ChemBioChem

Research Article
doi.org/10.1002/cbic.202400495

Chemistry Europe
European Chemical Societies Publishing

www.chembiochem.org

Machine Learning Guided Rational Design of a Non-Heme Iron-Based Lysine Dioxygenase Improves its Total Turnover Number

R. Hunter Wilson,[a] Daniel J. Diaz,[b, c] Anoop R. Damodaran,*[a] and Ambika Bhagi-Damodaran*[a]

Highly selective C–H functionalization remains an ongoing challenge in organic synthetic methodologies. Biocatalysts are robust tools for achieving these difficult chemical transformations. Biocatalyst engineering has often required directed evolution or structure-based rational design campaigns to improve their activities. In recent years, machine learning has been integrated into these workflows to improve the discovery of beneficial enzyme variants. In this work, we combine a structure-based self-supervised machine learning framework, MutComputeX, with classical molecular dynamics simulations to down select mutations for rational design of a non-heme iron-dependent lysine dioxygenase, LDO. This approach consistently resulted in functional LDO mutants and circumvents the need for extensive study of mutational activity before-hand. Our rationally designed single mutants purified with up to 2-fold higher expression yields than WT and displayed higher total turnover numbers (TTN). Combining five such single mutations into a pentamutant variant, LPNYI LDO, leads to a 40% improvement in the TTN (218 ± 3) as compared to WT LDO ($TTN = 160 \pm 2$). Overall, this work offers a low-barrier approach for those seeking to synergize machine learning algorithms with pre-existing protein engineering strategies.

Introduction

Selective and catalytic activation of aliphatic C–H bonds remains a long-standing challenge in synthetic chemistry.[1–4] Biocatalysis has emerged as a potential solution as enzymes can perform C–H bond activation with high degrees of regio-, chemo-, and stereoselectivity.[5–9] Late-stage, biocatalytic incorporation of desired functional groups streamline synthetic pathways and offer more sustainable solutions for challenging synthetic transformations. In particular, installation of hydroxyl groups into inert C–H bonds have been achieved with both heme-containing P450s and 2-oxoglutarate(2OG)-dependent non-heme iron metalloenzymes.[10–15] Protein engineering strategies such as structure-based rational design or directed evolution have been implemented to improve yields and enzyme stability.[16–21] P450 enzymes have been engineered to catalyze a diverse scope of reactions; however, a significant drawback to these systems is the requirement of a reductase protein partner or the presence of a costly NADPH regeneration system.[22,23] By contrast, 2OG-dependent non-heme iron-dependent enzymes only require the relatively inexpensive 2OG co-substrate to catalyze their reactions. Despite this convenience, engineering campaigns involving 2OG-dependent enzymes involving non-native substrates often have limited turnovers.[24–26] However, recent work from Maloney and co-workers demonstrated the evolution of Pip4H to catalyze the hydroxylation of a non-native substrate for >15000 turnovers; though, these results were only achieved after screening thousands of mutants and sampling the sequence space of every amino acid position.[19]

Machine learning (ML) is being increasingly involved in protein engineering campaigns to decrease the overwhelming sampling of sequence space and produce improved biocatalysts.[27–32] Leveraging these new ML tools could enable more rapid discovery of hotspot residues primed for mutagenesis via rational design or library generation for directed evolution. While ML tools can be integrated into engineering workflows, they often require extensive training datasets either from enzyme-specific reaction assays or non-structural genetic data.[33–38] Tailoring of extensive datasets for each specific enzyme still requires burdensome front-work from researchers. Thus, synergizing self-supervised pretrained ML frameworks with existing rational design strategies offers a low-barrier solution for identifying potentially beneficial mutations in an engineering campaign.

Herein, we explore a ML-guided rational design strategy for the engineering of LDO, a recently characterized 2OG-depend-

[a] R. Hunter Wilson, A. R. Damodaran, A. Bhagi-Damodaran
Department of Chemistry, University of Minnesota, Twin Cities Minneapolis, MN-55455, United States
E-mail: rdanoop@umn.edu
ambikab@umn.edu

[b] D. J. Diaz
Department of Chemistry, Department of Computer Science, University of Texas at Austin, Austin, TX-78705, United States

[c] D. J. Diaz
Institute for Foundations of Machine Learning, University of Texas at Austin, Austin, TX-78705, United States

Supporting information for this article is available on the WWW under https://doi.org/10.1002/cbic.202400495

ent lysine dioxygenase (Figure 1A; referred to as 'Hydrox' in the previous work).[39] LDO catalyzes the installation of a hydroxyl group into a chemically inert C–H bond on the C_4 (γ) carbon of lysine. γ-hydroxy-lysine is a useful building block for high-value scaffolds in pharmaceutical and polymer industries; thus, engineering biocatalysts for its efficient production is desirable.[40–45] We synergized the structure-based self-supervised framework MutComputeX[46] with molecular dynamics (MD) simulations to identify mutations capable of increasing the total turnover number (TTN) of LDO. Overall, our ML-guided rational design of LDO demonstrates a readily adoptable strategy for improving the catalytic performance of metalloenzyme biocatalysts.

