ARTICLE

A Computational Approach To Design and Evaluate Enzymatic Reaction Pathways: Application to 1-Butanol Production from Pyruvate

Di Wu,[†,⊥] Qin Wang,[‡,⊥] Rajeev S. Assary,[†,§] Linda J. Broadbelt,[*,†] and Goran Krilov[*,‡,∥]

[†]Department of Chemical and Biological Engineering, Northwestern University, Evanston, Illinois 60208, United States
[‡]Department of Chemistry, Boston College, Chestnut Hill, Massachusetts 02467, United States
[§]Materials Science Division, Argonne National Laboratory, Argonne, Illinois 60439, United States
[∥]Schrödinger, Inc., New York, New York 10036, United States

Ⓢ *Supporting Information*

ABSTRACT: We present a new computational strategy for the design and evaluation of novel enzymatic pathways for the biosynthesis of fuels and chemicals. The approach combines the use of the Biochemical Network Integrated Computational Explorer (BNICE) framework and a structure-based screening method for rapid generation and evaluation of novel enzymatic reactions and pathways. The strategy is applied to a case study of 1-butanol production from pyruvate, which yielded nine novel biosynthetic pathways. Using screening criteria based on pathway length, thermodynamic feasibility, and metabolic flux analysis, all nine novel pathways were deemed to be attractive candidates. To further assess their feasibility of implementation, we introduced a new screening criterion based on structural complementarity using molecular docking methods. We show that this approach correctly reproduces the native binding poses for a wide range of enzymes in key classes related to 1-butanol production and provides qualitative agreement with experimental measures of catalytic activity for different substrates. In addition, we show that the structure-based methods can be used to select specific proteins that may be promising candidates to catalyze novel reactions.
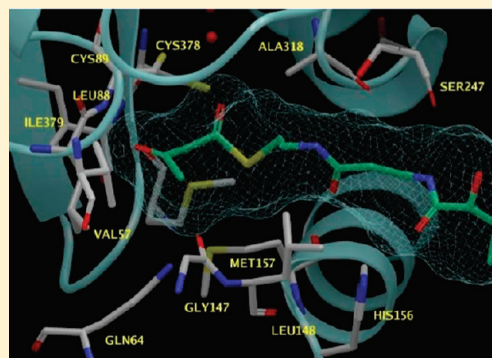
1. INTRODUCTION

1.1. Background. Growing energy demand, ongoing depletion of petroleum sources, geopolitical instability, and concerns over climate change have stimulated efforts to develop alternative energy sources as substitutes for today's liquid transportation fuels.[1] Bioethanol is one of the most widely used alternative energy sources today.[2] However, because of its low energy density and high hygroscopicity, ethanol is not an ideal fuel to replace or even to blend with gasoline at a higher ratio for existing engines and storage and distribution systems.[2] 1-Butanol, on the other hand, has a higher energy density, lower water absorption, better blending properties, and shows more similarity to gasoline.[2] It has several advantages over ethanol as a biofuel for blending with, or even replacing, gasoline as a primary bioenergy source for the current liquid fuel transportation system.[2]

Traditionally, 1-butanol is produced via fermentation through acetone—butanol—ethanol fermentation (ABE fermentation) by *Clostridium acetobutylicum*, which is an anaerobic, gram-positive, heterofermentative, and spore-forming bacterium. *Clostridium acetobutylicum* is the most widely studied starchy-material bacterium suited to perform such fermentation.[2,3] *Clostridium*

acetobutylicum is able to utilize a number of carbon resources.[4−6] Numerous enzymes secreted by *Clostridia* break complex sugars down to monosaccharides.[6] Monosaccharides are transported into the cell via a phosphoenolpyruvate-dependent phosphotransferase system.[2] Hexoses are converted to pyruvate through the Embden—Meyerhof—Parnas (EMP) pathway.[7] Pyruvate is a key intermediate in *Clostridium acetobutylicum* metabolism. Although pyruvate can be converted to lactate by lactate dehydrogenase under certain conditions, most of the pyruvate is transformed to acetyl-CoA and CO_2 by pyruvate—ferredoxin oxidoreductase. Acetyl-CoA undergoes a series of transformations to produce either oxidized products such as acetone, acetate, or CO_2, or reduced products including 1-butanol, ethanol, or butyrate.[8,9] The native pathway of *Clostridium acetobutylicum* takes seven steps to produce 1-butanol from pyruvate.[2] Pyruvate ferredoxin oxidoreductase cleaves the carboxyl group of pyruvate and transfers it to coenzyme A to form acetyl-CoA. Two acetyl-CoAs interact with acetyl-CoA acetyltransferase to form acetoacetyl-CoA. 3-Hydroxybutyryl-CoA dehydrogenase hydrogenates the keto group that is away from the CoA

moiety to produce 3-hydroxybutyryl-CoA. Crotonase and butyryl-CoA dehydrogenase convert 3-hydroxybutyryl-CoA to crotonyl-CoA and butyryl-CoA subsequently. Aldehyde/alcohol dehydrogenase acts on butyryl-CoA to remove CoA and generates 1-butanal. Lastly, aldehyde/alcohol dehydrogenase hydrogenates the aldehyde group to form the final product 1-butanol.

Metabolic engineering has highlighted the feasibility of manipulating microbial host systems for production of industrially high-valued chemicals.[10,11] However, our knowledge of other novel biosynthetic pathways that remain to be discovered or potentially engineered is still very limited.[12] Even for well studied organisms such as *E. coli*, novel pathways continue to be discovered.[12,13] Atsumi et al. demonstrated an excellent example of metabolic engineering by hijacking and altering the 2-keto acid biosynthetic pathway to produce a number of short chain alcohols in *E. coli*.[12] Novel pathways for industrial chemicals such as 3-hydroxypropanoate and 7-carboxyindole and novel biodegradation pathways for 1,2,4-trichlorobenzene have been identified using in silico methods that will be employed here.[14−17] Given the diversity and wealth of biochemistry and enzymes, alternative novel biosynthetic pathways from pyruvate to 1-butanol may exist and offer advantages over the native pathway.[17] In this work, we present a computational approach to design, evaluate, and screen novel biosynthetic pathways from pyruvate to 1-butanol at both pathway and reaction levels.

1.2. Computational Framework. Computational tools are increasingly used to complement experimental approaches in pathway design and evaluation. There are two major groups of tools that are used to discover alternative novel reaction pathways. The first group is mainly based on databases of known biochemical compounds and reactions such as the Kyoto Encyclopedia of Genes and Genomes (KEGG),[18] Encyclopedia of Metabolic Pathways (Metacyc),[19] and BRENDA.[20,21] The tools of the second group are able to generate novel enzymatic reaction pathways consisting of novel reactions based on a set of reaction rules. They are the RDM method,[22] the META method,[23] the PPS system,[24] and the Biochemical Network Integrated Computational Explorer (BNICE) framework.[14,17,25,26] The META method, the PPS system, and the BNICE framework rely on reaction rules that are curated from databases of known biochemical reactions. The RDM method generates reaction rules using a molecular structure comparison algorithm.[22] Unlike the META method and the PPS system that focus on metabolism of xenobiotics, the BNICE framework uses reaction rules based on curation of known biochemical reactions.[27] The BNICE framework is based on graph theory that creates complex networks of compounds and reactions using the generalized reaction rules or operators.[17,25,26,28−34] Substrates and products are represented as matrices, and reactions are carried out through addition of matrices of substrates and reaction rules.[17,25,26,31,32,34−36] Each enzyme is given a four digit classification (*i.j.k.l*) derived from the Enzyme Commission (EC) depending on its specific functional group transformations of the substrates.[37] The first index of the classification specifies the main class of the enzyme and identifies its primary function. The second index indicates the functional groups on which the enzyme acts. The third and fourth indices identify the cofactors and the substrates associated with the enzymatic reactions, respectively. The generalized enzyme reaction rules are designed based on the first three levels of enzyme classification. The generalized enzyme reaction rules are not substrate specific; therefore, they describe the transformation of functional groups. This concept allows enzyme functions to be

applied to not only the specific substrates but also to a broader range of substrates that share the same functional groups.[17,25,26,33,38−40] When applied to a starting substrate and its progeny, all possible enzymatic reactions given these generalized reaction rules are unfurled, including a large number of novel reactions.[15−17,25−27,33,38−43]

For a number of different systems, it has been shown that the computational framework identifies hundreds to thousands of novel pathways leading from a specified starting compound to a particular target compound.[14,26,33] Various screening methods were applied to prune this host of novel pathways to a manageable set of attractive candidates, including limits on pathway length and the number of novel intermediates, as well as analysis of thermodynamic landscape and pathway flux analysis.[14,39,42] To further reduce the number of novel pathway candidates, it would be valuable to assess the feasibility of engineering the reactions comprising novel pathways in a host organism. A critical piece of this effort is to identify specific proteins from a given *i.j.k* enzyme class that can be engineered to carry out catalysis of non-native substrates using a structure-based screening method.

