Main Manuscript for

Functional Hotspots Identification *via* a Hybrid NMR-Computational Approach Facilitating Directed Evolution of Large Enzyme

Yihao Chen[a], Zhou Gong[a,b], Xiaoling Zhao[c,d], Ali Raza[a,b], Mingjun Zhu[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 Spectroscopy and Imaging, 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

Email: helichun@apm.ac.cn

Author Contributions: Y.C., M.L. 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.

Competing Interest Statement: The authors declare no conflict of interest.

Classification: Biological Sciences; Biochemistry.

Keywords: Directed evolution, NMR Spectroscopy, DNA Polymerase, Functional hotspots

This PDF file includes:

Main Text
Figures 1 to 4

Abstract

Directed evolution has transformed enzyme engineering, yet large, multi-domain proteins remain difficult to evolve due to combinatorial sequence complexity and dynamic conformational exchanges. Here, we present a hybrid NMR-computational strategy that identifies functional hotspots in large enzyme by combining NMR-derived chemical shift perturbations with AlphaFold3-docked distance constraints. By utilizing side-chain methyl and aromatic probes overcome the size limitations of NMR spectroscopy for large proteins. Meanwhile the integrated approach circumvents the need for exhaustive complete NMR assignments while providing residue-specific insights into catalytic hotspots. Using the ~90 kDa DNA polymerase from *Pyrococcus furiosus* as a model, we assigned spatially distinct hotspots and applied saturation mutagenesis on these sites, yielding substantial improvements in multiple catalytic functions. Our results demonstrate the feasibility of applying NMR spectroscopy in targeted evolution of large enzymes and establish a potential approach for focusing mutational searches, offering a powerful route toward efficient biocatalyst optimization.

Significance Statement

Directed evolution has revolutionized enzyme engineering, yet its application to large enzymes remains severely limited by the immense combinatorial sequence space. We developed a hybrid NMR-computational approach that overcomes complete *de novo* resonance assignment and utilizes side-chain methyl and aromatic probes to resolve NMR spectroscopy size limitation for large proteins. This approach allowed us to efficiently pinpoint functional hotspots in a ~90 kDa DNA polymerase from *Pyrococcus furiosus.* By restricting mutational sampling to only 0.5% of residues, we generated variants with substantial improvements in activity and substrate tolerance. This approach overcomes the need for exhaustive screening and provides a transformative platform for engineering large enzymes in industrial biocatalysis and synthetic biology.

Main Text

Introduction

Native enzymes, refined over hundreds of millions of years of evolution, possess functional cores optimized for organismal survival, however, they are not necessarily optimized for maximal catalysis of specific chemical reactions under industrial conditions (1–4). Consequently, researchers are no longer limited to discovering natural enzymes but are actively engaged in engineering or designing enzymes tailored for specific applications (5). To this end, directed evolution has emerged as a powerful strategy to address key limitations such as insufficient catalytic efficiency, narrow substrate scope, and poor stability under non-physiological conditions by enhancing critical biocatalyst properties including stability, activity, and selectivity (6–10). These advances have facilitated the broad application of engineered enzymes, thereby revolutionizing fields such as biotechnology and biomedical research (11–13). Developing broadly applicable methodologies for directed enzyme engineering is therefore a significant pursuit. However, evolving native enzymes remains challenging due to their high molecular weight, multi-domain architectures, and complex substrates (14,15). Even when the substrate binding pocket is known, it still takes a lot of time to mutagenize all the residues that cover the pocket, not to mention the combinatorial explosion that occurs. (16). Moreover, enzymes depend on dynamic residue interactions beyond static active sites. These features avoid conventional structural analysis (17–21). Therefore, without high throughput screening capabilities, experimental interrogation of exhaustive mutant libraries demands prohibitive resources, particularly when addressing the combinatorial complexity of large enzymes (9,21,22). Therefore, narrowing sequence search spaces is supreme, driving substantial efforts to develop methods for designing focused libraries to guide directed evolution (19–21,23,24).

In this context, nuclear magnetic resonance (NMR) spectroscopy provides a complementary strategy by enabling direct analysis of catalytic parameters and residue specific chemical environments under catalytic conditions (17,19,25). By exploring enzymes in their apo- and holo- states in solution, NMR can capture functional hotspots, thereby focusing evolutionary efforts on catalytic relevant residues and facilitating directed evolution of enzymes (19,20). Moreover, NMR spectroscopy could also resolve residue specific chemical environments of enzymes under catalytically relevant conditions with cofactors (26). 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 (27–29), 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. Nevertheless, the assignment of these hot spots is challenging, especially for large enzymes, where fast transverse relaxation rates make scalar magnetization transfer inefficient (30). Consequently, there is an urgent need for more effective assignment methods to facilitate NMR analysis of hot spots of large enzymes.

Here, we developed a hybrid NMR-computational approach for mapping functional hotspots and guiding directed evolution of large enzymes. This method integrates NMR derived chemical shift perturbations (CSPs) with computationally derived distance constraints. By matching CSPs to these distance constraints and deriving structural insights, it predicts and assigns key functional hotspots, providing an alternative while complete NMR assignment is impractical. The high-fidelity DNA polymerase from the thermophilic archaeon *Pyrococcus furiosus* (PfuPol) with molecular weight ~90 kDa was applied as a model system (31). Hot spots across spatially distinct regions of PfuPol were identified, effectively narrowing sequence search space. Saturation mutagenesis of identified hot spots yielded substantial improvements in separate biocatalytic functions, demonstrating the feasibility and effectiveness of the integrative approach in directed evolution of large enzymes, paving the way for new fundamental studies on enzyme functions and evolution.

Results

Hybrid NMR–Computational Strategy and Workflow. NMR directed evolution of proteins including enzymes usually relied on the complete assignment of the target protein. However, for large enzyme, the complete assignment is hard to achieve. Therefore, to enable the NMR aided identification of functional hotspot, we developed a hybrid NMR-computational approach encompassing five sequential steps: (a)

Determination of catalytic hotspots *via* NMR spectroscopy derived ^{1}H-^{13}C chemical shift perturbation of target enzymes upon adding substrates, neutralization antibodies or cofactors (Fig. 1*A*); (b) Computational prediction of amino acid types based on Biological Magnetic Resonance Bank (BMRB) database (Fig. 1*B*); (c) Integrated interpretation of CSPs with AlphaFold3-docked distance constraints to assign the hotspots. (Fig. 1*C*); (d) Site directed saturation mutagenesis of assigned hot spots for screening (Fig. 1*D*); and (e) Exploration of combinatorial mutations coupled with functional characterization (Fig. 1*E*). We applied this strategy to PfuPol which is central to molecular biology technologies, driving the translation of theoretical insights into applications (32–34).

