

A EUROPEAN JOURNAL OF CHEMICAL BIOLOGY

CHEMBIOCHEM

SYNTHETIC BIOLOGY & BIO-NANOTECHNOLOGY

Accepted Article

Title: Computational design of enantiocomplementary epoxide hydrolases for asymmetric synthesis of aliphatic and aromatic diols

Authors: Hesam Arabnejad, Elvira Bombino, Dana I. Colpa, Peter A. Jekel, Milos Trajkovic, Hein J. Wijma, and Dick Janssen

This manuscript has been accepted after peer review and appears as an Accepted Article online prior to editing, proofing, and formal publication of the final Version of Record (VoR). This work is currently citable by using the Digital Object Identifier (DOI) given below. The VoR will be published online in Early View as soon as possible and may be different to this Accepted Article as a result of editing. Readers should obtain the VoR from the journal website shown below when it is published to ensure accuracy of information. The authors are responsible for the content of this Accepted Article.

To be cited as: *ChemBioChem* 10.1002/cbic.201900726

Link to VoR: <http://dx.doi.org/10.1002/cbic.201900726>

WILEY-VCH

www.chembiochem.org

A Journal of



Computational design of enantiocomplementary epoxide hydrolases for asymmetric synthesis of aliphatic and aromatic diols

Hesam Arabnejad, Elvira Bombino, Dana I. Colpa, Peter A. Jekel, Milos Trajkovic, Hein J. Wijma, Dick B. Janssen*

Biotransformation and Biocatalysis, Groningen Biomolecular Sciences and Biotechnology Institute, University of Groningen, Nijenborgh 4, 9747AG, Groningen, The Netherlands

* corresponding author: d.b.janssen@rug.nl

Accepted Manuscript

Abstract

The use of enzymes in preparative biocatalysis often requires tailoring enzyme selectivity by protein engineering. Here, we explore the use of computational library design and MD simulations to create variants of limonene epoxide hydrolase that produce enantiomeric diols from *meso*-epoxides. Three substrates of different sizes were targeted: 2,3-butene oxide, cyclopentene oxide and *cis*-stilbene oxide. Most of the 28 designs tested were active and showed the predicted enantioselectivity. Excellent enantioselectivities were obtained for the bulky substrate *cis*-stilbene oxide, and enantiocomplementary mutants produced (*S,S*)- and (*R,R*)-stilbene diol with >97% enantiomeric excess. An (*R,R*)-selective mutant was used to prepare (*R,R*)-stilbene diol with high enantiopurity (98% conversion to diol, >99 % e.e.). Some variants displayed higher catalytic rates (k_{cat}) than the original enzyme, but in most cases K_{M} values increased as well. The results demonstrate the feasibility of computational design and screening to engineer enantioselective epoxide hydrolase variants with very limited laboratory screening.

Keywords

Computational design, enantioselectivity, epoxide hydrolase, molecular dynamics, stilbene oxide

Introduction

The use of biocatalysis in chemistry is an attractive option for many synthetic processes, especially for preparing fine chemicals and bioactive compounds.^{[1]-[7]} Although nature provides an enormous diversity of industrially useful enzymes, they often must be engineered to meet industrial process requirements.^{[8], [9]} In case of pharmaceutical synthesis, of special importance are chemoselectivity, compatibility with harsh reaction conditions and product enantiopurity.^{[1], [10]} Consequently, extensive studies have been carried out on controlling and improving enzyme selectivity by protein engineering, often through directed evolution,^[11] which led to enzymes with improved selectivity in kinetic resolution of enantiomers and better performance in asymmetric transformation of prochiral compounds.^{[8] [10] [12]}

Whereas directed evolution is very successful, it requires high-throughput screening methods. In case of enzyme enantioselectivity, screening is possible by chiral chromatography or by the use of quasi enantiomers in NMR or MS^[13] but this may be time-consuming and expensive. Directed evolution becomes complicated when no high-throughput expression is available, such as in case of enzymes that must be produced in fungi. Several methods have been proposed to overcome these bottlenecks, such as optimizing strategies for library construction, e.g. by focusing mutations in different combinations around the active site,^{[14] [15] [16]} and by incorporating structural^[9] or

phylogenetic information.^{[17] [18]} Another option is the use of computational tools to design improved enzymes.^{[19] -[23]} Statistical methods have also been used.^{[24] [25]} Biophysics-based computational protocols have emerged as powerful platforms for the engineering of thermostable and organic-solvent compatible enzyme variants.^{[26] -[31]} Multiple mutations can be explored simultaneously, allowing larger jumps in sequence space compared to directed evolution and structure-based rational mutagenesis. Furthermore, *in silico* screening of enzyme variants by docking and high-throughput molecular dynamics simulations makes it possible to predict enzyme properties and reduce library size for experimental evaluation from thousands to dozens.^{[32] [33]}

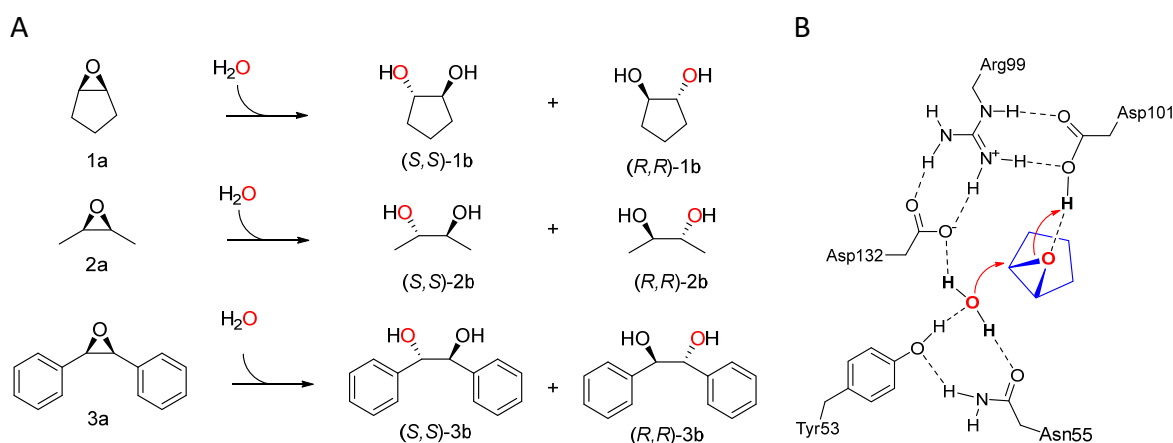
In this study, we explore a computational framework (CASCO, catalytic selectivity by computational design)^[32] for obtaining enantiocomplementary epoxide hydrolases. This framework uses the Rosetta scoring function and search algorithm^{[19] , [34]} to generate libraries of primary designs. Next, high-throughput molecular dynamics simulations (MD) with scoring the frequency of occurrence of reactive (or near-attack) conformations (NACs)^{[35] -[37]} are used for ranking and to select a small set of variants that qualify for laboratory testing. The frequency of reactive poses during MD simulations can explain the selectivity of computationally designed enzyme variants.^{[38] -[40]} For that purpose, generating multiple MD trajectories with independent assignment of initial atom velocities gives much better sampling of accessible conformational space than the use of a single trajectory running over a long simulation time.^{[41] -[45]} Accordingly, to enable screening of thousands of Rosetta designs by MD, CASCO uses 20-80 of such short MD simulations for scoring conformational stability of designed reactive enzyme-substrate complexes. The MD step thus examines if the conformation of the enzyme substrate complex, which is partially constrained during the Rosetta design step (NAC), will be maintained in short MD runs, or whether that reactive conformation is immediately lost e.g. by movement of the substrate to a non-reactive pose. We used this approach earlier to predict enantioselectivity in kinetic resolutions catalyzed by haloalkane dehalogenases.^[46]

The potential of this computational approach was for limonene epoxide hydrolase (LEH) redesign was illustrated in a previous study^[32] where we observed that the performance of the best enzymes in a 37-variant library obtained by the CASCO framework was similar to that of the best variants obtained by screening approximately 4700 variants generated by the CASTing strategy for directed evolution.^[47] Nevertheless, computational library design for enzyme engineering has serious challenges, of which reliability of predictions is an important example. To further explore the possibilities and limitations of computational redesign, we designed and examined a novel set of enantioselective limonene epoxide hydrolase (LEH) variants.

LEH catalyzes the hydrolysis of epoxides by activating a bound water molecule for nucleophilic attack directly on one of the substrate's oxirane carbon.^[48] The water is positioned by H-bonds to Asn55 and Tyr53 while Asp132 abstracts a proton from the water,

which enables nucleophilic attack as a hydroxyl ion (Scheme 1). At the same time, Asp101 protonates the epoxide oxygen, which makes it a better leaving group. The reaction is concerted.^[49] In case of *meso*-epoxides, the stereochemical outcome is determined by regioselectivity of the attack and the products are enantiomers. LEH has been extensively used as a model system for exploring the use of directed evolution strategies to engineer enantioselectivity, and many variants have been described.^{[47] [50] - [55]} We previously examined the use of computational design and screening to improve stability and control enantioselectivity.^{[26] , [29] [32]}

In this work, we investigated three small sets of enantiocomplementary epoxide hydrolase variants for converting *meso*-epoxides (*cis*-2,3-butene oxide, cyclopentene oxide, and *cis*-stilbene oxide) to their corresponding (*R,R*)- or (*S,S*)-diols. For each target enantiomer, five top-ranked new variants were experimentally characterized to explore how challenging it is to position each of the substrates uniquely in the active site. We also measured k_{cat} and K_{M} values for selected variants. This showed that the k_{cat} of some variants was higher than of the thermostable LEH-P variant, from which all mutants were derived. However, the Michaelis constants (K_{M}) were almost always higher as well, which decreases catalytic efficiency. The results confirmed the working hypothesis that obtaining unique binding orientations, reflected in high enantioselectivity, was easier with the bulkier substrates. Mutants converted *cis*-stilbene oxide to diols with an enantiomeric excess (e.e.) of > 99%. Small-scale preparative reactions were carried out.