Molecular docking is one of the structure-based methods that is extensively used in drug design to evaluate the potential of small molecule inhibitors to bind to active sites of therapeutically interesting protein targets.[44−49] This approach is most often employed to predict the preferred orientation of small molecules to their targeting proteins in order to estimate the strength of association.[44] It has shown promise in predicting the substrates of enzymes.[50] Given that binding is necessary but not sufficient for catalysis, we developed a structure-based screening method to complement the BNICE framework in order to assess the feasibility of novel pathways. While the generalized reaction rules of BNICE can enumerate the enzyme-catalyzed reactions required to transform a given reactant to a specific product and identify the enzyme classes capable of achieving the transformations, the availability of particular proteins that can be used to catalyze the novel reactions is more difficult to assess. Recently, significant effort has been invested to extend docking methods to the study of enzyme—substrate specificity. In particular, docking approaches were successful in identifying the substrates of α/β hydrolases,[51] amidohydrolase,[48] and dehydrogenase/reductase,[49] as well as those of enzymes of the enolase superfamily.[52−55] Docking was also employed to study the function of unknown proteins.[56,57]

2. METHODS

2.1. BNICE Formalism and Reaction Network Generation. The native pathway of *Clostridium acetobutylicum* was used as a reference to select the generalized enzymatic reaction rules. Forty-one generalized reaction rules were chosen from the same families as the native enzymes at the *i.j.k* level. These 41 operators were created based on curation of the KEGG database,[17,26,58] bioremediation chemistry,[14,16,27] and metabolic models of *E. coli*.[39,59] Prototypical reactions that show the specific transformations encoded by each operator are summarized in Table S1. This set of operators reproduces 652 reactions or 75.03% of the reactions in the selected *i.j.k* families from the KEGG database (Table S2). The network generation was initiated using pyruvate as the starting carbon source, and the additional cofactor compounds shown in Table S3, which are actively involved in this type of biochemical transformation. Network generation was carried out for 10 generations. A pathway identification algorithm that is resident in BNICE searched all possible linear routes to
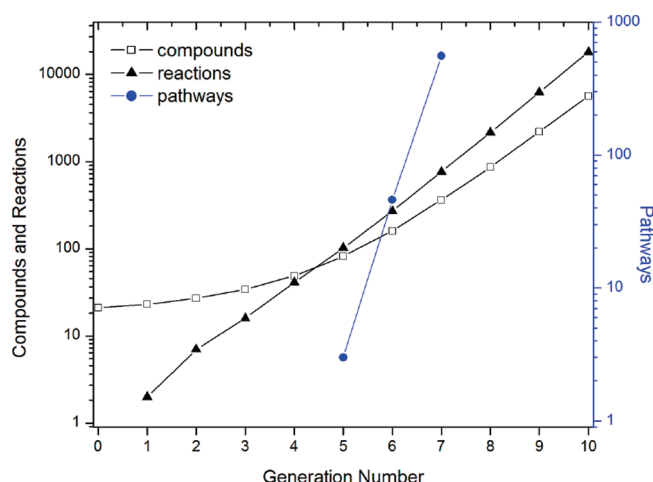
Figure 1. Number of compounds, reactions, and pathways generated by the BNICE protocol as a function of the generation number.

produce 1-butanol from pyruvate. Nonlinear branched pathways that converge pathways with different intermediates to a single intermediate were not considered by the algorithm.

2.2. Structure-Based Screening Methods. All 3D structure-based screening in this work was performed using the tools from the Schrödinger Suite 2009 of molecular modeling software.

2.2.1. Protein Selection. In the case of novel pathways, specific PDB structures of proteins within the enzyme classes appearing in the novel pathway reactions enumerated by BNICE were selected for structural analysis in the following way. For a given *i.j.k* class, all available PDB entries were grouped on the basis of the protein and ligand identity and sequence similarity. All PDB entries containing incomplete structures, the apo proteins, or the holo structures with inhibitors or ligands other than the native substrate were eliminated. PDBs corresponding to complexes of distinct proteins with their native substrates were retained, resulting in ~100 structures per PDB class. From these, a diverse sample was selected in the following manner. The catalytic site in each complex was identified by the position of the cocrystallized native substrate. The structural similarity of the catalytic binding sites for each pair of complexes from the pool was computed using the Pocketmatch program.[60] PDBs with similar binding pockets were grouped through clustering, and a single PDB entry was selected from each cluster. This provided ~10 complexes spanning a wide range of binding pockets within a given enzyme class to be tested for catalytic potential against novel ligands.

2.2.2. Protein Structure Preparation. The coordinates of the protein and bound substrate heavy atoms were obtained from X-ray crystal structures of the enzyme/substrate complexes available in the protein data bank (PDB). The Protein Preparation workflow within the Maestro molecular modeling environment was used to add hydrogens, assign bond orders and protonation states, and optimize the hydrogen bonding networks.[61] In cases where only mutant structures were available in the PDB, the mutated residues were converted to their wild type equivalents using Prime.[62] For computational efficiency, the protein domains far removed from the binding site of interest were truncated. All crystal waters beyond 5 Å from the binding ligands were deleted, and the remaining water orientations were sampled exhaustively. Finally, the protein−ligand−water complexes were minimized within an rmsd of 0.30 Å with the force field OPLS 2005.
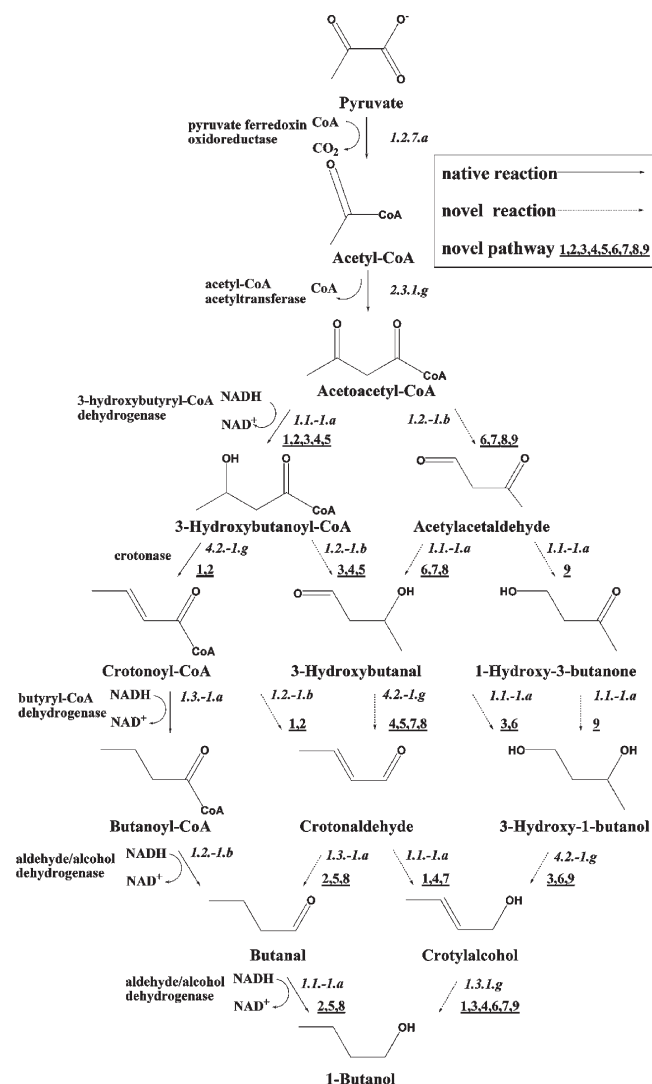

Figure 2. Enzyme-catalyzed reactions comprising the native and novel pathways for synthesis of 1-butanol from pyruvate.

2.2.3. Ligand Preparation. The ligands were prepared using the LigPrep module in the Schrodinger Suite.[63] All possible ionization states were generated at a target pH of 7.0 ± 2.0 using the ionizer tool. For chiral ligands, specified chiralities were retained. The ligand conformations were then generated using the MacroMo-del package by a mixed torsional/low-mode sampling method with the OPLS 2005 force field in implicit water without any constraints.[64] The five lowest potential energy conformers were retained as input for docking studies.

2.2.4. Docking. The docking of the ligands to their receptor targets was performed using the Glide package.[45,47,65] Various docking algorithms and parameters were tested to identify an optimal docking protocol that can achieve the best accuracy, efficiency, and feasibility in the large receptor−ligand complex reservoir. In each case, the protein domains, waters, enzyme metal ions, and cofactors were mapped onto a grid representation of the receptor. The grids were centered on the centroid of the cocrystallized ligand from the corresponding PDB structure, and the docked ligands were assumed to be of similar size to the cocrystallized ligand. The docking itself was performed in the extra-precision (XP) mode flexibly using expanded sampling,
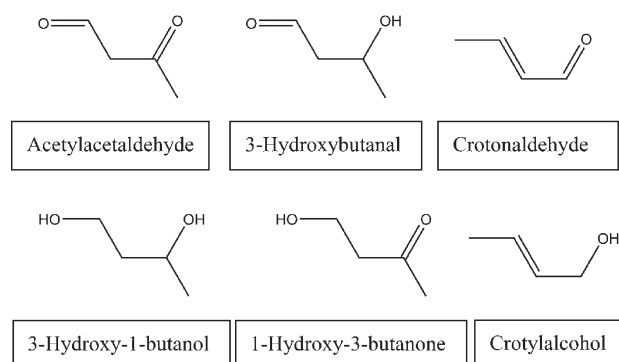
Figure 3. Compounds discovered as intermediates in the novel pathways that are novel to the KEGG database.

Figure 4. Thermodynamic landscapes for the native and novel pathways to produce 1-butanol from pyruvate.

which bypassed the elimination of poses on the basis of the rough docking score. Only the low-energy ring conformations were kept for ring conformation sampling. Ring conformations were discarded if their energies were higher than that of the lowest conformation by more than 2.5 kcal/mol. No positional, H-bond/metal, or hydrophobic constraints were applied. The atomic partial charges of the ligand were obtained directly from the docking force field. The van der Waals radii of nonpolar atoms, whose partial charges were less than 0.15, of the ligand and the receptor were scaled by a factor of 0.8 to decrease penalties for close contacts. The docked poses were scored using the XP GScore scoring function, and the best scoring pose was used for subsequent analysis. In the case of novel pathways, additional grids were generated with all crystal waters removed, and a separate set of dockings into these grids was performed using the same procedure as described above.

3. RESULTS AND DISCUSSION

3.1. Analysis of Reaction Network. A reaction network consisting of 5706 compounds and 18136 reactions was generated by BNICE after 10 iterations using the selected 41 generalized enzyme reaction rules. The growth in both the numbers of compounds and reactions was dramatic as the generation number increased. The pathway identification algorithm identified 3 five-step, 46 six-step and 556 seven-step pathways. All the above results are shown in Figure 1. These pathways were then pruned on the basis of the criteria that they must be shorter or equal in length to the native pathway with only pyruvate as the sole carbon source. Nine novel pathways that fit these screening criteria resulted. The nine novel pathways and the native pathway are summarized in Figure 2. All novel pathways are comprised of seven steps and share the same overall reaction as the native pathway. All pathways share the same first two enzymatic reaction steps, which first convert pyruvate to acetyl-CoA by operator 1.2.7 and then convert acetyl-CoA to acetoacetyl-CoA by operator 2.3.1. The reactions and compounds of the novel pathways were compared to those in the KEGG database to determine their novelty in the context of known biochemical reactions. Six compounds novel to KEGG were identified as summarized in Figure 3. In addition, all 12 of the reactions that were in the novel pathways but not part of the native pathway were novel to KEGG.

