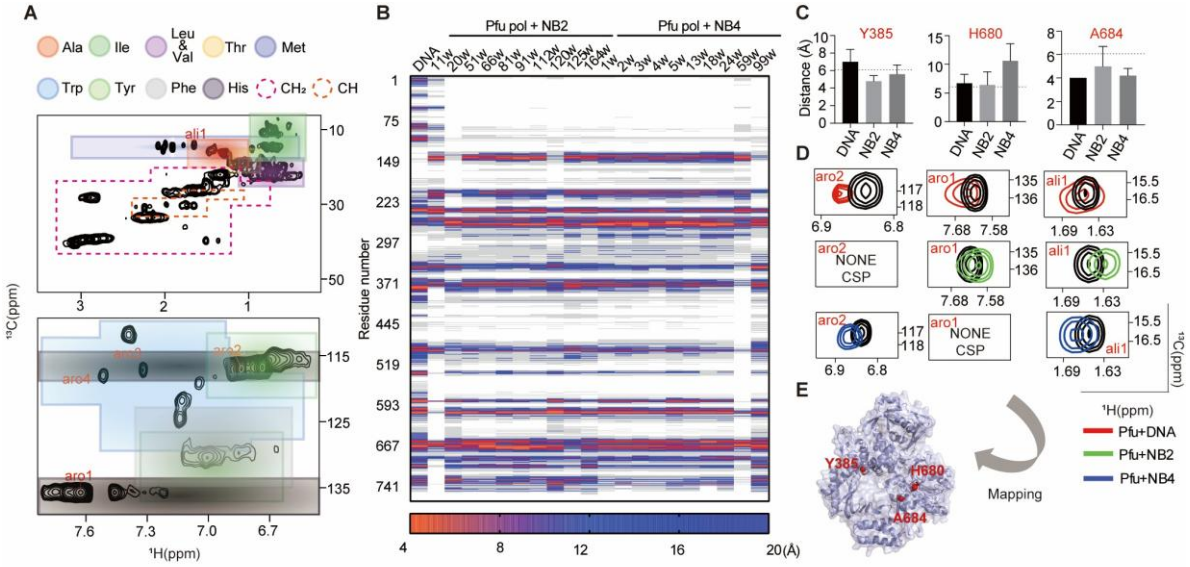



**Figure 3.** Identification of the hot spots of *Pfu* pol. (**A**) Two-dimensional [$^{13}$C,$^{1}$H]-HMQC spectra of 50 μM [$^{13}$C]-labeled *Pfu* pol acquired on the 600 MHz spectrometer at 298 K, where the resonances of the methyl groups and aromatic groups are coloured by different colours. Ala, alanine-$CH_3$ (red); Ile, isoleucine-$CH_3$ (pale green); Leu & Val, leucine-$CH_3$ & valine-$CH_3$, (pink); Thr, threonine-$CH_3$ (yellow); Met, methionine-$CH_3$ (blue); Trp, tryptophan-Ind (pale blue); Tyr, tyrosine-PhO (green); Phe, Phenylalanine-Ph (grey); His, histidine-Im (light black). The central point of each rectangle represents the mean values of $^{1}$H and $^{13}$C chemical shifts of the corresponding amino acid, while the width and height correspond to twice the standard deviations of the $^{1}$H and $^{13}$C chemical shifts, respectively. (**B**) Hierarchical clustering of distance of interacting residues of DNA substrate and nanobodies with *Pfu*

pol. The heatmap is colored from red to blue to grey based on the distance between the *Pfu* pol and its ligand. Residues at a distance of 30 Å or greater are indicated in white. The positions where hot spots of amino acid types were identified are marked with red (right). (**C**) Bar chart showing the distance of assigned aro1(H680 Cε), aro2(Y385 Cε) and ali1 (A684 $C_\beta$) with hairpin DNA, nanobody NB2 and NB4 respectively. Dashed line represents the averaged distance of hairpin DNA with all 8 cluster, which is 6.08 Å. (**D**) Enlarged view of NMR peaks of aor2, aro1 and ali1 residues, which show chemical shift perturbations upon adding 50 μM hairpin DNA (red), 75 μM NB2 (pale green) and 75 μM NB4 (blue). (**E**) The surface representation of *Pfu* Pol displays the positions of the identified hot spots within. The protein surface is colored by light blue, and the hot spots is colored by red. $CH_3$, methyl; Ind, indole; PhO, phenyl-OH; Ph, phenyl; Im, Imidazole.

**Directed evolution of *Pfu* pol based on the assigned hot spots.**

In a further step, we introduced a saturation mutagenesis library for the positions Y385, H680 and A684 to all other 20 alternatives for the evolution of *Pfu* pol. Mutants of *Pfu* pol was transformed into *E. coli* BL21 strain individually. Single colony was picked from LB agar plates and cultured in the 96-deepwell plate. Leveraging the high thermostability of *Pfu* pol, we directly utilized the recombinant *E. coli* cells as a source of *Pfu* pol in the PCR system. The PCR reaction with *E. coli* cell expressing *Pfu* pol mutants was performed by using a 10 kbp long template for amplification. The optimal amount of *E. coli* cell used in each PCR reaction was determined by using the wild type *Pfu* pol, which was add 1 μL of *E. coli* cell with OD (600nm) value of 0.33 in a 20μL reaction system (Figure S10). To our surprise, for all three assigned residues Y385, H680, and A684, we screened mutants with superior performance, mainly higher yield, than the wide type *Pfu* pol (Figure 4A). The mutation of Y385F revealed an enhanced activity in amplifying the 10kb template. Nevertheless, the mutations of aromatic residue Y385 to a non-aromatic residue almost completely lost its polymerase function (Figure 4B), which revealed the key role of Y385 for the function of *Pfu* pol. Further experiments also revealed the Y385F mutant significantly enhanced the amplification performance of *Pfu* pol in amplifying templates with varied length (Figure 5A). Meanwhile, the mutation of H680K not only showed improved activity, but also revealed enhanced specificity of during the amplification (Figure 5C). For the position A684, we found that when alanine is mutated to glutamine, the A684Q mutant reduced the polymerase's salt tolerance (Figure. S10), thereby enhancing its compatibility with high GC templates (Figure. 5B). As the reduction of salt concentration weakens the stability of DNA duplexes, leading to a decrease in the melting temperature (Tm) of GC-rich regions. Additionally, lower salt concentrations can suppress the formation of secondary structures, such as hairpins, by reducing the propensity of DNA to self-fold. Consequently, this allows for primer binding at lower annealing temperatures, thereby mitigating amplification failures caused by the high GC content of the template [44].

