Pervasive cooperative mutational effects on multiple catalytic enzyme traits emerge via long-range conformational dynamics

Carlos G. Acevedo-Rocha1,11, Aitao Li2,11, Lorenzo D’Amore3,11, Sabrina Hoebenreich4, Joaquin Sanchis5, Paul Lubrano1,10, Matteo P. Ferla6, Marc García-Borràs3, Silvia Osuna3,7✉ & Manfred T. Reetz4,8,9✉

Multidimensional fitness landscapes provide insights into the molecular basis of laboratory and natural evolution. To date, such efforts usually focus on limited protein families and a single enzyme trait, with little concern about the relationship between protein epistasis and conformational dynamics. Here, we report a multiparametric fitness landscape for a cytochrome P450 monooxygenase that was engineered for the regio- and stereoselective hydroxylation of a steroid. We develop a computational program to automatically quantify non-additive effects among all possible mutational pathways, finding pervasive cooperative signs and magnitude epistasis on multiple catalytic traits. By using quantum mechanics and molecular dynamics simulations, we show that these effects are modulated by long-range interactions in loops, helices and β-strands that gate the substrate access channel allowing for optimal catalysis. Our work highlights the importance of conformational dynamics on epistasis in an enzyme involved in secondary metabolism and offers insights for engineering P450s.
Directed evolution constitutes a powerful tool for optimizing protein properties, including activity, substrate scope, selectivity, stability, allostery or binding affinity. By applying iterative rounds of gene mutagenesis, expression and screening (or selection), proteins have been engineered for developing more efficient industrial biocatalytic processes1-4. Directed evolution has also provided important insights into the relationship between protein sequence and function5-8, yet understanding the intricacies of non-additive epistatic effects remains a challenge2. Epistasis means that the phenotypic consequences of a mutation depend on the genetic background9-11. Epistatic effects can be negative (antagonistic/deleterious) or positive (synergistic/cooperative) if the respective predictive value is smaller or greater in sign/magnitude than the expected value under additivity. Sign epistasis (SE) occurs when a mutation has a deleterious or beneficial effect alone but an opposite effect when combined with other(s) mutation(s), whereas in magnitude epistasis (ME) a mutation has a deleterious or beneficial effect in isolation and in combination with other mutation(s). Based on studies of natural and laboratory protein evolution, negative9,10 or positive11 epistasis is more widespread than originally thought7. Importantly, positive epistasis increases the evolution of new protein functions because it allows access to mutational pathways that avoid deleterious downfalls. On the other hand, negative epistasis has been associated with a higher tolerance for mutations, which is important because this mutational robustness enables protein stability and evolution12. For fundamental and practical reasons, it is thus important to determine the existence, type and molecular basis of epistasis in protein evolution. Epistatic effects can arise between residues that are located closely or away from each other via long-range indirect interactions, both mechanisms involving sometimes direct or indirect substrate binding11. These global epistatic effects may be mediated by changes in the protein conformational dynamics.

Proteins have the inherent ability to adopt a variety of thermally accessible conformational states, which play a key role in protein evolvability and activity13,14. Along the catalytic cycle, enzymes can adopt multiple conformations important for substrate binding or product release15,16, and conformational change can be rate-limiting in some cases17,18. Much debated is the existence of a link between active site dynamics and the chemical step19,20. Some studies have suggested that mutations remote from the enzyme active site may directly impact the energetically accessible conformational states, thereby influencing catalysis21-23. This has been shown by means of crystal structures and nuclear magnetic resonance (NMR) spectra of mutants along evolutionary pathways21,24,25 together with computational assistance26-30. Molecular dynamics (MD) simulations, which are highly complementary to NMR analyses31, allow the partial reconstruction of the enzyme conformational landscape, and how this is altered by mutations introduced by laboratory evolution29,30,32. Tuning the enzyme conformational dynamics can play an important role in the emergence of novel activities22,25,30,32,33.

The connection between conformational dynamics and epistasis has been studied in proline isomerase (cyclophilin A)24, phosphotriesterase24 and β-lactamases34-37. For example, negative SE between two distal mutations limited dynamics of active site loops mediating substrate accessibility in a β-lactamase35. These studies provide fascinating insights, but they are limited to a single protein trait (usually activity) as a measure of fitness. This term originally refers to the reproductive success of organisms, but it can be applied to protein activity, selectivity or stability34-37. This contrasts with directed evolution where often two or more traits (e.g. activity and selectivity or stability) are sought for practical purposes4,38. Therefore, connecting epistasis to conformational dynamics increases our understanding of proteins. In turn, analysing non-additive epistatic effects can be expected to benefit in silico-directed evolution39.

In the present work, we used a combination of enzyme kinetics and computational approaches to investigate epistatic effects and conformational dynamics in the stepwise evolution of a cytochrome P450 monooxygenase (CYP) engineered for the highly active, regioselective and stereoselective oxidative hydroxylation of a steroid as a non-natural substrate40.

To determine epistatic effects effectively, we developed a Python-based script and freely accessible web-app (https://epistasis.mutanalyst.com/), which can be used for any enzyme and catalytic trait (or for any protein and parameter). Unexpectedly, we found pervasive positive epistatic effects on multiple catalytic traits, with selectivity and activity being generally characterized by SE and ME. We found that the analysis of the link between protein epistasis and conformational dynamics reveals the increasing optimization of activity and selectivity along all evolutionary trajectories through fine tuning of loops, helices and β-strands that gate active site entrance and modulate the active site by long-range networks of interactions. Our study offers guiding principles for the simultaneous engineering of both activity and selectivity in a model CYP member.

