
Peer reviewed version

Link to published version (if available): 10.1021/acscatal.8b04846

Link to publication record in Explore Bristol Research

PDF-document

This is the author accepted manuscript (AAM). The final published version (version of record) is available online via ACS at https://pubs.acs.org/doi/10.1021/acscatal.8b04846. Please refer to any applicable terms of use of the publisher.

University of Bristol - Explore Bristol Research

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:

http://www.bristol.ac.uk/pure/about/ebr-terms
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

ABSTRACT: 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-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 computationally cost-effective atomistic simulations to rationalize these trends in WT, P94L, and V151L actKR, using trans-1-decalone (I) 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 stereocatalysis: 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-decanol 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.

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

1. Introduction

Typically found in actinomycete bacteria, type II polyketide synthases (PKSs) are large multi-enzyme complexes consisting of multiple standalone units. 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).

A key component of PKSs is their ketoreductase (KR) domain, 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 biocatalysts for the manufacture of small chiral alcohols from achiral ketones; and even for the synthesis of “unnatural natural products” with potentially improved antibiotic properties.

Such rise to prominence was significantly aided by the advent of novel protein reengineering techniques, which have revolutionized biocatalysis by speeding up the design and screening of evolved KR mutants.

Actinorhodin KR (actKR) from Streptomyces coelicolor has been attracting interest as a potential biocatalyst since at least the mid-2000s. Examination of its sequence and crystal structures 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). 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 stereocatalysis and activity of Lactobacillus kefir short-chain alcohol dehydrogenase. 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).

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 tetrad Asn114-Ser144-Tyr157-Lys161. Another salient structural element is the flexible α6-α7 loop which, by advent of novel protein reengineering techniques, has revolutionized biocatalysis by speeding up the design and screening of evolved KR mutants.

Actinorhodin KR (actKR) from Streptomyces coelicolor has been attracting interest as a potential biocatalyst since at least the mid-2000s. Examination of its sequence and crystal structures 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).

Another salient structural element is the flexible α6-α7 loop which, by advent of novel protein reengineering techniques, has revolutionized biocatalysis by speeding up the design and screening of evolved KR mutants.

Actinorhodin KR (actKR) from Streptomyces coelicolor has been attracting interest as a potential biocatalyst since at least the mid-2000s. Examination of its sequence and crystal structures 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).

Another salient structural element is the flexible α6-α7 loop which, by advent of novel protein reengineering techniques, has revolutionized biocatalysis by speeding up the design and screening of evolved KR mutants.

Actinorhodin KR (actKR) from Streptomyces coelicolor has been attracting interest as a potential biocatalyst since at least the mid-2000s. Examination of its sequence and crystal structures 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).

Another salient structural element is the flexible α6-α7 loop which, by advent of novel protein reengineering techniques, has revolutionized biocatalysis by speeding up the design and screening of evolved KR mutants.

Actinorhodin KR (actKR) from Streptomyces coelicolor has been attracting interest as a potential biocatalyst since at least the mid-2000s. Examination of its sequence and crystal structures 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).

Another salient structural element is the flexible α6-α7 loop which, by advent of novel protein reengineering techniques, has revolutionized biocatalysis by speeding up the design and screening of evolved KR mutants.
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 \(\alpha_6\)-\(\alpha_7\) loop in monomers A-D (A, C, open: light blue; B, D, closed: pink). (d) Monomer B superimposed onto monomer A.

Scheme 1. Rate-limiting step of actKR (reductive hydride transfer) and its stereochemistry.\(^{a,b}\)

\(^{a}\) Ketone substrate orientation (left) determines whether H\(^-\) attack is \(\text{pro-R}\) or \(\text{pro-S}\), and accordingly gives rise to an \(R^*\) or \(S^*\) product (right). \(^{b}\) Magenta lines denote hydrogen bonds.
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\textsuperscript{14–17} 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.,\textsuperscript{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)\textsuperscript{14,16–17} since other potential candidates, notably the part-aromatic α-tetralone, were found to be turned over very sluggishly.\textsuperscript{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-3 / S-3; Chart 1, right)\textsuperscript{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-3; and V151L, has exclusive specificity for R-3. By comparison, WT actKR only has a very mild preference for S-3 over R-3 (3.5:1).