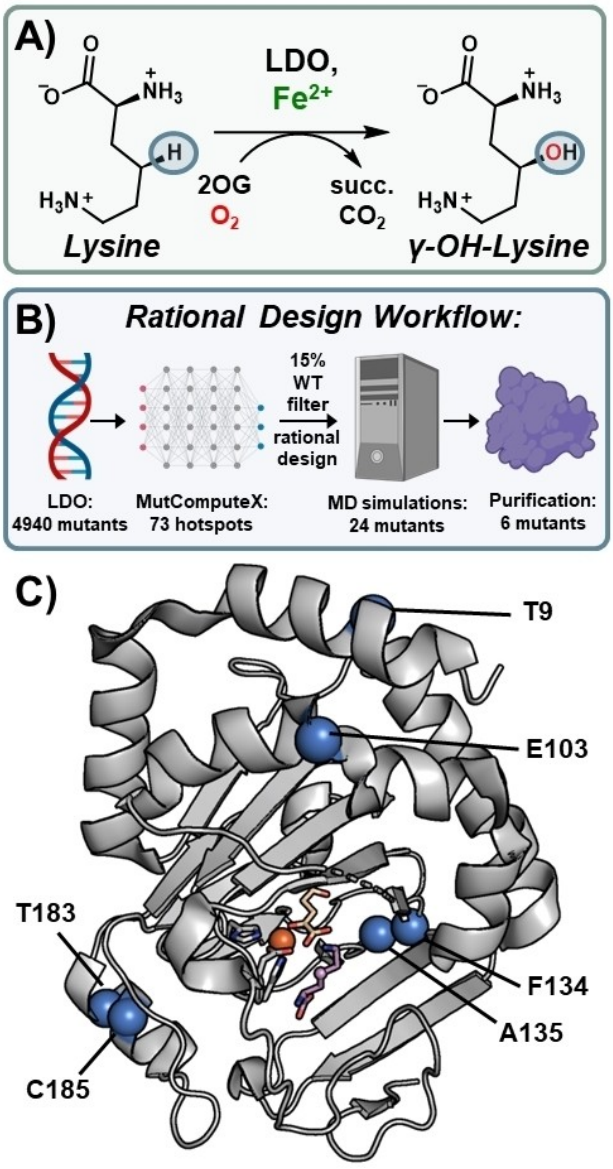

Figure 1. Machine learning-guided protein engineering of a non-heme iron 2OG-dependent dioxygenase. **A**) The non-heme iron enzyme, Lysine dioxygenase (LDO) catalyzes the hydroxylation of an aliphatic C–H bond at the C_4 position of lysine. 2OG and molecular oxygen are co-substrates; succinate and CO_2 are generated as byproducts. **B**) Rational design workflow overview starting from the single mutant sequence space of LDO and ending with our six designed single mutants. **C**) Crystal structure of LDO (PDB: 7JSD) showing ML-identified mutational hotspots as blue spheres. Iron displayed as an orange sphere, 2OG and Lysine are shown as beige and purple sticks, respectively.

Results and Discussion

Computational Design of LDO Variants

In the absence of extensive mutational data with LDO, application of data-driven ML algorithms for designing variants with enhanced catalytic activity was not possible. Nevertheless, with the crystal structure of lysine-bound LDO in hand,[39] we could use a structure-based self-supervised framework, MutComputeX, to enhance its developability and catalytic activity. MutComputeX is trained on >21000 sequence-balanced protein structures and can identify residues that are chemically incongruent with their surrounding chemical environment (microenvironment). Zero-shot predictions by the MutCompute platform have previously been shown to increase the turnovers and thermostability of several enzymes across a wide range of functions and we posited that it will generalize to LDO.[31,47–50] Using the LDO crystal structure (PDB: 7JSD) as input, MutComputeX identified 73 mutational hotspots (i.e. residues where the wildtype amino acid was not the top predicted amino acid) of 260 total residues in the protein (Figure 1B). Setting a cut-off filter of 15% WT probability, we down-selected 17 different mutational hotspots where the WT amino acid was disfavored (Table S1). We note that most of these amino acids are present on the surface of LDO. We inspected the crystal structure at the 17 down-selected mutational sites and developed 24 mutations by rational design based on the local protein environment (Table S2). Since MutComputeX suggests amino acid substitutions at each hotspot residue, those were considered in our rational design strategy. We intentionally avoided mutational hotspots proximal (<12.5 Å) to the iron center, as it's been demonstrated that enzyme activity in 2OG-dependent non-heme iron enzymes can be greatly diminished by mutations proximal to the iron center.[51–54] To verify potentially stabilizing interactions (due to increased H-bonding or hydrophobic packing interactions, etc.), we conducted MD simulations of our 24 rationally designed variants (Table S2). All mutational results were compared to simulations of WT LDO to discern whether new interactions were beneficial and did not diminish those present in the WT protein. This allowed us to narrow our experimental studies to six mutations (Figure 1C).

We began our MD analysis with the T9 L MutComputeX prediction. We rationalized from a structural analysis that this mutation will improve the hydrophobic packing with the surrounding hydrophobic pocket comprising of Leu10, Ileu6, Val99, and Leu49 (Figure 2A). This design was supported by a lower temperature factor (B-factor) observed across MD simulations compared to the WT (Figure 2A inset and Figure S1). Glu103 was the next target of our study and is located on a heavily strained turn connecting two alpha helices (Figure S2). MutComputeX predicted a proline at E103, and we anticipate that proline would help relax this strain and stabilize the helix-

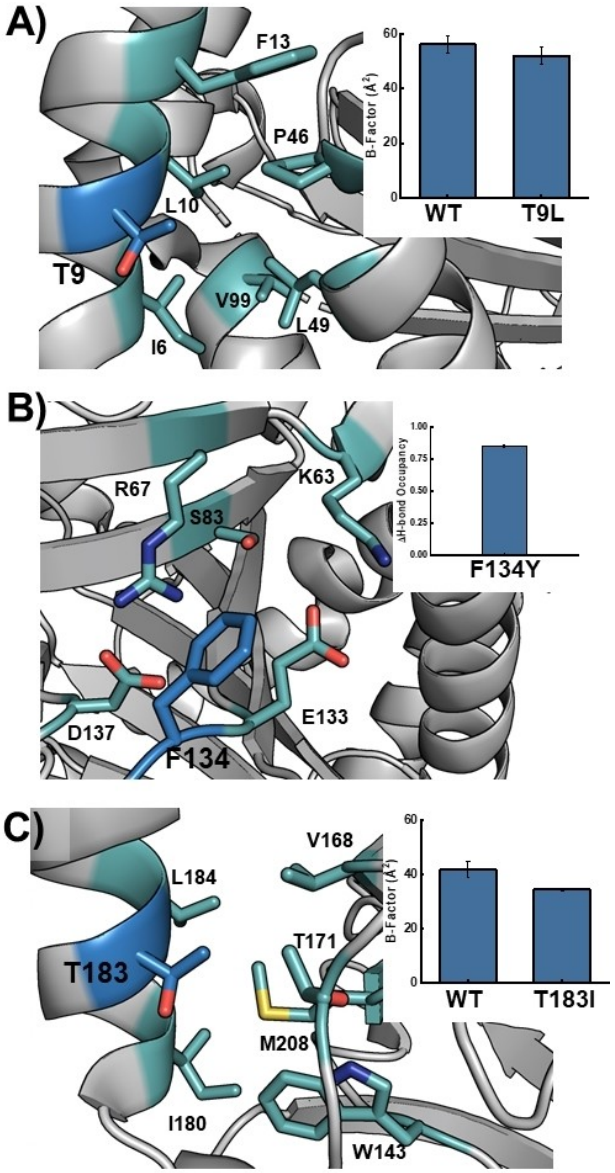

Figure 2. LDO variant design by Molecular Dynamics simulations. **A**) T9 L variant (blue) can increase hydrophobic interactions by burying into an adjacent pocket (teal); inset: B-factor of pocket from MD simulations. **B**) Mutation of Phe134 (blue) to Tyr increases the stability within a highly polar pocket (cyan); inset: change in solvent H-bonding upon mutation. **C**) T183I mutant (blue) can bury into an adjacent hydrophobic pocket (cyan) to increase hydrophobic interactions; inset: B-factor of pocket from MD simulations.