Scheme 1. Conversion of epoxides by limonene epoxide hydrolase. A) regioselective hydrolysis of *meso* epoxides examined in this study. **B)** Catalytic mechanism of LEH, illustrated with *proRR* hydrolysis of cyclopentene oxide.

Results

Computational design. The enantioselectivity of limonene epoxide hydrolase is dependent on the regioselectivity of water attack within the active site. To design epoxide hydrolase variants for enantioselective conversion of the three substrates (**1a**, **2a**, **3a**,

Scheme 1), we performed a series of design calculations for these substrates. Substrates were docked in the active site and placed in a reactive configuration using restraints. This included a hydrogen bond between the leaving oxygen and D101, a close distance between the nucleophilic water and the attacked oxirane carbon, and close to linear orientation of the nucleophilic water, the oxirane carbon, and the epoxide oxygen. Next, the Monte Carlo search algorithm of Rosetta was used to optimize the identity and side chain geometries of amino acids surrounding the active site for either *proRR* or *proSS* attack of the nucleophilic water on the epoxide carbon (Scheme 1). The reactive geometries were essentially defined

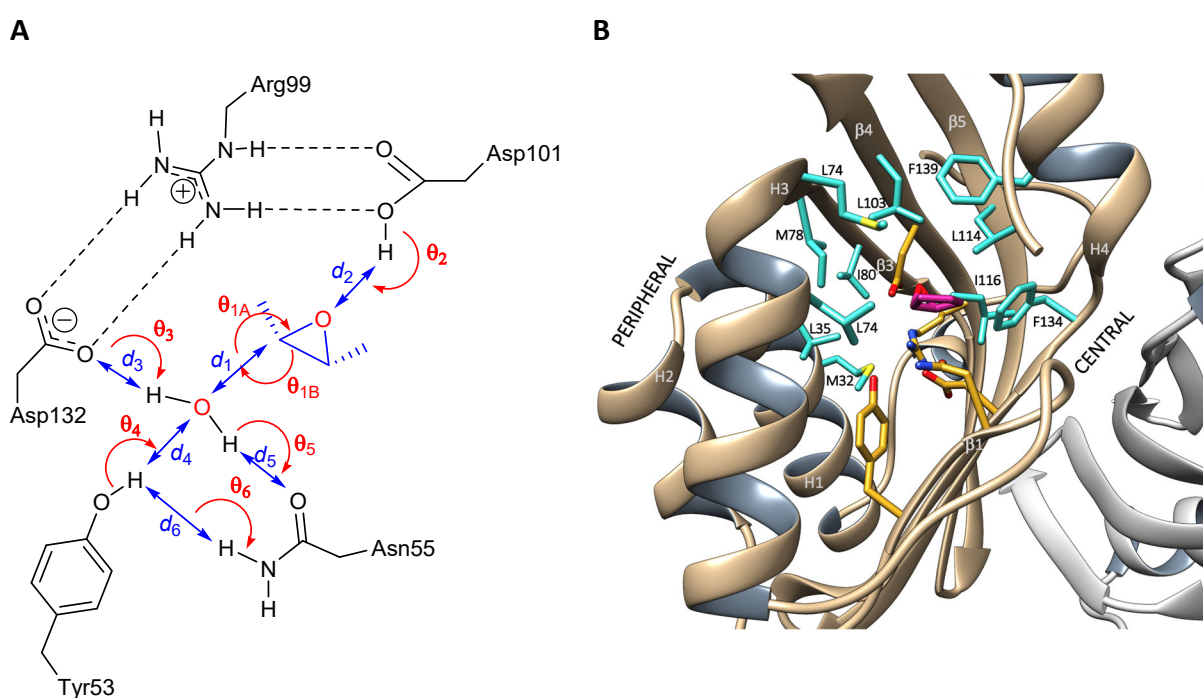


Figure 1. Design of limonene epoxide hydrolases for asymmetric conversion of meso epoxides. **A)** NAC criteria used to predict the enantioselectivity of diol formation from 2,3-butene oxide by MD-simulations. The same criteria were used for all three epoxides. Angles and distances are defined as follows: for the nucleophilic attack angle $\theta_{1A}=128-163^\circ$, $\theta_{1B}=128-163^\circ$ and $d_1=0-3.22 \text{ \AA}$. For the H-bonds $\theta_{2-6}=120-180^\circ$ and $d_{2-6}=0-3.50 \text{ \AA}$. **B)** selection of target positions (cyan) in PDB structure PDB 4R9K. Catalytic residues are shown in yellow, and the substrate in magenta. The targeted amino acid positions are either located in the peripheral structural elements H1 (M32, L35), H3 (L74), and β_3 (M78, I80, V83) which border the *proRR* side of the substrate binding pocket; or in the central region, which forms the *proSS* side of the binding pocket and consists of H4 (F139), β_4 (L103), β_5 (L114, I116) and β_6 (F134). The secondary structure elements are defined as follows: N-loop (residues 1 to 23), H1 (24 to 35), H2 (40 to 46), H3 (64 to 75), H4 (135 to 143), β_1 (52 to 56), β_2 (60 to 62), β_3 (79 to 91), β_4 (94 to 105), β_5 (111 to 123) and β_6 (126 to 133).

as a near attack conformation (Fig. 1A). The explored sequence space was created by targeting 11 selected positions around the active site (Fig. 1B) with randomization to any of the nine hydrophobic residues (AFGILMPVW). This way, Rosetta generated variants forming a low energy complex with the target substrate either in the *proRR* or *proSS* orientation, resulting in thousands of possible *proRR* and *proSS* designs per substrate (Table 1).

These sets of primary Rosetta designs represent mutant libraries enriched in the desired phenotype.

To computationally screen these libraries in an orthogonal manner by the likelihood of showing the correct regioselectivity of water attack, we performed molecular dynamics simulations. From MD trajectories, enantioselectivities were predicted by scoring the fraction of time that the enzyme-substrate complex is in the same reactive *proRR* or *proSS* conformation used during the Rosetta design step (near-attack conformations, NACs, Fig. 1A). Instead of a single or a few long MD simulations, we performed a large number of parallel MD simulations with independently assigned initial atom velocities (HTMI-MD), since this gives more extensive conformational sampling and better agreement with experimental results than single long simulations.^{[41]–[46]} For each of the 15,000 Rosetta designs, at least 5 independently initialized MD runs of 10 ps were performed (Table 1). Designs that passed initial selection rounds were subjected to more MD simulations, up to 80×10 ps and 5×100 ps. NACs were counted on the fly and their frequency was averaged for each design. The ratio between averaged NAC frequencies for *proRR* and *proSS* attack conformations was used as a predictor for enantioselectivity using Equation 1. This MD screening of the primary libraries reduced the number of designs by a factor of 15 to 70 (Table 1).

Table 1. Molecular dynamics screening of designed enzyme variants.

CASCO step	Criteria	No. of designs remaining for substrate		
		1a	2a	3a
Total Rosetta designs		28795	33252	20714
Rosetta designs with a unique sequence		4125	4732	6232
Designs remaining after MD screening				
5×10 ps	e.e. pred >97%	711	990	2504
10×10 ps	e.e. pred >97%	524	723	2205
20×10 ps	e.e. pred >97%	378	587	1631
40×10 ps	e.e. pred >98%	170	314	1284
80×10 ps	e.e. ^{pred} >98% + [NAC] ^{pref} > 5%	142	283	978 ^[a]
5×100 ps	e.e. ^{pred} > 98%	63	181	442 ^[b]
Rosetta designs passing all criteria		1.5%	3.8%	7.1%

^[a] [NAC]^{pref} indicates the NAC frequency for the preferred product enantiomer. Additional criterion [NAC]^{pref} > 10%. ^[b] Additional criterion [NAC]^{pref} > 5%.

It was noticed during in silico screening by MD that a larger fraction of the stilbene oxide designs displayed high NAC frequencies and high predicted enantioselectivities than what was found with cyclopentene and butene oxide designs. As a result, more stilbene oxide designs survived the final selection (7.1 %) than designs for cyclopentene oxide **1a** (1.5 %) and butene oxide **2a** (3.8 %), even though the selection criteria were set significantly more strict for stilbene oxide than for the other substrates (Table 1). This suggests a better occupancy of reactive orientations for stilbene oxide, and more restricted conformations of

the stilbene oxide designs than of the designs with the two smaller substrates, at least during MD simulations.

Visual inspection of the top-ranked designs was carried out to verify that there were no noticeable structural problems that would decrease catalytic activity or enantioselectivity. For every target enantiomer, the variants predicted to have a high e.e.^{pred} for that product were ranked, those with the highest [NAC]^{pref} first, and variants were inspected until five variants were identified for each target enantiomer. It was noticed that for cyclopentene oxide **1a** and butene oxide **2a** there were few designs with obvious errors. In 25 inspected designs for those two substrates there were only five with recognizable structural errors (one was unusually flexible, three had such a spacious active site that reorientation of the substrate seemed likely, and in one mutant there was a new H-bond to the epoxide oxygen, Table S1). In contrast, 10 of the 20 inspected variants for stilbene oxide displayed structural problems. Four designs appeared too spacious (Supporting Information) and 6 designs had a water positioned such that attack on the non-intended carbon atom of the epoxide seemed likely, even though the NAC analysis did not suggest this.