Identification of Hotspots. To enable NMR chemical shift perturbation experiments, we developed a hairpin DNA substrate (35) for NMR titration with PfuPol (Fig. 1*A*(i) and *SI Appendix*, Fig. S2*A*). Notably, we observed that the hairpin DNA exhibits a concentration-dependent inhibitory effect in conventional PCR reactions (*SI Appendix*, Fig. S2*B*). Therefore, we conclude that at micromolar concentrations, this hairpin DNA effectively occupies the active site of PfuPol in our NMR chemical shift perturbation experiments, thereby facilitating the identification of its functional hotspots. As shown in Fig. 1*A*(iii), (iv)), we quantified CSPs ($\Delta\delta_{comb}$) by calculating standard deviation from the differences in the ^{13}C and ^{1}H chemical shifts of the PfuPol upon binding the substrate analogue. Following the established protocol, 2D [^{13}C,^{1}H]- Heteronuclear Multi-Quantum Coherence (HMQC) spectra of [^{13}C]-labeled PfuPol were recorded both in the absence and presence of the hairpin DNA template (Fig. 2*A* and *SI Appendix*, Fig. S3*A*). Significant CSPs were observed for one aliphatic residue ('ali1') and three aromatic residues ('aro1', 'aro2', 'aro3') (36,37), with CSP magnitudes exceeding three times the standard deviation (>3σ) (Fig. 2*D*).

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 PfuPol, two neutralizing nanobodies (Nb2/Nb4) derived from our prior work (38) were then employed in NMR titration to filter non-catalytically relevant hotspot (Fig. 1*A*(ii) and *SI Appendix*, Fig. S2 *C–E*). Similarly, 2D [^{13}C,^{1}H]-HMQC spectra of [^{13}C]-labeled PfuPol were recorded in the absence and presence of unlabeled nanobodies Nb2 and Nb4, respectively (*SI Appendix*, Fig. S3 *B* and *C*). The presence of the nanobody Nb2 induced notable CSPs (>3σ) of peaks ali1, aro1, and aro3 (Fig. 2 *B* and *E*), while the presence of Nb4 caused pronounced significant CSPs (>3σ) of peaks ali1, aro2, and aro3 (Fig. 2 *C* and *F*). Residues ali1, aro1, aro2, and aro3 revealed communal chemical shift perturbation while adding the hairpin DNA and neutralization nanobody Nb2 or Nb4. Thus, these four residues were identified as hotspots to probe the activity related residues of PfuPol subsequent experiments. The CSP patterns resulting from the binding of PfuPol to Nb2, Nb4, or DNA are presented in the bottom panels of Fig. 2 *A–C*.

Assignment of Hotspots. To achieve assignments of these hotspots, we first determine amino acid types of the identified hotspot residues. Using average chemical shifts and standard deviations from the Biological Magnetic Resonance Data Bank (BMRB) data (39), we generated residue specific chemical shift ranges on the 2D [^{13}C,^{1}H]-HMQC spectrum of [^{13}C]-labeled PfuPol (*SI Appendix*, Fig. S4). Through statistical analysis of ^{1}H and ^{13}C chemical shifts of identified hotspots, potential amino acid type(s) were defined using probability density distributions and Bayesian probability calculations ((Fig. 1*B*(i)), assigning the probability of the amino acid type of identified hotspot (Fig. 1*B*(ii), (iii)). Ali1 exhibits a 78.8% probability corresponding to Ala Cβ versus 21.2% for Met Cε (Fig. 3*G*). Aro1 was unambiguously assigned to His Cε based on its unique chemical shift (Fig. 3*H*). Aro2 shows 70.3% probability for Tyr Cζ, 25.3% for Trp Cδ, and 4.4% for His Cδ (Fig. 3*I*), while aro3 has 85.1% probability for Trp Cδ and 14.9% for His Cδ (Fig. 3*J*).

To further precisely assign these hotspots, we introduce an integrated approach with the experimental data and computational distance restraints. Firstly, we obtained the interaction patterns between different binders (DNA, Nb2, Nb4) with PfuPol from CSPs (Fig. 1*C*(i)). Then, AlphaFold3 (40) was employed to generate the structure of the complex between the experimental hairpin DNA and PfuPol, while those with nanobodies (Nb2 and Nb4) were obtained through molecular docking (41). The top ten lowest-energy structures of the complex of Nb2 and Nb4 with PfuPol were selected respectively. The distance (r) of each residue of PfuPol with the surface of hairpin DNA, Nb2 and Nb4 were calculated for each conformer (Fig. 1*C*(ii) and *SI Appendix*, Fig. S5). Based on previous studies (42,43), the minimal distance between any atom of the perturbed residue and any atom of the binder conformer was assumed to be <6.0 Å. For complex models generated by AlphaFold3 and HADDOCK2.4, a positional uncertainty of 1.65 Å was applied (40,41).

Consequently, an upper distance limit of 7.65 Å was adopted as the reference threshold for binders induced CSPs on PfuPol.

By integrating the probabilistic assessment of amino acid types for each hotspot with their spatial distances to the hairpin DNA, Nb2, and Nb4, a systematic analysis was conducted. For the ali1 hotspot, residue A684 exhibited distances of 4.8 Å, 4.6 Å, and 4.2 Å to the hairpin DNA, Nb2, and Nb4, respectively (Fig. 2*K* and *SI Appendix*, Table S2), which aligns well with the observed CSP pattern. This clear distance profile strongly supported its assignment. In contrast, other candidate residues across all hotspots often showed contradictory trends either displaying significant CSPs despite long distances or showing no notable CSPs despite being in close proximity (*SI Appendix*, Fig. S6*A* and Table S2). Although spatial proximity does not necessarily guarantee CSPs generation, and residues distant from binding interfaces may occasionally display CSPs due to indirect effects, such cases require stronger assumptions and more complex interpretations. The CSP data for A684 were fully explicable without resorting special hypotheses (Fig. 2 *A–C*, bottom panels). Similarly, aro1, aro2, and aro3 were evaluated using the same integrated approach, taken the possibility of each hot spot belonging to a certain amino acid type, CSP patterns, and distances of hot spots to hairpin DNA, Nb2 and Nb4, all this information together (Fig. 1*C*(iii)). The consolidated evidence supported the of ali1 to Cβ of A684 (Fig. 2*K*), aro1 to Cε of H680 (Fig. 2*L*), aro2 to Cζ of Y393 (Fig. 2*M*), and aro3 to Cδ of W624 (Fig. 2*N* and *SI Appendix*, Fig. S6 *L–M*). Detailed assignment parameters are provided in *SI Appendix*, Fig. S6 and Table S2.

Characterization of Assigned Hotspots. All four assigned hotspots (Y393, W624, H680, and A684) were mapped onto the PfuPol structural model (Fig. 3*A*). Notably, Y393 residue located within the conserved "Y-GG/A" motif of a flexible loop bridging the 3'-5' exonuclease and polymerase domains. Although not a direct catalytic residue, this motif modulates the polymerase-exonuclease activity balance (44,45). W624 lies within a loop region in close proximity to DNA. A684 engaged in the DNA minor groove near the active site, while H680 occupied a peripheral position of the PfuPol-DNA interaction cluster, contributing to the local electrostatic potential.

