MANUSCRIPT TITLE

An NMR-guided approach for identifying co-factors in boosting the Pfu DNA polymerase

AUTHORS

Yihao Chen[a,b#], Mingjun Zhu[a,b#], Xiaoling Zhao[c,d], Zhiqing Tao[a,b], Dan Xiang[a], 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.

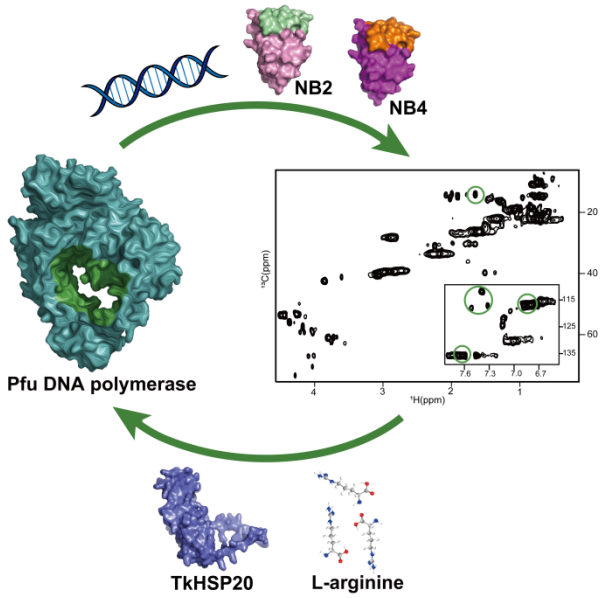
#. These authors contributed equally to this work.

* Corresponding author:

Lichun He

E-mail: helichun@apm.ac.cn

GRAPHICAL ABSTRACT



ABSTRACT

With rapid developments of emerging technologies like synthetic biology, the demand for DNA polymerases with superior activities including higher thermostability and processivity has increased significantly. Thus, rational optimization of the performance of DNA polymerase is of great interest. Nuclear magnetic resonance spectroscopy (NMR) is a powerful technique used for studying protein structure and dynamics. It provides the atomic resolution information of enzymes under its functional solution environment to reveal the active sites (hot spots) of the enzyme, which could be further used for optimizing the performance of enzymes. Here we applied NMR spectroscopy to determine hot spots of the Pfu polymerase. Employing these hot spots as probes, two new co-factors, the heat shock protein TkHSP20 from *Thermococcus Kodakaraensis* and the chemical chaperone L-arginine, are identified to interact with Pfu polymerase to boost its performance in amplifying long DNA fragments by enhancing the thermal stability and the processivity of the Pfu polymerase. This NMR-guided approach requires no prior assignment information of target enzymes, simplifying the exploration of novel co-factors for Pfu polymerase. Moreover, our approach is not dependent on structural data or bioinformatics. Therefore, it has significant potential for application in various enzymes to expedite the progress in enzyme engineering.

INTRODUCTION

PCR technology serves as the cornerstone for sequencing, molecular diagnostics, and synthetic biology (1-4). Despite its apparent simplicity, certain challenges persist, particularly in the context of long-range PCR. Several approaches have been developed to improve the amplification of lengthy DNA fragments, encompassing optimized buffer components, primer design, amplification protocols, and fusion domain engineering (5-8). Of these approaches, the utility of additives directly aiding the DNA polymerase to facilitate the synthesis of long DNA fragments is critical. One major problem in amplifying long DNA fragments is the long elongation time required at high temperatures (37), which will diminish the activity of DNA polymerases (9). DNA polymerases must endure throughout the entire polymerase chain reaction to enable DNA amplification. Additionally, with the augmentation of DNA length, there is an increased probability of secondary structure formation and premature amplicon generation, potentially impeding the PCR process. DNA polymerase with higher processivity is required to address this issue.

Chaperones are highly conserved molecular machines that play a vital role in maintaining the functionality of the proteome under both normal and stressful conditions. They participate in various cellular processes, including the folding of newly synthesized proteins (10), prevention of the aggregation of misfolded proteins (11,12), and the restoration of denatured proteins (13,14). Chaperones essentially serve as guides for client proteins, reducing the likelihood of denaturation, thereby, prolonging the lifetime and activity of client proteins. Additionally, plenty of studies also suggest that small, low-molecular-weight compounds can be effective in inhibiting protein aggregation or enabling proteins to restore their functions (15,16). These small molecules are referred to as chemical chaperones, and they are believed to non-specifically stabilize proteins, facilitating their proper folding. Among chemical chaperones, L-arginine (Arg) stands out as the most potent additive for suppressing protein aggregation induced by heat and reduced reagents, as well as aggregation during *in vitro* folding. Further to this, Bovine serum albumin (BSA), known for its chaperone-like properties (17,18), is widely employed in PCR systems (19-22), although its mechanism is not thoroughly clear. Thereby, we are intrigued by the potential functions of chaperones or other chaperone like additives in improving the performance of DNA polymerase. However, how to identify and prove the role of additives in aiding the activities of the DNA polymerase needs a rational detectable approach.

Nuclear magnetic resonance (NMR) spectroscopy is a powerful and versatile tool used for protein structure and dynamics analysis. It has the capability to decipher intricate structures and interactions of proteins in real-time, even within its native heterogeneous multiphase working environments, such as in cells and cell-like environments (23). Thus, NMR holds immense promise in biomolecule research, offering profound insights into a wide spectrum of enzyme catalytic processes at the molecular level. This includes the identification of binding sites (hot spots) for specific substrates in the working solution and the exploration of changes of hot spots on both enzyme and its substrates. Here ^{1}H-^{13}C HMQC spectra of Pfu polymerase were measured in the absence and presence of a substrate hairpin DNA to identify the hot spots, which were further confirmed by applying home screened hot-start nanobodies (24). By targeting these hotspots, we identified the chemical chaperone, L-arginine, and the heat shock protein, HSP20 from *Thermococcus kodakaraensis* engaged with Pfu polymerase to bolster its thermal resilience, processivity and enhance its activity in amplifying lengthy DNA fragments. This approach simplifies the exploration of novel additives for Pfu polymerase and demonstrates significant potential for application in other enzymes, thereby expediting progress in enzyme engineering.

MATERIAL AND METHODS

Cloning, expression, and purification of proteins

Genes of Pfu DNA polymerase, Pfu-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 gene of TkHSP20 was codon-optimized and synthesized by Sangon Biotech Co., Ltd. and cloned in fusion with the 10×His-SUMO-tag in a modified pET21a vector using FasHifi Super DNA polymerase (Affinibody Lifescience Co., Ltd.). The genes of 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 LB broth or M9 medium supplemented with 100 mg/ml ampicillin and induced for expression with a final concentration of 0.5 mM isopropyl-D-1-thiogalactopyranoside 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) (41). Cells were harvested by centrifugation at 5000g for 25 min at 4 °C.

