Computational Study of Enantioselective Carboligation Catalyzed by Benzoylformate Decarboxylase

Ferran Planas[a], Michael J McLeish[b] and Fahmi Himo[a*]*

[a] Department of Organic Chemistry, Arrhenius Laboratory, Stockholm University, SE-10691, Stockholm, Sweden.

[b] Department of Chemistry and Chemical Biology, Indiana University-Purdue University Indianapolis, IN 46202, Indianapolis, USA.

E-mail: fahmi.himo@su.se; mcleish@iupui.edu

Abstract

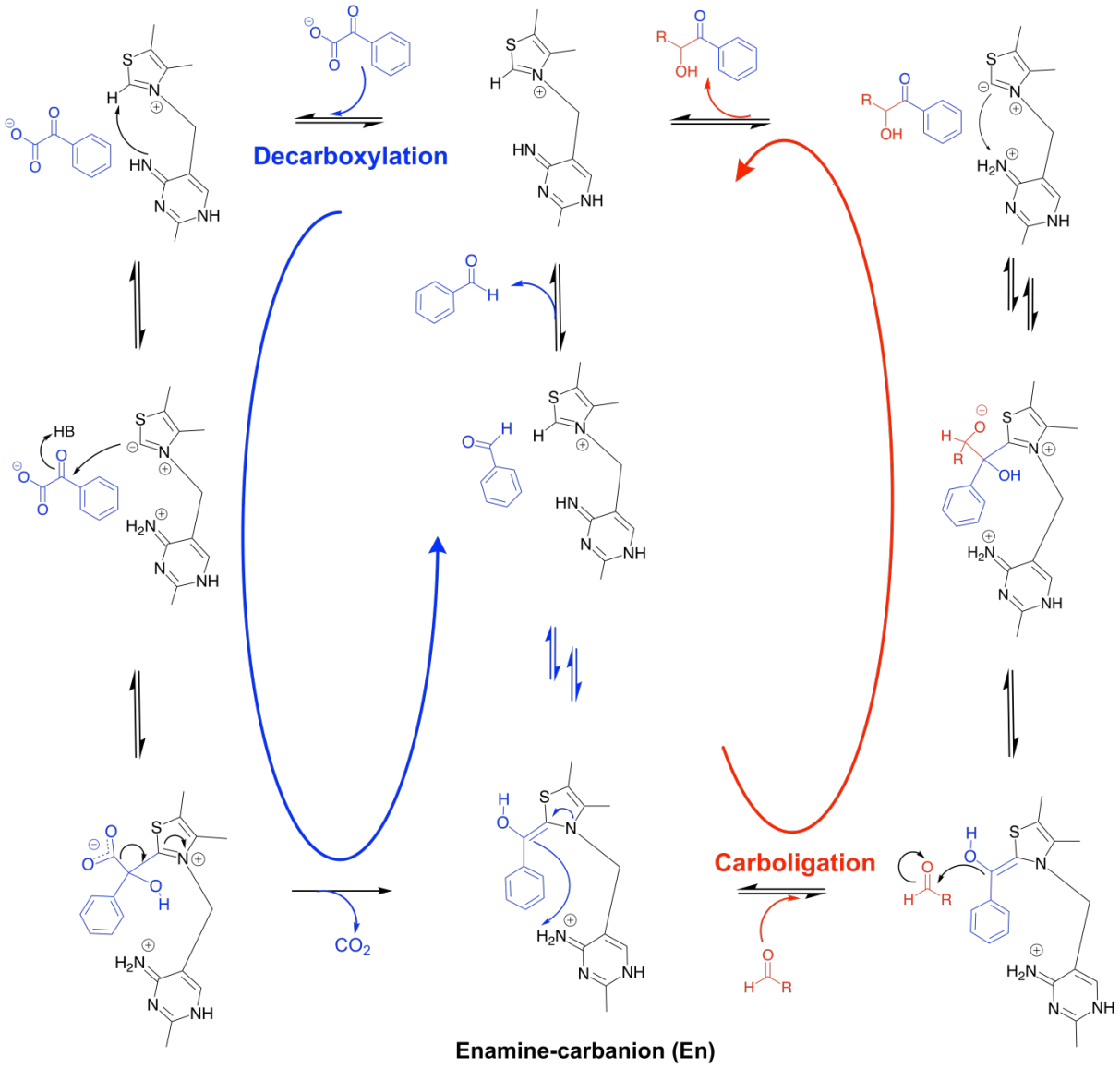
Benzoylformate decarboxylase (BFDC) is a thiamin-diphosphate enzyme that catalyzes the decarboxylation of benzoylformate to yield benzaldehyde and carbon dioxide. In addition to its natural reaction, BFDC is able to catalyze carboligation reactions in a highly enantioselective fashion, making the enzyme a potentially important biocatalyst. Here, we use density functional theory calculations to investigate the detailed mechanism of BFDC-catalyzed carboligation and to elucidate the sources of the enantioselectivity. Benzaldehyde and acetaldehyde are studied as acceptors, for, when reacting with a benzaldehyde donor, they yield products with opposite enantiospecificity. For each of the acceptors, several possible binding modes to the active site are initially examined before the individual reaction paths leading to the two enantiomeric products are followed. The calculated energies are in good agreement with the experimental results, and analysis of the transition states gives insight into the origins of the enantioselectivity.

Keywords: Biocatalysis, Enantioselectivity, Density functional theory, Carboligation, Benzoylformate decarboxylase, Reaction mechanism, Enzymology, Asymmetric synthesis.

I. Introduction

Benzoylformate decarboxylase (BFDC) is a thiamine diphosphate (ThDP)-dependent lyase that catalyzes the non-oxidative decarboxylation of benzoylformate into benzaldehyde and carbon dioxide as part of the mandelic acid degradation pathway.[1,2] The reaction mechanism of this enzyme has attracted attention, initially because BFDC showed some notable differences in the active site compared to other ThDP-dependent decarboxylases. Pyruvate decarboxylase (PDC), the archetypal member of the group, has two contiguous histidine residues, sometimes referred to as the HH motif,[3] as well as two ionizable acidic residues all of which have been shown to be catalytically important.[4-6] These residues are conserved in most other group members.[7-9] However, they are not present in BFDC, which lacks the two ionizable residues and has the two histidines differently placed both in sequence and space.[10] While plausible roles for the BFDC active site residues were initially reported,[11-12] subsequent studies brought those roles into question.[13-14]

Very recently, we reported a detailed computational study of the mechanism of the BFDC reaction (Scheme 1) characterizing the intermediates and transition states of the catalytic cycle in terms of geometries and energies.[15] The roles of the different active site residues were described and, very importantly, we demonstrated the existence of an off-cycle, yet kinetically relevant, tricyclic form of the ThDP cofactor not considered previously for this enzyme.[15] The decarboxylation step was calculated to be irreversible, and the resultant enamine-carbanion form of the cofactor, was found to be the resting state of the reaction. The following steps, i.e. enamine protonation, cofactor reprotonation and formation of benzaldehyde in the active site, were all found to be reversible.[15] Subsequently, we identified possible ThDP states and calculated their relative stabilities with different ligands bound to the enzyme.[16] The calculations showed that the nature of the bound ligand can have a dramatic effect on the energies of the cofactor states, and the aforementioned tricyclic state was demonstrated to be lowest in energy in the presence of the substrate.



