JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

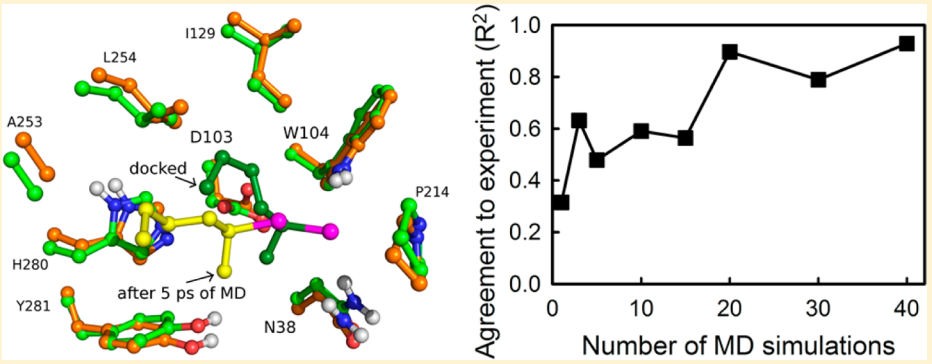
Computationally Efficient and Accurate Enantioselectivity Modeling by Clusters of Molecular Dynamics Simulations

Hein J. Wijma,[†] Siewert J. Marrink,[†,‡] and Dick B. Janssen*[,†]

[†] Groningen Biomolecular Sciences and Biotechnology Institute, University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands

[‡] Zernike Institute for Advanced Materials, University of Groningen, 9747 AG Groningen, The Netherlands

Supporting Information



ABSTRACT: Computational approaches could decrease the need for the laborious high-throughput experimental screening that is often required to improve enzymes by mutagenesis. Here, we report that using multiple short molecular dynamics (MD) simulations makes it possible to accurately model enantioselectivity for large numbers of enzyme−substrate combinations at low computational costs. We chose four different haloalkane dehalogenases as model systems because of the availability of a large set of experimental data on the enantioselective conversion of 45 different substrates. To model the enantioselectivity, we quantified the frequency of occurrence of catalytically productive conformations (near attack conformations) for pairs of enantiomers during MD simulations. We found that the angle of nucleophilic attack that leads to carbon−halogen bond cleavage was a critical variable that limited the occurrence of productive conformations; enantiomers for which this angle reached values close to 180° were preferentially converted. A cluster of 20−40 very short (10 ps) MD simulations allowed adequate conformational sampling and resulted in much better agreement to experimental enantioselectivities than single long MD simulations (22 ns), while the computational costs were 50−100 fold lower. With single long MD simulations, the dynamics of enzyme−substrate complexes remained confined to a conformational subspace that rarely changed significantly, whereas with multiple short MD simulations a larger diversity of conformations of enzyme−substrate complexes was observed.

INTRODUCTION

Enzymes can provide excellent regioselective and stereoselective catalysis and thereby produce enantiopure precursors for the synthesis of pharmaceuticals or other fine chemicals.[1,2] For such enzyme-catalyzed processes to be economically feasible, the enzyme should be sufficiently stable,[1−6] catalytically active, and stereoselective[2,7−10] with the target substrate. In many cases, protein engineering can be used to improve the stereoselectivity or substrate scope of an existing enzyme.[11−16] However, often tens of thousands of mutant variants need to be experimentally screened to obtain sufficiently improved variants.[9,17−19] Such high-throughput screening is costly and time consuming. Therefore, a key challenge[2] in protein engineering is the development of computational methods that are able to accurately predict[20−24] the substrate specificity and enantioselectivity of large numbers of enzyme variants without requiring such an amount of computational resources that experiments remain faster or cheaper. The availability of such *in silico* screening methods would decrease the need for experimental screening and thereby decrease costs while speeding up the engineering of improved enzyme variants.

A useful method to explain and predict the catalytic rates of enzymes is to monitor a series of distances and angles between the substrate and catalytically important atoms in the enzyme active site during a molecular dynamics (MD) simulation.[25−31] By such an analysis of both well and poorly converted enzyme substrates, it typically turns out that the best converted substrates are bound more often in a catalytically productive orientation. The latter can be quantified, for example, by computing an average distance between two reacting atoms, which is observed to be larger for the unproductively bound

Received: February 27, 2014

   dx.doi.org/10.1021/ci500126x | *J. Chem. Inf. Model.* XXXX, XXX, XXX−XXX

nonconverted substrates than for the productively bound substrates that are well converted.[25−30]

A similar but theoretically more justified method is the near attack conformation (NAC) approach.[32−36] The NAC is normally defined as a conformation having angles between the reacting atoms within 20° of those in the quantum mechanically modeled transition state and distances between the reacting atoms of less than the sum of their van der Waals radii.[32,37] Because NACs are conformations that are close to the transition state of the reaction, the enzyme−substrate complex necessarily must go through these conformations for a reaction to occur.[32,37] There is some disagreement about whether an enzyme stabilizes the transition state of a reaction by binding the substrate selectively in a NAC[38] or whether the high frequencies of NACs found during MD simulations of enzyme−substrate complexes are a side effect of transition state stabilization through a different mechanism, for example, electrostatic preorganization.[39,40] This difference in opinion is irrelevant for the modeling described below because in both interpretations a high frequency of NACs occurring among the conformations sampled by an MD simulation corresponds to a high reaction rate of the enzyme−substrate complex.

An important challenge for *in silico* screening of catalytic activity is the sampling and time scale problem in enzyme MD simulations. With existing computer facilities, MD simulations are mostly limited to the nanosecond and microsecond time scale,[41] while enzyme catalytic cycles occur on a time scale of milliseconds to seconds. Because conformational changes in enzymes occur on similarly long time scales, longer MD simulations give a more representative sampling of catalytically productive and unproductive conformations of the enzyme−substrate complex. However, if it is necessary to analyze tens of thousands of enzyme variants, long MD simulations would be too computationally expensive to compete with experimental screening.

A different approach that can lead to a more complete sampling of protein conformational space is to carry out multiple short MD simulations that start with the same three-dimensional protein structure but with different random initial atom velocities[42−44] or with a different water sphere surrounding the protein.[45] The resulting MD simulations are observed to sample different regions of conformational space while during a single MD simulation a jump from one conformational subspace (around a local minimum in the conformational landscape) to another region in conformational space rarely occurs because the energy barrier for such a transition is high.[42−44] Therefore, by using several shorter MD simulations, a larger conformational space of a protein can be sampled than during a single long MD simulation that takes equal computational resources.[42−44] Such potentially more effective MD approaches are rarely explored for enzymes; single long MD simulations are commonly used.[25−37,46] The use of up to five independent long MD simulations (5−10 ns) instead of one long MD simulation was reported for P450 enzymes, which have high flexibility and thus may adopt quite diverse conformations.[23,47] A quantitative analysis of the extent to which the sampling of enzyme−substrate complexes can be improved by independent MD simulations has not been reported.

In the work described here, we systematically investigated the differences in conformational sampling of enzyme−substrate complexes using single long simulations versus multiple shorter MD simulations. Enantioselectivity (E) was selected as the catalytic property that had to be reproduced by the MD simulations. With an achiral catalyst, the energy barrier for reaction of either enantiomer of the substrate is the same. Thus, in an enantioselective enzyme, the difference in reactivity between the substrate enantiomers is determined solely by their interactions with the chiral active site. This makes enantioselectivity prediction ideal for testing how completely the MD simulations sample the possible conformations of the enzyme−substrate complexes.

Haloalkane dehalogenases were selected as the model system because a large set of experimental data on enantioselective conversion of structurally diverse substrates (Figures 1 and 2) is

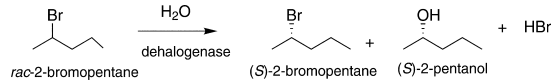


Figure 1. Example of an enantioselective conversion by a dehalogenase.

available.[48−50] Haloalkane dehalogenases also have potential biotechnological relevance because they may catalyze the kinetic resolution of racemic mixtures to produce enantiopure compounds.[51] The four haloalkane dehalogenases for which both an X-ray structure and experimental data on enantioselectivity have been reported are DhlA from *Xanthobacter autotrophicus*,[52] DhaA from *Rhodococcus rhodochrous*,[53] DbjA from *Bradyrhizobium japonicum*,[54] and LinB from *Sphingomonas paucimobilis*.[55] Data on enantioselectivity and conversion rates are available for 45 substrates of these enzymes. This includes over 80 different combinations of enzyme and racemic or prochiral substrate, with E varying from 1 to >200. For 32 of these enzyme−substrate combinations, it has experimentally been determined which enantiomer is preferentially converted. The other combinations, for which it is unknown which enantiomer is preferentially converted, are still useful for detecting undersampling of catalytically productive conformations during the MD simulations (see the Materials and Methods section).

Mechanistically, all these dehalogenases operate by nucleophilic attack of an aspartate carboxyl oxygen atom on the halogenated carbon atom of the substrate, with the halogen acting as a leaving group (step 2 with rate k_2 in Scheme 1).[56,57] The resulting alkyl-enzyme intermediate is hydrolyzed by attack of a water molecule (with rate k_3, Scheme 1) on the carbonyl carbon of the ester group. It has been proposed that the initial nucleophilic attack is most important for the enantioselectivity, with the substrate enantiomer that undergoes nucleophilic attack faster being preferentially converted.[50] Therefore, this step was modeled to predict E for all substrates.

We show here that the enantioselectivities calculated from the ratio of NACs observed during MD simulations with the (R) and (S) substrates agree well with the experimental results for all four dehalogenases, if at least 20 short MD simulations are used. Such a cluster of short MD simulations provided much better sampling of conformations than a single longer MD simulation. The computational costs of 20−40 of such short MD simulations are low enough to enable high-throughput *in silico* screening of enzyme variants. We also investigated the source of enantioselectivity in dehalogenases, which is currently under debate.[48,50]

dx.doi.org/10.1021/ci500126x | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

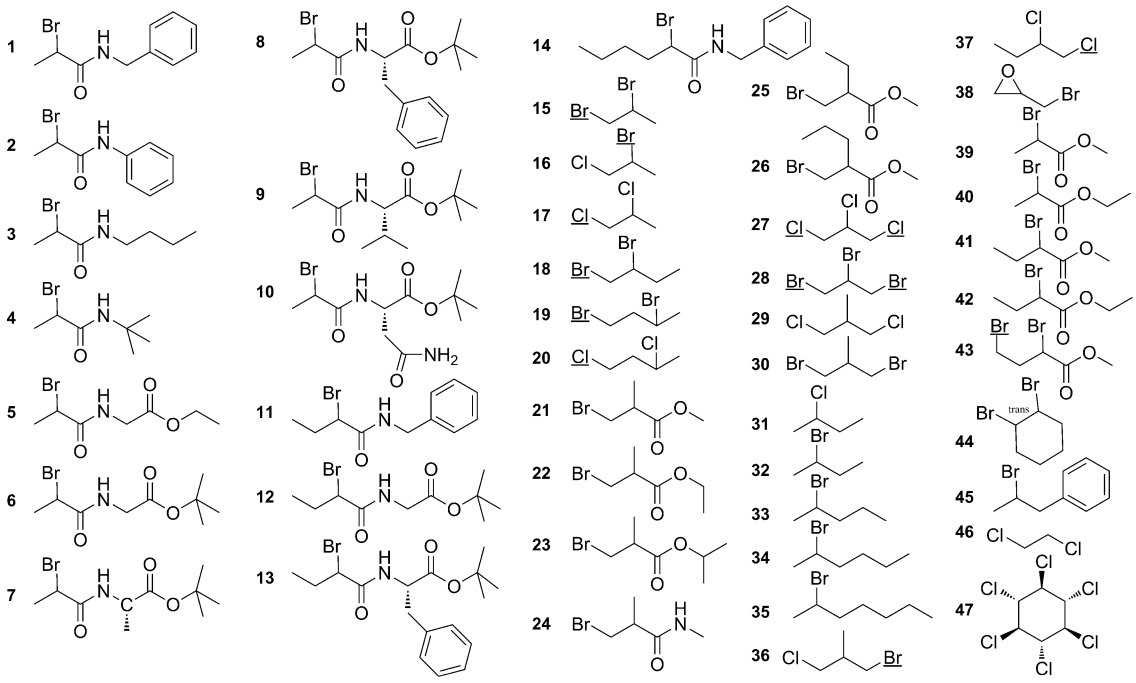


Figure 2. Dehalogenase substrates examined in this study. Substrates 1−26 and 31−45 are chiral. Substrates 7−10 and 13 occur as diastereomers. Substrates 27−30 are prochiral and product chirality depends on the orientation in which the substrate is bound during nucleophilic attack. Substrates 46−47 remain achiral upon conversion. If nonequivalent halogen atoms are present in the substrate, the leaving group halogen is underlined.