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

<table>
<thead>
<tr>
<th>Variant</th>
<th>( k_{\text{cat}}/K_M ) (1)\textsuperscript{a,b}</th>
<th>( k_{\text{cat}}/K_M ) (R-3)\textsuperscript{c}</th>
<th>( k_{\text{cat}}/K_M ) (S-3)\textsuperscript{c}</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT\textsuperscript{d}</td>
<td>3.23 ± 0.32</td>
<td>0.010 ± 0.001</td>
<td>0.035±0.006</td>
</tr>
<tr>
<td>V151L\textsuperscript{e}</td>
<td>1.28 ± 0.18</td>
<td>0.026 ± 0.005</td>
<td>inhibition</td>
</tr>
<tr>
<td>P94L\textsuperscript{d}</td>
<td>1.02 ± 0.59</td>
<td>inhibition</td>
<td>0.036 ± 0.011</td>
</tr>
</tbody>
</table>

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

The “forwards” reaction is further complicated by the fact that I (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\textsuperscript{+} in pro-R orientation, SR-1 can only yield (1R,4aS,8aR)-2 (henceforth RSR-2; black in Chart 1), and RS-1 can only yield (1R,4aR,8aS)-2 (RRS-2; grey); conversely, when attacked in pro-S orientation, SR-1 and RS-1 can only yield (1S,4aS,8aR)-2 and (1S,4aR,8aS)-2, respectively (SSR-2 and SRS-2; 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.

Due to increases in computer speed and improvements in algorithms, parameters, and usability over the past decades,\textsuperscript{18} in silico biomolecular simulation is increasingly used to complement experimental enzymatic studies,\textsuperscript{19–21} including those focusing on the design of novel enzyme variants relevant for biocatalysis.\textsuperscript{22–25} Indeed, enzyme reactivity and stereocontrol are often the result of a subtle interplay between different factors that are difficult to examine separately experimentally.\textsuperscript{26}

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 F1) 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, F1 describes the energetics of reaching and maintaining a reactive pose.\textsuperscript{27} 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.\textsuperscript{26} 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.\textsuperscript{28,29}

Factor II (henceforth F2) can be seen as part of F1, 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. \( \Delta \Delta G_{\text{bind}} \neq 0 \)), 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,\textsuperscript{30} enhanced Monte Carlo,\textsuperscript{31} and MM/PBSA.\textsuperscript{32–35}

Finally, Factor III (henceforth F3) 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;\textsuperscript{36–38} a popular option is the use of a hybrid quantum / classical (QM/MM) approach,\textsuperscript{39} where the computationally costly QM treatment can be limited to regions of chemical change (without time-intensive parameterization).\textsuperscript{37–39}

In the present work, we describe our computational efforts to rationalize the stereoselectivity of WT, P94L, and V151L actKR towards 1, determining the role of F1-F3 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 \((1S)\text{-trans-1-decalol}; V151L actKR to preferentially form \((1R)\text{-trans-1-decalol}; and a slight preference for \((1S)\text{-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.}
Chart 1. actKR substrates discussed in this work.\textsuperscript{a,b}

\begin{center}
\includegraphics[width=\textwidth]{chart1}
\end{center}

\textsuperscript{a} 1 and 2 color-coded as in Figure 4 and Figure 6, respectively. \textsuperscript{b} Underlined R or 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).

Figure 2. Comparison of (a, b) pro-R and pro-S poses of SR-1 in WT actKR; and (c, d) as before, but with RS-1. In the top left of each panel, we indicate the enantiomer of 2 generated in each case (cf. Chart 1). Snapshots are taken from the opposite angle to that in Figure 1b, and salient protons are shown in light yellow; colors and labels are otherwise identical.
2. Computational Strategy and Details

General Procedure. For F1, 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 F1, 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-$\mathcal{R}$ and pro-$\mathcal{S}$ poses may be directly extracted from these runs. To avoid substrate diffusion from the active site, a one-sided harmonic restraint ($k = 50$ kcal mol$^{-1}$ 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-$\mathcal{R}$ and pro-$\mathcal{S}$ 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-$\mathcal{R}$ and pro-$\mathcal{S}$ (see Supporting Information for further details).