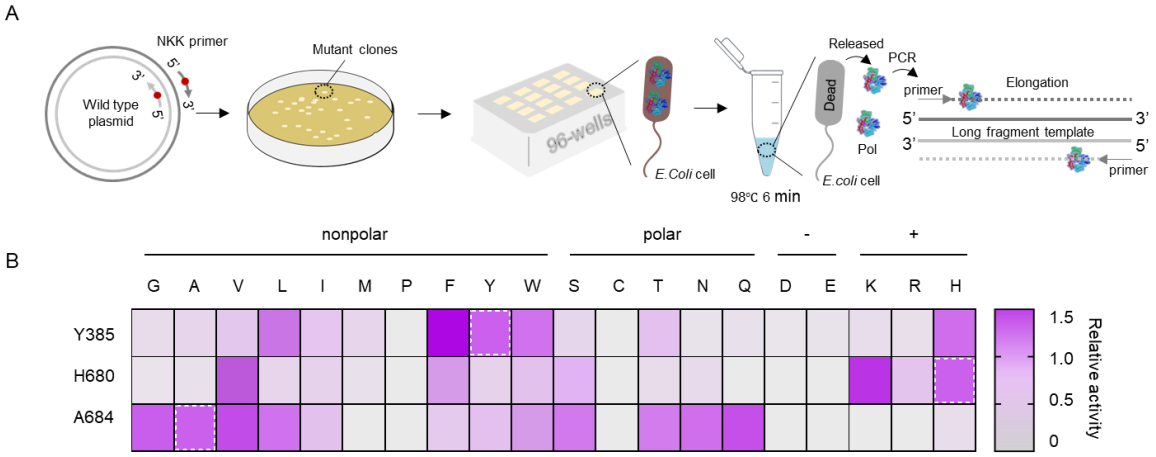



**Figure 4.** (**A**) Schematic illustration of screening for high-performance *Pfu* pol mutants. NKK mutations of residue Y385, H680 and A676 are introduced into the plasmid encoding the protein of wild type *Pfu* pol, followed by transformation into *E.coli* BL21 (DE3) competent cells. After growth, the recombinant cells are diluted to an appropriate OD and used directly as a source of polymerase in the PCR reaction system. After a pre-heating step of 6 minutes, the cells release variants of *Pfu* polymerase, which can catalyze the PCR reaction under specified conditions for screening. (**B**) The yield of the final PCR products was plotted as heatmap to represent the relative activity of *Pfu* pol and its variant. Wild-type *Pfu* pol are labeled with a white dashed rectangle.

**Biochemical and structural characterization of screened *Pfu* pol mutants**

The fidelity and enzymatic kinetics of the screened *Pfu* pol mutants were further investigated. A plasmid-based lacZ gene assay was performed to measure the fidelity [40]. The ratio of blue to white colonies or plaques was used to calculate the polymerase error rate which was expressed as the number of errors per base incorporated. Y385F mutant revealed an increased fidelity from 3.80 ± 0.37 to 2.68 ± 0.28. The other two mutants 3.24 ± 0.46 and 3.99 ± 0.46 showed a similar fidelity with wide type *Pfu* pol (Table S2). Moreover, Y385F revealed an enhanced $k_{cat}$ value from 190.7 ± 11.1 to 258.0 ± 8.22 $s^{-1}$. This hot spot-guided evolution method produced mutants with high catalytic rate constants ($k_{cat}$) values. which are typically challenging to achieve through traditional directed evolution methods [9]. In traditional directed evolution, high catalytic efficiency is usually obtained by reducing the Michaelis constant ($K_m$). The H680K mutant resulted in a decreased Michaelis constant (*Km*) from 0.45 ± 0.07 to 0.31 ± 0.06. From the structures of wild-type and H680K mutant of *Pfu* pol generated by AlphaFold3, the mutation of H to K increased the positive surface electrostatic potential of *Pfu* pol, which may lead to a tighter binding with negative charged DNA (Figure 5F). Similarly, comparing the structures of the wild-type and A684Q mutant of *Pfu* pol generated by AlphaFold3, it showed the side chain of glutamine extends deeper into the minor groove of the DNA double helix. This may enhance *Pfu* pol's interaction affinity to DNA in a low salt concentration buffer, thus improving its performance (Figure 5E).

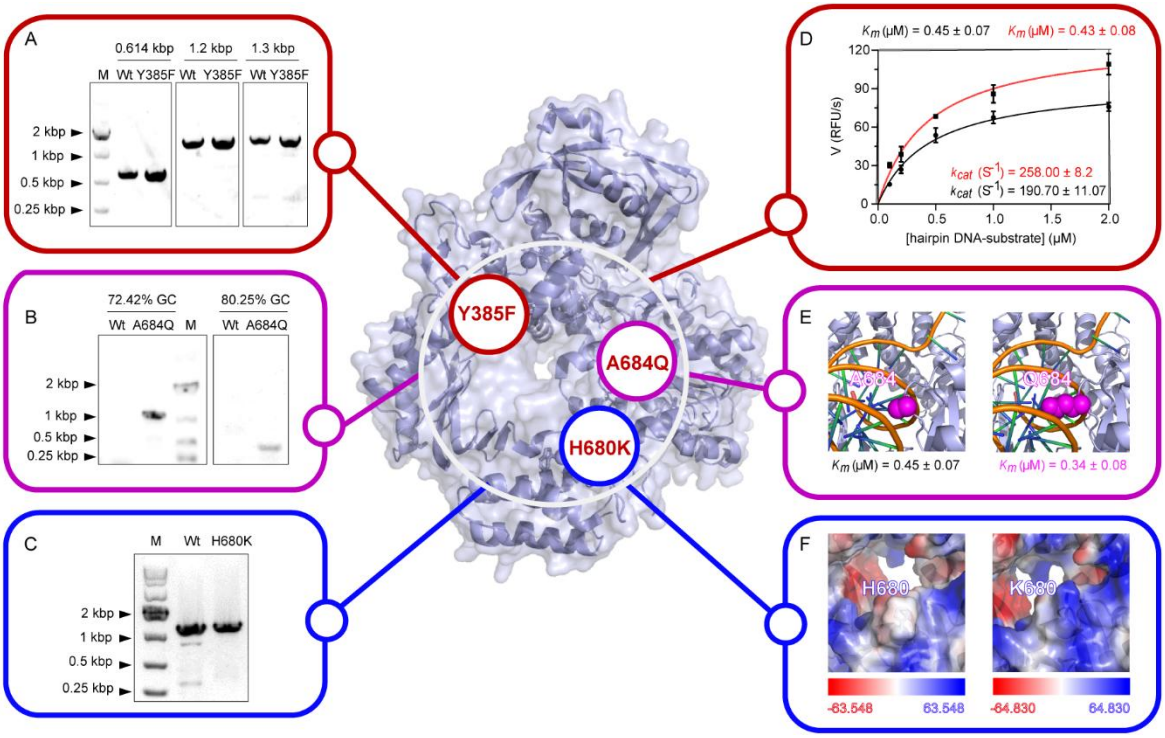



**Figure 5.** Biochemical and structural analysis of the selected *Pfu* pol mutants. The structure shows the 3 amino acid positions mutated from *Pfu* Pol. The left panels showed the PCR amplification performance of Y385F (red), H680H (blue)and A684Q (purple) mutant of *Pfu* pol compared with the wide type *Pfu* pol. The final PCR products were visualized on a 1% agarose gel. (**A**) PCR amplification of human gDNA substrate (with target length 614 bp, 1.2 kb and 1.3 kb). (**B**) PCR amplification of rice cDNA substrate (with target length 471 bp, 80.25% GC, and 1.17 kb ,72.42% GC). (**C**) PCR amplification of cotton gDNA substrate (with target length 1.3 kb). The right panels showed that biochemical and structural analysis of the selected *Pfu* pol mutants. (**D**) Initial reaction rates ($V_0$) of the *Pfu* pol were plotted versus different concentrations of the hairpin DNA substrate. Assays contained 1 nM wide type or Y385F *Pfu* pol and varied concentrations (0.1, 0.2, 0.5, 1 and 2 μM) of hairpin DNA templates respectively. The error bars represent the standard error from three replicates. Data was fitted to the Michaelis–Menten kinetics model. (**E**) Structural illustration of the DNA substrate binding pocket of the wild-type *Pfu* pol and the mutant Y385F. *Pfu* pol residues are shown in ball and stick representation colored by light blue. The DNA substrate is shown in ball and stick representation colored by orange. The key residue of WT and Y385F was drawn as spheres in purple. (**F**) Cross-section view of the positions of the mutated residues in *Pfu* pol. *Pfu* pol is shown as an electrostatic potential surface (red, negative; blue, positive).