For Pfu or Pfu-S, 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 SUMO-TkHSP20, the pellet was resuspended in 50 ml of Lysis Buffer B (pH 9.0, 50 mM Tris-HCl, 300 mM NaCl, 0.1 mM PMSF and 0.01 mg/ml DNase I). For NB2 or NB4, the 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 or Pfu-S, 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 or Pfu-S was eluted with 25 ml Elution Buffer A (pH 8.0, 50 mM Tris-HCl, 50 mM KCl, 300 mM imidazole). SUMO-TkHSP20 was eluted with 25 ml Elution Buffer B (pH 9.0, 50 mM Tris-HCl, 300 mM NaCl, 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 or Pfu-S, 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 or Pfu-S was 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 SUMO-TkHSP20, the eluted fraction was then dialyzed against SEC Buffer A (pH 9.0, 50 mM Tris-HCl, 300 mM NaCl) containing 1 mM DTT. SUMO-TkHSP20 was digested with SUMO protease overnight at 4 °C to get rid of the SUMO-tag. Afterwards, a reverse Ni-NTA column was applied for TkHSP20. The flow-through fractions, which contained the untagged TkHSP20, were collected and fractionated on a size exclusion Superdex-200 column (GE Bioscience) equilibrated in SEC Buffer A. For 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.

NMR Titration Experiments

All NMR experiments of the Pfu polymerase 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 polymerase 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 polymerase in the presence of 20 mM L-arginine or 1 M betaine or 0.05% Tween-20 were recorded to measure whether and where these additives interacted with the Pfu polymerase. Arginine was prepared in the phosphate buffer at pH 7.4 to avoid introducing additional salt ions. 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 (42). The combined chemical shift (δ_{comb}) difference was calculated for each residue using the formula:

$$\Delta\delta_{comb} = \sqrt{\omega_H(\Delta\delta_H)^2 + \omega_C(\Delta\delta_C)^2}$$

Where, $\Delta\delta_H$ and $\Delta\delta_C$ are chemical shift changes (in ppm) in ^{1}H and ^{13}C dimensions, respectively, and ω_H and ω_C are normalization factors (ω_H = 1.00, ω_C = 0.34 for aliphatic or 0.07 for aromatic). The binding affinity of the Pfu polymerase with TkHSP20 was measured *via* the NMR titration. NMR spectra of 50 μM ^{13}C labeled Pfu with a serial of different concentrations of TkHSP20 (0, 25, 50, 100 and 200 μM) were collected at 298K and 313K respectively. The magnitude of normalized CSP for residues aro3 of 50 μM ^{13}C labeled Pfu was plotted as a function of the concentration of TkHSP20 to obtain the K_d (dissociation constant) by fitting the curve according to the published protocol (43).

Fluorescence-based thermal shift assay

The thermal shift experiments were measured *via* QuantStudio 3 Real-Time PCR instrument from Thermo Fisher Scientific. 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 (Sigma) and 10 μM of every individual protein. The signal was collected per second. To measure whether L-arginine and TkHSP20 affect the stability of Pfu, we used the following procedure: Incubation the samples at 95 °C for 120 min. Meanwhile, the fluorescence signals were collected and plotted against the time. The first derivative of the fluorescence signal curve was calculated using GraphPad Prism 8.

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-S 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 or Pfu-S, 4 ng (plasmid DNA) or 50 ng (Alpaca cDNA) or 100 ng (Mouse tail gDNA) or 30 ng (λ DNA) 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 °C for 3 min; For Pfu, followed by 33 thermal cycles of 95 °C for 30 s, 55 °C for 30 s, and 72°C for 45 sec to 10 min depending on the target DNA length(1 kb/min); For Pfu-S, followed by 33 thermal cycles of 95 °C for 15 s, 60 °C for 15 s, and 72 °C 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% 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.

Thermal stability PCR assay

Apo-enzyme (50 nM Pfu) and holo-enzyme (50 nM Pfu supplemented with 100 pg/μl TkHSP20) were incubated at 72 °C for varied periods of time in the presence of the 2× Pfu DNA polymerase reaction buffer (pH 8.8, 100 mM Tris-HCl, 20 mM KCl, 12 mM ammonium sulfate, 4 mM $MgCl_2$). At each time point, 25 μl of the sample was mixed with 10 ng plasmid DNA templates (with target length 10 kb), primers (0.2 μM each), dNTPs (0.2 mM), and finally diluted into 50 μl reactions with ddH_2O. Amplification reactions, electrophoresis and analysis were set up as described above. 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 to estimate the half-life time of the Pfu polymerase in the absence and presence of 100 pg/μl TkHSP20.

Steady-state kinetic analyses

Steady-state kinetic data were collected and analyzed using the EvaGreen-based fluorometric polymerase activity assay (38). A hairpin DNA (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. Different amounts of hairpin DNA from 0.1-2 μM were then used as the template to mixed with dNTPs and 2 nM Pfu DNA polymerase in the presence and absence of 250 pg/μl TkHSP20 respectively. $MgCl_2$ was added to initiate the DNA synthesis at 72°C. 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 of each reaction from the fluorescence value. 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.

RESULTS

Determination of 'hot spots' of the Pfu polymerase

To identify the active sites (hot spots) of the Pfu polymerase, 2D [^{13}C,^{1}H]- heteronuclear multi-quantum coherence (HMQC) spectra of [^{13}C]-labeled Pfu protein was recorded in the absence and presence of a hairpin DNA template (See Table S1). Substantial chemical shift perturbations (CSPs) were observed for six aliphatic and aromatic residues (Fig. 1 and Fig. S1), named 'ali1', ali2, 'aro1', 'aro2', 'aro3', and 'aro4' (Fig. 1B, Fig. S1A), which were defined as potential DNA binding sites. As the chemical shift perturbation may arise from the binding site and the allosteric conformational changed region, we then employed two home-screened hot-start nanobodies (NB2 and NB4) (24) to further confirm the binding hot spots of the hairpin DNA on the Pfu polymerase. Adding unlabeled NB2 or NB4 nanobodies into the ^{13}C labeled Pfu polymerase resulted in substantial CSPs of NMR resonance of a set of peaks (Fig. S1B and C). The presence of NB2 induced significant chemical shift perturbations of the peaks ali1, aro1, aro3, and aro4 (Fig. 1C), while the presence of NB4 caused pronounced chemical shift perturbations of the peaks ali1, aro2, aro3, and aro4 (Fig. 1D). Therefore, we identified the five residue ali1, aro1, aro2, aro3, and aro4 as hotspots serving as probes to mark the active sites of Pfu polymerase

for subsequent experiments, whereas the chemical shift perturbation of ali2 might arise from the conformation changes of the Pfu polymerase upon synthesizing DNA after adding the hairpin DNA.

Betaine and Tween-20 interact with hot spots of the Pfu polymerase to improve its performance.