To determine the role of FII, $\Delta G_{\text{bound}}$ of SR-1 and RS-1 in their two prochiral orientations is calculated through a series of MM/PBSA calculations$^{32-35}$ (vide infra) 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).$^{38}$

For simplicity, all classical and QM/MM MD simulations are run in two distinct sets, treating enantiomers SR-1 and RS-1 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 [acrKR-(NADPH)$_2$-I] are set up with all four sites occupied: SR-1 or RS-1 in WT, P94L, or V151L acrKR. All starting structures are constructed from our reference WT crystal structure (PDB ID: 2RH4),$^{14}$ to demonstrate the possibility of assessing enzyme variants for which no crystal structure has been obtained. Existing mutant crystal structures (V151L; PDB ID: 4DBZ$^{17}$ and P94L; PDB IDs: 3RI3, 3QRW$^{16}$) are checked to ensure that constructed mutants retain plausibly oriented sidechains. Constructed structures are henceforth labelled as follows: WT-SR-1, (cf. Figure 2a and 2b), V151L-SR-1, and P94L-SR-1, for one enantiomer of I; and WT-RS-1 (cf. Figure 2c and 2d), V151L-RS-1, and P94L-RS-1, for the other.

In all six cases, active sites in monomers A and B are populated with the substrate in pro-$\mathcal{R}$ orientation whereas those in monomers C and D are populated with the substrate in pro-$\mathcal{S}$ orientation. All four monomers are modelled from residues 1 to 261. All residues are modelled in their standard protonation states, in line with pK$_\text{a}$s predicted by PROPKA 3.1.$^{40}$ Hydrogens are added with AmberTools’ (version 17)$^{41}$ reduce utility, resulting in His162 being singly protonated on N61; His153 and His201 on N2. Using tleap,$^{42}$ 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 acrKR-(NADPH)$_1$-I$_1$ systems in explicit water are run with the AMBER software package (version 16, 2017 distribution),$^{41-42}$ taking advantage of GPU acceleration where applicable.$^{43}$ Postprocessing and analysis of MD trajectories are carried out with the CPPTRAJ utility.$^{44}$ and visual inspection is conducted with VMD.$^{45}$ The protein and ions are described by the ff14SB forcefield,$^{46}$ the NADPH cofactor by the forcefield from Holmberg and coworkers,$^{47}$ and water using the TIP3P model.$^{48}$ The GAFF forcefield$^{49}$ with AM1-BCC charges derived with antechamber$^{50-51}$ 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.$^{52-53}$

For both the “free” and “restrained” MD simulations (vide supra) we carry out eight independent simulations (different random seeds) for each of the six acrKR-(NADPH)$_1$-I$_1$ 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$^{64}$ and a constant temperature of 298 K is enforced via the Langevin thermostat$^{55}$ (collision frequency set at 5 ps$^{-1}$). Bonds containing hydrogen are constrained by employing the SETTLE and SHAKE algorithms.$^{56}$ (See Supporting Information for details on settings in the pre-production stages).

MM/PBSA Calculations. MM/PBSA calculations$^{32-35}$ are run on 500 snapshots from each restrained MD trajectory, on 16 processors, using AmberTools’ MMPBSA.py.MPI utility,$^{57}$ with default atomic radii,$^{58}$ settings, and parameters.$^{52, 57}$ 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 $\varepsilon_\text{int}$, which is set to 4.0 as advised by Wang and coworkers.$^{35}$ The six acrKR-(NADPH)$_1$-I$_1$ topologies are preprocessed using the ante-MMPBSA.py tool.$^{57}$

The eight trajectories of each of the six acrKR-(NADPH)$_1$-I$_1$ systems are parsed by MMPBSA.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 I (i.e. one in each active site), four separate MM/PBSA runs are required on each “supertrajectory” (i.e. 24 in total, giving 24 $\Delta G_{\text{bound}}$ values). Entropic corrections to these values are calculated using the interaction entropy method reported by Duan et al.$^{59}$

QM/MM MD with Umbrella Sampling. All QM/MM MD simulations are run with AMBER’s sander MD engine.$^{60-61}$ Apart from the time-step (here 1 fs) and the absence of restraints other than the reaction coordinate, all other conditions
are identical to the classical MD production runs. The QM region (with explicit treatment of electrons) comprises: the reacting substrate in its entirety; NADPH’s nicotinamide moiety up to the first ribose; the Ser144 and Tyr157 sidechains (from C9); and three hydrogen link-atoms at the QM-MM boundary (Figure 3). All remaining atoms are included in the classical (MM) region. The SHAKE algorithm is switched off in the QM region, and atoms are treated using the semiempirical method PM6. PM6 was chosen over other semiempirical methods after benchmarking the reaction on a small model with higher-level QM calculations (SCS-MP2/aug-cc-pVTZ single-point energy calculations on B3LYP/aug-cc-pVTZ optimized potential energy surfaces); PM6 correctly identified the mechanism with a highly similar transition state, even though it significantly overestimates the barrier for reaction and the reaction energy. Details of benchmarking calculations are provided as Supporting Information; relevant coordinates and frequency calculations are also available for all stationary points in the ioChem-BD repository.63

Figure 3. QM region for QM/MM MD US simulations (rendered as sticks, except for three capping H atoms rendered as white spheres). Key: white: H; grey: C; blue: N; red: O. Secondary structure of MM region rendered as yellow ribbon. The chosen reaction coordinate (x – y) is marked with blue and black lines; atoms included in it are defined in the text.

