Unpicking the Cause of Stereoselectivity in Actinorhodin Ketoreductase Variants with Atomistic Simulations

Stefano Artin Serapian, Marc van der Kamp

Submitted date: 03/12/2018 • Posted date: 04/12/2018
Licence: CC BY-NC-ND 4.0

Citation information: Serapian, Stefano Artin; van der Kamp, Marc (2018): Unpicking the Cause of Stereoselectivity in Actinorhodin Ketoreductase Variants with Atomistic Simulations. ChemRxiv. Preprint.

Ketoreductase enzymes (KRs) with a high degree of regio- and stereoselectivity are useful biocatalysts for the production of small, specific chiral alcohols from achiral ketones. Actinorhodin KR (actKR), part of a type II polyketide synthase involved in the biosynthesis of the antibiotic actinorhodin, can also turn over small ketones. In vitro studies assessing stereocontrol in actKR have found that, in the “reverse” direction, the wild-type (WT) enzyme’s mild preference for S-α-tetralol is enhanced in certain mutants (e.g. P94L); and entirely reversed in others (e.g. V151L) in favor of R-α-tetralol. Here, we employ efficient atomistic simulations to rationalize these trends in WT, P94L, and V151L actKR, using trans-1-decalone (1) as the model substrate. Three potential factors (FI-FIII) are investigated: frequency of pro-R vs. pro-S reactive poses (FI) is assessed with classical molecular dynamics (MD); binding affinity of pro-R vs. pro-S orientations (FII) is compared using the binding free energy method MM/PBSA; and differences in reaction barriers towards trans-1-decalol (FIII) are assessed by hybrid semiempirical quantum / classical (QM/MM) MD simulations with umbrella sampling, benchmarked with density functional theory. No single factor is found to dominate stereocontrol: FI largely determines the selectivity of V151L actKR, whereas FIII is more dominant in the case of P94L. It is also found that formation of S-trans-1-decalol or R-trans-1-decalol mainly arises from the reduction of the trans-1-decalone enantiomers (4aS,8aR)-1 or (4aR,8aS)-1, respectively. Our work highlights the complexity of enzyme stereoselectivity as well as the usefulness of atomistic simulations to aid the design of stereoselective biocatalysts.

File list (3)

- Serapian_vanderKamp_Unpicking the cause of stereosele... (1.24 MiB)  view on ChemRxiv  download file
- SupportingInformation.pdf (1.51 MiB)  view on ChemRxiv  download file
- starting_top_rst_pdb_frcmod_prepc_files.zip (24.29 MiB)  view on ChemRxiv  download file
Unpicking the cause of stereoselectivity in actinorhodin ketoreductase variants with atomistic simulations

Stefano A. Serapian† and Marc W. van der Kamp*†‡

†School of Biochemistry, University of Bristol, University Walk, Bristol, BS8 1TD, United Kingdom
‡Centre for Computational Chemistry, University of Bristol, Cantock’s Close, Bristol, BS8 1TS, United Kingdom

KEYWORDS. Biocatalysis, Computational enzymology, QM/MM, Molecular dynamics, Ketoreductases, Chiral alcohols

ABSTRACT: Ketoreductase enzymes (KR s) with a high degree of regio- and stereoselectivity are useful biocatalysts for the production of small, specific chiral alcohols from achiral ketones. Actinorhodin KR (actKR), part of a type II polyketide synthase involved in the biosynthesis of the antibiotic actinorhodin, can also turn over small ketones. In vitro studies assessing stereocontrol in actKR have found that, in the “reverse” direction, the wild-type (WT) enzyme’s mild preference for S-a-tetralol is enhanced by certain mutations (e.g. P94L); and entirely reversed by others (e.g. V151L) in favor of R-a-tetralol. Here, we employ efficient atomistic simulations to rationalize these trends in WT, P94L, and V151L actKR, using trans-1-decalone (1) as the model substrate. Three potential factors (FI-FIII) are investigated: frequency of pro-R vs. pro-S reactive poses (FI) is assessed with classical molecular dynamics (MD); binding affinity of pro-R vs. pro-S orientations (FII) is compared using the binding free energy method MM/PBSA; and differences in reaction barriers towards trans-1-decalol (FIII) are assessed by hybrid semiempirical quantum / classical (QM/MM) MD simulations with umbrella sampling, benchmarked with density functional theory. No single factor is found to dominate stereocontrol: FI largely determines the selectivity of V151L actKR, whereas FIII is more dominant in the case of P94L. It is also found that formation of S-trans-1-decalol or R-trans-1-decalol mainly arises from the reduction of the trans-1-decalone enantiomers (4aS,8aR)-1 or (4aR,8aS)-1, respectively. Our work highlights the complexity of enzyme stereoselectivity as well as the usefulness of atomistic simulations to aid the design of stereoselective biocatalysts.

1. Introduction

Typically found in actinomycete bacteria, type II polyketide synthases (PKSs) are large multi-enzyme complexes consisting of multiple standalone units.1-3 In a perfect example of enzymatic teamwork, each of these PKS domains is highly specialized in catalyzing a particular biosynthetic step, first “growing” a reactive polyketide chain, and then processing it into a variety of natural products; these include anticancer agents (e.g. daunorubicin and doxorubicin), and antibiotics (e.g. actinorhodin, tetracycline, and doxycycline).1,2

A key component of PKSs is their ketoreductase (KR) domain,4-7 which typically exhibits some degree of regio- and stereoselectivity in catalyzing the reduction of polyketide chains. Together with several other KRs unrelated to PKSs, these enzymes have emerged as promising commercial biocatalysts8-10 for the manufacture of small chiral alcohols from achiral ketones; and even for the synthesis of “unnatural natural products” with potentially improved antibiotic properties.2 Such rise to prominence was significantly aided by the advent of novel protein reengineering techniques,11-12 which have revolutionized biocatalysis by speeding up the design and screening of evolved KR mutants.8-10

Actinorhodin KR (actKR) from Streptomyces coelicolor has been attracting interest as a potential biocatalyst since at least the mid-2000s.5,13,14 Examination of its sequence and crystal structures5,13,14 confirmed the enzyme’s homotetrameric nature (Figure 1a), and its structural and mechanistic similarity with enzymes of the short-chain dehydrogenase/reductase (SDR) family, especially fatty acid synthases (FASs). Essential for its activity are the presence in each active site of the NADPH cofactor, and of the catalytic tetrad Asn114-Ser144-Tyr157-Lys161 (Figure 1b).13

Another salient structural element is the flexible α6-α7 loop (Figure 1c and d), which varies substantially in size and composition across SDRs, and was proposed to play some role in the stereocontrol and activity of Lactobacillus kefir short-chain alcohol dehydrogenase.10 Conformational variation of actKR’s α6-α7 loop between “open” and “closed” is evident when superimposing monomers A and B of the wild-type actKR crystal structure (PDB: 2RH4; Figure 1d).14

The catalytic cycle of actKR begins when a ketone or polyketide substrate binds to its active site so that the oxygen of the carbonyl to be reduced is sandwiched between the catalytic Ser144 and Tyr157, with whose hydroxyl groups it forms hydrogen bonds (Figure 1b). Reduction then kicks off (Scheme 1; Figure 1b) when a hydride (H-) from NADPH transfers to the carbonyl carbon, yielding the oxidized form NADP+; conversely (but asynchronously) with respect to the hydride transfer, the carbonyl oxygen abstracts a proton from Tyr157 (to which it remains hydrogen-bonded). The deproto- nated Tyr157 is further stabilized by retaining its hydrogen bond to one of the cofactor’s ribose hydroxyl groups (and is reproto- nated in subsequent steps via H+ transfers from the ribose and Lys161).15 During this fundamental mechanistic step, which was shown to be rate-limiting in a human ketoacyl reductase,15 the substrate’s orientation with respect to the plane of NADPH’s nicotinamide moiety determines whether the H+ attacks in a pro-R or pro-S fashion (Scheme 1 top vs. bottom), and thus ultimately dictates the stereochemical outcome of the reaction (R- or S- alcohol).
Figure 1. Structural features of wild-type actinorhodin ketoreductase (based on PDB ID 2RH4).14 (a) Overall assembly of the four monomers A-D (grey: A; C; red: B, D) showing: the catalytic tetrad N114-S144-Y157-K161 in A (green sticks); and the NADPH cofactor in A-D (yellow spheres and sticks). H atoms omitted for clarity. (b) Closeup of the active site of monomer A, showing NADPH (C atoms in yellow); labelled tetrad residues (C in green); and a ketone fragment (C in magenta) docked in reactive position. O, N, P are rendered in red, blue, and orange, respectively. H atoms omitted for clarity, except for: hydroxyl H atoms on S144, Y157, NADPH ribose (grey spheres); and the reductive H\(^+\) (black). Relevant hydrogen bonds are shown in magenta; black line denotes direction of reductive H\(^+\) transfer. (c) Side view of actKR as shown in (a), highlighting the a6-a7 loop in monomers A-D (A, C: open: light blue; B, D: closed: pink). (d) Monomer B superimposed onto monomer A.

In nature, the enzyme typically forms a complex with an acyl carrier protein (ACP) bearing a 16-carbon polyketide chain (octaketide; cf. Chart 1 top), which is then unsheathed into one of the active sites, likely cyclized between the C7 and C12 positions and reduced specifically at the C9 position,\(^{14,16-17}\) before further processing by other PKS modules.

In vitro, wild-type (WT) actKR and 29 strategically chosen mutants were extensively examined by Korman, Javidpour, et al.,\(^{14,16-17}\) looking for improved stereocontrol in the turnover of model ketone and alcohol substrates (i.e. in the “forwards” and “reverse” directions, respectively). More specifically, to study the “forwards” reduction reaction, the authors’ preferred substrate is trans-1-decalone (1; Chart 1, left),\(^{14,16-17}\) since other potential candidates, notably the part-aromatic α-tetralone, were found to be turned over very sluggishly.\(^{14}\) Rather than employing 1’s directly corresponding alcohol trans-1-decalol (2) with its four possible stereoisomers (Chart 1, center), the authors assess the “reverse” oxidation reaction using the diastereomeric pair R-α-tetralol / S-α-tetralol (\(R\cdot3 / S\cdot3\); Chart 1, right),\(^{14,16-17}\) both of which are readily available commercially (in contrast to stereoisomers of 2).

Assuming that the selectivity of the “forwards” and “reverse” reactions are still directly comparable microscopically (despite the slight change in substrates), two of the 29 mutants demonstrate particularly high stereoselectivity (as per data in Table 1): P94L has exclusive specificity for \(S\cdot3\); and V151L, has exclusive specificity for \(R\cdot3\). By comparison, WT actKR only has a very mild preference for \(S\cdot3\) over \(R\cdot3\) (3.5:1).

Table 1. In vitro specificity constants of key actKR variants towards reduction of 1 and oxidation of 3.

| Variant | \(k_{cat}/K_m (1)\) \(^a,b\) | \(k_{cat}/K_m (R\cdot3)\) \(^*\) | \(k_{cat}/K_m (S\cdot3)\) \(^*\) |
|----------|-----------------|-----------------|-----------------|
| WT\(^c,d\) | 3.23 ± 0.32 | 0.010 ± 0.001 | 0.035± 0.006 |
| V151L\(^e\) | 1.28 ± 0.18 | 0.026 ± 0.005 | inhibition |
| P94L\(^d\) | 1.02 ± 0.59 | inhibition | 0.036 ± 0.011 |

\(^a\) s\(^{-1}\) mM\(^{-1}\). \(^b\) Racemate: (4aS,8aR)-1 and (4aR,8aS)-1 (cf. text). \(^*\) Reference \(^14,16\) Reference \(^16,17\).

