



HHS Public Access

Author manuscript

Nat Chem Biol. Author manuscript; available in PMC 2014 December 01.

Published in final edited form as:

Nat Chem Biol. 2014 June ; 10(6): 431–436. doi:10.1038/nchembio.1503.

The Role of Distant Mutations and Allosteric Regulation on LovD Active Site Dynamics

Gonzalo Jiménez-Osés^{1,5}, Sílvia Osuna^{1,5}, Xue Gao^{2,5}, Michael R. Sawaya³, Lynne Gilson⁴, Steven J. Collier⁴, Gjalb W. Huisman^{4,*}, Todd O. Yeates^{3,*}, Yi Tang^{1,2,*}, and K. N. Houk^{1,*}

¹Department of Chemistry and Biochemistry, University of California, Los Angeles, Los Angeles, CA 90095-1569

²Department of Chemical and Biomolecular Engineering, University of California, Los Angeles, Los Angeles, CA 90095

³Molecular Biology Institute, University of California, Los Angeles, Los Angeles, CA 90095

⁴Codexis, Inc., 200 Penobscot Drive, Redwood City, California 94063

Abstract

Natural enzymes have evolved to perform their cellular functions under complex selective pressures, which often require their catalytic activities to be regulated by other proteins. We contrasted a natural enzyme, LovD, which acts on a protein-bound (LovF) acyl substrate, with a laboratory-generated variant that was transformed by directed evolution to accept instead a small free acyl thioester, and no longer requires the acyl carrier protein. The resulting 29-mutant variant is 1000-fold more efficient in the synthesis of the drug simvastatin than the wild-type LovD. This is the first non-patent report of the enzyme currently used for the manufacture of simvastatin, as well as the intermediate evolved variants. Crystal structures and microsecond molecular dynamics simulations revealed the mechanism by which the laboratory-generated mutations free LovD from dependence on protein-protein interactions. Mutations dramatically altered conformational dynamics of the catalytic residues, obviating the need for allosteric modulation by the acyl carrier LovF.

Directed evolution encompasses iterative cycles of genetic diversification and screening to generate novel proteins with desired functions.^{1,2} Due to the large number of mutations that are generally introduced to generate the dramatic increases sought after, deriving scientific

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use:http://www.nature.com/authors/editorial_policies/license.html#terms

*Correspondence and requests for materials should be addressed to K.N.H., T.O.Y., Y.T. and G.H. hok@chem.ucla.edu, yitang@ucla.edu, yeates@mbi.ucla.edu, gjalb.huisman@codexis.com.

²These authors contributed equally to this work.

Author contributions: K.N.H., Y.T., G.H. and T.O.Y. planned research; L.G. planned directed evolution experiments; X.G. and M.S. did crystallography; X.G. did kinetics, G.J.O. and S.O. performed MD simulations; all authors analyzed data; G.J.O., S.O., K.N.H., Y.T., T.O.Y., and G.H. wrote paper.

Competing financial interests: L. G. and S. C. are shareholders in Codexis. G. W. H. is a Codexis employee and holds shares and stock options in Codexis.

Additional information: Supplementary information is available in the online version of the paper. Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>.

insights from such experiments is difficult. In particular, it remains poorly understood how beneficial mutations far from the active site confer improved catalytic properties. If such an understanding could be developed, it would aid the *ab initio* design of enzymes using computational tools.³ In this study, we tracked the evolutionary trajectory by which a natural enzyme, LovD, is converted to an efficient catalyst for an unnatural product of high pharmaceutical value (simvastatin).

LovD from *Aspergillus terreus* was previously discovered to catalyze the transfer of an α -methylbutyrate side chain to the C8-hydroxy position of monacolin J acid (MJA) to yield lovastatin (Fig. 1).^{4,5} The reaction depends strongly on another protein, LovF; the acyl carrier protein (ACP) domain of LovF acts as substrate to deliver the α -methylbutyrate group to LovD, and then to MJA. The acyl transfer reaction in the LovD active site proceeds via a ping-pong mechanism, involving a covalent acyl intermediate at Ser76 (Fig. 1). The sequential acylation-deacylation reactions are promoted by acid/base catalysis involving hydrogen bonding and proton shuttling by catalytic residues Tyr188 and Lys79. Simvastatin, an important cholesterol-lowering agent, differs from the natural LovD product (lovastatin) by only one methyl group and was prepared by whole-cell biocatalysis by supplying cells with a non-natural acyl donor.⁶ However, wild-type LovD displayed poor activity towards the non-natural acyl donor.

Envisioning a potential enzymatic manufacturing route to simvastatin, laboratory-directed evolution was applied to transform LovD to a variant that accepts as a substrate the free unnatural acyl donor, α -dimethylbutyryl-S-methylmercaptpropionate (DMB-SMMP) (Fig. 1). The improvement on the catalytic proficiency of the evolved enzymes was rationalized through structural comparisons of crystal structures of LovD and evolved variants and via molecular dynamics (MD) simulations. Crystallography and nanosecond MD did not provide an explanation for the improved performance of the evolved LovD; however, microsecond simulations provided dramatic new insights.

RESULTS

Direct evolution of LovD for simvastatin production

Through limited mutagenesis studies reported earlier, we evolved LovD to a simvastatin synthase, improving its k_{cat} five-fold for the unnatural reaction.⁷ In subsequent experiments, nine rounds of directed evolution using ProSAR-based directed evolution technologies⁸ provided an enzyme that exhibits a turnover number of $\sim 25,000$ in cell lysates over the course of the reaction, representing an increase of three orders of magnitude compared to the natural enzyme, and thereby enabling large scale simvastatin manufacture^{9,10} (Online Methods). A summary of the mutations incorporated, the turnover numbers measured in cell lysates, and the kinetic data for the purified enzymes of rounds 1, 3, 6, and 9 are given in Table 1. Over the nine rounds of evolution, diversity obtained from targeted NNK libraries or from homologous sequences was recombined. A total of 61,779 LovD variants were screened from an aggregate of 200 different libraries. In rounds 1 and 3 the best variants were identified from random mutagenesis libraries (L361M in round 1 and G275S in round 3); in all other rounds, the best variant was identified from combinatorial libraries and had 2 to 6 mutations. While the binding affinities (K_M values) for the wild type LovD, LovD3 and

LovD9 fell in a narrow range (within 0.2 to 0.5 mM), we indeed observed a dramatic increase of the k_{cat} values. As a consequence, the k_{cat}/K_M value of LovD3 was more than 20 times higher than wild-type LovD, while those of LovD6 and LovD9 were around 230 and 330 times higher. The optimized mutant, LovD9, contained 29 mutations resulting in a high reactivity towards the unnatural substrate ($k_{cat}/K_M \sim 430 \text{ min}^{-1} \text{ mM}^{-1}$), and simultaneously showed complete loss of activity towards the natural α -methylbutyryl-ACP substrate. The 29 mutations found in the final optimized variant (LovD9) were scattered throughout the entire enzyme (Fig. 2a,b and Supplementary Results, Supplementary Fig. 1 and 2), in accordance with previous directed evolution studies, in which changes in residues far from the active site had a pronounced effect on enzyme activity.¹¹

Crystallographic structures of evolved mutants

Crystal structures were obtained for LovD9 and for an evolutionary intermediate from the sixth round of selection (LovD6), which displays a 63-fold improvement in k_{cat} over the wild type (Online Methods and Supplementary Table 1). Both structures displayed the expected α/β hydrolase fold, which consists of a central seven-stranded antiparallel β -sheet flanked by α -helices on either side (Supplementary Fig. 3). Different packing arrangements were found in different crystal forms, *i.e.* four monomer subunits in wild-type LovD⁷ and LovD9 and two monomer subunits in LovD6. The dimeric arrangements shared an overall similarity but were different in detail (Supplementary Fig. 4), and were presumed to be non-biological.

The active site is gradually more buried in the evolved proteins, as shown in Fig. 2c–e. Simultaneously, the width of the substrate access channel was reduced throughout the evolutionary process, as indicated by the channel solvent-accessible volumes calculated for the X-ray structures of wild type LovD and LovD6 (Supplementary Fig. 5). As a consequence, as the enzyme evolves, the active site became inaccessible to potential binding proteins such as the ACP domain of LovF. However, while this constriction explained the decrease in activity towards protein-bound α -methylbutyryl-ACP, the increase in activity toward DMP-SMMP presented an intriguing mystery.

Despite the notable changes in the shape of the binding cleft (Supplementary Fig. 5 and 6), the crystal structures showed only very minor differences in the arrangement of the catalytic residues involved in the reaction (Fig. 3a and Supplementary Fig. 7). The positions of both the backbone and side chain orientations of the key residues Lys79, Tyr188 and Ser76 are maintained in LovD, LovD6 and LovD9, and are very close to the optimum positions predicted from the quantum mechanical optimizations of the resting state (Fig. 3b) and the transition state for acylation of the serine residue (Fig. 3c). Notably, the energy barrier associated with this transition state was calculated to be the rate-limiting step of the transacylation reaction (data not shown). This comparison of crystal structure active sites therefore failed to provide an explanation for the significant increase in k_{cat} along the series of evolved mutants.