Simulations are started from a series of representative reactive snapshots (see Supporting Information) from the restrained MD simulations, with retention of atom velocities. Ten such snapshots are chosen for each of the twelve reaction paths to be compared: reduction of SR-1 to either (1) RSR-2 or (2) SSR-2; and reduction of RS-1 to either (3) RRS-2 or (4) SSR-2; in WT; (5-8) P94L; and (9-12) V151L actKR. Only one of the four substrates present in each snapshot is considered for umbrella sampling (US).

Hydride transfer (i.e. the reaction coordinate) is monitored using the difference between two key distances (Figure 3): the distance NADPH:H−→C1(y) is subtracted from the distance NADPH:H−→NADPCHO−(x). Umbrella sampling (US) is performed by lengthening a restraint on the reaction coordinate (x – y) in 0.1 Å steps (“windows”) starting from a value close to that found in the initial snapshot, and thus gradually “forcing” the system to react until (x – y) reaches a value of 1.8 Å. Each US window is simulated with QM/MM MD for 2 ps, and (x – y) is restrained at the desired length through a harmonic biasing potential (k = 100 kcal mol−1 Å2); the last configuration of each US window is used to start the QM/MM MD simulation for the following one. Depending on the starting value of (x – y), the simulation time for each individual US simulation ranges from 66 ps (33 windows) to 90 ps (45 windows), with the total cumulative simulation time for each free energy profile ranging from 728 ps to 840 ps. 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) with the eponymous program by Grossfield; 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 (Δpro-R−pro-S[lR-1] and RS-1 (Δpro-R−pro-S[lS-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-1, V151L-SR-1, P94L-SR-1, WT-RS-1, V151L-RS-1, and P94L-RS-1), statistics for Δpro-R−pro-S[lR-1] or Δpro-R−pro-S[lS-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. Δpro-R−pro-S[lR-1] or Δpro-R−pro-S[lS-1] are calculated for each variant using:

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

and

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

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 (Δpro-R−pro-S[lR-1]) is then obtained by summing Δpro-R−pro-S[lR-1] and Δpro-R−pro-S[lS-1].
A second noteworthy observation is that for the overall difference in pro-R / pro-S reactive poses (∆pro-R vs. pro-S; 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. 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. Nonetheless, if FII 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-I systems. Using snapshots taken at 24 ps-intervals from 8 independent restrained MD runs of 12 ns each, ∆Gbind 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).

For SR-1, ∆Gbind 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 (∆ΔGbind ≈ 0), this is clearly not the case in the two mutants: RS-1 in pro-R orientation has a significantly higher affinity (compared to its pro-S orientation) in both V151L actKR (∆ΔGbind = 1.9 kcal mol⁻¹) and P94L actKR (∆ΔGbind = 3.6 kcal mol⁻¹). This increase in ∆ΔGbind when going from V151L to P94L actKR is mainly caused by a loss in affinity for the pro-S orientation.
the use of PM6 for the QM region (see Supporting Information, Figure S2); in addition, the small QM region, the approximate reaction coordinate and, in case of reaction energies, the fact that we do not sample subsequent proton transfers, will likely contribute to overestimation. We note that differences in reaction barrier height are what is relevant for investigation of the influence of **FIII** on stereoselectivity.

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

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 S<sub>SR-2</sub> 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);<sup>17</sup> 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 S<sub>RS-2</sub> (orange; 2.1 to 3.6 kcal mol<sup>-1</sup> lower than the other three). None of the other three conversions stand out for being particularly “easier”. This may agree with only a mild preference for S-3 being observed experimentally.<sup>14, 16</sup> Finally, we note that the reaction barriers for the expected preferred reactions are similar between the three enzyme variants. This is consistent with experiment,<sup>6, 14, 17</sup> considering our approximate reaction simulation approach: the P94L and V151L mutations only have very minor effects on the measured conversion rate (k<sub>on</sub>) of racemic 1 to 2, which translate (using transition state theory) to subtle increases in the free energy of activation (ΔG<sup>‡</sup>) of 0.8 kcal mol<sup>-1</sup> and 0.4 kcal mol<sup>-1</sup>, respectively.