The “forwards” reaction is further complicated by the fact that 1 (e.g. as purchased from Sigma Aldrich) exists as a racemate of the enantiomers (4aS,8aR)-1 and (4aR,8aS)-1 (henceforth SR-1 and RS-1, and with dark vs. light blue color codes, respectively, in Chart 1 and subsequent figures). Consequently, reduction of 1 can yield four different stereoisomers of 2 (Chart 1, center): when attacked by H\(^+\) in pro-R orientation, SR-1 can only yield \((1R,4aS,8aR)\cdot2\) (henceforth \(RSR\cdot2\); black in Chart 1), and RS-1 can only yield \((1R,4aR,8aS)\cdot2\) (\(RRS\cdot2\); grey); conversely, when attacked in pro-S orientation, SR-1 and RS-1 can only yield \((1S,4aS,8aR)\cdot2\) and \((1S,4aR,8aS)\cdot2\), respectively (\(SSR\cdot2\) and \(SRS\cdot2\); red and orange in Chart 1). Thus, when simulating the enzymatic reaction to study its stereoselectivity in the “forwards” direction, all four outcomes need to be considered. For illustration, putative Michaelis complexes in WT actKR for these scenarios are shown in Figure 2.
Scheme 1. Rate-limiting step of actKR (reductive hydride transfer) and its stereochemistry.\textsuperscript{a,b}

Priority of substituents: \textit{OH} > \textit{R'} > \textit{R} > \textit{H}

\textsuperscript{a} Ketone substrate orientation (left) determines whether \textit{H}– attack is \textit{pro-R} or \textit{pro-S}, and accordingly gives rise to an \textit{R}- or \textit{S}- product (right). \textsuperscript{b} Magenta lines denote hydrogen bonds.

Chart 1. actKR substrates discussed in this work.\textsuperscript{a,b}

\textsuperscript{a} 1 and 2 color-coded as in Figure 4 and Figure 6, respectively. \textsuperscript{b} Underlined \textit{R} or \textit{S} labels in 2 and 3 (and throughout the text) mark stereocenters introduced/removed by actKR. \textsuperscript{c} C7-C12 cyclisation prior to reduction at C9. \textsuperscript{d} [S] denotes link to ACP (see text).
Due to increases in computer speed and improvements in algorithms, parameters, and usability over the past decades, in silico biomolecular simulation is increasingly used to complement experimental enzymatic studies, including those focusing on the design of novel enzyme variants relevant for biocatalysis. Indeed, enzyme reactivity and stereocontrol are often the result of a subtle interplay between different factors that are difficult to examine separately experimentally. In short, enzyme variants conferring stereoselectivity can either do so by favoring formation of a particular reactive complex; or by favoring efficient reaction of the stereochemistry-determining step (or subsequent steps); or a combination of both. Here, we assess the stereoselectivity of actKR variants using three different factors that can easily be assessed individually through simulation.

Factor I (henceforth FI) is the likelihood of the enzyme-substrate complex attaining a “reactive” pose (or Michaelis complex); or, in other words, how likely it is for the substrate to reach an orientation and conformation at which key interatomic distances are sufficiently close for a reaction to occur. Any imbalance of this factor in favor of reactive pro-R or pro-S orientations will thus affect an enzyme’s stereocontrol by giving a head start to one reaction pathway over the other. In principle, FI describes the energetics of reaching and maintaining a reactive pose. This incorporates the interplay between conformational fluctuations in the enzyme and substrate (“dynamics”), and may even be affected by mutations far from the active site. Computationally, this can be probed by classical molecular dynamics (MD) simulations; in some cases (e.g., with sufficient structural information and particular prochiral binding orientations precluded), molecular docking may suffice.

Factor II (henceforth FII) can be seen as part of FI, but focusing solely on the end point, i.e. the binding affinity (or binding free energy; \( \Delta G_{\text{bind}} \)) of a specific substrate in its reactive pose within an enzyme’s active site. Assuming that reactive poses can be attained by both prochiral orientations, one prochiral orientation may be preferred over the other (i.e. \( \Delta G_{\text{bind}} \neq 0 \)), thereby favoring formation of the product associated with the preferred prochirality. In silico methods for calculating \( \Delta G_{\text{bind}} \) include (but are by no means limited to) WaterSwap, enhanced Monte Carlo, and MM/PBSA.

Finally, Factor III (henceforth FIII) is the height of the free energy barrier separating reactants and products of different chirality: if a substrate in a reactive pro-R or pro-S orientation is closer in energy to the corresponding transition state, then the associated product will be more accessible than its counterpart. A range of computational approaches are available to determine such barriers in enzyme reactions; a popular option is the use of a hybrid quantum / classical (QM/MM) approach, where the computationally costly QM treatment can be limited to regions of chemical change (without time-intensive parameterization).

In the present work, we describe our computational efforts to rationalize the stereoselectivity of WT, P94L, and V151L actKR towards I, determining the role of FI-FIII in each case. Based on the assumption that the stereospecific oxidation of 3 is comparable to the stereoselective reduction of 1, we expect P94L actKR to preferentially form (\( \text{S} \))-trans-1-decalol; V151L actKR to preferentially form (\( \text{R} \))-trans-1-decalol; and a slight preference for (\( \text{S} \))-trans-1-decalol by WT actKR. Focusing on a potential biocatalyst that has thus far been little studied in silico, we aim to show that physically realistic but relatively inexpensive computational simulations can be readily employed in this and similar contexts, to explain the behavior of other promising biocatalysts, and helping to direct the design of novel mutants with enhanced stereocontrol.

### 2. Computational Strategy and Details

#### General Procedure

For FI, FII, and FIII investigations alike, the first step entails running a number of independent classical MD simulations, to sample conformational space at relatively low computational cost.

In MD simulations for FI, which we henceforth refer to as “free”, the substrate is left free to explore as many orientations as possible in the binding site: the frequency of reactive pro-R and pro-S poses may be directly extracted from these runs. To avoid substrate diffusion from the active site, a one-sided harmonic restraint (\( k = 50 \text{ kcal mol}^{-1} \text{ Å}^{-2} \)) is employed when the center of mass of the substrate moves farther than 8 Å from the center of mass of NADPH’s nicotinamide moiety.

To provide starting points for evaluating FII and FIII, MD simulations were also performed of the “Michaelis” or “reactive” pro-R and pro-S enzyme-substrate complexes, from which, in principle, the chemical reaction can readily begin. These “restrained” MD simulations include one-sided harmonic restraints on three key interatomic distances, and an additional dihedral angle restraint to prevent the substrate from ‘flipping’ between pro-R and pro-S (see Supporting Information for further details).

To determine the role of FII, \( \Delta G_{\text{bind}} \) of SR-I and RS-I in their two prochiral orientations is calculated through a series of MM/PBSA calculations on the resulting “restrained” trajectories. Probing the role of FIII requires adequate sampling of free energy barriers for the conversion of I’s two enantiomers to one of the four accessible stereoisomers of 2. To do this, we select a total of 120 representative snapshots from the “restrained” MD simulations and carry out, on each snapshot, hybrid quantum/classical MD (QM/MM MD) with umbrella sampling (US) along a reaction coordinate (vide infra).

For simplicity, all classical and QM/MM MD simulations are run in two distinct sets, treating enantiomers SR-I and RS-I as separate substrates and, thus, with only one or the other enantiomer occupying all four active sites in each set.

#### Starting Structures

Six starting structures [actKR-(NADPH)\(_{2}\), I] are set up with all four sites occupied: SR-I or RS-I in WT, P94L, or V151L actKR. All starting structures are constructed from our reference WT crystal structure (PDB ID: 2RH4), to demonstrate the possibility of assessing enzyme variants for which no crystal structure has been obtained. Existing mutant crystal structures (V151L; PDB ID: 4DBZ and P94L; PDB IDs: 3RI3, 3QRW) are checked to ensure that constructed mutants retain plausibly oriented sidechains. Constructed structures are henceforth labelled as follows: WT-SR-I, (cf. Figure 2a and 2b), V151L-SR-I, and P94L-SR-I for one enantiomer of I; and WT-RS-I, (cf. Figure 2c and 2d), V151L-RS-I, and P94L-RS-I for the other.

In all six cases, active sites in monomers A and B are populated with the substrate in pro-R orientation whereas those in monomers C and D are populated with the substrate in pro-S orientation. All four monomers are modelled from residues 1 to 261. All residues are modelled in their standard protonation
states, in line with pKₐs predicted by PROPKA 3.1. Hydrogens are added with AmberTools' (version 17) reduce utility, resulting in His162 being singly protonated on Nδ1; His153 and His201 on Nε2. Using lmpwps, standard N- and C-termini are introduced and the structure is solvated in a truncated octahedral box of water extending at least 11 Å from any protein atom, with 40 Na⁺ ions added to neutralize the system. Starting structure files and further details regarding starting structure generation are available as Supporting Information.

**Classical MD.** MD simulations of our six actKR-(NADPH)₁₄₁ systems in explicit water are run with the AMBER software package (version 16, 2017 distribution), taking advantage of GPU acceleration where applicable. Postprocessing and analysis of MD trajectories are carried out with the CPPTRAJ utility; and visual inspection is conducted with VMD. The protein and ions are described by the ff14SB forcefield, the NADPH cofactor by the forcefield from Holmberg and coworkers, and water using the TIP3P model. The GAFF forcefield with AM1-BCC charges derived by antechamber is used for SR⁻₁ and RS⁻₁ (see also Supporting Information). The default cutoff of 8 Å is used to compute Lennard-Jones and Coulomb interactions, with Coulomb interactions beyond this limit computed using the Particle Mesh Ewald method.

For both the “free” and “restrained” MD simulations (vide supra) we carry out eight independent simulations (different random seeds) for each of the six actKR-(NADPH)₁₄₁ starting structures. This results in a total of 8 × 6 = 48 MD runs for both the “free” and “restrained” sets (96 in total). Each of the MD runs is carried out with the following general procedure (restraints retained throughout): minimization (600 steps); solvent equilibration (9 ps, NVT); heating (20 ps, NVT); equilibration (2040 ps, NpT); and production (12 ns, NpT). Production runs are conducted in the NpT ensemble with a time-step of 2 fs. A constant pressure of 1 atm is enforced via the Berendsen barostat and a constant temperature of 298 K is enforced via the Langevin thermostat (collision frequency set to 5 ps⁻¹). Bonds containing hydrogen are constrained by employing the SETTLE and SHAKE algorithms. (See Supporting Information for details on settings in the pre-production stages).

**MM/PBSA Calculations.** MM/PBSA calculations are run on 500 snapshots from each restrained MD trajectory, on 16 processors, using AmberTools' MM/PBSA, py.MPI utility, with default atomic radii, settings, and parameters. The only exceptions are the ionic strength, which is set at 0.025 mol dm⁻³ to reflect the 40 Na⁺ ions originally present in the simulations; and the internal dielectric constant εᵣₐₜ, which is set to 4.0 as advised by Wang and coworkers. The six actKR-(NADPH)₁₄₁ topologies are preprocessed using the ante-MM/PBSA, py tool.

The eight trajectories of each of the six actKR-(NADPH)₁₄₁ systems are parsed by MM/PBSA, py.MPI as six individual 96 ns “supertrajectories”, and each MM/PBSA calculation then runs on snapshots taken every 24 ps. Since each system contains four instances of 1 (i.e. one in each active site), four separate MM/PBSA runs are required on each “supertrajectory” (i.e. 24 in total, giving 24 ΔGₗ₀⁽⁾values). Entropic corrections to these values are calculated using the interaction entropy method reported by Duan et al.
one. The potential of mean force (PMF) along (x − y) amounts to the free energy profile of the reduction, and was obtained from the US runs using the weighted histogram analysis method (WHAM)64-65 with the eponymous program by Grossfield.66 This also carries out error analysis through Monte-Carlo bootstrapping. PMFs are extracted both individually for all 120 US runs, and cumulatively for the 12 processes sampled. A number of runs were discarded as outliers, for example when they did not manage to capture the concerted proton abstraction as observed in our benchmarking studies; all discarded runs were replaced. Details regarding reruns, WHAM, and error analysis are given in the Supporting Information.

3. Results and Discussion

In the subsections below, we first discuss findings from our calculations regarding the individual effects of FI-FIII on actKR stereoselectivity upon reduction of 1 to 2. We then examine which combinations of factors come into play when determining stereocontrol in WT, P94L, and V151L actKR, drawing comparisons with corresponding in vitro observations on the oxidation of 3,14, 16-17 and including trends between enantiomers of 1. Finally, we discuss our protocols and their potential for high-throughput screening, focusing on the simulation lengths required to achieve appreciable accuracy.