Results
Multiple parameters define the biocatalytic landscape in P450<sub>MO</sub>. The CYP super protein family has ~300,000 members that are involved, among others, in the biosynthesis of steroids, fatty acids and natural products as well as in the degradation of drugs in humans and of xenobiotics in the environment41,42. Thus this is an important enzyme class with relevant applications in biocatalysis, biomedicine, pharmacology, toxicology and biotechnology42-44. Previously, we achieved the stereoselective and regioselective hydroxylation of testosterone (1) by evolving the self-sufficient <i>Bacillus megaterium</i> cytochrome P450<sub>MO</sub> monooxygenase40. P450<sub>MO</sub> is one of the most active and versatile CYPs that oxidizes long fatty acids as the natural substrates, and there are various three-dimensional structures of the haem domain alone without the reductase domain45. However, wild type does not accept steroids, which is the reason why we chose mutant F87A as the starting enzyme. While F87A accepts 1, it provides in a whole-cell system only ~20% conversion with formation of a 1:1 mixture of 2β-hydroxytestosterone (2) and 15β-hydroxytestosterone (3)40. Combinatorial saturation mutagenesis at the randomization site R47/T49/Y51 allowed the evolution of mutant R47I/T49I/Y51I/F87A (III) displaying 94% 2β-selectivity and 67% conversion of 1 (1 mM) in 24 h whole-cell reactions40 (Fig. 1a). The mechanism is known to involve a radical process in which the catalytically active haem-Fe=O (Cpd I) abstracts an H atom from aliphatic C-H followed by a fast C-O bond formation, which requires a precise substrate positioning, as in other cases46 (Supplementary Fig. 1). The three mutated residues are located next to each other (distances of C<sub>α</sub> is ~6 Å) lining the large binding pocket but relatively far away (~15–20 Å) from haem-Fe=O, assuming the absence of dynamic effects (Fig. 1b). Complete deconvolution of variant III starting from parental F87A entails 3! = 6 theoretical pathways, which we constructed by generating the respective 6 intermediate mutants (Fig. 1c/d). The key question is how these residues determine selectivity and activity and whether they interact epistatically.

All intermediate mutants were generated, overexpressed in Escherichia coli BL21-Gold(DE3) and purified (Supplementary Fig. 2). Parent F87A (---) and variant III were also included, resulting in a total of 8 enzymes. Using defined substrate and NADPH concentrations, multiple parameters

---

ARTICLE NATURE COMMUNICATIONS | https://doi.org/10.1038/s41467-021-21833-w
were determined (Supplementary Note 1 and Supplementary Table 1).

The most 2β-selective variants (~67–91%) contain mutation Y51I (--I, I-I, -II and III), while the remaining ones proved to be 15β-selective, with substrate conversion being highest (35%) in mutants III and -II, and poor (~6–10%) in the remaining ones (Fig. 2a). NADPH leak rate without substrate is higher than NADPH consumption rate (NCR) in all mutants except -II and III, which display a respective <2- and <3-fold increased NCR (Fig. 2b). Mutants -II and III also showed a respective ~5- and ~10-fold improvement in product formation rates (PFRs) compared to the remaining variants (Fig. 2c), suggesting that variants -II and III have a good coupling efficiency (CE). CE describes how well the reductase domain delivers electrons from NADPH via the flavin cofactors to the substrate in the haem domain. A low CE value indicates futile NADPH usage, resulting in the formation of reactive species during the catalytic cycle (Supplementary Fig. 1) that can inactivate the enzyme47. Low CE values of 15–30% were found for all mutants, except for -II and III that display higher values of 37% (Fig. 2d). The total turnover number (TTN) is highest in mutants -II and III (Fig. 2e), whereas the total turnover frequency (TTF), PFR and NCR are highest in III (Fig. 2f).

Distal mutations enable conformational changes at the active site required for regioselectivity. To gain insights about the origin of selectivity and activity, we performed computational
Fig. 2 Multiple enzymatic parameters of deconvolution mutants. a Selectivity and conversion data are obtained from HPLC data and shown in percentage. b NADPH leak and consumption rate were measured in the absence and presence of testosterone (1) substrate, respectively. c Product formation rate (PFR) is calculated by multiplying the NADPH consumption rate by coupling efficiency. d Coupling efficiency (CE) is the ratio between NADPH consumption and production formation, and it is reported in percentage. e TTN describes the total moles of products per moles of enzyme after NADPH depletion. f TTF normalizes TTN by time after NADPH depletion. See Fig. 1d for mutant abbreviations. Other products mainly include the 15β-alcohol and other regioisomers. The data represent the average of two independent experiments (n = 2). Source data are provided with this paper.