In the next step, we wanted to know if these identified hot spots could be applied as probes to guide the screening and identification of co-factors or additives for Pfu polymerase to improve its PCR performance. Two well-known PCR additives, Tween-20 and betaine, were investigated if they acted on hot spots of Pfu polymerase. The addition of 0.01-0.1% Tween-20 (Fig. 2C) or 0.5-2.5 M Betaine (N,N,N-trimethylglycine, Fig. 2A) into the Pfu polymerase reaction system could enhance its performance in amplifying different templates with varied length (Fig.S2A and B). Meanwhile, NMR data in Fig. 2 revealed the addition of 0.5% Tween-20 into the Pfu polymerase resulted in CSPs of the residues ali1, aro1, aro3, and aro4 (Fig. 2D and Fig. S3B) and modest perturbations of non-hot-spot residues aro5, aro6 and aro7. Similarly, significant chemical shift perturbations of Pfu polymerase were mainly observed on the hot spot residues aro1, aro2, aro3, and aro4 upon the addition of 1 M betaine, while slight perturbations of residues aro5 and aro6 (Fig. 2B and Fig. S3A) were detected. These results showed both Tween-20 and betaine interacted with Pfu polymerase on the hot spots, strengthening the applicability of these hot spots as probes for identifying co-factors in boosting the Pfu DNA polymerase.

L-arginine enhances the performance of Pfu in long-range PCR

We then continued our study to identify new additives for the Pfu polymerase. L-arginine is a chemical chaperone commonly used as an additive for protein stabilization and refolding (25,26). In our preliminary tests, we found the addition of L-arginine (Fig. 3A) in the concentration range of 1-14 mM had little effect on the amplification of DNA fragments (Fig. S5A), and there was neither previous report on PCR enhancing effect by arginine to the best of our knowledge. Nevertheless, the observation of significant CSPs induced by L-arginine at the hotspots of ali1, aro1, aro2, and aro4 encouraged us for further investigation (Fig. 3B and Fig. S4A). More challenging amplification of lengthy templates (10 kb) was carried out in the presence and absence of L-arginine. The endpoint PCR revealed the presence of L-arginine can significantly improve the yield of the Pfu polymerase in amplifying long fragments in a concentration-dependent way, where ~30% more final PCR products were produced at the concentration of 2 mM L-arginine (Fig. 3C and D).

TkHSP20 boosts Pfu polymerase in amplifying lengthy templates

To further explore the application of chaperones in PCR assays, we cloned and expressed small heat shock protein HSP20 from *Thermococcus kodakaraensis* species (TkHSP20) to study its interaction with the Pfu polymerase, as small heat shock proteins are conserved across species and function in an ATP-independent manner (27-30). The chemical shift perturbations revealed the interaction surface of

TkHSP20 on the Pfu polymerase including residues aro2, aro3 and aro4, covering 3 out 5 of the active sites of the Pfu polymerase. Moreover, the intensity of the residue ali1 decreased significantly upon adding TkHSP20, implying this residue was involved in the interaction with TkHSP20 with an intermediate exchange rate (Fig. 4B). Notably, 1 non-hot spot residue aro5 of the Pfu polymerase was perturbated by adding TkHSP20. The different non-hot spots perturbated by TkHSP20 with betaine and L-arginine suggested they interacted with Pfu polymerase on the same region but not identical sites (Fig. S3-4). PCR assays were then performed to evaluate the effects of TkHSP20. Similar to the L-arginine, the amplification performance of short DNA fragments by Pfu polymerase was not improved significantly by adding TkHSP20. However, TkHSP20 showed significant enhancement of the performance of Pfu polymerase in amplify long DNA fragments. Within the range of concentrations from 0.1 pg/μl to 250 pg/μl, TkHSP20 significantly boosted the amplification of long DNA fragments by the Pfu polymerase (Fig. 4C and D) in a dose dependent manner. The best performance of PCR was achieved at the concentration of 10 pg/μl TkHSP20, where the yield of the final 10 kb PCR products was increased up to ~4 folds. The overall of performance of Pfu polymerase upon adding TkHSP20 is better than that of L-arginine (Fig. S6). Moreover, the addition of TkHSP20 and L-arginine together in the Pfu polymerase PCR systems showed synergistic effects (Fig. S6A and B) and enabled the amplification of 40 kb or longer fragments (Fig. S6C), indicating TkHSP20 and L-arginine improve the performance of Pfu polymerase through a cooperative way.

Both L-arginine and TkHSP20 prevent the heat-induced deactivation of the Pfu polymerase in lengthy fragment PCR

To further investigate the mechanism of TkHSP20 applied in ameliorating Pfu polymerase in long fragment PCR. We first quantified the affinity of TkHSP20 and Pfu polymerase at different temperature. NMR titrations of 5 different concentrations of TkHSP20 (0 μM, 25 μM, 50 μM, 100 μM and 200μM) into 50 μM Pfu polymerase were carried out on a 600 MHz spectrometer at 298K and 313K respectively. The CSPs of residues aro3 were used for global fitting to determine the apparent dissociation constant (K_d). The higher affinity (K_d = 18.6 ± 6.0 μM) of the interaction between TkHSP20 and Pfu polymerase at 313K compared to the binding affinity value measured at 298K (K_d=36.7 ± 4.8 μM), revealed TkHSP20 interacted tighter with Pfu polymerase at a higher temperature. We assumed this tighter binding of TkHSP20 with the Pfu polymerase at higher temperatures served as preventing the heat-induced denaturation of Pfu polymerase (Fig. 5A and Fig. S7). To confirm this, a protein thermal shift (PTS) assay of Pfu polymerase was performed by continuous heating of Pfu polymerase at 98°C with SYPRO Orange dye in the presence and absence of TkHSP20, where protein denaturation is monitored *via* an increase in fluorescence of SYPRO Orange dye due to its binding with the heat induced exposure of hydrophobic residues. Pfu polymerase in the presence of TkHSP20 showed an excellent tolerance of heating for 120 mins, with almost no increasement of the SYPRO Orange fluorescence intensity, while

Pfu alone showed a substantial denaturation with ~2 times increased SYPRO Orange fluorescence intensity (Fig. 5D). The plot of the first derivative of the protein thermos shift data clearly showed TkHSP20 delayed the rate of the denaturation of Pfu polymerase (Fig. 5D and 5E). Data of the thermal shift assay for Pfu mixed with L-arginine also revealed that L-arginine protected Pfu polymerase from heat-induced denaturation as well (Fig. S8). Furthermore, in the accelerated ageing test of Pfu polymerase at 72°C, the addition of TkHSP20 extended the half-life of Pfu polymerase activity in amplifying a 10 kb fragment from 1.4 hour to 2.8 hour (Fig. 5C).