Scheme 1. General Catalytic Scheme for a Dehalogenase[a]

$$E + R\text{-}X \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} E.R\text{-}X \xrightarrow{k_2} E\text{-}R.X^- \xrightarrow{k_3} E + R\text{-}OH + X^-$$

substrate binding; alkyl-enzyme formation; product release

[a]E is the enzyme. R-X is the halogenated substrate. E.R-X is the enzyme−substrate complex. E-R.X$^-$ is the covalent alkyl-enzyme intermediate with halide bound. R−OH is the produced alcohol.

MATERIALS AND METHODS

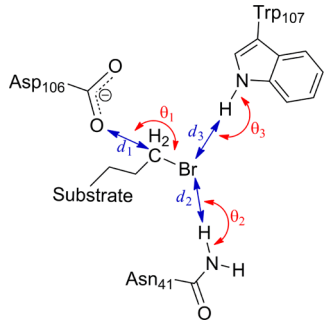
Enzymes, Substrates, and Docking. The following PDB files were selected for simulations: 1MJ5[58] (LinB, resolution 0.95 Å), 1BN6[59] (DhaA, 1.50 Å), 3A2M[50] (DbjA, 1.84 Å), and 1B6G[60] (DhlA, 1.15 Å). DbjA was simulated as its physically relevant dimer.[61] Protonation states of the Asp, Glu, His, Lys, Tyr, and Cys residues at pH 8, the typical buffer pH for activity assays, were initially predicted with Yasara (www.yasara.org).[62] Subsequently, all these protonation states were visually inspected and if needed corrected. All water molecules present in the X-ray structure were preserved with the exception of water molecules in the active site that could prevent the binding of the substrate. Because the DhaA variant for which enantioselectivity had been determined differs by three mutations from the DhaA of which the X-ray structure is available (V172A/I209L/G222A), the structure of DhaA was modeled based on 1BN6 as described elsewhere.[15] A set of variants of LinB[63] that differed in catalytic activity for substrate 47 were modeled similarly[64] based on 1MJ5.

To prepare the structures of the 92 substrates [Figure 2, 45 × 2 (pro)-chiral substrates and 2 achiral substrates] for docking and subsequent MD simulations, a semiautomated procedure was applied. With Chem3D (CambridgeSoft Corporation), a three-dimensional structure of the substrates was created. A script under Yasara automatically carried out a geometry optimization with semiempirical (AM1) quantum mechanics[65] with implicit solvation. Partial charges for docking and MD simulations were calculated with the AM1-BCC method,[66] which results in an accuracy similar to the RESP procedure at a fraction of the computational cost. If multiple halogens were present in the substrate, the correct leaving group halogen was set manually. Docking was carried out with Autodock 4[67] with 9990 separate docking runs of 25,000 energy evaluations each; these parameters made it possible to obtain a larger number of the substrates (Figure 2) in a NAC or close to a NAC than parameters that we used earlier.[48] For example, with these docking parameters, the known substrate 38 could be bound in a catalytically productive conformation, while with the earlier parameters, both enantiomers of substrate 38 were docked with their epoxide oxygen placed between Asn and Trp (Scheme 2), a position where the halogen atom of the substrate should be located for catalysis to proceed. After docking, for every substrate enantiomer, a favorable orientation for catalysis was automatically selected for MD simulation based on a geometric ranking of the 9,990 different docked orientations. If one or more substrate orientations had been docked in a NAC (defined in the caption of Scheme 2), the one that had the lowest score $D \equiv 2d_1 + d_2 + d_3$ (Scheme 2) was selected. Otherwise, simply, the conformation with the lowest D was selected. This protocol does not force substrates to dock into a catalytically productive orientation but merely selects from all docked orientations the pose that is most suitable for catalysis.

MD Simulations. The MD simulations were executed and analyzed automatically. A home-scripted protocol carried out the setup of the simulation box, pressure and temperature control, collection of geometric data during the MD simulation, and conversion of raw data to results (e.g., NAC percentages, average angles). The docked enzyme−substrate complexes

dx.doi.org/10.1021/ci500126x | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

Scheme 2. Geometric Criteria for Catalytically Productive Conformations[a]



[a]The enzyme–substrate complex was defined to be in a NAC when all geometric criteria were simultaneously met (d_1 < 3.41 Å, d_2 < 3.50 Å, d_3 < 3.50 Å, θ_1> 157°, θ_2 > 120°, θ_3 > 120°). The criteria were described previously.[34−36,48] Throughout this work, these geometries were never imposed as constraints; it was merely monitored whether NAC conformations were present. Residue numbers correspond to DhaA. For LinB, the corresponding residues are, respectively, Asn38, Asp108, and Trp109; for DbjA, they are Asn38, Asp103, and Trp104; for DhlA, they are Trp175, Asp124, and Trp125.

were positioned in a rectangular simulation cell with distances between the protein and the periodic boundaries of at least 7.5 Å. Water (TIP3P) and salt ions to neutralize the cell and provide ionic strength (NaCl 0.5%) were positioned by Yasara, with modifications made in an automated script to prevent salt ions and newly added water molecules from being positioned inside the enzyme while water molecules from the X-ray structures were all preserved. Prior to the MD simulations, an energy minimization was carried out as described elsewhere[48] to remove steric clashes and conformational stress. The MD simulations were run under Yasara with a leapfrog integration scheme and a Berendsen thermostat[68,69] under pressure control. Hydrogen atoms were constrained with LINCS[70] and SETTLE[71] algorithms. The time step was 1.33 fs with the nonbonded interactions list updated every three time steps. Long range (>7.86 Å) electrostatic interactions were calculated with a particle mesh Ewald algorithm with fourth degree B-spline functions.[72] The force field for the simulations was Amber ff03[73] or Yamber3.[68] Unless specifically mentioned otherwise in the text, the results are those with Yamber3.

To carry out multiple independent MD simulations per enzyme–substrate complex, simulations were started with different initial atom velocities according to a Maxwell–Boltzmann distribution.[42] At the start of the longer MD simulations, the temperature was gradually increased from 5 to 298 K during 30 ps, after which the simulation was allowed to equilibrate for 1970 ps before the production phase simulation of 5000 or 20,000 ps. Thus, the total simulation time of the single trajectories was 7 or 22 ns. For the shorter MD simulations, the system was gradually heated during 3 ps and allowed to equilibrate for 2 ps (the temperature equilibrated in this time frame), while the production phase was 5 ps, amounting to a total of 10 ps simulation time. To analyze RMSD, RMSF, and energy, snapshots were obtained every 25 ps for the longer MD simulations. The three distances (d) and three angles (θ) depicted in Scheme 2 were sampled on the fly every 100 fs for the long MD simulations and every 20 fs for the short MD simulations.

Enantioselectivity Modeling. On the basis of the general catalytic cycle of dehalogenases (Scheme 1), it can be derived using standard methods[48,74] that eqs 1–5 apply. In these equations, K_d is the dissociation constant of the substrate, k_{cat} is the catalytic rate at saturating substrate concentrations, K_M is the Michaelis constant, P and D superscripts denote the preferred and disfavored enantiomers, and k_1 to k_3 are the rates depicted in Scheme 1. Equation 6 follows from eq 4 and 5 if K_d is equal for both enantiomers. Enantioselectivity for substrates was predicted with eq 7, in which [NAC] is the percentage of NACs observed during MD simulation of the enzyme–substrate complex. As further evaluated in the discussion, eq 7 was derived assuming that K_d is approximately equal for the different enantiomers (eq 6) and that the relative rate of step k_2 is modeled satisfactory by the NAC approach. The highest experimental enantioselectivities were reported[48−50] as >200, which reflects that conversion of the disfavored enantiomer was not detectable. Therefore, if the calculated $[NAC]^P/[NAC]^D$ exceeded 200, it was rounded down to 200 to enable comparison to experimental results. Predictions of E were considered to be correct within error if the predicted values reproduced the experimental results within a factor 10 (see Discussion for a justification of this error margin). To assess which fraction of the E predictions were correct, we used data for substrates for which it is experimentally clear which enantiomer is preferentially converted as well as data for substrates for which the E was close to one (i.e., ≤ 1.5). Substrates for which it is unclear which enantiomer is preferentially converted were only used to test for false negatives (see below).