Fig. 3 Conformational population analysis of key geometric parameters for hydroxylation. Distances determined between the oxygen atom of haem-Fe=O and the C-atom (C2 or C15) of 1 (x-axis) and angles formed by O(Fe = O) – (1)-H(C2/15) – (1)-C(2/15) (y-axis) from the first replica of the MD dataset of parent mutant (a), “single” mutants (b-d), “double” mutants (e-g) and “triple” mutant III (h) (see Supplementary Figs. 6 and 7 for additional replicas). Geometric parameters measured for C-2 and C-15 are shown in red and blue, respectively. The ideal distance and angle for the transition state TS (black dot) corresponds to the Density Functional Theory (DFT) optimized geometry for the C–H abstraction by haem-Fe=O using a truncated computational model (Supplementary Note 2). See Fig. 1d for mutant abbreviations.
studies on all mutants. Given the identification of comparable reaction barriers for hydrogen atom abstraction from C2 and C15 by using Density Functional Theory (DFT) calculations on truncated models (difference of <1.0 kcal/mol, see Supplementary Note 2, Supplementary Fig. 3 and Supplementary Table 2), we carried out MD simulations. For each mutant, we started from pose 15 and from pose 2 (i.e. positioning C15 or C2 closer to the Fe=O, respectively) to analyse whether the binding pose of I in the active site determines the experimentally observed selectivity (Fig. 3 and Supplementary Note 3). Starting from parent ---, pose 2 (presenting C2 close to the catalytic Cpd I) and pose 15 (C15 close to Cpd I) generated from manual dockings are possible (Supplementary Figs. 4 and 5). Further analysis of these binding poses along MD simulations in --- indicate that substrate I in pose 15 explores near attack conformations (NACs)48 closer to the quantum mechanics-predicted ideal transition state geometry for H abstraction than in pose 2 (Fig. 3a), thus making pose 15 more productive towards 15β- hydroxylation. Introducing mutations R47I and/or T49I does not have any effect on selectivity (Fig. 3b, c, e), i.e. the selectivity is retained due to the catalytically competent conformation inherent in pose 15 along MD simulations (pose 2 adopts a reduced number of catalytically competent conformations). However, the picture completely changes when mutation Y51I is induced: the substrate bound in pose 15 becomes unstable and leaves the active site in 1 out of 3 replicas (ca. >15 Å C2-O distances explored, Fig. 3d), whereas pose 2 is highly stabilized and explores short C2-O distances for the incipient C-H eventually leading to 2β-hydroxytestosterone in 2 out of 3 replicas (Fig. 3d and Supplementary Figs. 6 and 7). As experimentally determined, 2β-selectivity is retained in variants I-I, -II and III that contain mutation Y51I (Fig. 3f-h). This is even more dramatic in variant III, in which pose 15 is highly unstable and I rapidly rotates to position C2 close to the catalytic Cpd I for 2β-hydroxylation (Supplementary Movie 1). Instead, pose 2 in variant III is stable and adopts near attack conformations in all MD replicas (Fig. 3d and Supplementary Figs. 6 and 7).

Notwithstanding, mutant -II and III only differ for the R47I mutation in the latter case, yet the re-orientation of the substrate from pose 15 to pose 2 is observed only during the MD simulation of mutant III. To further investigate the specific effect of R47I mutation on substrate rotation inside the haem pocket, we performed a Principal Component Analysis (PCA) on the substrate-bound MD trajectories of mutant -II and III, finding that pc2 indeed describes an increased flexibility of residues A87, T260, G265 and T327 in mutant III, as compared to -II (Supplementary Fig. 8). Thus R47I may modulate via long-range conformational dynamic effect the flexibility of such residues, which have been shown to be instrumental to promote substrate re-orientation in mutant III (Supplementary Movie 2). Moreover, mutant III presents a substantially wider active site pocket as compared to the other variants: the active site volume in the --- variant is 89 Å3, which is expanded to 235 Å3 in III (Supplementary Fig. 9). We hypothesized that, in all variants, except III, selectivity must be determined by the orientation adopted by the substrate while entering the haem pocket. In the productive trajectory corresponding to mutant I-I, I accesses the haem with the correct orientation for 15β-hydroxylation (Fig. 4b and Supplementary Fig. 10). In this case, residue Y51 establishes a hydrogen bond with the carbonyl group of 1 (Supplementary Fig. 11), constraining the substrate in such way that it can only progress into the active site pocket pointing its C15 ahead towards haem-Fe=O50. Thus Y51 is instrumental in promoting the observed C15-selectivity in --- and in I-I, -I- and II- variants. It should be noted, that in previous studies, R47 and especially Y51 were found to interact with the terminus end of long-chain fatty acids while bound at the P450BM3 active site51. Such direct interaction with testosterone and Y51 is only possible at the pre-binding pocket, which is lost after the retreat of β4 sheet, allowing substrate access to the haem pocket. Additionally, the higher C2-selectivity observed in variant III occurs due to the flipping and motion of the β4 sheet, destabilizing pose 15 while favouring pose 2 (Fig. 4c).

Interestingly, the analysis of the most relevant conformational changes in each independent variant through PCA predicts that the most active mutant III shows the highest flexibility of the β4 sheet (Fig. 4d). This higher flexibility related to activity has, however, no impact on the B' helix and B'-loop conformational dynamics (Supplementary Figs. 12 and 13). These flexible regions, responsible of controlling substrate binding as described above, are likely to influence activity, as mutant III shows the highest TTF, NCR and PFR numbers. These findings suggest that favouring a more efficient substrate binding in a catalytically competent pose increases enzyme TTF, while NADPH leak is reduced due to a more efficient interaction between the substrate and the catalytically active Fe=O species once generated.

Pervasive epistatic effects on multiple parameters are cooperative. According to Tokuriki11 and Bendixsen et al.32, non-additive mutational effects can occur in different forms (Fig. 5), which can be calculated with additivity equations (Supplementary Note 5). Aiming at exploring the existence of epistatic effects in an effective manner, we developed and applied a Python-based computational program to automatically determine the type and intensity of amino acid interactions among all possible mutational combinations for the three mutations introduced (Supplementary Note 6).

We quantified all amino acid interactions among all 6 trajectories leading from parent --- to mutant III for multiple parameters
focused on the evolution towards 2β-hydroxytestosterone (Table 1).

All combinations on substrate conversion show synergistic effects, with 6 (86%) and 1 (14%) cases of SE and ME, respectively (Supplementary Table 3). For 2β-selectivity, all interactions are likewise synergistic, with most of them showing positive SE and only one case of positive ME (combination of R47I and T49I). For example, the combination of the single mutations R47I, T49I and Y51I (in parent mutant --) is expected to contribute $-3.45 \pm 0.25$ kJ/mol. The two former mutations confer 15β-selectivity in mutants I-- and I-I, while the latter one induces 2β-selectivity in variant --I.