Scheme 1. *Partitioning between the decarboxylation and carboligation in the mechanism of BFDC.*

The reversibility of the second part of the reaction has implications for a very important feature of BFDC, namely its ability to catalyze carboligation reactions. These can occur when the enamine intermediate reacts with an aldehyde acceptor molecule, as shown in Scheme 1. The ability of BFDC to catalyze a carboligation reaction was initially reported after 2-hydroxypropiophenone (2-HPP) was obtained as a side product of benzoylformate decarboxylation with acetaldehyde as cosubstrate.[17] The fact that 2-HPP was obtained in

parallel with benzaldehyde formation indicated that the post-decarboxylation intermediate, the enamine-carbanion, could either be protonated, ultimately producing benzaldehyde (i.e., the natural reaction), or intercepted by the acetaldehyde to yield 2-HPP (carboligation). It was further observed that, at high concentrations of benzaldehyde, 2-HPP was also formed. This demonstrated that the enamine intermediate could be obtained from either benzoylformate or from benzaldehyde (see Scheme 1) and provided irrefutable evidence that all post-decarboxylation steps must be reversible.[17] Further, since acetaldehyde is a prochiral molecule, a stereogenic center will be formed in this reaction. It is now well established that *S*-HPP is formed, with ~90% enantiomeric excess (*ee*).[17-19] Intriguingly, when carried out with benzaldehyde lyase (BAL), another ThDP-dependent enzyme, the same reaction provided *R*-HPP in better that 99% *ee*.[20,21] Conversely, if benzaldehyde is used as the acceptor, both BFDC and BAL produce solely the *R*-enantiomer of benzoin.[18,22-24]

The source of the enantioselectivity of the BFDC-catalyzed carboligations has been addressed by both experimental mutagenesis studies[19,25,26] and molecular modeling.[21] Site-directed mutagenesis showed that Ala460 and Phe464 were relevant for the enantioselectivity, as simultaneous mutation of both these positions to isoleucine resulted in a change of enantioselectivity for the carboligation producing *R*-HPP in ~50% *ee*.[19]

In the molecular modeling study, selected conformations of the acceptor were built into the active site, which enabled the identification of residues likely to control the stereochemical outcome.[21] Two pockets in the BFDC active site were identified, one large and highly solvent-exposed cavity lying on top of the phenyl group of the donor, and a much smaller cavity lying on top of the thiazole ring (see Figure 1). It was noted that the residues forming the two cavities were consistent with those shown to affect enantioselectivity in the mutagenesis studies, and suggested that differential binding of acceptor molecules in the two pockets was crucial for the stereoselectivity of the carboligation. The modeling study also revealed Leu461 as a potentially important position, something that was later confirmed experimentally.[27]

Here, we use accurate quantum chemical methods to investigate the mechanism and origins of the enantioselectivity of the carboligation reaction of BFDC. It is important to appreciate that the modeling of enantioselective reactions requires high accuracy in relative transition state energies in order to reproduce and rationalize observed selectivities. In general, this is a daunting task for any kind of computational approach. That said, the adopted quantum

chemical cluster approach has been demonstrated to be a valuable tool in this field,[28-30] and has been successfully applied to explore the reaction mechanisms and enantioselectivities of several enzymes, including epoxide hydrolases,[31,32] arylmalonate decarboxylase,[33] phenolic acid decarboxylase,[34,35] and secondary alcohol dehydrogenase.[36]

As with our earlier studies on the BFDC decarboxylation reaction[15] and the energetics of the cofactor states,[16] we employ a large model of the active site designed on the basis of the X-ray structure. Using benzoylformate as the donor, we considered two acceptor substrates, benzaldehyde (**BA**) and acetaldehyde (**AA**) which yield benzoin and 2-HPP, respectively. These two substrates provide products with opposite enantioselectivities,[18] and we address the reasons for this.