$$k_{cat} = \frac{k_2 k_3}{k_2 + k_3} \quad (1)$$

$$K_M = K_d \times \frac{k_3}{k_2 + k_3} \quad (2)$$

$$K_d = \frac{k_{-1}}{k_1} \quad (3)$$

$$k_{cat}/K_M = \frac{k_2}{K_d} \quad (4)$$

$$E = \frac{{k_{cat}}^P/{K_M}^P}{{k_{cat}}^D/{K_M}^D} \quad (5)$$

$$E = \frac{{k_2}^P}{{k_2}^D} \quad (6)$$

$$E^{predicted} = \frac{[NAC]^P}{[NAC]^D} \quad (7)$$

It was also evaluated which fraction of the substrate enantiomers in the data set are known to be experimentally converted, whereas the MD simulations did not produce catalytically productive orientations. Such false negative results can be caused by insufficient sampling of the conformational space of the enzyme–substrate complex. Two criteria were used to quantify such false negatives. If not a single NAC was found during MD simulations of an enzyme complexed with its known preferred substrate enantiomer, this combination was a false negative. Furthermore, the absence of NACs during MD

dx.doi.org/10.1021/ci500126x | J. Chem. Inf. Model. XXXX, XXX, XXX–XXX

Table 1. Results of a Single Long and 40 Short MD Simulations for Modeling Enantioselectivity of LinB[a]

	single long MD simulations (1 × 22 ns)			40 short MD simulations (average of 40 × 10 ps)			experimental enantioselectivity[48,50]	
	NAC fraction (%)			NAC fraction (%)				
substrate	(R)	(S)	$E^{predicted}$	(R)	(S)	$E^{predicted}$	E	preference
1	3.76	0.08	47.0	8.98	0.08	112	67	(R)
2	0.03	0.16	0.2	0.29	0.28	1	28	(R)
3	1.99	0.35	5.7	5.53	0.15	36.9	74	(R)
5	1.84	0.43	4.3	11.91	0.85	14	10	(R)
6	3.87	0.62	6.2	16.82	1.77	9.5	11	(R)
11	2.67	0	>200	3.28	0	>200	>200	(R)
15	0	0	NP	0.12	1.01	0.1	1	none
19	0.07	0.06	1.2	1.95	1.15	1.7	5	unknown
20	2.33	0.07	33.3	3.31	2.54	1.3	3	unknown
21	0.87	0	>200	0.68	0.55	1.2	3	unknown
22	0.05	0.37	0.1	0	1.23	0	1	none
31	1.95	1.77	1.1	6.61	6.36	1	1	none
32	0.26	8.01	0.0	3.47	1.87	1.9	2	unknown
35	0	0	NP	8	0.63	12.7	3	unknown
36	1.1	0.12	9.2	0.62	0.42	1.5	2	unknown
38	0	0	NP	0	0	NP	1	none
43	0	2.07	0.0	9.36	0.16	58.5	2	unknown
44	0	0	NP	0	0.04	0	3	unknown
45	0	0	NP	0	0.46	0	2	unknown

[a]The predicted E was calculated from [NAC-(R)]/[NAC-(S)]. NP: No prediction possible because no NACs were observed for both the (R) and (S) enantiomer of the substrate. If the preferred enantiomer was unknown, it was assumed to be the (R) enantiomer for calculation of the predicted E. Gray backgrounds reveal which modeling results are in disagreement with experimental results, either because $E^{predicted}$ is incorrect or because of false negative results (see Materials and Methods for criteria).

simulation for either the (R) or the (S) enantiomer was classified as a false negative if that substrate was converted with an $E \leq 10$. Such a low enantioselectivity implies that the enzyme converts both substrate enantiomers.

RESULTS

Long Single MD Simulations versus Shorter Multiple MD Simulations. Enantioselectivities were modeled by MD simulations of enzyme−substrate complexes that were obtained by docking substrates into the dehalogenase X-ray structures. Using geometric filters, the catalytically most productive poses from 9990 docking runs per substrate enantiomer were selected. Due to this large number of dockings per substrate, it was possible to find also rare substrate orientations. In this manner, we obtained potentially catalytically productive poses for the large majority of the substrates, even for many substrate enantiomers that are poorly converted when tested experimentally (Tables S1−S4, Supporting Information). Of the 175 enzyme−substrate complexes selected for subsequent MD simulation, 156 were found to satisfy the narrow geometric limits of a NAC (89%). It was concluded that hand-docking was not required to start the MD simulations with catalytically relevant substrate orientations.

Initially, single long MD simulations (22 ns) were carried out for 38 complexes of LinB with different substrate enantiomers (Table 1) and 46 complexes of DhaA with different enantiomers (DhlA and DbjA were addressed later because of the computational expense, see below). The resulting predictions for E, calculated from the ratio of NACs for a pair of enantiomers during the MD simulation (eq 7), agreed with experimental data for the majority of the enzyme−substrate combinations (Table 2). However, 20% of the single long MD simulations produced false negatives, i.e., trajectories in which no NACs were observed even though a substrate enantiomer is known to be converted from experimental data. Such cases gave a wrongly predicted very high (>200) or a very low E (see Table 1 for examples), which negatively influenced the correlation between predictions and experiments.

Table 2. Agreement between Experimentally Obtained and Modeled Results for Different Sampling Regimes[a]

	predictions that (dis)agree with experimental data		
time regime[b]	correct E prediction (%)	false negatives (%)	Pearson's R^2 correlation between experimental and modeled E
1 × 22 ns	72	20	0.51
3 × 7 ns	78	13	0.51
40 × 10 ps	67	14	0.93

[a]Results are those for the experimental data of LinB and DhaA with converted substrates for which either E was equal to one or for which the preferred enantiomer was experimentally determined. The total number of substrates for which this applied was 18 (see Tables S1 and S2, Supporting Information). [b]Number of MD simulations per enzyme−substrate complex followed by the length of the individual simulations.

It was investigated whether the false negatives were due to undersampling of substrate conformations and if additional simulations would sample different regions of conformational space. Specifically, two additional 7 ns MD simulations were started with the same structure but with a different set of initial atom velocities, which were generated using different random number seeds (seed B and seed C; seed A was the first 7 ns of the original 22 ns MD simulation). For individual enzyme−

dx.doi.org/10.1021/ci500126x | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

substrate combinations, shortening simulations from 22 to 7 ns hardly changed the observed fraction of NACs. The only two exceptions were LinB with (*S*)-5 (4-fold increase, Table 3) and DhaA with (*R*)-15 (10-fold increase, Table S1, Supporting Information). More importantly, the results showed that the assignment of different initial velocities can produce significantly different NAC frequencies during parallel simulations and reduced the occurrence of false negatives when the results were averaged (Table 3, Table S1, Supporting Information). For example, for (*R*)-43, which is a known substrate of LinB, no NACs were observed with seed A for up to 22 ns, while for the other two seeds, NAC percentages of 9.8 and 8.4 were found (Table 1). The use of three shorter simulations decreased the percentage of false negatives (from 20 to 13%) and produced better agreement with experimental data than the 22 ns simulation (Table 2, Figure 3A,B).

Table 3. Differences in Frequency of Catalytically Relevant Orientations during Nanosecond Time Scale MD Simulations of LinB Resulting from Different Set of Initial Atom Velocities

	NACs in MD simulation (%)			
	seed A		seed B	seed C
substrate	20 ns	5 ns	5 ns	5 ns
examples with greater than 4-fold difference between 5 and 20 ns MD simulations[a]				
(*R*)-2	0.03	0.04	**0.28**	0.02
(*S*)-5	0.43	**0.10**	**0.07**	**0.04**
(*R*)-15	0.00	0.00	**0.44**	**0.94**
(*S*)-19	0.06	0.21	0.20	**0.94**
(*S*)-20	0.07	0.15	**1.31**	**1.71**
(*R*)-21	0.87	1.40	**0.00**	**0.00**
(*R*)-22	0.05	0.18	**0.00**	**0.00**
(*S*)-22	0.37	0.62	**0.08**	**0.06**
(*R*)-31	1.95	4.84	2.72	**0.02**
(*R*)-32	0.26	0.51	**2.18**	**2.33**
(*R*)-35	0.00	0.00	0.00	**0.09**
(*R*)-36	1.10	0.82	**0.15**	0.53
(*S*)-36	0.12	0.35	**0.69**	0.40
(*R*)-43	0.00	0.00	**9.81**	**8.36**
(*R*)-45	0.00	0.00	**0.07**	0.00
(*S*)-45	0.00	0.00	0.00	**0.91**
all examples with less than 4-fold difference between 5 and 20 ns MD simulations				
(*R*)-1	3.76	4.51	3.33	3.23
(*S*)-1	0.08	0.04	0.11	0.05
(*S*)-2	0.16	0.12	0.24	0.08
(*R*)-3	1.99	1.59	1.65	2.05
(*S*)-3	0.35	0.44	0.27	0.10
(*R*)-5	1.84	4.60	7.27	6.69
(*R*)-6	3.87	6.57	10.68	8.42
(*S*)-6	0.62	0.62	0.34	0.33
(*R*)-11	2.67	3.24	0.91	1.63
(*S*)-11	0.00	0.00	0.01	0.00
(*S*)-15	0.00	0.00	0.00	0.00
(*R*)-19	0.07	0.15	0.13	0.02
(*R*)-20	2.33	3.32	1.63	4.87
(*S*)-21	0.00	0.00	0.00	0.00
(*S*)-31	1.77	7.07	3.87	5.38
(*S*)-32	8.01	10.34	2.45	15.71
(*S*)-35	0.00	0.00	0.00	0.00
(*R*)-38	0.00	0.00	0.00	0.00
(*S*)-38	0.00	0.00	0.00	0.00
(*S*)-43	2.07	2.55	2.72	7.76
(pro*RS*)-44	0.00	0.00	0.00	0.00
(pro*SR*)-44	0.00	0.00	0.00	0.00

[a]NAC percentages for the 5 ns MD simulations that differ more than a factor 4 from the result of the 20 ns MD simulation are indicated in **bold**. The abbreviation proRS stands for prochiral substrate in a binding mode that results in a (1*R*,2*S*) product. The tested substrates are known to be (enantioselectively) converted by LinB.

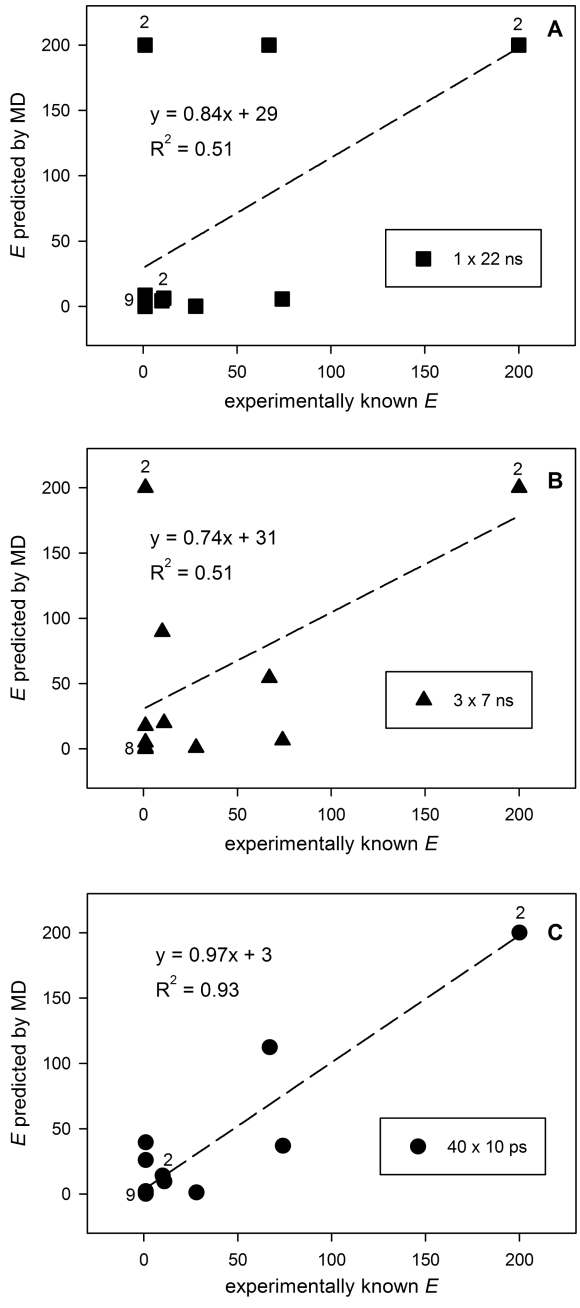


Figure 3. Improved correlation between predicted and experimentally determined *E* by switching from a single long simulation to multiple shorter MD simulations. Predicted values are for LinB and DhaA and obtained from a single 22 ns trajectory (A) or by averaging NAC ratios from three 7 ns (B) or 40 10 ps simulations (C). The numbers in the graph indicate the number of overlapping data points. The dashed line is the fit. For further information, see Table 2.