turn-helix motif. Indeed, MD simulations with this variant displayed an average 0.5 Å lower root mean square deviation (RMSD) of the helix-turn-helix backbone atoms with E103P mutant as compared to WT LDO, suggesting this variant could relax the apparent strain. The F134Y predictions is located on the protein surface embedded in a hydrogen bond network (Lys65, Arg67, Asp137, Glu133, and Ser83 residues) and is participating in a cation-π interactions with Arg67. Thus, we rationalized that mutating this residue to a polar tyrosine would not only increase solubility but would strengthen the solvent-exposed hydrogen bond network while maintaining the cation-π interaction with Arg67 (Figure 2B). Indeed, we observe that the F134Y variant can form new H-bonds with water molecules present in the pocket (Figure 2B inset and Figure S3). Residue Thr183 is semi-solvent exposed and adjacent to the hydrophobic residues Ile180, Trp143, and Val168 (Figure 2C). Here, MutComputeX predicts Ile, which we rationalize will improve hydrophobic packing. MD simulations of T183I demonstrate a 20% lower B-factor for the hydrophophic pocket, supporting that the T183I mutant would experience tighter hydrophobic packing (Figure 2C and Figure S4). In addition to predictions with the LDO crystal structure, we performed MutComputeX predictions on frames from MD simulations of WT LDO to account for the effects of protein dynamics into our rational design strategy. As a result, two more mutational hotspots were identified. First, substitution of Ala135 with a polar Asn was anticipated to form new H-bonds with residues on an adjacent loop (Figure S5). Indeed, the MD simulations predicted new H-bonds with the side chains of Ser136 and Asp217. Additionally, mutating surface exposed Ala to Asn increases H-bonds with water, likely improving solubility. Finally, MutComputeX flagged the solvent-exposed Cys185 for mutagenesis. Removing solvent exposed cysteine residues is critical for developing enzymes with high turnovers because it preempts promiscuous oxidation that can deactivate the enzyme. We mutated C185 to its isosteric, redox-inactive counterpart: serine. MD simulations demonstrate the major impact of this mutation is increased H-bonding interactions with Asn181 and Glu182, located within the residing alpha helix (Figure S6).

As all these residues are distal (>12.5 Å) from the LDO active site, they have the potential to form stabilizing interactions without sacrificing enzyme activity. Parsing through all residues distal from the active site would require time-consuming front-end labor. By leveraging MutComputeX, non-intuitive mutational hotspots were readily identified, enabling rapid structural rationalization and curation of mutational designs. Since MutComputeX relies solely on a static structure and does not account for steric and electronic changes induced by a mutation, it is better at identifying residues primed for mutagenesis rather than predicting specific mutations. Thus, our MD simulations were instrumental in screening viable mutations created by ML-guided rational design (Table S2). For example, despite position Tyr18 having a low WT probability (1.9%), mutation of this residue to Phe (as suggested by MutComputeX) resulted in the elimination of H-bonds with the backbone carbonyls of Met42/Arg43 as well as an increased B-factor of the local hydrophobic pocket (Figure S7). Another MutComputeX prediction, N181 W, resulted in a sharp increase in the B-factor of its adjacent hydrophobic pocket, weakening local hydrophobic interactions (Figure S8). Given these observed deleterious interactions, neither of these designs were expressed or purified. By synergizing MutComputeX with MD simulations, we were able to propose six mutants for experimental evaluation of hydroxylation activity (Figure 1B and C).

Purification of LDO Variants and Thermostability Analysis

WT LDO and all designed variants were expressed as His-Tagged constructs and purified by affinity chromatography. Interestingly, five of six variants were expressed in greater yields than WT, implying they exhibited increased solubility and/or folding dynamics[47] (Figure S9). The C185S mutant was isolated with approximately half the yield compared to WT, indicating that Cys185 may serve a structural role. Notably, Cys185 is the only cysteine residue in LDO, and it cannot form any intramolecular disulfide bonds. Disulfide bonds are also absent between monomeric chains in the crystal structure. Given that the other five mutations appeared to increase the protein yields, we also designed and purified a combinatorial mutant containing the other five mutations (T9 L/E103P/F134Y/A135N/T183I), which resulted in a variant with ~2.5-fold improvement in protein expression. We refer to this quintuple variant as LPYNI LDO.

MutComputeX-guided protein engineering has been shown to enhance the thermostability of designed variants, resulting in increased melting points (T_m) and turnovers.[31,48,49] To assess the melting points of WT and designed LDO variants, we employed a thermal shift assay (TSA) to investigate how individual mutations impacted overall protein stability. In the apo forms of the protein, WT LDO had a T_m of 31.0 °C with most of the other variants containing slightly lower T_m (within 2 °C) (Figure S10). Only A135 N and C185S displayed melting points drastically lower than WT (−4.0 °C), while T183I and the combinatorial LPYNI variant displayed elevated T_m values (+2.2 and +2.7 °C, respectively). We note that since the variants were designed and simulated in a Fe/2OG-bound state, we would not know how our designs would behave in an apo state.

To validate our designs, we performed the TSA in the presence of metal and 2OG substrates to capture a state more similar to our simulations (Figure S11). Interestingly, the T_m of the WT increases to 50.7 °C, a +19.7 °C shift relative to the apo form of the protein. This implies that metal and 2OG binding result in new protein-stabilizing interactions to form a more stable and catalytically active state. Lending credence to this, no crystal structure of a 2OG-dependent dioxygenase has been solved without the active site metal, suggesting an important structural role.[55–61] Additionally, this iron-dependent enzyme family forms extensive H-bonding interactions with the 2OG substrate, securing its position within the active site.[39,55,56,58,59] A similar increase in T_m is observed across all variants, indicating that they can still bind cofactors necessary for catalysis. Relative to WT, most variants exhibit modest increases in T_m (+0.2–0.5 °C for T9 L, E103P, and F134Y) while T183I exhibits a T_m 3.5 °C greater than WT. Analogous to the apo protein, A135N and C185S have lowered melting temperatures (−2.4 °C and −4.2 °C, respectively) relative to the ligand-bound WT. The combinatorial mutant, LPYNI displays a modest decrease in T_m of 0.9 °C relative to WT, indicating that the stabilizing/destabilizing effects of individual mutations are not additive when stacked.