**Different Factors Dominate for Different Variants.** Taking into account the data for FII-FIII arising from our different simulations (Figure 4, Figure 5, and Figure 6), it becomes clear that the causes driving stereosecontrol 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 RS-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 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 sterics and thermodynamics prior to the reduction step (FI and FII). As such, our calculations indicate that this variant will predominantly form RS-2 (instead of R<sub>SR-2</sub>).

In the case of the WT enzyme, mild selectivity can be expected towards S-2 (based on its experimentally determined specificity towards S-3). The pro-S 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 S<sub>SR-2</sub> product. Thus, for S<sub>SR-2</sub> formation (compared to R<sub>SR-2</sub>), this “steric head start” is likely important. Formation of the other S-2 isomer, S<sub>RS-2</sub>, 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 S-2 selectivity (as expected from its in vitro specificity towards S-3) mostly arises due to a change in relative reaction barriers (FIII). Specifically, formation of S<sub>SR-2</sub> is clearly preferred (Figure 6). FI may play some additional role, since SR-1 forms more pro-S reactive poses. Altogether, the simulations further indicate that this variant will predominantly form S<sub>SR-2</sub> (instead of R<sub>SR-2</sub>); pro-S RS-1 reactive poses occur less frequently and have lower affinity than pro-S SR-1 reactive poses (Figure 4; Figure 5), and they are also much less favored to react to S<sub>SR-2</sub> (FIII; Figure 6).

<table>
<thead>
<tr>
<th>Enantiomer</th>
<th>SR-1</th>
<th>RS-1</th>
<th>RS-1</th>
<th>RS-1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prochirality</td>
<td>pro-R</td>
<td>pro-S</td>
<td>pro-R</td>
<td>pro-S</td>
</tr>
<tr>
<td>Non-CO ring&lt;sup&gt;a&lt;/sup&gt;</td>
<td>→ 94</td>
<td>→ 151</td>
<td>→ 94</td>
<td>→ 151</td>
</tr>
<tr>
<td>H in 2&lt;sup&gt;b&lt;/sup&gt;</td>
<td>Axial</td>
<td>Equatorial</td>
<td>Equatorial</td>
<td>Axial</td>
</tr>
</tbody>
</table>

<sup>a</sup> Approximate direction towards which 1’s C4a-C5-C6-C7-C8-C8a ring points: P/L94 vs. V/L151. <sup>b</sup> Final position of the transferred hydride in the resulting isomer of 2.
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-S RS-1 to RRS-2 in WT actKR; (b) pro-S SR-1 to SSR-2 in WT actKR; (c) pro-R RS-1 to RRS-2 in V151L actKR; and (d) pro-S 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.

Our extensive atomistic simulations allow investigation of geometrical characteristics within each active site that may help rationalize the above findings. Considering the orientation of RS-1 and SR-1 in the active site (Table 2), as well as representative transition states for their reaction to 2 (Figure 7), it emerges that (pro)chiralities enhanced by each mutation (R for V151L; S for P94L) are those which do not entail a steric clash between the non-carbonyl ring of 1 and the extra bulk introduced by that mutation. We note that, in this case, this effect could not be reproduced by standard molecular docking (likely due to the active site’s large volume compared to 1).

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-S 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 Enantiomers and Equatorial Attack.** Our work indicates that P94L actKR prefers facilitating the reaction of pro-S 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-S 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\(^7\), (with reagents such as [B(H)\(_2\)] and [AIH\(_4\)]), 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,\(^72\) 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.,\(^73\) 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 \(\alpha\)-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 \(\alpha\)-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 \textit{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 \(\alpha\)-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 (R/S)-2 remains (through a degree of microscopic reversibility)
comparable to the oxidation of (R/S)-3 to α-tetralone.
(This issue is reminiscent of the “experimental problem” re-
viewed by Van Gunsteren and colleagues,74 whereby experi-
mental 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 ‘chair-
flip’ 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 RSR-2/RRS-2). The only major (steric) dif-
fERENCE 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 carefully chosen computational methods to
dissect the influence of FI−FIII—something that is very hard
to achieve experimentally. We have deliberately used proto-
cols 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 coworkers24, 75−76 to efficiently screen newly sug-
gested stereoselective enzyme variants in silico, prior to as-
essment by experiment. We note that the HTMI-MD ap-
proach 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 re-
sources required should be modest. In this respect, the fact that
our relatively low-cost protocols have been able to character-
ize 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 sub-
section, we discuss the rationale for the protocols to assess
FI−FIII and evaluate if computational time can be reduced fur-
ther.