dx.doi.org/10.1021/ci500126x | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

We next examined what causes the differences between trajectories that share the same starting structure but are attributed with different sets of initial atom velocities. Structural inspection revealed that different substrate orientations are

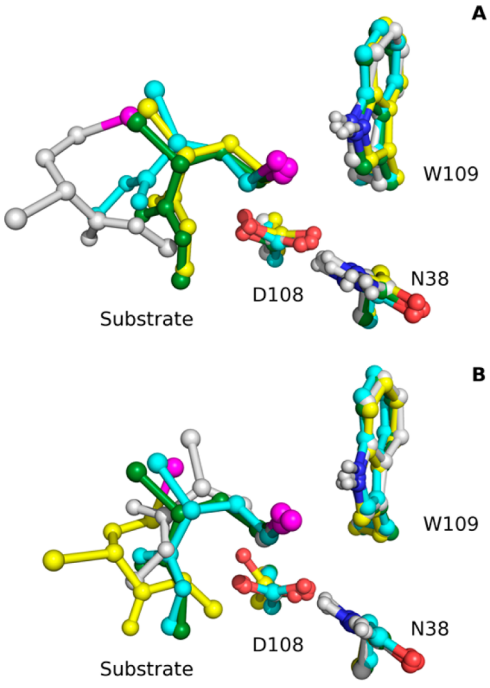


Figure 4. Different substrate orientations observed in three MD simulations of LinB with (*R*)-43 with different initial atom velocities. (A) Structures after 7 ns of MD simulation. (B) Structures after 5 ps of MD simulation. Only the substrate and the catalytic residues are shown (see Figure 6 for other surrounding residues). Carbon atoms in green correspond to the starting structure, gray spheres correspond to seed A, light blue spheres correspond to seed B, and yellow spheres correspond to seed C. The magenta sphere represents the leaving group halide, red spheres represent oxygen atoms, dark blue spheres show nitrogen, and hydrogen atoms are in light gray.

present after equilibration of the MD simulations (0−2 ns), which are subsequently maintained and cause differences in the occurrence of NACs. For example, in the three independent simulations of LinB with (*R*)-43 described above, which exhibited rather different NAC frequencies, different substrate orientations were observed (Figure 4A). Thus, the additional MD simulations improved the sampling of the conformational space accessed by the enzyme−substrate complexes.

Multiple Ultrashort MD Simulations for Sampling Near Attack Conformations. We next investigated whether it was possible to further improve conformational sampling by using larger numbers of MD simulations of only 10 ps. With 40 of these independently initialized simulations per enzyme−substrate complex, a significantly better correlation (Table 2) to experimental data (Figure 3C) was obtained than with 3 orders of magnitude longer single (Figure 3A) or triple MD simulations (Figure 3B). The correlation coefficients are relatively high (0.51−0.93) because the highest experimentally determined enantioselectivities were reported as >200, which were rounded down to 200 to enable analysis (Table 2). The improved correlation obtained with a large number of very short MD simulations is due to better predictions for enzyme−substrate combinations with low enantioselecitivity (compare Figure 3A,B to C). The fraction of false negatives (14%) and the fraction of *E* that were correctly predicted (67%) are within error equal to those obtained from a triple 7 ns MD simulation, but at much lower computational costs.

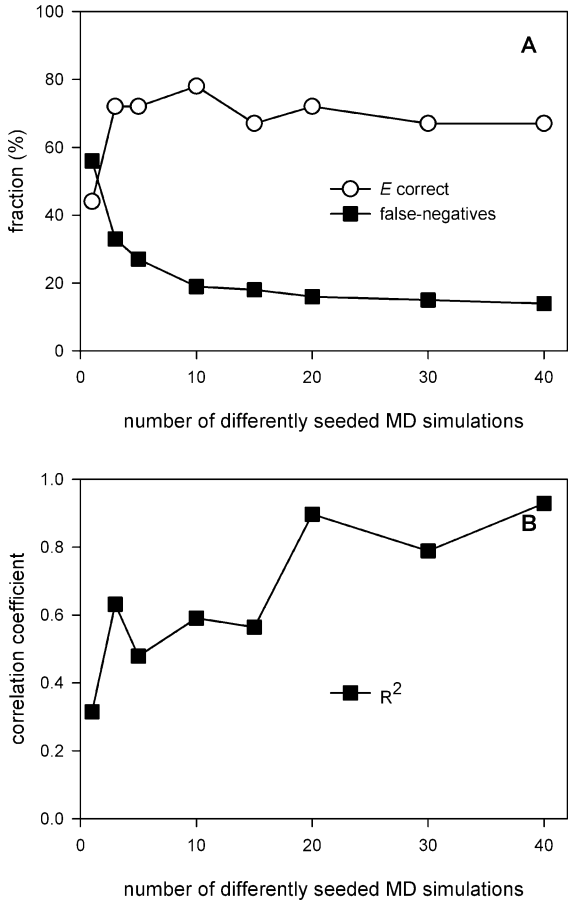


Figure 5. Improved correlation between predictions and experimentally observed *E* obtained by increasing the number of MD simulations. The results are for LinB and DhaA. The MD simulations were 10 ps long. (A) Correct predictions (*i.e.*, within factor 10 from experimental result) and false negative results (*i.e.*, no NAC observed for a converted substrate enantiomer). (B) Pearson's R^2 coefficients for the correlation between experimental and predicted *E*.

The quality of the predictions obtained by averaging NAC frequencies from different trajectories increased for up to at least 20 different MD runs (Figure 5). This is due to a better averaging because individual simulations produced rather different NAC percentages (Table 4, Figure S1, Supporting Information). The NAC percentages obtained from 40 MD simulations appear to follow a multinomial distribution rather than a normal distribution (Figure S1, Supporting Information), which is in agreement with distinct conformations of the enzyme−substrate complex being sampled by the different MD simulations. Inspection of the snapshots of individual 10 ps MD simulations showed that the different NAC percentages corresponded to different substrate orientations in the active site (Figure 4B), as was also observed for the much longer 7 ns MD simulations (Figure 4A).

This 40 × 10 ps MD simulation protocol was also tested for the other two dehalogenases, DhlA and DbjA (Tables S4 and S5, Supporting Information). DbjA and DhlA are only distantly related to LinB and DhaA (31−51% sequence identity) and differ in shape and size of their active sites.[52−55] Also for these

dx.doi.org/10.1021/ci500126x | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

Table 4. Differences in Frequency of Catalytically Relevant Orientations Resulting from Different Sets of Initial Atom Velocities during Ultrashort (10 ps) MD Simulations of LinB

	NACs in MD simulation (%)				
	examples individual trajectories				
substrate	seed A	seed B	seed C	seed D	average of 40 trajectories
all examples greater than 4-fold difference with seed A[a]					
(*R*)-2	0	0	**11.2**	0	**0.29**
(*R*)-3	4.0	3.6	9.6	**0.8**	5.53
(*S*)-5	0.4	0.8	0.4	**0**	0.85
(*S*)-6	0.8	**0**	0.4	**0**	1.77
(*R*)-11	0	**6.8**	**2.4**	**13.2**	**3.28**
(*R*)-15	0	0	**0.8**	0	**0.12**
(*S*)-15	0	**1.2**	0	0	**1.01**
(*R*)-19	0.4	1.2	**3.6**	**0**	**1.95**
(*S*)-19	1.2	1.2	**0**	0.4	1.15
(*R*)-21	0.8	1.6	**0**	**0**	0.68
(*S*)-21	0.4	**0**	**3.2**	**0**	0.55
(*S*)-22	0.8	**3.6**	**0**	0.8	1.23
(*R*)-31	6.8	**0**	3.2	2.8	6.61
(*R*)-32	0	0	**0.8**	**7.6**	**3.47**
(*S*)-32	1.2	**0**	2.0	2.8	1.87
(*R*)-35	19.6	**3.2**	8.4	14.0	8.00
(*S*)-35	0	**0.8**	**3.2**	0	**0.63**
(*R*)-36	0	0	**1.2**	**2.8**	**0.62**
(*S*)-36	0	0	**0.4**	**0.4**	**0.42**
(*R*)-43	58.4	**6.0**	**0**	**0**	**9.36**
(*S*)-45	0	0	**0.8**	**0.8**	**0.46**
all examples less than 4-fold difference with seed A					
(*R*)-1	5.6	8.8	10.4	7.2	8.98
(*S*)-1	0	0	0	0	**0.08**
(*S*)-2	0	0	0	0	**0.28**
(*S*)-3	0	0	0	0	**0.15**
(*R*)-5	10.8	9.6	13.2	12.4	11.91
(*R*)-6	16.0	12.0	12.0	20.0	16.82
(*S*)-11	0	0	0	0	0
(*R*)-20	2.8	0.8	4.8	0.8	3.31
(*S*)-20	4.4	2.8	5.2	5.2	2.54
(*R*)-22	0	0	0	0	0
(*S*)-31	2.4	2.8	4.4	3.6	6.36
(*R*)-38	0	0	0	0	0
(*S*)-38	0	0	0	0	0
(*S*)-43	0	0	0	0	**0.16**
(pro*RS*)-44	0	0	0	0	0
(pro*SR*)-44	0	0	0	0	**0.04**
(*R*)-45	0	0	0	0	0

[a]NAC percentages that differ more than a factor 4 from the result with seed A are shown in bold. The length of the MD simulations was 10 ps. The abbreviation pro*RS* stands for a binding mode of a prochiral substrate that results in a (1*R*,2*S*) product.

two enzymes the predicted and experimental *E* agreed well (78% correct predictions for 14 different pairs of enzyme−substrate enantiomers), and the fraction of false negatives was very low (4% of 50 enzyme−substrate complexes). The combined results for the four dehalogenases showed that enantioselectivity predictions were correct for 72% of the enzyme−substrate combinations and the fraction of false negatives was 10%.

One might have expected that for nonconverted substrates MD simulations longer than 10 ps are required for observing substrate escaping from a NAC in which it was initially docked. However, for the majority of the poorly converted substrate enantiomers that are docked in a NAC conformation at the start the simulation (Tables S1−S4, Supporting Information), the ligand rapidly reorients during the initial 5 ps equilibration, and no NACs are adopted during the production phase of the MD simulation. This behavior was observed, for example, with DbjA complexed with (*S*)-1, (*S*)-2, (*S*)-6, (*S*)-34, and (*S*)-35 (see Tables S1−S4, Supporting Information for more examples). Also, the opposite occurred, i.e., a substrate enantiomer that is preferentially converted by the enzyme adopted NACs during MD simulation, while the initial enzyme−substrate complex did not obey NAC criteria because the preceding docking step gave no substrate orientation in a NAC (e.g., for DbjA in complex with the well converted substrate (*R*)-2, Table S5, Supporting Information). Thus, a converted substrate can reorient itself during MD simulation to a pose more favorable for nucleophilic attack within 5 ps or remain in such a pose (Figure 6A). On the same time scale, nonconverted substrate enantiomers were observed to adopt nonproductive orientations even when its docked starting structure was in a NAC (Figure 6B).