Experimental characterization. The top-5 designs for each product enantiomer of the three substrates were investigated experimentally. Two designs (26A, 45A) were selected twice, both for *proRR* hydrolysis of cyclopentene oxide **1a** and butene oxide **2a** (Table 2). We also included 3 designs predicted to have *proRR* selectivity by molecular dynamics simulation while they were originally designed using Rosetta to have *proSS* selectivity. The 28 new LEH variants were constructed in the thermostable LEH-P template described earlier^[56] by sequential rounds of QuikChange mutagenesis. The use of a thermostable parent enzyme increases the chance that protein function is maintained upon introduction of mutations that may be too destabilizing in a mesostable template and is also reported for directed evolution protocols.^{[57] [59]} The LEH template used here (LEH-P, T_{m,app}=70°C, pdb 4R9K) is not the most stable enzyme from our previous work, since it lacks the disulfide bonds of the most stable variant (T_{m,app}=85°C, pdb 4R9L).^[56]

After sequence verification genes were expressed in *E. coli* Top10 or 10β for enzyme production. All mutants were well expressed and could be isolated by His-tag metal affinity chromatography. Typical yields were 50-150 mg per L of culture. The redesigned LEHs were very stable and could be stored for over 2 years at -80°C without loss of activity. Most variants showed a somewhat lower apparent melting temperature but more stable variants were also found (Table 2). Overall thermostability was well maintained with an average T_{m,app} of the redesigned enzymes of 62.7°C as compared to 70°C for the template LEH-P and 50°C for the wild-type LEH. Activity assays were done by mixing purified enzyme with epoxide and measuring diol formation. Of the obtained 28 designs, 27 showed catalytic activity on the substrate they were designed for. Of these active variants, 66% have an activity that is between 10-164% of the wild-type activity for their respective substrate.

Table 2. predicted and observed activities of computationally redesigned limonene epoxide hydrolase variants.

Enzyme	Mutations ^[a]		Computational		Experimental			
	cyclopentene oxide (1a)		NAC% proRR	NAC% proSS	ee _{pred} (%)	ee _{exp} (%) ^[b]	Act. (%) ^[c]	T _{m,app} (°C)
LEH-P						14	0.12 U/mg	70
Designs for (<i>R,R</i>)- 1b	1A	M32L_M78W_I80V_L103F_F139W	5.63	0.000	100	32	2	63.5
	43A	L74I_L103V_F134Y_F139W	3.86	0.040	98	9	29	67.0
	59A	M78L_L103V_L114G_I116F_F139I	2.59	0.008	99	6	1	67.0
	45A	M32L_L74I_L103V_L114W_I116L_F134G_F139W	2.44	0.000	100	- ^d	<1	-
	46C	M32L_L103V_L114A_I116F_F139W	2.08	0.000	100	46	24	57.0
Designs for (<i>S,S</i>)- 1b	3A	M32L_L35W_I80A_I116V_F139W	0.280	36.5	-98	-74	41	-
	4C	M32L_L35W_I80G_V83I_I116V_F139W	0.208	24.1	-98	-80	55	48.5
	24A	M32A_M78I_I80F_L103I_I116V_F139L	0.248	23.6	-98	-85	15	68.5
	25A	M78I_I80F_L103I_I116V_F139L	0.080	19.4	-99	-84	16	73.5
	26A	L35F_M78F_I80G_I116V_F139W	0.240	19.3	-98	-60	164	45.5
<i>cis</i> -butene oxide (2a)							Act. (%) ^c	
LEH-P						-2	0.23 U/mg	
Designs for (<i>R,R</i>)- 2b	47B	M32L_L35M_M78I_I80L_V83L_I116M_F134Y	24.1	0.176	99	57	3	56.5
	48A ^e	M32L_L35G_M78L_I80W_L103I_F139L	9.00	0.152	97	24	60	56.0
	45A	M32L_L74I_L103V_L114W_I116L_F134G_F139W	5.72	0.112	96	18	1	-
	49A	L103V_L114W_I116L_F134G_F139W	4.14	0.048	98	31	6	59.0
	50A	M32L_L103V_F134Y_F139M	2.90	0.040	97	-3	15	73.0
Designs for (<i>S,S</i>)- 2b	30A	L74W_I80F_L103I_I116V_F139L	0.496	25.2	-96	-73	11	70.5
	31A	L35F_M78F_I80A_I116V_F139W	0.320	23.7	-97	-41	100	56.0
	32A	L35W_L74F_I80G_I116V_F139L	0.080	21.8	-99	-82	15	57.5
	33B	M32L_I80W_L103I_F139L	0.168	21.0	-98	6	1	78.0
	26A	L35F_M78F_I80G_I116V_F139W	0.080	20.4	-99	-60	87	45.5
<i>cis</i> -stilbene oxide (3a)							Act. (%) ^c	
LEH-P						92	0.35 U/mg	
Designs for (<i>R,R</i>)- 3b	51A ^e	M32L_L35G_I80W_L103V_F139W	28.2	0.104	99	91	51	59.5
	52A	M32L_L35M_M78I_L103I_L114M_I116F	24.9	0.144	99	-97	22	61.0
	60A ^e	M32L_L35G_I80W_L103V_F139L	22.7	0.048	100	>99	52	65.0
	61B	M32A_I80V_L103V_L114W_I116V_F134G_F139L	22.1	0.136	99	92	3	66.5
	38A	M32L_M78L_I80V_L103V_F134W_F139L	22.0	0.352	97	>99	11	75.5
Designs for (<i>S,S</i>)- 3b	62A	M32L_I80V_L103V_L114W_I116A_F134G_F139L	0.016	15.5	-100	-	<1	71.0
	63B	M78I_I80L_L103V_L114W_I116V_F134G_F139L	0.240	15.4	-97	40	7	67.0
	64C	M32A_L103V_L114W_I116A_F134G_F139L	0.152	15.3	-98	-	<1	57.5
	41B	M32L_L35M_L103I_L114M_I116F_F139L	0.072	15.0	-99	-95	21	62.0
	65B	L74I_M78F_L103V_L114A_I116V_F134W_F139M	0.008	14.7	-100	-28	2	66.0

^[a] Mutations at the peripheral (*proRR* side) of the substrate binding pocket are marked with a gray background.

Other mutations line the center (*proSS* side) of the substrate binding pocket.

^[b] Positive numbers: (*R,R*)-diol preference; negative numbers: (*S,S*)-diol preference.

^[c] Relative activities expressed in percentage of the activity with the template enzyme (indicated). Data from duplicate measurements with the same enzyme batch.

^[d] -, no activity.

^[e] Variants designed by Rosetta to exhibit (*S,S*)-product selectivity but predicted by HTMI-MD and found experimentally to produce (*R,R*)-diol **3b**.

Chiral analysis of the formed diols revealed that 21 (77%) of the designed active variants showed the enantioselectivity predicted by MD simulations (10% e.e. cutoff, Table 2).

The cyclopentene oxide (**1a**) designs were generated to examine if the design and selection protocol gave comparable results to an earlier study in which 34 LEH variants were tested for the same substrate.^[32] Of the 10 new cyclopentene oxide designs, nine were active and had the predicted enantioselectivity. The inactive design was 45A, which was also selected as a design for butene oxide **2a**, for which it did have a low catalytic activity. The five *proSS* designs produced (*S,S*)-diol **1b** with e.e. values of 56-85%, while lower e.e.'s were obtained with *proRR* designs for (*R,R*)-**1b**. The higher enantioselectivity of *proSS* designs is in agreement with previous observations.^[32]

The butene oxide designs were generated to test the possibility to control regioselectivity of water attack with a very small prochiral epoxide substrate. All 10 variants were active and 8 of them had the predicted enantioselectivity. As with cyclopentene oxide, the *proSS* designs performed better than the *proRR* designs, with the best mutant (32A) providing (*S,S*)-diol **2b** with 77% e.e. The wrong predicted designs had very low enantioselectivity (e.e. of 2-6 %). Thus, for both of the small substrates the MD simulations were an effective tool for predicting LEH enantioselectivity.

For the largest substrate, stilbene oxide (**3a**), 6 of the 8 active variants showed the predicted enantioselectivity. This included two variants (51A, 60A) that were originally designed using Rosetta to have (*S,S*)-**3b** selectivity but for which the HTMI-MD screening predicted preferential formation of (*R,R*)-diol, which was in agreement with what was observed experimentally. In these cases, MD corrected the Rosetta design prediction. On the other hand, the combination of Rosetta and MD for design and prediction of enantioselectivity still gave 2 mismatches between prediction and experiment in case of **3a** designs. Of these, variant 52A was exceptional since it was predicted to give (*R,R*)-diol **3b** whereas experimentally it produced (*S,S*)-diol with high e.e. (97%).