In a further step, we introduced a saturation mutagenesis library for the position Y393, W624, H680, and A684 individually to all other 19 alternatives for the evolution of PfuPol. Mutants of PfuPol was transformed into *E. coli* BL21 strain (Fig. 1*D*(i)). Single colonies were picked from LB agar plates (Fig. 1*D*(ii)) and cultured in 96-deepwell plates (Fig. 1*D*(iii)). Owing to the high thermostability of PfuPol (31), we directly utilized the recombinant *E. coli* cells as the source of PfuPol in the PCR system (Fig. 1*D*(iv)). Thermal pre-lysis of the *E. coli* resuspension reduces degradation of oligonucleotides and dNTPs by components in the lysate. The PCR reaction with *E. coli* cell expressing PfuPol mutants was performed by using a long template for amplification (Fig. 1*D*(v)). Surprisingly, saturation mutagenesis of these hotspot residues resulted in dramatic alterations in catalytic activity, providing strong evidence for the accuracy of the assignments and confirming the functional relevance of these hotspots to PfuPol activity. Notably, for all four assigned residues Y393, H680, and A684, we screened mutants with superior performance, mainly higher yield, than the wide type PfuPol (Fig. 3*D* and *SI Appendix*, Fig. S7*A*) without compromising fidelity (*SI Appendix*, Table S3). In contrast, for W624, only the W624S mutant retained activity comparable to that of the wild-type enzyme (*SI Appendix*, Fig. S7*A*). This outcome demonstrated the identified hotspots were relevant to the activity of PfuPol (19,20).

To elucidate mechanisms underlying enhanced activity and further optimize PfuPol, we determined the kinetic parameters *via* Michaelis–Menten fitting (Fig. 1*E*(iii)) using an EvaGreen-based (46) fluorescent polymerase activity assay (Fig. 1*E*(ii)). These parameters were subsequently analyzed in conjunction with the AlphaFold3-generated structures of PfuPol variants complexes with DNA. For PfuPol-A684Q variant, structural analysis revealed deeper penetration of the glutamine side chain into the DNA minor groove compared to wild-type (Fig. 3 *B* and *C*), suggesting enhanced DNA stabilization during long-template amplification and increased processivity. This result was consistent with a reduced K_M value of 0.34 ± 0.03 μM of PfuPol-A684Q variant compared with the K_M value of 0.45 ± 0.07 μM of PfuPol (Fig. 3*D*). For PfuPol-Y393F, substitution with phenylalanine improved the activity of PfuPol in amplify the long template, whereas non-aromatic mutations abolished polymerase function (*SI Appendix*, Fig. S7*A*), showing the essential role of aromaticity in this position. Structural comparisons with the wild-type PfuPol (*SI Appendix*, Fig. S7 *B* and *C*) revealed subtle conformational rearrangements and altered intramolecular interactions within the flexible loop harboring the Y-GG/A motif, consistent with the observed increased k_{cat} (190.7 ± 11.1 s^{-1} to 258.0 ±

8.22 s^{-1}). Conversely, compared to wild-type, the PfuPol-H680K variant enhanced positive surface electrostatic potential (*SI Appendix*, Fig. S7 *D* and *E*), likely strengthening DNA backbone contacts and reducing K_M (0.45 ± 0.07 to 0.31 ± 0.06 μM), demonstrating dominance of charge over aromaticity at this site. In the absence of DNA, the local loop region containing W624 in the PfuPol-W624S mutant adopts a spatial conformation that is markedly different from that of the wild-type enzyme (RMSD = 3.317 Å) (*SI Appendix*, Fig. S7*F*). Because of DNA binding, the flexible loop in the mutant becomes locked into a specific active conformation that is highly similar to the wild-type structure (RMSD = 0.556 Å) (*SI Appendix*, Fig. S7*G*). This conformational stabilization results in a modest increase in the k_{cat} value of PfuPol-W624S (Fig. 3*D* and *SI Appendix*, Table S3). Collectively, these functional enhancements validate the catalytic relevance of the hotspot assignments.

Exploration of Hotspot Synergy in Catalytic Activity. To elucidate the underlying mechanisms of catalytic enhancement and identify optimal mutants (Fig. 1*E*(i)), we introduced combinatorial mutagenesis of PfuPol, generating variants including three double mutants PfuPol-Y393F/H680K, PfuPol-Y393F/A684Q, and PfuPol-H680K/A684Q and two triple mutants PfuPol-Y393F/H680K/A684Q and PfuPol-Y393F/W624S/A684Q. We prepared these variants and determined their Michaelis-Menten kinetics of rate as a function of substrate concentration (Fig. 4*A*). With the exception of W624S, which proved incompatible in combinatorial mutants (*SI Appendix*, Table S4), the combinatorial mutants exhibited a 2- to 10-fold modulation in catalytic efficiency relative to the wild-type enzyme. Specifically, PfuPol-Y393F/A684Q and PfuPol-Y393F/H680K/A684Q showed a comparable twofold enhancement, with catalytic efficiencies of $(1.10 \pm 0.25) \times 10^9$ $M^{-1}s^{-1}$ and $(1.03 \pm 0.16) \times 10^9$ $M^{-1}s^{-1}$, respectively, compared to the wild-type value of $(0.52 \pm 0.14) \times 10^9$ $M^{-1}s^{-1}$. In contrast, the double mutants PfuPol-Y393F/H680K and PfuPol-H680K/A684Q—each containing the distal side-chain mutation H680K in combination with an additional hotspot mutation—displayed a more substantial improvement of over 10-fold, achieving values of $(5.81 \pm 1.52) \times 10^9$ $M^{-1}s^{-1}$ and $(5.70 \pm 1.00) \times 10^9$ $M^{-1}s^{-1}$, respectively (Fig. 4*B*), indicating strong synergistic effects. Notably, a high correlation was found between catalytic efficiency (k_{cat}/K_M) and the catalytic rate constant (k_{cat}) across the combinatorial mutants (Fig. 4*C*), suggesting that the mutations primarily enhance catalytic turnover rather than substrate binding. While tight substrate binding (a low K_M) can confer high catalytic efficiency in principle, the reaction rate is largely dictated by k_{cat} under the high substrate conditions typical of many applied settings (22).

Discussion

We developed a hybrid NMR-computational approach integrating NMR experiments with computational distance constraints to map functional hotspots and guide the directed evolution of large enzymes. Applying this approach to the multi-domain PfuPol, we successfully assigned four experimentally identified hotspots, demonstrating the feasibility of integrating distance constraints with CSPs analysis for assigning functional hotspots in large enzymes. Directed evolution confirmed the catalytic relevance of these hotspots, reducing reliance on exhaustive combinatorial screening and highlighting the advantage of compact, precisely guided libraries. Critically, this approach eliminates dependency on high-throughput screening. Single-round rapid selection efficiently eliminates deleterious mutations while combination of the single mutant enabling cost-effective identification of optimal variants.

This study provides a streamlined pipeline for identification of hotspots in large native enzymes without complete *de novo* NMR assignments. The use of ^{1}H-^{13}C moieties in NMR spectroscopy extends protein analysis to mega-Dalton complexes (47), enabling active site mapping in high-molecular-weight enzymes. Our approach employs distinct ^{1}H-^{13}C chemical shift signatures of amino acid side chains. When conventional NMR analysis becomes infeasible due to protein size, ^{1}H-^{13}C correlation spectra provide an effective alternative. Given potential spectral overlap challenges, distortionless enhancement by polarization transfer (DEPT) could further help to resolve methyl and methylene spectral overlap (48). Titration with substrates or neutralization nanobodies resolves functional residues *via* CSPs analysis (27,28,49). Betaine, a well-known PCR additive (50,51), was also investigated if they acted on hot spots of PfuPol. The addition of 0.5-2.5 M Betaine (N,N,N-trimethylglycine) into the PfuPol reaction system was shown to enhance its performance in amplifying different templates with varied length (*SI Appendix*, Fig. S8 *C–E*) respectively. Meanwhile, significant CSPs of PfuPol were observed on the hot spot residues aro1, aro2, and aro3 upon the

addition of 1 M betaine (*SI Appendix*, Fig. S8 *A* and *B*) were detected. This result showed if neutralization antibody is not available, co-factors of the target enzyme could also be used to identify the functional hot spots. Additionally, natural-abundance [^{13}C,^{1}H]-HMQC spectra of PfuPol showed a high resolution and signal to noise ratio signals, enabling hotspots detection in unlabeled enzymes (*SI Appendix*, Fig. S9), demonstrating the potential of broad application of our proposed approach.