To further investigate the conformational dynamics of enzyme−substrate complexes, it was monitored how often

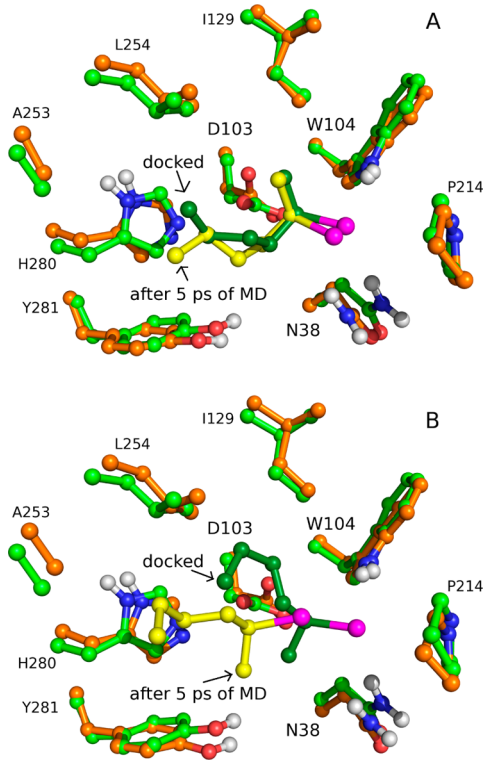


Figure 6. Short MD simulations eliminating productive conformations for a nonconverted substrate enantiomer, but not for a converted substrate enantiomer, within 5 ps. The leaving group halide is magenta, nitrogen atoms are blue, oxygen atoms are red, and hydrogen atoms are gray. Carbon atoms are indicated with green for DbjA with docked substrate (carbon atoms are dark green). Carbon atoms depicted in orange represent the same DbjA after 5 ps of MD simulation with carbon atoms of the substrate depicted in yellow. D103, W104, and N38 are the residues directly involved in the first reaction step. (A) Well-converted substrate enantiomer of DbjA, (*R*)-34. (B) Poorly converted substrate (*S*)-34.

dx.doi.org/10.1021/ci500126x | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

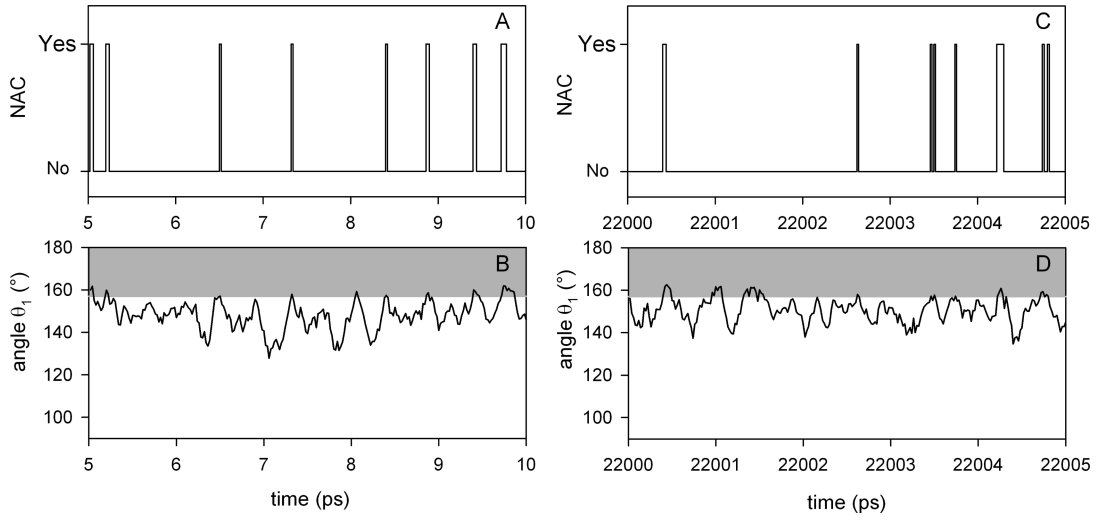


Figure 7. Highly dynamic exchange of productive and nonproductive conformations at a fs time scale. At 20 fs intervals, it was determined whether a NAC was present or not. Both examples are for LinB with (*R*)-1 as a substrate. For angle θ_1, a gray background indicates that the angle satisfies the NAC criterion as described in the caption of Scheme 2. (A) Occurrence of NACs from 5 to 10 ps. (B) θ_1 angle of the same MD simulation. (C) Occurrence of NACs on the same time scale after 22 ns. (D) θ_1 angle of the same MD simulation after 22 ns.

the bound substrates enter and leave a NAC. Analysis of the 10 ps MD simulations showed that the substrate routinely slips in and out of a NAC on a 20 fs time scale (Figure 7A). This was not an artifact due to the shortness of the MD simulation because also after 22 ns of simulation the substrate still enters and leaves the NAC on the same time scale (Figure 7C). The occurrence of a NAC typically corresponds to the angle θ_1 reaching a value that satisfied the NAC criteria (Figure 7B,D) while the other geometric criteria are satisfied most of the time (see below). These rapid movements of the substrate on a fs time scale can explain why even ps time scale simulations can provide sufficient sampling of the locally accessible conformational space of the enzyme−substrate complex.

Structural Basis of Dehalogenase (Enantio)selectivity. Good agreement between experimentally observed and predicted *E* was obtained by NAC analysis of the MD simulations for a large data set with four different dehalogenases and more than 80 enzyme−substrate combinations. Therefore, it is justified to use the geometries that were collected for the NAC analysis to investigate what causes the enantioselectivity in dehalogenases. For all 175 different enzyme−substrate combinations, it was found that the geometric component that most frequently disagreed with the formation of a productive substrate conformation was the angle θ_1, which needs to be close to 180° for nucleophilic attack to occur but was predominantly much lower. This angle between the nucleophilic aspartate oxygen, carbon atom, and halide leaving group (Scheme 2) was only 5% of the time within the requirements for a NAC while the other five criteria were satisfied ≥80% of the time (on average 80% of the time for d_1, 93% for d_2, 85% for d_3, 97% for θ_2, and 94% for θ_3). A low nucleophilic attack angle θ_1 will lower rate the k_2 (Scheme 1). If θ_1 differs enough between the enantiomers, this results in enantioselective catalysis (eq 6).

Another proposed explanation for stereoselective conversion by dehalogenases is that the preferred enantiomers are bound stronger than the disfavored enantiomers.[50] However, binding energies predicted by Autodock4[75] showed no noteworthy differences (<2 kJ mol^{-1}) for (*R*) and (*S*) enantiomers, even in case of substrates that were converted with high stereoselectivity (Table S6, Supporting Information). To explain the enantioselectivities that exceed 200 would require differences in binding energy of 13.1 kJ mol^{-1} (see Discussion).

To investigate whether the fraction of NACs during an MD simulation can be correlated to the catalytic rate of an enzyme variant for a target substrate, further tests were carried out with β-hexachlorocyclohexane (substrate 47). Nucleophilic attack on this particular Lindane isomer is sterically hindered, and conversion is poor with all known dehalogenases. There are two natural variants of LinB[63] that have some catalytic activity for this substrate. They differ 10-fold in catalytic rate, and also six intermediate variants have been experimentally characterized.[63] For these LinB variants, the fraction of NACs observed by MD simulation correlated positively (Figure 8, R^2 = 0.46)

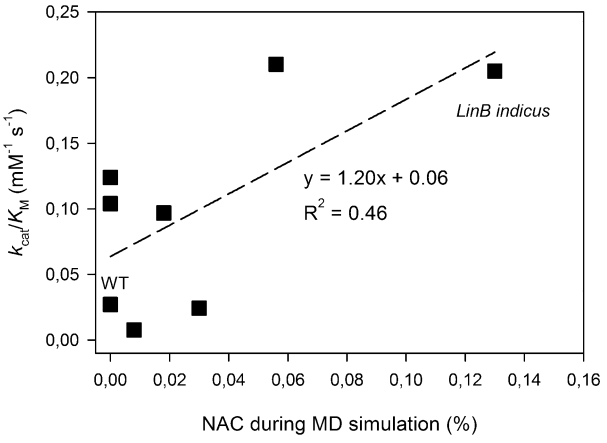


Figure 8. Correlation between experimentally determined specificity constants (k_{cat}/K_M) and the occurrence of catalytically productive substrate conformations in modeled LinB variants. The substrate is 47. LinB from *Sphingomonas paucimobilis indicus*, which differs by 7 mutations (A81T/A112V/I134V/A135T/I138L/A247H/M253I) from the WT LinB, which is from *Sphingomonas paucimobilis japonicus*. These WT dehalogenases and experimentally characterized intermediate variants were used for MD simulations. Experimental data are from Ito et al.[63] Only for the WT, an X-ray structure was available; all other variants were modeled based on the X-ray structure of LinB WT.

dx.doi.org/10.1021/ci500126x | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

with their experimentally determined k_{cat}/K_M (which should be correlated to k_2, see eq 5). Also, for β-hexachlorocyclohexane, the critical geometric component is θ_1 (Scheme 2), which for all of the variants was on average <140°, which is far from optimal, and causes low (maximally 0.13%) NAC percentages, in agreement with the experimental catalytic activities (maximally 0.2 mM^{-1} s^{-1}, Figure 8). In contrast, for DhlA, with its native substrate 1,2-dichloroethane (substrate 46), an extremely high fraction of NACs was present during the MD simulation (56%, Table S5, Supporting Information). This is a higher NAC percentage than for all other 174 tested enzyme−substrate combinations. These results confirm that the degree to which an enzyme is suitable for conversion of a particular substrate can correlate with the percentage of NACs during MD simulations.

DISCUSSION

Modeling Enantioselectivity and Specificity by NACs. By quantifying catalytically productive conformations observed during short MD simulations, it was possible to successfully model the enantioselectivities displayed by four dehalogenases for a large set of structurally diverse substrates (Figure 2). No fitting of correlation constants or scoring criteria to experimental data was applied in this NAC approach, while this is required in other approaches to model enzyme specificity and enantioselectivity.[21,22,79,80] The criteria to define a NAC are based on quantum mechanical calculations,[35,36] which explains why it can identify productive conformations without calibration with experimentally data. The principle behind the modeling described here is that the reactivity of a substrate in an active site correlates with the frequency with which productive conformations [NAC] occur (eq 8). The constant C in eq 8 differs per substrate, which would require QM calculations to accurately predict the relative rates for different substrates. However, constant C should be similar for enantiomers of the same substrate; thus, the ratio of occurrence of NACs during MD simulation can provide accurate modeling of enantioselectivity (eq 9, P and D denote preferred and disfavored enantiomers). The low CPU costs of the method (approximately 30 min for 20 simulations, Table S8, Supporting Information) makes it possible to investigate large numbers of different enzyme−substrate complexes.

$$k_2 = C \times [\mathrm{NAC}] \tag{8}$$

$$E = \frac{{k_2}^{\mathrm{P}}}{{k_2}^{\mathrm{D}}} = \frac{C \times [\mathrm{NAC}]^{\mathrm{P}}}{C \times [\mathrm{NAC}]^{\mathrm{D}}} \tag{9}$$

When employing the 40 × 10 ps protocol to the data set of four haloalkane dehalogenases with different substrates, good correlations between predictions and experimental results were observed (Figure 3C, Table 2). Also, 72% of the modeled enzyme−substrate complexes have a predicted E which agrees within a factor of 10 with the experimentally determined enantioselectivities. The latter can have similar errors because in most cases they are based on single time point measurements.[48−50] When a more accurate assay with a series of time points was employed for a dehalogenase, the resulting E differed 3-fold from the value determined with a single time-point measurement.[48] Furthermore, the model assumes that only the rate k_2 (Scheme 1) differs for enantiomers; differences in K_d are ignored (eq 6). Considering these limitations of the modeling, modest accuracy of experimental data, and inherent errors of simulating proteins and substrates with a force field, it is very satisfactory that for 72% of the data the difference between experimentally determined and predicted E was less than 10-fold with a computationally inexpensive protocol.