Microsecond dynamics in solution

The minimal differences observed crystallographically between the active sites structures of LovD variants prompted us to explore MD simulations to analyze potential effects of remote mutations on the active site dynamics.^{12,13} Whereas LovD wild-type and some variants are dimers in the crystal form, the existence of a single peak in all gel filtration chromatography traces (Supplementary Fig. 8) and a single spot on all SDS-PAGE gels (Supplementary Fig. 9), both corresponding to a molecular weight of 46 kDa, indicated that all the LovD proteins were monomers in aqueous solution. Hence, monomeric forms of the LovD structure were extracted from the observed crystal structures and used for unbiased all-atom MD simulations including explicit water (Online Methods). Picosecond or nanosecond simulations are generally not long enough to describe transitions between active and inactive protein conformations¹², and in our studies nanosecond simulations similarly did not provide new insights (data not shown). Therefore we analyzed the dynamics of several mutants along the evolutionary pathway at longer timescales (1–1.5 μ s) using the special purpose protein dynamics computer, ANTON.^{14–16}

MD simulations reproduced the shrinkage of the active site channel noted in the crystallographic structures (Supplementary Fig. 10) of both LovD6 and LovD9 with respect to wild-type LovD. More significantly, in the simulations of the wild-type enzyme sequence, the apparently optimum catalytic conformation of the catalytic Ser76–Lys79–Tyr188 triad observed in the crystal structure (Fig. 4a) underwent a rapid transition to an inactive conformation characterized by a very long Lys79–Tyr188 distance (~ 9 Å, Fig. 4b), which was maintained along the whole microsecond trajectory (Fig. 4c). In aqueous solution, the non-catalytic conformation of Tyr188 around χ^1 dihedral angle was stabilized by the neighboring residue Ile325, producing an unproductive hydrogen-bonded arrangement which differs dramatically from the one observed in the crystalline state. Simultaneously, the protein backbone at Tyr188 changes from an extended conformation ($\psi \sim 150^\circ$) to a turn ($\psi \sim 50^\circ$), which traps catalytic Lys79 in a non-productive conformation as well (Supplementary Fig. 11a). Additional interactions not observed in the crystal structures, such as between Tyr327 and Gly364 (Supplementary Fig. 11b,c), create a completely different hydrogen bond network in the active site.

The first round of evolution (LovD1) introduced a single mutation (L361M), which was maintained in subsequent rounds of evolution. This residue is located in a β -sheet 15 Å from the α -helix holding the active residue Ser76 (Fig. 2b). In the simulations, this single mutation relative to the wild type sequence restored the Tyr188 backbone in its catalytic conformation and maintained it for about 300 ns (Fig 4c). Thereafter, the Tyr188 side chain underwent the same detrimental conformational change observed in the wild type LovD and remained in the inactive arrangement for an additional 800 ns before returning to the active conformation. Therefore, in LovD1 the arrangement of the catalytic triad was organized for catalysis about 30% of the simulation time, which agrees with the modest increase in activity measured for this enzyme. In LovD3, which displays a 10-fold increase in k_{cat} compared to LovD, the correct arrangement of the catalytic triad was conserved $\sim 56\%$ of the simulation time, and the conformational change to the inactive arrangement occurs only after 800 ns. In LovD6 and LovD9, the catalytically active conformation of the Ser76–

Tyr188–Lys79 triad was maintained during the whole simulations. The averaged Tyr188–Lys79 distance was, in both cases, close to the optimized QM value of 3.0 Å (3.6±0.4, and 3.7±0.7 Å for LovD6 and LovD9, respectively). As depicted in Supplementary Fig. 12, the progressive weakening of the detrimental interaction of Tyr188 with Ile325, and the subsequent restoration of the catalytic triad in the catalytic conformation, correlate with the increase in the activity of the enzyme in the absence of the LovF-ACP partner, especially in the initial stages of the evolution process. While LovD6 and LovD9 both maintain the catalytic triad in an optimal conformation, LovD9 showed a $k_{\text{cat}} \sim 3$ times higher than that of LovD6 and is ~ 2 -fold more productive in simvastatin synthesis, indicating that other factors beyond the active site configuration impact catalysis.

Besides L361M, other remote mutations in the β -sheet region were found along the evolution trajectory (Fig. 2b and Supplementary Fig. 1). These buried mutations, all involving nonpolar and structurally similar residues, modify the protein backbone dynamics in such a way that the distance between the α -carbons of Gly364 in the mutated β -sheet and Tyr327, located in the same flexible loop as the catalytic Tyr188, was progressively reduced from 9.2±0.4 Å in the wild-type to 7.6±0.7 Å in LovD9 (Supplementary Fig. 13). Fig. 5a–e shows how the multiple, often non-catalytic, arrangements of Tyr188 and Tyr327 in wild-type LovD became gradually more ordered in mutants LovD1, LovD3 and LovD6 and then localized in the optimal catalytic arrangement in LovD9, as a result of an apparent lowering of the free energy of the catalytic arrangement relative to the non-catalytic conformations. The progressive tightening of the second shell residues prevents Tyr188 from drifting into non-catalytic conformations, and as a consequence the catalytic activity of the triad is retained.

Thus, according to the MD simulations the catalytically-competent conformations predicted by QM calculations were progressively stabilized and appear with increasing frequency as a result of the dynamic effects of remote mutations on the active site residues (Fig. 4c and Supplementary Fig. 14). The relative free energy of the catalytic conformation, reflected in the population observed in dynamics simulations, is lowered gradually during enzyme optimization.

Role of interprotein interactions in active site dynamics

As noted above, the crystal structures of different enzyme variants all showed a very similar, catalytically competent arrangement of the active site, irrespective of their catalytic proficiencies (Fig. 3a). MD models for the monomeric enzymes in solution showed wide variations in the population of these active site conformers as evolution progresses. A possible source of this different behavior is that associations of proteins in the crystalline state somehow influence the conformational states of the active site of LovD, in the same way that association of LovF-ACP influences the active site. The effects of protein-protein interactions on the structure and dynamics of the active site were analyzed by performing MD simulations on both the LovD dimer as observed in the crystal form, and the complex formed between LovD and the LovF acyl carrier domain (ACP). While a model for simulating the dynamics of the LovD dimer was directly available from the X-ray structure (Fig. 6a), no structure of the LovD/LovF-ACP complex was available; complexes involving

other modular synthases are likewise unresolved.^{17,18} Computational models of LovD/LovF-ACP were therefore generated (Fig. 6b), both including and lacking the 4'-phosphopantetheine (PPN) acylating group, using homology modelling and protein docking techniques.¹⁹ MD simulations demonstrated (Fig. 6c) that protein-protein interactions lead to stabilization of the active site residues in an optimal catalytic conformation, which was conserved permanently in the LovD dimer, and for most of the simulation in the LovD/ACP complex. In the LovD dimer, the flexibility of Tyr188 was substantially reduced compared to the monomeric state (Fig. 5f and Supplementary Fig. 15). The presence of a flexible loop (323–329) from another protein chain in the vicinity of the active site precluded the great distortion of the active site occurring in the monomer state in solution, although some mobility is still observed in the catalytic residues.

In the LovD-ACP complexes, the detrimental conformational change of Tyr188 side chain did not take place along the whole trajectory either (Supplementary Fig. 16). The distance between Tyr188 and Lys79 was reduced from 8.7 ± 0.5 Å for monomeric LovD to 3.7 ± 0.6 Å for LovD-ACP-PPN. In the case of LovD-ACP lacking the PPN group, Tyr188 was still very flexible and the Tyr188-Lys79 distance was elongated to 5.2 ± 1.6 , indicating the key role of the PPN acylating arm to properly orient Tyr188 for catalysis. In agreement with the trend exhibited by the evolved LovD1–9 variants, the detrimental Tyr188-Ile324 interaction was elongated from 3.0 ± 0.8 Å in monomeric LovD to 6.8 ± 1.2 Å and 9.3 ± 1.5 Å for LovD-ACP and LovD-ACP-PPN, respectively. Interestingly, after 850 ns, the long acylating PPN arm abandoned the active site, and the Tyr188-Ile324 interaction was substantially reduced to approximately 7 Å, approaching the non-catalytic arrangement. The backbone arrangement of Tyr188 remained in an extended conformation in both LovD-ACP-PPN and LovD-ACP complexes. Overall, these simulations supported the important role protein-protein interactions play in stabilizing the active site of wild type LovD.

DISCUSSION

Enzymes may be improved as a result of various features, including lowering of the activation barrier (increasing k_{cat}) and increasing substrate binding (lowering K_M), but also protein expression, folding, stability, and other characteristics that do not directly relate to the intrinsic activity of the enzyme. Given the scattered distribution of mutations throughout the protein, the specific role of every mutation introduced in LovD1–9 cannot be assessed exactly; moreover, most of these positions are located far away from the active site. Based on their close location to the active site, mutations N191S, N191G, L192I (rounds 2–5) are likely to affect catalysis directly. These residues are situated at the end of an α -helix nearby the catalytic triad (Supplementary Fig. 17a) and prevent the detrimental interaction between the long side chain amide group of the natural Asn191 and the backbone carbonyl of Tyr188. The role of the three other mutations located close to the active site (L174F, A178L, S172N) which also appear early in the evolutionary pathway (rounds 2–4) is more difficult to assess since they are located in flexible loops, and could be affecting catalysis indirectly. Alternatively, these three mutations may affect folding kinetics as they appear to be part of a loop structure that wraps around a second loop. Triggered by the first key L361M mutation, up to four modifications were made at four consecutive strands of the internal β -sheet; three of them (V370I, A383V, I35L) were introduced sequentially at the last stages of the

evolution process (rounds 5–8) (Supplementary Fig. 17b). Such an evolution pattern where subsequent mutations are spatially adjacent to other mutations has been observed previously,¹¹ and may increase enzyme stability through improved hydrophobic packing of the core. Another group of mutations (A247S, A9V, K26E, H404K, I4N, R28S) is located in the ACP-PPN binding motif (Supplementary Fig. 17b) hypothesized from our homology model/docking studies. This motif includes two antiparallel α -helices flanked by flexible loops, and the mobile tail (residues 1–12) found to contribute to ACP binding in our MD simulations. These mutations often include charged residues and might disrupt the formation of the protein-protein complex occurring in the natural enzyme; the conserved mutation K26E,⁷ which changes the electrostatic environment of this local domain drastically, is probably key to produce this effect. Finally, based on ProSAR analysis during the evolution work we believed that another set of late mutations (round 5–9) Q241M, A261V and A261H provide thermostability to the protein, and N43R, D96R, H404K reduce aggregation.