## 3. DISCUSSION

Here, we effectively showcase the capabilities of the proposed RAMA method by pinpointing five key hot spots and assigning 3 out of 5 recognized hot spots on the multi-domain Pfu polymerase, which has

a molecular weight of up to 90 kDa. The use of $^{1}H$-$^{13}C$ moieties in NMR spectroscopy has expanded the range of proteins suitable for analysis to those in the Mega-Dalton range, enabling the identification of active sites in high-molecular-weight enzymes. The distinct chemical shift characteristics of $^{1}H$-$^{13}C$ in amino acid residues make it relatively straightforward to classify these hot spots into specific amino acid types. Although the chemical shifts of valine and leucine side chains exhibit significant overlap, the side chains of other aliphatic amino acids possess distinct chemical shift characteristics that are distributed widely, making it straightforward to narrow them down to 1 to 3 amino acid types. In contrast, aromatic amino acid side chains display a broader range of $^{13}C$ chemical shifts, significantly reducing the issue of spectral overlap. In our experiments, the widely dispersion of the methyl group chemical shift enable us to narrow down the amino acid type of hot spot residues within 3, and with a high possibility from 70.3 to almost 100% to be one amino acid type. Combining with the structures of complexes of targeted enzymes with either specific substrate, inhibitors, or neutralization antibodies, the search space of interested residues could be further narrowed down. The distance information and chemical shift perturbations (CSPs) provided by complex structures can offer valuable constraints for assigning amino acid types, facilitating the unambiguous assignment of hot spot residues. Meanwhile, the known cofactors of an enzymes can also help to pinpoint hot spots directly involved in catalytic activity. Additionally, characteristic chemical shifts of specific groups—such as Cβ signals of Thr and Ser, Cα signals of Gly and Pro, and methylene/methine groups—provided further support for hot spot identification and assignment. In case of overlapping of the methyl group peaks, such as leucine and valine, cross-selective isotopic labeling could also be employed to resolve ambiguities. In our study, the aro3 and aro4 in our case could not be unambiguously assigned, as the chemical shift perturbation of hot spot could arise from the conformation changes of *Pfu* pol. Although the proposed RAMA approach has a limitation in unambiguous assignment of hot spots caused by the conformational change of the target enzymes, the narrowed search space of aro3 and aro4 to either tryptophan with a possibility of at least 85% or histidine with less than 15% probability, enabling a researcher to conduct further screening within a much smaller set of variants.

Three of the mutants we obtained increased *Pfu* pol activity in different ways. This highlights the advantage of using a "compact and precise" library guided by hot spot identification, which eliminates the need for high-throughput screening. Instead, a single round of rapid selection sufficed to remove deleterious mutations. Previous studies indicated that the Y-GG/A motif, located in a solvent-accessible loop between the 3'-5' exonuclease and polymerization domains of B-family DNA polymerases, is linked to polymerase catalytic activity. The tyrosine 387 of *Pfu* pol identified in our study was also located in the Y-GG/A motif. The mutation of Y387 to phenylalanine enhanced *Pfu* pol activity, demonstrating a high success rate in identifying hot spot amino acid types and positions in large enzymes. In contrast to Y387, H680 located at the periphery of the interaction cluster. It was still

successfully identified and screened to mutate to lysine for an improved performance, showing our approach retained the capacity to probe proximal residues around the interaction interface.

Overall, this method showcases strong compatibility, expanding the use of NMR spectroscopy in the directed evolution of large proteins. It is also worth noting the $^{1}H$-$^{13}C$ HMQC spectra of unlabeled enzymes could also be recorded through the natural abundant $^{13}C$ carbons, which is advantageous given the difficulty and expense of obtaining labeled enzymes (Figure S11). Thus, this proposed RAMA approach holds great promise for advancing enzyme evolution, paving the way for the development of more efficient and versatile biocatalysts in the future.

## CRediT authorship contribution statement

Y.C., and L.H. planned the study and designed experiments. Y.C., and Z.G. performed experiments. X.Z., M.L., and L.H. acquired funds. All authors analyzed and interpreted the results. Y.C., and L.H. wrote the manuscript with input from all authors.

## Funding Statement

This work is supported by the Strategic Priority Research Program of the Chinese Academy of Sciences XDB0540000, Natural Science Foundation of China grants 22327901, 22174151 and 21991080, and Hubei Provincial Natural Science Foundation of China 2023AFA041.

## Declaration of competing interest

The authors declare no conflict of interest.

## Appendix A. Supplementary data

Supplementary material.

## Data Availability

Data will be made available on request.

## REFERENCES


(1) Bell EL, Smithson R, Kilbride S, et al. Directed evolution of an efficient and thermostable PET depolymerase. Nat Catal. 2022, 5:673-681.

(2) Zhang S, Zhu J, Fan S, Xie W, Yang Z, Si T. Directed evolution of a cyclodipeptide synthase with new activities *via* label-free mass spectrometric screening. Chem Sci. 2022, 13:7581-7586.

(3) Carr R, Alexeeva M, Enright A, Eve TS, Dawson MJ, Turner NJ. Directed evolution of an amine oxidase possessing both broad substrate specificity and high enantioselectivity. Angew Chem Int Ed Engl. 2003, 42(39):4807-4810.

(4) Soskine M, Tawfik DS. Mutational effects and the evolution of new protein functions. Nat Rev Genet. 2010, 11(8):572-582.

(5) Geddie ML, Matsumura I. Rapid evolution of beta-glucuronidase specificity by saturation mutagenesis of an active site loop. J Biol Chem. 2004, 279(25):26462-26468.

(6) Wang Y, Xue P, Cao M, Yu T, Lane ST, Zhao H. Directed Evolution: Methodologies and Applications. Chem Rev. 2021, 121(20):12384-12444.

(7) Wu Z, Kan SBJ, Lewis RD, Wittmann BJ, Arnold FH. Machine learning-assisted directed protein evolution with combinatorial libraries. Proc Natl Acad Sci USA. 2020, 117(1):788-789.

(8) Marshall LR, Bhattacharya S, Korendovych IV. Fishing for Catalysis: Experimental Approaches to Narrowing Search Space in Directed Evolution of Enzymes. JACS Au. 2023, 3(9):2402-2412.

(9) Bhattacharya S, Margheritis EG, Takahashi K, Kulesha A, D'Souza A, Kim I, Yoon JH, Tame JRH, Volkov AN, Makhlynets OV, Korendovych IV. NMR-guided directed evolution. Nature. 2022, 610(7931):389-393.

(10) Otten R, Pádua RAP, Bunzel HA, et al. How directed evolution reshapes the energy landscape in an enzyme to boost catalysis. Science. 2020, 370(6523):1442-1446.

(11) Acevedo-Rocha CG, Li A, D'Amore L, et al. Pervasive cooperative mutational effects on multiple catalytic enzyme traits emerge via long-range conformational dynamics. Nat. Commun. 2021, 12:1621.

(12) Osuna S. The challenge of predicting distal active site mutations in computational enzyme design. WIREs Comput. Mol. Sci. 2020, 11:e1502.

(13) Mainz A, Religa TL, Sprangers R, Linser R, Kay LE, Reif B. NMR spectroscopy of soluble protein complexes at one mega-dalton and beyond. Angew Chem Int Ed Engl. 2013, 52(33):8746-8751.