TkHSP20 enhances the processivity of the Pfu polymerase

The processivity of a DNA polymerase is commonly indicative of its synthesis speed, rate, and its affinity for substrates. It refers to the quantity of nucleotides processed during a singular binding event. Thus, it could be indirectly measured by recording the amount of newly synthesized double-strand DNA in a given time (40). Here, we used the EvaGreen fluorescence to indicate the amount of the newly incorporated nucleotides in the hairpin DNA template during the whole PCR reaction (Fig. 5F). The final products were loaded on an agarose gel. A clear single target band excluded the contribution of the EvaGreen fluorescence signals from unspecific double-strand DNA by-products (Fig. S9F). The faster increasement of the EvaGreen fluorescence in a given initial reaction time (4 min) in Fig. 6B and Fig. S9 for the PCR with Pfu polymerase in the presence of 250 pg/ul TkHSP20, compared to the control group with buffer, revealed that TkHSP20 enhanced the processivity of the Pfu polymerase. The fit of the enzyme kinetic data with the Michaelis-Menten equation also showed a significantly reduced K_m value 0.26±0.05 μM for the group with 250 pg/ul TkHSP20 compared to the K_m value (0.45±0.07 μM) of the group with buffer. This result further implied the presence of chaperone TkHSP20 not only stabilized the Pfu polymerase but also improved its affinity with the substrate and enhanced the processivity of the Pfu polymerase (Fig. 6C-D). DNA polymerases with high processivity prove advantageous in amplifying lengthy templates, sequences featuring secondary structures and high GC content, and under conditions containing PCR inhibitors. Therefore, our previously observed PCR enhancing effects by TkHSP20 could also be stemmed from the improved processivity of the Pfu polymerase.

DISCUSSION

Although plenty of PCR-derived techniques have been devised and widely adopted across various domains of molecular biology and biotechnology. The limitations related to long-range amplification inherent in current PCR methodologies impede a broader utilization of these related techniques. The challenges in amplifying lengthy DNA fragments involve several factors, which may intensify with more complex templates (31). Firstly, prolonged heating during PCR can compromise the activity of

DNA polymerases, which must remain catalytically active throughout the PCR process to support DNA amplification. Secondly, high temperatures can also lead to the degradation of DNA or newly synthesized amplicons to the point where templates become severely fragmented for amplification (32). Thirdly, as the length of DNA increases, the likelihood of forming secondary structures and premature amplicon increases which may further hinder PCR (33). Finally, long-range PCR is sensitive to even minor variations in reaction conditions, which may not always be obvious and, thus, are challenging to detect. Developing a long PCR-based assay can be a burdensome and unrewarding endeavour (34). Here, we employed NMR spectroscopy to identify both TkHSP20 and L-arginine could improve the activity of the Pfu polymerase in amplifying long DNA fragments. Especially, TkHSP20 almost completely prevents the denaturation of the Pfu polymerase at 98 ℃ for 2 hours, significantly enhancing Pfu's thermostability and its processivity. The chemical L-arginine, which can not only stabilize protein, but also nucleic acid (35). Thus, the combined usage of TkHSP20 and L-arginine showed a more significant enhancement in amplifying the long DNA fragments (Fig. S6A and B). To overcome the problem of unwanted secondary structure formation and premature amplicon, a recent suppression thermo-interlaced (STI) PCR method was developed to repress the amplification of smaller non-specific products (36). Using the STI-PCR method, we observed clear amplified PCR bands of 40 kb DNA fragments with the presence of TkHSP20 and L-arginine, while the control group failed by using solely Pfu polymerase (Fig. S6C). Additionally, the increased stability of the DNA polymerase by TkHSP20 also showed an obvious potential in the application for storage and handling of DNA polymerase, which may also be extended to use as an additive for preserving and transporting other heat-sensitive enzymes.

The NMR-based approach for discovering new additives or co-factors for DNA polymerase in our work is innovative in its clarity and generalizability. This versatile method can be extended to optimize other enzymes, including co-factor identification and buffer optimization. It doesn't rely on the pre-determined assignment information. Measurements of ^{1}H-^{13}C moieties extended the size of proteins suitable for NMR spectroscopy up to Mega-Dalton (44), enabling the identification of active sites of high-molecular-weight enzymes. CSPs of NMR peaks of the target enzyme upon adding its corresponding substrate or the activity blocking antibody aid the determination of hot spots of the enzyme. This NMR guided approach provides the experimentally determined hots spot and it is not dependent on the structural data or bioinformatics, comparing to the current computational methods for predicting hot spots. Furthermore, these determined hot spots could not only be used as probes for rational optimization of the target enzyme involved reactions, but also enable the full potential of directed enzyme evolution with the assignment information (45). Last but not the least, the newly identified TkHSP20 and L-arginine in our work facilitate the application of Pfu polymerase in the lengthy DNA fragment PCR, benefiting various cloning, functional analysis, and nanopore based long fragment sequencing and representing a significant technological advancement that promises to accelerate the fields of synthetic biology, molecular biology, and biotechnology.

DATA AVAILABILITY

Data as supplementary files to be published alongside the article.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR online.

AUTHOR CONTRIBUTIONS

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

ACKNOWLEDGEMENTS

We thank Ziming Liu and Jingwen Liu from Qinhe Life Science Ltd., Wuhan 430000, China for the critical reading of the manuscript and technical help.

FUNDING

This work was supported by National Key R&D Program of China 2018YFE0202301, 2018YFE0202300, National Natural Sciences Foundation of China grants 22174151 and 21991080, the Strategic Priority Research Program of the Chinese Academy of Sciences XDB0540000, and Hubei Provincial Natural Science Foundation of China 2023AFA041.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