Despite the 2 prediction errors, the results show that highly enantioselective variants could also be designed computationally for stilbene oxide. The thermostable template enzyme had 92% (*R,R*)-diol selectivity, and 4 of the 8 active new variants also displayed very high (*R,R*)-preference (>91% e.e.) whereas 2 other designed variants displayed high (*S,S*)-diol preference (>91% e.e.). The more extensive protein-substrate interactions with a bulky substrate likely result in more restricted reactive conformations of enzyme-substrate complexes, accompanied by high product enantioselectivity. Of the experimentally characterized variants for all three substrates, 3 have a relatively large (> 100 Å³) increase in volume of active site, as calculated from reduced side chain volume of introduced amino acids, and indeed these three variants have low or no catalytic activity (see Supporting Information). When variants with a predicted increase of the active site volume of >100 Å³ would have been removed from the libraries, only the primary Rosetta libraries for *cis*-stilbene oxide would have shrunk (elimination of 178 of the 442 variants listed in Table 1).

Catalytic properties of the best variants. To examine how the use of Rosetta for redesign of LEH towards production of a specific diol enantiomer influences catalytic activity, we examined the kinetic properties (k_{cat} and K_{M}) of the best mutants (Table 3). The LEH variants were produced using 1 L cultures, giving again 50-150 mg purified protein/L broth, which was comparable to the yield of the parent thermostable enzyme. Variants RR8 and SS16 were selected earlier as the best variants from a set of 37 designs for cyclopentene oxide **1a** [32] and were included for comparison (Table 3). The results showed that 46C and RR8 had catalytic rate constants (k_{cat}) for cyclopentene oxide that were comparable to that of the thermostable template enzyme LEH-P. Variants 43A and 24A had lower catalytic constants (2.5-fold and 7-fold respectively). Variant SS16 variant displayed an almost 2-fold higher k_{cat} than the template LEH-P. Thus, catalytic rates were quite well maintained. However, in most cases the K_{M} values were higher (8- to 76-fold) for the redesigned LEHs. This might be due to an increase in the volume of the substrate

Table 3. Kinetic properties of computationally redesigned epoxide hydrolases.

Variant	Designed to produce	Assay substrate	e.e. ^[a] (%)	Preference	k_{cat} (s^{-1}) ^[b]	K_{M} (mM) ^[b]	$k_{\text{cat}}/K_{\text{M}}$ ($\text{M}^{-1} \text{s}^{-1}$)
LEH-P	-	1a	13	(R,R)	0.035±0.004	4.2±0.2	7.9
RR8	(R,R)-1b	1a	85	(R,R)	0.039±0.009	189±20	0.20
46C	(R,R)-1b	1a	34	(R,R)	0.022±0.002	344±17	0.06
43A	(S,S)-1b	1a	8	(R,R)	0.017±0.005 ^[c]	35±4 ^[c]	0.48
SS16	(S,S)-1b	1a	-90	(S,S)	0.062±0.001	54±4	1.14
24A	(S,S)-1b	1a	-85	(S,S)	0.005±0.001	3±0.2	1.9
LEH-P	-	2a	24	(R,R)	0.056±0.010	18±1	3.1
32A	(S,S)-2b	2a	-82	(S,S)	0.063±0.002	225±9	0.28
LEH-P	-	3a	92	(R,R)	0.147±0.012	0.37±0.02	406
60A	(R,R)-3b	3a	>99	(R,R)	0.052±0.006	0.06±0.01	890
52A	(R,R)-3b	3a	-89	(S,S)	0.003±0.001	0.16±0.03	19
41B	(S,S)-3b	3a	-94	(S,S)	0.002±0.001	0.19±0.02	10.5

^[a] Calculated from multiple data points i.e. product enantiomers in different reactions.

^[b] Averages of duplicate measurements with standard deviation.

^[c] Single measurement; margins from average coefficient of variation for **1a** data.

binding cavity, which grew by 37 Å³ in mutant 46C that had a 76-fold increase in K_{M} for **1a**. The exception was 24A, which has a 1.4-fold lower K_{M} . In all designs for **1a**, the specificity constant ($k_{\text{cat}}/K_{\text{M}}$) was reduced (Table 3). The kinetic measurements show that the main cause of the lower catalytic activities of most designs found during initial tests (Table 2) is due to a higher K_{M} , not a lower k_{cat} .

The results for the variants designed to convert the other small substrate, *cis*-butene oxide **2a**, are similar. For this substrate, we examined 32A, which has the highest enantioselectivity for producing (*S,S*)-diol and inverted enantioselectivity in comparison to the template. It showed an insignificant increase in k_{cat} , but the K_{M} was much higher, resulting in a drop in $k_{\text{cat}}/K_{\text{M}}$ in comparison to the template LEH-P.

The results were different for the variants designed to convert the bulky epoxide stilbene oxide **3a** to (*R,R*)- or (*S,S*)-diol. Here, k_{cat} values were lower than with wild-type, whereas K_{M} values were better. The observation that k_{cat} values are lowered in the designs for **3a** whereas NAC percentages during MD simulations (Table 2) were fine indicates that prediction of reactivities across different substrates using such short MD simulations is troublesome. This is not unexpected as MD does not account for energy barriers along reaction coordinates.

The high enantioselectivities obtained for stilbene oxide variants, in comparison to the designs for the two other substrates, are likely due a more restricted conformational freedom in case of the bulky stilbene oxide. The better K_{M} values relative to those with small substrates might be due to the design procedure making the active site too spacious for small substrates, preventing a snug fit with good hydrophobic binding interactions. Stilbene oxide is also more bulky than the natural substrate limonene epoxide and the mutations created enough additional space (ca. 67 Å³ for mutant 60A, calculated from reduced side chain volumes) for tighter binding and a low K_{M} . Consequently, the 3-fold lower k_{cat} of mutant 60A with **3a** was accompanied by a 6-fold better K_{M} , leading to an improved catalytic efficiency in 60A. Furthermore, the tight substrate binding caused the specificity constants ($k_{\text{cat}}/K_{\text{M}}$) to be higher for stilbene oxide **3a** than for cyclopentene oxide **1a** and *cis*-2,3-butene oxide **2a** (Table 3). The improved K_{M} values for **3a** in comparison to wild type was seen with all 3 tested stilbene oxide designs.

Preparative scale conversions. The stilbene oxide enantioselectivities of designs 41B and 52A are higher than reported for other LEH variants tested on this substrate.^{[50] [52]} To examine if these redesigned LEHs could be used in preparative scale conversions, the conversion of *cis*-stilbene oxide **3a** to the (*R,R*)- and (*S,S*)-diols by variants 60A and 41B was examined under different reaction conditions, including varying temperatures and cosolvents (Table 4). Cosolvents were tested because the solubility of the substrates and products in water is low. Even in the presence of 10% dioxane in 50 mM HEPES, pH 8.0, both *cis*-stilbene oxide and the diols were only partially soluble when added at 50 mM. Under these conditions, the *cis*-stilbene oxide remained visible as globular crystals while the (*R,R*)-diol and the (*S,S*)-diol formed needles. The LEH variants 41B and 60A were active in this suspension. Addition of cosolvents 1,4-dioxane and THF and the presence of a biphasic system with a layer of ethyl acetate were studied. No conversion was observed with variant

60A in the biphasic system with ethyl acetate. Addition of 10% dioxane gave the best conversion, yielding up to 78% diol in 44 h at 30 °C. Remarkably, adding THF as cosolvent (similar properties as 1,4-dioxane) gave lower conversion when compared to reaction conditions without cosolvent or with 10% dioxane. The conversion of *cis*-stilbene oxide by variant 60A was further improved (63% to 80%) by increasing the reaction temperature from 30 to 40 °C. The best conditions (10% dioxane and 40 °C) were combined and gave a conversion of 86% and 63% for variants 60A and 41B, respectively. Increasing the dioxane concentration to 15% was beneficial for the *proRR* variant 60A, yielding a conversion of 98%, but drastically decreased conversion of **3a** by the *proSS* variant 41B. The enantiomeric excess of the stilbene diol products was analyzed by chiral HPLC (Fig. S2-S4). For the best conversions the following results were obtained: >99% e.e. for (*R,R*)-stilbene diol (LEH 60A) and 88% e.e. for the (*S,S*)-stilbene diol (LEH 41B).

Table 4. Asymmetric synthesis of stilbene diols by computationally engineered enantioselective epoxide hydrolases.^[a]

	Temp (°C)	Time (h)	Conversion (%)	e.e. (%)
Production of (<i>R,R</i>)-diol by LEH 60A				
HEPES 50 mM	30	44	63	_ ^[b]
10% 1,4-dioxane	30	44	78	-
10% THF	30	44	52	-
At 40 °C	40	44	80	-
10% 1,4-dioxane	40	24	76	-
10% 1,4-dioxane	40	48	86	-
15% 1,4-dioxane	40	48	98	>99% (<i>R,R</i>)
Production of (<i>S,S</i>)-diol by LEH 41B				
HEPES 50 mM	30	44	38	-
10% 1,4-dioxane	40	24	50	-
10% 1,4-dioxane	40	48	63	88% (<i>S,S</i>)
15% 1,4-dioxane	40	48	18	-

^[a] Reaction mixtures (total volume 1 ml) contained 10 mg *cis*-stilbene oxide (final conc. 50 mM, suspension) and 6.37 mg enzyme (final conc. 320 μM) in 50 mM HEPES, pH 8.0.

^[b] -, not determined.

Structural origin of enantioselectivity. In order to provide a structural explanation for the observed mutant enantioselectivities we inspected the Rosetta designed structures and the average HTMI-MD structures of selected mutants. According to these models, the nucleophilic water molecule stays in virtually the same position (Fig. 1, 2). This agrees with the X-ray structure, in which the catalytic water has an unusually low B-factor^[48] indicating a precise orientation due to H-bonds from Tyr53, Asn55, and Asp132. The enantioselectivity of the enzyme is therefore determined by the positioning of substrate relative to this water.