Yet the experimental value of mutant III yields 5.6 ± 0.0 kJ/mol, which represents a difference of about 9 kJ/mol between the experimental and theoretical values (Supplementary Table 4). Whereas NCR has 6 cases of positive epistatic effects (86%) and one negative case (14%), PFR likewise shows 6 cases of positive effects but one additive case (Supplementary Tables 5 and 6). Similarly, CE shows 6 cases of synergistic epistatic effects and 1 antagonistic case (Supplementary Table 7). Overall, these results indicate that an efficient consumption of NADPH and oxidation of testosterone towards formation of 2β-hydroxytestosterone requires pervasive cooperative effects among R47, T49 and R51 regardless of mutational combination.

Conformational dynamics shape the evolution of the fitness landscape. The complete deconvolution of a multi-mutational variant enables the exploration of all possible pathways from parental enzyme to the evolved mutant, thus determining a full multidimensional fitness landscape. Such landscapes provide insights on the different routes that evolution can take. Additionally, engineering proteins by single mutational steps is a highly successful strategy in directed evolution. To explore the step-wise accessibility in the evolution of parental towards III, we constructed a fitness “pathway” landscape based on both activity and selectivity (Fig. 6a). This system is a 4-dimensional surface (3 sets of mutations as independent vectors and ΔΔG‡ as the dependent variable obtained from the experimental selectivities). Two kinds of trajectories can be noted: those lacking local minima (favoured) and those characterized by at least one local minimum (disfavoured). Pathways 1–4 are characterized by a decrease in both selectivity and activity at the first step, indicating that they are evolutionarily disfavoured (pathway 3 is highlighted in red). Pathway 5 (highlighted in green) and 6...
are favoured because –I enables conformational changes in the active site and has implications on the substrate binding (as discussed above). In the two latter pathways, activity improves slightly in the evolution of -I towards -II (TTF = 11 → 21) at the first step, but at the second and third steps of pathway 6 it increases significantly towards -I and III (TTF = 21 → 72 → 158). This is due to the β4 sheet that shows an increased flexibility in the most active mutants -II and III, highlighting the key role of β4 sheet for activity (Fig. 6b). Interestingly, when all other parameters are considered, pathways 5 and 6 are the only ones that remain accessible (Supplementary Note 7 and Supplementary Fig. 14), with selectivity and CE showing the strongest non-additive effects, thus indicating the importance of residue Y51 towards efficient 2β-hydroxytestosterone formation.

To identify the most important conformational changes in all evolutionary pathways and to describe how distal mutations influence them, we performed extensive MD simulations in the absence of substrate of each variant and applied the dimensionality reduction technique PCA to the whole data set (Figs. 6c and 7a). A conformational population analysis resulting from all the accumulated simulation data was generated in terms of principal components (PCs) 1 and 3, which describe the first and third most important conformational differences among all variants (for PC2, see Supplementary Fig. 16). Notably, a clear distinction between 2β- and 15β-selective mutants is revealed through their separation with respect to PC1 (x-axis), suggesting that changes in selectivity are linked to the impact that the introduced mutations have on the enzyme conformational dynamics (Fig. 6c). These conformational changes related to selectivity mainly involve the G helix, the F–G loop, the β1 hairpin and the B’ helix (located at the entrance of the 2a/b channels) as well as

---

**Table 1 Epistatic analysis of all possible mutational combinations on multiple parameters towards formation of product 2β-hydroxytestosterone.**

| Typea | Combination of mutationsb | Resulting mutantb | Parameterc,d,e |
|-------|---------------------------|-------------------|---------------|
|       |                           |                   | Conv. | Sel. | NCR | PFR | CE | TTN | TTF |
| B     | -I + +l                   | II                | +SE 0.4 | +ME 2.8 | -SE -13 | +SE 1 | -RSE 3 | +SE 9 | +SE 2 |
|       | -I + +l                   | II                | +SE 25  | +SE 5.6 | +ME 91  | +ME 39 | +ME 16 | +ME 490 | +ME 51 |
|       | +I + -I                   | I                 | +SE 2   | +SE 2.6 | +SE 23  | ADD -0.2 | -SE -7 | +SE 38 | +SE 7 |
|       | +II + +I                  | III               | +SE 3   | +SE 3.5 | +SE 192 | +SE 70 | +SE 2  | +SE 58 | +SE 90 |
|       | +I + +I + (+I)            | III               | +SE 27  | +SE 6.3 | +SE 284 | +ME 108 | +ME 15 | +ME 539 | +ME 139 |
|       | +II + +I                  | III               | +ME 25  | +SE 6.4 | +ME 260 | +ME 109 | +ME 24 | +ME 509 | +ME 133 |
|       | +I + +I + (+I)            | III               | +SE 27  | +SE 9.0 | +SE 283 | +SE 109 | +SE 18 | +SE 548 | +SE 140 |

This shortened data set originates from Supplementary Tables 3–9.

aBinary (B) and tertiary (T) combinations.

bSee Fig. 1c for nomenclature of mutations and mutants.

cThe parameters (units) are: Conversion (Conv.), selectivity (Sel.), NADPH consumption rate (NCR), product formation rate (PFR), coupling efficiency (CE), total turnover number (TTN), total turnover frequency (TTF). The units can be found in Supplementary Tables 3–9.

dThe types of epistatic effects are: Sign Epistasis (SE), Magnitude Epistasis (ME), Reciprocal Sign Epistasis (RSE), which can be positive (+) or negative (−). Additivity (ADD) means absence of epistatic effects.