1. Moore,P. (2005) Replicating success. *Nature*, **435**, 235.
2. Shendure,J., Balasubramanian,S., Church,G.M., Gilbert,W., Rogers,J., Schloss,J.A. and Waterston,R.H. (2017) DNA sequencing at 40: past, present and future. *Nature*, **550**, 345–353.
3. Kevadiya,B.D., Machhi,J., Herskovitz,J., Oleynikov,M.D., Blomberg,W.R., Bajwa,N., Soni,D., Das,S., Hasan,M., Patel,M., *et al.* (2021) Diagnostics for SARS-CoV-2 infections. *Nat. Mater.*, **20**, 593–605.
4. Gibson,D.G., Young,L., Chuang,R.Y., Venter,J.C., Hutchison,C.A. and Smith,H.O. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods*, **6**, 343–345.
5. Cheng,S., Chang,S.Y., Gravitt,P. and Respess,R. (1994) Long PCR. *Nature*, **369**, 684–685.
6. Cheng,S., Fockler,C., Barnes,W.M. and Higuchi,R. (1994) Effective amplification of long targets from cloned inserts and human genomic DNA. *Proc. Natl. Acad. Sci. U.S.A.*, **91**, 5695–5699.
7. Chua, E. W., Maggo, S. and Kennedy, M. A. (2017). Long Fragment Polymerase Chain Reaction. *Methods Mol. Biol.*, **1620**, 65–74.
8. Wang,Y., Prosen,D.E., Mei,L., Sullivan,J.C., Finney,M. and Horn,P.B.V. (2004) A novel strategy to engineer DNA polymerases for enhanced processivity and improved performance in vitro. *Nucleic Acids Res.*, **32**, 1197–1207.
9. Dietrich,J., Schmitt,P., Zieger,M., Preve,B., Rolland,J.-L., Chaabihi,H. and Gueguen,Y. (2002) PCR performance of the highly thermostable proof-reading B-type DNA polymerase from Pyrococcus abyssi. *FEMS Microbiol. Lett.*, **217**, 89–94.
10. Epstein,C.J., Goldberger,R.F. and Anfinsen,C.B. (1963) Genetic control of tertiary protein structure–studies with model systems. *Cold Spring Harb. Sym.*, **28**, 439–449.
11. Bukau,B.,Weissman,J. and Horwich A. (2006) Molecular chaperones and protein quality control. *Cell*, **125**, 443–451.
12. Hinault,M.P., Ben-ZviA. and Goloubinoff,P. (2006) Chaperones and proteases: cellularfold-controlling factors of proteins in neurodegenerative diseases and aging. *J. Mol. Neurosci.*, **30**, 249–265.
13. Zolkiewski,M. (1999) ClpB cooperates with DnaK, DnaJ, and GrpE in suppressing protein aggregation. A novel multi-chaperone system from Escherichia coli. *J. Biol. Chem.*, **274**,28083–28086.
14. Goloubinoff,P., Mogk,A., Ben-ZviA., Tomoyasu,T. and Bukau B. (1999). Sequential mechanism of solubilization and refolding of stable protein aggregates by a bichaperone network. *Proc. Natl. Acad. Sci. U.S.A.*, **96**, 13732–13737.
15. Huang,Y., Wen,J., Ramirez,L.M., Gümüşdil,E., Pokhrel,P., Man,V.H., Ye,H., Han,Y., Liu,Y., Li,P., Su,Z., Wang,J., Mao,H., Zweckstetter,M., Perrett,S., Wu,S. and Gao,M. (2023) Methylene blue accelerates liquid-to-gel transition of tau condensates impacting tau function and pathology. *Nat. Commun.*, **14**, 5444.
16. Fiedorczuk,K. and Chen,J. (2022) Mechanism of CFTR correction by type I folding correctors. *Cell*, **185**, 158–168.
17. Jin,Y., Yu,G., Yuwen,T., Gao,D., Wang,G., Zhou,Y., Jiang,B., Zhang,X., Li,C., He,L. and Liu,M. (2022) Molecular Insight into the Extracellular Chaperone Serum Albumin in Modifying the Folding Free Energy Landscape of Client Proteins. *J. Phys. Chem. Lett.*, **13**, 2711–2717.
18. Ahmed,R., Huang,J., Weber,D.K., Gopinath,T., Veglia,G., Akimoto,M., Khondker,A., Rheinstädter,M.C., Huynh,V., Wylie,R.G., Bozelli,J.C.,Jr, Epand,R.M. and Melacini,G. (2020) Molecular Mechanism for the Suppression of Alpha Synuclein Membrane Toxicity by an Unconventional Extracellular Chaperone. *J. Am. Chem. Soc.*, **142**, 9686-9699.
19. Abu Al-Soud,W. and Rådström,P. (2000) Effects of Amplification Facilitators on Diagnostic PCR in the Presence of Blood, Feces, and Meat. *J. Clin. Microbiol.*, **38**, 4463–4470.

20. Höss,M. and Pääbo,S. (1993) DNA extraction from Pleistocene bones by a silica-based purification method. *Nucleic Acids Res.*, **21**, 3913–3914.
21. Farell,E.M. and Alexandre,G. (2012) Bovine serum albumin further enhances the effects of organic solvents on increased yield of polymerase chain reaction of GC-rich templates. *BMC Res. Notes*, **5**, 257.
22. Zhang,Y., Li,X., Zou,R., Xue,Y., Lou,X. and He,M. (2014) Bovine thrombin enhances the efficiency and specificity of polymerase chain reaction. *BioTechniques*, **57**, 289–294.
23. Wang,Y., Schwieters,C. D. and Tjandra,N. (2012) Parameterization of solvent-protein interaction and its use on NMR protein structure determination. *J. Magn. Reson.*, **221**, 76–84.
24. Tao,Z., Zhao,X., Wang,H., Zhang,J., Jiang,G., Yu,B., Chen,Y., Zhu,M., Long,J., Yin,L., Zhang,X., Liu,M. and He,L. biorXiv https://doi.org/10.1101/2023.02.15.528753, 15 February 2023, pre-print: not peer-reviewed.
25. Lange,C. and Rudolph,R. (2009) Suppression of Protein Aggregation by L-arginine. *Curr. Pharm. Biotechnol.*, **10**, 408–414.
26. Shiraki,K., Kudou,M., Fujiwara,S., Imanaka,T. and Takagi,M. (2002) Biophysical Effect of Amino Acids on the Prevention of Protein Aggregation. *J. Biochem.*, **132**, 591–595.
27. Finka,A., Mattoo,R.U.H. and Goloubinoff,P. (2016) Experimental Milestones in the Discovery of Molecular Chaperones as Polypeptide Unfolding Enzymes. *Annu. Rev. Biochem.*, **85**, 715–742.
28. Muchowski,PJ. and Clark,J.I. (1998) ATP-enhanced molecular chaperone functions of the small heat shock protein human αB crystallin. *Proc. Natl. Acad. Sci. USA.*, **95**, 1004–1009.
29. Biswas,A. and Das,K.P. (2004) Role of ATP on the Interaction of α-Crystallin with Its Substrates and Its Implications for the Molecular Chaperone Function. *J. Biol. Chem.*, **279**, 42648–42657.
30. Stengel,F., Baldwin,A.J., Painter,A.J., Jaya,N., Basha,E., Kay,L.E., Vierling,E., Robinson,C.V. and Benesch,J.L.P. (2010) Quaternary dynamics and plasticity underlie small heat shock protein chaperone function. *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 2007–2012.
31. Kalle,E., Kubista,M. and Rensing,C. (2014) Multi-template polymerase chain reaction. *Biomol. Detect. Quantif.*, **2**, 11e29.
32. Cheng,S., Chen,Y., Monforte,J.A., Higuchi,R. and Van,H.B. (1995) Template integrity is essential for PCR amplification of 20- to 30-kb sequences from genomic DNA. *PCR Methods Appl.*, **4**, 294e298.
33. Fan,H., Wang,J., Komiyama,M. and Liang,X. (2019) Effects of secondary structures of DNA templates on the quantification of qPCR. *J. Biomol. Struct. Dyn.*, **37**, 2867–2874.
34. Briscoe,A.G., Goodacre,S., Masta,S.E., Taylor,M.I., Arnedo,M.A., Penney,D., Kenny,J. and Creer,S. (2013) Can long-range PCR Be used to amplify genetically divergent mitochondrial genomes for comparative phylogenetics? A case study within spiders (Arthropoda: araneae). *PLoS One*, **8**, e62404.
35. Nicholson,D.A., Sengupta,A., Sung,H.L. and Nesbitt,D.J. (2018) Amino Acid Stabilization of Nucleic Acid Secondary Structure: Kinetic Insights from Single-Molecule Studies. *J. Phys. Chem. B*, **122**, 9869–9876.
36. Zhao,Z., Xie,X., Liu,W., Huang,J., Tan,J., Yu,H., Zong,W., Tang,J., Zhao,Y., Xue,Y., Chu,Z., Chen,L. and Liu,Y.G. (2022). STI PCR: An efficient method for amplification and de novo synthesis of long DNA sequences. *Mol Plant*, **15**, 620–629.
37. Green,M.R., Sambrook,J. Long and Accurate Polymerase Chain Reaction (LA PCR). (2019) *Cold Spring Harb Protoc*., **2019**.
38. Mao,F., Leung,W.Y. and Xin,X. (2007) Characterization of EvaGreen and the implication of its physicochemical properties for qPCR applications. *BMC Biotechnol*, **7**, 76.
39. Tunyasuvunakool,K., Adler,J., Wu,Z., Green,T., Zielinski,M., Žídek,A., Bridgland,A., Cowie,A., Meyer,C., Laydon,A., Velankar,S., Kleywegt,G.J., Bateman,A., Evans,R., Pritzel,A., Figurnov,M., Ronneberger,O., Bates,R., Kohl,S.A.A., Potapenko,A., Ballard,A.J., Romera-Paredes B., Nikolov,S., Jain,R., Clancy,E., Reiman,D., Petersen,S., Senior,A.W., Kavukcuoglu,K., Birney,E., Kohli,P., Jumper,J. and Hassabis,D. (2021) Highly accurate protein structure prediction for the human proteome. *Nature*, **596**, 590-596.