In all of the modelled structures, the positional differences that influence enantioselectivity can globally be described as a sliding motion of the epoxide carbon atoms in front of the nucleophilic water molecule (Fig. 2). Mutations causing the substrate to reside more towards the center of the dimeric enzyme (i.e. near β strands $\beta 4$, $\beta 5$, $\beta 6$, and helix H4, see legend of Fig. 1 for residue numbers) will lead to attack on the (*R*)-configured carbon of the epoxide ring, resulting in an (*S,S*)-diol. *Vice versa*, *proRR* attack will dominate if the substrate is positioned more towards the peripheral side (i.e. at near β strand $\beta 3$ and helices H1 and H3).

In the models of the *proSS* selective variants that convert substrates **1a** and **2a** with high enantioselectivity (e.e. >75%), the dominant substrate orientations are achieved by steric hindrance introduced by mutations at the *proRR* side (e.g. L35W/F, L74W/F, M78F, and I80W/F). These mutations will promote positioning of the substrate more towards the central (*proSS*) side. At the same time, space-creating mutations on the *proSS* side (I116V and F139L) will further increase the preference for (*S,S*)-diol formation. Indeed, designs for substrate **1a** and **2a** carrying both L35W and I116V gave an (*S,S*)-diol preference of > 73% e.e. (Table 2). Furthermore, also in mutant 32A the steric hindrance mutations L35F and L74F on the peripheral (*proRR*) side are accompanied by I116V, leading to (*S,S*)-butanediol formation with the best e.e. of 77%.

The origin of the opposite (*R,R*)-selectivity of variants with substrates **1a** and **2a** can be explained in a similar way. It is likely that increased steric hindrance due to mutations on the central side of the substrate binding cavity (e.g. mutations L114W, I116F/M) encourage positioning of the substrate closer to the peripheral (*proRR*) side of the binding pocket. This is visible in mutant 46C for substrate **1a** (e.g. mutation I116F) and possibly in 47B for substrate **2a** (e.g. mutation I116M). However, such a steric effect is not clear in all predicted *proRR* mutant structures, and some designs indeed show low enantioselectivity (e.g. 45A, 50A). The weakly (*R,R*)-diol selective variant 63B contains mutation I116V, creating space on the central (*proSS*) side, but it is accompanied by L114W, reducing that space. The combined effect of such mutations appears difficult to rationalize in view of effects of side chain interactions and dynamics.

The mutants designed for stilbene oxide **3a** have high product enantioselectivities but the mutations that cause them appear to not only involve steric effects. The majority of the **3a** designs showed (*R,R*)-preference, including variant 63B designed for (*S,S*)-enantioselectivity. Surprisingly, the two most (*S,S*)-diol selective mutants for **3a** (52A, 41B) carried mutation I116F, a mutation that introduces steric hindrance at the central side and in case of substrates **1a** and **2a** favors (*R,R*)-selectivity (see above). The unexpected (*S,S*)-diol preference might be due to π - π interactions between the substrate and the newly introduced aromatic ring of Phe116. The same selectivity by attraction might hold for one of the best (*R,R*) selective mutants: variant 60A (e.e. >99%) carries the I80W mutation at the peripheral (*proRR*) side and no steric hindrance introducing mutation on the *proSS* site. For

the other strongly (*R,R*)-selective mutant, 38A with e.e. > 99%, the introduction of steric hindrance at the *proSS* side (F134W) can explain the improvement in enantioselectivity

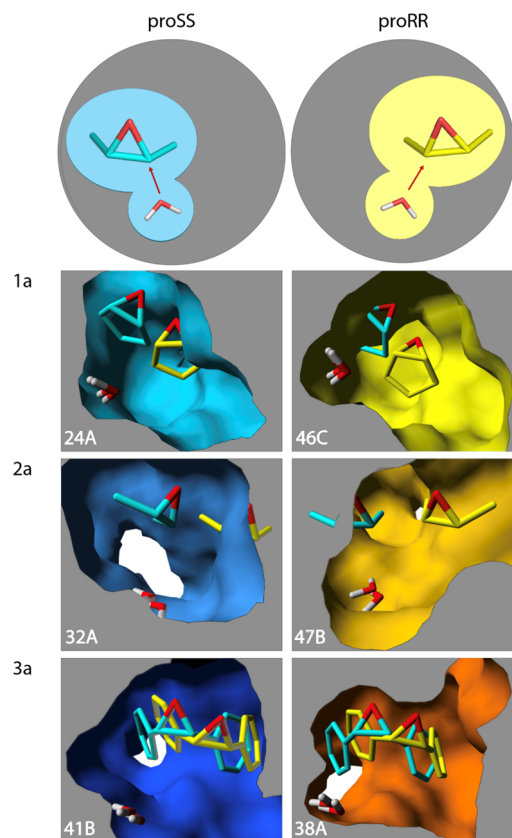


Figure 2. Structural basis of redesigned enantioselectivity. Shown are the active-site cavities of three *proSS*-variants (blue shades, names indicated) and three *proRR* variants (yellow-orange shades). The variants were designed for the substrates indicated at the left of each pair of panels. The reacting water molecules are shown. Hydrogen atoms of substrates are hidden for clarity. In each panel, both the designed enzyme with its substrate is shown, as well the position of the same substrate in the opposite design (substrates in *proSS* designs in cyan, substrates in *proRR* designs in yellow). This shows pronounced differences in substrate positioning and how steric hindrance steers product enantioselectivity.

along the same lines as for substrates **1a** and **2a**. The wild type (e.e. >90% (*R,R*)-diol) also has aromatic functionality with Phe134 at the peripheral (*proRR*) region. Thus, it appears possible that enantiomeric preference with stilbene oxide is partially determined by an influence of attractive π - π interactions on binding orientations of the substrate.

Instead of a translational shift in the position of substrate, the stereoselectivity of LEH variants could also be influenced by rotating the substrate in the binding pocket by 180° along an axis formed by the epoxide oxygen and the spot in between the two epoxide carbon atoms. This would switch the orientation of the epoxide carbon atoms relative to the nucleophilic water. However, such substrate rotations were not observed in any of the

models or during MD simulations. Also attempts by us to dock substrates into productive orientations featuring such a rotation were unsuccessful.

Discussion

Changing enantioselectivity of limonene epoxide hydrolase by directed evolution has been extensively investigated by Reetz, Sun and coworkers, focusing on improving directed evolution strategies including optimizing target positions and positional diversity in libraries.^{[47] [50] [52]} Such well-optimized directed evolution protocols still rely on substantial experimental screening by chiral chromatography, which triggered us to examine if computational methods could be developed to enrich libraries and replace most of the laboratory screening by computational screening of protein libraries.^[60] In silico screening methods have been used earlier to design cocaine hydrolyzing enzymes with improved catalytic efficiency,^{[61] [62]} but also to design libraries of a cytochrome P450 harboring variants with controlled selectivity,^[63] to increase amidase activity in an esterase.^[64]

The results show that design of small sets of mutants with Rosetta and screening by MD simulations could indeed generate LEH variants with desired enantioselectivity. Rosetta design targeted 11 positions at the same time, with a 9 amino acid diversity per position. MD screening was done by multiple ultra-short simulations with on the fly scoring of reactive conformations and allowed to screen of thousands of Rosetta designs. This CASCO protocol^[32] reduced the size of the library required, and for each substrate only 5 variants per desired enantiomer were used to find mutants with enhanced enantioselectivity. In case of cyclopentene oxide **1a**, directed evolution^{[47] [50] - [52]} and previous computational design^[32] gave (*S,S*)-selective LEH variants producing diol with similar e.e. as found here (60%-95% e.e., Table 2).

A comparison of predicted enantioselectivities as calculated from NAC percentages (Table 2) shows that there is good overall agreement only in qualitative terms, i.e. (*R,R*)- or (*S,S*)-diol stereopreference was correctly predicted for 77% of the variants using multiple short MD simulations with independent initialization. On the other hand, within a set of 5 designs, experimental activities for individual variants did not correlate with their computed NAC%, showing that the MD simulations as performed here do not provide quantitative information on catalytic rates. Note, however, that rates shown in Table 2 are strongly influenced by both K_m values and do not reflect k_{cat} . Furthermore, in a broader sense, for each of the three substrates examined, the set of designs that gave the highest NAC% in MD also showed the highest average activity. Thus, on average the *proRR* designs for **3a** were on average more active than its *proSS* designs and also give the highest NAC percentages for **3a**. For the other two substrates *proSS* designs were more active and gave higher NAC%.

In this study, the best enantioselectivities were clearly obtained with stilbene oxide **3a**. Whereas most designs produced (*R,R*)-diol **3b** (with e.e. >99% for 2 variants), enantiocomplementary mutants yielding (*S,S*)-**3b** were also found, with a highest e.e. of

97%. With butene oxide **1a**; [51] [52] the obtained variants showed high and modest enantioselectivity in the production of (*R,R*)- and (*S,S*)-diols, respectively. The LEH mutants found for stilbene oxide could be used to produce enantiopure (*R,R*)-diol and (*S,S*)-diol at preparative scale, indicating the potential of this approach to generate a practically useful biocatalyst. Ring opening of stilbene oxide was tested earlier using mutants optimized on cyclopentene oxide and cyclohexene oxide, which resulted in highly (*R,R*) selective variants (with e.e. 99%) but only modest (*S,S*)-selective variants (e.e. 44%). [65] The results suggest that the likelihood of obtaining high enantioselectivity is much better with the bulkier stilbene oxide than with the smaller substrates, with the two phenyl rings of stilbene oxide offering more opportunities for steric hindrance as well as Van der Waals and π - π interactions.