In natural LovD, the interaction with LovF-ACP plays a key role in ordering the active site for catalysis. The interaction with the ACP partner provides an induced fit to organize the catalytic triad. This interaction must be crucial for substrate specificity towards the acyl group bound to LovF. Such a mechanism suppresses LovD activity towards other acyl donors in the cell, such as those involved in fatty acid biosynthesis. Some remote LovD homologs, such as EstB,²⁰ are active on small molecule substrates, while LovD is prevented from reacting with these substrates because the catalytic machinery is not properly poised without binding to the LovF-ACP carrier. Distinct evolutionary pressures have apparently led to diverse enzymes in this superfamily having varied substrate specificities and dependencies on protein interactions. We used directed evolution to convert an enzyme in this family that is naturally activated by protein binding into one that does not require a protein partner for catalysis.

The influence of remote mutations on the free energy of the catalytic conformation of the active site were revealed in our study by long timescale all-atom unbiased simulations that illuminated how numerous mutations remote from the active site increased, by three orders of magnitude, the catalytic activity in an enzyme acting on an unnatural substrate. Directed evolution increases activity for the small molecule substrate DMP-SMMP by rendering the catalytic site organization independent of LovF-ACP. This was accomplished by distributed mutations that lower the free energy of the catalytic arrangement of the triad relative to non-catalytic arrangements. The dynamics that produce this type of preorganization²¹ have been the focus of many investigations of enzyme catalysis.²² In previous studies, remote mutations introduced in dihydrofolate reductase have been shown to interfere with the network of protein motions coupled to the reaction coordinate.²³ We have shown how mutations incorporated into LovD are the evolutionary response to increase activity when the protein partner LovF is absent. These findings showed how features of natural enzymes that are necessary for effective chemistry in the cell can be modified to provide more effective catalysts for a manufacturing setting. The influence of protein-protein interactions in natural enzymes must be compensated for in the engineering of new or improved biocatalysts by evolution or design.

The computational design of stable and functional enzymes is a longstanding and very challenging goal. Our approach is to learn how to predict beneficial mutations and to incorporate that understanding into the computational design process. Major achievements in enzyme engineering reported recently^{11,24} and the results reported here for LovD share a common and intriguing feature: many mutations leading to high activity are scattered remotely from the active site. By contrast, design protocols currently in use often involve manual reversion of second-shell mutations to the native residues, based on the idea that such mutations will be deleterious to folding. The results presented here demonstrated the crucial role that remote mutations can have on modulation of the catalytic proficiency of the active site by conformational changes transmitted throughout the protein backbone. Future design protocols will need to predict and allow remote mutations, evaluated by μ s MD. While computationally expensive, μ s MD simulations are capable of predicting the structural consequences of mutations, particularly how mutations remote from the active site can alter the active site geometry. Experience shows that precise control of catalytic group geometries at the ideal positions for catalysis is crucial for high activity, and the remote mutation/MD evaluation strategy is a way to build this into the computational design process.

ONLINE METHODS

Production of the LovD gene

The acyltransferase encoding gene *lovD* from wild-type *Aspergillus terreus* (GenBank AAD34555.1) was designed for expression in *E. coli* using standard codon optimization. Genes were synthesized using oligonucleotides composed of 42 nucleotides and cloned into expression vector pCK110900,²⁵ under the control of a *lac* promoter. Resulting plasmids were transformed into *E. coli* W3110 or *E. coli* BL21 using standard methods.

Production of LovD libraries

LovD variants were obtained via directed evolution using APS methods for library generation,²⁶ HTP screening (below) and ProSAR analysis²⁷ of the data. Over the course of nine iterative rounds of *in vitro* evolution, 216 libraries were constructed and 61,779 variants were screened resulting in improved LovD variant such as LovD1, LovD3, LovD6 and LovD9 as representative hits from the screens in rounds 1, 3, 6 and 9. Complete protein sequences and detailed procedures are described.^{27,9} Mutations in LovD were generated by methods described previously.¹¹ Beneficial mutations identified in HTP screening using the ProSAR algorithm²⁷ were recombined via semi-synthetic shuffling methods.¹¹

Preparation of soluble LovD cell lysate

Single colonies of *E. coli* containing a plasmid encoding a variant LovD polypeptide of interest was inoculated into 50 mL 2xYT broth containing 30 μ g/ml chloramphenicol and 1% glucose. Cells were grown overnight at 30 °C with shaking at 250 rpm. The culture was diluted into 250 ml 2xYT broth containing 30 μ g/ml chloramphenicol to an OD600 of 0.2 and expression of *lovD* was subsequently induced at OD600 0.6–0.8 by addition of isopropyl-D-thiogalactoside (IPTG) to a final concentration of 1 mM. Incubation was then continued overnight. Cells were harvested by centrifugation (2400 g, 15 min, 4 °C) and the

cell pellet re-suspended with an equal volume of cold (4°C) 50 mM phosphate buffer (pH 8.5) and harvested by centrifugation as above. The washed cells were resuspended in two volumes of the cold phosphate buffer and passed through a French Press (18,000 psi, 4 °C). Cell debris was removed by centrifugation (7700 g, 30 min, 4 °C). The cleared lysate supernatant was collected and stored at -20°C. Lyophilization of frozen clear lysate provides a dry shake-flask powder of crude LovD polypeptide. Alternatively, the cell pellet (before or after washing) can be stored at 4°C or -80°C.

Purification of LovD polypeptides for kinetic assays and crystallization

The wild type LovD as well as variants were cloned into pET28a vector with an *N*-terminal 6Xhis tag. The Cys40 and Cys60 were mutated to Ala and Asn to prevent oxidative multimerization as previously demonstrated.¹⁰ The resulting plasmids were transferred into *E. coli* BL21 for protein expression. All LovD proteins were purified to homogeneity by Ni-NTA agarose affinity chromatography and eluted with an increasing gradient of imidazole in buffer A (50 mM Tris-HCl, 500 mM NaCl, pH 7.9). Protein concentrations were qualitatively assessed by SDS-PAGE and quantitatively determined by the Bradford protein assay using bovine serum albumin as the standard. (Supplementary Fig. 9). Samples for gel filtration chromatography were prepared at 10 mg/ml concentration and loaded onto a Superdex 200 (GE Healthcare) column and eluted at a flow rate of 0.5 ml min⁻¹ in 50 mM Tris pH 8.0 and 2 mM DTT. The final concentration of samples was in the 20–40 µM range.

Synthesis of DMB-SMMP

α -dimethylbutyryl-*S*-methyl 3-mercaptopropionate (DMB-SMMP) was prepared according to reported procedures (ref. 6).

HPLC method for determining conversion of Na⁺-monacolin J to Na⁺-simvastatin

At 10 g/L LovD and 4 equivalents of DMB-SMMP, 50% conversion of 3 g/L monacolin J was obtained in an 18 h reaction, corresponding to a volumetric productivity of 0.0024 g/L h g_{LovD}. Cell free lysates of LovD variants were screened under increasingly stringent conditions over nine rounds of evolution, during which the concentration of monacolin J was gradually increased, and the concentration of DMB-SMMP gradually decreased. The best LovD9 variant produced simvastatin at a rate of 2.70 g/L h g_{LovD} (75 g/L monacolin J, was 97% converted in a 36 h reaction using 0.75 g/L LovD9). Cell lysates made in 100 mM triethanolamine buffer were pre-incubated for 1 to 24 h at 35 to 45 °C prior to the simvastatin forming reaction. Over the course of the LovD evolution the pH of the buffer was increased from pH 8 to 9.5. LovD activity assays were set-up in 96-well plates using 10–20 µL cell free extract, 3–75 g/L Na⁺-monacolin J and incubated at room temperature to 45 °C for 24 hrs. Conversion of Na⁺-monacolin J to Na⁺-simvastatin was determined using an HPLC 1200 (Agilent) equipped with a Gemini[®] C18 column (4.6 × 50 mm). For the assay, 10 µL samples were eluted with a 52% (v/v) aqueous solution of acetonitrile containing 0.1 % trifluoroacetic acid (TFA) at a flow rate of 1.5 mL/min and a temperature of 30 °C. The eluate was monitored at 238 nm. Under these conditions, the retention times of monacolin J acid sodium salt, DMB-SMMP, and simvastatin sodium salt are approximately 0.8, 2.9, and 3.9 min respectively.

Kinetic characterization of the LovD variants

To access the kinetic parameters (k_{cat} and K_M), of the different variants, the concentration of MJA was varied from 0.25 to 5 mM, while the DMB-SMMP concentration was fixed at 2 mM to obtain a Michaelis-Menten kinetic curve. The assays were performed at room temperature in 50 mM HEPES (pH = 7.9) with 10 mM MgCl_2 . Dimethyl sulfoxide (DMSO) was added to a final concentration of 10% to facilitate the solubilization of DMB-SMMP. Since LovD3, LovD6 and LovD9 were expected to convert MJA to SVA much faster than wild-type, reaction time and enzyme concentration were optimized to get the initial reaction velocity. LovD wild-type and LovD1 were set up to a final concentration 10 μM , while only 1 μM LovD3, LovD6 and LovD9 were used in the assays. Also, the reaction time for LovD wild-type and LovD1 was 1 h, but only 5 min for LovD3, LovD6 and LovD9. At the desired time point, an aliquot of the reaction mixture was removed, quenched and clarified of protein using with ethyl acetate (EA) containing 1% trifluoroacetic acid (TFA). The organic phase was separated, dried, resolubilized by acetonitrile (ACN), and samples were analyzed by HPLC (Shimadzu) using a Luna 5 μm , 2.0 mm \times 100 mm C18 reverse-phase column (Phenomenex). Samples were separated on a linear gradient of 5–95% (v/v) ACN in water (0.1% (v/v) TFA) for 30 min at a flow rate of 0.1 ml min^{-1} followed by isocratic 95% (v/v) ACN in water (0.1% (v/v) TFA) for 5 min. Conversion of MJA to SVA was measured by integration of the peaks at 238 nm. Measurements were made in triplicate.