The millimolar Gibbs free energy change of each reaction comprising all 10 pathways was determined using the group contribution method to assess the thermodynamic feasibility.[14,39,41,66]
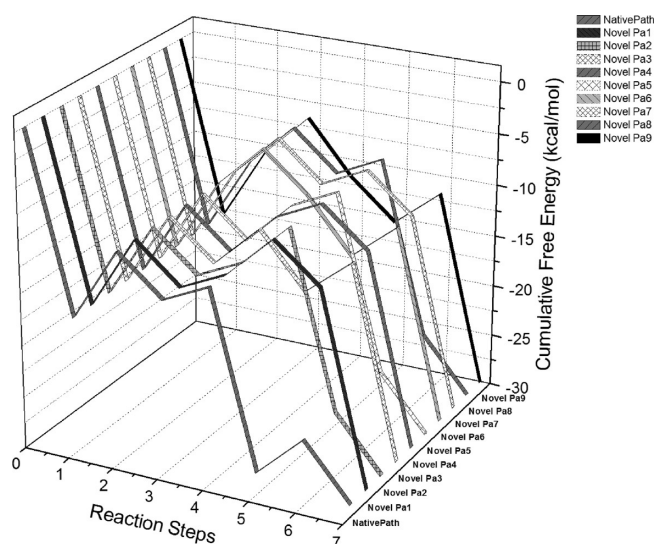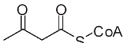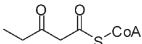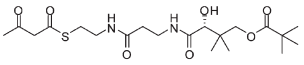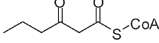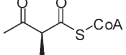
The group contribution method was developed for fast estimation of ΔG_f and ΔG_r for biological compounds and reactions in aqueous phase at pH 7 and room temperature.[66−69] The reference state of 1.0 mM was used to better represent cellular physiological conditions.[70] All pathways have the same cumulative free energy ($\Delta G_{rxn} = -28.78$ kcal/mol) because they adhere to the same overall reaction. However, as shown in Figure 4, the thermodynamic landscapes are distinct. If there were substantial endogonic barriers to be surmounted for a given pathway, then it could be discarded as thermodynamically infeasible. However, only modest barriers were observed for each of the novel pathways, so all nine were retained for the next stage of screening.

The next screening criterion used thermodynamics-based metabolic flux analysis (TMFA) to predict the maximum yield of 1-butanol from glucose in *Clostridium acetobutylicum*.[33,39,41,42,59,71] There are 422 intracellular metabolites, 552 reactions, and 80 membrane transport reactions in the genome-scale *Clostridium acetobutylicum* model.[9] The concentrations of the extracellular nutrient sources in the model were specified on the basis of a synthetic medium in Table S4.[72] Novel pathway reactions replaced the native pathway reactions in the metabolic model to establish the potential of the novel pathways to rival the native pathway. Thermodynamics-based metabolic flux analysis (TMFA) was applied to assess the yield of 1-butanol for each distinct pathway in turn on the basis of anaerobic conditions with glucose set to be 10 mmol/g DCW as the only carbon source.[73] The maximum yield of 1-butanol for the native pathway was 14.17 mmol/g DCW hour, and all novel pathways resulted in the same yield as the native pathway. Therefore, to further estimate the feasibility of the alternative novel pathways, the potential of existing proteins to catalyze new reactions appearing in these pathways was evaluated using structural screening.

3.2. Structure-Based Screening of Reaction Pathways. To validate our approach, we first examined several reactions along the native pathway for fermentation synthesis of 1-butanol from pyruvate. In particular, we considered (i) the conversion of acetyl-CoA to acetoacetyl-CoA, catalyzed by acetyltransferase (EC 2.3.1); (ii) dehydrogenation of acetoacetyl-CoA to 3-hydroxybutanoyl-CoA by alcohol dehydrogenase (EC 1.1.-1); and (iii)

Table 1. Summary of Docking Results for Select Substrates of Acetyltransferase

Substrate		Catalytic activity[a]	Docking score	EFF(Å²)[b]
acetoacetyl-CoA		100%	-16.75	0.044
C5-3-oxoacyl-CoA		50.0%	-14.03	0.015
acetoacetyl-S-patetheine-11-pivalate		33.0%	-12.65	0.230
C6-3-oxoacyl-CoA		6.00%	-11.87	10.495
2-methyl-acetoacetyl CoA		<0.100%	-11.29	2.717

[a] Catalytic activity was derived based on thiolisis Vmax ratios measured for *Z. ramigera* and *A. eutrophus* thiolase.[76] [b] The EFF was computed based on the following key distances: O1(ligand carbonyl)-O(water 82), O1(ligand carbonyl)- Nε2(His348), O3(ligand carbonyl)-N(Cys89), and O3(ligand carbonyl)-N(Gly380).
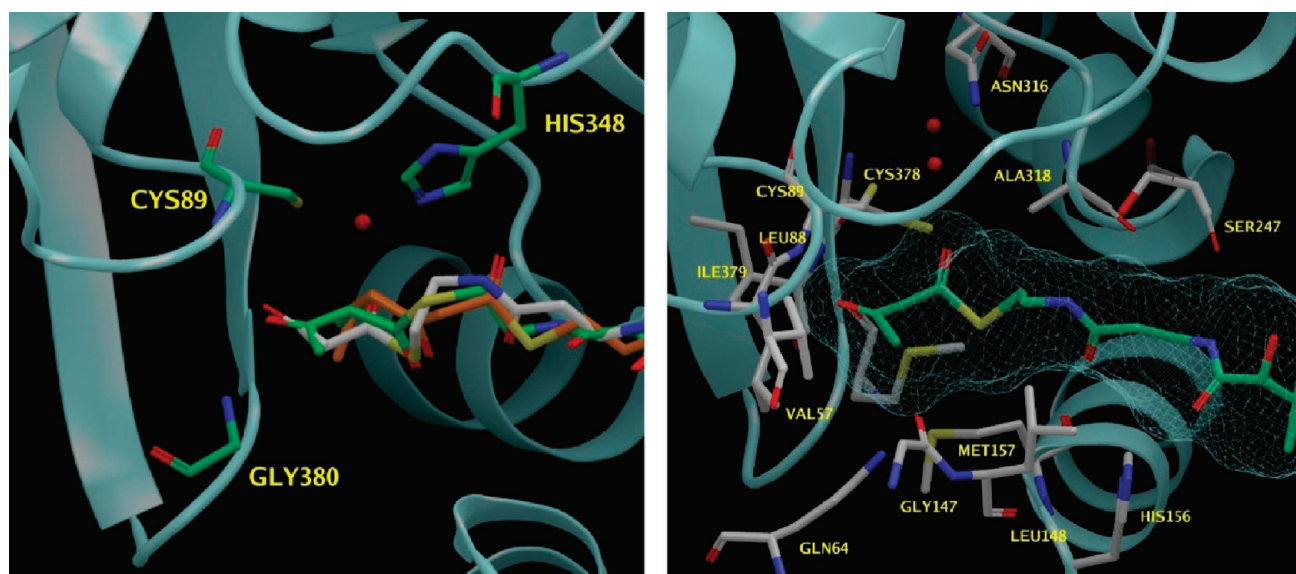

Figure 5. Binding pocket of the acetyltransferase with PDB code 1M1O. The left figure shows the three residues, Cys89, His348, and Gly380, involved in the key interactions of catalysis and the crystal geometry of the substrate acetoacetyl CoA (white), the docking geometries of the ligands 2-methyl-acetoacetyl CoA (green), and C6-3-oxoacyl CoA (orange). The right figure presents the rigid binding pocket of the protein 1M1O with the native substrate acetoacetyl CoA.

conversion of 3-hydroxybutanoyl-CoA to crotonoyl-CoA by dehydrolyase (EC 4.2.1), as illustrated in Figure 2. In each case, specific enzymes were chosen for which the structure of the protein in complex with its native substrate was available in the PDB and for which substantial biochemical data on both active and inactive ligands was published. While the docking scoring functions such as the Glide XP GScore were primarily designed to estimate the binding affinity of the ligand to a target receptor, the catalytic activity of an enzyme is strongly dependent on the binding geometry of the reactant substrates and cofactors relative to the residues involved in the reaction. In particular, Macchiarulo et al. have recently shown through a series of cross-docking studies that native substrates do not necessarily form the most stable complexes with the cognate enzymes.[74] Hence, to facilitate comparison of the docked substrates with respect to their potential for

catalytic activity, we have introduced the enzyme fitness function (EFF)

$$EFF(L) = \frac{1}{N} \sum_{j=1}^{N} (d_j^L - d_j^S)^2 \qquad (1)$$

where d_j^L are the distances between the N key atoms j of the ligand and the protein known to be involved in the catalytic reaction, while d_j^S are the distances between the corresponding substrate—protein atom pairs as observed in the crystal structure. Hence, the EFF measures the deviation of key ligand atoms involved in the reaction from the ideal native substrate geometry.