We also investigated the effect of adding the substrate lysine to the protein complex (Figure 3A–C). Upon binding the substrate, the melting point of the WT LDO protein complex increases to 54.5 °C (+3.8 °C relative to LDO without lysine). Once again, it appears that structural changes are occurring upon substrate binding that increase the stability of the overall complex. Recent structural studies in an analogous 2OG-dependent dioxygenase, HalD, demonstrated that upon substrate binding, a lid closes over the active site, securing the substrate's orientation near the reactive iron center.[62] Given our observed increase in melting temperatures upon lysine binding, a similar conformational change might also occur in LDO. Similar to WT, all mutants display an increase in T_m upon binding substrate lysine (Figures 3A–C and S11). Most lysine bound mutants (T9L, E103P, and F134Y) display similar melting temperatures to the WT protein (within 0.5 °C), while T183I displays an increased T_m (+3.0 °C). While the quintuple LPYNI

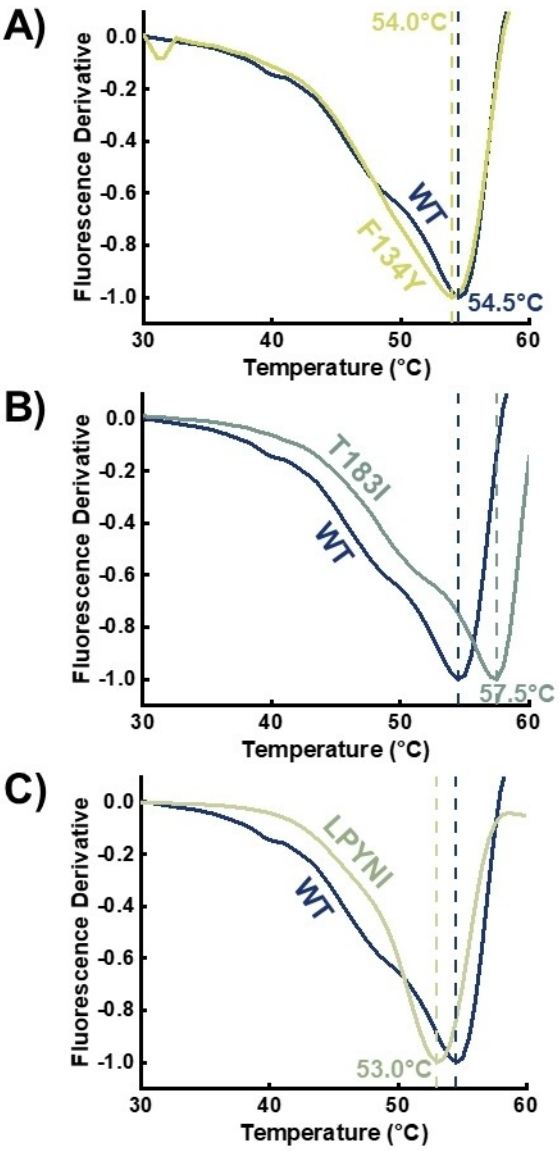

Figure 3. Melting point analysis for designed variants (**A** – F134Y, **B** – T183I, **C** – LPYNI) as compared to WT. All proteins were constituted with Mn^{2+}, 2OG, and Lysine. Vertical lines are references for the WT and variant melting points.

variant displayed a slightly lower overall T_m than WT, the onset of melting was observed at temperatures higher than WT indicating higher stability under ambient conditions (Figure 3C). Overall, these thermostability measurements and expression yields reveal that these mutations do not necessarily enhance the T_m of the enzyme-substrate-cofactor ternary complex but rather improve solubility and/or folding dynamics.

Catalytic Assessment of Designed LDO Variants

While our TSA measurements indicate that the designed enzymes can bind their co-substrates, we wanted to assess their catalytic viability. Specifically, we wanted to assess their ability to hydroxylate lysine relative to the WT scaffold. To that end, we designed hydroxylation assays in which LDO variants were incubated with all necessary substrates (iron, 2OG, L-lysine) under aerobic conditions (Figure 1A). After separation of protein from the reaction mixture, the lysine reactants and products were derivatized with a hydrophobic 6-aminoquinolyl-N-hydroxysuccinimidyl carbamate (AQC) tag. Derivatization with the AQC tag provides a chromophore for UV detection and enables retention of highly polar amino acids for reverse-phase HPLC analysis (Figure 4A). In control reactions lacking LDO, a single peak for AQC-tagged lysine is present in the chromatogram ($R_t = 4.8$ min). The predominant peak at 2.2 min (Bis-AMQ) is a common byproduct of derivatization with AQC.[63] In reactions incubated with WT LDO, a new peak appears at an earlier retention time ($R_t = 4.3$ min), indicative of the installation of a polar functional group into lysine. This implies the successful formation of γ-hydroxy-lysine, which we validated via accurate mass measurements: the reaction product exhibited an m/z of 252.1055, with a mass error of 2.8 ppm, validating its identity (Figure S12). From the reaction, we determined a total turnover number (TTN) of 160 ± 2 for WT LDO, which is in good agreement with the previously reported 136 as determined by mass spectrometric analysis.[39]

After establishing our assay with WT LDO, we evaluated the catalytic efficiency of our designed LDO variants. HPLC chromatograms displayed lower amounts of reactant lysine and increased amounts of γ-hydroxy-lysine present in the reaction mixture relative to the WT reaction (Figure 4A). Interestingly, all variants displayed a greater TTN than the WT enzyme (Figure 4B). Even mutations that decreased the T_m of LDO (A135N and C185S), resulted in increased TTN over WT LDO. While the individual mutants modestly increased TTN by 16–40 turnovers, we were intrigued as to whether the combinatorial LPYNI variant would perform better than the individual mutants. Indeed, LPYNI LDO displayed a TTN of 218 ± 3, performing 1.4-fold better than WT. Despite this improvement in activity, we observe diminishing returns in TTN from the incorporation of individual mutants, analogous to results obtained from directed evolution campaigns. In a recent study, a 2OG-dependent hydroxylase was evolved to function on a non-native substrate after sampling thousands of variants for activity.[19] While the increases in activity here can be perceived as marginal compared to gains from directed evolution, our engineering strategy still produced functional, improved variants with only 7 purified enzymes. Additionally, our rational design strategy circumvents the need for extensive front-end mutagenesis campaigns for generating training data for activity-specific ML algorithms. Overall, our rational design strategy results in biocatalysts with significantly improved expression yields and more favorable TTN while avoiding mutations in the primary and secondary coordination spheres of the iron center.