(14) Frueh DP. Practical aspects of NMR signal assignment in larger and challenging proteins. Prog Nucl Magn Reson Spectrosc. 2014, 78:47-75.

(15) Lundberg KS, Shoemaker DD, Adams MW, Short JM, Sorge JA, Mathur EJ. High-fidelity amplification using a thermostable DNA polymerase isolated from Pyrococcus furiosus. Gene. 1991, 108(1):1-6.

(16) Biles BD, Connolly BA. Low-fidelity Pyrococcus furiosus DNA polymerase mutants useful in error-prone PCR. Nucleic Acids Res. 2004, 32(22):e176.

(17) Moore P. PCR: replicating success. Nature. 2005, 435(7039):235-238.

(18) Shendure J, Balasubramanian S, Church GM, et al. DNA sequencing at 40: past, present and future. Nature. 2017, 550(7676):345-353.

(19) He L, Lührs T, Ritter C. Solid-state NMR resonance assignments of the filament-forming CARD domain of the innate immunity signaling protein MAVS. Biomol NMR Assign. 2015, 9(2):223-7.

(20) Vranken WF, Boucher W, Stevens TJ, et al. The CCPN data model for NMR spectroscopy: development of a software pipeline. Proteins. 2005, 59(4):687-696.

(21) Abramson J, Adler J, Dunger J, et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature. 2024, 630(8016):493-500.

(22) Honorato RV, Trellet ME, Jiménez-García B, et al. The HADDOCK2.4 web server for integrative modeling of biomolecular complexes. Nat Protoc. 2024, 19(11):3219-3241.

(23) Mao F, Leung WY, Xin X. Characterization of EvaGreen and the implication of its physicochemical properties for qPCR applications. BMC Biotechnol. 2007, 7:76.

(24) Tao Z, Zhao X, Wang H, et al. A method for rapid nanobody screening with no bias of the library diversity. iScience. 2024, 27(2):108966.

(25) Henke W, Herdel K, Jung K, Schnorr D, Loening SA. Betaine improves the PCR amplification of GC-rich DNA sequences. Nucleic Acids Res. 1997, 25(19):3957-3958.

(26) Green MR, Sambrook J. Polymerase Chain Reaction (PCR) Amplification of GC-Rich Templates. Cold Spring Harb Protoc. 2019, 2019(2):10.1101/pdb.prot095141.

(27) Ulrich EL, Akutsu H, Doreleijers JF, et al. BioMagResBank. Nucleic Acids Res. 2008, 36(Database issue):D402-D408.

(28) Mishra AK, Mariuzza RA. Insights into the Structural Basis of Antibody Affinity Maturation from Next-Generation Sequencing. Front Immunol. 2018, 9:117.

(29) Yin R, Feng BY, Varshney A, Pierce BG. Benchmarking AlphaFold for protein complex modeling reveals accuracy determinants. Protein Sci. 2022, 31(8):e4379.

(30) Gaudreault F, Corbeil CR, Sulea T. Enhanced antibody-antigen structure prediction from molecular docking using AlphaFold2. Sci Rep. 2023, 13(1):15107.

(31) Xu Y, Zhang D, Rogawski R, Nimigean CM, McDermott AE. Identifying coupled clusters of allostery participants through chemical shift perturbations. Proc Natl Acad Sci U S A. 2019, 116(6):2078-2085.

(32) Wu FJ, Rieder PS, Abiko LA, Rößler P, Gossert AD, Häussinger D, Grzesiek S. Nanobody GPS by PCS: An Efficient New NMR Analysis Method for G Protein Coupled Receptors and Other Large Proteins. J Am Chem Soc. 2022, 144(47):21728-21740.

(33) Bloch JS, Mukherjee S, Kowal J, Filippova EV, Niederer M, Pardon E, Steyaert J, Kossiakoff AA, Locher KP. Development of a universal nanobody-binding Fab module for fiducial-assisted cryo-EM studies of membrane proteins. Proc Natl Acad Sci U S A. 2021, 118(47):e2115435118.

(34) Robertson MJ, Papasergi-Scott MM, He F, Seven AB, Meyerowitz JG, Panova O, Peroto MC, Che T, Skiniotis G. Structure determination of inactive-state GPCRs with a universal nanobody. Nat Struct Mol Biol. 2022, 29(12):1188-1195.

(35) Yu Z, Li P, Merz KM Jr. Using Ligand-Induced Protein Chemical Shift Perturbations To Determine Protein-Ligand Structures. Biochemistry. 2017, 56(18):2349-2362.

(36) Ten Brink T, Aguirre C, Exner TE, Krimm I. Performance of protein-ligand docking with simulated chemical shift perturbations. J Chem Inf Model. 2015, 55(2):275-83.

(37) Krishnamoorthy J, Yu VC, Mok YK. Auto-FACE: an NMR based binding site mapping program for fast chemical exchange protein-ligand systems. PLoS One. 2010,5(2):e8943.

(38) Li A, Wang B, Ilie A, et al. A redox-mediated Kemp eliminase. Nat Commun. 2017, 8:14876.

(39) Irobalieva RN, Manolaridis I, Jackson SM, Ni D, Pardon E, Stahlberg H, Steyaert J, Locher KP. Structural Basis of the Allosteric Inhibition of Human ABCG2 by Nanobodies. J Mol Biol. 2023, 435(19):168234.

(40) Keith BJ, Jozwiakowski SK, Connolly BA. A plasmid-based lacZα gene assay for DNA polymerase fidelity measurement. Anal Biochem. 2013, 433(2):153-61.

(41) Truniger V, Lázaro JM, Salas M, Blanco L. A DNA binding motif coordinating synthesis and degradation in proofreading DNA polymerases. EMBO J. 1996, 15(13):3430-41.

(42) Böhlke K, Pisani FM, Vorgias CE, et al. PCR performance of the B-type DNA polymerase from the thermophilic euryarchaeon Thermococcus aggregans improved by mutations in the Y-GG/A motif. Nucleic Acids Res. 2000, 28(20):3910-3917.

(43) Wynne SA, Pinheiro VB, Holliger P, Leslie AG. Structures of an apo and a binary complex of an evolved archeal B family DNA polymerase capable of synthesising highly cy-dye labelled DNA. PLoS One. 2013, 8(8):e70892.

(44) Gruenwedel DW, Hsu CH. Salt effects on the denaturation of DNA. Biopolymers. 1969, 7(4):557-70.

(45) Jones S, Thornton JM. Principles of protein-protein interactions. *Proc Natl Acad Sci U S A*. 1996, 93(1):13-20.

(46) Brooks BR, Brooks CL 3rd, Mackerell AD Jr, et al. CHARMM: the biomolecular simulation program. *J Comput Chem*. 2009, 30(10):1545-1614.

(47) Doddrell DM, Pegg DT, Bendall MR. Distortionless enhancement of NMR signals by polarization transfer. *J Magn Reson (1969)*. 1982, 48(2):323-327.