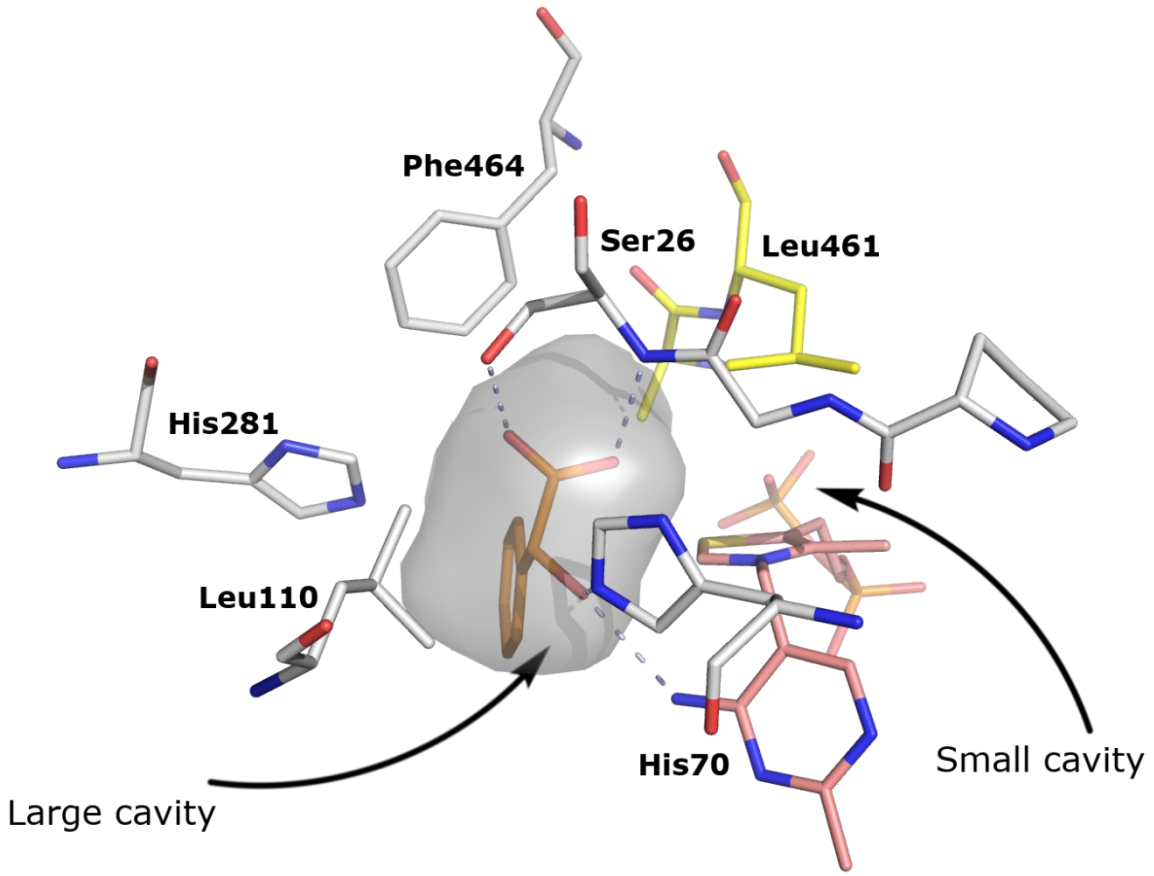


Figure 1. *X-Ray structure of BFDC active site with a bound substrate analog, R-mandelate (PDB 1MCZ). The large and small cavities are indicated.*

II. Computational Methods

II.A. Quantum chemical technical details

All the calculations were performed using the Gaussian 09 package[37] and the B3LYP-D3(BJ) method.[38-41] The 6-31G(d,p) basis set was used for the geometry optimizations and the electronic energies of the stationary points were refined by single-point calculations with the 6-311+G(2d,2p) basis set. Solvation energies were calculated with the SMD implicit solvent method[42] and a dielectric constant of ε=4. The zero-point energy corrections were done at the same level of theory as the geometry optimizations.

II.B. Active site model

A model of the BFDC active site was built on the basis of the crystal structure (PDB 1MCZ) and, as with the two previous studies,[15,16] it comprises the catalytically important Ser26, Glu28, Glu47, His70 and His281, along with the residues that define the active site cavities, i.e., Asn26, Pro24, Asn27, Ala44, Leu45, Asn77, Leu110, Gly401, Gly402, Leu403, Tyr433, Tyr458, Ala460, and Phe464. The residues were truncated as shown in Figure 2. In order to mimic the restraints that the protein surrounding imposes on the active site residues, the truncated atoms were kept fixed at their crystallographic positions. This approximation that has been employed in numerous studies and has been shown to yield good results.[28-36] In this instance it is particularly appropriate as it is well-established that the active site of BFDC is remarkably pre-organized.[12,14] The acceptor substrates were manually placed in the cavity assuming different binding modes and orientation (see below). The total number of atoms in the model was 318 with benzaldehyde as acceptor, and 311 for acetaldehyde as acceptor, and the overall charge was 0 in both cases.

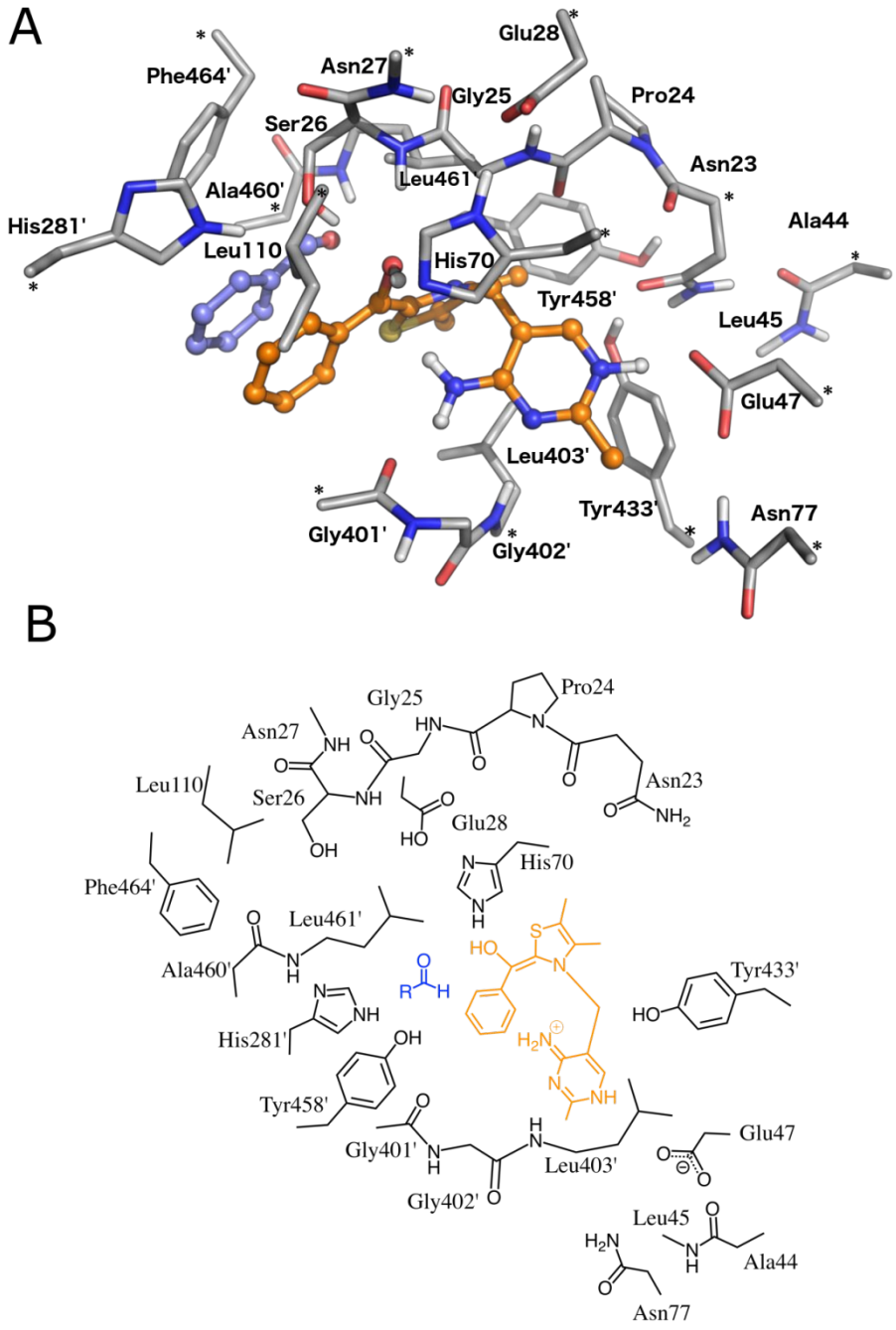


Figure 2. *A) Optimized structure of the active site model with benzaldehyde bound. Asterisks mark the atoms fixed in the geometry optimizations. For clarity, the non-polar hydrogens of the residues are omitted. B) Schematic representation of the active site model. In both A and B the enamine of ThDP derived from reaction with either benzoylformate or benzaldehyde is shown in orange. The acceptor molecule is shown in blue (R = Ph, Me).*