The short MD simulations and NAC analysis also allow predicting enantioselectivity of substrates for which the preferred enantiomer is unknown. For LinB, there are four similar substrates (compounds 39−42, Figure 2) for which high enantioselectivities (up to E > 200) were measured, while the preferred enantiomer was not experimentally identified. The results of 40 × 10 ps MD simulations (Table S2, Supporting Information) predict that for these substrates the (R) enantiomer is preferentially converted.

Furthermore, a correlation was observed between NAC frequencies and the k_{cat}/K_M of different LinB variants for substrate 47 (Figure 8). Such a correlation may be expected because k_{cat}/K_M should correlate to k_2 (eq 4). The X-ray structures of these variants, which carry up to 7 mutations, were unknown and experimentally the slowest and fastest variants differ only by a factor of 10 (Figure 8). This result bodes well for *in silico* screening of improved enzyme variants using computationally inexpensive clusters of short MD simulations.

Origin of Enantioselectivity of Dehalogenases. The correlation between NAC frequencies and enantioselectivities for the four dehalogenases suggest that enantioselectivity is governed by faster initial nucleophilic attack on the preferred enantiomers by the active site aspartate (rate k_2, Scheme 1). For this initial reaction, angle θ_1 (Scheme 2) was limiting the formation of NACs for the entire set of 175 different dehalogenase−substrate complexes. This angle needs to be close to 180° for nucleophilic attack to occur. Only for conversion of 1,2-dichlorethane by DhlA, the enzyme that naturally evolved to convert this substrate,[52,81] θ_1 is close to ideal during the most of the MD trajectory resulting in >50% NAC. Apparently, DhlA strongly favors the ideal positioning of its native substrate.

It can be rationalized why θ_1 is the predominantly limiting geometric variable. For the negatively polarized halogen atom of the substrate, it will be energetically favorable to be positioned between the hydrogen bond-donating atoms of the active site tryptophan and asparagine side chains (Scheme 2), which together act as a specific halogen binding site in most haloalkane dehalogenases. Because the substrate halogen spontaneously remains in this binding site, the criteria for θ_2, θ_3, d_2, and d_3 are satisfied, as was observed for over 85% of the recorded geometries of all enzyme−substrates complexes during MD simulation. With the halogen atom dynamically bound in this site, the carbon atom is pulled close to the aspartate oxygen atom, which is positioned at approximately the required distance from the halide binding site; this satisfied d_1 for more than 80% of the time. While located approximately between the halide binding site and the aspartate oxygen atom, this carbon atom of the substrate can still occupy positions that give poor alignment for nucleophilic attack. As a result, θ_1 is only satisfied 5% of the time, 20-fold less than all other geometric criteria.

If selective binding of the preferred enantiomer would govern the enantioselectivity of dehalogenases,[50] for the E to exceed 200, a difference in binding energy between the enantiomers of >13.1 kJ mol^{-1} would be required (eqs 10 and 11, in which R is the gas constant and T the absolute temperature). However, binding energy predictions for substrates that are converted with such high enantioselectivity do not support such large

dx.doi.org/10.1021/ci500126x | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

differences (Table S6, Supporting Information). The experimental observation that the K_M values are lower for the preferred enantiomers than for the nonpreferred enantiomers[48,50] can also be explained by faster initial nucleophilic attack because under the kinetics of Scheme 1 the K_M becomes smaller with a larger k_2 (eq 2). An influence of the rate of hydrolysis and product release can also be excluded because the equations show that the corresponding rate k_3 does not affect k_{cat}/K_M (eq 4) and thus also does not influence E (eq 5). While these results strongly support a role for the angle θ_1 in determining enantioselectivity for the large majority of compounds, it cannot be excluded that for other dehalogenase substrates selective binding of one of the enantiomers does determine enantioselectivity.[82]

$$E = \frac{K_d^{\,P}}{K_d^{\,D}} \tag{10}$$

Equations 10 and 11 only hold if k_2 is equal for both enantiomers, which is unlikely.

$$\Delta\Delta G^{binding} = RT \times \ln E \tag{11}$$

Multiple Independent MD Simulations for Modeling Enzyme–Substrate Complexes. Our results support the conclusion that a cluster of short (10 ps) independent MD simulations sample the conformational space of the enzyme–substrate complex more thoroughly than a single long (up to 22 ns) MD simulation. Apparently, the use of multiple short simulations with different initial velocities gave more complete coverage of conformational space adopted by enzyme–substrate complexes than single MD simulations that were run over a much longer time scale (Figure 4, Table 2). Better sampling by a cluster of MD simulations was found both with the Yamber3 and Amber ff03 force fields (Supporting Information). During equilibration, the substrate can reorient, and an enzyme–substrate complex will adopt a structure representing one of the conformational subspaces, allowing complete sampling of NACs. The short trajectories can be adequately averaged because substrates move in and out of a NAC on a very short time scale (e.g., 20 fs, with simulations covering 5 ps). Because the average speed of a carbon atom at room temperature is approximately 250 m s^{-1} (2.5 Å ps^{-1}), the sampling of movements of a substrate in an active site on this time scale seems physically realistic. However, even during prolonged simulations, there is little exchange between different protein conformations that are attained during equilibration, and these trajectories show a different behavior of substrate in terms of NAC frequencies. This also explains that we observed a multinomial distribution of NAC frequencies over the different MD simulations instead of a normal distribution (Figure S1, Supporting Information). Fitting standard errors according to a normal distribution can be used to determine the convergence of sampling protein–ligand complexes,[86] but was inapt for enzyme–substrate complexes because it would cover a more narrow range of NAC frequencies than the original data (Figure S1, Supporting Information) and thereby overestimates the obtained degree of convergence. For this reason, the convergence of sampling was only appraised based on the agreement with experimental enantioselectivity (Figure 3).

These observations can be explained by the existing idea that that once an MD simulation has started the protein structure will quickly relax to one of the many local energy minima in the conformational landscape, while jumps from one local minimum to another are rare.[42−44] This is relevant because enzyme catalysis typically occurs on the time scale of seconds (e.g., LinB[48] with substrate (*R*)-6 k_{cat} = 0.082 s^{-1}; DhlA[83] with substrate 46, k_{cat} = 3.3 s^{-1}). It is likely that only a fraction of the conformations that a single dehalogenase samples in solution during a second can be observed on the nanosecond time scale of an MD simulation. It was recently shown that even an extremely long 1 ms MD simulations of bovine pancreatic trypsin inhibitor had far from completely sampled all of its conformations,[84,85] which further shows that even the longest single MD simulations can be inadequate for conformational sampling and that the use of shorter parallel MD simulations may be preferred if long equilibrations are not necessary. For dehalogenases, the protein structure observed by X-ray crystallography is catalytically active,[56] which can explain why 5 ps equilibrations were sufficient for the work reported here. This would not apply for enzymes that crystallize in a catalytically inactive conformation.

Other recent methodological investigations that explored the use of multiple independent MD simulations[45,86−88] used 1 to 4 orders of magnitude longer MD simulations (100 ps to 10 ns) than employed here (10 ps). These studies focused either on reproducing the flexibility of proteins in solution or on predicting ligand-binding affinity. For the latter task, the analysis of the entropy and electrostatic contributions to binding based on MD snapshots were as computationally expensive as generating the 100 ps MD trajectory. Thus, for these studies, it was not worthwhile to shorten the trajectories further. For the NAC analysis and enantioselectivity predictions, merely a few distances and angles are recorded, and generating the MD trajectory requires by far most CPU time, making it desirable to keep the MD trajectories short.

The CPU expense of 20 × 10 ps MD simulations is a fraction of that required for a 1 × 22 ns MD simulation and low enough for routine *in silico* screening of 10^3–10^4 enzyme variants. With alternative techniques to predict enzyme activity, specifically QM/MM, it is possible to obtain accurate predictions as well[89,90] but at a much higher CPU expense. Clusters of short MD simulations could select in a high-throughput manner enzyme variants in which catalytically favorable geometries occur before employing QM/MM or experimental screening.

CONCLUSIONS

The use of a cluster of short MD simulations allowed accurate modeling of enantioselectivity for a set of four different dehalogenases. The enantioselectivity for most of the 45 dehalogenase substrates appears to be caused by differences in the angle for nucleophilic attack (θ_1, Scheme 2) for the separate enantiomers; this angle seemed suboptimal for almost all substrates and strongly influenced by the three-dimensional structure of the substrate, resulting in enantioselective conversions. More complete conformational sampling was obtained by clusters of ≥20 very short MD simulations than by single long MD simulations. The combination of accurate modeling and low computational cost of such MD simulations enables high-throughput *in silico* screening. These insights are important for engineering the substrate scope and enantioselectivity of dehalogenases and other enzymes.

ASSOCIATED CONTENT

Supporting Information

Available are descriptions of control experiments with an alternative force field, and Figures S1–S6 and Tables S1–S8.

dx.doi.org/10.1021/ci500126x | J. Chem. Inf. Model. XXXX, XXX, XXX–XXX

This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*E-mail: d.b.janssen@rug.nl. Tel: +31 50 363 4008. Fax: +31 50 363 4165.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

Support for this work came from the European Union (seventh Framework) through the Kyrobio (289646, KBBE-2011-5) and Micro B3 (287589, OCEAN.2011-2) projects and from NWO (Netherlands Organization for Scientific Research) through an ECHO grant.

REFERENCES

(1) Schmid, A.; Dordick, J.; Hauer, B.; Kiener, A.; Wubbolts, M.; Witholt, B. Industrial biocatalysis today and tomorrow. *Nature* **2001**, *6817*, 258−268.

(2) Bornscheuer, U. T.; Huisman, G. W.; Kazlauskas, R. J.; Lutz, S.; Moore, J. C.; Robins, K. Engineering the third wave of biocatalysis. *Nature* **2012**, *7397*, 185−194.

(3) Eijsink, V.; Gaseidnes, S.; Borchert, T.; Van den Burg, B. Directed evolution of enzyme stability. *Biomol. Eng.* **2005**, *1*−*3*, 21−30.