## SUPPLEMENTARY DATA

This file contains:

- Supplementary Table S1-S2
- Supplementary Figures S1–S11

**Table S1.** Synthesized oligonucleotides.

| Name | Sequence |
|---|---|
| Hairpin DNA | 5’-CCAGCATTATGAAAGTGACACGTGCACCATTGGTGCACGTG-3’ |
| Hairpin substrate | 5’-CCAGCATTATGAAAGTGACACGTGCACCATTGGTGCACGTGTCACTTTCATAATGCTGG-3’ |
| 10K F | 5’-ATCCTGGACGTGGATTACATCACC-3’ |
| 10K R | 5’-CATATGACGACCTTCGATATGGCC-3’ |
| Ap cDNA F | 5’-GTCCTGGCTCTCTTCTACAAGG-3’ |
| Ap cDNA R | 5’-GGTACGTGCTGTTGAACTGTTCC-3’ |
| Mt gDNA F | 5’-AACCCAGAAGACAGGTGGAAAG-3’ |
| Mt gDNA R | 5’-GCCAGATTACGTATATCCTGGCAG-3’ |
| M13 F | 5’-AAGCCATCCGCAAAAATGACCTCT-3’ |
| M13 R | 5’-GTCAGAAGCAAAGCGGATTGCATC-3’ |
| λDNA PS F | 5’-GCCTGGCTCCACGCTCCGAGTGGCGAAAATTCGTCGCATCCCGTCAGG-3’ |
| λDNA PS R | 5’-GCCTGGCTCCACGCTCCGAGTTTCTGGCTGAGTTCGTGCTTACCGCAG-3’ |
| Y385N F | 5’-CGCCTGCGCGAATCCNNKACCGGCGGTTTCGTG-3’ |
| Y385N R | 5’-CACGAAACCGCCGGTMNNGGATTCGCGCAGGCG-3’ |
| H680N F | 5’-ATCACGCGTCCGCTGNNKGAATACAAAGCCATC-3’ |
| H680N R | 5’-GATGGCTTTGTATTCMNNCAGCGGACGCGTGAT-3’ |

| | |
|---|---|
| A684N F | 5’-CTGCACGAATACAAANNKATCGGTCCGCATGTT-3’ |
| A684N R | 5’-AACATGCGGACCGATMNNTTTGTATTCGTGCAG-3’ |
| 250 F | 5’-CTTGGATCCGGTCAGCTGAAACAGCGTC-3’ |
| 250 R | 5’-GGGAAGCTTTTAGCCCTGAATCTG-3’ |
| Os pro5 F | 5’-ATAACACATCCCTCCTTTTGTTATTTTATATAC-3’ |
| Os pro5 R | 5’-AGCTACAGCCTACACAGC-3’ |
| Os pro6 F | 5’-GCCCTTTGGGATCAAGCAAATTAAACGC-3’ |
| Os pro6 R | 5’-GGCTCCTGCTCTCTCCCC-3’ |
| Os pro7 F | 5’-CTTCTATATATAAAGTCATTTTCAACAATTATATTAATTCC-3’ |
| Os pro7 R | 5’-CTCTTCCTCTTGCTGTTCTTGCTTCTTC-3’ |
| Os CDS1 F | 5’-ATGGCGTCCTACGACAAGGCC-3’ |
| Os CDS1 R | 5’-TTATTTGCTTGCATCTTGGGGATCTTGATTCTC-3’ |
| Gh gDNA F | 5’-GGTACCGAGATACAGAAGTTTTAG-3’ |
| Gh gDNA R | 5’-GTCGACCATCTAGAACAGGAGTGAC-3’ |
| Puc F | 5’-GAATTCGAGCTCGGTACCC-3’ |
| Puc R | 5’-ACTGGCCGTCGTTTTAC-3’ |

**Table S2** Summary of the steady-state kinetics analyses

| Enzyme | $k_{cat}$ ($s^{-1}$) | $K_m$ (μM) | Error rate ($\times 10^{-6}$)[a] |
|---|---|---|---|
| ***Pfu*$^{WT}$-Pol** | 190.70 ± 11.07 | 0.45 ± 0.07 | 3.80 ± 0.37 |
| ***Pfu*$^{Y385F}$-Pol** | 258.00 ± 8.22 | 0.43 ± 0.08 | 2.68 ± 0.28 |
| ***Pfu*$^{H680K}$-Pol** | 173.28 ± 5.36 | 0.31 ± 0.06 | 3.24 ± 0.46 |
| ***Pfu*$^{A684Q}$-Pol** | 194.90 ± 2.36 | 0.34 ± 0.03 | 3.99 ± 0.46 |

[a] The circular pUC19 plasmid was linearized by digestion with EcoR I. The linearized pUC19 plasmid were amplified by the *Pfu* pol or its mutants in the buffer containing 25 mM Tris-HCl pH 8.8, 10 mM KCl, 6 mM $(NH_4)_2SO_4$, 2 mM $MgCl_2$, 200 μM dNTPs, and 0.05% Triton X-100. The concentrations of the DNA polymerase and the linearized plasmid were 50 nM and 2 nM, respectively. Amplification reactions were performed using a thermos cycler (C1000 Touch, Bio-rad): First, incubation at 95 ℃ for 2 min; For *Pfu* pol, followed by 26 thermal cycles of 95 ℃ for 15 s, 55 ℃ for 15 s, and 72℃ for 3 min. The linearized pUC19 plasmid PCR product was examined using1% agarose gel electrophoresis and recovered by gel excision, following by a ligation assay with the T4 DNA ligase. Afterwards, 2.5 μl of the ligation product was transformed into *Turbo* competent cells (NEB). Transformed cells were plated on LB agar with 100 mg/ml ampicillin, 1 mg/mL X-Gal and 1.5 Mm IPTG. Blue/white colonies were counted to calculate the error rates using the equation [40]: $ER = mf/(bp \times d)$, Where, *mf* is the mutation frequency determined by dividing the total number of white plaques by the total number of plaques. *d* is the number of DNA duplications and *b* is the effective target size of the linearized pUC19 plasmid, which is 323 bp.

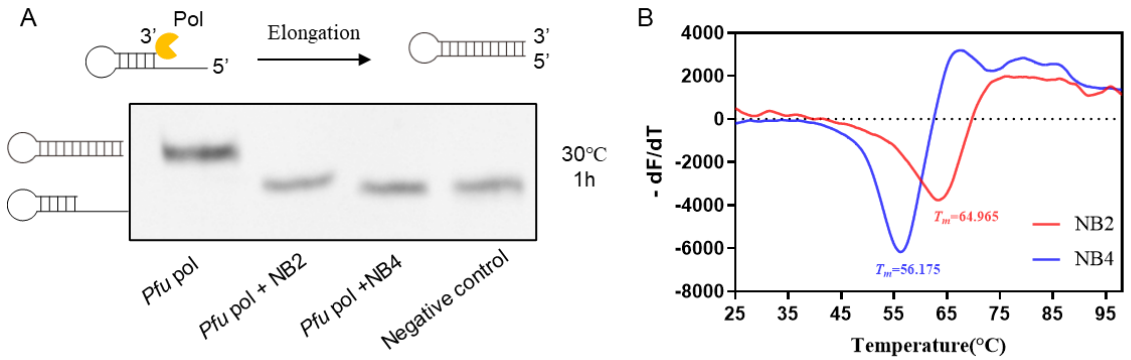



**Figure S1.** (**A**) Two different nanobodies NB2 and NB4 block the DNA polymerase activity of *Pfu* Pol. The hairpin substrate (sequence in Table S1, 100 μM in water) was prepared in advance by heating at 98°C for 5 min and annealing on the ice for 30 min. 2 μM hairpin substrate were then used as the template to mix with dNTPs and 2 nM *Pfu* pol in the presence and absence of 10 nM NB2 or NB4 respectively. $MgCl_2$ was added to initiate the DNA synthesis at 72°C. The reaction buffer contained 10 mM Tris-HCl pH 8.8, 10 mM KCl, 2 mM $MgCl_2$, 200 μM dNTPs, and 0.05% Triton X-100. The lanes of PCR products with either nanobody NB2 or NB4 showed polymerase blocking activity, where no elongation of hairpin DNA occurs. (**B**) Thermostability of two different nanobodies NB2 (red) and NB4 (blue). The thermostability experiments were measured via QuantStudio 3 Real-Time PCR instrument from Thermo Fisher Scientific. The commercial dye SYPRO Orange (Sigma) was used to monitor the thermodynamic stability of both proteins. The melting temperature was programmed from 25 °C to 98 °C with a 0.15 °C temperature increment. Each reaction was performed in a final volume of 20 μl containing 50 mM Tris-HCl pH 8.8, 10 mM KCl, 10× SYPRO Orange and 10 μM of every individual protein. The signal was collected per second. The first derivative of the fluorescence signal curve was calculated using GraphPad Prism 8.