3.2.1. Acetyltransferase. Acetyltransferase catalyzes the addition of an acetyl group to acetyl-CoA to form acetoacetyl-CoA. Structural, biochemical, and mutant studies on *Zooglea ramigera*
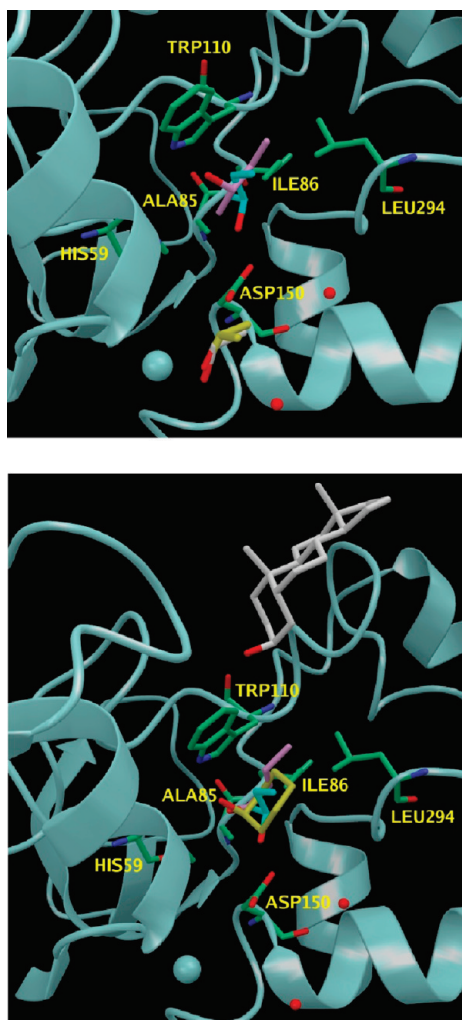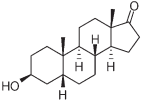
Figure 6. Binding pocket of the alcohol dehydrogenase 1BXZ. The enzyme residues His59, Ala85, Ile86, Trp110, Asp150, and Leu294 are involved in the key interactions of catalysis. The top figure shows the protein in ribbon and the crystal geometry of the native substrate sec-butanol (blue), the docking geometries of ligands sec-butanol (plum), isopropanol (yellow), and ethanol (white). The bottom figure includes the crystal geometry of the native substrate sec-butanol (blue) and the docking geometries of ligands 1-butanol (plum), cyclohexanol (yellow), and 5β-A-3β-OH(white).

biosynthetic thiolase indicate that the catalytic site is located in a very narrow and rigid cavity, which directs the reactive O3 of the acetoacetyl group into an oxyanion hole formed by the NH groups of Cys89 and Gly380, while the thioester O1 is stabilized by hydrogen bonds to a nearby tightly bound water molecule and an NH of His348.[75] The reaction is primarily facilitated through the interaction of the substrate with Cys89, Gly380, and tightly bound crystal waters, conferring a high degree of substrate specificity on this enzyme. Starting with the *Zooglea ramigera* biosynthetic thiolase mutant complexed with the native substrate (PDB 1M1O), we have prepared the protein as described in the methods section, with crystal waters proximal to the binding site retained, and docked the native substrate acetoacetyl-CoA and several additional substrates summarized in Table 1, for which the catalytic activity with this enzyme had been measured. To better quantify the catalytic potential of bound poses, the EFF

scores were computed based on the key distances described above and listed in Table 1 along with the Glide XP GScore docking scores. While the docking scores were useful in selecting the most likely binding geometry, in general they show poor correlation with the catalytic activity data. This is not unexpected because, as pointed out above, formation of a stable, strongly bound complex does not necessarily imply a catalytically active state can be achieved. Hence, the subsequent analysis focuses on the EFF score, while the docking scores are provided for completeness. The top scoring docking pose of the native acetoacetyl-CoA was found to closely resemble the crystal structure geometry, which is also indicated by an excellent EFF score of 0.04. On the other hand, closely related 2-methyl-acetoacetyl-CoA binds with a somewhat different geometry because the extra methyl group prevents the key substrate oxygens from reaching deep enough into the narrow binding pocket, which is reflected in a higher EFF score. This agrees well with experimental measurements, which show 2-methyl-acetoacetyl-CoA to have less than 0.1% of the catalytic activity of the native substrate. The narrow cavity of the active site allows some tolerance in the acetoacetyl moiety of the substrate. In particular, the binding pose of C6-3-oxoacyl-CoA (Figure 5), which is nearly inactive, exhibits a large EFF score of 10.50 because the bulky acetoacetyl moiety is too large to fit into the binding pocket, preventing the oxygens from reaching the catalytically active geometry, while the binding pose of C5-oxoacyl-CoA, which shows approximately half of the activity, presents a lower EFF score with one carbon less. On the other hand, acetoacetyl-S-patetheine-11-pivalate, which shows modest activity, exhibits an intermediate EFF score. Hence, our docking approach was able to correctly reproduce the catalytically active binding geometry of the native substrate and distinguish the active substrates from decoys through EFF analysis of the binding pose geometry.

3.2.2. Alcohol Dehydrogenase. Unlike the thiolase, alcohol dehydrogenases have less specific binding pockets generally capable of accommodating several small alcohol substrates. The hydrophobically lined binding site assures some functionality is retained even in organic solvent environments, while the catalytic activity is usually centered on the Zn^{2+} in the rear of the pocket.[77] Here, we considered the thermophilic bacteria alcohol dehydrogenase (TBAD) because of the ample structural and biochemical data available for this enzyme. Crystallographic data shows that the native substrate sec-butanol binds such that the C1 carbon fits well into the hydrophobic cavity formed by Leu294, Trp110, Ala85, His59, and Asp150. This in turn positions the reactive oxygen of the alcohol hydroxyl group to interact with Asp150 and His59 via hydrogen bonds, which facilitates the catalysis and is likely the reason why secondary alcohols show better catalytic activity. Interestingly, unlike in eukaryotic alcohol dehydrogenases, the Zn^{2+} in TBAD is ~7 Å distant and does not form direct contacts with the substrate. Starting with the crystal structure of TBAD in complex with sec-butanol (PDB 1BXZ), we have docked both the native substrate as well as several other small alcohol and steroid substrates for which the bioactivity has been measured. In Figure 6(top), we show the docked poses of sec-butanol, isopropanol, and ethanol overlaid on the crystal structure, while in Figure 6(bottom) we show those of butanol, cyclohexanol and 5β-A-3β-OH. The corresponding EFF scores based on the key distances described above are summarized along with the Glide XP GScore docking scores in Table 2. The docked pose of the native substrate is well-aligned with the crystal structure coordinates, and the low EFF

Table 2. Summary of Docking Results for Select Substrates of Alcohol Dehydrogenase

Substrate		Km (mM)[a]	Docking Score	EFF(Å²)[b]
sec-butanol		-	-2.65	2.408
5β-A-3β-OH		0.0185	-4.18	41.168
cyclohexanol		1.15	-3.87	1.334
ethanol		2.13	-3.03	28.924
butanol		8.75	-3.44	0.778
isopropanol		76.4	-2.81	26.685

[a] Km (mM) values for rat liver alcohol dehydrogenase, as reported in ref 78. [b] The EFF was computed based on the following key distances: O(ligand hydroxyl)-Nε2(His59), O(ligand hydroxyl)-OD1(Asp150), C1(ligand)-Cε2(Trp110), and C1(ligand)-CD1(Ile86).

score indicates that our docking approach reproduces the correct catalytically active geometry of sec-butanol very well. This is especially encouraging as small molecules with few heavy atoms are very challenging for docking methods because of the multitude of possible orientations such substrates can take in the binding site of the receptor. Ethanol, which is marginally reactive, and isopropanol, which has very low measured catalytic activity, both show higher values of EFF than sec-butanol. In particular, isopropanol is found to be unable to assume a catalytically active geometry due to trapping in the electrophilic zinc ion site, which shows that our docking approach can successfully screen out poor substrates even in cases where the latter are structurally very similar to the native, catalytically active compound. Of the small alcohol substrates tested, cyclohexanol and butanol show lower EFF scores than the native substrate, and the former was also found to be the most active of these alcohols with rat liver alcohol dehydrogenase. On the other hand, the steroid 5β-A-3β-OH had failed to dock with a geometry presumably necessary for catalytic reactivity, even though it showed substantial activity in the RLADH assay. There are several reasons for these discrepancies with the assay data. First, structural differences between the thermophilic bacteria ADH, the crystal structure of which was used in docking studies, and that of RLADH used in biochemical measurements, particularly the positioning of the Zn²⁺ in the binding site, would likely lead to differences in the catalytically active binding geometry. This is particularly true for steroids, which bind in a highly stereospecific manner. Moreover, difficulty with protein stabilization and sample purity in the early study introduces a fair degree of uncertainty in the reliability of assay data. In spite of these issues, the strong correlation of EFF scores with biological data testifies to the robustness of our approach at identifying catalytically active substrate/enzyme pairs.

3.2.3. Enoyl-CoA Hydratase. The enoyl-CoA hydratase catalyzes the hydration of α,β-unsaturated-CoA in a stereospecific fashion. Structural and mutant studies[79] indicate that the oxyanion hole formed by Ala98 and Gly141 backbone amides as well
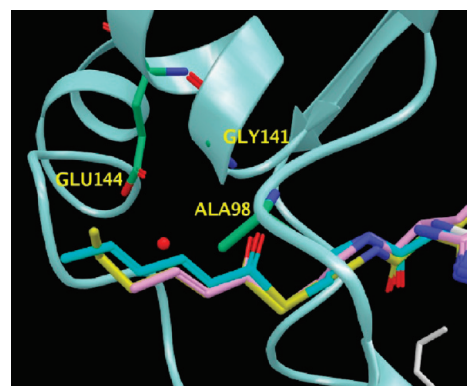
Figure 7. Binding site of protein 1MJ3. The figure shows the key interacting residues Ala98, Gly141, and Glu144 in catalysis and the crystal geometry of the native substrate hexanoyl CoA (blue) and the docking geometries of ligands crotonoyl CoA (plum), hexenoyl CoA (yellow), and octenoyl CoA (white).

as Glu144 and Glu164 are essential for stabilizing the transition state and binding the catalytic water, respectively, and thus are crucial for catalytic activity of the enzyme. Stereospecificity is believed to be achieved via destabilization of the *cis* conformer ground state with respect to the *trans* conformer through polarization of the substrate by the local enzyme field. Starting with the crystal structure of the rat mitochondrial enoyl-CoA hydratase (PDB 1MJ3) cocrystallized with the substrate analog hexanedienoyl-CoA, we have docked the native substrate, *trans*-2-crotonoyl-CoA, hexanedienoyl-CoA, and several longer chain fatty acid ligands and analyzed the best scoring binding poses. In Figure 7, we show the binding poses of *trans*-2-crotonoyl-CoA, hexanedienoyl-CoA, and 8-enoyl-CoA overlaid on the crystal structure coordinates. Both the native substrate and the self-docked analog dock in a pose closely resembling the crystal structure. In particular, the reactive carbonyl is well positioned with the oxygen of the reactive carbonyl occupying the oxyanion hole and the carbonyl C3 carbon properly oriented for the catalytic addition of a water molecule, as demonstrated by the respective EFF scores shown along with the Glide XP GScore docking scores in Table 3. A similar result is obtained for 6-enoyl-CoA. On the other hand, the longer carbon chain of 8-enoyl-CoA causes the latter ligand to assume a very different binding geometry with a much higher EFF score. The remaining ligands with still longer chains fail to dock in the binding site of 1MJ3. This is in agreement with experimental observations that show that the catalytic hydration rate falls rapidly with increase in the length of the fatty acid tail.