Crystallization of LovD6 and LovD9

Additional C40A and C60N mutations were introduced in LovD6 and LovD9 to facilitate crystallization. LovD6 was cloned into pET28a vector with an *N*-terminal 6Xhis tag. However, well-ordered crystals of LovD9 could not be obtained from this version, so we moved the 6Xhis tag to the *C*-terminus. Both LovD6 and LovD9 were purified as previously described. The proteins were then dialysed overnight into 50 mM Tris pH 8.0, 150 mM NaCl and 10 mM DTT (using Spectra/Por molecular porous membrane tubing MWCO 6~8,000 Da). LovD proteins were concentrated to 20 mg ml^{-1} (using Amicon Ultra 15 MWCO 30,000 Da) for crystallization experiments. Crystals of LovD6 were grown at room temperature by a hanging drop vapor diffusion method using a 1:1 protein to reservoir solution ratio for a total drop size of 2 μl . Diffraction quality crystals were obtained in 1~2 days when using 50 mM sodium citrate, 16% propanol, 18% PEG 4,000, 1 mM DTT, as a reservoir solution. In preparation for data collection, crystals were briefly soaked in 30%/70% mixture of glycerol/reservoir solution and flash frozen. Crystals of LovD9 were grown at room temperature by the hanging drop vapor diffusion method using a 2:1 protein to reservoir solution ratio for a total drop size of 3 μl . Diffraction quality crystals were obtained in 1~2 days when using 0.2 M MgCl_2 , 0.1 M Bis-Tris pH 5.5, 18% PEG3,350 and 10 mM DTT, as a reservoir solution.

Data collection and structure determination

The LovD9 crystal belonged to space group P1 with four protein molecules in the asymmetric unit. X-ray diffraction data were collected at the Advanced Photon Source (Argonne National Laboratory), beamline 24-ID-C, using a Quantum 315 CCD detector (ADSC). Crystals were cryo-protected by a quick dip in a solution consisting of 6.5 μl

reservoir and 3.5 μl 50% (w/v) D-(+)-trehalose. Crystals were cooled to 100 K during the data collection. One-hundred-twenty 1.0° oscillation frames were collected at a wavelength of 0.9791 \AA . Data reduction and scaling were performed using XDS.²⁸ Diffraction to 3.2 \AA resolution was observed. The LovD6 crystal belonged to space group $P2_1$ with two protein molecules in the asymmetric unit. X-ray diffraction data were collected at the same beamline as LovD9. Crystals were cryo-protected by a quick dip in a solution consisting of 6.5 μl reservoir and 3.5 μl glycerol. Crystals were cooled to 100 K during the data collection. One-hundred-eighty 1.0° oscillation frames were collected at a wavelength of 0.9641 \AA . Data reduction and scaling were performed using XDS. Diffraction to 1.8 \AA resolution was observed. The crystal structures were determined by the molecular replacement method using the program PHASER²⁹ and search model LovD G5 mutant, PDB entry 3HLC⁴. The models were refined using REFMAC5³⁰ and Buster/TNT³¹ with TLS parameterization of domain disorder.³² After each refinement step, the models were visually inspected in COOT,³³ using both 2Fo-Fc and Fo-Fc difference maps. The models were validated with the following structure validation tools: PROCHECK,³⁴ ERRAT³⁵ and VERIFY3D.³⁶ For the LovD9 structure, 90.4% of the residues are within the most favored region of the Ramachandran plot, 9.3% were in additional allowed regions, and 0.3% were in generously allowed regions. There were no residues in disallowed regions. For the LovD6 structure, 91.9% of the residues are within the most favored region of the Ramachandran plot, 7.8% were in additional allowed regions, and 0.3% were in generously allowed regions. There were no residues in disallowed regions. The Errat scores, 93.7% and 94.5%, for LovD9 and LovD6 structures respectively, indicates that these percentages of residues fall within the 95% confidence limit for correctly modeled regions. Verify3D reports 100% and 96.2% of the residues have an averaged 3D-1D score greater than 0.2 for LovD9 and LovD6 structures, respectively. Data collection and refinement statistics are reported in Supplementary Table 2.

DFT calculations

All calculations were carried out with the M06-2X hybrid functional³⁷ and 6-31G(d,p) basis set. Full geometry optimizations and transition state searches were carried out with the Gaussian 09 package (Frisch, M. J. *et al.*, Gaussian, Inc., 2009). Frequency analyses were carried out at the same level used in the geometry optimizations, and the nature of the stationary points was determined in each case according to the appropriate number of negative eigenvalues of the Hessian matrix. Bulk solvent effects were considered implicitly during optimization through the IEF-PCM polarizable continuum model³⁸ as implemented in Gaussian 09. The internally stored parameters for diethylether ($\epsilon = 4$) were used as an estimation of the dielectric permittivity in the enzyme active site.^{39,40}

Homology modeling and protein-protein docking

A homology model of the PPN binding region of the LovF sequence⁴ was obtained using a solution (NMR) structure of a fungal type I polyketide synthase ACP domain⁴¹ as a template and using MOE 2012 (Chemical Computing Group, 2013). This small protein was then docked into a hypothesized suitable binding region of wild-type LovD using ClusPro⁴²⁻⁴⁵ and refined with Rosie (RosettaDock Online Server).⁴⁶ Finally, the 2-methylbutyryl form of PPN was modeled at the active ACP serine (Ser44) and docked in a proper orientation along

the binding channel and into the active site close to the catalytic triad of LovD using Gold 5.1.⁴⁷ To dissect the influence of protein-protein contacts and the presence of the long acylating arm, a control model lacking the 2MB-PPN function was also built in a similar way.

Generation of computational models for the mutants

The crystallographic structure of G5 mutant obtained previously⁷ (PDB ID 3HLG) was used as a template for defining a starting model for MD simulation of the various DE mutants. This was the only structure among all crystallized LovD variants for which the entire protein chain could be visualized by X-ray crystallography. Mutations corresponding to the DE variants were introduced into the model using RosettaDesign⁴⁸ to obtain low-energy conformations after stochastic repacking and minimization of the side chains. Structures were then refined through 80 ns MD simulations in explicit water using AMBER (see below). After the MD equilibration, the generated structures were in good agreement with the partially resolved X-ray structures.

MD simulations

Long-timescale Molecular Dynamics (MD) simulations were performed using either the special-purpose ANTON (D. E. Shaw Research)¹⁴ machine or an in-house GPU cluster. Unless otherwise stated, monomer proteins in explicit water were subjected to simulation, according to experimental observations. When necessary, homo- and heterodimer protein-protein complexes were also modeled. Parameterization and preparation for the simulations were carried out externally prior to production-level MD on ANTON. Substrate parameters for the MD simulations were generated within the *antechamber* module of AMBER 12 (Case, D. A. *et al.*, UCSF, 2012) using the general AMBER force field (GAFF),⁴⁹ with partial charges set to fit the electrostatic potential generated at the HF/6–31G(d) level by the RESP⁵⁰ model. The charges were calculated according to the Merz-Singh-Kollman scheme^{51,52} using Gaussian 09. Each enzyme was immersed in a pre-equilibrated truncated cuboid box with a 10 Å buffer of TIP3P⁵³ water molecules using the *leap* module, resulting in the addition of around 25,000 solvent molecules. The systems were neutralized by addition of explicit counter ions (Na⁺ and Cl⁻). All subsequent calculations were done using the widely tested Stony Brook modification of the Amber 99 force field.⁵⁴ A two-stage geometry optimization approach was performed. The first stage minimizes the positions of solvent molecules and ions imposing positional restraints on solute by a harmonic potential with a force constant of 500 kcal mol⁻¹ Å⁻², and the second stage is an unrestrained minimization of all the atoms in the simulation cell. The systems are gently heated using six 50 ps steps, incrementing the temperature 50 K each step (0–300 K) under constant-volume and periodic-boundary conditions. Water molecules are treated with the SHAKE algorithm such that the angle between the hydrogen atoms is kept fixed. Long-range electrostatic effects are modeled using the particle-mesh-Ewald method.⁵⁵ An 8 Å cutoff was applied to Lennard-Jones and electrostatic interactions. Harmonic restraints of 10 kcal/mol are applied to the solute, and the Langevin equilibration scheme is used to control and equalize the temperature. The time step is kept at 1 fs during the heating stages, allowing potential inhomogeneities to self-adjust. Each system is then equilibrated without restraints for 4 ns with a 2 fs timestep at a constant pressure of 1 atm and temperature of 300 K. After the

systems were equilibrated in the NPT ensemble, the frame of these simulations with the volume closest to the average volume were chosen as the starting structures for the subsequent long MD simulations on the ANTON hardware. Then 1.2–1.6 μ s production MD simulations on ANTON specialized hardware were performed for each of the systems in the NVT ensemble and periodic-boundary conditions using the Nose-Hoover thermostat with a relaxation time of 1.0 ps. A cutoff of 9.5 Å for the Lennard-Jones and the short-range electrostatic interactions was used. For the long-range electrostatic interactions the k-Gaussian split Ewald method⁵⁶ with a $32 \times 32 \times 32$ grid and the Particle Mesh Ewald method⁵⁷ with a $32 \times 32 \times 32$ grid and a fifth-order interpolation scheme were used in the Desmond simulations. A multistep r-RESPA scheme⁵⁸ was employed for the integration of the equations of motion with time steps of 2.5 fs, 2.5 fs and 5.0 fs for the bonded, short-range non-bonded and long-range non-bonded interactions, respectively. In all simulations performed on ANTON, geometry and velocity snapshots were saved every 100 ps to provide 10,000 data points upon completion of the simulation. Trajectories were merged and converted into AMBER format using the VMD software⁵⁹ and were analyzed using the AMBERTOOLS utilities (*ptraj* module). Solvent-accessible volumes of active site entrance channels⁶⁰ were measured using ³V (<http://3vee.molmovdb.org/>).