III. Results and Discussion

The starting point for the study of the carboligation reaction was the enamine-carbanion derived from the decarboxylation of benzoylformate, here abbreviated **En** (Scheme 1). The steps leading up to this intermediate were established in detail in the previous computational study on the BFDC reaction.[15] Further, as it has been established that all the post-decarboxylation steps are reversible,[15,17] **En** may also be considered to be the intermediate obtained in a carboligation reaction when benzaldehyde is employed as the donor substrate. Two acceptor substrates were considered in this study, benzaldehyde (**BA**) and acetaldehyde (**AA**), which lead to the benzoin and 2-HPP products, respectively. After the potential active site binding modes and orientations were examined, the lowest-energy binding mode for each acceptor was set to zero. Subsequently, the individual reactions were studied by calculating the transition states and intermediates leading to the two enantiomeric outcomes that were possible for each acceptor. We start by discussing the results relating to benzoin formation, as this reaction has been experimentally observed to take place with effectively complete enantiospecificity.[18]

III.A. Reaction of benzaldehyde

As an acceptor, benzaldehyde can bind to the active site with either its *pro*-S or *pro*-R face toward the enamine donor. A variety of orientations of the phenyl ring are also possible, therefore it was necessary to determine the relative energies of a number of binding modes. Overall, the results can be described in three modes depending on the orientations of the carbonyl and phenyl moieties (Figure 3). In the lowest-energy binding mode, called **En(BA)-R1**, the benzaldehyde is bound with its *pro*-R face toward the cofactor, with the carbonyl group on top of the hydroxyl group of the donor and forming a hydrogen bond with the side chain of Ser26. The phenyl group of the acceptor is in the large pocket and located on top of the phenyl ring of the donor.

The second binding mode, **En(BA)-S1**, has a calculated energy of +1.1 kcal/mol relative to **En(BA)-R1**. In this pose, the phenyl ring remains in the large pocket but the acceptor now faces the donor with its *pro*-S face. Overall, this mode resembles that in **En(BA)-R1**, but the

carbonyl lies in opposite direction to the hydroxyl of Ser26, instead forming a hydrogen bond with the backbone amide of Leu461.

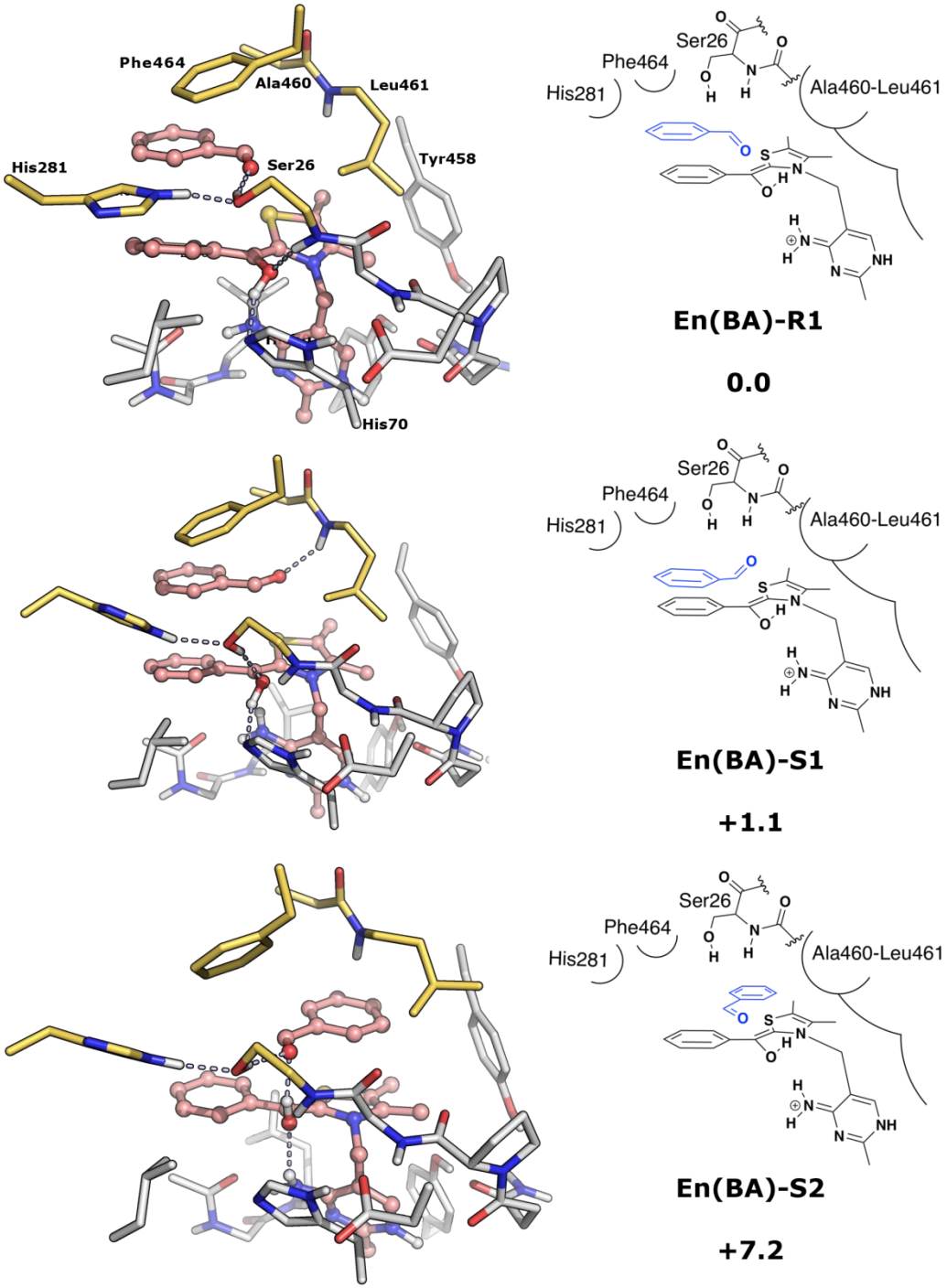


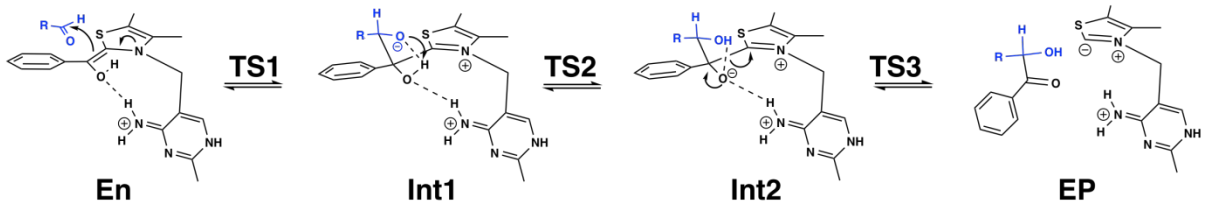
Figure 3. *Optimized geometries and schematic drawings of different binding modes for benzaldehyde. Relative energies are given in kcal/mol. For clarity, only a part of the active site model is shown here.*