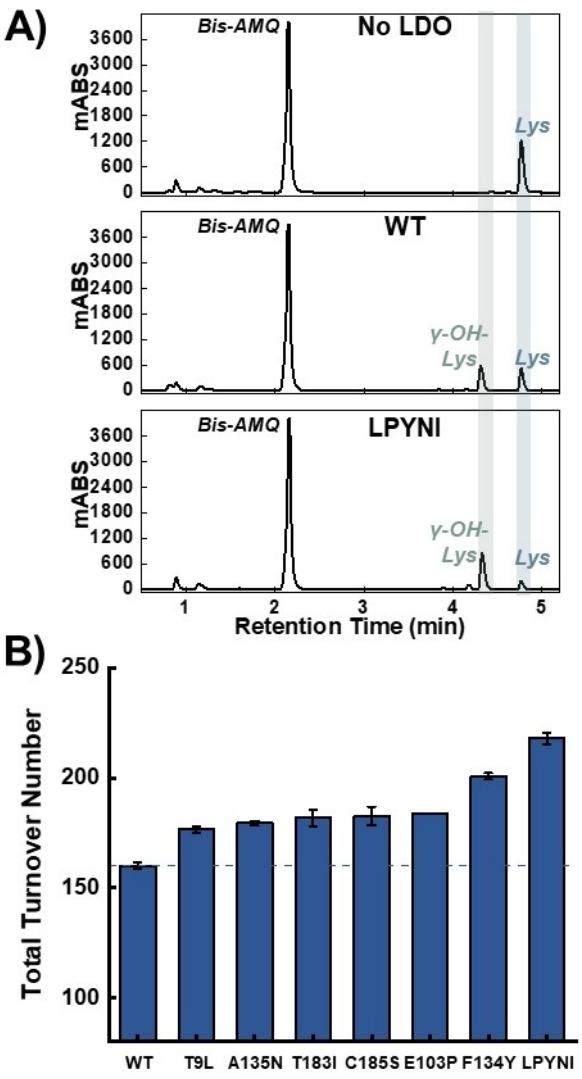

Figure 4. Hydroxylation assay analysis for LDO and designed variants. **A**) HPLC-UV detection of AQC-tagged lysine and γ-OH-lysine from a control reaction lacking LDO (top), a reaction with WT LDO (middle), and a reaction with LPYNI LDO (bottom). **B**) Total Turnover Number (TTN) for WT and LDO variants in the hydroxylation assay ($n = 3$). Horizontal line is a reference for the WT TTN. Error bars are the standard deviation from three independent reactions.

Despite these variants not displaying enhanced melting points, all were still able to improve the total turnover number (Figure 4B). Notably, mutations distal from the active site have been shown to influence protein conformational dynamics to enhance enzymatic activity.[64,65] To probe if an altered dynamics was the molecular basis for this improved activity, we again employed MD simulations to characterize WT and LPYNI LDO

over longer timescales (triplicate 500 ns productions). Analysis of the RMSD of the protein backbone atoms between WT and LPYNI shows that the pentamutant displayed ~0.5 Å smaller deviation than the WT scaffold (Figure S13). This suggests that the mutant complex displays more ordered dynamics than WT which could enable greater catalyst lifetimes. Furthermore, the time to reach an equilibrated state is notably shorter in LPYNI LDO, indicating the pentamutant readily equilibrates to a stable conformation. While this provides rudimentary evidence for the enhanced turnovers of the LPYNI variant, we were curious if other dynamic effects were at play. More specifically, we wanted to know if the residues targeted by our rational design strategy were involved in correlated protein motions, as it's been demonstrated that distal residues can enhance catalytic efficiency.[66,67] To this end, we used our MD simulations to create Shortest Path Maps[68] to identify correlated residue motions for both WT and LPYNI LDO (Figure S14 and 15). Intriguingly, none of the WT residues constituting the pentamutant were shown to be involved in significant correlated protein motions. However independent correlation networks were formed between the two scaffolds indicating that LPYNI LDO produces significant variations in the overall protein dynamics. Previous work incorporating MutCompute has shown that the platform can target (and propose stabilizing mutations for) residues that aid in protein folding dynamics.[47] Notably, significant residues identified by the Shortest Paths Maps did not overlap with those identified by MutComputeX, suggesting an alternative rational design approach could be implemented to target these residues for increased catalytic activity.

Next, we assessed the Root-Mean-Square Fluctuation (RMSF) of WT and LPYNI LDO residues to ascertain whether certain residues may contribute to protein instability (Figure S16). Residues with larger fluctuations in their positioning suggest either poor packing within the protein or a lack of stabilizing interactions to prevent aberrant protein motions. Notably, the majority of residues in LPYNI LDO display lower RMSF than the WT, suggesting that the scaffold exhibits more ordered protein dynamics. Upon inspection of the residues of greatest RMSF stabilization, one region corresponds to a highly flexible loop on the protein exterior, while the other region is associated with substrate binding and recognition (Figure S16).[62] Due to the greater stability of the C-terminal lysine-binding loop, it appears that LPYNI LDO may have allowed the protein to complete more substrate binding/product release steps before unfolding or other non-catalytic kinetic events could occur.[69] From our TSA analysis, we note that the onset of melting of LPYNI LDO was 10 °C higher than the WT scaffold, suggesting greater overall stability at ambient temperatures (Figure 3C). In summary, incorporation of our rationally designed mutants resulted in more ordered protein conformations which in turn provided greater turnovers.

Conclusions

We have developed a ML-guided rational design strategy for enhancing the catalytic TTN of metalloenzymes. As a proof of concept, we targeted an iron/2OG-dependent hydroxylase which activates an inert C–H bond on the C_4 carbon of lysine to produce γ-hydroxy-lysine. By identifying mutational hotspots with MutComputeX and screening rational designs with MD simulations, we were able to reduce the overwhelming sampling of sequence space and identify six potential mutations for purification (Figure 1B). The MutComputeX identified hotspots throughout the structure enabling facile generation of rational designs that hedge against deleterious mutations near the active site. Most variants exhibited similar thermostability as the WT protein and all variants were able to bind the necessary substrates for catalysis. All designed variants displayed increased TTN, further validating our design methodology. Further computational analysis suggests that the LPYNI variant afforded a more ordered scaffold enabling greater turnovers before catalyst deactivation. While our present work focuses on engineering a 2OG-dependent hydroxylase, we anticipate that our rational design method could be applied for improving other members of this superfamily such as halogenases, cyclases, and desaturases.[70,71] Beyond 2OG-dependent enzymes, this design method can be used towards engineering metalloenzymes containing more complicated cofactors such as porphyrins and iron-sulfur clusters.[72,73] While this strategy requires access to an initial crystal structure for ML predictions and computing resources for MD simulations, these computational investments are significantly less expensive than high-throughput strategies for enzyme evolution. Overall, our ML-guided rational design method enhances the expression and catalytic activity of industrially relevant biocatalysts.