Statistics

Data are represented as mean values \pm s.d.

The coordinates of the X-ray structures of mutants LovD6 and LovD9 have been deposited in the Protein Data Bank (<http://www.rcsb.org/>) with PDB codes 4LCL and 4LCM, respectively.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was funded by the National Institutes of Health through grant 1R01GM097200 (to Y.T., T.O.Y. and K. N. H.) and GM075962 (to K.N.H.). G.J.O and S.O. were supported by MCE (EX2012-1063) and EU Marie Curie (PIOF-GA-2009-252856) postdoctoral fellowships, respectively. We thank the LovD evolution team at Codexis, Inc. for generating the variants used for this work. ANTON simulations were performed on the National Resource for Biomedical Supercomputing (NRBSC) at the Pittsburgh Supercomputing Centre with funding from the National Institute of General Medical Sciences under grant RC2GM093307. Additional calculations were performed on the MinoTauro cluster at BSC, and Hoffman2 and Dawson2 GPU clusters at UCLA and the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by the National Science Foundation (OCI-1053575).

References

1. Arnold FH, Volkov AA. Directed evolution of biocatalysts. *Curr. Opin. Chem. Biol.* 1999; 3:54–59. [PubMed: 10021399]
2. Benkovic SJ, Hammes-Schiffer S. A perspective on enzyme catalysis. *Science.* 2003; 301:1196–1202. [PubMed: 12947189]
3. Kiss G, Çelebi-Ölçüm N, Moretti R, Baker D, Houk KN. Computational enzyme design. *Angew. Chem. Int. Ed.* 2013; 52:5700–5725.

4. Kennedy J, et al. Modulation of polyketide synthase activity by accessory proteins during lovastatin biosynthesis. *Science*. 1999; 284:1368–1372. [PubMed: 10334994]
5. Xie X, Meehan MJ, Xu W, Dorrestein PC, Tang Y. Acyltransferase mediated polyketide release from a fungal megasynthase. *J. Am. Chem. Soc.* 2009; 131:8388–8389. [PubMed: 19530726]
6. Xie X, Tang Y. Efficient synthesis of simvastatin by use of whole-cell biocatalysis. *Appl. Environ. Microbiol.* 2007; 73:2054–2060. [PubMed: 17277201]
7. Gao X, et al. Y. Directed evolution and structural characterization of a simvastatin synthase. *Chem. Biol.* 2009; 16:1064–1074. [PubMed: 19875080]
8. Fox RJ, et al. Improving catalytic function by ProSAR-driven enzyme evolution. *Nat Biotech.* 2007; 25:338–344.
9. Gilson, L., et al. Variant LovD polypeptides and their uses. Patent. WO2011/041231. 2011.
10. Collier, SJ.; Teo, EL.; Sukurmaran, J.; Wilson, RJ.; Xu, J. Improved LovD acyltransferase mediated acylation. Patent. WO2011/041233. 2011.
11. Savile CK, et al. Biocatalytic asymmetric synthesis of chiral amines from ketones applied to sitagliptin manufacture. *Science*. 2010; 329:305–309. [PubMed: 20558668]
12. Shan Y, Arkhipov A, Kim ET, Pan AC, Shaw DE. Transitions to catalytically inactive conformations in EGFR kinase. *Proc. Natl. Acad. Sci. U. S. A.* 2013; 110:7270–7275. [PubMed: 23576739]
13. Bhabha G, et al. A dynamic knockout reveals that conformational fluctuations influence the chemical step of enzyme catalysis. *Science*. 2011; 332:234–238. [PubMed: 21474759]
14. Shaw DE, et al. Anton, a special-purpose machine for molecular dynamics simulation. *Commun. ACM.* 2008; 51:91–97.
15. Lindorff-Larsen K, Piana S, Dror RO, Shaw DE. How fast-folding proteins fold. *Science*. 2011; 334:517–520. [PubMed: 22034434]
16. Shaw DE, et al. Atomic-level characterization of the structural dynamics of proteins. *Science*. 2010; 330:341–346. [PubMed: 20947758]
17. Maier T, Jenni S, Ban N. Architecture of mammalian fatty acid synthase at 4.5 Å resolution. *Science*. 2006; 311:1258–1262. [PubMed: 16513975]
18. Jenni S, et al. Structure of fungal fatty acid synthase and implications for iterative substrate shuttling. *Science*. 2007; 316:254–261. [PubMed: 17431175]
19. Agarwal V, Lin S, Lukk T, Nair SK, Cronan JE. Structure of the enzyme-acyl carrier protein (ACP) substrate gatekeeper complex required for biotin synthesis. *Proc. Natl. Acad. Sci. U. S. A.* 2012; 109:17406–17411. [PubMed: 23045647]
20. Wagner UG, Petersen EI, Schwab H, Kratky C. EstB from *Burkholderia gladioli*: A novel esterase with a β -lactamase fold reveals steric factors to discriminate between esterolytic and β -lactam cleaving activity. *Protein Sci.* 2002; 11:467–478. [PubMed: 11847270]
21. Adamczyk AJ, Cao J, Kamerlin SCL, Warshel A. Catalysis by dihydrofolate reductase and other enzymes arises from electrostatic preorganization, not conformational motions. *Proc. Natl. Acad. Sci. U. S. A.* 2011; 108:14115–14120. [PubMed: 21831831]
22. Schulenburg, C.; Hilvert, D. *Top. Curr. Chem. Vol. Ch. 411.* Berlin Heidelberg: Springer; 2013. p. 1-27.
23. Wong KF, Selzer T, Benkovic SJ, Hammes-Schiffer S. Impact of distal mutations on the network of coupled motions correlated to hydride transfer in dihydrofolate reductase. *Proc. Natl. Acad. Sci. U. S. A.* 2005; 102:6807–6812. [PubMed: 15811945]
24. Coelho PS, Brustad EM, Kannan A, Arnold FH. Olefin cyclopropanation via carbene transfer catalyzed by engineered cytochrome P450 enzymes. *Science*. 2013; 339:307–310. [PubMed: 23258409]
25. Davis, S.; Huisman, G.; Jenne, S.; Krebber, A.; Newman, L. Ketoreductase polypeptides and related polynucleotides. Patent. US2006/0195947. 2006.
26. Colbeck, J., et al. Combined automated parallel synthesis of polynucleotide variants. Patent. US8383346. 2013.
27. Fox RJ, et al. Improving catalytic function by ProSAR-driven enzyme evolution. *Nat. Biotech.* 2007; 25:338–344.

28. Kabsch W. XDS. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* 2010; 66:125–132. [PubMed: 20124692]
29. McCoy AJ, et al. Phaser crystallographic software. *J. Appl. Crystallogr.* 2007; 40:658–674. [PubMed: 19461840]
30. Murshudov GN, Vagin AA, Dodson EJ. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* 1997; 53:240–255. [PubMed: 15299926]
31. Blanc E, et al. Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* 2004; 60:2210–2221. [PubMed: 15572774]
32. Winn MD, Murshudov GN, Papiz MZ. Macromolecular TLS refinement in REFMAC at moderate resolutions. *Macromolecular Crystallography, Pt D.* 2003; 374:300–321.
33. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of COOT. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* 2010; 66:486–501. [PubMed: 20383002]
34. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* 1993; 26:283–291.
35. Colovos C, Yeates TO. Verification of protein structures: Patterns of nonbonded atomic interactions. *Protein Sci.* 1993; 2:1511–1519. [PubMed: 8401235]
36. Luthy R, Bowie JU, Eisenberg D. Assessment of protein models with three-dimensional profiles. *Nature.* 1992; 356:83–85. [PubMed: 1538787]
37. Zhao Y, Truhlar D. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theor. Chem. Acc.* 2008; 120:215–241.
38. Scalmani G, Frisch MJ. Continuous surface charge polarizable continuum models of solvation. I. General formalism. *J. Chem. Phys.* 2010; 132:114110. [PubMed: 20331284]
39. Schutz CN, Warshel A. What are the dielectric “constants” of proteins and how to validate electrostatic models? *Proteins: Struct., Funct., Bioinf.* 2001; 44:400–417.
40. Li L, Li C, Zhang Z, Alexov E. On the dielectric “constant” of proteins: smooth dielectric function for macromolecular modeling and its implementation in Delphi. *J. Chem. Theory Comput.* 2013; 9:2126–2136. [PubMed: 23585741]
41. Wattana-amorn P, et al. Solution structure of an acyl carrier protein domain from a fungal type I polyketide synthase. *Biochemistry.* 2010; 49:2186–2193. [PubMed: 20136099]
42. Kozakov D, et al. Achieving reliability and high accuracy in automated protein docking: Cluspro, PIPER, SDU, stability analysis in CAPRI rounds 13–19. *Proteins: Struct., Funct., Bioinf.* 2010; 78:3124–3130.
43. Kozakov D, Brenke R, Comeau SR, Vajda S. PIPER: An FFT-based protein docking program with pairwise potentials. *Proteins: Struct., Funct., Bioinf.* 2006; 65:392–406.
44. Comeau SR, Gatchell DW, Vajda S, Camacho CJ. ClusPro: an automated docking and discrimination method for the prediction of protein complexes. *Bioinformatics.* 2004; 20:45–50. [PubMed: 14693807]
45. Comeau SR, Gatchell DW, Vajda S, Camacho CJ. ClusPro: a fully automated algorithm for protein–protein docking. *Nucleic Acids Res.* 2004; 32:W96–W99. [PubMed: 15215358]
46. Lyskov S, Gray JJ. The RosettaDock server for local protein–protein docking. *Nucleic Acids Res.* 2008; 36:W233–W238. [PubMed: 18442991]
47. Verdonk ML, Cole JC, Hartshorn MJ, Murray CW, Taylor RD. Improved protein–ligand docking using GOLD. *Proteins: Struct., Funct., Bioinf.* 2003; 52:609–623.
48. Richter F, Leaver-Fay A, Khare SD, Bjelic S, Baker D. De novo enzyme design using Rosetta3. *PLoS One.* 2011; 6:e19230. [PubMed: 21603656]
49. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA. Development and testing of a general Amber force field. *J. Comput. Chem.* 2004; 25:1157–1174. [PubMed: 15116359]