Distance constraints derived from enzyme-binder complexes correlate with CSPs to strongly assign hotspot residues, drastically narrowing mutational search space (19–21,52,53). In structural bioinformatics, CSPs are widely recognized as a valuable source of medium to low-resolution restraint information (36,37,54). Unlike the precise geometric data provided by nuclear Overhauser effects (NOEs), CSPs offer ambiguous, qualitative, and probabilistic spatial constraints. However, they play an indispensable role in identifying the overall binding mode of protein–ligand complexes. In this work, CSP-derived restraints were used as a coarse-grained filtering step an important advancement when full resonance assignments were unavailable. Although "false negatives" (i.e., proximal residues showing no significant CSPs) can occur, clusters of residues with substantial CSPs are statistically enriched at true binding sites across the protein interface. While data from a single binder may occasionally be misleading due to random effects, the consistent CSP patterns observed for three distinct binders provide multidimensional cross-validation, making the possibility of random matches exceedingly low. This greatly strengthens the reliability of the resulting conclusions. Notwithstanding its limitation in resolving certain residues, the method leverages the abundant data from widely distributed side-chain moieties to enable effective screening for key functional sites on the target protein.

Notably, our assigned hotspots encompass the tyrosine residue within the previously reported "Y-GG/A" motif of Family B polymerases (44,45), which aligns with our Y393 hotspot, demonstrating the method's reliability in identifying residues that indirectly regulate active sites. While A684 directly engages DNA and Y393 modulates activity allosterically, H680 occupies a peripheral interaction cluster. These results demonstrate the capacity of our method to probe residues proximal to functional interfaces. Variants characterization revealed some minor functional trade-offs: Lysine substitution at PfuPol-H680K improved pH stability over wild-type histidine, but arginine substitution abolished activity, indicating electrostatic interactions with DNA must remain balanced within a defined range (*SI Appendix*, Fig. S7*A*). Hotspot-guided evolution generated variants with elevated k_{cat} values, diverging from traditional directed evolution strategies that typically enhance catalytic efficiency by reducing K_M (19). PCR assays showed the PfuPol-Y393F variant enhanced amplification efficiency across diverse template lengths (*SI Appendix*, Fig. S10*A*), PfuPol-H680K improved specificity for select templates (*SI Appendix*, Fig. S10*C*), and PfuPol-A684Q boosted amplification of high-GC templates (*SI Appendix*, Fig. S10*B*). Further enhancement of the activity of PfuPol *via* combinatorial mutagenesis also confirmed the non-redundant roles of identified hotspots, demonstrating the high efficiency of our proposed approach. The W624S mutation does not completely eliminate enzymatic activity, but it increases the flexibility and induces conformational changing in the corresponding loop region under apo conditions (*SI Appendix*, Fig. S7 *F* and *G*). This enhanced flexibility may allow the loop to sample multiple conformational states, making it more challenging to maintain the proper architecture in the absence of substrate. DNA binding appears to compensate for the loss in intrinsic stability caused by the mutation. Thus, W624 may act as a critical "switching" residue that helps maintain the local conformation of PfuPol in the substrate-free state, while the functional catalytic conformation is largely induced and stabilized by substrate binding.

In summary, this hybrid NMR–computational framework markedly broadens the utility of NMR spectroscopy in guiding the directed evolution of large enzymes. By integrating NMR methods with computational analysis, it eliminates the need for full *de novo* NMR assignments, thereby advancing enzyme engineering toward greater catalytic efficiency and functional diversity.

Materials and Methods

Cloning, expression, and purification of proteins. Genes of PfuPol was codon-optimized and synthesized by Sangon Biotech Co., Ltd. (China) and cloned in a modified pET16b vector (Novagen). The mutants of PfuPol were directly obtained through library screening (see Library construction) or generated via site-directed mutagenesis PCR (see *SI Appendix*, Table 1). The genes of nanobodies Nb2 and Nb4 were cloned in a modified pET22b vector containing N-terminal His-tag (38). 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 5 hours at 37 °C. Uniformly [^{13}C]-labeled protein was prepared by growing cells in M9 medium containing [U-^{13}C] glucose (3 g/liter) (55). Cells were harvested by centrifugation at 5000g for 25 min at 4 °C.

For the purification of polymerase, the pellet was resuspended in 50 mL of Lysis Buffer A (50 mM Tris-HCl pH 9.0, 300 mM KCl, 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. 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 25 min. The supernatant was then diluted 6-fold and loaded onto a 5 mL HiTrap Heparin column (GE Bioscience) that was pre-equilibrated with Heparin Buffer (50 mM Tris-HCl pH 8.0, 50 mM KCl). The polymerase was eluted with a 0–500 mM KCl gradient (50 mM Tris-HCl pH 8.0). The eluted protein was immediately buffer exchanged via a PD-10 column (GE Bioscience) to the Storage Buffer (25mM Tris-HCl pH 7.4, 50 mM KCl, 0.1 mM EDTA, 1 mM DTT, 50% glycerol) and stored at –20 °C.

For the purification of nanobody, the cell pellet was resuspended in 50 mL of Lysis Buffer B (20 mM Tris-HCl pH 7.4, 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. The supernatant containing each target protein was loaded onto a Ni-NTA (nitrilotriacetic acid) column (Qiagen) which was pre-equilibrated with Lysis Buffer B. The column was washed with 100 mL of Wash Buffer (Lysis Buffer B with 30 mM imidazole), and eluted with 25 mL Elution Buffer C (20 mM Tris-HCl pH 7.4, 150 mM NaCl, 0.1 mM DTT, 300 mM imidazole). The eluted fraction was then fractionated on a size exclusion Superdex-75 column (GE Bioscience) equilibrated in SEC Buffer (20 mM Tris-HCl pH 7.4, 150 mM NaCl, 0.1 mM DTT, 0.1 mM EDTA) to further purify the proteins. The proteins (Nb2 and Nb4) were concentrated and store at –80 °C for further usage.

NMR spectroscopy. All NMR experiments of the PfuPol were performed in the NMR buffer (20 mM NaH_2PO_4/Na_2HPO_4 pH 7.4, 35 mM NaCl, 0.02% NaN_3, 8% D_2O). 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 PfuPol 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 hotspots. 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 (56). The combined chemical shift perturbation ($\Delta\delta_{comb}$) was calculated for each residue using the formula:

$$\Delta\delta_{comb} = \sqrt{\omega_H(\Delta\delta_H)^2 + \omega_C(\Delta\delta_C)^2} \quad (1)$$

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 ($\omega_H = 1.00$, $\omega_C = 0.34$ for aliphatic or 0.07 for aromatic).