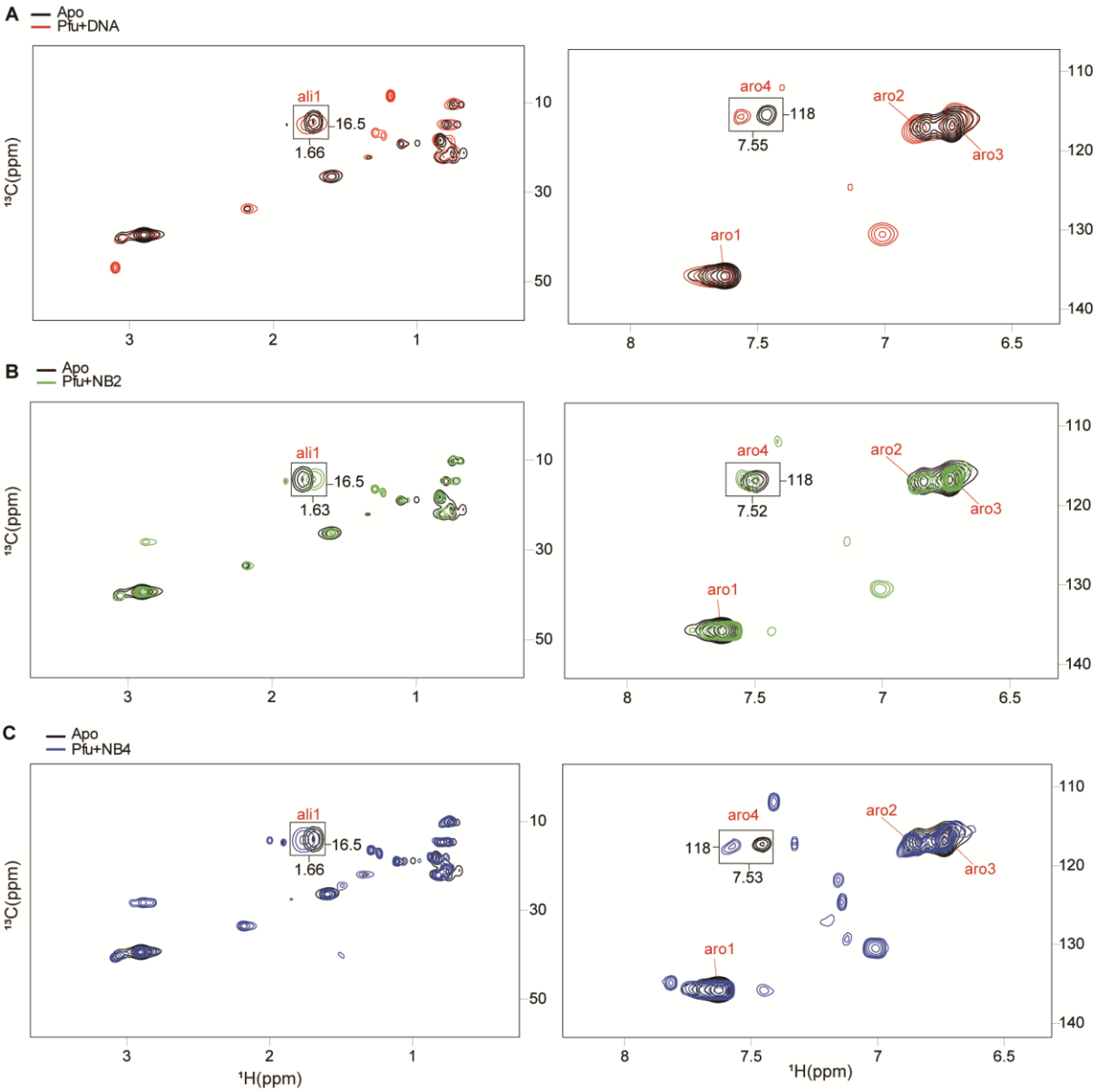



**Figure S2.** 2D [$^{13}$C,$^{1}$H]-HMQC spectra of 50 µM [$^{13}$C]-labeled *Pfu* pol acquired on the 600 MHz ($^{1}$H frequency) spectrometer at 298 K in the absence and presence of 50 µM hairpin DNA (**A**), or 75 µM nanobody NB2 (**B**) or 75 µM nanobody NB4 (**C**).

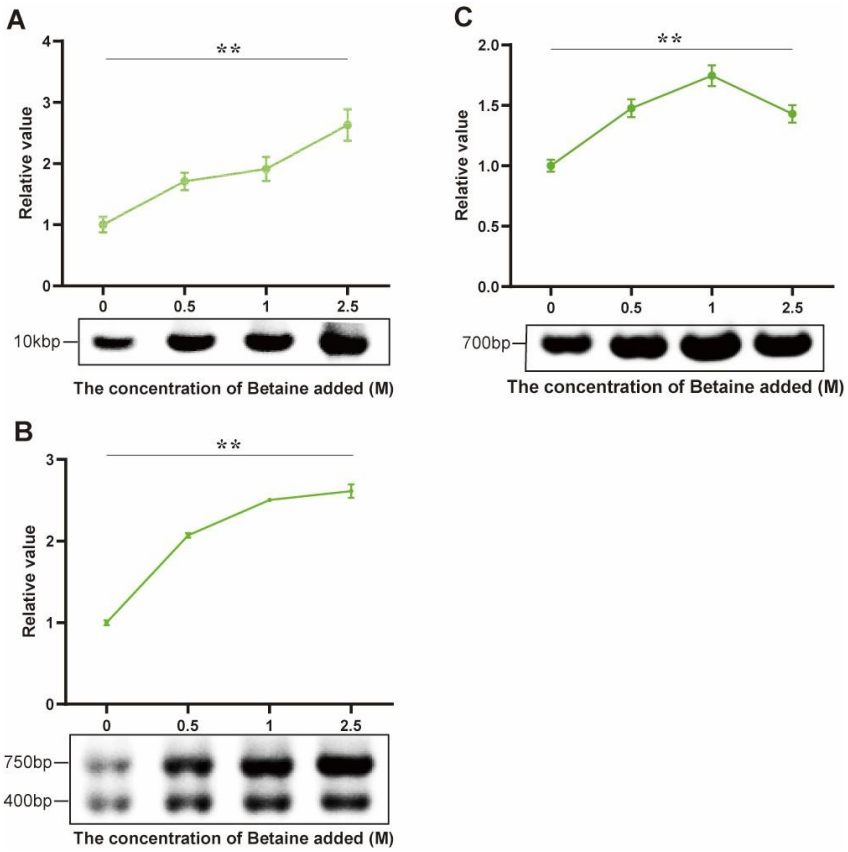



**Figure S3.** The amplification of different DNA templates by *Pfu* pol in the presence of different concentrations of betaine ranging from 0 M to 2.5 M. All PCR assays were performed as described in the methods. The PCR amplification with different templates: a 10 kb plasmid DNA template (**A,** with a target length of 10 kb), an Alpaca cDNA template (**B**, with target length of 750 bp and 400 bp) and a mouse tail gDNA template (**C**, with a target length of 700 bp). Intensities of gel bands were quantified and plotted against the concentration of betaine, showing on top of the corresponding gel image. Results expressed as mean ± SD of 3 independent biological replicates. Data were analyzed using unpaired Student's test, **$P < 0.01$.