The data represent the average of two independent experiments (n = 2). The standard error mean can be found in Supplementary Tables 3–9.
the α-A loop and the β4 sheet (located at the entrance of the 2f channel) (Fig. 7b). In variant -I-, the channel 2a has a narrower substrate access entrance due to a closed state of the F–G loop (ca. 9.3 Å determined between the Cα of R47 and N192). Conversely, the combination of mutations introduced in III favours an open conformational state of the same F–G loop (ca. 12.4 Å measured between the Cα of I47 and N192), enlarging the access channel 2a, which is mainly responsible for allowing access to the enzyme-binding pocket (Fig. 7c). Indeed, the area surrounding the access channel 2a in mutant III is calculated to have a volume of 140 Å³ with respect to 44 Å³ in -I-. 
To further study the link between epistasis and conformational dynamics, we applied the shortest path map (SPM) analysis\textsuperscript{22} (Supplementary Note 9) using the accumulated 1.8 µs MD simulation performed on parent --- in the absence of substrate. SPM considers the different conformations that the enzyme samples along the MD simulation and identifies which residues are those that are more important for the observed conformational changes, which in this case are associated with different selectivities and activities\textsuperscript{22}. In the parent --- enzyme, the generated SPM identifies residues Y51 as well as V78 and A330, known from earlier studies\textsuperscript{40}, to be important for enzyme activity and to be interconnected in terms of Ca correlated movements, thus highly contributing to the enzyme inactive-to-active conformational interconversion. This highlights why these
three distal positions are found to be key during the evolutionary pathway for improving catalysis, in line with what we observed for the laboratory-evolved retro-aldolases. Importantly, the SPM also describes strong connections between all the five-stranded β1 sheet with the β4-2 strand and the B’ helix, which we showed to be crucial for substrate binding and gating (Fig. 7d). This long-distance communicating pathway between β1 and β4 sheets directly relates the mutated positions on β1-2 strand (positions R47, T49 and Y51) and the increased flexibility of the β4 sheet. This shows how evolutionary pathways take advantage of networks of residue–residue interactions to fine tune the conformational dynamics along the evolutionary pathways for improving enzyme function.

Discussion
The identification of epistasis and its molecular mechanism are crucial for understanding protein function, but these are hardly ever explored in laboratory evolution studies of multiparametric optimization. For example, the directed evolution of activity and selectivity in plant sesquiterpene synthases or in P450BM3 (ref. 60) did not consider epistatic effects, while “stability-mediated epistatic effects” were observed in a P450BM3 study some years later. Conversely, various studies in evolutionary biology have determined the contribution of multiple parameters on epistasis and organismal fitness. Shakhnovich et al., for instance, found that bacterial growth depends on the activity and expression of adenylate kinase or on the activity, binding and folding stability of dihydrofolate reductase.

The construction of fitness landscapes using a single catalytic parameter has been reported in two main research areas. While such landscapes have revealed that usually many pathways are accessible in laboratory evolution of enzymes as catalysts in organic chemistry, different conclusions have been made in evolutionary biology. In the present study, we observed that only a few trajectories (2/6) are accessible to both selectivity and activity. Interestingly, the two accessible pathways for selectivity correspond to mutation Y51I. The addition of mutation R47I has almost no effect on activity and selectivity, while mutation T49I, which is closer to Y51I, significantly improves both parameters as it alters the enzyme conformational dynamics. T49I and Y51I enhance the flexibility of the β4 sheet, and both combined with R47I reshape the active site for enhanced 2β-hydroxylation. The triple mutant III excels in all parameters compared to all double and single mutants. Unexpectedly, upon going from the “parent” enzyme to mutant III, cooperative interactions at each step in the evolution of selectivity and activity (i.e. TTF) remain pervasive, with SE and ME characterizing those effects. Residues R47I, T49I and Y51I are located at the entrance of a long substrate channel far away from the active site. Since the mutated residues were not observed to interact directly with the substrate in our MD simulations, we propose that the observed epistatic effects, which are mediated by long-range interactions, can occur via one main mechanism: direct effects between mutations but no direct interaction between the substrate and the mutations.

Our computational exploration of the mutation-induced conformational changes on F87A variants provide key insights concerning the importance for P450BM3 evolution towards more active and selective variants. These simulations predict that activity is dictated by the flexibility of the β4 sheet, which acts as a gate and modulates substrate access to the catalytic haem pocket for efficient hydroxylation. Our simulations also highlight the key role of the F–G loop in open–close conformational transitions involved in substrate binding, as shown for P450BM3 by Shaik and for P450p450 by Houk and Sherman. The substrate-binding simulations show that selectivity is dictated by how the substrate is oriented when accessing the haem pocket through the β4 sheet. By tuning the open/close conformational states of the F–G loop and the β1 hairpin, the substrate access channels are altered, which impact substrate orientation and thus selectivity. This rich conformational heterogeneity observed for P450BM3, which is important for substrate binding, is in line with previous reports and also with the selective stabilization of discrete conformational states of P450cip19 and P450p450 upon ligand binding. However, our simulations contrast to what was previously observed in P450cam, which does not depend on open/closed conformational changes of the F, G helices and loop for allowing substrate binding. It should also be mentioned that P450cam complexed with its redox partner adopts an open conformation that stabilizes the active site key for the proton relay network.

Using SPM analysis, the most important positions that participate in the open/closed conformational conversions that dictate selectivity and activity were predicted. Of relevance is that the key residue Y51I found to be essential for both activity and selectivity in this study is not connected in the SPM path, as well as the previously described V78 and A330 positions. SPM also highlights a long-distance communicating pathway between β4 and β1 where positions R47, T49 and Y51 are located, which is exploited along the evolutionary pathway for altering BM3 protein function.

This study provides evidence that in P450BM3 epistasis is intrinsically linked to conformational dynamics, which fine-tunes multiple functions in a protein involved in secondary metabolism. Our findings on the conformational changes connected to CYP activity and selectivity and residue networks that modulate such conformational conversions can be expected to facilitate future rational evolution of these enzymes for diverse practical applications.