Formation of reactive complexes: FI. We measure the difference (Δ) between the frequency of reactive pro-R and reactive pro-S poses of SR-1 (Δ(pro-R − pro-S)SR-1) and RS-1 (Δ(pro-R − pro-S)RS-1) in WT, P94L, and V151L actKR as they occur in our “free” MD simulations. Criteria to determine whether or not a pose is “reactive”, and whether its prochirality is pro-R or pro-S, are based on the restraints imposed on “restrained” MD simulations used to study FII and FIII (Supporting Information, Table S1).

For each of our six simulated systems (WT-SR-1a, V151L-SR-1a, P94L-SR-1a, WT-RS-1a, V151L-RS-1a, and P94L-RS-1a), statistics for Δ(pro-R − pro-S)SR-1 or Δ(pro-R − pro-S)RS-1 are measured cumulatively: in other words, the 8 MD replicas conducted for each of the six systems are considered collectively, as are all 4 active sites. Δ(pro-R − pro-S)SR-1 and Δ(pro-R − pro-S)RS-1 are calculated for each variant using:

\[
\Delta \text{(pro-R − pro-S)SR-1} = \left( \frac{\text{reactive pro-R SR-1 poses}}{96000} \right) \times 100
\]

and

\[
\Delta \text{(pro-R − pro-S)RS-1} = \left( \frac{\text{reactive pro-R RS-1 poses}}{96000} \right) \times 100 ;
\]

where the denominator 96000 reflects the fact that each of the eight replicas we are dealing with has 3000 frames (250 per ns), and that each of these frames has 4 active sites (and 4 instances of 1). A general Δ for trans-1-decalone as a whole (Δ(pro-R − pro-S)1) is then obtained by summing Δ(pro-R − pro-S)SR-1 and Δ(pro-R − pro-S)RS-1.

Figure 4. Difference (Δ) in the frequency of reactive pro-R and reactive pro-S poses in “free” MD simulations of: (dark blue) SR-1; (light blue) RS-1; and (brown) 1 overall, within P94L, V151L, and WT actKR. Negative values indicate excess of pro-R reactive poses (red points), positive values excess of pro-S poses (black point). See also color codes in Chart 1; lines are to guide the eye. Errors indicated are based on leave-one-out procedure (details in Supporting Information).

Although the overall frequency of pro-R and pro-S reactive poses is generally low (the highest being 8.3% across the 8 V151L-RS-1a MD replicas), significant trends still emerge (Figure 4). The first observation to make is that, regardless of actKR variant, simulations with SR-1 (dark blue) consistently show an excess of reactive pro-S poses (most marked in WT-SR-1a at 51%), whereas RS-1 systems (light blue) all favor reactive pro-R poses instead (most markedly in V151L-RS-1a at 7.8%).

A second noteworthy observation is that for the overall difference in pro-R / pro-S reactive poses (Δ(pro-R − pro-S)1; brown line in Figure 4), V151L actKR has an excess of reactive pro-R poses (3.9%; black diamond in Figure 4), whereas both WT and P94L actKR have an excess of reactive pro-S poses (red diamonds).

The excess of reactive pro-R poses seen in V151L actKR is encouragingly in line with the mutant’s observed in vitro specificity towards R-3 over S-3 in the reverse direction.17 Similarly, the slight excess of pro-S reactive poses emerging from simulations of P94L and WT actKR is in line with in vitro observations that both are preferentially turning over S-3.16 Nonetheless, if FI were the predominant factor driving P94L actKR’s exclusive preference for S-3, one would expect a far greater excess of reactive pro-S poses to emerge.

Binding free energy of reactive poses: FII. The role of FII was investigated through separate MM/PBSA binding free energy calculations on each of the four active sites of our six actKR-(NADPH)2-1a systems (using snapshots taken at 24 ps-intervals from 8 independent restrained MD runs of 12 ns each). ΔGbind values for active sites A and B (containing pro-R poses) and those for active sites C and D (containing pro-S poses) are averaged and plotted in (Figure 5).
Figure 5. Binding free energies (\(\Delta G_{\text{bound}}\)) predicted by MM/PBSA calculations on (left) SR-1; and (right) RS-1; either in pro-\(R\) poses (black or grey, respectively); or in pro-\(S\) poses (red or orange, respectively). Observations for pro-\(R\) poses are averaged between monomers A and B; those for pro-\(S\) between C and D. See also color codes in Chart 1; lines are guides for the eyes. Error bars correspond to the standard error of the mean.

For SR-1, \(\Delta G_{\text{bound}}\) values for pro-\(R\) and pro-\(S\) poses are highly similar (left-hand side of Figure 5), indicating no particular preference for either prochirality in any of the enzyme variants. The only significant difference is that the P94L mutant indiscriminately shows a higher affinity for both orientations (more negative by \(-1.5\) kcal mol\(^{-1}\) with respect to WT and V151L actKR). More variation is observed between the binding energies for pro-\(R\) and pro-\(S\) poses of RS-1 (right-hand side of Figure 5). Whilst both orientations retain similar binding affinities in WT actKR (\(\Delta G_{\text{bound}} = 0\)), this is clearly not the case in the two variants: RS-1 in pro-\(R\) orientation has a significantly higher affinity (compared to its pro-\(S\) orientation) in both V151L actKR (\(\Delta G_{\text{bound}} = 1.9\) kcal mol\(^{-1}\)) and P94L actKR (\(\Delta G_{\text{bound}} = 3.6\) kcal mol\(^{-1}\)). This increase in \(\Delta G_{\text{bound}}\) when going from V151L to P94L actKR is mainly caused by a loss in affinity for the pro-\(S\) orientation.

In the case of WT actKR, enantiomers of 1 all have similar binding affinities regardless of orientation (with a slight preference of SR-1 in either pose over RS-1, \(\Delta G_{\text{bound}} = 0.75\) kcal mol\(^{-1}\)). In the case of V151L, only the pro-\(R\) RS-1 pose has significantly higher affinity, with pro-\(S\) RS-1 at the same level as both SR-1 orientations (in line with the preferential formation of reactive poses for pro-\(R\) RS-1, vide supra). For P94L, there is a significant loss in affinity of pro-\(S\) RS-1 (orange), whereas pro-\(R\) RS-1 and both SR-1 poses have similar affinities (within 1 kcal mol\(^{-1}\)). Based on the stereoselectivity observed in vitro,\(^{14, 16}\) one might expect to see a greater (more negative) binding affinity of pro-\(S\) poses for at least one of SR-1 / RS-1 (red/orange) in the case of P94L (as well as a greater binding affinity of pro-\(R\) poses in at least one of SR-1 / RS-1 in the case of V151L, as is observed here). Thus, FIII is not likely to be involved in determining the stereoselectivity of P94L actKR, but may contribute instead to that of V151L actKR.

Free energy barrier of reaction: FIII. To probe the effects of FIII on the stereoselectivity of WT actKR and its mutants, we have used QM/MM MD umbrella sampling simulations to obtain free energy profiles for each of the four possible reductions of trans-1-decalone to trans-1-decalol (1 to 2, Figure 6). Profiles have been normalized so that the free energy of 1 on the “reactants” side is equal to zero, and each plot has been color-coded according to the stereoisomer of 2 obtained (cf. color codes in Chart 1). Note that the unusually high energy barriers are mainly due to the use of PM6 for the QM region (see Supporting Information); we note that differences in reaction barrier height are what is relevant for investigation of the influence of FIII on stereoselectivity.

Figure 6. Free energy profiles resulting from QM/MM MD US simulations (PM6/ff14SB) of the conversion of 1 to 2 in: (left) P94L actKR; (center) V151L actKR; and (right) WT actKR. In each panel, each curve is the cumulative result of 10 individual simulations run for the 4 possible combinations of: SR-1 or RS-1 (darker- vs. lighter-colored curves); and final \(R\)-2 or \(S\)-2 stereochemistry (black/grey vs. red/orange, respectively; cf. Chart 1). Regions close to the transition state (blue dotted line) are magnified in the blue insets, with error bars; errors are obtained through Monte-Carlo bootstrapping (see Supporting Information for details).

For P94L actKR (Figure 6, left panel), \(\Sigma SR\)-2 (red) is indicated as the most easily accessible stereoisomer of 2, through a free energy barrier that is at least 5 kcal mol\(^{-1}\) lower than those required to access either \(R\)-2 stereoisomer, and about 4 kcal mol\(^{-1}\) lower than that required for the formation of \(\Sigma RS\)-2. This finding is in line with the \(S\)-specificity detected in vitro,\(^{16}\) and is further substantiated by the fact that reaching either of the two \(R\)-2 stereoisomers (black and grey) is much more difficult in this variant than in either V151L or WT actKR. The fact that the barrier to reach the other \(S\)-2 stereoisomer (\(\Sigma RS\)-2; orange) is also higher than in the other forms of actKR—together with the aforementioned lower binding affinity for the pro-\(S\) RS-1 reactive pose (Figure 5)—suggests that \(S\)-2 selectivity in P94L actKR is likely to arise from formation of \(\Sigma SR\)-2 from SR-1.

Findings for V151L (Figure 6, center panel) are less clear: although barriers to reach the two \(R\)-2 stereoisomers are indeed significantly lower than in P94L and WT actKR, barriers to reach the \(S\)-2 stereoisomers are both even lower, with formation of \(\Sigma RS\)-2 exhibiting the lowest free energy barrier out of all investigated combinations. Taken on its own, this contradicts the in vitro observation (specificity towards \(R\)-3 vs. \(S\)-3):\(^{17}\) other factors are thus likely to play a more dominant role (vide infra).

For WT actKR (Figure 6, right panel), a slightly lower free energy barrier is only detected for the conversion of RS-1 to
Different factors dominate for different variants. Taking into account the data for FI-FIII arising from our different simulations (Figure 4, Figure 5, and Figure 6), it becomes clear that the causes driving stereocore in each actKR variant are likely to be different.

For V151L actKR, expected to have a strong R-preference, the RS-1 enantiomer is significantly more prone to form reactive pro-R poses in the enzyme active site (FI). In addition, binding of reactive pro-R RS-1 poses is thermodynamically favored over the other possibilities (FII). It is true that reaching RRS-2 from pro-R RS-1 is significantly easier in V151L actKR than in any of the other variants investigated (FIII; Figure 6), however, this in itself this would not enhance stereoselectivity, as barriers to other species are equally reduced. It therefore appears that for V151L actKR, stereoselectivity is mainly determined by the steric and thermodynamics prior to the reduction step (FI and FII). As such, our calculations indicate that this variant will predominantly form RRS-2 (instead of RSR-2).

In the case of the WT enzyme, mild selectivity can be expected towards ±2 (based on its experimentally determined specificity towards ±3). The pro-± reactive poses are somewhat more easily attained within its active site (FI; Figure 4), with SR-1 being the best-placed enantiomer to reach its SSR-2 product. Thus, for SSR-2 formation (compared to SSR-2), this “steric head start” is likely important. Formation of the other ±-isomer, SSR-2, is favored by its lower free energy barrier (FIII; Figure 6). FII (Figure 5) appears to play no significant role for stereoselectivity.

For P94L actKR, the situation is significantly different: its strong ±-selectivity (as expected from its in vitro specificity towards ±-3) mostly arises due to a change in relative reaction barriers (FIII). Specifically, formation of SSR-2 is clearly preferred (Figure 6). FI may play some additional role, since SR-1 forms more pro-± reactive poses. Altogether, the simulations further indicate that this variant will predominantly form SSR-2 (instead of SSR-2): pro-± RS-1 reactive poses occur less frequently and have lower affinity than pro-SR-1 reactive poses (Figure 4; Figure 5), and they are also much less favored to react to SSR-2 (FIII; Figure 6).

**Table 2. Key Structural Features of Reactive Poses of 1**

| Enantiomer | SR-1 | RS-1 | RS-1 | RS-1 |
|------------|------|------|------|------|
| Prochirality | pro-R | pro-± | pro-R | pro-± |
| Non-CO ring** | → 94 | → 151 | → 94 | → 151 |
| H-attack | Axial | Equatorial | Equatorial | Axial |

**Table 2. Key Structural Features of Reactive Poses of 1**

**Figure 7.** Representative snapshots of approximate transition states from QM/MM MD US simulations (obtained from clustering on the substrate RMSD after alignment on the cofactor), (a) pro-R RS-1 to RRS-2 in WT actKR; (b) pro-± SR-1 to SSR-2 in WT actKR; (c) pro-R RS-1 to RRS-2 in V151L actKR; and (d) pro-± SR-1 to SSR-2 in P94L actKR. In each case, snapshots are taken from the 0.2 simulation window; all four choices depict equatorial attack. Key and orientation: same as Figure 2; mutations in (c) and (d) are marked in magenta.