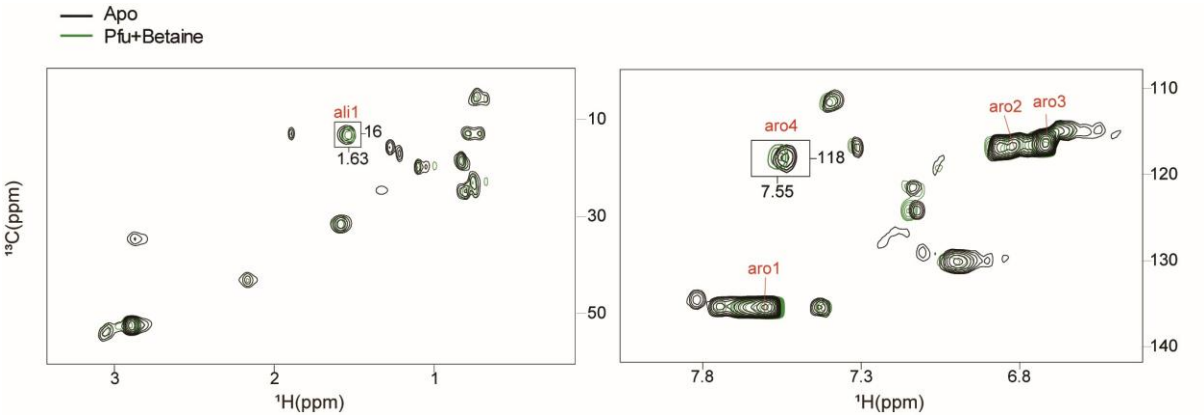



**Figure S4.** 2D [$^{13}$C,$^{1}$H]-HMQC spectra of 50 μM [$^{13}$C]-labeled *Pfu* pol acquired on the 600 MHz ($^{1}$H frequency) spectrometer at 298 K in the absence and presence of 1M betaine.

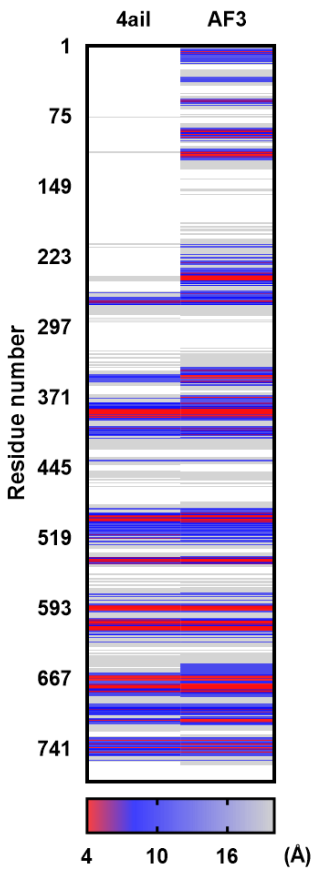



**Figure S5.** Hierarchical clustering of the distance between residues of *Pfu* pol with its corresponding DNA template. The heatmap is colored from red to blue to white based on the distance between the *Pfu* pol and its ligand. 'AF3' is the lowest energy structure of *Pfu* pol with a hairpin DNA substrate generated by Alphafold3.

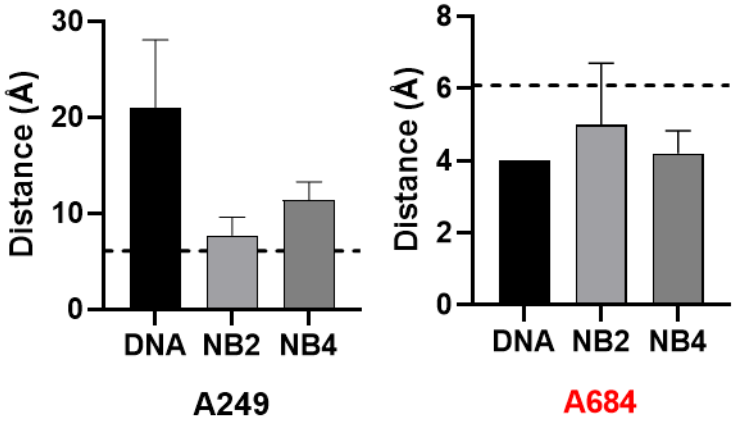



**Figure S6.** Bar chart plots of the distance between residue A249 and A684 of *Pfu* pol with the hairpin DNA, nanobody NB2 and NB4 respectively. Hashed bars represent a reference distance of 6.08 Å.

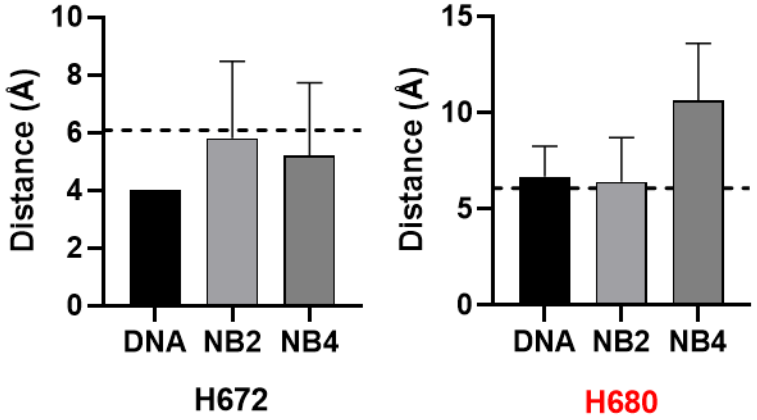



**Figure S7.** Bar chart plots of the distance between residue H672 and H680 of *Pfu* pol with the hairpin DNA, nanobody NB2 and NB4 respectively. Hashed bars represent a reference distance of 6.08 Å.