40. Peng,Q., Xie,Y., Kuai,L., Wang,H., Qi,J., Gao,G.F. and Shi,Y. Structure of monkeypox virus DNA polymerase holoenzyme. (2023) *Science*, **379**, 100-105.
41. He,L., Lührs,T. and Ritter,C. (2015) Solid-state NMR resonance assignments of the filament-forming CARD domain of the innate immunity signaling protein MAVS. *Biomol. NMR Assign.*, **9**, 223–227.
42. Vranken,W. F., Boucher,W., Stevens,T. J., Fogh,R.H., Pajon,A., Llinas,M., Ulrich,E.L., Markley,J.L., Ionides,J. and Laue,E.D. (2005) The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins*, **59**, 687–696.
43. Fielding,L. (2003) NMR methods for the determination of protein-ligand dissociation constants. *Curr. Top. Med. Chem.*, **3**, 39–53.
44. Mainz,A., Religa,T.L., Sprangers,R., Linser,R., Kay,L.E. and Reif,B. (2013) NMR spectroscopy of soluble protein complexes at one mega-dalton and beyond. *Angew. Chem. Int. Ed. Engl.*, **52**, 8746-8751.
45. Bhattacharya,S., Margheritis,E. G., Takahashi,K., Kulesha,A., D'Souza,A., Kim,I., Yoon,J. H., Tame,J.R.H., Volkov,A.N., Makhlynets,O.V. and Korendovych,I.V. (2022) NMR-guided directed evolution. *Nature*, **610**, 389–393.

TABLE AND FIGURES LEGENDS

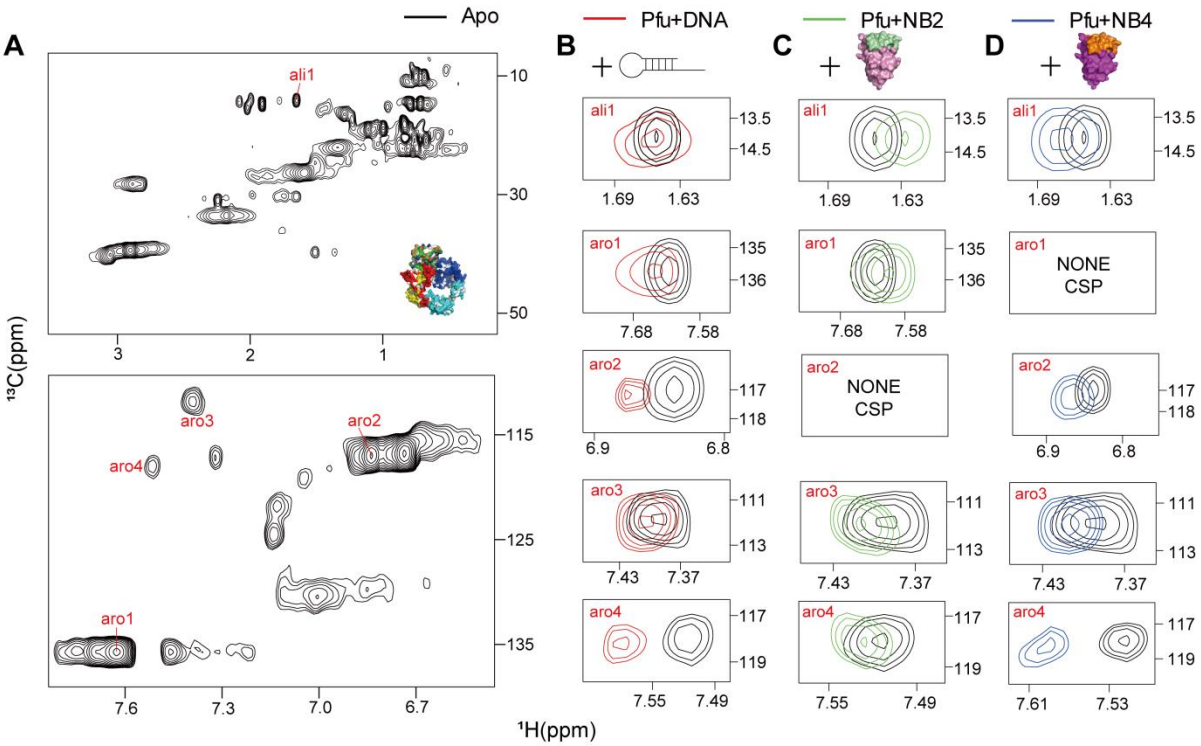


Figure 1. Determination of active 'hot spots' of Pfu by NMR titration. (**A**) Two-dimensional [^{13}C,^{1}H]-HMQC spectra of 50 µM [^{13}C]-labeled Pfu polymerase acquired on the 600 MHz spectrometer at 298 K, where the active hot spots involved in interactions with the hairpin DNA are labeled in red (ali1, aro1, aro2, aro3, and aro4). (**B-D**) Enlarged view of NMR peaks of hot spot residues, which show chemical shift perturbations upon adding 50 µM hairpin DNA (**B**), 75 µM NB2 (**C**) and 75 µM NB4 (**D**).