The third binding mode, **En(BA)-S2**, sees also a *pro*-S binding of benzaldehyde. However, the phenyl substituent of the acceptor now points into the small binding pocket and the carbonyl oxygen forms hydrogen bonds to both the hydroxyl of Ser26 and the hydroxyl of the cofactor (Figure 3). In order for the phenyl substituent to fit into the small space on top of the thiazole ring, the side chain of Leu461 rotates away. Concomitantly, the thiazole ring is pushed towards the pyrimidine ring and the V-shape of the cofactor becomes somewhat distorted (see SI for a superposition of the structures of the three binding modes). In this binding mode, the donor and acceptor are much closer to each other than the other two binding modes, and due to the steric clashes, the energy of **En(BA)-S2** is 7.2 kcal/mol higher than **En(BA)-R1**.

It should be mentioned that a fourth class of binding modes was also examined in which the acceptor faces the donor with its *pro*-R face and with the phenyl substituent pointing into the small pocket. However, all attempts to optimize such geometry resulted in the benzaldehyde leaving the active site entirely.

At this point it is important to remember that we here only evaluate the *relative* energies of the various binding modes. That is, the absolute binding free energies are not calculated, as these cannot be accurately determined by the current approach.

The carboligation mechanism obtained from the calculations is shown in Scheme 2. Upon binding of the **BA** acceptor, the first step is a C-C bond formation that leads to an alkoxide tetrahedral intermediate, **Int1**, which has a newly formed chiral center. Next, an intramolecular proton transfer takes place from the hydroxyl group of the donor moiety to the alkoxide, resulting in **Int2**. Finally, the last step is a C-C bond cleavage to recover the ylid form of the cofactor and to form the enzyme-product complex (**EP**).



Scheme 2. *General mechanism for the carboligation reaction starting at the enamine.*

As seen from the calculated potential energy profiles, the reactions leading to the *R*- and *S*-products display very large energy differences (Figure 4), starting as early as initial C-C bond formation. While the barrier for this step for the *R*-pathway is calculated to be 10.7 kcal/mol (**TS1(BA)-R**), the barrier for the *S*-pathway is as high as 25.1 kcal/mol (**TS1(BA)-S**). The energies of the resulting intermediates also show very large difference and are calculated to be +9.4 kcal/mol for **Int1(BA)-R** and +22.9 kcal/mol for **Int1(BA)-S**. Intriguingly, the pathway leading to the *S*-product does not start from the lowest-energy *pro*-S binding mode **En(BA)-S1**, rather from the higher energy **En(BA)-S2**. Attempts to form the C-C bond starting from **En(BA)-S1** always resulted in the formation of the enamine. Presumably, reaction following this binding mode lacks the hydrogen bonds that can stabilize the charge of the alkoxide developed at the acceptor upon C-C formation.

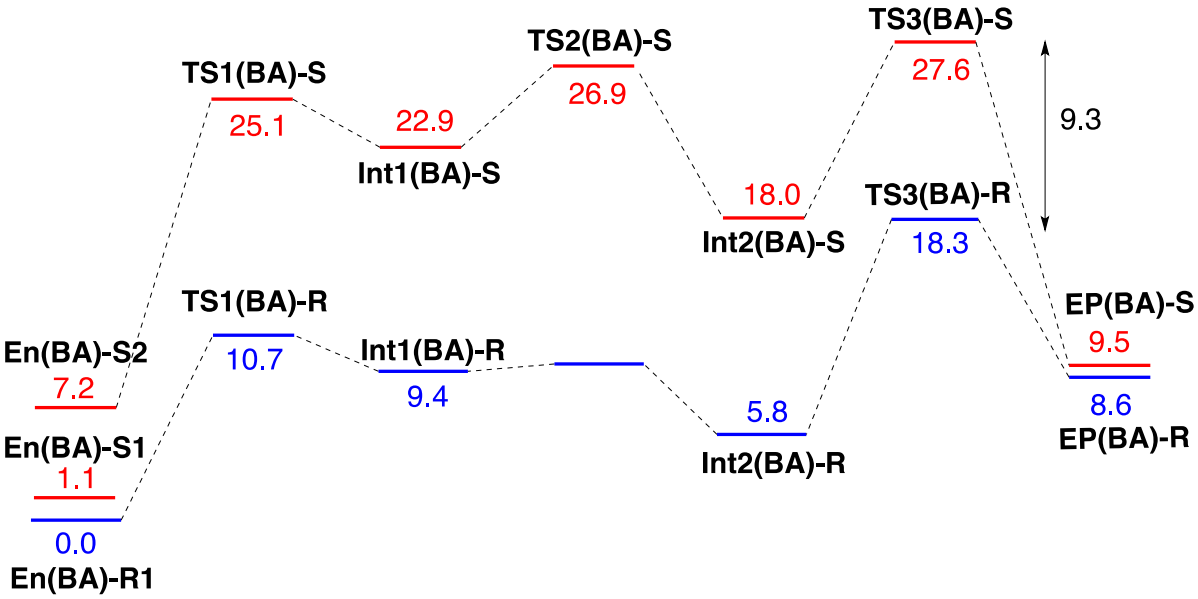


Figure 4. *Calculated potential energy profiles for the carboligation reaction of benzaldehyde. Energies are given in kcal/mol.*

The calculated energy difference of 7.2 kcal/mol in the binding (Figure 4) has thus increased to 14.4 kcal/mol at the TS of the C-C bond formation step. Comparison of the optimized geometries of the **TS1(BA)-R** and **TS1(BA)-S** (Figure 5) reveals some differences that explain this increase. In **TS1(BA)-R**, the charge developing at the alkoxide oxygen is stabilized by an

oxyanion hole consisting of three hydrogen bonds (to the hydroxyl and backbone NH of Ser26, and to the hydroxyl group of the donor), while in **TS1(BA)-S** only the first two of these are possible. We note also that, in **TS1(BA)-S**, the phenyl ring of the acceptor has moved from the small pocket toward the large pocket. Due to the formation of a tetrahedral carbon center at the acceptor it now experiences steric clashes with the Phe464, Ala460 and Leu461 residues.