3.3. Structural Analysis of Novel Pathways. We next considered the novel pathways for production of 1-butanol from pyruvate generated by BNICE. A total of 10 pathways that passed the initial thermodynamic screening were found to contain 12 novel reactions, six of which belonged to the 1.1.-1 or 4.2.1 class. The six new intermediates involved in these reactions are shown in Figure 3. Four of these reactions are catalyzed by enzymes from the 1.1.1 class and two from the 4.2.1 class. However, the BNICE classification does not specify the specific enzymes within the given class to be used. Another advantage offered by the structural screening is to help identify which particular proteins, if any, within a class identified by BNICE most closely "fit" the novel ligands and provide the greatest potential to be engineered into viable platforms to

Table 3. Summary of Docking Results for Select Substrates of Enoyl-CoA Hydratase

Substrate		Km (mM)[a]	Docking score	EFF(Å^2)[b]
hexanoyl-CoA		-	-17.11	0.024
crotonoyl-CoA		0.020	-14.58	0.042
hexenoyl-CoA		0.24	-16.64	0.047
octenoyl-CoA		0.28	-9.56	180.365
decenoyl-CoA		0.30	-10.95	148.423
dodecenoyl-CoA		0.40	-12.64	245.969
hexadecenoyl-CoA		0.50	-6.15	209.594

[a] As reported in ref 80. [b] EFF was computed based on the following key distances: C3(ligand)-Wat960, O(ligand carbonyl)-N(Ala98), and O(ligand carbonyl)-N(Gly141).

catalyze the novel reactions. We note that structural screening as presented in this study only evaluates the ability of the ligands to bind to the receptor in a geometry consistent with the known catalytic mechanism. This in itself, however, does not guarantee that catalysis will actually take place, but merely confirms that the prerequisites for making the reaction possible have been met. For example, the chemical features of the ligand may still make the reaction energetically or kinetically unfavorable. One approach for refining the structure-based screening is to incorporate detailed reaction energetics (such as obtained from quantum mechanics/molecular mechanics (QM/MM) calculations) as well as information on known reaction patterns into our scoring function. Both of these will be included in a future study. A survey of the PDB showed several hundred entries for each of the above enzyme classes. Out of these, a subset of enzymes was selected representing a diverse set of binding sites for which the binding mode of the native substrate is known, as described in the Methods section. This process yielded eight enzyme structures for the 1.1.1 class and four for the 4.2.1 class. Each of the reactants participating in the novel reactions catalyzed by a given enzyme class was then docked into each PDB structure from that class.

3.3.1. Novel Reactions Catalyzed by 4.2.1 Enzymes. The novel reactions catalyzed by the 4.2.1 enzymes include the transformation of 3-hydroxybutanal to crotonaldehyde and 3-hydroxy-1-butanol to crotonaldehyde. Because both of the reactants contain a chiral center, both the *S* and *R* enantiomers were considered in the structural screen. The four proteins of the 4.2.1 enzyme class that were selected as representatives from the PDB are summarized in Table 4. Also shown are the cocrystallized native substrates and the distances between the key substrate and protein atoms experimentally determined to be important in the respective catalytic reactions.

Both enantiomers of 3-hydroxybutanal and 3-hydroxy-1-butanol were then docked to each of the above proteins as described in the Methods section. Two grids were used to represent each receptor: in one case the crystal waters present in the binding site

Table 4. Structures of Representative 4.2.1 Class Enzyme−Substrate Complexes

protein	PDB	native substrate	key distances (Å)	
enolase	2ONE	2-phospho-D-glycerate	C2-Nζ(Lys345)	4.211
			HO-Oε2(Glu211)	2.846
aconitase	7ACN	isocitrate	HO-Oε2(Glu168)	2.802
			OHA-Oδ1(Asp165)	3.125
			OHA-Nε2(His101)	3.240
			Cα-NH2(Arg452)	4.744
			Cβ-Oγ(Ser642)	3.140
tartrate dehydratase	2DW7	S,R meso-tartaric acid	OH-NH2(Lys184)	3.215
			Cβ-Nδ1(His322)	4.227
enolase-Mg^{2+}	7ENL	phosphoglycerate	C2−O(Wat 525)	3.492
			O3−Oε2(Glu168)	3.005
			O3−Mg^{2+}	1.984

were retained, and in the other they were deleted. Arguments can be made for and against both of these approaches. On one hand, water molecules are known to participate in the catalytic reaction, and hence, proper ligand−water interactions are likely necessary for the reaction to proceed. On the other hand, the location of crystal waters is specific to the native substrates cocrystallized with the protein. It is plausible that in the case of novel ligands, catalytic waters may rearrange into a more accommodating conformation, and retaining crystal waters may place too stringent a restraint on the possible novel ligand binding geometries.

In Figure 8, we show the representative binding poses of 3-hydroxybutanal(R) and 3-hydroxy-1-butanol(R) superimposed on the crystal structure coordinates of 7ENL and 2ONE. Clearly, both ligands dock closer to the putative catalytically active geometry in 2ONE than in 7ENL. This is reflected in the EFF

1641

dx.doi.org/10.1021/ci2000659 |*J. Chem. Inf. Model.* 2011, 51, 1634–1647

Figure 8. Binding pockets of protein 7ENL (left) and 2ONE (right). Both panels include the crystal geometry of the native substrates (blue), the docking geometries of ligands 3-hydroxybutanal(R) (plum), and 3-hydroxy-1-butanol(R) (yellow).

scores which are summarized along with the respective Glide XP GScore docking scores for all the structures analyzed in Table 5. For both enantiomers of 3-hydroxybutanal and 3-hydroxy-1-butanol, the best overall EFF scores are obtained for 7ENL, followed by 2DW7, while the scores for 2ONE and 7ACN are significantly poorer. Only 3-hydroxy-1-butanol(S) was able to achieve a pose close to the active geometry in 2ONE, while all ligands showed significant deviation in 7ACN. Qualitatively similar results are obtained when docking in the presence of crystal waters and are shown in Table S5, but with overall poorer EFF scores due to restrictions on the binding geometry imposed by the bound water molecules. These results show that the two novel reactions predicted by BNICE are probably feasible, and the best candidates for engineering this 4.2.1 catalyzed reaction are the enolase 7ENL followed by tartrate dehydratase 2DW7. We note that the four enzymes examined all belong to the enolase superfamily. It is generally believed that the substrates in this family require a carboxylate group to be stabilized by the magnesium ion in order to form enolate intermediates.[81] Therefore, it is quite possible that 7ENL could not be engineered to catalyze 3-hydroxybutanal and 3-hydroxy-1-butanol as substrates. This example further demonstrates that while the structural screening approach is successful in identifying complexes which permit substrate binding geometries consistent with catalytically active states, additional strategies, such as QM/MM calculations and statistics-based scoring terms based on known reactivity patterns, are needed to filter out reactions that are energetically or kinetically infeasible.

3.3.2. Novel Reactions Catalyzed by 1.1.1 Enzymes. The novel pathways for 1-butanol synthesis generated by BNICE also contain five reactions catalyzed by the 1.1.1 enzymes, including the transformation of acetylacetaldehyde to 3-hydroxybutanal and 1-hydroxy-3-butanone, the reactions of 3-hydroxybutanal and 1-hydroxy-3-butanone to form 3-hydroxy-1-butanol, and the reaction of crotonaldehyde to form crotylalcohol. As in the case of 4.2.1, both enantiomers of 3-hydroxybutanal were included in the structural analysis. The eight proteins from the 1.1.1 class selected for structural screening are summarized in Table 6, along with the cocrystallized native substrates and the distances between key ligand and protein atoms identified as important for catalysis in the literature.

The five ligands (including both enantiomers of 3-hydroxybutanal) were docked into each of the proteins in Table 6.

Table 5. Docking and EFF Scores for Novel Reactants Catalyzed by 4.2.1 Enzymes

PDB	ligand	docking score	EFF (Å²)
2ONE	3-hydroxybutanal(R)	−5.54	4.956
	3-hydroxybutanal(S)	−6.10	4.478
	3-hydroxy-1-butanol(R)	−6.65	4.245
	3-hydroxy-1-butanol(S)	−6.64	0.887
7ACN	3-hydroxybutanal(R)	−7.63	4.283
	3-hydroxybutanal(S)	−6.61	1.896
	3-hydroxy-1-butanol(R)	−9.13	2.278
	3-hydroxy-1-butanol(S)	−9.33	1.954
2DW7	3-hydroxybutanal(R)	−4.60	2.742
	3-hydroxybutanal(S)	−4.61	0.594
	3-hydroxy-1-butanol(R)	−4.33	0.010
	3-hydroxy-1-butanol(S)	−4.36	0.648
7ENL	3-hydroxybutanal(R)	−5.37	0.207
	3-hydroxybutanal(S)	−5.15	0.069
	3-hydroxy-1-butanol(R)	−5.05	0.226
	3-hydroxy-1-butanol(S)	−4.58	0.069

Note that acetylacetaldehyde has two reactive carbonyl centers, which participate in the reactions of acetylacetaldehyde → 3-hydroxybutanal and acetylacetaldehyde → 1-hydroxy-3-butanone, respectively. Hence, each of these was evaluated in turn, and labeled acetylacetaldehyde(a) and acetylacetaldehyde(b). In Figure 9, we show representative poses for acetylacetaldehyde and 1-hydroxy-3-butanone superimposed onto the crystal structures of 1PJ2 and 3EEW. It is clear that 1-hydroxy-3-butanone fits much better into the latter structure, with the binding geometry closely matching the presumably catalytically active mode of the native substrate, whereas in 1PJ2 the reactive carbonyl oxygen is substantially displaced from the catalytic metal center. While similar behavior is observed for acetylacetaldehyde(a), conversely, the second reaction center of the novel ligand acetylacetaldehyde(b) appears to be much better positioned in 1PJ2 than in 3EEW, closely mimicking the geometry of the native malate. This observation is supported by the EFF scores, which are, along with those of other 1.1.1 ligands, given in Table 7. Hence, 3EEW would be a promising starting point for