There appears to be a slight difference in the effect brought about by mutant sidechains. The L151 sidechain (Figure 7c) extends more directly into the active site, and thus influences prochirality by ‘stealing’ volume from prospective pro-± poses (from the side of the non-carbonyl ring), making such poses less likely (FI). The L94 sidechain appears to have a subtler effect (Figure 7d), potentially affecting the orientation of the substrate during reaction (thereby affecting FIII).

**Trans-1-decalone enantiomer:**

Our work indicates that P94L actKR prefers facilitating the reaction of pro-± SR-1, and V151L actKR prefers facilitating pro-R RS-1. A significant difference in SR-1 and RS-1 complexes with actKR already emerged from investigation of FI (Figure 4), indicating a general preference of pro-± reactive poses with SR-1 (dark blue), and pro-R reactive poses with RS-1 (light blue), regardless of the actKR variant considered. These observations, in combination with the geometrical features from the simulations (Table 2 and Figure 7), indicate that both enantiomers of 1 are prone to react in the prochirality that favors equatorial rather than axial hydride attack. This contrasts with previous literature on the non-enzymatic reduction of 1 in vitro**69,70** (with reagents such as [BH₄⁻] and [AH₄⁻]⁻), which shows that the favored product is instead the one resulting from axial attack. A similar preference for the axial product was found computationally for the non-enzymatic reduction of cyclohexanone,**68** with the transition state for axial attack found to be about 1.8 kcal mol⁻¹ more stable.

The preference we predict for equatorial hydride attack in the enzymatic reduction by actKR is, however, entirely in line with experimental findings by Østergaard et al.,**69** which
showed SR-1 (from a racemate of 1) to preferentially undergo equatorial attack (to form SSR-2) within the ketoreductase module of the erythromycin polyketide synthase. The authors further concluded that, in the reverse direction, the enzyme preferentially turns over a racemate of RRS-2 and SSR-2 (as opposed to one of RSR-2 and SSR-2).

Does selectivity for trans-1-decalone reduction equal specificity for α-tetralol oxidation? The significant difference between actKR complexes with RS-1 and SR-1 prompts a deeper discussion on the general assumption that selectivity towards trans-1-decalone and α-tetralone are directly comparable. Resolving the stereochemical outcome of the “forwards” reaction with trans-1-decalone (by assessing S-2 : R-2 ratio) was not successful, despite trying several approaches (personal communication with Prof. SC Tsai, UC Irvine). Korman, Javidpour et al.14, 16-17 thus investigate specificity in the reverse reaction for S-3 and R-3 (commercially available, unlike the stereoisomers of 2); studying the “forwards” reduction of α-tetralone is hampered by actKR’s very sluggish turnover of that substrate.

To obtain detailed insight into (the stereoselectivity of) reduction of trans-1-decalone, which is readily turned over by the three enzyme variants studied here (Table 1), we opted to investigate the “forwards” reaction (reduction of 1 to 2). Note that had we chosen to study the reaction using α-tetralone/3 instead, a complication would have arisen because the aliphatic carbonyl ring can chair-flip freely (due to aromaticity in its non-carbonyl ring): chair-flipping—virtually unachievable in 1—would have introduced a considerable degree of complexity in our simulations. We thus rely on the assumption (also made by Korman, Javidpour et al.)14, 16-17 that the reduction of 1 to (RS)-2 remains (through a degree of microscopic reversibility) comparable to the oxidation of (RS)-3 to α-tetralone. (This issue is reminiscent of the “experimental problem” reviewed by Van Gunsteren and colleagues,70 whereby experimental data is sometimes scarcer than desirable when setting up biomolecular simulations). We believe this assumption to be reasonable in this case, because in each of the two ‘chairflip’ conformers of α-tetralone, the carbonyl and three aliphatic carbons are likely to have a near-identical arrangement to atoms C1-C4 (Chart 1) in RS-1 or SR-1 (there would be the same such correspondence between S-3 and SSR-2/SRS-2; and between R-3 and RRS-2/RRS-2). The only major (steric) difference would be the axial atoms in 1 and 2 jutting out above and below the plane of the non-carbonyl ring: as a result, we speculate that the 1/2 pair could be more susceptible to steric hindrance at its non-reactive (non-carbonyl) end, possibly increasing the degree of stereoselectivity towards 1 compared to α-tetralone/3.

Towards an efficient approach for understanding enzymatic stereocontrol. The procedure that we employ in this study relies on state-of-the-art methods in computational chemistry to dissect the influence of FI-FIII—something that is very hard to achieve experimentally. We have deliberately used protocols that (1) require input of the WT structure only; (2) use relatively limited computational resources (short simulations and semiempirical QM treatment); and (3) can be automated. We thus envisage that these protocols can be used in a way that is conceptually similar to the ‘high-throughput–multiple independent MD simulations’ (HTMI-MD) approach used by Wijma and coworkers23, 71-72 to efficiently screen newly suggested stereoselective enzyme variants in silico, prior to assessment by experiment. We note that the HTM-MD approach only assesses FI, whereas we have shown here that, in particular, FIII can be crucial to understand stereoselectivity in certain ketoreductase variants (P94L actKR).

The protocols serve to obtain prediction of and insight into stereoselectivity in novel mutants relatively rapidly, and could thus aid the design of novel mutants. For the use of such tools to become more widespread, and to allow screening a large number of enzyme variants, the computational time and resources required should be modest. In this respect, the fact that our relatively low-cost protocols have been able to characterize the stereoselectivity of WT, P94L and V151L actKR with encouraging matches to experimental data shows promise for future applications to other enzymes and mutants. In this subsection, we discuss the rationale for the protocols to assess FI-FIII and evaluate if computational time can be reduced further.

To study FI (and to later generate trajectories for FII and FIIII) we employed classical MD simulations, previously employed in several examples of interest with other KRs and SDRs.10, 73 In contrast to many previous studies that have used a small number of longer MD simulations to collect their data,10, 24, 73-74 our strategy was to run a large set of independent MD simulations of shorter length (8 × 12 ns, essentially becoming 32 × 12 ns due to the four active sites) to maximize conformational sampling. This length is still significantly greater than those employed by Wijma et al. (20-40 × 10 ps or 10 × 100 ps),23, 72 because this may be necessary to capture relevant enzyme conformational changes. The importance of carrying out multiple independent MD replicas (at least 5-10 as a rule-of-thumb) has been recently restated.75

In the case of ketoreductases/SDRs, the α6-α7 loop can adopt alternative conformations (Figure 1c and d), and was proposed to play a role in determining reactivity and selectivity in a different enzyme.10 We thus confirmed (vide infra) that the 12 ns simulation length is sufficient for “closed” chains (B, D) to sample open α6-α7 loop conformations and vice versa for “open” chains (A, C). We further explored the possibility of investigating FI with simulations shorter than 12 ns, and our data (Supporting Information, Figure S4) show that this is feasible: trends are identical with simulations shorter than 0.3 ns, leaving the conclusions unaltered.

The MM/PBSA option was chosen to study FII in light of its relatively modest computational cost, as well as its good performance in previous ΔGpred-prediction problems across several enzyme-ligand systems.29, 76-77 Alternative options such as absolute binding free energy calculations using alchemical perturbations27-28 would have raised the computational cost significantly.70 Our decision to run MM/PBSA calculations on a large set of independent MD simulations was to maximize conformational sampling, and has been shown to improve performance.70-80 In fact, though in our case we employed 12 ns MD replicas to adequately sample conformational changes in the α6-α7 loop, previous work shows that much shorter MD simulations (in the order of hundreds of ps to a few ns) are typically sufficient.79-80 In our case, tests suggest that using MD trajectories shorter than 4-8 ns may have repercussions on accuracy (Supporting Information, Figure S5); likely in part due to the use of fewer snapshots overall. (The 24 ps-interval at which snapshots are extracted is roughly in line with previous MM/PBSA and MM/GBSA indications, but could be reduced).79-81
To study FIII, we employed a semiempirical QM/MM approach, to avoid specific parameterization of models (e.g. as required for empirical valence bond methods) and limit computational cost (e.g. as compared to QM/MM or a QM cluster approach with DFT methods). To further limit this cost, we chose to use a small QM region with only groups directly electronically involved in the rate-limiting step (Figure 3), rather than a region that incorporates possible further proton transfers (as used in previous QM/MM studies of actKR-related enzymes). A wide range of options is available for calculating reaction barriers with QM/MM. In our case, it was not known a priori what the orientation of the substrate would be in the different variants (due to its small size in comparison to the relatively large active site in actKR). Therefore, an approach using QM/MM optimization alone was not ideal, since both the initial choice of substrate orientation and enzyme conformation may have a large influence on the resulting activation energy barriers, and this would have meant collecting a very large number of starting conformations. We thus opted for QM/MM MD reaction simulations to allow sampling of many orientations, as well as full consideration of entropic effects. Conformational sampling was further enhanced by performing 10 independent reaction simulations (leading to 20 ps sampling per umbrella sampling window), for each of the four possible reactions, across the three actKR variants (120 in total); indeed, our tests with 36 independent simulations (3 per case, 6 ps per window; Supporting Information, Figure S6) show that results become significantly less reliable (with larger errors).

Notably, QM/MM simulations used a range of starting structures from the many independent classical MD simulations performed: these capture a range of α6-α7 loop conformations (between ‘open’ and ‘closed’). As mentioned earlier, in contrast to previous work on Lactobacillus kefir short-chain alcohol dehydrogenase, we found little correlation between the loop openness (as measured by the average of three interatomic loop-enzyme distances) and barrier height (Supporting Information, Figures S7 and S8), indicating that in this case, the α6-α7 loop conformation has little or no influence on actKR’s catalytic reactivity when turning over small substrates.

4. Conclusions

Biocatalyst (stereo)selectivity can arise due to a combination of effects, including substrate binding (and its affinity in different orientations), as well as reaction barrier. Such effects, and their interplay, can be subtle, especially when dealing with a relatively open active site (as in the case of actKR, the enzyme studied here). One route towards stereoselective enzymes (for use in biocatalysis) can thus be to reduce their active site volume (e.g. in such a way that substrates can only bind in specific orientations). However, such changes may also reduce their capability for turning over a range of substrates (of different sizes) as well as reduce efficiency.

Here, we have presented a detailed computational study of three factors that may affect stereoccontrol in the wild-type ketoreductase actKR and in two key variants with similar efficiency (P94L, V151L). We separately investigated the formation (FI), the binding free energy (FII) and the reactivity (free energy barrier, FIII) of reactive complexes, using classical MD simulations, MM/PBSA calculations and QM/MM MD reaction simulations, respectively. We demonstrate that the strict stereoselectivity observed in the P94L and V151L actKR variants (both leading to only a small decrease in active site volume) is arising through different mechanisms in each case. For the 5-selective P94L actKR variant, stereoselectivity is driven by the relative difference between activation free energy barriers (FIII), whereas for the R-selective V151L actKR variant, the formation and binding affinity of reactive substrate poses (FI, FII) are mainly responsible.

The observation that the main effect (or factor) determining stereoccontrol can differ from mutant to mutant has important implications for obtaining stereoselective enzymes through redesign: either substrate orientation or efficiency of catalysis (via transition state stabilization), or both, may be either altered to confer stereoselectivity.

Our simulations further indicate that each enantiomer of the substrate trans-1-decalone (1) has a preference for particular reactive poses: pro-5 poses are predominantly found for SR-1, and pro-R poses for RS-1. We note that this was observed in all three actKR variants studied, including the mildly 5-selective WT actKR. For the stereoselective reduction of 1, it is thus the combined effects of the actKR variant and substrate enantiomer that determine the final product, with P94L favoring formation of SSR-2 and V151L favoring formation of RSS-2.