Similar to what we found with redesigned aspartases catalyzing asymmetric hydroamination of acrylates, [33] we observed that the lower activity found with most designs for **1a** and **2a** as compared to the template (Table 2) was not due to a reduced k_{cat} but due to an increase in K_M (Table 3). From a practical point of view, this is less disturbing than the opposite, because for reasons of process economy preparative-scale applications must be carried out at high substrate concentrations anyway. Recently, Sun et al. [65] attributed low activity of LEH variants obtained by directed evolution to misalignment of the nucleophilic water, epoxide carbon and epoxide oxygen for an S_N2 reaction, caused by increased flexibility in the active site, but without reporting kinetic data. This explanation likely does not hold for the computationally redesigned enzyme studied by us because such a misalignment would reduce k_{cat} and probably also K_M , which is not what we observe, except for some stilbene oxide designs (Table 3). Proper alignment of the nucleophilic water and reacting substrate atoms for S_N2 is one of the constraints in the Rosetta design process and a NAC criterion during MD screening.

Surprisingly, four of the observed enantioselectivities with stilbene oxide **3a** were opposite to the Rosetta design target (Table 2). Two of these were corrected in MD simulations. To understand how these incorrect designs could emerge, we examined sequences and structures of Rosetta-optimized enzyme-substrate complexes of all substrates (see above). For substrates **1a** and **2a**, the observed enantioselectivities could be explained by the combination of steric hindrance and space-creating mutations from the peripheral and central side of the substrate binding pocket, acting together to steer the substrate in a *proRR* or *proSS* binding mode. This did not hold for several of the mutants designed for stilbene oxide **3a**. For both variants that are strongly *proSS* selective for **3a** the only bulk introducing mutation is I116F, which would be expected to introduce steric hindrance at the central side of the cavity and thus stimulate *proRR* selectivity. Also, for one of the two most *proRR* selective variants (60A, e.e. > 99%) the I80W mutation would be expected to decrease space at the peripheral side and thereby stimulate *proSS* selectivity. Thus, it appears that effects of mutations on bulkiness are poorly related to

stereopreference in case of aromatic substrate **3a**, suggesting that electronic effects that are not well modeled, such as π - π interactions, dominate over steric factors in determining substrate positioning or reactivity.

Standard computational design and MD simulations do not explicitly account for π - π interactions to save computation time. MD simulations can give more realistic results in cases where aromatic interactions play a role when the force field is adapted with an additional noncovalent interaction term.^[66] Whether the main effect of π - π interactions is on binding, conformational dynamics, or reactivity of bound substrates is unclear at present. Recently, Zaugg et al.^[67] investigated the origin of enantioselectivity of *Aspergillus niger* epoxide hydrolase in the conversion of the chiral substrate phenyl glycidyl ether (PGE). Molecular dynamics simulations suggested that the protein does not differentiate enantiomers based on binding mode, and free energy calculations did not show significant differences between (*R*)- and (*S*)-PGE binding either. The authors suggested that the enantioselectivity is due to kinetic differences. For such an α/β -hydrolase fold epoxide hydrolase a computational analysis is more complicated due to the multiplicity of reaction pathways and chemical steps. Earlier, Lau et al.^[68] studied murine epoxide hydrolase with (1*S*,2*S*)-*trans*-2-methylstyrene oxide using ab initio and density functional calculations, and suggested the importance of interactions between the substrate's phenyl group and aromatic residues in the binding pocket. Moreover, Lind and Himo^[69] published the reaction mechanism of a soluble epoxide hydrolase (StEH1) converting styrene oxide. They proposed coplanarity of the oxirane C1-C2 carbons with the substrate's phenyl substituent, and π - π interactions between this phenyl group and a histidine and phenylalanine to be important for the stabilization of the transition state and for the selectivity of the enzyme. Rinaldi et al.^[70] proposed that substrate-dependent LEH regioselectivity is related to reorganization of the active site towards each ligand. Based on QM/MM calculations, they confirmed that substrate-specific LEH regioselectivity is due to both conformational and electronic parameters.

We conclude that computational design and MD simulations are well able to predict and screen enantioselectivity of LEH variants in case of small aliphatic substrates. Whereas highly selective variants for production of aromatic diols can be obtained, prediction accuracy is lower. In view of the effect of interactions involving aromatic groups on epoxide hydrolase enantioselectivity, rapid scoring methods that more accurately include effects of π - π interactions appear necessary to further improve computational screening of LEH variants acting on aromatic substrates.

Experimental Section and Computational Methods

Materials. The meso-epoxides and their corresponding diols, oligonucleotides for mutagenesis, organic solvents and glycerol were purchased from Sigma-Aldrich. Restriction enzymes and PfuUltra Hotstart PCR Master Mix were obtained from New England Biolabs and Agilent,

respectively. Ni-NTA resin was purchased from GE Healthcare Life Sciences. SYPRO orange was obtained from ThermoFisher Scientific. Complete protease inhibitor cocktail tablets were bought from Roche. Media components were obtained from Difco (BD Biosciences).

Computational design. To design LEH variants for production of highly enantioenriched diols from meso-epoxides the previously developed CASCO strategy was used with only minor modifications.^[32] The X-ray structure of the wild type LEH (Protein Databank 1NWW) was used for computational design. Eleven positions around the active site (M32, L35, L74, M78, I80, V83, L103, L114, I116, F134 and F139) were selected to mutate simultaneously to any of the nine hydrophobic residues (AFGILMPVW). Each of the three substrates was docked in the enzyme active site, either in a *proRR* or *proSS* conformation using Rosetta enzyme design^{[34] [71]} Catalytically productive binding modes were defined using a constraint file as previously.^[32] This geometric description of how the substrate should be bound included the obligation to form H-bonds between the epoxide oxygen and D101 and between the nucleophilic water and D132, Y53, and N55. Furthermore, the water oxygen had to be close (1.8 Å) to the attacked carbon atom while the angle of nucleophilic attack (*i.e.* from water oxygen, attacked carbon, and epoxide oxygen) should be close to 180°. Another constraint was that the distance between the nucleophilic water and the non-attacked epoxide carbon atom should be > 3.8 Å. To hinder undesired substrate-binding orientations, a bulky residue (W, F or Y) was introduced at one of the eleven target positions, since this may reduce binding poses not contributing to the desired selectivity.^[32] Rosetta enzyme design was used to simultaneously mutate the remaining ten residues to any of the nine hydrophobic residues and sequence-conformational space was searched for substrate-bound structures with low energy and a catalytically productive binding mode.

High-throughput-multiple independent MD simulations (HTMI-MD) were used for in silico screening of the generated libraries and to rank the primary designs with an orthogonal tool (Wijma et al., 2014). Independent initialization of multiple trajectories increases the conformational space sampled by molecular dynamics and reduces the computational cost of the screening step as compared to a single long MD run.^[46] The reactivity and selectivity of each mutant were predicted by scoring the fraction of snapshots in which the enzyme-substrate complex is in a *proRR* or *proSS* near-attack conformation (NAC). The latter are defined by geometric constraints (Fig. 1), which should be fulfilled for a reaction to become feasible. The geometric criteria for *proRR* and *proSS* attack conformations were as defined using published quantum mechanical modelling.^[72] The ratio of *proRR* and *proSS* NAC frequencies were considered to reflect regioselectivity of attack and thus product enantioselectivity according to Equation 1,

$$e.e. = \frac{([NAC]^{proRR} - [NAC]^{proSS})}{([NAC]^{proRR} + [NAC]^{proSS})} \quad (\text{Equation 1})$$

in which e.e. is the predicted enantiomeric excess, $[NAC]^{proRR}$ and $[NAC]^{proSS}$ indicate the fraction of snapshots in which the enzyme-substrate complex is in a *proRR* or *proSS* conformation, respectively. Positive values indicate predicted (*R,R*)-diol preference, negative values (*S,S*) selectivity.

The only modification from the existing procedures are listed in this paragraph. More design calculations were done than previously, and also more seeds per MD simulation.^[32] For the current study, approximate 25 thousand design calculations were run per target substrate (Table 1), which is two times more than previously. Like earlier,^[32] a stepwise scheme to rank the variants was adopted in which variants were eliminated as soon as they failed a criterion (Table 1). The selection criteria

were based on e.e.^{pred} and [NAC] values for the preferred enantiomer. The criteria differed per target substrate and are listed in Table 1. For each designed new variant 20-80 independently started MD simulations of 10 ps were used (previously maximally 20 MD simulations). For the final variants also 5 MD simulations of 100 ps were performed.

Finally, the best ranked mutants were visually inspected. For each of the targeted product enantiomers, only variants predicted to have a high enantioselectivity were visually inspected, starting with those variants that were predicted to have the highest fractions of NACs. The main reasons for elimination of designs were a too spacious active site cavity or an orientation of the substrate relative to the water that seemed in disagreement with the predicted enantioselectivity (Table S1). Only 15 of the 45 inspected designs were eliminated at the stage of visual inspection. Furthermore, no mutations were added at this stage even though this is common in the field.^[73] As a result, the visual inspection only took a few hours.