Methods
Chemicals, materials and software. All commercial chemicals were purchased with the highest purity grade (e.g., high-performance liquid chromatography (HPLC)) from Sigma-Aldrich (St. Louis, US) unless otherwise indicated. For protein purification, lysozyme and DNAse I was purchased from Applichem (Darmstadt, Germany). For PCRs, KOD Hot-Start DNA Polymerase was obtained from Novagen (Merck, Darmstadt, Germany). Restriction enzyme DpnI was bought from New England Biolabs (Ipswich, US). The E. coli BL21-Gold(DE3) strain, obtained from Novagen (Merck-Millipore) and generally cultured in lysogeny broth (LB) with 30 µg/mL kanamycin (Kan456) as marker (Kan456), both obtained from Carl Roth, was used for transformation of site-directed mutagenesis reactions as well as for protein overexpression experiments. According to standard molecular biology protocols, electro-competent E. coli cells were prepared using 10% glycerol (Applichem) and transformed with the corresponding plasmids using a “MicroPulser” electroporator (BioRad, Hercules, US) following the manufacturer’s instructions. Oligonucleotides were purchased from Metabion (Martinsried, Germany). The analysis of sequencing reads was performed using the commercial software MegaAlign from DNASTAR Lasergene version 11 (Madison, US) and the freeware ApE plasmid editor version 2.0.44 by Wayne Davis. The software used for constructing the fitness pathway landscapes is Surfer version 8 (Golden, US) and the graphs and dot plots were done with GraphPad Prism version 9 (La Jolla, US).

Site-directed mutagenesis. The mutants were created using the MegaPrimer method as reported in ref. 49. Briefly, the P450BM3 mutant F87A gene, already cloned in the pETM11-BM3 plasmid, was amplified by PCR by using <25 ng template with 2.5 µM of both silent and mutagenic oligos depending on mutant (Supplementary Table 12) in 50 µL of 1× KOD hot start buffer, 2 mM dNTPs (each), 25 mM MgSO4, and 0.5 units of KOD hot start polymerase. The PCR programme started with 1 cycle of 95 °C for 3 min, 5 cycles of 95 °C for 30 s, 62 °C for 1 min, 72 °C for 6 min, 20 cycles of 95 °C for 3 min, 5 cycles of 95 °C for 30 s, 62 °C for 1 min, 72 °C for 6 min, 20 cycles of 95 °C for 3 min, 68 °C for 8.5 min, 1 cycle of 68 °C for 10 min and cooling. The samples were treated with 1 µL DpnI and incubated at 37 °C overnight to remove the parent plasmid. For each mutant, 5 colonies were incubated in 4 mL LB and the plasmids were extracted using the commercial DNA sequencing was conducted with the four respective oligos listed in Supplementary Table 12 by service provider GATC (now Eurofins, Constance, Germany).
control unit and UV-Vis-NIR Spectra Manager software II. All assays were performed in 100 mM sodium phosphate buffer (pH 8.0) at 25 °C using quartz cuvette samples (900 rpm). NADPH consumption was determined by measuring NADPH depletion monitored at 340 nm (ε = 6.22 mM⁻¹ cm⁻¹). A concentration of 0.24 mM NADPH was used in the reaction mixture. Due to uncoupling reactions, where NADPH is consumed without substrate hydroxylation, the rates were calculated by subtracting the rate of NADPH consumption in the absence of substrate. Reactions containing 0.2 mM testosterone dissolved in DMF with a final solvent concentration of 1% (v/v) were started with addition of 100 nM P4502A3 enzyme in a final volume of 1 mL and these were monitored until NADPH was completely depleted or a measured change in absorbance at 340 nm (completion of the reaction). Afterwards, the reaction mixture was immediately transferred into 96-well plates and frozen at −20 °C. Reaction mixtures of 600 µL were taken and mixed with ethyl acetate (2 x 150 µL) with the LiHa of the Tecan robot platform (dispensing speed, 600 µL/s). The organic phase was extracted using the Te-MO multi-pipette option and transferred to 500 µL MTPs (Nunc, Roskilde, Denmark). The solvent was dried overnight, and the next day the steroid was resuspended in 150 µL acetonitrile and passed through a PTFS 96-well plate filter to remove solid particles (Pall, VWR, Germany) into a new 500 µL MTP (Nunc). The MTPs, which were closed using silicon lids for the corresponding plates, were stored at 4 °C prior to screening.

Steroid hydroxylation screening by HPLC. A LC-2010 HPLC system (Shimadzu, Japan) equipped with four MTPs racks was used employing a reverse-phase “250 Eclipse XDB C18 column of 250 mm (1.8 µm size particle) together with a corresponding pre-column bought from Agilend (Waldbrenn, Germany) as a stationary phase and installed in the oven at 40 °C. The mobile phase was composed of a mixture of high-purity water generated from the local deionized water supply using a TKA MicroLab water purification system, acetonitrile (CH3CN) and methanol (MeOH). For testosterone (T), a programme of 8 min based on a CH3CN:MeOH:H2O mixture was used: 0 → 3 min (15:15:70), 3 → 5 min (20:20:60), 5 → 6 min (30:30:40), 6 → 7 min (15:15:70). This protocol allows the separation of >14 oxidation products of T. The retention times of the known and unknown compounds can be found elsewhere. Data acquisition was done using the Shimadzu LCSolution software version 3, while data analysis was performed with Microsoft Excel 365 MSO version 2012 (16:0.13.35.20054) 32-bit.