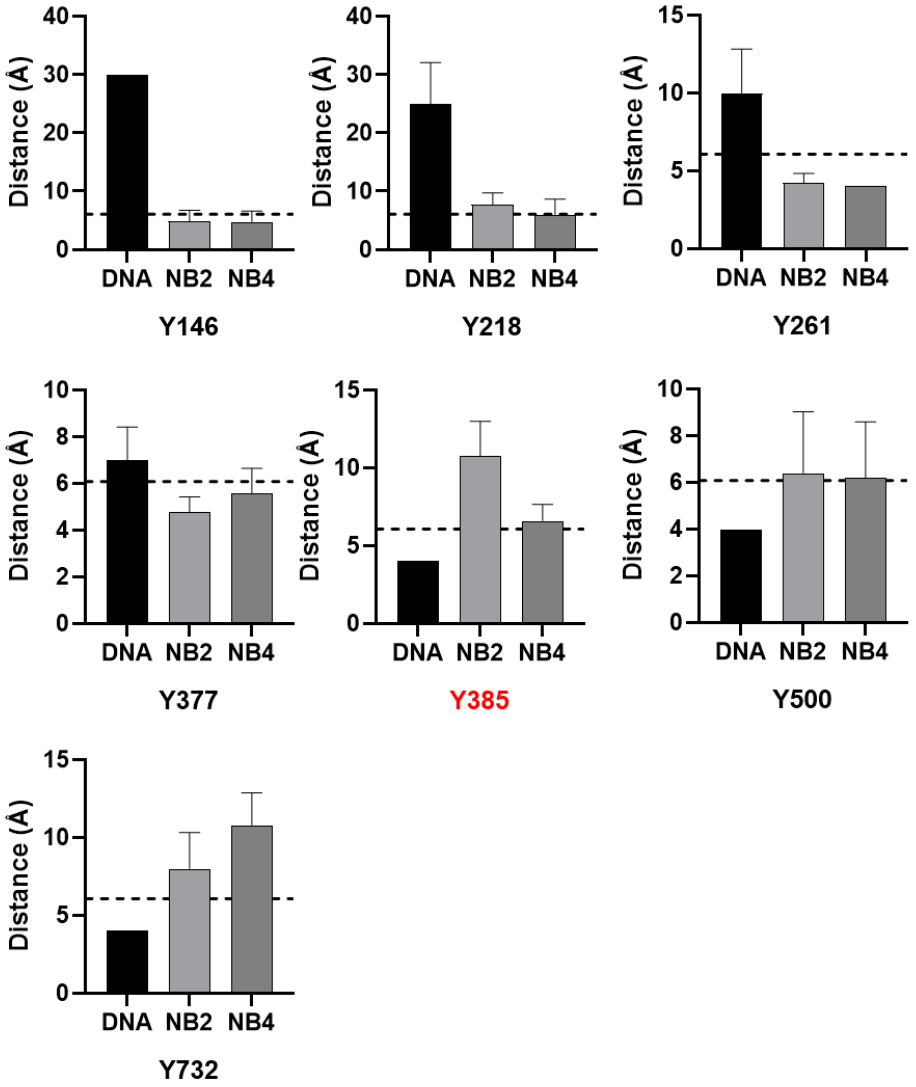



**Figure S8.** Bar chart plots of the distance between tyrosine residues of *Pfu* pol with the hairpin DNA, nanobody NB2 and NB4 respectively. Hashed bars represent a reference distance of 6.08 Å.

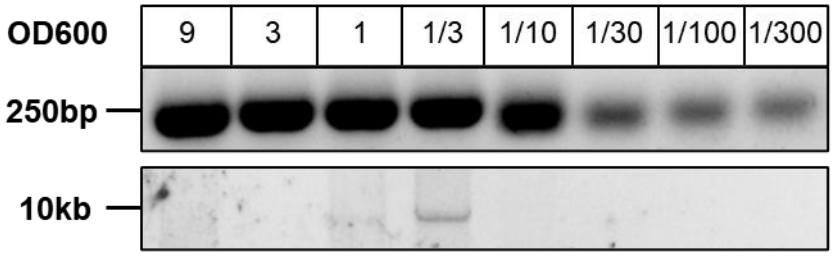



**Figure S9.** The optimal amount of *E. coli* cell used in each PCR reaction was determined by using the wild type *Pfu* pol. In a 20 μL reaction system, add 1 μL of *E. coli* cells with OD600 values of 9, 3, 1, 0.33, 0.1, 0.033, 0.01, and 0.0033, respectively. PCR amplification of different DNA templates (with target length of 250 bp and 10 kb) with different amount *E.coli*. expressing wild-type *Pfu* pol. PCR products are visualized on a 1% agarose gel.

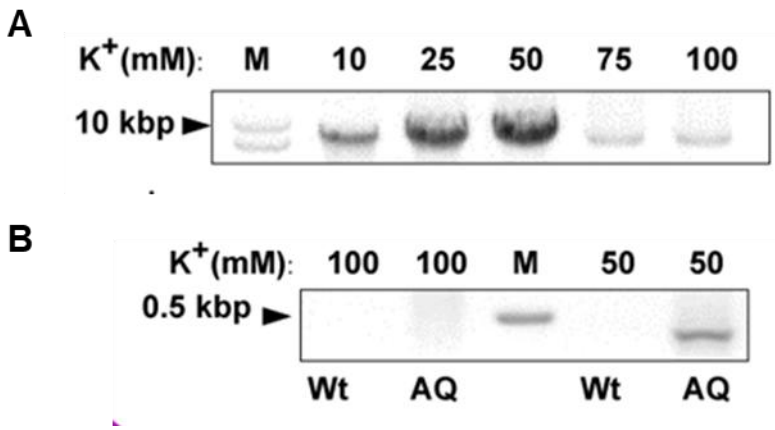



**Figure S10.** PCR amplification with different templates. Products of PCR with a plasmid (A, with target length 10 kb) and rice gDNA template (B, with target length 471 bp, 80.25% GC) in the presence of different concentrations of potassium salt from 10 mM to 100 mM.

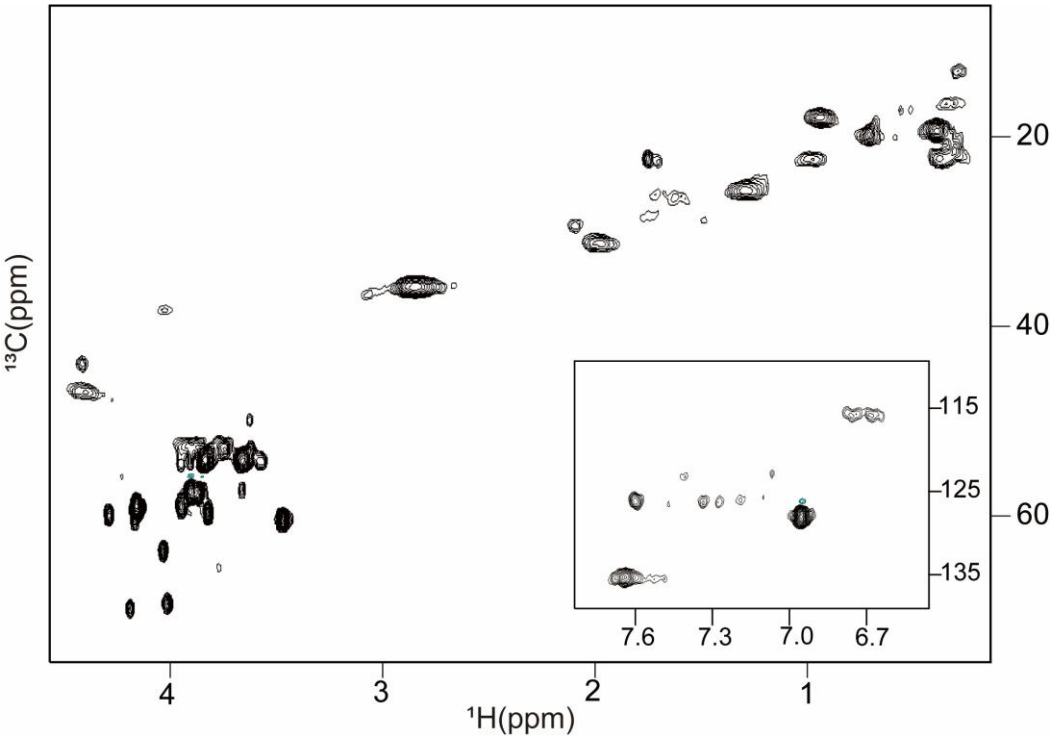



**Figure S11.** 2D [$^{13}$C,$^{1}$H]-HMQC spectra of 1 mM unlabeled *Pfu* pol acquired on the 800 MHz ($^{1}$H frequency) spectrometer at 298 K.