Table 6. Structures of Representative 1.1.1 Class Enzyme−Substrate Complexes

| protein	PDB	native substrate	key distances (Å)	
mannitol dehydrogenase	1M2W	D-mannitol	O2-Nδ2(Asn191)	3.170
			O2-Nζ(Lys295)	2.896
			O2-Nδ2(Asn300)	3.009
NAD-dependent malic enzyme	1PJ2	malate	O3-Nz(Lys183)	2.783
			O3-Mn	2.213
gluconate dehydrogenase	3CXT	D-glucarate	O5-OH(Tyr163)	3.567
			C5-C4N(NAP273)	4.408
hydroxybutyrate dehydrogenase	3EEW	3S-3-hydroxybutanoic acid	O8-OG(Ser142)	2.707
			O8-OH(Tyr155)	2.926
20alpha-hydroxysteroid dehydrogenase	1MRQ	progesterone	O20-OH(Tyr55)	5.678
			O20-Nε2(His222)	3.049
			C20-C4N(NAP)	4.297
histidinol dehydrogenase	1KAE	histidinol	O-O(His367)	2.782
			O-Ne2(His327)	3.334
			C-C4N(NAD)	4.164
PdxA	1PS6	4-hydroxyl-L-threonine 5-monophosphate	OG1-Zn	2.211
			OG1-Nε2(His166)	2.979
			OG1-Nε2(His211)	3.252
D-lactate dehydrogenase	3KB6	lactic acid	O2-Nε2(His294)	2.778
			O2-NH1(Arg231)	2.918
			C2-C4N(NAD400)	3.459
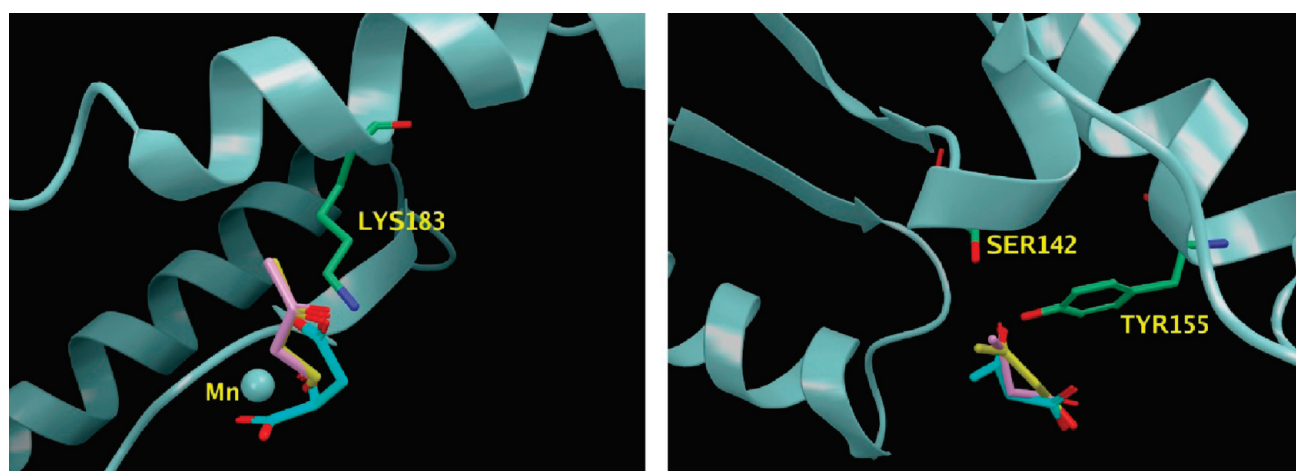
Figure 9. Binding pockets of protein 1PJ2 (left) and 3EEW (right). Both panels include the crystal geometry of the native substrates: malate and 3S-3-hydroxybutanoic acid (blue), the docking geometries of ligands acetylacetaldehyde (a/b) (plum), and 1-hydroxy-3-butanone (yellow).

bioengineering the enzyme catalyst for acetylacetaldehyde(a) and 1-hydroxy-3-butanone, while 1PJ2 would be a better choice for acetylacetaldehyde(b) and 3-hydroxybutanal(R). On the other hand, 3CXT exhibits low EFF scores for all novel ligands except acetylacetaldehyde(b) and may be the best all-around platform, while 1PS6 and 1KAE are clearly inferior with respect to all ligands. Qualitatively similar results are again obtained with crystal waters retained and are shown in Table S6, except in cases where the water molecules are clearly blocking the binding site of the novel ligand. Overall, the existence of 1.1.1 enzymes with binding sites that fit each of the ligands in the novel pathways predicted by BNICE suggests that the pathways are potentially

Table 7. Docking and EFF Scores for Novel Reactants Catalyzed by 1.1.1 Enzymes

PDB	ligand	docking score	EFF (Å^2)	PDB	ligand	docking score	EFF (Å^2)
1M2W	3-hydroxybutanal(R)	−5.75	6.711	1MRQ	3-hydroxybutanal(R)	−5.23	37.577
	3-hydroxybutanal(S)	−6.28	0.051		3-hydroxybutanal(S)	−5.43	0.272
	acetylacetaldehyde(a)	−6.32	6.548		acetylacetaldehyde(a)	−5.43	5.171
	acetylacetaldehyde(b)	−6.32	6.577		acetylacetaldehyde(b)	−5.43	0.034
	1-hydroxy-3-butanone	−6.14	0.025		1-hydroxy-3-butanone	−5.40	37.049
	crotonaldehyde	−4.49	6.639		crotonaldehyde	−4.97	0.053
1PJ2	3-hydroxybutanal(R)	−7.81	0.08	1KAE	3-hydroxybutanal(R)	−4.73	6.571
	3-hydroxybutanal(S)	−8.26	8.844		3-hydroxybutanal(S)	−5.33	14.777
	acetylacetaldehyde(a)	−7.80	9.371		acetylacetaldehyde(a)	−4.39	108.603
	acetylacetaldehyde(b)	−7.80	0.05		acetylacetaldehyde(b)	−4.39	48.199
	1-hydroxy-3-butanone	−8.00	9.336		1-hydroxy-3-butanone	−5.35	13.430
	crotonaldehyde	−7.26	0.079		crotonaldehyde	−4.16	48.510
3CXT	3-hydroxybutanal(R)	−5.32	0.625	1PS6	3-hydroxybutanal(R)	−6.17	23.957
	3-hydroxybutanal(S)	−5.08	0.359		3-hydroxybutanal(S)	−6.26	24.507
	acetylacetaldehyde(a)	−5.26	0.514		acetylacetaldehyde(a)	−5.43	29.841
	acetylacetaldehyde(b)	−5.26	2.92		acetylacetaldehyde(b)	−5.43	42.624
	1-hydroxy-3-butanone	−5.38	0.475		1-hydroxy-3-butanone	−5.91	25.774
	crotonaldehyde	−3.56	0.059		crotonaldehyde	−4.76	24.441
3EEW	3-hydroxybutanal(R)	−6.35	6.971	3KB6	3-hydroxybutanal(R)	−5.74	1.728
	3-hydroxybutanal(S)	−6.29	9.945		3-hydroxybutanal(S)	−4.91	0.661
	acetylacetaldehyde(a)	−7.19	0.051		acetylacetaldehyde(a)	−5.96	4.216
	acetylacetaldehyde(b)	−7.19	7.55		acetylacetaldehyde(b)	−5.96	0.181
	1-hydroxy-3-butanone	−6.35	0.039		1-hydroxy-3-butanone	−5.60	0.238
	crotonaldehyde	−5.62	7.433		crotonaldehyde	−3.82	4.441

viable and worth exploring further, both experimentally and computationally.

4. CONCLUSIONS

In this study, we have presented a strategy for in silico prediction of viable biosynthetic pathways, which combines a graph theory-based pathway generation framework with a structure-based screening methodology centered on molecular docking. Using the fermentation of 1-butanol from pyruvate as an example, we have shown that our approach is capable of generating competitive novel pathways and reactions. While thermodynamic analysis based on group contribution and metabolic flux analysis can be very useful for a first pass screen of the proposed pathways, a more detailed structural evaluation is necessary in order to assess whether the proposed reactions can be engineered using known enzymes and to identify particular protein candidates.

Molecular docking provides an ideal balance of accuracy and throughput to make it a desirable method for structure-based screening of novel biosynthetic pathways. While the native docking scoring functions such as Glide XP GScore are useful in ranking compounds on the basis of their binding affinity, in general they are not expected to correlate well with catalytic activity, a behavior that was observed in the present work. However, with the introduction of the enzyme fitness function to characterize the quality of the docked poses, the approach presented in this study was shown to be capable of distinguishing

the native substrate and other active ligands from inactive decoys for a diverse set of enzymatic reactions along the native 1-butanol synthesis pathway. More importantly, analysis of the novel pathways has shown that our structure-based method can be used to identify specific proteins within a given enzyme class that are most likely to catalyze a given novel reaction on the basis of the fitness with the respective substrate. This is particularly important as it allows prioritization of specific enzyme—substrate pairs for experimental testing and greatly improves the likelihood of finding a viable novel biosynthetic pathway. Overall, the results are very encouraging and point to docking as a feasible method for structure-based screening of biosynthetic pathways.

Still, further work is needed to develop the structure-based screening approach presented here into an automated high-throughput method capable of processing thousands of pathways. This includes the construction of a comprehensive database of protein—substrate complex structures spanning the known space of enzyme-catalyzed reactions. This can be accomplished through extensive mining of the crystal structures available in the PDB. It is possible that for some enzyme classes few or even no suitable crystal structures may exist. In these cases homology modeling would be used to fill the gaps. Kalyanaraman and Jacobson have shown in their very recent study of *E. coli* glycolysis pathways that homology models can be effectively used in conjunction with docking to separate native enzyme substrates from decoys.[50] While the ability to bind in the correct geometry is required for substrates to be catalytically reactive, it is not always a sufficient condition. To further qualify the potential of reactions

1644

dx.doi.org/10.1021/ci2000659 |*J. Chem. Inf. Model.* 2011, 51, 1634–1647

along novel pathways and eliminate false positives, such as the enolase catalysis of hydroxybutanal discussed in Section 3.3.1, it would be useful to obtain more detailed information about the barriers along the reaction path for specific enzyme–substrate complexes. QM/MM methods employed in this context have already been shown to provide both useful qualitative and quantitative results.[82] The pathway scoring can be further refined by incorporating information on known reactivity patterns. Finally, a quantitative measure of binding site specificity would be desirable as an estimate of the "engineerability" of a given protein. All of these studies are currently in progress and will be the subject of future publications.