Acknowledgements

RHW acknowledges the support of the National Institute of Health Chemical Biology Training Grant (T32GM132029). The characterization of LDO and its biochemical assays were supported by NSF CBET and CLP (Grant # 2046527). The rational design work was supported by UMN Start up funds. Mass spectrometry analysis was performed at The University of Minnesota Department of Chemistry Mass Spectrometry Laboratory (MSL), supported by the Office of the Vice President of Research, College of Science and Engineering, and the Department of Chemistry at the University of Minnesota. The content of this work is the sole responsibility of the authors and does not represent endorsement by MSL personnel. We would like to thank NSF AI Institute for Foundations of Machine Learning (IFML) for their support and AMD for the donation of hardware and support resources from its HPC fund.

Conflict of Interests

D.J.D. has a financial relationship with Intelligent Proteins LLC, which uses AI models for protein engineering.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Keywords: 2OG-dependent · Non-heme iron · Hydroxylase · Machine-learning · Rational design

[1] D. A. Petrone, J. Ye, M. Lautens, *Chem. Rev.* **2016**, *116*, 8003–8104.
[2] N. Holmberg-Douglas, D. A. Nicewicz, *Chem. Rev.* **2022**, *122*, 1925–2016.
[3] L. Guillemard, N. Kaplaneris, L. Ackermann, M. J. Johansson, *Nat. Rev. Chem.* **2021**, *5*, 522–545.
[4] H. M. L. Davies, J. D. Bois, J.-Q. Yu, *Chem. Soc. Rev.* **2011**, *40*, 1855–1856.
[5] E. L. Bell, W. Finnigan, S. P. France, A. P. Green, M. A. Hayes, L. J. Hepworth, S. L. Lovelock, H. Niikura, S. Osuna, E. Romero, K. S. Ryan, N. J. Turner, S. L. Flitsch, *Nat. Rev. Methods Primers* **2021**, *1*, 1–21.
[6] J. Dong, E. Fernández-Fueyo, F. Hollmann, C. E. Paul, M. Pesic, S. Schmidt, Y. Wang, S. Younes, W. Zhang, *Angew. Chem. Int. Ed.* **2018**, *57*, 9238–9261.
[7] S. P. France, R. D. Lewis, C. A. Martinez, *JACS Au* **2023**, *3*, 715–735.
[8] J. Tian, A. A. Garcia, P. H. Donnan, J. Bridwell-Rabb, *Biochemistry* **2023**, *62*, 1807–1822.
[9] M. Brimberry, A. A. Garcia, J. Liu, J. Tian, J. Bridwell-Rabb, *Curr. Opin. Chem. Biol.* **2023**, *72*, 102227.
[10] K. Zhang, A. Yu, X. Chu, F. Li, J. Liu, L. Liu, W.-J. Bai, C. He, X. Wang, *Angew. Chem. Int. Ed.* **2022**, *61*, e202204290.
[11] P. R. Ortiz de Montellano, *Chem. Rev.* **2010**, *110*, 932–948.
[12] S. N. Charlton, M. A. Hayes, *ChemMedChem* **2022**, *17*, e202200115.
[13] R.-Y. Zhang, T. Ma, D. Liu, Y.-L. Yang, L. Gao, H.-B. Cui, Z.-Q. Wang, Y.-Z. Chen, *Mol. Catal.* **2024**, *553*, 113791.
[14] Y. Zhao, B. Zhang, Z. Q. Sun, H. Zhang, W. Wang, Z. R. Wang, Z. K. Guo, S. Yu, R. X. Tan, H. M. Ge, *ACS Catal.* **2022**, *12*, 9839–9845.
[15] M. Hirsilä, P. Koivunen, V. Günzler, K. I. Kivirikko, J. Myllyharju, *J. Biol. Chem.* **2003**, *278*, 30772–30780.
[16] R. Agudo, G.-D. Roiban, M. T. Reetz, *ChemBioChem* **2012**, *13*, 1465–1473.
[17] K. Neufeld, B. Henßen, J. Pietruszka, *Angew. Chem. Int. Ed.* **2014**, *53*, 13253–13257.
[18] H. Chen, M. Huang, W. Yan, W.-J. Bai, X. Wang, *ACS Catal.* **2021**, *11*, 10625–10630.
[19] W. L. Cheung-Lee, J. N. Kolev, J. A. McIntosh, A. A. Gil, W. Pan, L. Xiao, J. E. Velásquez, R. Gangam, M. S. Winston, S. Li, K. Abe, E. Alwedi, Z. E. X. Dance, H. Fan, K. Hiraga, J. Kim, B. Kosjek, D. N. Le, N. S. Marzijarani, K. Mattern, J. P. McMullen, K. Narsimhan, A. Vikram, W. Wang, J.-X. Yan, R.-S. Yang, V. Zhang, W. Zhong, D. A. DiRocco, W. J. Morris, G. S. Murphy, K. M. Maloney, *Angew. Chem. Int. Ed.* **2024**, *63*, e202316133.
[20] J. H. Mills, S. D. Khare, J. M. Bolduc, F. Forouhar, V. K. Mulligan, S. Lew, J. Seetharaman, L. Tong, B. L. Stoddard, D. Baker, *J. Am. Chem. Soc.* **2013**, *135*, 13393–13399.
[21] P. Hosseinzadeh, G. Bhardwaj, V. K. Mulligan, M. D. Shortridge, T. W. Craven, F. Pardo-Avila, S. A. Rettie, D. E. Kim, D.-A. Silva, Y. M. Ibrahim, I. K. Webb, J. R. Cort, J. N. Adkins, G. Varani, D. Baker, *Science* **2017**, *358*, 1461–1466.
[22] Y. Wei, E. L. Ang, H. Zhao, *Curr. Opin. Chem. Biol.* **2018**, *43*, 1–7.
[23] H. Renault, J.-E. Bassard, B. Hamberger, D. Werck-Reichhart, *Curr. Opin. Plant Biol.* **2014**, *19*, 27–34.
[24] C. R. Zwick III, M. B. Sosa, H. Renata, *Angew. Chem. Int. Ed.* **2019**, *58*, 18854–18858.
[25] F. Meyer, R. Frey, M. Ligibel, E. Sager, K. Schroer, R. Snajdrova, R. Buller, *ACS Catal.* **2021**, *11*, 6261–6269.
[26] J. Büchler, S. H. Malca, D. Patsch, M. Voss, N. J. Turner, U. T. Bornscheuer, O. Allemann, C. Le Chapelain, A. Lumbroso, O. Loiseleur, R. Buller, *Nat. Commun.* **2022**, *13*, 371.
[27] D. Patsch, R. Buller, *CHIMIA* **2023**, *77*, 116–121.
[28] R. Buller, S. Lutz, R. J. Kazlauskas, R. Snajdrova, J. C. Moore, U. T. Bornscheuer, *Science* **2023**, *382*, eadh8615.
[29] Z. Wu, S. B. J. Kan, R. D. Lewis, B. J. Wittmann, F. H. Arnold, *Proc. Natl. Acad. Sci.* **2019**, *116*, 8852–8858.
[30] B. L. Hie, K. K. Yang, *Curr. Opin. Struct. Biol.* **2022**, *72*, 145–152.