50. Bayly CI, Cieplak P, Cornell W, Kollman PA. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J. Phys. Chem.* 1993; 97:10269–10280.
51. Besler BH, Merz KM, Kollman PA. Atomic charges derived from semiempirical methods. *J. Comput. Chem.* 1990; 11:431–439.
52. Singh UC, Kollman PA. An approach to computing electrostatic charges for molecules. *J. Comput. Chem.* 1984; 5:129–145.
53. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 1983; 79:926–935.
54. Wang J, Cieplak P, Kollman PA. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.* 2000; 21:1049–1074.
55. Darden T, York D, Pedersen L. Particle mesh Ewald: An N-log(N) method for Ewald sums in large systems. *J. Chem. Phys.* 1993; 98:10089–10093.
56. Shan Y, Klepeis JL, Eastwood MP, Dror RO, Shaw DE. Gaussian split Ewald: A fast Ewald mesh method for molecular simulation. *J. Chem. Phys.* 2005; 122:054101.
57. Essmann U, et al. A smooth particle mesh Ewald method. *J. Chem. Phys.* 1995; 103:8577.
58. Tuckerman M, Berne BJ, Martyna GJ. Reversible multiple time scale molecular dynamics. *J. Chem. Phys.* 1992; 97:1990.
59. Humphrey W, Dalke A, Schulten K. VMD: Visual molecular dynamics. *J. Mol. Graphics Modell.* 1996; 14:33–38.
60. Voss NR, Gerstein M, Steitz TA, Moore PB. The geometry of the ribosomal polypeptide exit tunnel. *J. Mol. Biol.* 2006; 360:893–906. [PubMed: 16784753]

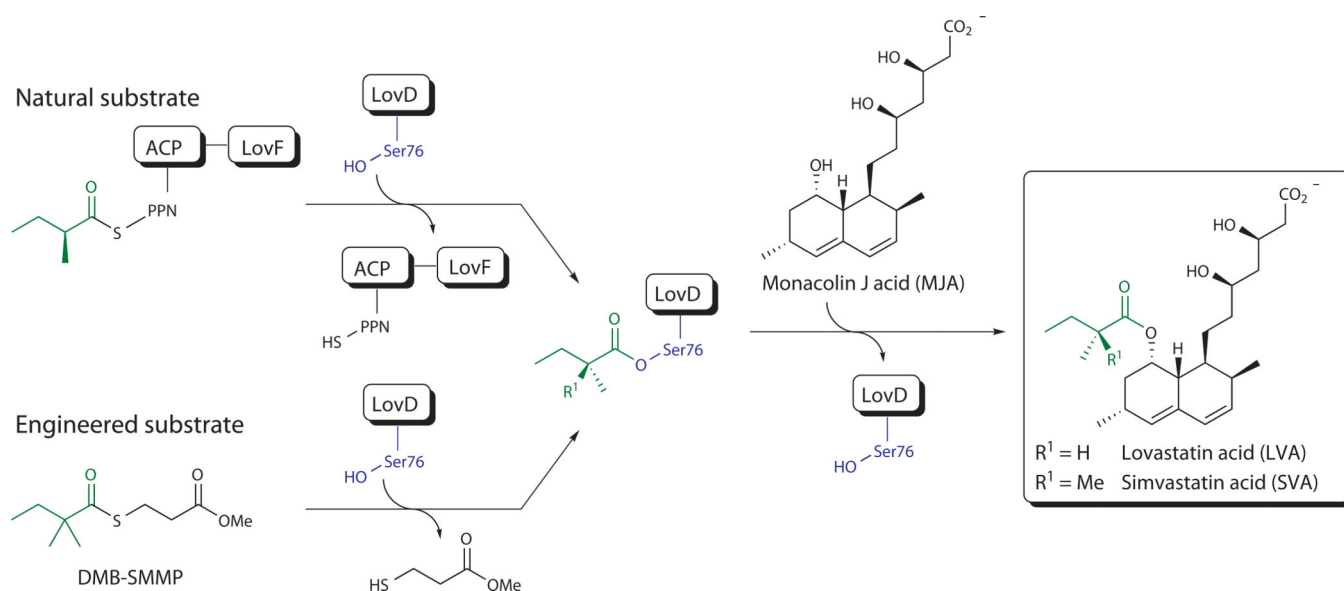


Figure 1. Chemical reactions catalyzed by LovD

The natural reaction in which LovD is acylated with the α -methylbutyrate group from the megasynthase LovF, and the engineered reaction in which LovD accepts 2,2-dimethylbutyrate, DMB, from small acyl carriers, are shown. The final step involves an acyl transfer reaction from acylated LovD to MJA to yield either LVA or semisynthetic SVA.

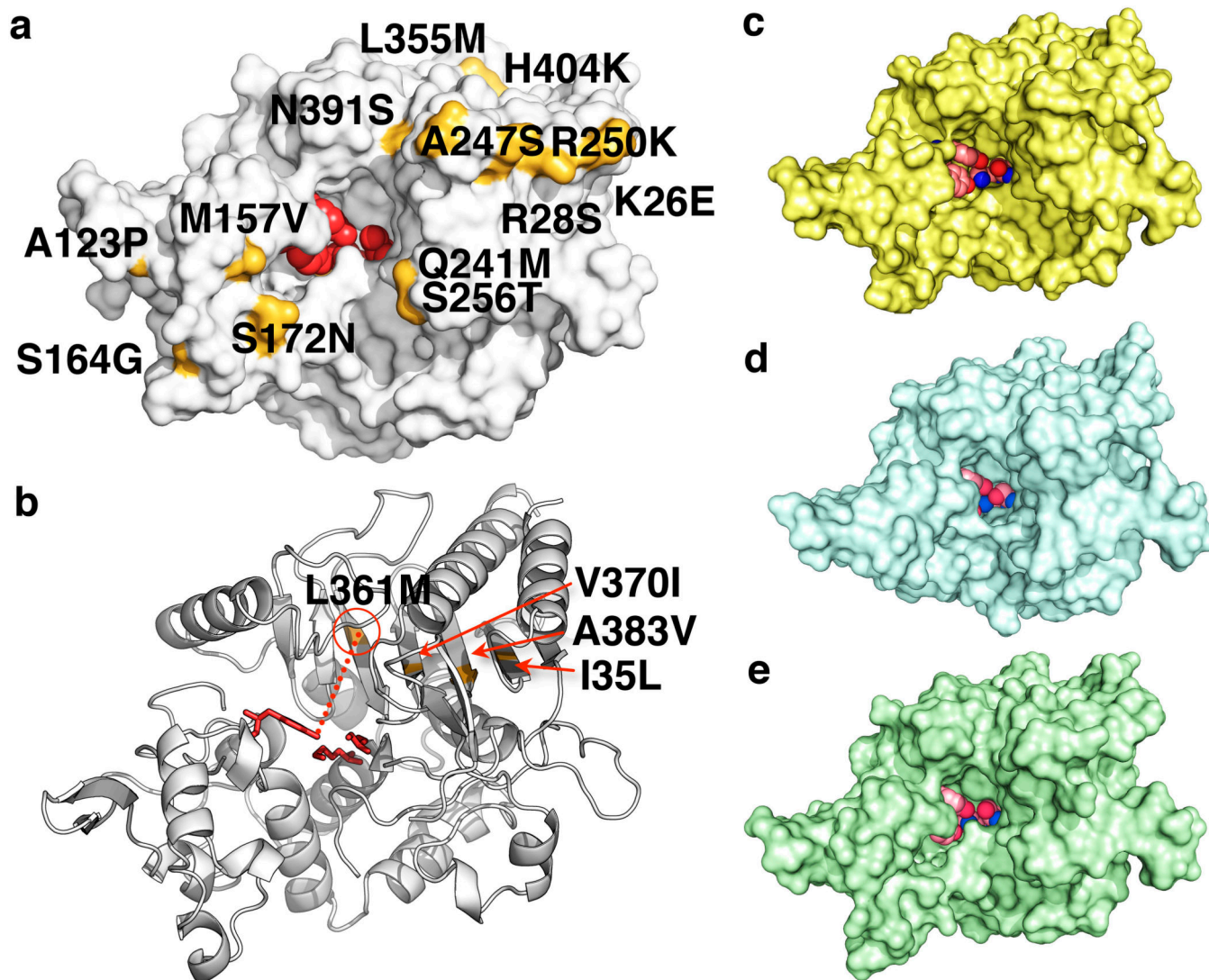


Figure 2. Location and structural effects of laboratory-evolved mutations in LovD

(a) Surface and (b) β -sheet mutations in LovD9. The catalytic triad (Ser76–Lys79–Tyr188) is shown in red. Mutations are shown in gold. One of the critical mutations, L361M, is 15 Å away (dashed line) from the catalytic triad (Ser76–Lys79–Tyr188). (c) Surface representation of wild-type LovD,⁷ (d) LovD6 and (e) LovD9 X-ray structures showing active site channel reduction. Active site residues are rendered in space-filling models.