To study FI (and to later generate trajectories for FII and
FIII) we employed classical MD simulations, previously em-
ploired in several examples of interest with other KR's and
SDRs.10, 77 In contrast to many previous studies that have used
a small number of longer MD simulations to collect their da-
ta,10, 26, 77−78 our strategy was to run a large set of independent
MD simulations of shorter length (8 × 12 ns, essentially be-
coming 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),24, 76 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.79

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 ∆Gtrue/prediction problems across
several enzyme-ligand systems.32, 30−31 Alternative options such
as absolute binding free energy calculations using alchemical
perturbations30−31 would have raised the computational cost
significantly.82 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.83−84 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.83−84 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 previ-
ous MM/PBSA and MM/GBSA indications, but could be re-
duced).83−85

To study FIII, we employed a semiempirical QM/MM ap-
proach, to avoid specific parameterization of models (e.g. as
required for empirical valence bond methods)36 and limit com-
putational 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).10, 15 A wide range of options is available for calculat-
ing reaction barriers with QM/MM.35−38 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 acti-
vation energy barriers, and this would have meant collecting a
very large number of starting conformations.86 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 larg-
er errors).

Notably, QM/MM simulations used a range of starting
structures from the many independent classical MD simula-
tions performed: these capture a range of $\alpha_6$-$\alpha_7$ loop confor-
mations (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 intera-
tomic loop-enzyme distances) and barrier height (Supporting
Information, Figures S7 and S8), indicating that in this case,
that the $\alpha_6$-$\alpha_7$ loop conformation has little or no influence on
actKR’s catalytic reactivity when turning over small sub-
strates.

4. Conclusions

Biocatalyst (stereo)selectivity can arise due to a combina-
tion 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 en-
zyme studied here). One route towards stereoselective en-
zymes (for use in biocatalysis) can thus be to reduce their ac-
tive 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 sub-
strates (of different sizes) as well as reduce efficiency.

Here, we have presented a detailed computational study of
three factors that may affect stereocontrol in the wild-type
ketoreductase actKR and in two key variants with similar effi-
ciency (P94L, V151L). We separately investigated the for-
mation (F1), the binding free energy (FII) and the reactivity
(free energy barrier, FIII) of reactive complexes, using clas-
sical 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 $\Sigma$-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 (F1, FII) are mainly responsible.

The observation that the main effect (or factor) determining
stereocontrol 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 al-
tered to confer stereoselectivity.

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

The complexity of our findings is representative of the mul-
tifaceted origins of catalysis and selectivity in enzyme biocata-
lysts, which arises from a subtle interplay of steric, dynamic,
and electronic effects. Detailed biomolecular simulation pro-
vides valuable means to break down the causes of stereocon-
trol (e.g. in terms of the individual factors FII-FIII). The compu-
tational procedures we have used here to determine the ori-
gins of stereocontrol in actKR are deliberately generically
applicable and employ limited computer resources. They can
thus be used for similar evaluations of different enzyme bio-
catalysts in a time- and cost-effective manner (especially in
combination with further automation); although it should be
noted that evaluation of FIII does require the reaction mech-
nism to be known in advance. Indeed, we believe that an accu-
rate, yet computationally efficient assessment of the different
effects (or factors) dictating an enzyme’s stereocontrol is high-
ly valuable for the (re)design of biocatalysts with enhanced
stereoselectivity or -specificity, as well as for in-depth under-
standing 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 ob-
tained from benchmarking (available at https://doi.org/10.19061/iochem-bd-6.12): starting structures and
forcefield topologies for the six actKR-(NADPH)$_2$-I$_4$ 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 experi-
ments, MKW devised the project. Both authors analyzed results,
wrote the manuscript and have given approval to the final version
of the manuscript.

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

ACKNOWLEDGEMENT

The authors thank Prof. Shiu-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 compu-
tational facilities of the Advanced Computing Research Centre of
the University of Bristol.
REFERENCES


65. A collection of relevant calculations is available on the iochemBD server. [https://doi.org/10.19061/iochem-bd-6-12] Last accessed: December 3rd 2018.