The complexity of our findings is representative of the multifaceted origins of catalysis and selectivity in enzyme biocatalysts, which arises from a subtle interplay of steric, dynamic, and electronic effects. Detailed biomolecular simulation provides valuable means to break down the causes of stereoccontrol (e.g. in terms of the individual factors FIII). The computational procedures we have used here to determine the origins of stereoccontrol in actKR are deliberately generically applicable and employ limited computer resources. They can thus be used for similar evaluations of different enzyme biocatalysts in a time- and cost-effective manner (especially in combination with further automation). Indeed, we believe that an accurate, yet computationally efficient assessment of the different effects (or factors) dictating an enzyme’s stereoccontrol is highly valuable for the (re)design of biocatalysts with enhanced stereoselectivity or -specificity, as well as for in-depth understanding of selectivity in existing enzyme variants.

ASSOCIATED CONTENT

Supporting Information. Details on the construction of starting structures, force-field parameters and MD simulations (restraints and equilibration procedure); additional details for QM/MM MD umbrella sampling and QM/MM benchmarking; error analysis; Gaussian09 *.log files for all optimized stationary points obtained from benchmarking (available at https://doi.org/10.19061/sochem-bd-6-12); starting structures and forcefield topologies for the six actKR-(NADPH)1L systems (* .zip folder). This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author
* E-mail: Marc.VanderKamp@bristol.ac.uk

Orcid
Marc W. van der Kamp: 0000-0002-8060-3359
Stefano A. Serapian: 0000-0003-0122-8499
Author Contributions
Both authors designed the experiments, SAS performed experiments, MWK devised the project. Both authors analyzed results, wrote the manuscript and have given approval to the final version of the manuscript.

Funding Sources
MWK is a BBSRC David Phillips Fellow, and the authors wish to thank BBSRC for funding (grant BB/M026280/1).

ACKNOWLEDGEMENT
The authors thank Prof. Shioi-Chan Tsai (Sheryl) and Gabriel O. Moreno (UC Irvine), as well as Prof. Matt Crump (Bristol), for helpful discussions. This work was conducted using the computational facilities of the Advanced Computing Research Centre of the University of Bristol.

REFERENCES
1. Hertweck, C.; Luzhetskyy, A.; Rebets, Y.; Becthold, A., Type II polyketide synthases: gaining a deeper insight into enzymatic teamwork. Nat. Prod. Rep. 2007, 24 (1), 162–190.
2. Zhang, Z.; Pan, H. X.; Tang, G. L., New insights into bacterial type II polyketide biosynthesis. F1000Research 2017, 6, 172.
3. Dui, D.; Katsuyama, Y.; Shin-ya, K.; Ohnishi, Y., Reconstitution of a Type II Polyketide Synthase that Catalyzes Polyene Formation. Angew. Chem. Int. Ed. 2018, 57 (7), 1954–1957.
4. Javidpour, P.; Das, A.; Khosla, C.; Tsai, S.-C., Structural and Biochemical Studies of the Hedamycin Type II Polyketide Ketoreductase (HedKRR). Biochim. Biophys. Acta 2011, 50 (34), 7426–7439.
5. Korman, T. P.; Hill, J. A.; Vu, T. N.; Tsai, S.-C., Structural Analysis of Actorinohodin Polyketide Ketoreductase: Cofactor Binding and Substrate Specificity. Biochemistry 2004, 43 (46), 14529–14538.
6. Tang, Y.; Lee, H. Y.; Tang, Y.; Kim, C.-Y.; Mathews, I.; Khosla, C., Structural and Functional Studies on SCO1815: A β-Ketoacyl-Acyl Carrier Protein Reductase from Streptomyces coelicolor A3(2). Biochemistry 2006, 45 (47), 14085–14093.
7. Valentic, T. R.; Jackson, D. R.; Brady, S. F.; Tsai, S.-C., Comprehensive Analysis of a Novel Ketoreductase for Pentangular Polyphenol Biosynthesis. ACS Chem. Biol. 2016, 11 (12), 3421–3430.
8. Moore, J. C.; Pollard, D. J.; Kosjek, B.; Devine, P. N., Advances in the Enzymatic Reduction of Ketones. Acc. Chem. Res. 2007, 40 (12), 1412–1419.
9. Huisman, G. W.; Liang, J.; Kreebber, A., Practical chiral alcohol manufacturing using ketoreductases. Curr. Opin. Chem. Biol. 2010, 14 (2), 122–129.
10. Noey, E. L.; Tiberwal, N.; Jiménez-Osés, G.; Osuna, S.; Park, J.; Bond, C. M.; Cascio, D.; Liang, J.; Zhang, X.; Huisman, G. W.; Tang, Y.; Houk, K. N., Origins of stereoselectivity in evolved ketoreductases. Proc. Natl. Acad. Sci. U. S. A. 2015, 112 (51), E7065–E7072.
11. Bornscheuer, U. T.; Huisman, G. W.; Kazlauskas, R. J.; Lutz, S.; Moore, J. C.; Robins, K., Engineering the third wave of biocatalysis. Nature 2012, 485, 185.
12. Bornscheuer, U. T., The fourth wave of biocatalysis is approaching. Phil. Trans. R. Soc. A 2018, 376 (2110), 20170063.
13. Hadfield, A. T.; Limpkin, C.; Teartasin, W.; Simpson, T. J.; Crosby, J.; Crump, M. P., The Crystal Structure of the activH Actinohodin Polyketide Reductase: Proposed Mechanism for ACP and Polyketide Binding. Struct. 2004, 12 (10), 1865–1875.
14. Korman, T. P.; Tan, Y.-H.; Wong, J.; Luo, R.; Tsai, S.-C., Inhibition Kinetics and Emdinococystal Structure of a Type II Polyketide Ketoreductase. Biochem. 2008, 47 (7), 1837–1847.
15. Medina, F. E.; Neves, R. P. P.; Ramos, M. J.; Fernandes, P. A., A QM/MM study of the reaction mechanism of human [small beta]-ketoacyl reductase. Phys. Chem. Chem. Phys. 2017, 19 (1), 347–355.
16. Javidpour, P.; Korman, T. P.; Shakya, G.; Tsai, S.-C., Structural and Biochemical Analyses of Regio- and Stereospecificities Observed in a Type II Polyketide Ketoreductase. Biochim. 2011, 50 (21), 4638–4649.
17. Javidpour, P.; Bruegger, J.; Srithanan, S.; Korman, T. P.; Crump, Matthew P.; Crosby, J.; Burkart, Michael D.; Tsai, S.-C., The Determinants of Activity and Specificity in Actinorhodin Type II Polyketide Ketoreductase. Chem. Biol. 2013, 20 (10), 1225–1234.
18. Byrne, M. J.; Lees, N. R.; Han, L.-C.; van der Kamp, M. W.; Mulholland, A. J.; Stach, J. E. M.; Willis, C. L.; Race, P. R., The Catalytic Mechanism of a Natural Diels–Alderase Revealed in Molecular Detail. J. Am. Chem. Soc. 2016, 138 (19), 6095–6098.
19. Daniels, A. D.; Campeotto, L.; van der Kamp, M. W.; Bolt, A. H.; Trinh, C. H.; Phillips, S. E. V.; Pearson, A. R.; Nelson, A.; Mulholland, A. J.; Berry, A., Reaction Mechanism of N-Acetylneuraminic Acid Lyase Revealed by a Combination of Crystallography, QM/MM Simulation, and Mutagenesis. ACS Chem. Biol. 2014, 9 (4), 1025–1032.
20. Leferink, N. G. H.; Ranaghan, K. E.; Karuppiah, V.; Currin, A.; van der Kamp, M. W.; Mulholland, A. J.; Scrutton, N. S., Experiment and Simulation Reveal How Mutations in Functional Plasticity Regions Guide Plant Monoterpenes Synthase Product Outcome. ACS Catal. 2018, 8 (5), 3780–3791.
21. Amrein, B. A.; Steffen-Munsberg, F.; Szelé, I.; Purg, M.; Kulkarni, Y.; Kamerlin, S. C. L., CADEE: Computer-Aided Directed Evolution of Enzymes. IURC 2017, 4 (1), 50–64.
22. Hediger, M. R.; De Vico, L.; Rames, J. B.; Jäckel, C.; Besenmatters, W.; Svendsen, A.; Jensen, J. H., In silico screening of 393 mutants facilitates enzyme engineering of amide hydroxyl activity in CalB. PeerJ 2013, 1, e145.
23. Wijma, H. J.; Floor, R. J.; Bjelic, S.; Marrink, S. J.; Baker, D.; Janssen, D. B., Enantioselective Enzymes by Computational Design and In Silico Screening. Angew. Chem. 2015, 127 (12), 3797–3801.
24. Rinaldi, S.; van der Kamp, M. W.; Ranaghan, K. E.; Mulholland, A. J.; Colombo, G., Understanding Complex Mechanisms of Enzyme Reactivity: The Case of Limonene-1,2-Epoxide Hydrolases. ACS Catal. 2018, 8 (7), 5698–5707.
25. Janfalk Carlsson, Å.; Bauer, P.; Dobritszch, D.; Nilsson, M.; Kamerlin, S. C. L.; Widersten, M., Laboratory-Evolved Enzymes Provide Snapshots of the Development of Enantioconvergence in Enzyme-Catalyzed Epoxide Hydrolysis. ChemBioChem 2016, 17 (18), 1693–1697.
26. Sun, Z.; Lonsdale, R.; Kong, X.-D.; Xu, J.-H.; Zhou, J.; Reddy, M. T., Reshaping the Enzyme Binding Pocket for Enhanced and Inverted Stereoselectivity: Use of Smallest Amino Acid Alphabets in Directed Evolution. Angew. Chem. Int. Ed. 2015, 54 (42), 12410–12415.
27. Woods, C. J.; Malaisere, M.; Hannonbua, S.; Mulholland, A. J., A water-swap reaction coordinate for the calculation of absolute protein–ligand binding free energies. J. Chem. Phys. 2011, 134 (5), 054114.
28. Cabeza de Vaca, I.; Qian, Y.; Vilseck, J. Z.; Tirado-Rives, J.; Jorgensen, W. L., Enhanced Monte Carlo Methods for Modeling Proteins Including Computation of Absolute Free Energies of Binding. J. Chem. Theory Comput. 2018.
29. Genhedhen, S.; Ryde, U., The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. Expert Opin. Drug Discovery 2015, 10 (5), 449–461.
30. Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E., Calculating Structures and Free Energies of Complex Molecules: Combining Molecular Mechanics and Continuum Models. Acc. Chem. Res. 2000, 33 (12), 889–897.
31. Rastelli, G.; Del Rio, A.; Degliesposti, G.; Sgobbo, M., Fast and accurate predictions of binding free energies using MM-PBSA and MM-GBSA. J. Comput. Chem. 2010, 31 (4), 797–810.
32. Wang, C.; Nguyen, P. H.; Pham, K.; Huynh, D.; Le, T.-B.; Wang, H.; Ren, P.; Luo, R., Calculating protein–ligand binding
71. Li, R.; Wijma, H. J.; Song, L.; Cui, Y.; Otzen, M.; Tian, Y.; Du, J.; Li, T.; Niu, D.; Chen, Y.; Feng, J.; Han, J.; Chen, H.; Tao, Y.; Janssen, D. B.; Wu, B., Computational redesign of enzymes for regio- and enantioselective hydroamination. Nat. Chem. Biol. 2018, 14 (7), 664-670.
72. Wijma, H. J.; Marrink, S. J.; Janssen, D. B., Computationally Efficient and Accurate Enantioselectivity Modeling by Clusters of Molecular Dynamics Simulations. J. Chem. Inf. Model. 2014, 54 (7), 2079-2092.
73. Mugnai, M. L.; Shi, Y.; Keatinge-Clay, A. T.; Elber, R., Molecular Dynamics Studies of Modular Polyketide Synthase Ketoreductase Stereospecificity. Biochem. 2015, 54 (14), 2346-2359.
74. Sun, Z.; Wu, L.; Boccola, M.; Chan, H. C. S.; Lonsdale, R.; Kong, X.-D.; Yuan, S.; Zhou, J.; Reetz, M. T., Structural and Computational Insight into the Catalytic Mechanism of Limonene Epoxide Hydrolase Mutants in Stereoselective Transformations. J. Am. Chem. Soc. 2018, 140 (1), 310-318.
75. Knapp, B.; Ospina, L.; Deane, C. M., Avoiding false positive conclusions in molecular simulation: the importance of replicas. J. Chem. Theory Comput. 2018.
76. Kuhn, B.; Kollman, P. A., Binding of a Diverse Set of Ligands to Avidin and Streptavidin: An Accurate Quantitative Prediction of Their Relative Affinities by a Combination of Molecular Mechanics and Continuum Solvent Models. J. Med. Chem. 2000, 43 (20), 3786-3791.
77. Wang, J.; Morin, P.; Wang, W.; Kollman, P. A., Use of MM-PBSA in Reproducing the Binding Free Energies to HIV-1 RT of TIBO Derivatives and Predicting the Binding Mode to HIV-1 RT of Efavirenz by Docking and MM-PBSA. J. Am. Chem. Soc. 2001, 123 (22), 5221-5230.
78. Aldeghi, M.; Bodkin, M. J.; Knapp, S.; Biggin, P. C., Statistical Analysis on the Performance of Molecular Mechanics Poisson–Boltzmann Surface Area versus Absolute Binding Free Energy Calculations: Bromodomains as a Case Study. J. Chem. Inf. Model. 2017, 57 (9), 2203-2221.
79. Adler, M.; Beroza, P., Improved Ligand Binding Energies Derived from Molecular Dynamics: Replicate Sampling Enhances the Search of Conformational Space. J. Chem. Inf. Model. 2013, 53 (8), 2065-2072.
80. Sadiq, S. K.; Wright, D. W.; Kenway, O. A.; Coveney, P. V., Accurate Ensemble Molecular Dynamics Binding Free Energy Ranking of Multidrug-Resistant HIV-1 Proteases. J. Chem. Inf. Model. 2010, 50 (5), 890-905.
81. Genheden, S.; Ryde, U., How to obtain statistically converged MM/GBSA results. J. Comput. Chem. 2010, 31 (4), 837-846.
82. Ryde, U., How Many Conformations Need To Be Sampled To Obtain Converged QM/MM Energies? The Curse of Exponential Averaging. J. Chem. Theory Comput. 2017, 13 (11), 5745-5752.
SUPPORTING INFORMATION