ASSOCIATED CONTENT

ⓢ Supporting Information. Operators that were used in the generation of the reaction network are provided in Table S1. A summary of the reactions in the Kyoto Encyclopedia of Genes and Genomes (KEGG) that can be reproduced by the operators in Table S1 is provided in Table S2. The cofactor compounds included during network generation are summarized in Table S3, and Table S4 reports the concentrations of the different components in the synthetic media used for flux analysis. Finally, Tables S5 and S6 report the EFF scores for novel reactants catalyzed by enzymes with the presence of crystal waters for the 4.2.1 and 1.1.1 classes, respectively. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author
*E-mails: broadbelt@northwestern.edu (L.J.B.); goran.krilov@schrodinger.com (G.K.).

Author Contributions
[⊥]Both authors contributed equally to this work.

ACKNOWLEDGMENT

REFERENCES

(1) Himmel, M. E.; Ding, S.; Johnson, D. K.; Adney, W. S.; Nimlos, M. R.; Brady, J. W.; Foust, T. D. Biomass recalcitrance: Engineering plants and enzymes for biofuels production. *Science* **2007**, *315* (5813), 804–807.

(2) Durre, P. Biobutanol: An attractive biofuel. *Biotechnol. J.* **2007**, *2* (12), 1525–1534.

(3) Ezeji, T. C.; Qureshi, N.; Blaschek, H. P. Bioproduction of butanol from biomass: From genes to bioreactors. *Curr. Opin. Biotechnol.* **2007**, *18* (3), 220–227.

(4) Ezeji, T.; Qureshi, N.; Blaschek, H. P. Butanol production from agricultural residues: Impact of degradation products on *Clostridium beijerinckii* growth and butanol fermentation. *Biotechnol. Bioeng.* **2007**, *97* (6), 1460–1469.

(5) Ezeji, T.; Qureshi, N.; Blaschek, H. P. Production of acetone–butanol–ethanol (ABE) in a continuous flow bioreactor using degermed corn and *Clostridium beijerinckii*. *Process Biochem.* **2007**, *42* (1), 34–39.

(6) Ezeji, T. C.; Qureshi, N.; Blaschek, H. P. Production of acetone butanol (AB) from liquefied corn starch, a commercial substrate, using *Clostridium beijerinckii* coupled with product recovery by gas stripping. *J. Ind. Microbiol. Biotechnol.* **2007**, *34* (12), 771–777.

(7) Burgard, A. P.; Maranas, C. D. Review of the enzymes and metabolic pathways (EMP) database. *Metab. Eng.* **2001**, *3* (3), 193–194.

(8) Lee, J.; Yun, H.; Feist, A. M.; Palsson, B. O.; Lee, S. Y. Genome-scale reconstruction and in silico analysis of the *Clostridium acetobutylicum* ATCC 824 metabolic network. *Appl. Microbiol. Biotechnol.* **2008**, *80* (5), 849–862.

(9) Senger, R. S.; Papoutsakis, E. T. Genome-scale model for *Clostridium acetobutylicum*: Part I. Metabolic network resolution and analysis. *Biotechnol. Bioeng.* **2008**, *101* (5), 1036–1052.

(10) Connor, M. R.; Liao, J. C. Microbial production of advanced transportation fuels in non-natural hosts. *Curr. Opin. Biotechnol.* **2009**, *20* (3), 307–315.

(11) Connor, M. R.; Cann, A. F.; Liao, J. C. 3-Methyl-1-butanol production in *Escherichia coli*: Random mutagenesis and two-phase fermentation. *Appl. Microbiol. Biotechnol.* **2010**, *86* (4), 1155–1164.

(12) Atsumi, S.; Hanai, T.; Liao, J. C. Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels. *Nature* **2008**, *451* (7174), 86–U13.

(13) Atsumi, S.; Higashide, W.; Liao, J. C. Direct photosynthetic recycling of carbon dioxide to isobutyraldehyde. *Nat. Biotechnol.* **2009**, *27* (12), 1177–U1142.

(14) Finley, S. D.; Broadbelt, L. J.; Hatzimanikatis, V. Thermodynamic analysis of biodegradation pathways. *Biotechnol. Bioeng.* **2009**, *103* (3), 532–541.

(15) Henry, C. S.; Broadbelt, L. J.; Hatzimanikatis, V. Discovery and analysis of novel metabolic pathways for the biosynthesis of industrial chemicals: 3-Hydroxypropanoate. *Biotechnol. Bioeng.* **2010**, *106* (3), 462–473.

(16) Finley, S. D.; Broadbelt, L. J.; Hatzimanikatis, V. In silico feasibility of novel biodegradation pathways for 1,2,4-trichlorobenzene. *BMC Syst. Biol.* **2010**, *4*, Article 7.

(17) Li, C. H.; Henry, C. S.; Jankowski, M. D.; Ionita, J. A.; Hatzimanikatis, V.; Broadbelt, L. J. Computational discovery of biochemical routes to specialty chemicals. *Chem. Eng. Sci.* **2004**, *59* (22–23), 5051–5060.

(18) Kanehisa, M.; Goto, S.; Hattori, M.; Aoki-Kinoshita, K. F.; Itoh, M.; Kawashima, S.; Katayama, T.; Araki, M.; Hirakawa, M. From genomics to chemical genomics: New developments in KEGG. *Nucleic Acids Res.* **2006**, *34*, D354–D357.

(19) Schomburg, D.; Schomburg, I. Enzyme databases. *Methods Mol. Biol.* **2010**, *609*, 113–128.

(20) Pharkya, P.; Nikolaev, E. V.; Maranas, C. D. Review of the BRENDA database. *Metab. Eng.* **2003**, *5* (2), 71–73.

(21) Vaughn, M. G.; Howard, M. O. Combining medication and psychosocial treatments for addictions: The BRENDA approach. *J. Social Service Res.* **2002**, *29* (1), 85–87.

(22) Oh, M.; Yamada, T.; Hattori, M.; Goto, S.; Kanehisa, M. Systematic analysis of enzyme-catalyzed reaction patterns and prediction of microbial biodegradation pathways. *J. Chem. Inf. Model.* **2007**, *47* (4), 1702–1712.

(23) Klopman, G.; Dimayuga, M.; Talafous, J. META 1. A program for the evaluation of metabolic transformation of chemicals. *J. Chem. Inf. Comp. Sci.* **1994**, *34* (6), 1320–1325.

(24) Hou, B. K.; Ellis, L. B. M.; Wackett, L. P. Encoding microbial metabolic logic: Predicting biodegradation. *J. Ind. Microbiol. Biotechnol.* **2004**, *31* (6), 261–272.

(25) Hatzimanikatis, V.; Li, C. H.; Ionita, J. A.; Broadbelt, L. J. Computational framework for the discovery of novel biotransformations. In *Abstracts of Papers of the American Chemical Society*; 225th National Meeting of the American Chemical Society, New Orleans, LA, March 23–27, 2003; pp U201–U201.

(26) Hatzimanikatis, V.; Li, C. H.; Ionita, J. A.; Henry, C. S.; Jankowski, M. D.; Broadbelt, L. J. Exploring the diversity of complex metabolic networks. *Bioinformatics* **2005**, *21* (8), 1603–1609.

(27) Finley, S. D.; Broadbelt, L. J.; Hatzimanikatis, V. Computational framework for predictive biodegradation. *Biotechnol. Bioeng.* **2009**, *104* (6), 1086–1097.

(28) Ugi, I.; Bauer, J.; Brandt, J.; Friedrich, J.; Gasteiger, J.; Jochum, C.; Schubert, W. New applications of computers in chemistry. *Angew. Chem., Int. Ed.* **1979**, *18* (2), 111–123.

(29) Balaban, A. T. Applications of graph-theory in chemistry. *J. Chem. Inf. Comp. Sci.* **1985**, *25* (3), 334–343.

(30) Balaban, A. T. Graph-theory and theoretical chemistry. *THEOCHEM* **1985**, *21* (Feb.), 117–142.

(31) Broadbelt, L. J.; Stark, S. M.; Klein, M. T. Computer-generated pyrolysis modeling on-the-fly generation of species, reactions, and rates. *Ind. Eng. Chem. Res.* **1994**, *33* (4), 790–799.

(32) Broadbelt, L. J.; Stark, S. M.; Klein, M. T. Computer generated reaction networks: On-the-fly calculation of species properties using computational quantum chemistry. *Chem. Eng. Sci.* **1994**, *49* (24B), 4991–5010.

(33) Gonzalez-Lergier, J.; Broadbelt, L. J.; Hatzimanikatis, V. Theoretical considerations and computational analysis of the complexity in polyketide synthesis pathways. *J. Am. Chem. Soc.* **2005**, *127* (27), 9930–9938.

(34) Klein, M. T.; Neurock, M.; Broadbelt, L.; Foley, H. C. Reaction Pathway Analysis: Global Molecular and Mechanistic Perspectives; In *Selectivity in Catalysis*; Davis, M. E., Suib, S. L., Eds.; ACS Symposium Series 517; 1993; pp 290–315.

(35) Broadbelt, L. J.; Stark, S. M.; Klein, M. T. Computer generated reaction modelling: Decomposition and encoding algorithms for determining species uniqueness. *Comput. Chem. Eng.* **1996**, *20* (2), 113–129.

(36) Klein, M. T.; Hou, G.; Quann, R. J.; Wei, W.; Liao, K. H.; Yang, R. S. H.; Campain, J. A.; Mazurek, M. A.; Broadbelt, L. J. BioMOL: A computer-assisted biological modeling tool for complex chemical mixtures and biological processes at the molecular level. *Environ. Health Perspect.* **2002**, *110*, 1025–1029.

(37) Tipton, K.; Boyce, S. History of the enzyme nomenclature system. *Bioinformatics* **2000**, *16* (1), 34–40.

(38) Broadbelt, L. J.; Pfaendtner, J. Lexicography of kinetic modeling of complex reaction networks. *AIChE J.* **2005**, *51* (8), 2112–2121.

(39) Henry, C. S.; Jankowski, M. D.; Broadbelt, L. J.; Hatzimanikatis, V. Genome-scale thermodynamic analysis of *Escherichia coli* metabolism. *Biophys. J.* **2006**, *90* (4), 1453–1461.