(4) Polizzi, K. M.; Bommarius, A. S.; Broering, J. M.; Chaparro-Riggers, J. F. Stability of biocatalysts. *Curr. Opin. Chem. Biol.* **2007**, *2*, 220−225.

(5) Magliery, T. J.; Lavinder, J. J.; Sullivan, B. J. Protein stability by number: High-throughput and statistical approaches to one of protein science's most difficult problems. *Curr. Opin. Chem. Biol.* **2011**, *3*, 443−451.

(6) Wijma, H. J.; Floor, R. J.; Janssen, D. B. Structure- and sequence-analysis inspired engineering of proteins for enhanced thermostability. *Curr. Opin. Struct. Biol.* **2013**, *23*, 588−594.

(7) Munoz Solano, D.; Hoyos, P.; Hernaiz, M. J.; Alcantara, A. R.; Sanchez-Montero, J. M. Industrial biotransformations in the synthesis of building blocks leading to enantiopure drugs. *Bioresour. Technol.* **2012**, *151*, 196−207.

(8) Sukumaran, J.; Hanefeld, U. Enantioselective C−C bond synthesis catalysed by enzymes. *Chem. Soc. Rev.* **2005**, *6*, 530−542.

(9) Reetz, M. T. Laboratory evolution of stereoselective enzymes: A prolific source of catalysts for asymmetric reactions. *Angew. Chem., Int. Ed.* **2011**, *1*, 138−174.

(10) Koudelakova, T.; Bidmanova, S.; Dvorak, P.; Pavelka, A.; Chaloupkova, R.; Prokop, Z.; Damborsky, J. Haloalkane dehalogenases: Biotechnological applications. *Biotechnol. J.* **2013**, *1*, 32−45.

(11) Pikkemaat, M. G.; Janssen, D. B. Generating segmental mutations in haloalkane dehalogenase: A novel part in the directed evolution toolbox. *Nucleic Acids Res.* **2002**, *8*, E35−5.

(12) Chang, C. H.; Schindler, J. F.; Unkefer, C. J.; Vanderberg, L. A.; Brainard, J. R.; Terwilliger, T. C. *In vivo* screening of haloalkane dehalogenase mutants. *Bioorg. Med. Chem.* **1999**, *10*, 2175−2181.

(13) Bosma, T.; Damborsky, J.; Stucki, G.; Janssen, D. B. Biodegradation of 1,2,3-trichloropropane through directed evolution and heterologous expression of a haloalkane dehalogenase gene. *Appl. Environ. Microbiol.* **2002**, *7*, 3582−3587.

(14) Chaloupkova, R.; Sykorova, J.; Prokop, Z.; Jesenska, A.; Monincova, M.; Pavlova, M.; Tsuda, M.; Nagata, Y.; Damborsky, J. Modification of activity and specificity of haloalkane dehalogenase from *Sphingomonas paucimobilis* UT26 by engineering of its entrance tunnel. *J. Biol. Chem.* **2003**, *52*, 52622−52628.

(15) Van Leeuwen, J. G. E.; Wijma, H. J.; Floor, R. J.; Van der Laan, J.; Janssen, D. B. Directed evolution strategies for enantiocomplementary haloalkane dehalogenases: From chemical waste to enantiopure building blocks. *ChemBioChem.* **2012**, *1*, 137−148.

(16) Koudelakova, T.; Chaloupkova, R.; Brezovsky, J.; Prokop, Z.; Sebestova, E.; Hesseler, M.; Khabiri, M.; Plevaka, M.; Kulik, D.; Kuta Smatanova, I.; Rezacova, P.; Ettrich, R.; Bornscheuer, U. T.; Damborsky, J. Engineering enzyme stability and resistance to an organic cosolvent by modification of residues in the access tunnel. *Angew. Chem., Int. Ed.* **2013**, *7*, 1959−1963.

(17) Arnold, F. H. Combinatorial and computational challenges for biocatalyst design. *Nature* **2001**, *6817*, 253−257.

(18) Fox, R. J.; Davis, S. C.; Mundorff, E. C.; Newman, L. M.; Gavrilovic, V.; Ma, S. K.; Chung, L. M.; Ching, C.; Tam, S.; Muley, S.; Grate, J.; Gruber, J.; Whitman, J. C.; Sheldon, R. A.; Huisman, G. W. Improving catalytic function by ProSAR-driven enzyme evolution. *Nat. Biotechnol.* **2007**, *3*, 338−344.

(19) Jackel, C.; Hilvert, D. Biocatalysts by evolution. *Curr. Opin. Biotechnol.* **2010**, *6*, 753−759.

(20) Juhl, P. B.; Doderer, K.; Hollmann, F.; Thum, O.; Pleiss, J. Engineering of *Candida antarctica* lipase B for hydrolysis of bulky carboxylic acid esters. *J. Biotechnol.* **2010**, *4*, 474−480.

(21) Pace, V.; Cortes Cabrera, A.; Ferrario, V.; Sinisterra, J. V.; Ebert, C.; Gardossi, L.; Braiuca, P.; Alcantara, A. R. Structural bases for understanding the stereoselectivity in ketone reductions with ADH from *Thermus thermophilus*: A quantitative model. *J. Mol. Catal. B* **2011**, *1*−*2*, 23−31.

(22) Braiuca, P.; Lorena, K.; Ferrario, V.; Ebert, C.; Gardossi, L. A three-dimensional quanititative structure-activity relationship (3D-QSAR) model for predicting the enantioselectivity of *Candida antarctica* lipase B. *Adv. Synth. Catal.* **2009**, *9*, 1293−1302.

(23) Seifert, A.; Antonovici, M.; Hauer, B.; Pleiss, J. An efficient route to selective bio-oxidation catalysts: an iterative approach comprising modeling, diversification, and screening, based on CYP102A1. *ChemBioChem.* **2011**, *9*, 1346−1351.

(24) Hediger, M. R.; De Vico, L.; Svendsen, A.; Besenmatter, W.; Jensen, J. H. A computational methodology to screen activities of enzyme variants. *PloS one* **2012**, *12*, e49849−e49849.

(25) Zheng, H.; Reetz, M. T. Manipulating the stereoselectivity of limonene epoxide hydrolase by directed evolution based on iterative saturation mutagenesis. *J. Am. Chem. Soc.* **2010**, *44*, 15744−15751.

(26) Stjernschantz, E.; van Vugt-Lussenburg, B. M.; Bonifacio, A.; de Beer, S. B.; van der Zwan, G.; Gooijer, C.; Commandeur, J. N.; Vermeulen, N. P.; Oostenbrink, C. Structural rationalization of novel drug metabolizing mutants of cytochrome P450 BM3. *Proteins* **2008**, *1*, 336−352.

(27) Kiss, G.; Rothlisberger, D.; Baker, D.; Houk, K. N. Evaluation and ranking of enzyme designs. *Protein Sci.* **2010**, *9*, 1760−1773.

(28) Syren, P. O.; Hendil-Forssell, P.; Aumailley, L.; Besenmatter, W.; Gounine, F.; Svendsen, A.; Martinelle, M.; Hult, K. Esterases with an introduced amidase-like hydrogen bond in the transition state have increased amidase specificity. *ChemBioChem.* **2012**, *5*, 645−648.

(29) Zheng, F.; Yang, W.; Ko, M. C.; Liu, J.; Cho, H.; Gao, D.; Tong, M.; Tai, H. H.; Woods, J. H.; Zhan, C. G. Most efficient cocaine hydrolase designed by virtual screening of transition states. *J. Am. Chem. Soc.* **2008**, *36*, 12148−12155.

(30) Pan, Y.; Gao, D.; Yang, W.; Cho, H.; Yang, G.; Tai, H. H.; Zhan, C. G. Computational redesign of human butyrylcholinesterase for anticocaine medication. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *46*, 16656−16661.

(31) Privett, H. K.; Kiss, G.; Lee, T. M.; Blomberg, R.; Chica, R. A.; Thomas, L. M.; Hilvert, D.; Houk, K. N.; Mayo, S. L. Iterative approach to computational enzyme design. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *10*, 3790−3795.

(32) Bruice, T. C. Computational approaches: Reaction trajectories, structures, and atomic motions. Enzyme reactions and proficiency. *Chem. Rev.* **2006**, *8*, 3119−3139.

(33) Negri, A.; Marco, E.; Damborsky, J.; Gago, F. Stepwise dissection and visualization of the catalytic mechanism of haloalkane dehalogenase LinB using molecular dynamics simulations and computer graphics. *J. Mol. Graph. Model* **2007**, *3*, 643−651.

(34) Hur, S.; Kahn, K.; Bruice, T. C. Comparison of formation of reactive conformers for the SN2 displacements by CH_3CO_2- in water