Large-scale protein expression and purification. The P4502A3 mutants were inoculated into 4 mL LBKas® broth and cultured overnight in the orbital shaker with tray with adhesive mating for shake flasks (Multitrion) at 37 °C and 220 rpm. The overnight culture (4 mL) was transferred into 200 mL TBKas® in 500 mL shaking flasks. The cultivation continued at 37 °C and 220 rpm for 2–3 h until the OD660 reached 0.6–0.8, then IPTG was added to a final concentration of 100 µM and the temperature was reduced to 25 °C. After 20 h expression, the cells were harvested by centrifugation at 1100 x g and 4 °C for 15 min. The cell pellets were stored at −80 °C until further processing. The cell pellets were dissoluted in buffer (50 mM KPi, 800 mM NaCl, pH 7.5) and disrupted by sonication under an ice bath. The disrupted by sonication for 15 min at 30,000 g for 45 min. The obtained brownish-red supernatant was filtered to sterility with a 0.45-µm filter to remove solid particles (Pall, VWR, Germany). The supernatant was added to the next day the steroid was resuspended in 150 µL acetonitrile and passed through a PTFS 96-well plate filter to remove solid particles (Pall, VWR, Germany) into a new 500 µL MTP (Nunc). The MTPs were stored at 4 °C prior to screening. The kinetic parameters are shown in Supplementary Table 1.

Received: 25 June 2020; Accepted: 29 January 2021; Published online: 12 March 2021

References

1. Zeymer, C. & Hilvert, D. Directed evolution of protein catalysts. Annu. Rev. Biochem. 87, 131–157 (2018).
2. Arnold, F. H. Innovation by evolution: bringing new chemistry to life (Nobel Lecture). Angew. Chem. Int. Ed. 58, 14420–14426 (2019).
3. Wang, Y., Yu, X. & Zhao, H. Biosystems design by directed evolution. AIChE J. 66, e16716 (2020).
4. Qu, G., Li, A., Sun, Z., Acevedo-Rocha, C. G. & Reetz, M. T. The crucial role of methodology development in directed evolution of selective enzymes. Angew. Chem. Int. Ed. 59, 13204–13231 (2020).
5. Romero, P. A. & Arnold, F. H. Exploring protein fitness landscapes by directed evolution. Nat. Rev. Mol. Cell Biol. 10, 866–876 (2009).
6. Currin, A. et al. Synthetic biology for the directed evolution of protein biocatalysts: navigating sequence space intelligently. Chem. Soc. Rev. 44, 1172–1239 (2015).
7. Reetz, M. T. The importance of additive and non-additive mutational effects in protein engineering. Angew. Chem. Int. Ed. 52, 2668–2666 (2013).
8. De Visser, J. A. G. M. & Krug, J. Empirical fitness landscapes and the predictability of evolution. Nat. Rev. Genet. 15, 480–490 (2014).
9. Hartl, D. L. What can we learn from fitness landscapes? Curr. Opin. Microbiol. 21, 51–57 (2014).
10. Starr, T. N. & Thornton, J. W. Epistasis in protein evolution. Protein Sci. 25, 1204–1218 (2016).
11. Miton, C. M. & Tokuriki, N. How mutational epistasis impacts predictability in protein evolution and design. Protein Sci. 25, 1260–1272 (2016).
12. Berestehin, S., Segal, M., Bekerman, R., Tokuriki, N. & Tawfik, D. S. Robustness–epistasis link shapes the fitness landscape of a randomly drifting protein. Nature 444, 929–932 (2006).
13. Tokuriki, N. & Tawfik, D. S. Protein dynamism and evolvability. Science 324, 203–207 (2009).
14. Pabis, A., Risso, V. A., Sanchez-Ruiz, J. M. & Kamerlin, S. C. Cooperativity and flexibility in enzyme evolution. Curr. Opin. Struct. Biol. 48, 83–92 (2018).
15. Boehr, D. D., Nussinov, R. & Wright, P. E. The role of dynamic conformational ensembles in biomolecular recognition. Nat. Chem. Biol. 5, 789–796 (2009).
16. Guengerich, F. P., Wilkey, C. J., Glass, S. M. & Reddick, M. J. Conformational selection dominates binding of steroids to human cytochrome P450 17A1. J. Biol. Chem. 274, 10028–10041 (2001).
17. Jackson, C. J. et al. Conformational sampling, catalysis, and evolution of the bacterial phosphotriesterase. Proc. Natl Acad. Sci. USA 106, 21631–21636 (2009).
18. Henzler-Wildman, K. A. et al. Intrinsic motions along an enzymatic reaction coordinate. Proc. Natl Acad. Sci. USA 100, 2246–2251 (2003).
19. Campbell, E. C. et al. Laboratory evolution of protein conformational dynamics. Nat. Rev. Mol. Cell Biol. 18, 397–407 (2017).
20. Benkovic, S. J. & Hammes-Schiffer, S. A perspective on enzyme catalysis. J. Am. Chem. Soc. 135, 8001–8015 (2013).
21. Lim, J. B. et al. Insights into electron leakage in the reaction cycle of cytochrome P450 BM3 revealed by kinetic modeling and mutagenesis. Protein Sci. 24, 1874–1883 (2015).
22. Wijma, H. J. et al. Enantioselective enzymes by computational design and in silico screening. Angew. Chem. Int. Ed. 54, 3726–3730 (2015).
23. Ahdawat, N. & Mondal, J. Mapping the substrate recognition pathway in cytochrome P450 2C9. J. Am. Chem. Soc. 140, 17743–17752 (2018).
24. Ruettiger, R. T. & Fulco, A. J. Epoxidation of unsaturated fatty acids by a soluble cytochrome P450-dependent system from Bacillus megaterium. J. Biol. Chem. 256, 5728–5734 (1981).
25. Li, H. & Poulos, T. L. The structure of the cytochrome p450BM-3 haem domain complexed with the fatty acid substrate, palmitoleic acid. Nat. Struct. Biol. 4, 140–146 (1997).
26. Bendixen, D. P., Østman, B. & Hayden, E. J. Negative epistasis in experimental RNA fitness landscapes. J. Mol. Evol. 85, 159–168 (2017).
27. Tracewell, C. A. & Arnold, F. H. Directed enzyme evolution: climbing fitness landscapes with a ratchet. Curr. Opin. Chem. Biol. 13, 3–9 (2009).
28. Bougioukos, J. D., Kille, S., Taglieber, A. & Reetz, M. Directed evolution of an enantioselective enolate-reductase: testing the utility of iterative saturation mutagenesis. Adv. Synth. Catal. 351, 3287–3305 (2009).
29. Sayers, V., Lubrano, P., Li, Y. & Acevedo-Rocha, C. G. Unbiased libraries in protein directed evolution. Biochim. Biophys. Acta 1868, 140321 (2020).
30. Reetz, M. T. & Sanchis, J. F. Constructing and analyzing the fitness landscape of an experimental evolutionary process. ChemBioChem 9, 2260–2267 (2008).
31. DeMars, M. D. et al. Exploring the molecular basis for substrate specificity in homologous macrolide biosynthetic cytochromes P450. J. Biol. Chem. 294, 15947–15961 (2019).
32. Dubey, K. D. & Shaik, S. Cytochrome P450 – the wonderful nanomachine revealed through dynamic simulations of the catalytic cycle. Acc. Chem. Res. 52, 389–399 (2019).
33. O’Maille, P. E. et al. Quantitative exploration of the catalytic landscape separating divergent plant sesquiterpene synthases. Nat. Chem. Biol. 4, 617–623 (2008).
34. Fasan, R., Mehtarena, Y. T., Snow, C. D., Poulos, T. L. & Arnold, F. H. Evolutionary history of a specialized p450 propanoic monooxygenase. J. Mol. Biol. 383, 1069–1080 (2008).
35. Bloom, J. D., Labhatavukal, S. T., Otey, C. R. & Arnold, F. H. Protein stability promotes evolvability. Proc. Natl Acad. Sci. USA 103, 5869–5874 (2006).
36. Aklar, B. V. et al. Optimization of ligad phase shapes the evolution of a bacterial enzyme. Nat. Evol. 1, 1–6 (2017).
37. Rodrigues, I. V. et al. Biophysical principles predict fitness landscapes of drug resistance. Proc. Natl Acad. Sci. USA 113, E1470–E1478 (2016).
38. Zhang, Z.-G., Lonsdale, R., Sanchis, J. & Reetz, M. T. Extreme synergistic mutational effects in the directed evolution of a Baeyer–Villiger monooxygenase as catalyst for asymmetric sulfoxidation. J. Am. Chem. Soc. 136, 17262–17272 (2014).
39. Narayan, A. R. H. et al. Enzymatic hydroxylation of an unactivated methylene C-H bond guided by molecular dynamics simulations. Nat. Chem. 7, 653–660 (2015).
40. Basudhar, D. et al. Analysis of cytochrome P450 CYP119 ligand-dependent conformational dynamics by two-dimensional NMR and x-ray crystallography. J. Biol. Chem. 290, 9400–9417 (2015).
41. Batabya, D., Richards, L. S. & Poulos, T. L. Effect of redox partner binding on cytochrome P450 conformational dynamics. J. Am. Chem. Soc. 139, 13193–13197 (2017).
42. Feira, M. P., Pagnamenta, A. T., Damerell, D., Taylor, J. C. & Marsden, B. D. MichelinaGLo: sculpting protein views on web pages without coding. Bioinformatics 36, 3268–3270 (2020).
43. Saha, A. & Deisenhofer, J. Structure and function of cytochromes P450: a comparative analysis of three crystal structures. Structure 3, 41–62 (1995).