(40) Li, C. H.; Ionita, J. A.; Henry, C. S.; Jankowski, M.; Broadbelt, L. J. Exploring the diversity of metabolism. In *Abstracts of Papers of the American Chemical Society*; 229th National Meeting of the American Chemical Society, San Diego, CA, March 13, 2005; p U1174.

(41) Henry, C. S.; Jankowski, M.; Broadbelt, L. J.; Hatzimanikatis, V. Analysis of free energy change and thermodynamic feasibility in a genome scale metabolic model. In *Abstracts of Papers of the American Chemical Society*. 229th National Meeting of the American Chemical Society, San Diego, CA, March 13, 2005; p U193.

(42) Henry, C. S.; Broadbelt, L. J.; Hatzimanikatis, V. Thermodynamics-based metabolic flux analysis. *Biophys. J.* **2007**, *92* (5), 1792–1805.

(43) Assary, R. S.; Broadbelt, L. J. Computational screening of novel thiamine-catalyzed decarboxylation reactions of 2-keto acids. *Bioprocess Biosyst. Eng.* **2011**, *34* (3), 375–388.

(44) Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J. Docking and scoring in virtual screening for drug discovery: Methods and applications. *Nat. Rev. Drug Discovery* **2004**, *3* (11), 935–949.

(45) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* **2004**, *47* (7), 1739–1749.

(46) Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L. Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J. Med. Chem.* **2004**, *47* (7), 1750–1759.

(47) Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren, T. A.; Sanschagrin, P. C.; Mainz, D. T. Extra precision glide: Docking and scoring incorporating a model of hydrophobic enclosure for protein−ligand complexes. *J. Med. Chem.* **2006**, *49* (21), 6177–6196.

(48) Hermann, J. C.; Ghanem, E.; Li, Y. C.; Raushel, F. M.; Irwin, J. J.; Shoichet, B. K. Predicting substrates by docking high-energy intermediates to enzyme structures. *J. Am. Chem. Soc.* **2006**, *128* (49), 15882–15891.

(49) Favia, A. D.; Nobeli, I.; Glaser, F.; Thornton, J. M. Molecular docking for substrate identification: The short-chain dehydrogenases/reductases. *J. Mol. Biol.* **2008**, *375* (3), 855–874.

(50) Kalyanaraman, C.; Jacobson, M. P. Studying enzyme−substrate specificity in silico: A case study of the *Escherichia coli* glycolysis pathway. *Biochem.* **2010**, *49* (19), 4003–4005.

(51) Tyagi, S.; Pleiss, J. Biochemical profiling in silico: Predicting substrate specificities of large enzyme families. *J. Biotechnol.* **2006**, *124* (1), 108–116.

(52) Kalyanaraman, C.; Bernacki, K.; Jacobson, M. P. Virtual screening against highly charged active sites: Identifying substrates of alpha−beta barrel enzymes. *Biochem.* **2005**, *44* (6), 2059–2071.

(53) Song, L.; Kalyanaraman, C.; Fedorov, A. A.; Fedorov, E. V.; Glasner, M. E.; Brown, S.; Imker, H. J.; Babbitt, P. C.; Almo, S. C.; Jacobson, M. P.; Gerlt, J. A. Prediction and assignment of function for a divergent N-succinyl amino acid racemase. *Nat. Chem. Biol.* **2007**, *3* (8), 486–491.

(54) Kalyanaraman, C.; Imker, H. J.; Fedorov, A. A.; Fedorov, E. V.; Glasner, M. E.; Babbitt, P. C.; Almo, S. C.; Gerlt, J. A.; Jacobson, M. P. Discovery of a dipeptide epimerase enzymatic function guided by homology modeling and virtual screening. *Struct.* **2008**, *16* (11), 1668–1677.

(55) Rakus, J. F.; Kalyanaraman, C.; Fedorov, A. A.; Fedorov, E. V.; Mills-Groninger, F. P.; Toro, R.; Bonanno, J.; Bain, K.; Sauder, J. M.; Burley, S. K.; Almo, S. C.; Jacobson, M. P.; Gerlt, J. A. Computation-facilitated assignment of the function in the enolase superfamily: a regiochemically distinct galactarate dehydratase from oceanobacillus iheyensis. *Biochem.* **2009**, *48* (48), 11546–11558.

(56) Hermann, J. C.; Marti-Arbona, R.; Fedorov, A. A.; Fedorov, E.; Almo, S. C.; Shoichet, B. K.; Raushel, F. M. Structure-based activity prediction for an enzyme of unknown function. *Nature* **2007**, *448* (7155), 775–U772.

(57) Xiang, D. F.; Kolb, P.; Fedorov, A. A.; Meier, M. M.; Fedorov, L. V.; Nguyen, T. T.; Sterner, R.; Almo, S. C.; Shoichet, B. K.; Raushel, F. M. Functional annotation and three-dimensional structure of dr0930 from deinococcus radiodurans, a close relative of phosphotriesterase in the amidohydrolase superfamily. *Biochemistry* **2009**, *48* (10), 2237–2247.

(58) Hatzimanikatis, V.; Li, C. H.; Ionita, J. A.; Broadbelt, L. J. Metabolic networks: Enzyme function and metabolite structure. *Curr. Opin. Struct. Biol.* **2004**, *14* (3), 300–306.

(59) Gonzalez-Lergier, J.; Broadbelt, L. J.; Hatzimanikatis, V. Analysis of the maximum theoretical yield for the synthesis of erythromycin precursors in *Escherichia coli. Biotechnol. Bioeng.* **2006**, *95* (4), 638–644.

(60) Yeturu, K.; Chandra, N. PocketMatch: A new algorithm to compare binding sites in protein structures. *BMC Bioinf.* **2008**, *9*, Article 543.

(61) *Maestro*, version 9.1; Schrödinger: New York, 2010.

(62) *Prime*, version 2.2; Schrödinger: New York, 2010.

(63) *LigPrep*, version 2.4; Schrödinger: New York, 2010.

(64) *MacroModel*, version 9.8; Schrödinger: New York, 2010.

(65) *Glide*, version 5.6; Schrödinger: New York, 2010.

(66) Mavrovouniotis, M. L. Group contributions for estimating standard Gibbs energies of formation of biochemical-compounds in aqueous-solution. *Biotechnol. Bioeng.* **1990**, *36* (10), 1070–1082.

(67) Benson, S. W. New methods for estimating the heats of formation, heat capacities, and entropies of liquids and gases. *J. Phys. Chem. A* **1999**, *103* (51), 11481–11485.

(68) Benson, S. W. Current status of group additivity. In *Abstracts of Papers of the American Chemical Society*; 212th National Meeting of the American Chemical Society, Orlando, FL, 1996; pp 56–COMP.

(69) Mavrovouniotis, M. L. Estimation of standard Gibbs energy changes of biotransformations. *J. Biol. Chem.* **1991**, *266* (22), 14440–14445.

(70) Tran, L. M.; Rizk, M. L.; Liao, J. C. Ensemble modeling of metabolic networks. *Biophys. J.* **2008**, *95* (12), 5606–5617.

(71) Feist, A. M.; Henry, C. S.; Reed, J. L.; Krummenacker, M.; Joyce, A. R.; Karp, P. D.; Broadbelt, L. J.; Hatzimanikatis, V.; Palsson, B. O. A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Mol. Syst. Biol.* **2007**, *3*, Article 121.

(72) Monot, F.; Martin, J. R.; Petitdemange, H.; Gay, R. Acetone and butanol production by *Clostridium acetobutylicum* in a synthetic medium. *Appl. Environ. Microbiol.* **1982**, *44* (6), 1318–1324.

(73) Monot, F.; Engasser, J. M.; Petitdemange, H. Influence of pH and undissociated butyric-acid on the production of acetone and butanol in batch cultures of *Clostridium acetobutylicum. Appl. Microbiol. Biotechnol.* **1984**, *19* (6), 422–426.

(74) Macchiarulo, A.; Nobeli, I.; Thornton, J. M. Ligand selectivity and competition between enzymes in silico. *Nat. Biotechnol.* **2004**, *22* (8), 1039–1045.

(75) Kursula, P.; Ojala, J.; Lambeir, A. M.; Wierenga, R. K. The catalytic cycle of biosynthetic thiolase: A conformational journey of an acetyl group through four binding modes and two oxyanion holes. *Biochemistry* **2002**, *41* (52), 15543–15556.

(76) Masamune, S.; Walsh, C. T.; Sinskey, A. J.; Peoples, O. P. Poly-(R)-3-hydroxybutyrate (Phb) biosynthesis: Mechanistic studies on the biological Claisen condensation catalyzed by beta-ketoacyl thiolase. *Pure Appl. Chem.* **1989**, *61* (3), 303–312.

(77) Li, C.; Heatwole, J.; Soelaiman, S.; Shoham, M. Crystal structure of a thermophilic alcohol dehydrogenase substrate complex suggests determinants of substrate specificity and thermostability. *Proteins Struct. Funct. Bioinf.* **1999**, *37* (4), 619–627.

(78) Markovic, O.; Theorell, H.; Rao, S. Rat liver alcohol dehydrogenase: Purification and properties. *Acta Chem. Scand.* **1971**, *25* (1), 195–205.

(79) Bell, A. F.; Feng, Y.; Hofstein, H. A.; Parikh, S.; Wu, J.; Rudolph, M. J.; Kisker, C.; Whitty, A.; Tonge, P. J. Stereoselectivity of enoyl-CoA hydratase results from preferential activation of one of two bound substrate conformers. *Chem Biol.* **2002**, *9* (11), 1247–1255.

(80) Waterson, R. M.; Hill, R. L. Enoyl coenzyme A hydratase (crotonase): Catalytic properties of crotonase and its possible regulatory role in fatty-acid oxidation. *J. Biol. Chem.* **1972**, *247* (16), 5258–5265.

(81) Gerlt, J. A.; Babbitt, P. C.; Rayment, I. Divergent evolution in the enolase superfamily: The interplay of mechanism and specificity. *Arch. Biochem. Biophys.* **2005**, *433* (1), 59–70.

(82) Friesner, R. A.; Guallar, V. Ab initio quantum chemical and mixed quantum mechanics/molecular mechanics (QM/MM) methods for studying enzymatic catalysis. *Annu. Rev. Phys. Chem.* **2005**, *56*, 389–427.

1647

dx.doi.org/10.1021/ci2000659 |*J. Chem. Inf. Model.* 2011, 51, 1634–1647