Unpicking the cause of stereoselectivity in actinorhodin ketoreductase variants with atomistic simulations

Stefano A. Serapian† and Marc W. van der Kamp*,†,‡

†School of Biochemistry, University of Bristol, University Walk, Bristol, BS8 1TD, United Kingdom
‡Centre for Computational Chemistry, University of Bristol, Cantock’s Close, Bristol, BS8 1TS, United Kingdom

Contents

1. MD Restraints and criteria for ‘reactive poses’ .................................................................................. 2
2. Construction of Starting Structures ................................................................................................. 2
   2.1 Construction of WT-SR-14 .......................................................................................................... 2
   2.2 Construction of V151L-SR-14 and P94L-SR-14, and resolution of steric clashes ................... 3
   2.3 Construction of WT-RS-14, V151L-RS-14, and P94L-RS-14 ..................................................... 3
3. Forcefield parameters for enantiomers of 1 ....................................................................................... 3
4. MD Simulation Details for Pre-Production Stages ............................................................................ 4
   4.1 Minimization ............................................................................................................................... 4
   4.2 Solvent equilibration .................................................................................................................. 4
   4.3 Heating ....................................................................................................................................... 4
   4.4 Equilibration .............................................................................................................................. 4
5. QM/MM Benchmarking ...................................................................................................................... 4
   5.1 QM/MM free energy profiles from semiempirical QM/MM MD US ......................................... 4
   5.2 Full QM potential energy surfaces (Semiempirical, DFT, MP2, and SCS-MP2) ....................... 5
   5.3 Benchmarking results ................................................................................................................ 6
6. Additional Details for QM/MM MD US Runs .................................................................................... 7
   6.1 Choice of the original 120 snapshots .......................................................................................... 7
   6.2 Outliers and repeats .................................................................................................................. 8
   6.3 WHAM analysis and error calculation (Monte-Carlo Bootstrap procedure) ........................... 8
7. Error Analysis for FI Statistics ........................................................................................................... 9
8. Typical CPU / GPU Times vs. Accuracy ........................................................................................... 9
9. α6-α7 Loop Opening and Barrier Height ......................................................................................... 12
10. References ....................................................................................................................................... 14
1. MD Restraints and criteria for ‘reactive poses’

Investigations on the role of FII and FIII used “restrained” classical MD simulations as their starting point (MM/PBSA and QM/MM MD, respectively; see main text). The purpose of the restraints is to keep the substrate in a “reactive” position with a specific prochiral orientation. Details of (one-sided) harmonic restraints are listed in Table S1.

Table S1. Restraints imposed in “restrained” classical MD simulations to investigate FII and FIII.

| Type   | Atoms Affected                                        | k   | Applied       |
|--------|-------------------------------------------------------|-----|---------------|
| Distance | 1:O, Ser144:O(H)                                         | 50^a| >3.75 Å       |
| Distance | 1:O, Tyr157:O(H)                                         | 50^a| >4.40 Å       |
| Distance | 1:C15, NADPH:H^-                                     | 50^a| >3.75 Å       |
| Dihedral | 1:O, 1:C15, 1:C8a, NADPH:H^-                        | 25^d| <190°; >350°  |
| Dihedral | 1:O, 1:C15, 1:C8a, NADPH:H^-                        | 25^d| <10°; >170°   |

^a kcal mol^-1 Å^-2. ^b For numbering and nomenclature of C atoms in 1 cf. Chart 1 in main text. ^c Only in monomers A and B (SR-1 or RS-1 kept in pro-R orientation; vide infra). ^d kcal mol^-1 rad^-2. ^e Only in monomers C and D (SR-1 or RS-1 kept in pro-¥ orientation; vide infra).

One-sided harmonic distance restraints between 1’s oxygen atom and hydroxyl oxygens of Ser144 and Tyr157 (adding an energy penalty when farther than 3.75 Å and 4.40 Å apart, respectively) are chosen by adding 0.2 Å to the mean values of these distances in the docked starting structures (based on the emodin inhibitor co-crystallized in PDB ID 2RH4, vide infra). The range of the 1:C1, NADPH:H^- restraint (adding an energy penalty when atoms are farther than 3.75 Å) is chosen so that atoms are free to cover the first peak in the radial distribution function g(1:C1, NADPH:H^-) obtained from “free” MD simulations. The single restraint imposed on the dihedral 1:O, 1:C1, 1:C8a, NADPH:H^- serves the purpose of enforcing a particular prochirality. In monomers A and B, the restraint ensures that substrate molecules preserve a pro-R orientation (180°−360°). In monomers C and D the restraint ensures that substrate molecules preserve a pro-¥ orientation (0°−180°).

When analysing our “free” MD simulations for reactive conformations (to examine the role of FII), the thresholds of the three distance restraints in Table S1 also serve as the basis to determine whether or not a particular pose of 1 can be considered “reactive” (if one of the three thresholds is violated, the pose is considered “unreactive”). Similarly, the value of the 1:O, 1:C1, 1:C8a, NADPH:H^- dihedral is used as the measure of prochirality in “free” MDs: pro-¥ (if it is in the range 0°−180°) or pro-R (if it is in the range 180°−360°).

2. Construction of Starting Structures

All six actKR-(NADPH)-1, starting structures (see main text for nomenclature) are constructed using the Biovia Discovery Studio modeling software (version 3.5), starting from our reference WT crystal structure (PDB ID: 2RH4). AMBER files (*.rst, *.top, and *.pdb) are available electronically in a *.zip directory alongside the present document. Steps taken are described in the following subsections.

2.1 Construction of WT-SR-1

One molecule of SR-1 was manually docked in each of the four active sites, in either pro-¥ or pro-R orientation with respect to H^- attack, resulting in the four combinations outlined in Table S2. Docking was guided by placing the SR-1 carbonyl and carbonyl-bearing aliphatic ring in the position of the carbonyl and central aromatic ring of the co-crystallized inhibitor emodin (present in monomers A and C). In monomers B and D, SR-1 is placed in the same positions as in monomers A and C, respectively.
Table S2 Criteria for setting up starting structures of actKR and 1 for classical MD simulations

| Active Site | Orientation of 1 | α6-α7 Loop |
|-------------|-----------------|------------|
| Monomer A   | Pro-R           | Open       |
| Monomer B   | Pro-R           | Closed     |
| Monomer C   | Pro-Ś           | Open       |
| Monomer D   | Pro-Ś           | Closed     |

*a* Applies to WT, V151L, and P94L actKR. *b* With respect to hydride attack. *c* Either RS-1 or SR-1. *d* cf. Figures 1c and 1d in main text.

Monomer lengths in the created WT-SR-1<sub>4</sub> structure were standardized so that they all began at Met1 and terminated at Tyr261 (*cf.* main text): extra residues were eliminated from monomers B and D and the Met1-Asp5 fragment was added to monomers A and C with the same conformation and orientation as in monomer B.

### 2.2 Construction of V151L-SR-1<sub>4</sub> and P94L-SR-1<sub>4</sub>, and resolution of steric clashes

Point mutations were applied to the obtained WT-SR-1<sub>4</sub> structure to create initial structures for V151L-SR-1<sub>4</sub> and P94L-SR-1<sub>4</sub>, in line with the crystal structures of V151L (PDB ID: 4DBZ)<sup>2</sup> and P94L (PDB IDs: 3RI3, 3QRW).<sup>3</sup> The resulting moderate steric clashes are resolved with two successive minimizations (300 steps each, using AMBER’s default combination of the steepest descent and conjugate gradient algorithms): the first minimization relaxes hydrogen atoms (all other atoms positionally restrained with a 5 kcal mol<sup>−1</sup> Å<sup>−2</sup> force constant); the second minimization relaxes all protein sidechains (positional restraints on heavy atoms in the backbone, cofactor, and SR-1).

### 2.3 Construction of WT-RS-1<sub>4</sub>, V151L-RS-1<sub>4</sub>, and P94L-RS-1<sub>4</sub>

To obtain structures in complex with RS-1, SR-1 was replaced with its enantiomer in each of the three structures previously generated (*i.e.* WT actKR and 2 minimized mutants), with pro-R and pro-Ś orientations preserved in each monomer. The two carbonyl atoms (C1) and the adjacent carbon atoms C2 and C8<sub>a</sub> (main text; Chart 1) are superimposed exactly. We thus end up with six distinct starting structures of actKR-(NADPH)<sub>4</sub>-1<sub>4</sub> to be used in our MD simulations.

### 3. Forcefield parameters for enantiomers of 1

Parameters for SR-1 and RS-1 in line with the protein force-field were obtained using the AmberTools antechamber utility:<sup>4</sup> this calculates point charges using the semi-empirical AM1-bond charge correction scheme (AM1-BCC) method,<sup>5,6</sup> and types atoms (Lennard-Jones and bonded parameters) according to the General Amber Force Field (GAFF; *cf.* main text).<sup>7</sup> Missing parameters for a single improper dihedral in both SR-1 and RS-1 were assigned by parmchk2; relevant *.prepc and *.frcmod files are also included in the supplementary *.zip folder provided.
4. MD Simulation Details for Pre-Production Stages

In this section, we discuss parameters for the pre-production stages of our 48 “free” and 48 “restrained” MD runs in greater detail. As discussed in the main text, these are: minimization (600 steps); solvent equilibration (9 ps, NVT ensemble); heating (20 ps, NVT); and equilibration (2040 ps, NpT).

4.1 Minimization

Minimization is carried out in two stages, of 300 steps each, using AMBER’s default algorithm. First, solvent and all hydrogens are minimized (positional restraints on all solute heavy atoms with a 5 kcal mol⁻¹ Å⁻² force constant). Second, minimization is performed without positional restraints (with only the usual restraints in place for the “restrained” MDs).

4.2 Solvent equilibration

Solvent is equilibrated with 9 ps of simulated annealing in the NVT ensemble (time-step, Δt, at 1 fs), with all other atoms position-restrained (k = 10 kcal mol⁻¹ Å⁻²). Starting at 25 K, the solvent is heated to 400 K over 3 ps; equilibrated at 400 K for a further 3 ps; and cooled back to 25 K in the final 3 ps. Only in this case, constant temperatures are enforced via the Berendsen thermostat; temperature coupling τT is tight (0.2 ps) for the first 6 ps, but is significantly relaxed for cooling (1.0 ps to 2.0 ps over the last 3 ps).