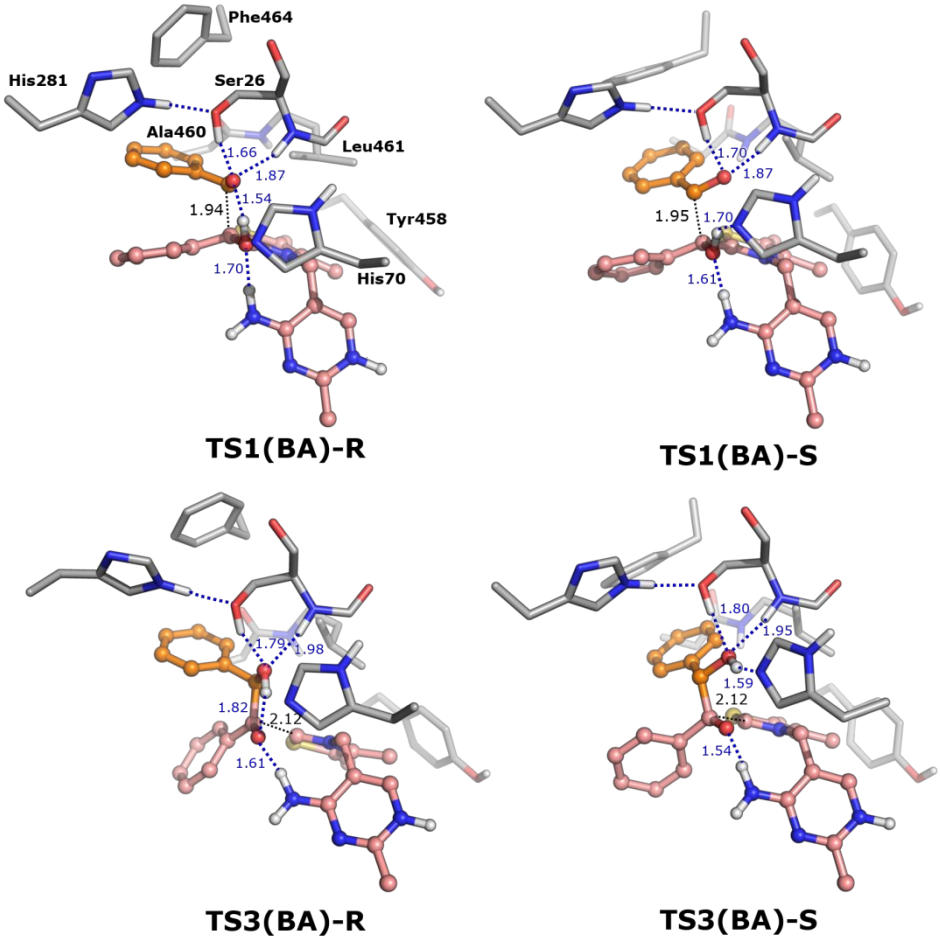


Figure 5. *Optimized geometries of selected transition states for the reaction with benzaldehyde as acceptor. For clarity, only a part of the active site model is shown here.*

The subsequent proton transfer step is predicted to have a barrier of 4 kcal/mol for the *S*-pathway (**TS2(BA)-S**, Figure 4). Conversely, for the *R*-pathway proton transfer appears to be barrierless or, at most, occur with a very low barrier. A transition state for the *R*-pathway could be optimized at the level of the theory of the geometry optimization. However, when

corrections such as the large basis set, solvation and zero-point energy corrections were added to the energy the barrier disappeared (i.e., became lower in energy than the intermediate). Consequently, the formation of **Int2(BA)-R** from **Int1(BA)-R** can be considered to be effectively barrierless.

The calculations suggest that the last step of the mechanism, the C-C bond cleavage to form benzoin and recover the ylid, will have the highest barrier of all studied steps. This is true for both the *R*- and *S*-pathways, with overall barriers of 18.3 and 27.6 kcal/mol, respectively, relative to **En(BA)-R1**. As with the initial C-C bond formation step, the large energy difference arises from the clashes of the phenyl ring of the acceptor with the Phe464, Ala460 and Leu461 groups in **TS3(BA)-S**. Conversely, as shown in Figure 5, in **TS3(BA)-R** the phenyl ring points to the more open side of the large pocket.

Interestingly, although TS3 for the two pathways differ considerably in energy, the enzyme-product complexes, **EP(BA)-R** and **EP(BA)-S**, are calculated to have quite similar energies (+8.6 and +9.8 kcal/mol, respectively). Most likely, this is because the strain and steric clashes present in **TS3(BA)-S** are alleviated once the C-C bond is cleaved and the product is detached from the cofactor.

It is important to remember that the present study focuses on the steps between the enamine-carbanion intermediate **En** (to which the donor molecule is bound) and the enzyme-product complex **EP** (in which the cofactor is in the ylid form, Scheme 2), as it is assumed that the selectivity is controlled in this part of the reaction. To continue the catalytic cycle, the benzoin product has to be released, a new donor molecule has to bind, the enamine intermediate has to be regenerated, and a new acceptor molecule has to bind (Scheme 1). While the mechanism for the enamine formation has been described in detail in our previous work on the natural reaction mechanism,[15] the non-chemical steps involving substrate binding and product release are not considered explicitly as their modeling is associated with large errors.

Since TS3 constitutes the highest energy transition state in both pathways, the energy difference between **TS3(BA)-R** and **TS3(BA)-S** will determine the stereochemical outcome. The large calculated difference of 9.3 kcal/mol is consistent with the full stereocontrol observed experimentally for this substrate.[18] It is likely, however, that the difference is somewhat overestimated in the calculations due to the model likely being somewhat too rigid to properly

adapt to the position of the large phenyl substituent, therefore causing more clashes than expected.

III.B. Reaction of acetaldehyde

With acetaldehyde as acceptor, a binding mode analysis was conducted in a similar manner to that employed for benzaldehyde. Four individual binding modes could be identified. In contrast to benzaldehyde, in the lowest-energy binding mode the substrate faces the enamine by its *pro*-S face (**En(AA)-S1**, Figure 6), with the carbonyl group forming a hydrogen bond with the side chain of Ser26. The small methyl substituent points into the much larger active site pocket on top of the phenyl ring of the donor, and the distance between the two carbons that will form the C-C bond is longer than in the case of the benzaldehyde acceptor.

The second binding mode, **En(AA)-R1**, has the *pro*-R face toward the enamine, with the methyl group lying above the aromatic ring of the donor. In that position the carbonyl group forms two hydrogen bonds, with Ser26 and with the hydroxyl group of the enamine (Figure 6). As the acceptor is now closer to the donor, this binding mode causes more steric crowding and the calculated energy is therefore higher than **En(AA)-S1**, by 4.7 kcal/mol.