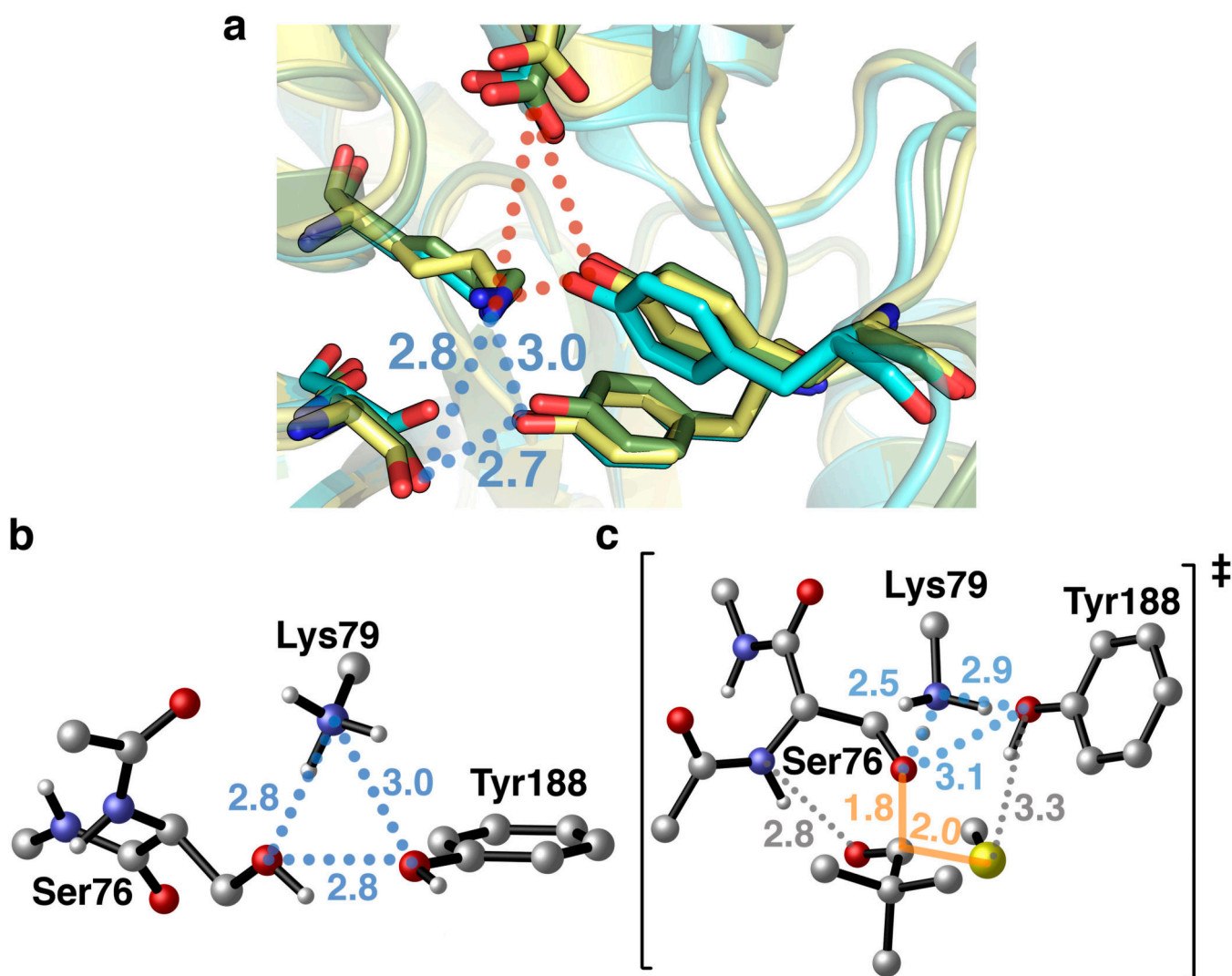


Figure 3. Structural features of the Ser76–Lys79–Tyr188 catalytic triad

(a) Overlays of active sites from X-ray structures in LovD (yellow), LovD6 (cyan) and LovD9 (green). (b) QM optimized model calculated for the catalytic triad resting state. (c) QM optimized TS model calculated for Ser76 acylation with methyl 2,2-dimethylpropanethioate. Aliphatic hydrogens have been omitted for clarity. Distances are in Angstroms.

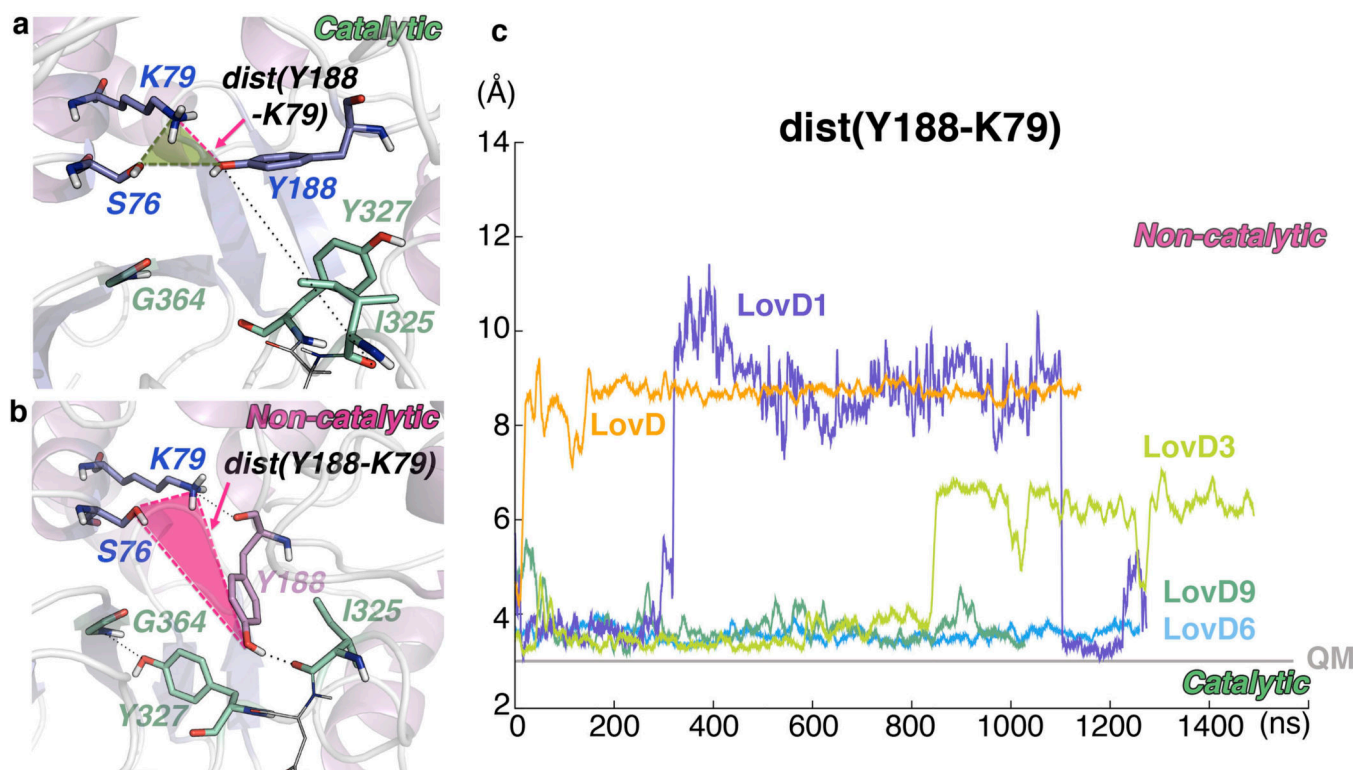


Figure 4. Active site dynamics of LovD variants determined through MD simulations
 Catalytically inactive (a) and Non-catalytic active (b) arrangements of the Ser76–Lys79–Tyr188 triad revealed by MD simulations on LovD1 (600 and 1200 ns, respectively). (c) The catalytic Tyr188–Lys79 distance along the MD trajectories. The optimal value for this distance determined through QM calculations has been used as reference and represented using a horizontal gray line. Catalytic and non-catalytic regimes have been depicted in green and magenta, respectively.