Mutagenesis, expression and purification. For the expression of LEH and variants thereof in *E. coli* a pBAD based expression vector was used. This vector contained the gene of the thermostable variant LEH-P with an N-terminal hexa-histidine tag.^[56] The computationally designed variants of LEH-P were constructed by QuikChange site-directed mutagenesis using *Pfu* Ultra Hotstart PCR Mastermix (Agilent), combining multiple mutations in a single primer when possible, and omitting sequence verification between individual mutation steps. PCR reactions, transformations, plating and final sequencing were done in microtiter plate format.^[74] The obtained plasmids were used to transform chemically competent *E. coli* Top10 or *E. coli* NEB10 β cells (Thermo Fischer Scientific). For expression, cells were grown overnight in 5 mL Luria-Bertani broth at 37°C. All cultures were supplemented with 50 $\mu\text{g mL}^{-1}$ ampicillin. The resulting culture was used to inoculate 500 mL Terrific Broth medium and incubated at 37°C and 135 rpm. When an OD₆₀₀ of 0.6 was reached, expression was induced by adding 0.04% (w/v) arabinose and growth was continued at 30°C and 135 rpm. After 24 h the cells were harvested by centrifugation at 6,700 *g* and 4°C for 15 min.

For protein isolation, cells were resuspended in 50 mM HEPES buffer, pH 8, containing 500 mM NaCl (3 mL per gram of cells) and half of a protease inhibitor cocktail tablet to prevent proteolysis (Roche Applied Science). After sonication (60 \times 10 s with 20 s intervals, Labsonic M), the extract was centrifuged at 35,200 *g* and 4°C for 1 h. The supernatant was collected and the enzyme of interest was purified by gravity-flow affinity chromatography under native conditions using Ni-NTA agarose resin (Thermo Fischer Scientific). The protein concentration of the collected fractions was determined by a Bradford assay, and selected fractions were desalted by Econo-Pac 10DG desalting columns (Bio-Rad). The purity of the prepared enzymes (yield 50-150 mg per L TB medium) was analyzed by SDS-PAGE. Enzymes were stored at -80°C until further use.

Catalytic properties. Chiral chromatography was used to determine the enantioselective hydrolysis of three *meso*-epoxides (*cis*-2,3-butene oxide, cyclopentene oxide and *cis*-stilbene oxide) to chiral diols. In case of *cis*-2,3-butene oxide and *cis*-stilbene oxide, 5 mg of purified enzyme was added to 50 mM substrate (virtual concentration of the suspension) in 50 mM HEPES pH 8 (800 μL total reaction volume). After incubation of the reaction mixture at 30°C for 1 h, 500 μL of 5 M NaCl was added and the samples were extracted three times with 600 μL ethyl acetate. The combined extracts containing diols were dried by adding anhydrous sodium sulfate, concentrated under vacuum, and dissolved in 100 μL ethyl acetate. For cyclopentene oxide, reactions were done in a similar way after which 250 mg K₂CO₃ was added to the reaction mixture followed by extraction for

two times by 600 μ L of n-butanol. The combined extract was dried, concentrated under vacuum and resuspended in 100 μ L n-butanol.

In case of *cis*-2,3-butene oxide and cyclopentene oxide, chiral analysis was carried out by injecting 2 μ L of the extracts into an Agilent gas chromatograph equipped with a flame ionization detector and a Hydrodex β -TBDAC column (Aurora Borealis, initial temp. 40°C, 10°C/min to 150°C, hold 20 min). For *cis*-stilbene oxide and its diols, samples were analyzed by HPLC on a Luxcellulose-3 column (Phenomenex, Utrecht, the Netherlands) with heptane/2-propanol (90/10) as the mobile phase (flow rate 1 ml/min, detection at 254 nm). Samples from preparative scale reactions with *cis*-stilbene oxide and the diols were also analyzed by HPLC on a Chiralpak AS-H column (Daicel Corp, Illkirch, France) with n-hexane/2-propanol (90:10 (v/v) as the mobile phase (1 ml/min, detection at 254 nm). Enantiomeric excess (e.e.) values were calculated from concentrations of the (*R,R*) and (*S,S*) product enantiomers.

To obtain steady-state kinetic parameters, initial velocities at different substrate concentrations were determined and fitted with the Michaelis-Menten equation.

Determination of the apparent melting temperature. The ThermoFluor assay was used to determine the apparent melting temperatures (T_m^{app}) of the purified enzyme variants.^[75] This method is based on monitoring the change in fluorescence of Sypro Orange dye during the thermal unfolding of a protein. The dye binds to the unfolded and exposed hydrophobic protein core, increasing its fluorescence signal. The assays were done as described before.^[26]

Synthesis of stilbene diols. Reaction mixtures (total volume 1 mL) contained 10 mg of *cis*-stilbene oxide (final conc. 50 mM, suspension) and 6.37 mg of enzyme (final conc. 320 μ M) in 50 mM HEPES, pH 8.0, and were incubated (at 30 or 40 °C, 135 rpm) for 48 h. Substrate and products were extracted three times by 4 mL ethyl acetate, dried over MgSO₄ and filtered. The solvent was removed by a rotary evaporator. The residue was analyzed by ¹H-NMR for conversion and by chiral HPLC to determine the enantiomeric excess. Chiral HPLC was used as described above.

Author's contribution

HJW designed the mutants; HA constructed, isolated and characterized the mutants, HA, PJ and DIC measured catalytic activities; HJW, HA, and EB interpreted the mutants; DIC and MT performed preparative experiments; HA, HJW, EB, MT and DBJ wrote the paper; HJW and DBJ supervised the work.

Acknowledgment

This research was supported by the Dutch Ministry of Economic Affairs through BE-Basic [FS07.001] and the European Union's Horizon 2020 Programme (Marie Curie Action - ITN Es-Cat) under GA No. 722610.

References

- [1] D. Muñoz Solano, P. Hoyos, M. J. Hernáiz, A. R. Alcántara, J. M. Sánchez-Montero, *Biores. Technol.* **2012**, *115*, 196-207.
- [2] B. M. Nestl, S. C. Hammer, B. A. Nebel, B. Hauer, *Angew. Chem. Int. Ed.* **2014**, *53*, 3070-3095.

- [3] R. N. Patel, *Bioorg Med Chem.* **2018**, *26*, 1252-1274.
- [4] M. T. Reetz, *J. Am. Chem. Soc.* **2013**, *135*, 12480-12496.
- [5] J. M. Choi, S. S. Han, H. S. Kim, *Biotechnol. Adv.* **2015**, *33*, 1443-54.
- [6] R. A. Sheldon, J. M. Woodley, *Chem. Rev.* **2018**, *118*, 801-838.
- [7] H. Sun, H. Zhang, E. L. Ang, H. Zhao, *Bioorg. Med^[7]. Chem.* **2018**, *26*, 1275-1284
- [8] N. J. Turner, *Nat. Chem. Biol.* **2009**, *5*, 567-573.
- [9] U. T. Bornscheuer, G. W. Huisman, R.J. Kazlauskas, S. Lutz, J.C. Moore, K. Robins, *Nature* **2012**, *485*, 185-194.
- [10] J. Lalonde, *Curr. Opin. Biotechnol.* **2016**, *42*, 152-158.
- [11] F. H. Arnold, *Angew. Chem. Int. Ed.* **2019**, *58*, 14420-14426.
- [12] M. T. Reetz, *Angew. Chem. Int. Ed.* **2011**, *50*, 138-174.
- [13] M. T. Reetz, *Methods Mol. Biol.* **2003**, *230*, 259-282.
- [14] Z. Sun, Y. Wikmark, J. E. Bäckvall, M. T. Reetz, *Chemistry* **2016**, *22*, 5046-54.
- [15] G. J. Williams, T. Woodhall, L. M. Farnsworth, A. Nelson, A. Berry, *J. Am. Chem. Soc.* **2006**, *128*, 16238-16247.
- [16] J. G. van Leeuwen, H. J. Wijma, R. J. Floor, J. M. van der Laan, D. B. Janssen, *ChemBiochem* **2012**, *13*, 137-48.
- [17] K. Steiner, H. Schwab, *Comput. Struct. Biotechnol. J.* **2012**, *2*, e201209010.
- [18] J. F. Chaparro-Riggers, K.M. Polizzi, A. S. Bommarius, *Biotechnol. J.* **2007**, *2*, 180-191.
- [19] R. Das, D. Baker, *Annu. Rev. Biochem.* **2008**, *77*, 363-82.
- [20] G. Kiss, N. Çelebi-Ölçüm, R. Moretti, D. Baker, K.N. Houk, *Angew. Chem. Int. Ed.* **2013**, *52*, 5700-5725.
- [21] J. Damborsky, J. Brezovsky, *Curr. Opin. Chem. Biol.* **2014**, *19*, 8-16
- [22] H. K. Privett, G. Kiss, T. M. Lee, R. Blomberg, R. A. Chica, L.M. Thomas, D. Hilvert, K. N. Houk, S. L. Mayo, *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 3790-3795.
- [23] N. S. Scrutton, *IUCrJ.* **2016**, *4*, 5-6.
- [24] R. J. Fox, S. C. Davis, E. C. Mundorff, L. M. Newman, V. Gavrilovic, S. K. Ma, L. M. Chung, C. Ching, S. Tam, S. Muley, J. Grate, J. Gruber, J. C. Whitman, R. A. Sheldon, G. W. Huisman, *Nat. Biotechnol.* **2007**, *25*, 338-344.
- [25] Z. Wu, S. B. J. Kan, R. D. Lewis, B. J. Wittmann, F. H. Arnold, *Proc. Natl. Acad. Sci.* **2019**, *116*, 8852-8858.
- [26] H. J. Wijma, R. J. Floor, P. A. Jekel, D. Baker, S. J. Marrink, D. B. Janssen, *Protein Eng. Des. Sel.* **2014**, *27*, 49-58.
- [27] H. Arabnejad, M. Dal Lago, P.A. Jekel, R.J. Floor, A.W.H. Thunnissen, A.C. Terwisscha van Scheltinga, H.J. Wijma, D.B. Janssen, *Protein Eng. Des. Sel.* **2017**, *30*, 173-187.
- [28] R.J. Floor, H.J. Wijma, D.I. Colpa, A. Ramos-Silva, P.A. Jekel, W. Szymański, B.L. Feringa, S.J. Marrink, D.B. Janssen, *ChemBiochem*, 2014, **15**, 1660-1672.
- [29] R. J. Floor, H. J. Wijma, P. A. Jekel, A. C. Terwisscha van Scheltinga, B. W. Dijkstra, D. B. Janssen, *Proteins* **2015**, *83*, 940-951.
- [30] A. Goldenzweig, M. Goldsmith, S. E. Hill, O. Gertman, P. Laurino, Y. Ashani, R. L. Lieberman, *Molec. Cell*, **2016**, *63*, 337-346.
- [31] D. Bednar, K. Beerens, E. Sebestova, J. Bendl, S. Khare, R. Chaloupkova, J. Damborsky, *PLoS Comp. Biol.* **2015**, *11*, e1004556.
- [32] H. J. Wijma, R. J. Floor, S. Bjelic, S. J. Marrink, D. Baker, D. B. Janssen, *Angew. Chem. Int. Ed.* **2015**, *54*, 3726-3730.
- [33] R. Li, H. J. Wijma, L. Song, Y. Cui, M. Otzen, Y. Tian, J. Du, T. Li, D. Niu, Y. Chen, J. Feng, J. Han, H. Chen, Y. Tao, D. B. Janssen, W. Wu. *Nature Chem. Biol.* **2018**, *14*, 664-670.
- [34] A. Leaver-Fay, M. Tyka, S. M. Lewis, O. F. Lange, J. Thompson, R. Jacak, K. Kaufman, P. D. Renfrew, C. A. Smith, W. Sheffler, I. W. Davis, S. Cooper, A. Treuille, D. J. Mandell, F. Richter, Y. E. Ban, S. J. Fleishman, J. E. Corn, D. E. Kim, S. Lyskov, M. Berrondo, S. Mentzer, Z. Popović, J. J.