In the third binding mode, **En(AA)-S2**, the acceptor lies above the enamine and the carbonyl forms two hydrogen bonds with the hydroxyl and backbone amide groups of Ser26. These interactions position the methyl group in the small cavity and close to Leu461, again resulting in steric clashes. Predictably, the calculated energy for this binding mode is also high, +7.2 kcal/mol.

The fourth binding mode, **En(AA)-R2**, has the carbonyl group of the acceptor facing Ala460 and the methyl group oriented toward Ser26. This prevents the formation of any stabilizing hydrogen bond to the carbonyl group and has been calculated to be as much as 14.1 kcal/mol higher in energy than **En(AA)-S1**. A superposition of the structures of the four binding modes is given in the SI.

Starting from these binding modes, the reactions with acetaldehyde as acceptor were followed. The lowest-energy pathways leading to each of the enantiomers of the 2-HPP product are displayed in Figure 7. Although the calculations show that the reactions follow the same mechanism as for benzaldehyde (Scheme 2), the potential energy profiles are significantly different (Figure 4). Chemically, acetaldehyde is a better electrophile than benzaldehyde, and

therefore will react faster. Accordingly, the energy barriers are expected to be lower. However, in this instance, an additional factor is at play. Replacing the phenyl ring of benzaldehyde with the considerably smaller methyl group in acetaldehyde, removes many of the less favorable steric interactions in the active site, resulting in much lower barriers and intermediate energies.

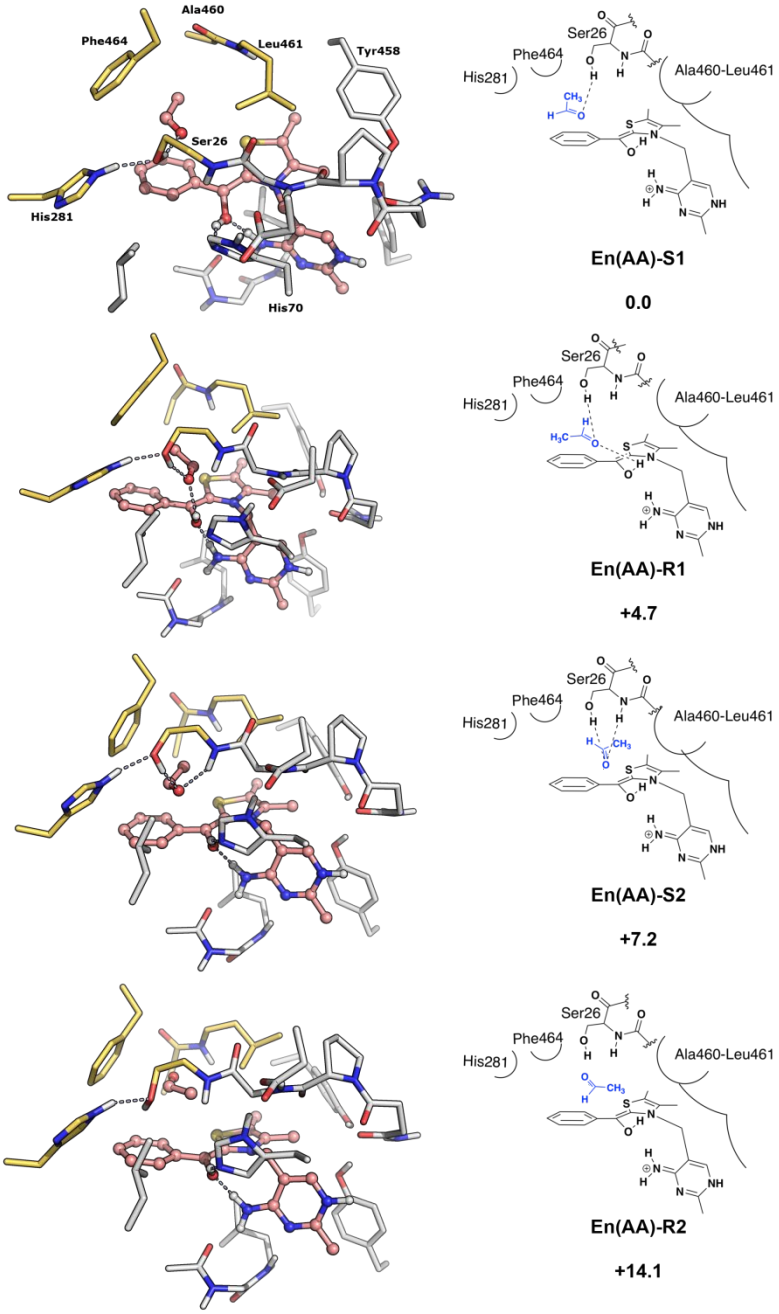


Figure 6. *Optimized geometries and schematic drawings of the different binding modes for acetaldehyde. Relative energies are given in kcal/mol. For clarity, only a part of the active site model is shown here.*

The geometries of the transition states for both the initial C-C bond formation (TS1) and the subsequent proton transfer (TS2) could be optimized but, after adding the large basis set, solvation, and zero-point energy corrections, their energies were calculated to be lower than those of the connecting intermediates. Therefore, these steps can be assumed to be barrierless, or occur with very low barriers. This is very different from the benzaldehyde reaction, where initial C-C bond formation had a quite high barrier (at least 10.7 kcal/mol, Figure 4).

The C-C bond formation results in **Int1(AA)-S** and **Int1(AA)-R**, which are calculated to be -4.8 and -3.1 kcal/mol, respectively, relative to the enamine **En(AA)-S1**. Interestingly, due to its smaller size, the methyl group of acetaldehyde can readily fit in both the large cavity in **Int1(AA)-R** and the small cavity in **Int1(AA)-S**, with the only minor rotations of the Phe464 and Leu461 side chains (Figure 8).

With acetaldehyde as the acceptor, as with benzaldehyde, C-C bond cleavage was calculated to have the highest energy barrier. Since the previous steps are reversible, the stereochemical outcome of the carboligation reaction with acetaldehyde will be determined by the energy difference between **TS3(AA)-R** and **TS3(AA)-S**. This is calculated to be slightly (0.3 kcal/mol) in favor of the *S*-product (Figure 7). In broad terms this number is consistent with the experimental observations of ~90% *ee* *S*-2-HPP,[17] but the calculated energy difference appears to be underestimated. It would not be unreasonable to attribute this discrepancy to the active site model being somewhat rigid, possibly not allowing proper relaxation of the various groups to fully adapt to the changes between the different binding modes.