4.3 Heating

From this stage onwards, Δt is 2 fs, the SETTLE and SHAKE algorithms are used to constrain all bonds containing hydrogen, and temperature is controlled with the Langevin thermostat. The collision frequency is set at 0.75 ps⁻¹ for heating. Systems are heated from 25 K to 298 K (NVT) over 20 ps, with harmonic positional restraints on backbone Cα atoms (k = 5 kcal mol⁻¹ Å⁻²), alongside the usual restraint(s) for “free” and “restrained” MDs.

4.4 Equilibration

Switching to the NpT ensemble, equilibration is then started at 298 K, 1 atm, over 2040 ps, as follows:

- for the first 20 ps, position restraints on backbone Cαs are relaxed to k = 3.75 kcal mol⁻¹ Å⁻²;
- for the next 20 ps, they are even further relaxed to k = 1.75 kcal mol⁻¹ Å⁻²; and
- for the final 2000 ps (2 ns) they are lifted completely, with only the usual restraint(s) remaining in place for “free” and “restrained” MDs.

- Pressure control with the Berendsen barostat.
- Temperature control with the Langevin thermostat (collision frequency 1 ps⁻¹).

5. QM/MM Benchmarking

To assess the performance of semiempirical methods that could be employed in QM/MM calculations used to study FIII, we performed a series of preliminary semiempirical QM/MM MD US simulations, and compare the resulting free energy profiles to the potential energy surfaces mapped at higher QM levels of theory. In addition, we assess how well each QM/MM method is able to reproduce the reduction mechanism in a small model system.

5.1 QM/MM free energy profiles from semiempirical QM/MM MD US

Preliminary QM/MM MD US simulations are run using one snapshot from one of the “restrained” MDs of WT-SR-1, forcing the pro-R RS-1 in monomer A to react to RS-R-2. Conditions and procedures are identical to the QM/MM MD US runs described in the main text, as are the procedures used to extract the PMF (i.e. the free energy profile) with WHAM. Three semiempirical methods (implemented in AMBER) were chosen for testing: PM6; DFTB2; and DFTB3.
5.2 Full QM potential energy surfaces (Semiempirical, DFT, MP2, and SCS-MP2)

Full QM potential energy surfaces (PESs) are mapped using the Gaussian09 package (Revision D01).

Figure S1. Examples of the simplified active site model used for QM benchmarking calculations, both initially extracted from monomer A of WT-SR-14, and both featuring the transition state (TS) between SR-1 and RSR-2. (a) TS located by B3LYP/6-31+G(d,p) calculations. (b) TS located by PM6 calculations. Incipient proton abstraction from Tyr157 is visible in both cases.

Calculations are run on two highly simplified models of the actKR active site: one extracted from monomer A of the starting structure of WT-SR-14 (see Figure S1) and the other from monomer A of the starting structure of WT-RS-14. Only three items are retained in each of these models: (1) SR-1 or RS-1, both in pro-R orientation; (2) the sidechain of the catalytic Tyr157 up to the Cβ, which is hydrogen-capped to give para-cresol; and (3) part of the cofactor, as nicotinamide riboside.

The PES of the reduction of 1 to 2 in our simplified models is mapped with several methods. For the SR-1 model’s conversion to RSR-2, these are: the semiempirical methods to be benchmarked (PM6, DFTB2, and DFTB3); the density functional theory (DFT) method B3LYP/6-31+G(d,p). In addition, at every geometry point on the B3LYP/6-31+G(d,p) PES, we also perform single-point corrections using spin-component-scaled 2nd-order Møller-Plesset perturbation theory (SCS-MP2). For RS-1 conversion to RRS-2, the PES is only mapped with PM6 and B3LYP/6-31+G(d,p). All DFTB2 and DFTB3 calculations were performed in the gas phase, and PM6 and B3LYP/6-31+G(d,p) in implicit water (using the polarized continuum solvation model, PCM).

PES mapping proceeds as follows, with the nature of minima and transition states (TSs) always confirmed through frequency calculations (vide infra). First, the SR-1 or RS-1 model is optimized with each of the chosen methods, using default settings. Next, product models for RSR-2 or RRS-2 are “guessed” from the corresponding optimized reactant model, and subsequently optimized (again with each of the chosen methods). Thereafter, the TS between the optimized models of 1 and 2 is located via a Synchronous Transit-Guided Quasi-Newton (STQN) calculation, again by optimizing a “guessed” structure. Starting from the located TS, an intrinsic reaction co-ordinate (IRC) calculation is then run to map the same extent of PES on either side, either over 80 points with a 0.05 Bohr step size, or over 40 points with a 0.1 Bohr step size, depending on the method: this is enough to span a similar region to the one covered by QM/MM MD US. Once the calculation is complete, Gaussian09’s IRC is converted to our geometric reaction coordinate [(x – y); see Figure 3 in main text], using printed NADPH:C–NADPH:H– and NADPH:H+–1:C1 distances at every step. Finally, the two structures at the “reactant” and “product” extremities of the IRC calculations are both optimized to minima, to check that they plausibly return to the minima from which the TS was initially located: it is from these optimized structures that we...
calculate all of our PES energetics, and normalize them for comparability (with energy of the “reactants” set at zero).

As additional Supporting Information (https://doi.org/10.19061/iochem-bd-6-12), we have included Gaussian09 *.log files for optimized stationary points in all of our semiempirical and B3LYP/6-31+G(d,p) PES maps. These include:

- optimized reactants (SR-1 or RS-1) from one extremity of the IRC calculation;
- optimized products (RSR-2 or RRS-2) from the other extremity of the IRC calculation; and
- optimized transition states from the STQN calculations for each PES.

All of these *.log files contain a frequency calculation on their optimized structure to confirm their nature as minima or transition states. Most files also record the full optimization; a minority only contain frequency calculations on an already-optimized structure.

5.3 Benchmarking results

Free energy profiles from the QM/MM MD US simulations, and full QM PESs of our simplified models of the actKR active site (SR-1 and RS-1), are compared in Figure S2.

![Figure S2. Benchmarking results. (a) Free energy profiles resulting from QM/MM MD US simulations (PM6 - blue; DFTB2 - magenta; DFTB3 - maroon), of the reduction of SR-1 to RSR-2 in monomer A of WT actKR. (b) PESs for (left) the conversion of SR-1 to RSR-2, and (right) the reduction of RS-1 to RRS-2 in a highly simplified actKR model (see text). In the left panel of (b), PESs mapped with the three semiempirical methods tested in (a) are compared to the full QM methods B3LYP/6-31+G(d,p) (green); MP2/6-31+G(d,p)//B3LYP/6-31+G(d,p) (orange); and SCS-MP2/6-31+G(d,p)//B3LYP/6-31+G(d,p) (light brown). In the right panel of (b), we compare the PM6 and B3LYP/6-31+G(d,p) PESs only. Scales are identical to Figure 6 in the main text. From the comparison of semi-empirical methods to the high-level benchmarks (B3LYP/6-31+G(d,p), MP2/6-31+G(d,p), and SCS-MP2/6-31+G(d,p)) it is clear that DFTB2 and DFTB3 fare well with respect to the barrier height. Yet, if we examine the geometries of the simplified RSR-2 models on the “product” side of the PESs, it is evident that neither DFTB2 nor DFTB3 are able to capture the (asynchronous) concerted proton abstraction from Tyr157 (Figure S3) that is known to occur in the enzyme mechanism, and that is instead correctly captured by B3LYP/6-31+G(d,p) (cf. TS in Figure S1a) as well as PM6; in the model TS shown in Figure S1b, incipient proton abstraction from Tyr157 can be clearly seen.}
Figure S3. Model products ($\text{R}_{-2}$), as found in the PES mapped with (a) DFTB2; and (b) DFTB3. Proton abstraction from Tyr157 is entirely absent in both cases.

The poorer ability of DFTB2 and DFTB3 to capture asynchronous proton abstraction from Tyr157 is further confirmed in the QM/MM MD US simulations shown in Figure S2 (including on additional snapshots; not shown here). Conversely, the vast majority of our PM6 QM/MM MD US simulations does correctly capture proton abstraction. Thus, in light of the poor qualitative performance of DFTB2 and DFTB3, PM6 was chosen to model the enzyme reaction in QM/MM MD US simulations (despite the less satisfactory performance for the barrier height).

As a final remark, it is interesting to note from Figure S2b that our simplified PES maps do not capture any appreciable energetic difference between the reduction of pro-$\text{R}$ SR-1 and that of pro-$\text{R}$ RS-1: the PM6 and B3LYP/6-31+G(d,p) barriers for the two processes are both predicted to be very close.

6. Additional Details for QM/MM MD US Runs
6.1 Choice of the original 120 snapshots
The 120 snapshots initially selected for our QM/MM MD US simulations are isolated from “restrained” MDs (see main text). The following criteria are always respected.

- To sample the reduction of pro-$\text{R}$ SR-1 to $\text{R}_{-2}$ and of pro-$\text{R}$ RS-1 to $\text{RS}_{-2}$ in WT, P94L, and V151L actKR:
  - In 5 snapshots, we always force to react the instance of 1 in monomer A.
  - In the other 5, we always force to react the instance of 1 in monomer B.

- Conversely, to sample the reduction of pro-$\text{S}$ SR-1 to $\text{SS}_{-2}$ and of pro-$\text{S}$ RS-1 to $\text{SRS}_{-2}$ in WT, P94L, and V151L actKR:
  - In 5 snapshots, we always force to react the instance of 1 in monomer C.
  - In the other 5, we always force to react the instance of 1 in monomer D.

We choose snapshots so that they satisfy the “reactive” distance criteria, as listed in Table S1. In addition, we make sure that chosen snapshots are uncorrelated by either picking them from different independent MD runs, or picking them to be at least 100 ps apart.
6.2 Outliers and repeats

The small proportion of outliers arising in the 120 original QM/MM MD US simulations are tabulated in Table S3, for each actKR mutant and product chirality. Outliers can be subdivided in three categories:

- Ones in which the substrate flips during the course of the US and generates 2 with opposite chirality to the expected one.
- Ones in which the proton fails to be abstracted from Tyr157, and the resulting trans-1-decanoate molecule attacks the cofactor, or abstracts the proton from Ser144.
- Ones in which proton abstraction occurs very “late” in comparison to the other snapshots and the QM benchmark (reaction coordinate value > 0.35 Å).

Table S3. Summary of outlier and valid QM/MM MD US simulations

|        | Flipped chirality | Wrong / no proton abstracted | Proton abstracted “late” | Valid | Total sampled | Success rate |
|--------|-------------------|------------------------------|--------------------------|-------|---------------|--------------|
| P94L pro-R SR-1 | 2                 | 3                            | 0                        | 10    | 15            | 67%          |
| P94L pro-S SR-1 | 0                 | 0                            | 0                        | 0     | 10            | 100%         |
| V151L pro-R SR-1 | 2                 | 9                            | 2                        | 10    | 23            | 43%          |
| V151L pro-S SR-1 | 0                 | 1                            | 1                        | 10    | 12            | 83%          |
| WT pro-R SR-1    | 2                 | 9                            | 0                        | 10    | 21            | 48%          |
| WT pro-S SR-1    | 0                 | 1                            | 0                        | 10    | 11            | 91%          |
| P94L pro-R RS-1  | 0                 | 3                            | 0                        | 10    | 13            | 77%          |
| P94L pro-S RS-1  | 2                 | 4                            | 0                        | 10    | 16            | 63%          |
| V151L pro-R RS-1 | 0                 | 1                            | 2                        | 10    | 13            | 77%          |
| V151L pro-S RS-1 | 0                 | 0                            | 0                        | 10    | 10            | 100%         |
| WT pro-R RS-1    | 0                 | 4                            | 0                        | 11    | 15            | 73%          |
| WT pro-S RS-1    | 0                 | 0                            | 1                        | 10    | 11            | 91%          |

Whenever an outlier is encountered, the entire QM/MM MD US simulation is discarded, and a new simulation is begun from a new snapshot. In some cases, as per Table S3, several attempts are necessary before 10 valid runs are obtained.