Assignments of hotspots. The peaks from the 2D [^{13}C,^{1}H]-HMQC spectra of PfuPol, showing common chemical shift perturbation upon the titration of hairpin DNA, nanobody Nb2 and Nb4 individually were identified as the hotspots. To assign the hotspots, we first compared the chemical shifts of hotspots with the statistics values of amino acids from the Biological Magnetic Resonance Bank (BMRB) (39) database to calculate the possibility of these hotspots belonging to a certain amino acid type. The probability density function $f(\mathrm{a},\mathrm{b})$ for a hotspot residue belongs to a certain amino acid types is given by:

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

where a is the chemical shift of the hotspot residue in the Hydrogen dimension, b is the chemical shift of the hotspot residue in the carbon dimension, μ_H is the mean chemical shift in the Hydrogen dimension for the

amino acid type, σ_H is the standard deviation of the chemical shift in the Hydrogen dimension for the amino acid type, μ_C is the mean chemical shift in the carbon dimension for the amino acid type, σ_C is the standard deviation of the chemical shift in the carbon dimension for the amino acid type. Using Bayes' theorem, the probabilities that the hotspot belongs to certain amino acid type are:

$$P_x(\mathrm{a},\mathrm{b}) = \frac{f_x(\mathrm{a},\mathrm{b})}{\sum f(\mathrm{a},\mathrm{b})} \tag{3}$$

where $f_x(\mathrm{a},\mathrm{b})$ is the probability density value for the hotspot 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 hotspot residue.

With the possibility of the hotspot belonging to which amino acid, we then applied the integrated methods with AlphaFold3 and HADDOCK2.4 to assign the hotspot. The structure of PfuPol-hairpin DNA complex was generated by AlphaFold3 (40). Based on 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 were docked with PfuPol via HADDOCK with a previously described protocol (41). 100 docking models for each PfuPol-nanobody complex were generated. According to haddock-score, the ensemble of 10 conformations with the lowest energy were selected. Interatomic distances (r) between residues of PfuPol and binding partners, specifically DNA or nanobodies Nb2 and Nb4, were measured using PyMOL. The minimum distance between any atom of the perturbed residue and any atom of the binder conformer was set to less than 6.0 Å, following previous studies (42,43). For complex structures generated by AlphaFold3 and HADDOCK2.4, a positional uncertainty of 1.65 Å was incorporated to account for model variability (40,41). Thus, an upper distance threshold of 7.65 Å was defined as the reference criterion for identifying binder-induced chemical shift perturbations (CSPs) in PfuPol. The distance (r) between each PfuPol residue of the assigned amino acid type with the corresponding binder—hairpin DNA, Nb2, or Nb4—was evaluated against the upper distance threshold of 7.65 Å. Residues that consistently had a distance with hairpin DNA, Nb2, and Nb4 less than 7.65 Å were assigned to the identified hotspot.

Library construction. Saturation mutagenesis library of PfuPol was created via site-directed PCR with wild-type PfuPol 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 393, 680 and 684. Primer sequences are provided in *SI Appendix*, 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.

Library Screening. *E.coli* BL21 (DE3) cells containing PfuPol mutants were plated on LB agar with 100 mg/mL ampicillin. A single colony was picked and inoculated in a 96 deep-well plate (with 2 mL LB medium and 100 mg/mL ampicillin per well), sealed using a breath-easy membrane. 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 3500g for 20 min at 4 °C. Pellets were resuspended in ddH_2O, centrifuged again, and the supernatant was discarded. Cells were resuspended in ddH_2O, and their OD_{600} values were measured. Subsequent stepwise dilution into ddH_2O enabled determination of the optimal *E. coli* cells concentration for PCR reactions using wild-type PfuPol. We established that adding 1 μL of *E. coli* cells suspension (OD_{600} = 0.33) to 20 μL reaction volumes yielded maximal activity. Screening was performed in two stages: first, short-fragment amplification (LucB primer sequences in *SI Appendix*, Table S1) promptly eliminated clones lacking detectable activity; second, long-fragment pressure screening (Long primer sequences in *SI Appendix*, Table S1) identified functionally robust variants. Specifically, first, incubation at 98 °C for 6 min to lyse cells for releasing the highly thermostability of PfuPol proteins. Then, followed by 34 thermal cycles of 98 °C for 15 s, 60 °C for 15 s, and 72 °C for 20 sec to 5 min depending on the target DNA length (1 kb/min). The reaction buffer contained 50 mM Tris-HCl pH 8.8, 10 mM KCl, 2.5 mM $MgCl_2$, 0.2 mM dNTPs, 1M betaine, 0.001% BSA and 0.05% Triton X-100. 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, China). 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 variants were calculated. The heatmap of relative activities was generated using the software GraphPad Prism 8.0 (GraphPad Software, Inc.). Plasmids extracted from the colonies demonstrating improved activity were sequenced by Sanger sequencing (Sangon Biotech Co., Ltd.) to determine the identities of beneficial mutations.

Steady-state kinetic analyses. Steady-state kinetic data were collected and analyzed using the EvaGreen-based fluorometric polymerase activity assay (46). The hairpin DNA (T) (sequence in *SI Appendix*, Table S1) 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 (T) at the indicated concentrations were then used as the template to mixed with dNTPs, 8 nM polymerase respectively. $MgCl_2$ was added to initiate the DNA synthesis at 72 °C. For PfuPol and its variants, the reaction buffer contained 25 mM Tris-HCl pH 8.8, 10 mM KCl, 2.5 mM $MgCl_2$, 1× EvaGreen dye, 0.2 mM 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 software GraphPad Prism 8.0. The reported values are the average of a biological duplicate each with three technical replicates.

Acknowledgments

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.

References

1. R. A. Sheldon, J. M. Woodley, Role of Biocatalysis in Sustainable Chemistry. *Chem. Rev.* **118**, 801–838 (2018).
2. S. L. Lovelock, *et al.*, The road to fully programmable protein catalysis. *Nature* **606**, 49–58 (2022).
3. T. M. Lister, *et al.*, Engineered enzymes for enantioselective nucleophilic aromatic substitutions. *Nature* **639**, 375–381 (2025).
4. L. Shi, *et al.*, Complete Depolymerization of PET Wastes by an Evolved PET Hydrolase from Directed Evolution. *Angew. Chem. Int. Ed.* **62**, e202218390 (2023).
5. F. H. Arnold, Directed Evolution: Bringing New Chemistry to Life. *Angew. Chem. Int. Ed.* **57**, 4143–4148 (2018).
6. S. M. Forget, *et al.*, Evolving a terminal deoxynucleotidyl transferase for commercial enzymatic DNA synthesis. *Nucleic Acids Res.* **53**, gkaf115 (2025).
7. X. Sun, *et al.*, Metabolic pathway assembly using docking domains from type I cis-AT polyketide synthases. *Nat. Commun.* **13**, 5541 (2022).
8. E. L. Bell, et al., Directed evolution of an efficient and thermostable PET depolymerase. *Nat. Catal.* **5**, 673–681 (2022).
9. S. Zhang, et al., Directed evolution of a cyclodipeptide synthase with new activities via label-free mass spectrometric screening. *Chem. Sci.* **13**, 7581–7586 (2022).
10. R. Carr, et al., Directed Evolution of an Amine Oxidase Possessing both Broad Substrate Specificity and High Enantioselectivity. *Angew. Chem. Int. Ed.* **42**, 4807–4810 (2003).
11. J. A. Doudna, E. Charpentier, The new frontier of genome engineering with CRISPR-Cas9. *Science* **346**, 1258096 (2014).
12. N. J. Turner, Directed evolution drives the next generation of biocatalysts. *Nat. Chem. Biol.* **5**, 567–573 (2009).
13. U. Hafeez, S. Parakh, H. K. Gan, A. M. Scott, Antibody–Drug Conjugates for Cancer Therapy. *Molecules* **25**, 4764 (2020).