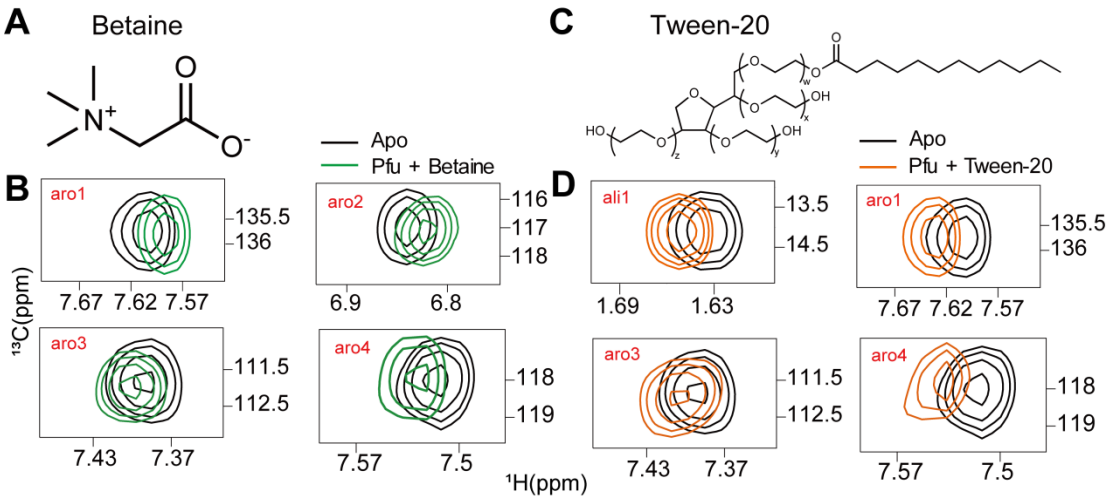


Figure 2. Both betaine and Tween20 interact with active sites of the Pfu polymerase. (**A**) Chemical structures of betaine. (**B**) Plots of chemical shift perturbations of 'hot spots' of 50 μM [^{13}C]-labeled Pfu upon adding 1M betaine. (**C**) Chemical structures of Tween-20. (**D**) Plots of chemical shift perturbations of 'hot spots' of 50 μM [^{13}C]-labeled Pfu upon adding 0.05% Tween-20.

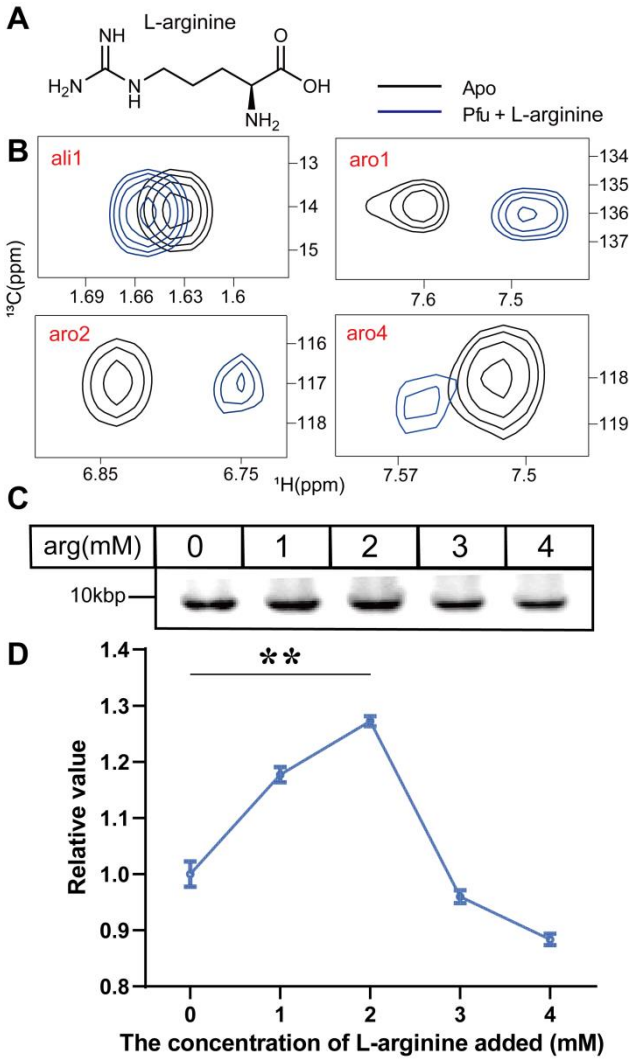


Figure 3. L-arginine enhances the performance of Pfu polymerase in the long-range PCR. (**A**) Chemical structures of L-arginine. (**B**) Plots of chemical shift perturbations of 'hot spots' of 50 μM [^{13}C]-labeled Pfu upon adding 20 mM L-arginine. (**C**) PCR amplification of 10 kb plasmid DNA by the Pfu polymerase in the presence of different concentrations of L-arginine from 1 mM to 4 mM. PCR assays were performed as described in the method, with the following L-arginine concentrations: 0, 1, 2, 3 and 4 mM (lane 1-4) (**D**) Intensities of gel bands were quantified and plotted against the concentration of L-arginine. Results expressed as mean ± SD of 3 independent biological replicates. Data were analyzed using unpaired Student's test, **$P < 0.01$.

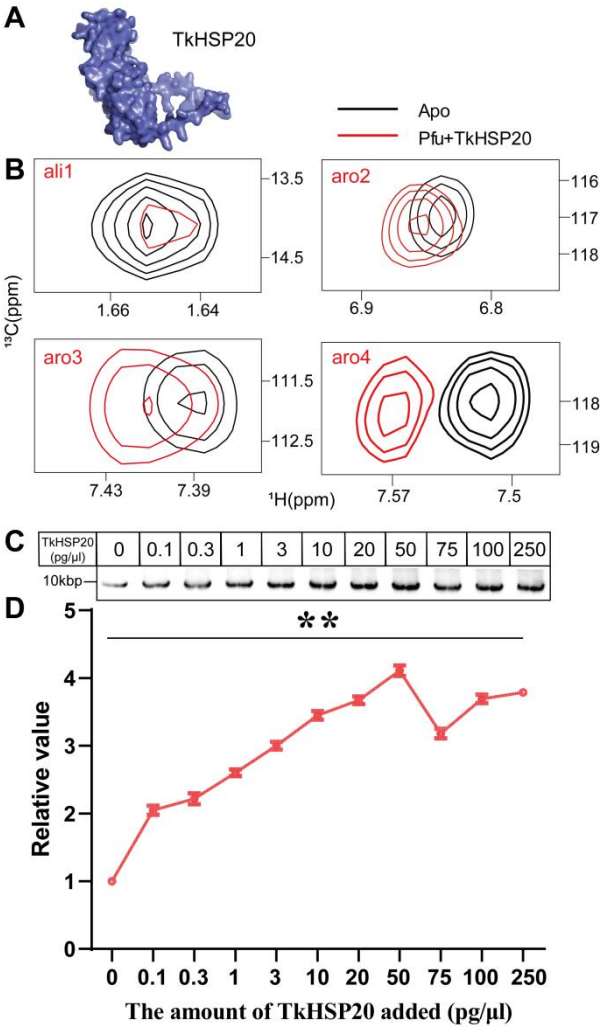


Figure 4. TkHSP20 improves the performance of Pfu polymerase in long-range PCR. (**A**) Cartoon representation of TkHSP20 generated by AlphaFold2 (39). (**B**) Plots of chemical shift perturbations of 'hot spots' of 50 µM [^{13}C]-labeled Pfu upon adding 50 µM TkHSP20. (**C**) PCR amplification of 10 kb plasmid DNA by the Pfu polymerase in the presence of different concentrations of TkHSP20 from 0.1 pg/µl to 250 pg/µl. The PCR assays were performed as described in the methods, with the following TkHSP20 concentrations: 0, 0.1, 0.3, 1, 3, 10, 20, 50, 75, 100, and 250 pg/µl (lane 1-11) (**D**) Intensities of gel bands were quantified and plotted against the concentration of TkHSP20. Results expressed as mean ± SD of 3 independent biological replicates. Data were analyzed using unpaired Student's test, **$P < 0.01$.