[31] H. Lu, D. J. Diaz, N. J. Czarnecki, C. Zhu, W. Kim, R. Shroff, D. J. Acosta, B. R. Alexander, H. O. Cole, Y. Zhang, N. A. Lynd, A. D. Ellington, H. S. Alper, *Nature* **2022**, *604*, 662–667.
[32] D. J. Diaz, A. V. Kulikova, A. D. Ellington, C. O. Wilke, *Curr. Opin. Struct. Biol.* **2023**, *78*, 102518.
[33] S. d'Oelsnitz, D. J. Diaz, W. Kim, D. J. Acosta, T. L. Dangerfield, M. W. Schechter, M. B. Minus, J. R. Howard, H. Do, J. M. Loy, H. S. Alper, Y. J. Zhang, A. D. Ellington, *Nat. Commun.* **2024**, *15*, 2084.
[34] J. C. Greenhalgh, S. A. Fahlberg, B. F. Pfleger, P. A. Romero, *Nat. Commun.* **2021**, *12*, 5825.
[35] M. J. Menke, Y.-F. Ao, U. T. Bornscheuer, *ACS Catal.* **2024**, *14*, 6462–6469.
[36] D. Probst, M. Manica, Y. G. Nana Teukam, A. Castrogiovanni, F. Paratore, T. Laino, *Nat. Commun.* **2022**, *13*, 964.
[37] E. Orsi, L. Schada von Borzyskowski, S. Noack, P. I. Nikel, S. N. Lindner, *Nat. Commun.* **2024**, *15*, 3447.
[38] T. Matsushita, S. Kishimoto, K. Hara, H. Hashimoto, H. Yamaguchi, Y. Saito, K. Watanabe, *ACS Catal.* **2024**, *14*, 6945–6951.
[39] M. E. Neugebauer, E. N. Kissman, J. A. Marchand, J. G. Pelton, N. A. Sambold, D. C. Millar, M. C. Y. Chang, *Nat. Chem. Biol.* **2021**, *18*, 1–9.
[40] C. Prell, S.-A. Vonderbank, F. Meyer, F. Pérez-García, V. F. Wendisch, *Curr. Res. Biotechnol.* **2022**, *4*, 32–46.
[41] X. Jing, H. Liu, Y. Nie, Y. Xu, *Syst. Microbiol. Biomanuf.* **2021**, *1*, 275–290.
[42] K. R. Herbert, G. M. Williams, G. J. S. Cooper, M. A. Brimble, *Org. Biomol. Chem.* **2012**, *10*, 1137–1144.
[43] B. Ho, T. M. Zabriskie, *Bioorg. Med. Chem. Lett.* **1998**, *8*, 739–744.
[44] D. Lamarre, G. Croteau, E. Wardrop, L. Bourgon, D. Thibeault, C. Clouette, M. Vaillancourt, E. Cohen, C. Pargellis, C. Yoakim, P. C. Anderson, *Antimicrob. Agents Chemother.* **1997**, *41*, 965–971.
[45] A. Amatuni, H. Renata, *Org. Biomol. Chem.* **2019**, *17*, 1736–1739.
[46] A. V. Kulikova, D. J. Diaz, J. M. Loy, A. D. Ellington, C. O. Wilke, *J. Biol. Phys.* **2021**, *47*, 435–454.
[47] R. Shroff, A. W. Cole, D. J. Diaz, B. R. Morrow, I. Donnell, A. Annapareddy, J. Gollihar, A. D. Ellington, R. Thyer, *ACS Synth. Biol.* **2020**, *9*, 2927–2935.
[48] I. Paik, P. H. T. Ngo, R. Shroff, D. J. Diaz, A. C. Maranhao, D. J. F. Walker, S. Bhadra, A. D. Ellington, *Biochemistry* **2023**, *62*, 410–418.
[49] A. Kunka, S. M. Marques, M. Havlasek, M. Vasina, N. Velatova, L. Cengelova, D. Kovar, J. Damborsky, M. Marek, D. Bednar, Z. Prokop, *ACS Catal.* **2023**, *13*, 12506–12518.
[50] Y. Liu, S. G. Bender, D. Sorigue, D. J. Diaz, A. D. Ellington, G. Mann, S. Allmendinger, T. K. Hyster, *J. Am. Chem. Soc.* **2024**, *146*, 7191–7197.
[51] R. H. Wilson, S. Chatterjee, E. R. Smithwick, J. J. Dalluge, A. Bhagi-Damodaran, *ACS Catal.* **2022**, *12*, 10913–10924.
[52] M. J. Ryle, K. D. Koehntop, A. Liu, L. Que, R. P. Hausinger, *PNAS* **2003**, *100*, 3790–3795.
[53] Y.-H. Chen, L. M. Comeaux, S. J. Eyles, M. J. Knapp, *Chem. Commun.* **2008**, *39*, 4768–4770.
[54] K. D. Koehntop, S. Marimanikkuppam, M. J. Ryle, R. P. Hausinger, L. Que, *J. Biol. Inorg. Chem.* **2006**, *11*, 63–72.
[55] M. A. McDonough, V. Li, E. Flashman, R. Chowdhury, C. Mohr, B. M. R. Liénard, J. Zondlo, N. J. Oldham, I. J. Clifton, J. Lewis, L. A. McNeill, R. J. M. Kurzeja, K. S. Hewitson, E. Yang, S. Jordan, R. S. Syed, C. J. Schofield, *Proc. Natl. Acad. Sci.* **2006**, *103*, 9814–9819.
[56] J. M. Elkins, M. J. Ryle, I. J. Clifton, J. C. Dunning Hotopp, J. S. Lloyd, N. I. Burzlaff, J. E. Baldwin, R. P. Hausinger, P. L. Roach, *Biochemistry* **2002**, *41*, 5185–5192.
[57] N. P. Dunham, A. J. Mitchell, J. M. Del Río Pantoja, C. Krebs, J. M. Jr. Bollinger, A. K. Boal, *Biochemistry* **2018**, *57*, 6479–6488.
[58] L. C. Blasiak, F. H. Vaillancourt, C. T. Walsh, C. L. Drennan, *Nature* **2006**, *440*, 368–371.
[59] A. J. Mitchell, Q. Zhu, A. O. Maggiolo, N. R. Ananth, M. L. Hillwig, X. Liu, A. K. Boal, *Nat. Chem. Biol.* **2016**, *12*, 636–640.
[60] L. Dai, X. Zhang, Y. Hu, J. Shen, Q. Zhang, L. Zhang, J. Min, C.-C. Chen, Y. Liu, J.-W. Huang, R.-T. Guo, *Appl. Environ. Microbiol.* **2022**, *88*, e02497–21.
[61] W. Chang, Y. Guo, C. Wang, S. E. Butch, A. C. Rosenzweig, A. K. Boal, C. Krebs, J. M. Bollinger, *Science* **2014**, *343*, 1140–1144.
[62] E. N. Kissman, M. E. Neugebauer, K. H. Sumida, C. V. Swenson, N. A. Sambold, J. A. Marchand, D. C. Millar, M. C. Y. Chang, *Proc. Natl. Acad. Sci.* **2023**, *120*, e2214512120.
[63] J. M. Armenta, D. F. Cortes, J. M. Pisciotta, J. L. Shuman, K. Blakeslee, D. Rasoloson, O. Ogunbiyi, D. J. Jr. Sullivan, V. Shulaev, *Anal. Chem.* **2010**, *82*, 548–558.
[64] E. Campbell, M. Kaltenbach, G. J. Correy, P. D. Carr, B. T. Porebski, E. K. Livingstone, L. Afriat-Jurnou, A. M. Buckle, M. Weik, F. Hollfelder, N. Tokuriki, C. J. Jackson, *Nat. Chem. Biol.* **2016**, *12*, 944–950.
[65] S. Osuna, *WIREs Comput. Mol. Sci.* **2021**, *11*, e1502.
[66] H. A. Bunzel, J. L. R. Anderson, D. Hilvert, V. L. Arcus, M. W. van der Kamp, A. J. Mulholland, *Nat. Chem.* **2021**, *13*, 1017–1022.