- Havranek, J. Karanicolas, R. Das, J. Meiler, T. Kortemme, J. J. Gray, B. Kuhlman, D. Baker, P. Bradley, *Methods Enzymol.* **2011**, *487*, 545-574.
- [35] T. C. Bruice. *Acc. Chem. Res.* **2002**, *3*, 139-148.
- [36] T. C. Bruice, S. Benkovic. *Biochemistry* **2000**, *21*, 6267-6274.
- [37] A. Warshel, P.K. Sharma, M. Kato, Y. Xiang, H. Liu, M.H.M. Olsson. *Chem. Rev.* **2006**, *8*, 3210-3235.
- [38] G. Kiss, D. Rothlisberger, D. Baker, K.N. Houk. *Protein Sci.* **2010**, *9*, 1760-1773.
- [39] G. Kiss, V.S. Pande, K.N. Houk. *Methods Enzymol.* **2013**, *523*, 145-170.
- [40] H.J. Wijma. In: A.S. Svendsen (ed.), *Understanding enzymes; function, design, engineering and analysis*. Taylor and Francis, New York, pp 805-833.
- [41] L.S.D. Caves, J.D. Evanseck, M. Karplus. *Protein Sci.* **1998**, *3*, 649-666.
- [42] L. Monticelli, E.J. Sorin, D. P. Tieleman, V.S. Pande, G. Colombo. *J. Comput. Chem.* **2008**, *11*, 1740-1752.
- [43] S. Genheden, U. Ryde. *J. Comput. Chem.* **2010**, *4*, 837-846.
- [44] S. Genheden, U. Ryde. *J. Comput. Chem.* **2011**, *2*, 187-195.
- [45] S. Genheden, C. Diehl, M. Akke, U. Ryde. *J. Chem. Theory Comput.* **2010**, *7*, 2176-2190.
- [46] H. J. Wijma, S. J. Marrink, D. B. Janssen, *Chem. Inf. Model.* **2014**, *54*, 2079-92.
- [47] H. Zheng, M. T. Reetz, *J. Am. Chem. Soc.* **2010**, *132*, 15744-15751.
- [48] M. Arand, B. M. Hallberg, J. Zou, T. Bergfors, F. Oesch, M. J. van der Werf, J. A. de Bont, T. A. Jones, S. L. Mowbray, *EMBO J.* **2003**, *22*, 2583-2592.
- [49] K. H. Hopmann, B. M. Hallberg, F. Himo, *J. Am. Chem. Soc.* **2005**, *127*, 14339-14347.
- [50] Z. Sun, R. Lonsdale, X. D. Kong, J. H. Xu, J. Zhou, M. T. Reetz, *Angew. Chem. Int. Ed.* **2015**, *54*, 12410-12415.
- [51] Z. Sun, R. Lonsdale, G. Li, M. T. Reetz, *ChemBiochem* **2016a**, *17*, 1865-1872.
- [52] Z. Sun, R. Lonsdale, L. Wu, G. Li, A. Li, J. Wang, M. T. Reetz, *ACS Catalysis* **2016**, *6*, 1590-1597.
- [53] A. Li, C. G. Acevedo-Rocha, Z. Sun, T. Cox, J. L. Xu, M. T. Reetz, *ChemBiochem* **2018**, *19*, 221-228.
- [54] G. Li, H. Zhang, Z. Sun, X. Liu, M. T. Reetz, *ACS Catalysis*, **2016**, *6*, 3679-3687.
- [55] G. Li, J. B. Wang, M.T. Reetz, *Bioorg Med Chem.* **2018**, *26*, 1241-1251.
- [56] R.J. Floor, H.J. Wijma, P. A. Jekel, A.C. Terwisscha van Scheltinga, B.W. Dijkstra, D.B. Janssen, *Proteins*, **2015**, *83*, 940-951.
- [57] J. Bloom, S. Labthavikul, C. Otey, F.H. Arnold, *Proc. Natl. Acad. Sci. USA*, **2006**, *103*, 5869-5874.
- [58] W. Besenmatter, P. Kast, D. Hilvert, *Protein Struct. Funct. Bioinform.*, 2007, *66*, 500-506.
- [59] N. Tokuriki, D.S. Tawfik, *Curr. Opin. Struct. Biol.*, **2009**, *19*, 596-604.
- [60] B. A. Amrein, F. Steffen-Munsberg, I. Szeler, M. Purg, Y. Kulkarni, S. C. Kamerlin, *IUCrJ.* **2017**, *50*-64.
- [61] D. Gao, H. Cho, W. Yang, Y. Pan, G. Yang, H. Tai, C. Zhan, *Angew. Chem. Int. Ed.* **2006**, *45*, 653-657.
- [62] F. Zheng, W. Yang, M.C. Ko, J. Liu, H. Cho, D. Gao, M. Tong, H.H. Tai, J.H. Woods, C.G. Zhan, *J. Am. Chem. Soc.* **2008**, *130*, 12148-12155.
- [63] A. Seifert, M. Antonovici, B. Hauer, J. Pleiss, *ChemBiochem* **2011**, *12*, 1346-1351.
- [64] P.O. Syren, P. Hendil-Forsell, L. Aumailley, W. Besenmatter, F. Gounine, A. Svendsen, M. Martinelle, K. Hult, *ChemBiochem* **2012**, *13*, 645-648.
- [65] Z. Sun, L. Wu, M. Bocola, H. C. S. Chan, R. Lonsdale, X. D. Kong, S. Yuan, J. Zhou, M. T. Reetz, *J. Am. Chem. Soc.* **2018**, *140*, 310-318.
- [66] H. Yuki, Y. Tanaka, M. Hata, H. Ishikawa, S. Neya, T. J. Comput. Chem. **2007**, *28*, 1091-1099.
- [67] J. Zaugg, Y. Gumulya, M. Bodén, A. E. Mark, A. K. Malde, *J. Chem. Inf. Model.* **2018**, *58*, 630-640.
- [68] Y. E. Lau, E. Z. Newby, T. C. Bruice, *J. Am. Chem. Soc.* **2001**, *123*, 3350-3357.
- [69] M. E. S. Lind, F. Himo, *ACS Catalysis* **2016**, *6*, 8145-8155.

- [70] S. Rinaldi, M. W. van der Kamp, K. E. Ranaghan, A. J. Mulholland, G. Colombo, *ACS Catalysis* **2018**, *8*, 5698-5707.
- [71] F. Richter, A. Leaver-Fay, S. D. Khare, S. Bjelic, D. Baker, *PLoS One* **2011**, *6*, e19230.
- [72] M. E. S. Lind, H. Fahmi, *Angew. Chem. Int. Ed.* **2013**, *52*, 4563-4567.
- [73] L. G. Nivón, S. Bjelic, C. King, D. Baker, *Proteins* **2014**, *82*, 858-866.
- [74] M. L. J. Fürst, C. Martin, N. Lončar, M. W. Fraaije, *Methods Enzymol.* **2018**, *608*, 151-87.
- [75] U. B. Ericsson, B. M. Hallberg, G. T. Detitta, N. Dekker, P. Nordlund, *Anal. Biochem.* **2006**, *357*, 289-98.