14. M. Soskine, D. S. Tawfik, Mutational effects and the evolution of new protein functions. *Nat. Rev. Genet.* **11**, 572–582 (2010).
15. M. L. Geddie, I. Matsumura, Rapid Evolution of β-Glucuronidase Specificity by Saturation Mutagenesis of an Active Site Loop. *J. Biol. Chem.* **279**, 26462–26468 (2004).
16. Y. Wang, et al., Directed Evolution: Methodologies and Applications. *Chem. Rev.* **121**, 12384–12444 (2021).
17. R. Otten, et al., How directed evolution reshapes energy landscapes in enzymes to boost catalysis. *Science* **370**, 1442–1446 (2020).
18. C. G. Acevedo-Rocha, et al., Pervasive cooperative mutational effects on multiple catalytic enzyme traits emerge via long-range conformational dynamics. *Nat. Commun.* **12**, 1621 (2021).
19. S. Bhattacharya, et al., NMR-Guided Directed Evolution. *Nature* **610**, 389–393 (2022).
20. L. I. Gutierrez-Rus, *et al.*, Enzyme Enhancement Through Computational Stability Design Targeting NMR-Determined Catalytic Hotspots. *J. Am. Chem. Soc.* **147**, 14978–14996 (2025).
21. L. R. Marshall, S. Bhattacharya, I. V. Korendovych, Fishing for Catalysis: Experimental Approaches to Narrowing Search Space in Directed Evolution of Enzymes. *JACS Au* **3**, 2402–2412 (2023).
22. H. Yu, S. Ma, Y. Li, P. A. Dalby, Hot spots-making directed evolution easier. *Biotechnol. Adv.* **56**, 107926 (2022).
23. O. Khersonsky, *et al.*, Automated Design of Efficient and Functionally Diverse Enzyme Repertoires. *Mol. Cell* **72**, 178-186.e5 (2018).
24. H. J. Wijma, *et al.*, Enantioselective Enzymes by Computational Design and In Silico Screening. *Angew. Chem. Int. Ed.* **54**, 3726–3730 (2015).
25. Y. Hu, *et al.*, NMR-Based Methods for Protein Analysis. *Anal. Chem.* **93**, 1866–1879 (2021).
26. A. Mainz, *et al.*, NMR Spectroscopy of Soluble Protein Complexes at One Mega-Dalton and Beyond. *Angew. Chem. Int. Ed.* **52**, 8746–8751 (2013).
27. V. Tugarinov, R. Sprangers, L. E. Kay, Probing Side-Chain Dynamics in the Proteasome by Relaxation Violated Coherence Transfer NMR Spectroscopy. *J. Am. Chem. Soc.* **129**, 1743–1750 (2007).
28. A. M. Ruschak, L. E. Kay, Methyl groups as probes of supra-molecular structure, dynamics and function. *J. Biomol. NMR* **46**, 75–87 (2010).
29. D. M. Korzhnev, *et al.*, Low-populated folding intermediates of Fyn SH3 characterized by relaxation dispersion NMR. *Nature* **430**, 586–590 (2004).
30. D. P. Frueh, Practical aspects of NMR signal assignment in larger and challenging proteins. *Prog. Nucl. Magn. Reson. Spectrosc.* **0**, 47–75 (2014).
31. K. S. Lundberg, et al., High-fidelity amplification using a thermostable DNA polymerase isolated from Pyrococcus furiosus. *Gene* **108**, 1–6 (1991).
32. P. Moore, Replicating success. *Nature* **435**, 235–235 (2005).
33. J. Shendure, *et al.*, DNA sequencing at 40: past, present and future. *Nature* **550**, 345–353 (2017).
34. D. G. Gibson, *et al.*, Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345 (2009).
35. E. R. Moody, R. Obexer, F. Nickl, R. Spiess, S. L. Lovelock, An enzyme cascade enables production of therapeutic oligonucleotides in a single operation. *Science* **380**, 1150–1154 (2023).
36. Z. Yu, P. Li, K. M. Merz, Using Ligand-Induced Protein Chemical Shift Perturbations To Determine Protein–Ligand Structures. *Biochemistry* **56**, 2349–2362 (2017).
37. C. Aguirre, T. Ten Brink, O. Cala, J.-F. Guichou, I. Krimm, Protein–ligand structure guided by backbone and side-chain proton chemical shift perturbations. *J. Biomol. NMR* **60**, 147–156 (2014).
38. Z. Tao, et al., A method for rapid nanobody screening with no bias of the library diversity. *iScience* **27**, 108966 (2024).
39. E. L. Ulrich, et al., BioMagResBank. *Nucleic Acids Res.* **36**, D402–D408 (2008).
40. J. Abramson, et al., Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 493–500 (2024).
41. R. V. Honorato, et al., The HADDOCK2.4 web server for integrative modeling of biomolecular complexes. *Nat. Protoc.* **19**, 3219–3241 (2024).

42. J. Stark, R. Powers, Rapid Protein−Ligand Costructures Using Chemical Shift Perturbations. *J. Am. Chem. Soc.* **130**, 535–545 (2008).
43. J. Liu, et al., NMR characterization of weak interactions between RhoGDI2 and fragment screening hits. *Biochim. Biophys. Acta Gen. Subj.* **1861**, 3061–3070 (2017).
44. V. Truniger, J. M. Lázaro, M. Salas, L. Blanco, A DNA binding motif coordinating synthesis and degradation in proofreading DNA polymerases. *EMBO J.* **15**, 3430–3441 (1996).
45. K. Böhlke, 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.* **28**, 3910–3917 (2000).
46. F. Mao, W.-Y. Leung, X. Xin, Characterization of EvaGreen and the implication of its physicochemical properties for qPCR applications. *BMC Biotechnol.* **7**, 76 (2007).
47. A. Mainz, et al., NMR Spectroscopy of Soluble Protein Complexes at One Mega-Dalton and Beyond. *Angew. Chem. Int. Ed.* **52**, 8746–8751 (2013).
48. D. M. Doddrell, D. T. Pegg, M. R. Bendall, Distortionless enhancement of NMR signals by polarization transfer. *J. Magn. Reson. (1969)* **48**, 323–327 (1982).
49. S. D. Gorman, D. Sahu, K. F. O'Rourke, D. D. Boehr, Assigning methyl resonances for protein solution-state NMR studies. *Methods* **148**, 88–99 (2018).
50. W. Henke, K. Herdel, K. Jung, D. Schnorr, S. A. Loening, Betaine improves the PCR amplification of GC-rich DNA sequences. *Nucleic Acids Res.* **25**, 3957–3958 (1997).
51. M. R. Green, J. Sambrook, Polymerase Chain Reaction (PCR) Amplification of GC-Rich Templates. *Cold Spring Harb. Protoc.* **2019**, pdb.prot095141 (2019).
52. F.-J. Wu, et al., Nanobody GPS by PCS: An Efficient New NMR Analysis Method for G Protein Coupled Receptors and Other Large Proteins. *J. Am. Chem. Soc.* **144**, 21728–21740 (2022).
53. Y. Xu, D. Zhang, R. Rogawski, C. M. Nimigean, A. E. McDermott, Identifying coupled clusters of allostery participants through chemical shift perturbations. *Proc. Natl. Acad. Sci. USA* **116**, 2078–2085 (2019).
54. G. Wagner, A. Pardi, K. Wuethrich, Hydrogen bond length and proton NMR chemical shifts in proteins. *J. Am. Chem. Soc.* **105**, 5948–5949 (1983).
55. L. He, T. Lührs, C. Ritter, Solid-state NMR resonance assignments of the filament-forming CARD domain of the innate immunity signaling protein MAVS. *Biomol. NMR Assign* **9**, 223–227 (2015).
56. W. F. Vranken, et al., The CCPN data model for NMR spectroscopy: Development of a software pipeline. *Proteins* **59**, 687–696 (2005).