Acknowledgements
Support from the Max-Planck-Society and the LOEWE Research cluster SynChemBio is gratefully acknowledged. A.I. thanks the support from the National Key Research and Development Program of China (2019YFA0905000). This study was also supported in part by the European Research Council Horizon 2020 research and innovation programme (ERC-2015-StG-679001, to S.O.), Spanish MINECO (project PGC2018-101292-M).
B-I00, to S.O.; project PID2019-111300GA-I00, to M.G.-B.; and Juan de la Cierva -
Incorporación fellowship IJCI-2017-33411, to M.G.-B.). UdG (predoctoral fellowship
IFUdG2016, to L.D.), and Generalitat de Catalunya AGAUR (SGR-1707, to S.O.;
and Beatriu de Pinós H2020 MSCA-Cofund 2018-BP-00204, to M.G.-B.). M.P.F. is
supported by the Wellcome Trust [203141/Z/16/Z] and the NIHR Biomedical Research
Centre Oxford.

Author contributions
C.G.A.-R., S.H., S.O. and M.T.R. conceived the project. C.G.A.-R., A.L. and S.H. created
and purified mutants and measured their activity. P.L. and M.F. wrote the Python code
for automated calculation of epistatic effects. J.S. constructed and analysed the
fitness pathway landscapes. L.D. performed the computational modelling and analysis with
support and guidance from M.G.-B. and S.O. All authors wrote, revised and approved the
manuscript.

Funding
Open Access funding enabled and organized by Projekt DEAL.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material
available at https://doi.org/10.1038/s41467-021-21833-w.

Correspondence and requests for materials should be addressed to S.O. or M.T.R.

Peer review information Nature Communications thanks the anonymous reviewer(s) for
their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in
published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons
Attribution 4.0 International License, which permits use, sharing,
adaptation, distribution and reproduction in any medium or format, as long as you give
appropriate credit to the original author(s) and the source, provide a link to the Creative
Commons license, and indicate if changes were made. The images or other third party
material in this article are included in the article’s Creative Commons license, unless
indicated otherwise in a credit line to the material. If material is not included in the
article’s Creative Commons license and your intended use is not permitted by statutory
regulation or exceeds the permitted use, you will need to obtain permission directly from
the copyright holder. To view a copy of this license, visit http://creativecommons.org/
licenses/by/4.0/.

© The Author(s) 2021