and by Asp124−CO_2− in a haloalkane dehalogenase. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *5*, 2215−2219.
(35) Lau, E. Y.; Kahn, K.; Bash, P. A.; Bruice, T. C. The importance of reactant positioning in enzyme catalysis: A hybrid quantum mechanics/molecular mechanics study of a haloalkane dehalogenase. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *18*, 9937−9942.
(36) Lightstone, F. C.; Zheng, Y. J.; Maulitz, A. H.; Bruice, T. C. Non-enzymatic and enzymatic hydrolysis of alkyl halides: a haloalkane dehalogenation enzyme evolved to stabilize the gas-phase transition state of an SN2 displacement reaction. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *16*, 8417−8420.
(37) Bruice, T. C. A view at the millennium: the efficiency of enzymatic catalysis. *Acc. Chem. Res.* **2002**, *3*, 139−148.
(38) Bruice, T.; Benkovic, S. Chemical basis for enzyme catalysis. *Biochemistry* **2000**, *21*, 6267−6274.
(39) Warshel, A.; Sharma, P. K.; Kato, M.; Xiang, Y.; Liu, H.; Olsson, M. H. M. Electrostatic basis for enzyme catalysis. *Chem. Rev.* **2006**, *8*, 3210−3235.
(40) Ranaghan, K. E.; Mulholland, A. J. Conformational effects in enzyme catalysis: QM/MM free energy calculation of the 'NAC' contribution in chorismate mutase. *Chem. Commun. (Camb)* **2004**, *10*, 1238−1239.
(41) Klepeis, J. L.; Lindorff-Larsen, K.; Dror, R. O.; Shaw, D. E. Long-timescale molecular dynamics simulations of protein structure and function. *Curr. Opin. Struct. Biol.* **2009**, *2*, 120−127.
(42) Caves, L. S. D.; Evanseck, J. D.; Karplus, M. Locally accessible conformations of proteins: Multiple molecular dynamics simulations of crambin. *Protein Sci.* **1998**, *3*, 649−666.
(43) Daggett, V. Long timescale simulations. *Curr. Opin. Struct. Biol.* **2000**, *2*, 160−164.
(44) Monticelli, L.; Sorin, E. J.; Tieleman, D. P.; Pande, V. S.; Colombo, G. Molecular simulation of multistate peptide dynamics: A comparison between microsecond timescale sampling and multiple shorter trajectories. *J. Comput. Chem.* **2008**, *11*, 1740−1752.
(45) Genheden, S.; Ryde, U. A comparison of different initialization protocols to obtain statistically independent molecular dynamics simulations. *J. Comput. Chem.* **2011**, *2*, 187−195.
(46) Kiss, G.; Pande, V. S.; Houk, K. N. Molecular dynamics simulations for the ranking, evaluation, and refinement of computationally designed proteins. *Methods in Protein Design* **2013**, *523*, 145−170.
(47) Keizers, P. H.; de Graaf, C.; de Kanter, F. J.; Oostenbrink, C.; Feenstra, K. A.; Commandeur, J. N.; Vermeulen, N. P. Metabolic regio- and stereoselectivity of cytochrome P450 2D6 towards 3,4-methylenedioxy-N-alkylamphetamines: *In silico* predictions and experimental validation. *J. Med. Chem.* **2005**, *19*, 6117−6127.
(48) Westerbeek, A.; Szymanski, W.; Wijma, H. J.; Marrink, S. J.; Feringa, B. L.; Janssen, D. B. Kinetic resolution of alpha-bromoamides: Experimental and theoretical investigation of highly enantioselective reactions catalyzed by haloalkane dehalogenases. *Adv. Synth. Catalys.* **2011**, *6*, 931−944.
(49) Pieters, R.; Spelberg, J.; Kellogg, R.; Janssen, D. The enantioselectivity of haloalkane dehalogenases. *Tetrahedron Lett.* **2001**, *3*, 469−471.
(50) Prokop, Z.; Sato, Y.; Brezovsky, J.; Mozga, T.; Chaloupkova, R.; Koudelakova, T.; Jerabek, P.; Stepankova, V.; Natsume, R.; van Leeuwen, J. G.; Janssen, D. B.; Florian, J.; Nagata, Y.; Senda, T.; Damborsky, J. Enantioselectivity of haloalkane dehalogenases and its modulation by surface loop engineering. *Angew. Chem., Int. Ed.* **2010**, *35*, 6111−6115.
(51) Szymanski, W.; Westerbeek, A.; Janssen, D. B.; Feringa, B. L. A simple enantioconvergent and chemoenzymatic synthesis of optically active alpha-substituted amides. *Angew. Chem., Int. Ed.* **2011**, *45*, 10712−10715.
(52) Keuning, S.; Janssen, D. B.; Witholt, B. Purification and characterization of hydrolytic haloalkane dehalogenase from *Xanthobacter autotrophicus* GJ10. *J. Bacteriol.* **1985**, *2*, 635−639.
(53) Kulakova, A. N.; Larkin, M. J.; Kulakov, L. A. The plasmid-located haloalkane dehalogenase gene from *Rhodococcus rhodochrous* NCIMB 13064. *Microbiology* **1997**, *Pt 1*, 109−115.
(54) Sato, Y.; Monincova, M.; Chaloupkova, R.; Prokop, Z.; Ohtsubo, Y.; Minamisawa, K.; Tsuda, M.; Damborsky, J.; Nagata, Y. Two rhizobial strains, *Mesorhizobium loti* MAFF303099 and *Bradyrhizobium japonicum* USDA110, encode haloalkane dehalogenases with novel structures and substrate specificities. *Appl. Environ. Microbiol.* **2005**, *8*, 4372−4379.
(55) Nagata, Y.; Miyauchi, K.; Damborsky, J.; Manova, K.; Ansorgova, A.; Takagi, M. Purification and characterization of a haloalkane dehalogenase of a new substrate class from a gamma-hexachlorocyclohexane-degrading bacterium, *Sphingomonas paucimobilis* UT26. *Appl. Environ. Microbiol.* **1997**, *9*, 3707−3710.
(56) Verschueren, K. H.; Seljee, F.; Rozeboom, H. J.; Kalk, K. H.; Dijkstra, B. W. Crystallographic analysis of the catalytic mechanism of haloalkane dehalogenase. *Nature* **1993**, *6431*, 693−698.
(57) Prokop, Z.; Monincova, M.; Chaloupkova, R.; Klvana, M.; Nagata, Y.; Janssen, D. B.; Damborsky, J. Catalytic mechanism of the haloalkane dehalogenase LinB from *Sphingomonas paucimobilis* UT26. *J. Biol. Chem.* **2003**, *46*, 45094−45100.
(58) Oakley, A. J.; Klvana, M.; Otyepka, M.; Nagata, Y.; Wilce, M. C.; Damborsky, J. Crystal structure of haloalkane dehalogenase LinB from *Sphingomonas paucimobilis* UT26 at 0.95 A resolution: dynamics of catalytic residues. *Biochemistry* **2004**, *4*, 870−878.
(59) Newman, J.; Peat, T.; Richard, R.; Kan, L.; Swanson, P.; Affholter, J.; Holmes, I.; Schindler, J.; Unkefer, C.; Terwilliger, T. Haloalkane dehalogenases: Structure of a *Rhodococcus* enzyme. *Biochemistry* **1999**, *49*, 16105−16114.
(60) Ridder, I. S.; Rozeboom, H. J.; Dijkstra, B. W. Haloalkane dehalogenase from *Xanthobacter autotrophicus* GJ10 refined at 1.15 A resolution. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **1999**, *Pt 7*, 1273−1290.
(61) Chaloupkova, R.; Prokop, Z.; Sato, Y.; Nagata, Y.; Damborsky, J. Stereoselectivity and conformational stability of haloalkane dehalogenase DbjA from *Bradyrhizobium japonicum* USDA110: The effect of pH and temperature. *FEBS J.* **2011**, *15*, 2728−2738.
(62) Krieger, E.; Nielsen, J. E.; Spronk, C. A.; Vriend, G. Fast empirical pK_a prediction by Ewald summation. *J. Mol. Graph. Model* **2006**, *4*, 481−486.
(63) Ito, M.; Prokop, Z.; Klvana, M.; Otsubo, Y.; Tsuda, M.; Damborsky, J.; Nagata, Y. Degradation of beta-hexachlorocyclohexane by haloalkane dehalogenase LinB from gamma-hexachlorocyclohexane-utilizing bacterium *Sphingobium* sp. MI1205. *Arch. Microbiol.* **2007**, *4*, 313−325.
(64) Richter, F.; Leaver-Fay, A.; Khare, S. D.; Bjelic, S.; Baker, D. *De novo* enzyme design using Rosetta3. *PLoS One* **2011**, *5*, e19230.
(65) Stewart, J. Special issue − MOPAC: A semiempirical molecular-orbital program. *J. Comput. Aided Mol. Des.* **1990**, *1*, 1−45.
(66) Jakalian, A.; Jack, D. B.; Bayly, C. I. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J. Comput. Chem.* **2002**, *16*, 1623−1641.
(67) Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.* **2009**, *16*, 2785−2791.
(68) Krieger, E.; Darden, T.; Nabuurs, S.; Finkelstein, A.; Vriend, G. Making optimal use of empirical energy functions: Force-field parameterization in crystal space. *Prot. Struct. Funct. Bioinform.* **2004**, *4*, 678−683.
(69) Berendsen, H. J. C.; Postma, J. P. M.; Van Gunsteren, W. F.; Dinola, A.; Haak, J. R. Molecular-dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *8*, 3684−3690.
(70) Hess, B.; Bekker, H.; Berendsen, H.; Fraaije, J. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **1997**, *12*, 1463−1472.
(71) Miyamoto, S.; Kollman, P. A. SETTLE − An analytical version of the Shake and Rattle algorithm for rigid water models. *J. Comput. Chem.* **1992**, *8*, 952−962.

(72) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle Mesh Ewald method. *J. Chem. Phys.* **1995**, *19*, 8577−8593.
(73) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.* **2003**, *16*, 1999−2012.
(74) Cook, P.; Cleland, W. W. *Enzyme kinetics and mechanism*; Garland Science: London, 2007.
(75) Huey, R.; Morris, G. M.; Olson, A. J.; Goodsell, D. S. A semiempirical free energy force field with charge-based desolvation. *J. Comput. Chem.* **2007**, *6*, 1145−1152.
(76) Wang, J.; Cieplak, P.; Kollman, P. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.* **2000**, *12*, 1049−1074.
(77) Lange, O. F.; van der Spoel, D.; de Groot, B. L. Scrutinizing molecular mechanics force fields on the submicrosecond timescale with NMR data. *Biophys. J.* **2010**, *2*, 647−655.
(78) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **2010**, *8*, 1950−1958.
(79) Kmunicek, J.; Bohac, M.; Luengo, S.; Gago, F.; Wade, R. C.; Damborsky, J. Comparative binding energy analysis of haloalkane dehalogenase substrates: modelling of enzyme-substrate complexes by molecular docking and quantum mechanical calculations. *J. Comput.-Aided Mol. Des.* **2003**, *5−6*, 299−311.
(80) Stjernschantz, E.; Vermeulen, N. P.; Oostenbrink, C. Computational prediction of drug binding and rationalisation of selectivity towards cytochromes P450. *Expert Opin. Drug Metab. Toxicol.* **2008**, *5*, 513−527.
(81) Pries, F.; van den Wijngaard, A. J.; Bos, R.; Pentenga, M.; Janssen, D. B. The role of spontaneous cap domain mutations in haloalkane dehalogenase specificity and evolution. *J. Biol. Chem.* **1994**, *26*, 17490−17494.
(82) Banas, P.; Otyepka, M.; Jerabek, P.; Petrek, M.; Damborsky, J. Mechanism of enhanced conversion of 1,2,3-trichloropropane by mutant haloalkane dehalogenase revealed by molecular modeling. *J. Comput.-Aided Mol. Des.* **2006**, *6*, 375−383.
(83) Schanstra, J. P.; Kingma, J.; Janssen, D. B. Specificity and kinetics of haloalkane dehalogenase. *J. Biol. Chem.* **1996**, *25*, 14747−14753.
(84) Shaw, D. E.; Maragakis, P.; Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Eastwood, M. P.; Bank, J. A.; Jumper, J. M.; Salmon, J. K.; Shan, Y.; Wriggers, W. Atomic-level characterization of the structural dynamics of proteins. *Science* **2010**, *6002*, 341−346.
(85) Genheden, S.; Ryde, U. Will molecular dynamics simulations of proteins ever reach equilibrium? *Phys. Chem. Chem. Phys.* **2012**, *24*, 8662−8677.
(86) Genheden, S.; Ryde, U. How to obtain statistically converged MM/GBSA results. *J. Comput. Chem.* **2010**, *4*, 837−846.
(87) Genheden, S.; Diehl, C.; Akke, M.; Ryde, U. Starting-condition dependence of order parameters derived from molecular dynamics simulations. *J. Chem. Theory Comput.* **2010**, *7*, 2176−2190.
(88) Zagrovic, B.; van Gunsteren, W. F. Computational analysis of the mechanism and thermodynamics of inhibition of phosphodiesterase 5A by synthetic ligands. *J. Chem. Theory Comput.* **2007**, *1*, 301−311.
(89) Lind, M. E. S.; Himo, F. Quantum chemistry as a tool in asymmetric biocatalysis: Limonene epoxide hydrolase test case. *Angew. Chem., Int. Ed.* **2013**, *17*, 4563−4567.
(90) Lonsdale, R.; Harvey, J. N.; Mulholland, A. J. A practical guide to modelling enzyme-catalysed reactions. *Chem. Soc. Rev.* **2012**, *8*, 3025−3038.