Figures and Tables

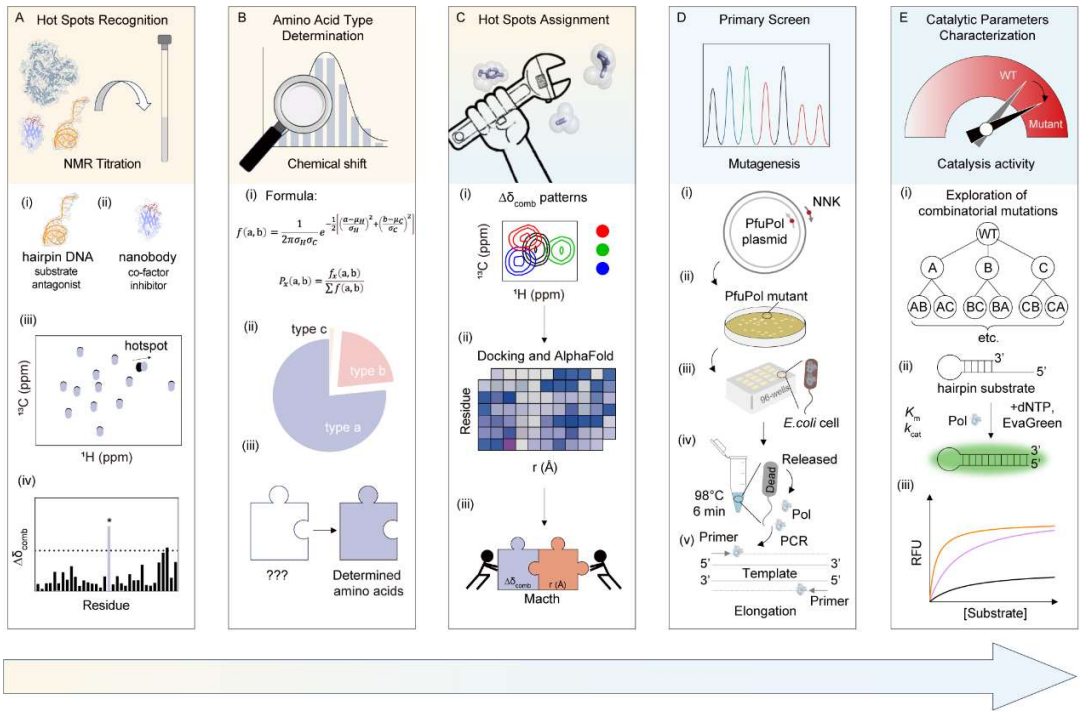

Fig. 1. Schematic series of the hybrid NMR-computational approach: from functional hotspot mapping to guided evolution of large enzymes. (*A*) Schematic representation of hotspots identification by NMR titration. (*B*) Schematic diagram of the determination of amino acid types of hotspot residues. (*C*) Assignments of hotspot residues. (*D*) Schematic pipeline for the primary screening of hotspot residue libraries. (*E*) Schematic workflow for characterization of catalytic parameters and exploration of combinatorial mutations.