6.3 WHAM analysis and error calculation (Monte-Carlo Bootstrap procedure)

Using Grossfield’s WHAM software,11 the 12 PMFs shown in Figure 6 are obtained through combined WHAM analysis of the 10 valid US runs existing for each of the 12 reduction processes listed in Table S3. We note that when obtaining these PMFs earlier US windows / reaction coordinates will have poorer collective sampling, because some of the snapshots have different starting points (with the reaction coordinate ranging from –2.5 Å to –1.5 Å). Error bars shown in Figure 6 are obtained via the Monte-Carlo Bootstrap (MCBS) procedure available in Grossfield’s WHAM software. (The general theory is discussed in the WHAM manual: http://membrane.urmc.rochester.edu/sites/default/files/wham/doc.pdf, accessed November 2018. Occasionally, the WHAM program failed to produce an error in PMF regions close to the TS: for these regions, we took the highest error found elsewhere along the PMF). MCBS requires an “autocorrelation decay time” for each US window, which we derived as follows:

- \((x – y)\) value files in each window, in each of the 120 valid US runs, have their autocorrelation function measured up to a lag of 1000 frames (i.e. half their length).
• Arithmetic-averaged autocorrelation functions are obtained for the 10 replicas in each window for each of the 12 reduction processes sampled.
• Maxima are extracted from peaks of the 12 averaged autocorrelation functions (which are assumed to start at 1 for lag = 0).
• The 12 series of maxima are fitted to an exponential decay curve.
• The “autocorrelation decay time” (required by the WHAM program) is defined as twice the time taken by this fitted exponential decay curve to reach $1/e$.

7. Error Analysis for Fl Statistics
Errors in Fl statistics, resulting in the error bars appearing in Figure 4 in the main text, are calculated as follows for each of our six actKR-(NADPH)$_4$-1$_4$ systems, using the leave-one-out procedure:
• The percentage of pro-R reactive poses is recalculated 8 times, but each time removing a different one of the eight MD replicas. This is done using the formula

$$\frac{\sum_{r=1}^{8} \text{reactive pro-R poses in all replicas except } r}{84000} \times 100;$$

where the denominator 84000 indicates that only seven trajectories are being considered.
• The percentage of pro-S reactive poses is quantified in an identical way, again by removing a different one of the 8 MD replicas, using the formula

$$\frac{\sum_{r=1}^{8} \text{reactive pro-S poses in all replicas except } r}{84000} \times 100.$$
• Two standard deviations $\sigma_{\text{pro-R}}$ and $\sigma_{\text{pro-S}}$ are calculated for the eight resulting pro-R and the eight resulting pro-S percentages, respectively.
• The final error $E$ on $\Delta_{\text{pro-S} - \text{pro-R}}$SR-1 or $\Delta_{\text{pro-S} - \text{pro-R}}$RS-1 is obtained using the formula

$$E_{SR-1} = \sqrt{\sigma_{\text{pro-S}}^2 + \sigma_{\text{pro-R}}^2} \quad \text{or} \quad E_{RS-1} = \sqrt{\sigma_{\text{pro-S}}^2 + \sigma_{\text{pro-R}}^2}.$$

For each of WT, P94L, and V151L actKR, overall errors $E$ for $\Delta_{\text{pro-S} - \text{pro-R}}$1 are then obtained using the formula

$$E = \sqrt{E_{SR-1}^2 + E_{RS-1}^2}.$$

8. Typical CPU / GPU Times vs. Accuracy
To assess the effects of using shorter simulation times / scarcer sampling, we have collected:
• data for Fl (Δ) on the first 264 ps, 1008 ps, 4008 ps, and 8016 ps of each “free” MD replica (Figure S4a to Figure S4d, with Figure S4e for comparison to the main text at 12 ns);
• data for FII (ΔGbind) on the first 264 ps, 1008 ps, 4008 ps, and 8016 ps of each “restrained” MD replica (chosen since they are multiples of 24 ps; Figure S5a to Figure S5d, with Figure S5e for comparison to the main text at 12 ns); and
• data for FIII (PMF), this time based on only 36 (Figure S6a) reaction profiles instead of the 120 used for the main batch of results (Figure S6b for comparison).

Trends emerging from Figure S4, Figure S5, and Figure S6 are briefly commented on in the main text.

Figure S4. Effects of shorter “free” MD simulation times on the difference (Δ) between the frequency of reactive pro-R and reactive pro-S poses (Φ). (a) 264 ps; (b) 1008 ps; (c) 4008 ps; (d) 8016 ps; (e) 12000 ps (corresponding to Figure 4 in the main text).
Figure S5. Effects of shorter “restrained” MD simulation times on the binding free energies ($\Delta G_{\text{bind}}$), predicted by MM/GBSA, on SR-1 and RS-1, either in pro-$R$ poses (black or grey, respectively); or in pro-$S$ poses (red or orange, respectively). (a) 264 ps; (b) 1008 ps; (c) 4008 ps; (d) 8016 ps; (e) 12000 ps (corresponding to Figure 5 in the main text).
9. α6-α7 Loop Opening and Barrier Height

To check for a possible correlation between the degree of closure of the α6-α7 loop and higher reactivity (lower reaction barrier), we monitor (Figure S7) reaction barrier height vs. average loop closure in the initial snapshot for each of the 120 energy profiles on which we conducted QM/MM US calculations (10 snapshots per 4 possible reduction routes per three actKR barriers). The average of three key distances (see e.g. Figure S8 for monomer A of the WT) was used as a measure of loop opening/closing:

- Pro94:N [non-loop] – Met194:Cα [loop] (or Leu94:N – Met194:Cα);
- Gly95:C [non-loop] – Val198:N [loop]; and
- Val151:Cα [non-loop] – Ile217:C [loop] (Leu151:Cα – Ile217:C).

Figure S6. Effects of sampling less reaction profiles on the resulting 12 PMFs: (a) 3 individual umbrella sampling (US) QM/MM MD simulations per reduction route per actKR variant instead of (b) the 10 per reduction route per actKR variant (corresponding to Figure 6 in the main text).
Figure S7. Plots of reaction barrier height vs. average α6-α7 loop opening in the starting snapshot for each of the 120 free energy profiles sampled with QM/MM MD with US. There appears to be no (significant) correlation for any of the 12 cases.

Figure S8. View of WT actKR (monomer A) highlighting the three key distances whose average we use to determine the degree of opening/closing of the α6-α7 loop: (blue; foreground) Pro94:N [non-loop] – Met194:Cα [loop]; (green, foreground) Gly95:C [non-loop] – Val198:N [loop]; and (red, background) Val151:Cα [non-loop] – Ile217:C [loop]. Key: same as Figure 1c in the main text; additionally, residues from which distances are measured are rendered as magenta sticks.
10. References

1. Korman, T. P.; Tan, Y.-H.; Wong, J.; Luo, R.; Tsai, S.-C., Inhibition Kinetics and Emodin Cocrystral Structure of a Type II Polyketide Ketoreductase. Biochem. 2008, 47 (7), 1837-1847.
2. Javidpour, P.; Bruegger, J.; Sridhahan, S.; Korman, Tyler P.; Crump, Matthew P.; Crosby, J.; Burkert, Michael D.; Tsai, S.-C., The Determinants of Activity and Specificity in Actinorhodin Type II Polyketide Ketoreductase. Chem. Biol. 2013, 20 (10), 1225-1234.
3. Javidpour, P.; Korman, T. P.; Shaka, G.; Tsai, S.-C., Structural and Biochemical Analyses of Regio- and Stereospecificities Observed in a Type II Polyketide Ketoreductase. Biochem. 2011, 50 (21), 4638-4649.
4. Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A., Automatic atom type and bond type perception in molecular mechanical calculations. J. Mol. Graphics Modell. 2006, 25 (2), 247-260.
5. Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I., Fast, efficient generation of high-quality atomic charges. AM1-BCC model: I. Method. J. Comput. Chem. 2000, 21 (2), 132-146.
6. Jakalian, A.; Jack, D. B.; Bayly, C. I., Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. J. Comput. Chem. 2002, 23 (16), 1623-1641.
7. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A., Development and testing of a general amber force field. J. Comput. Chem. 2004, 25 (9), 1157-1174.
8. Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R., Molecular dynamics with coupling to an external bath. J. Chem. Phys. 1984, 81 (8), 3684-3690.
9. Shuichi, M.; A., K. P., Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. J. Comput. Chem. 1992, 13 (8), 952-962.
10. Loncharich, R. J.; Brooks, B. R.; Pastor, R. W., Langevin dynamics of peptides: The frictional dependence of isomerization rates of N-acetylalanyl-N'-methylamide. Biopolymers 1992, 32 (5), 523-535.
11. Grossfield, A. WHAM: the weighted histogram analysis method, version 2.0.9.1; 2013.
12. Walker, R. C.; Crowley, M. F.; Case, D. A., The implementation of a fast and accurate QM/MM potential method in Amber. J. Comput. Chem. 2008, 29 (7), 1019-1031.
13. Seabra, G. d. M.; Walker, R. C.; Elstner, M.; Case, D. A.; Roitberg, A. E., Implementation of the SCC-DFTB Method for Hybrid QM/MM Simulations within the Amber Molecular Dynamics Package. J. Phys. Chem. A 2007, 111 (26), 5655-5664.
14. Stewart, J. J. P., Optimization of parameters for semiempirical methods V: Modification of NDDO approximations and application to 70 elements. J. Mol. Model. 2007, 13 (12), 1173-1213.
15. Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, T.; Suhai, S.; Seifert, G., Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties. Phys. Rev. B 1998, 58 (11), 7260-7268.
16. Gaus, M.; Cui, Q.; Elstner, M., DFTB3: Extension of the Self-Consistent-Charge Density-Functional Tight-Binding Method (SCC-DFTB). J. Chem. Theory Comput. 2011, 7 (4), 931-948.
17. Gaus, M.; Goez, A.; Elstner, M., Parametrization and Benchmark of DFTB3 for Organic Molecules. J. Chem. Theory Comput. 2013, 9 (1), 338-354.
18. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, M. A.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams-Young, D.; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Keith, T.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Cygler, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.
Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J., Gaussian 09, Revision D.01. Wallingford CT, 2009.

19. Kohn, W.; Sham, L. J., Self-Consistent Equations Including Exchange and Correlation Effects. *Phys. Rev.* **1965**, **140**(4A), A1133-A1138.

20. Becke, A. D., Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993**, **98**(7), 5648-5652.

21. Grimme, S.; Goerigk, L.; Fink, R. F., Spin-component-scaled electron correlation methods. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2012**, **2**(6), 886-906.

22. Tomasi, J.; Mennucci, B.; Cammi, R., Quantum Mechanical Continuum Solvation Models. *Chem. Rev.* **2005**, **105**(8), 2999-3094.

23. Li, X.; Frisch, M. J., Energy-Represented Direct Inversion in the Iterative Subspace within a Hybrid Geometry Optimization Method. *J. Chem. Theory Comput.* **2006**, **2**(3), 835-839.

24. Schlegel, H. B., Optimization of equilibrium geometries and transition structures. *J. Comput. Chem.* **1982**, **3**(2), 214-218.

25. Peng, C.; Ayala, P. Y.; Schlegel, H. B.; Frisch, M. J., Using redundant internal coordinates to optimize equilibrium geometries and transition states. *J. Comput. Chem.* **1996**, **17**(1), 49-56.

26. Peng, C.; Schlegel, H. B., Combining Synchronous Transit and Quasi-Newton Methods to Find Transition States. *Israel J. Chem.* **1993**, **33**(4), 449-454.

27. Maeda, S.; Harabuchi, Y.; Ono, Y.; Taketsugu, T.; Morokuma, K., Intrinsic reaction coordinate: Calculation, bifurcation, and automated search. *Int. J. Quant. Chem.* **2015**, **115**(5), 258-269.

28. https://doi.org/10.19061/iochem-bd-6-12.

29. Medina, F. E.; Neves, R. P. P.; Ramos, M. J.; Fernandes, P. A., A QM/MM study of the reaction mechanism of human [small beta]-ketoacyl reductase. *Phys. Chem. Chem. Phys.* **2017**, **19**(1), 347-355.

30. Noey, E. L.; Tibrewal, N.; Jiménez-Osés, G.; Osuna, S.; Park, J.; Bond, C. M.; Cascio, D.; Liang, J.; Zhang, X.; Huisman, G. W.; Tang, Y.; Houk, K. N., Origins of stereoselectivity in evolved ketoreductases. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, **112**(51), E7065-E7072.