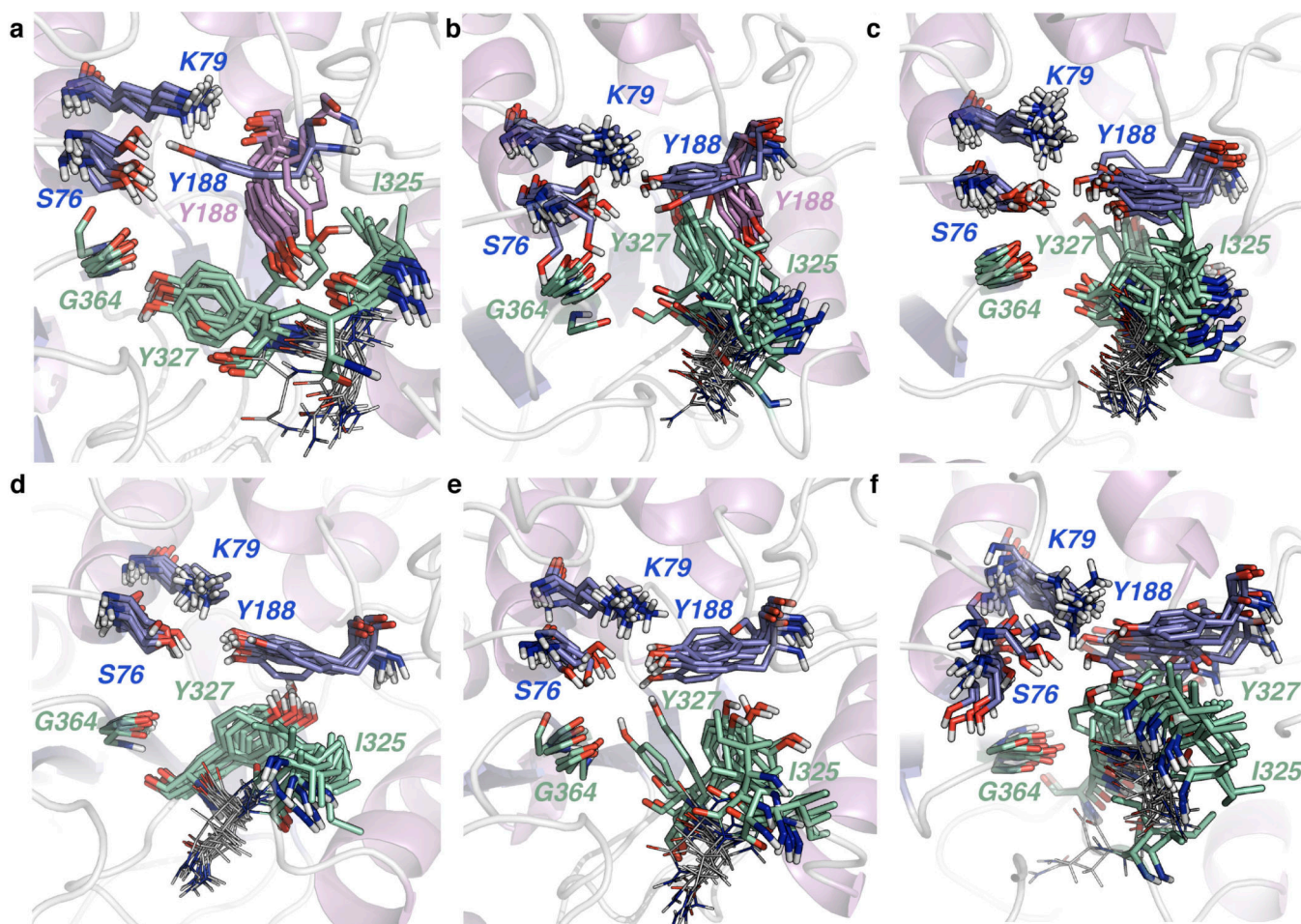


Figure 5. Conformational ensembles of the active site of LovD variants obtained through MD simulations

Overlay of 15 snapshots obtained from 1.2 μ s trajectories for: (a) wild-type LovD, (b) LovD1, (c) LovD3, (d) LovD6, (e) LovD9 and (f) LovD dimer. The catalytic triad (Ser76–Lys79–Tyr188) is represented in blue, together with other relevant non-catalytic residues (Gly364, Tyr327 and Ile325) in green. For clarity, Tyr188 has been represented in different colors depending on its catalytic (blue) or non-catalytic (purple) side chain conformation.

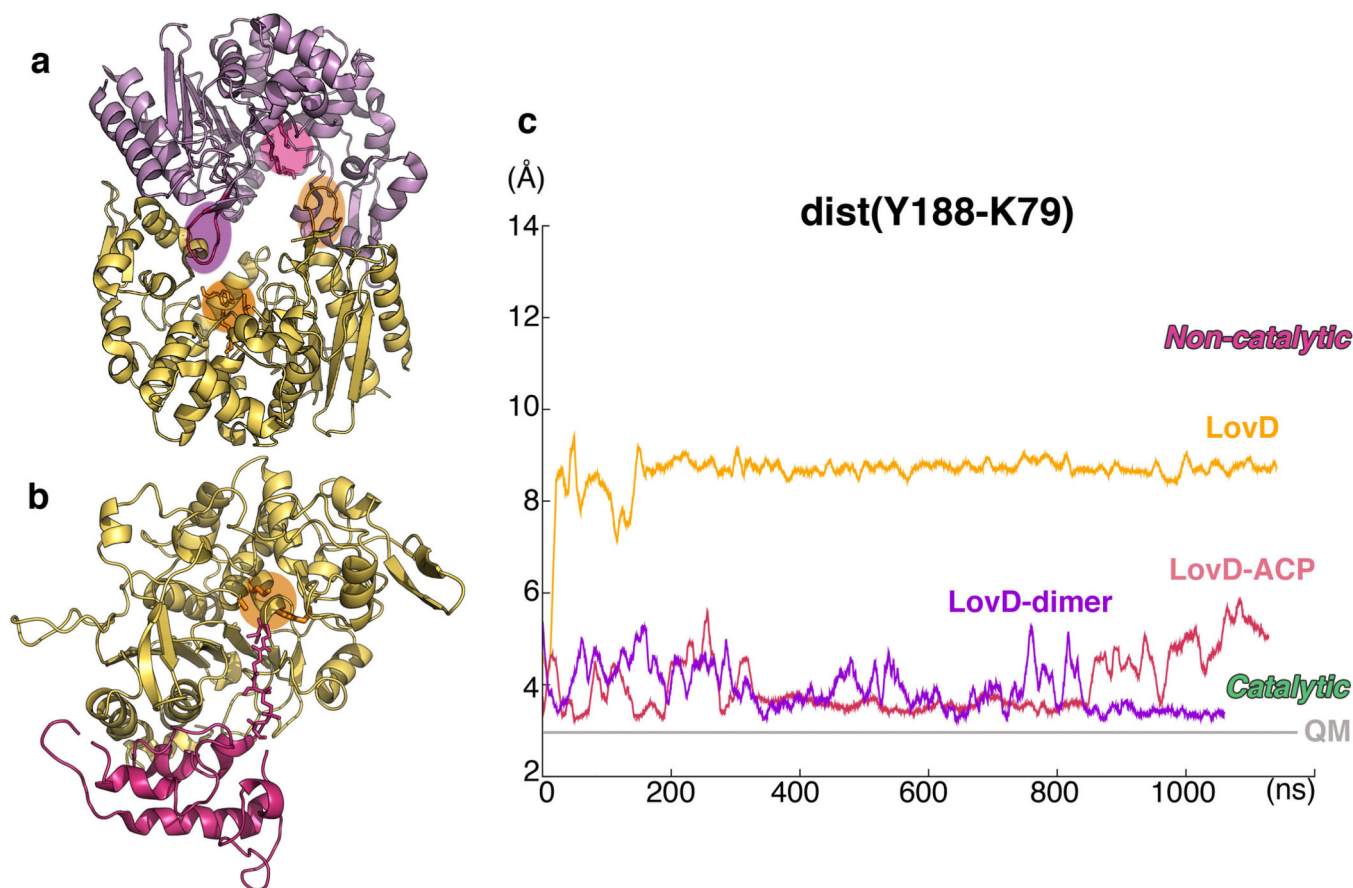


Figure 6. The role of protein-protein interactions on LovD active site dynamics

Models used in MD simulations: (a) Homodimer observed in the LovD X-ray structure, and (b) LovD-ACP complex constructed using homology modelling and protein-protein docking techniques. (c) The catalytic Tyr188–Lys79 distance along the MD trajectories. The optimal value for this distance determined through QM calculations has been used as reference and represented using a horizontal gray line. Catalytic and non-catalytic regimes have been depicted in green and magenta, respectively.

Table 1

Cumulative mutations produced through directed evolution of LovD and kinetic characterization of selected variants

Variant (mutations)	New mutations	k_{cat} (min ⁻¹) ^a	K_M (mM) ^a	k_{cat} / K_M (min ⁻¹ mM ⁻¹) ^a	Relative productivity ^b
LovD (0)	none	0.74±0.05	0.56±0.12	1.32±0.42	1
LovD1 (1)	L361M	0.95±0.02	0.46±0.04	2.07±0.50	2
LovD2 (6)	A123P; L174F; A178L; N191S; A247S	n.d.	n.d.	n.d.	12
LovD3 (7)	G275S	7.35±0.35	0.24±0.05	30.63±7.00	36
LovD4 (9)	M157V; S172N	n.d.	n.d.	n.d.	89
LovD5 (15)	A9V; K26E; L192I; R250K; Q297E; A383V	n.d.	n.d.	n.d.	220
LovD6 (20)	N43R; S164G; N191G; Q241M; V370I; H404K	42.02±2.33	0.14±0.04	300.14±58.25	595
LovD7 (23)	S256T; A261V; Q297G; N391S	n.d.	n.d.	n.d.	n.d.
LovD8 (25)	I4N; R28S; A261H	n.d.	n.d.	n.d.	n.d.
LovD9 (29)	I35L; D96R; S109C; L355M	120.46±6.79	0.28±0.06	430.21±113.17	1002

^a Measured using purified samples of the enzymes. K_M measured with respect to MJA. Data represent mean values ± s.d. Three replicates of each measurement were made.

^b Measured as relative turnover numbers in cell lysates. See Online Methods for further details. n.d.: not determined.