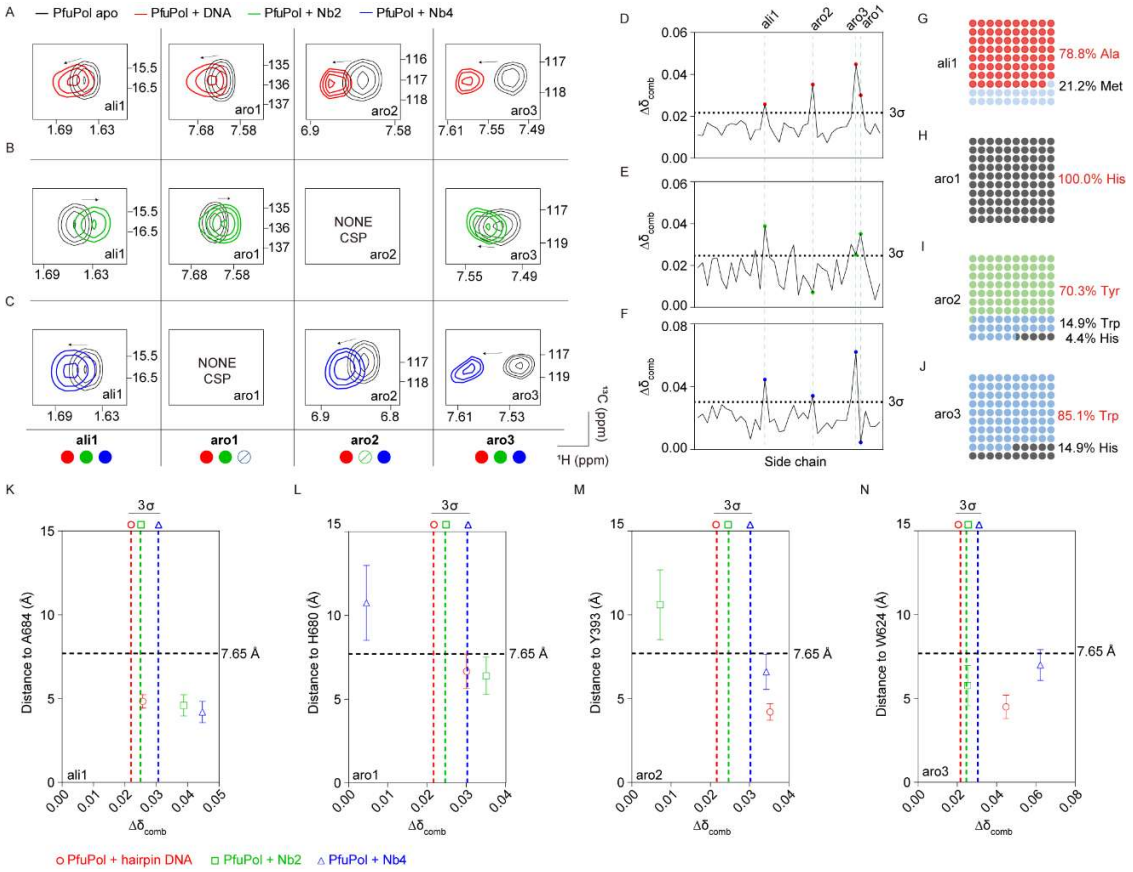

Fig. 2. Identification and assignment of hotspots of PfuPol. (*A–C*) Two-dimensional [^{13}C,^{1}H]-HMQC spectra of 50 µM [^{13}C]-labeled PfuPol were acquired on the 600 MHz spectrometer at 298 K. Enlarged view of NMR peaks of hotspot residues, which show CSPs upon adding 50 µM hairpin DNA (*A*, red), 75 µM Nb2 (*B*, green), 75 µM Nb4 (*C*, blue). Residues with CSPs were depicted by a colored circle at the bottom. Red circle indicated hairpin DNA induced CSPs. Green circle indicated Nb2 induced CSPs. Blue circle indicated Nb4 induced CSPs. Circle with a slash denoted no CSPs. (*D*) Side chain CSP of PfuPol upon addition of 50 µM hairpin DNA. (*E*) Side chain CSP of PfuPol upon addition of 75 µM Nb2. (*F*) Side chain CSP of PfuPol upon addition of 75 µM Nb4. (*G–J*) The probability of each amino acid type for residues ali1 (*G*), aro1 (*H*), aro2 (*I*), and aro3 (*J*) was determined based on calculations using probability density functions and Bayes' theorem. (*K–L*) The minimum distances (r) between residues in PfuPol and their interaction partners (DNA, Nb2, Nb4) were plotted alongside the chemical shift perturbations ($\Delta\delta_{comb}$) observed at the hotspots upon addition of the interaction partners. (*K*) The assignment of ali1 was made to residue A684, which exhibited a minimum distance of less than 7.65 Å from all three binders (hairpin DNA, Nb2, and Nb4) and showed significant perturbation upon interaction with each of them. (*L*) The assignment of aro1 was made to residue H680, which exhibited a minimum distance of less than 7.65 Å with each of hairpin DNA and Nb2 and showed significant perturbation upon interaction with each of them. (*M*) The assignment of aro2 was made to residue Y393, which exhibited a minimum distance of less than 7.65 Å with each of hairpin DNA and Nb4 and showed significant perturbation upon interaction with each of them. (*N*) The assignment of aro3 was made to residue W624, which exhibited a minimum distance of less than 7.65 Å from all three binders (hairpin DNA, Nb2, and Nb4) and showed significant perturbation upon interaction with each of them.

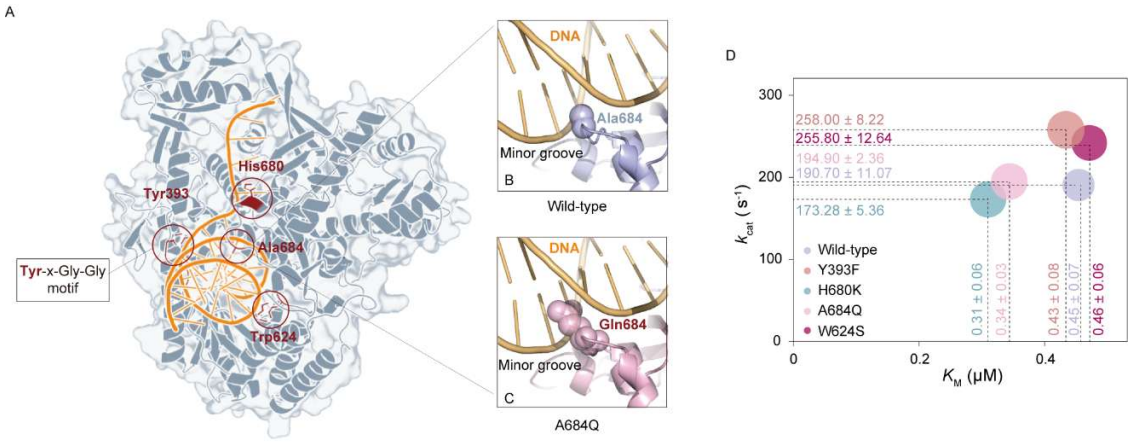

Fig. 3. Structural and biochemical analysis of variants of PfuPol. (*A*) Structural models taken from AlphaFold3 of complex of PfuPol with a hairpin DNA. Assigned hotspots were mapped on the structure of PfuPol showing as red sticks. (*B* and *C*) Structural illustration of the DNA substrate binding pocket of the wild-type PfuPol (*B*, lightblue) and the variant PfuPol-A684Q (*C*, lightpink), predicted by AlphaFold3. The DNA substrate is shown in ball and stick representation colored by orange. The mutated residue of PfuPol and PfuPol-A684Q was drawn as spheres. All structure figures were made using PyMOL. (*D*) Comparison of kinetic parameters for PfuPol and its variants. Bubble size represents fitted k_{cat}/K_M values. Wild-type (lightblue): PfuPol; Y393F (deepsalmon): PfuPol-Y393F; H680K (cyan): PfuPol-H680K; A684Q (pink): PfuPol-A684Q; W624S (purple): PfuPol-W624S.

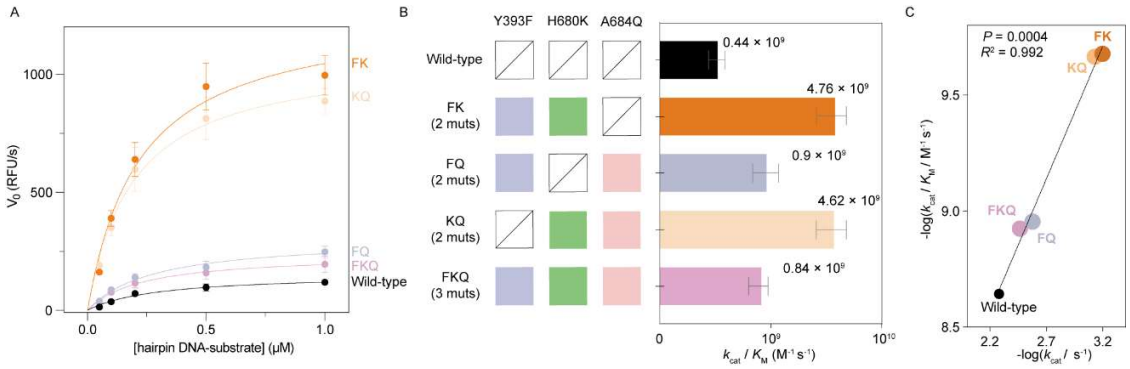

Fig. 4. Characterization of PfuPol with combinatorial mutations at three hotspot residues. (*A*) Michaelis-Menten plots for PfuPol and its variants. (*B*) The catalytic efficiency of PfuPol and its variants. Slashes denote unmutated hotspots. Data represent the mean ± SD of 3 biological replicates. Muts, number of mutations. (*C*) Plot of logarithm of catalytic efficiency vs logarithm of catalytic rate constant including PfuPol and its variants. Wild-type (black): PfuPol; FK (brightorange): PfuPol-Y393F/H680K; FQ (lightblue): PfuPol-Y393F/A684Q; KQ (lightorange): PfuPol-H680K/A684Q; FKQ (lightpink): PfuPol-Y393F/H680K/A684Q.