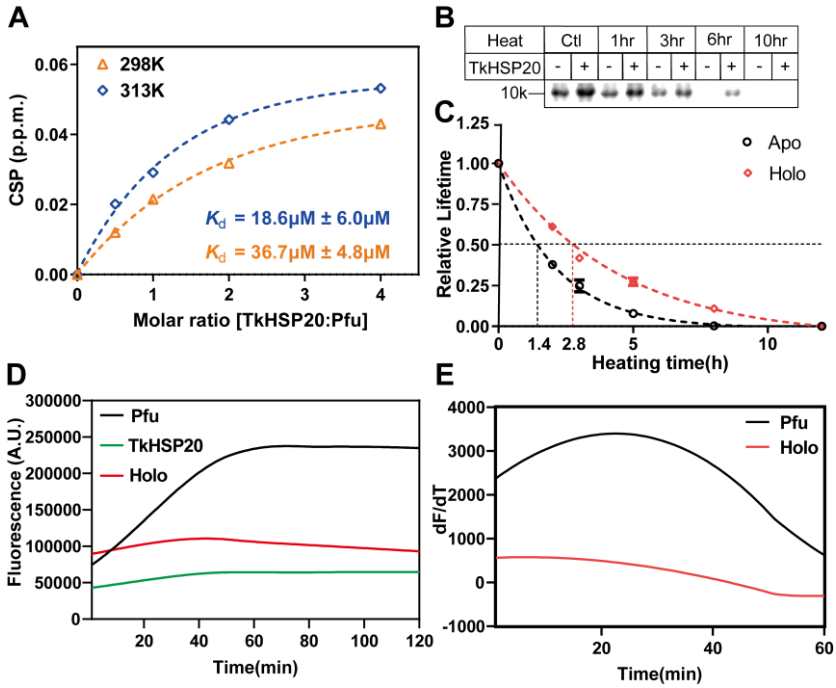


Figure 5. TkHSP20 improves the thermostability of the Pfu polymerase. (**A**) The magnitude of normalized CSP at 298K (orange) and 313K (blue) for the residue aro3 of 50 μM Pfu was plotted as a function of the molar concentration ratio of TkHSP20 and Pfu polymerase. Binding K_d was fitted for aro3. Mean K_d ± S.D. was 36.7 ± 4.8 μM (298K) or 18.6 ± 6.0 μM (313K). (**B**) PCR amplification of 10 kb DNA fragment, by using the Pfu polymerase after heating at 72 ℃ for 0, 1, 3, 6 and 10 hours (lane 1-2, lane 3-4, lane 5-6, lane 7-8, lane 9-10) in the presence and absence of 100 pg/μl TkHSP20 respectively. (**C**) Intensities of gel bands were quantified with Image J and plotted against the time of Pfu polymerase heated at 72 ℃. The lifetime of Pfu polymerase is represented by its activity in amplifying 10 kb DNA fragments. Results expressed as mean ± SD of 3 independent biological replicates. (**D** and **E**) Protein thermal shift (PTS) assay of Pfu polymerase in the presence (red) and absence (blank) of 30μM TkHSP20. The commercial dye SYPRO Orange was used to monitor the thermodynamic stability of the Pfu polymerase during the continuous heating at 98 ℃ for 120 mins. The first derivative of the SYPRO Orange fluorescence is plotted in panel (**E**).

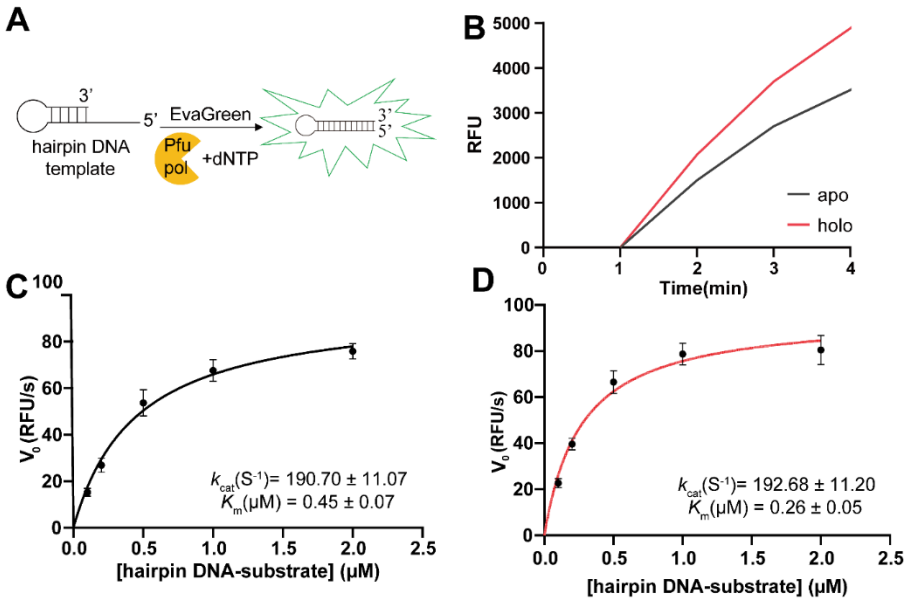


Figure 6. TkHSP20 enhances the processivity of the Pfu polymerase (**A**) Schematic overview of the EvaGreen-Based fluorometric polymerase activity assay. (**B**) Relative fluorescence values of EvaGreen plotted versus the reaction time in the absence (apo, black) and presence (holo, red) of 250 pg/µl TkHSP20. Assays contained 2 nM of Pfu and a concentration of 0.2 µM hairpin DNA template. (**C** and **D**) Initial reaction rates (V_0) of the Pfu polymerase were plotted versus different concentrations of the hairpin DNA substrate in the absence (**C**) and presence (**D**) of 250 pg/µl TkHSP20. Assays contained 2 nM of Pfu and different concentrations (0.1, 0.2, 0.5, 1 and 2 µM) of hairpin DNA templates respectively. The error bars represent the standard error of the mean from two biological replicates with three technical replicates for each biological replicate. The data was fitted to the Michaelis–Menten kinetics model.