[67] M. A. Maria-Solano, T. Kinateder, J. Iglesias-Fernández, R. Sterner, S. Osuna, *ACS Catal.* **2021**, *11*, 13733–13743.
[68] G. Casadevall, J. Casadevall, C. Duran, S. Osuna, *Protein Eng. Des. Sel.* **2024**, *37*, gzae005.
[69] J. W. Slater, C.-Y. Lin, M. E. Neugebauer, M. J. McBride, D. Sil, M. A. Nair, B. J. Katch, A. K. Boal, M. C. Y. Chang, A. Silakov, C. Krebs, J. M. Jr. Bollinger, *Biochemistry* **2023**, *62*, 2480–2491.
[70] R. H. Wilson, S. Chatterjee, E. R. Smithwick, A. R. Damodaran, A. Bhagi-Damodaran, *ACS Catal.* **2024**, *14*, 13209–13218.
[71] E. R. Smithwick, R. H. Wilson, S. Chatterjee, Y. Pu, J. J. Dalluge, A. R. Damodaran, A. Bhagi-Damodaran, *ACS Catal.* **2023**, *13*, 13743–13755.
[72] J. Liu, S. Chakraborty, P. Hosseinzadeh, Y. Yu, S. Tian, I. Petrik, A. Bhagi, Y. Lu, *Chem. Rev.* **2014**, *114*, 4366–4469.
[73] V. Van, J. B. Brown, C. R. O'Shea, H. Rosenbach, I. Mohamed, N.-E. Ejimogu, T. S. Bui, V. A. Szalai, K. N. Chacón, I. Span, F. Zhang, A. T. Smith, *Nat. Commun.* **2023**, *14*, 458.
[74] G. W. Larson, P. K. Windsor, E. Smithwick, K. Shi, H. Aihara, A. Rama Damodaran, A. Bhagi-Damodaran, *Biochemistry* **2023**, *62*, 3283–3292.
[75] N. Eswar, B. Webb, M. A. Marti-Renom, M. Madhusudhan, D. Eramian, M. Shen, U. Pieper, A. Sali, *Curr. Protoc. Bioinf.* **2006**, *15*, 5.6.1–5.6.30.
[76] M. A. Martí-Renom, A. C. Stuart, A. Fiser, R. Sánchez, F. Melo, A. Sali, *Annu. Rev. Biophys. Biomol. Struct.* **2000**, *29*, 291–325.
[77] A. Šali, T. L. Blundell, *J. Mol. Biol.* **1993**, *234*, 779–815.
[78] A. Fiser, R. K. G. Do, A. Šali, *Protein Sci.* **2000**, *9*, 1753–1773.
[79] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, T. E. Ferrin, *J. Comput. Chem.* **2004**, *25*, 1605–1612.
[80] D. A. Case, K. Belfon, I. Y. Ben-Shalom, S. R. Brozell, D. S. Cerutti, T. E. Cheatham III, V. W. D. Cruzeiro, T. A. Darden, R. E. Duke, G. Giambasu, M. K. Gilson, H. Gohlke, A. W. Goetz, R. Harris, S. Izadi, S. A. Izmailov, K. Kasavajhala, A. Kovalenko, R. Krasny, T. Kurtzman, T. S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, V. Man, K. M. Merz, Y. Miao, O. Mikhailovskii, G. Monard, H. Nguyen, A. Onufriev, F. Pan, S. Pantano, R. Qi, D. R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C. L. Simmerling, N. R. Skrynnikov, J. Smith, J. Swails, R. C. Walker, J. Wang, L. Wilson, R. M. Wolf, X. Wu, Y. Xiong, Y. Xue, D. M. York, P. A. Kollman, (**2020**), AMBER 2020, University of California, San Francisco.
[81] C. Tian, K. Kasavajhala, K. A. A. Belfon, L. Raguette, H. Huang, A. N. Migues, J. Bickel, Y. Wang, J. Pincay, Q. Wu, C. Simmerling, *J. Chem. Theory Comput.* **2020**, *16*, 528–552.
[82] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, D. A. Case, *J. Comput. Chem.* **2004**, *25*, 1157–1174.
[83] Z. Li, L. F. Song, P. Li, K. M. Merz, *J. Chem. Theory Comput.* **2020**, *16*, 4429–4442.
[84] J. Wang, W. Wang, P. A. Kollman, D. A. Case, *J. Mol. Graphics Modell.* **2006**, *25*, 247–260.
[85] P. Li, K. M. Jr. Merz, *J. Chem. Inf. Model.* **2016**, *56*, 599–604.
[86] S. Izadi, R. Anandakrishnan, A. V. Onufriev, *J. Phys. Chem. Lett.* **2014**, *5*, 3863–3871.
[87] D. R. Roe, T. E. Cheatham, *J. Chem. Theory Comput.* **2013**, *9*, 3084–3095.
[88] A. Nicholls, *J. Comput. Aided Mol. Des.* **2014**, *28*, 887–918.

Manuscript received: June 5, 2024
Revised manuscript received: September 5, 2024
Accepted manuscript online: October 6, 2024
Version of record online: December 5, 2024