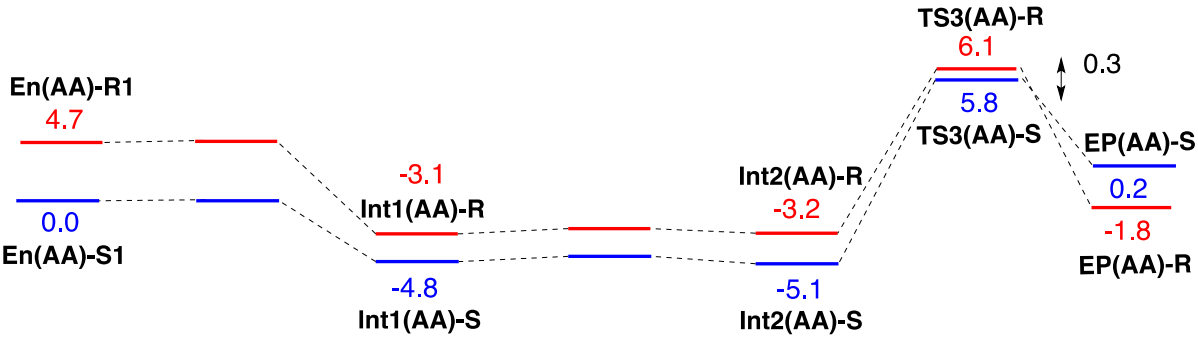


Figure 7. *Calculated potential energy profiles for the carboligation reaction between benzaldehyde (donor) and acetaldehyde (acceptor). Energies are given in kcal/mol.*

It is difficult to pinpoint the source of such a small energy difference. However, analysis of the geometries of the selectivity-determining TS3 (Figure 8) can give some hints. In **TS3(AA)-S** the methyl substituent of the acceptor points toward Leu461, while in **TS3(AA)-R** the methyl group points in the direction of Phe464. This suggests that these two residues are likely to influence the stereochemical outcome of the reaction. It is gratifying to note that this suggestion is in accord with site-directed mutagenesis studies which show that replacement of Leu461 by either Ala or Gly increases the *ee* of the *S*-product (to 98%). Conversely, but nonetheless consistent, mutation of Phe464 to Ile decreases the *ee* to 42%.[21]

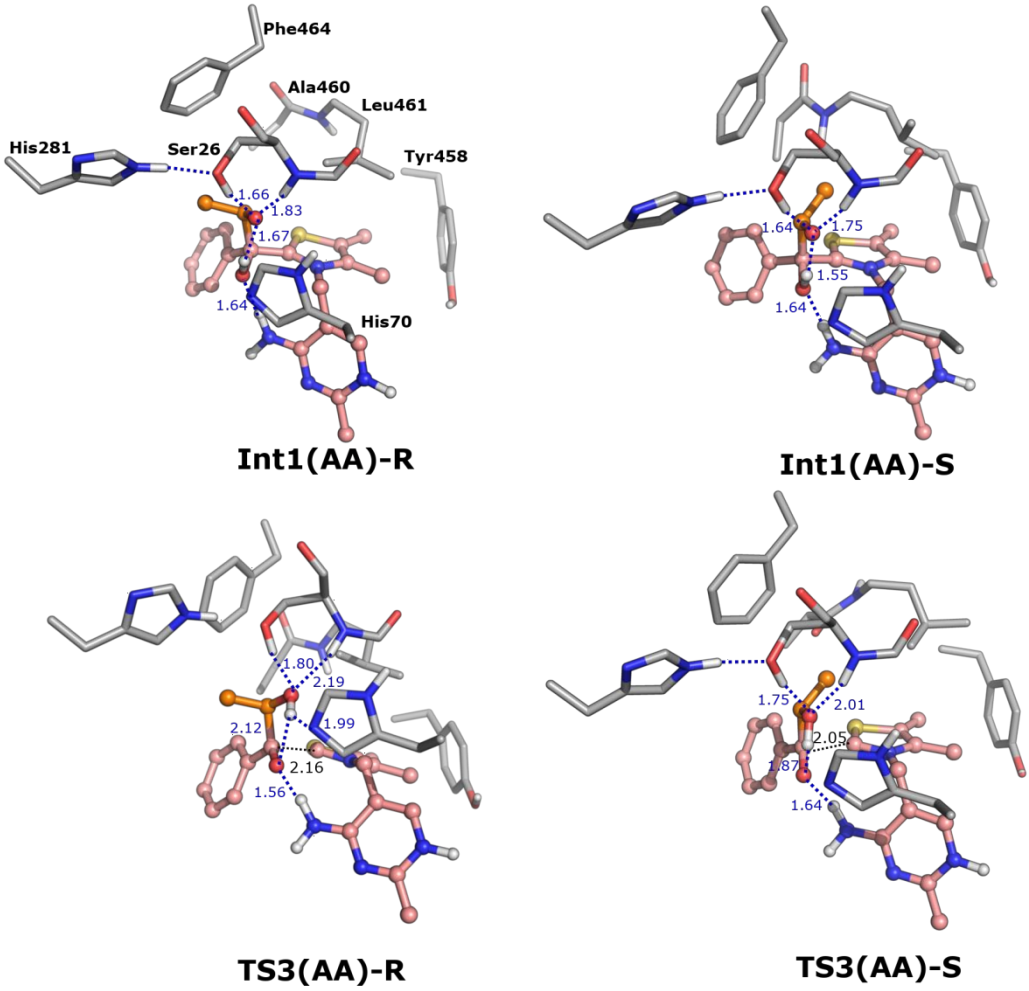


Figure 8. *Optimized geometries of selected stationary points along the paths for acetaldehyde as acceptor. For clarity, only a part of the model is shown here.*

IV. Conclusions

In this work, we have used quantum chemical calculations to investigate the reaction mechanism and the sources enantioselectivity of BFDC-catalyzed carboligation reaction. A large model of the active site, comprising more than 300 atoms, was designed on the basis of the crystal structure and the detailed mechanisms of reaction of benzaldehyde with two different acceptors, benzaldehyde and acetaldehyde, were examined, resulting in benzoin and 2-HPP, respectively. Experimentally, BFDC has been demonstrated to yield the opposite enantiomers of the two products.[17,23]

For each acceptor, several different binding conformations were first evaluated. These were then used as starting points for calculations of the chemical mechanism leading to the two possible enantiomeric outcomes of the products.

The reactions of both substrates are demonstrated to follow the same mechanism. A C-C bond is formed between the donor and acceptor, resulting in an alkoxide tetrahedral intermediate. Next, an intramolecular proton transfer takes place from the hydroxyl group of the donor moiety to the alkoxide. Finally, C-C bond cleavage yields the reaction product and recovers the ylid form of the ThDP cofactor.

Although the overall reaction mechanisms are the same, the energy profiles display some differences. The most significant differences are apparent in the barriers of the C-C bond formation and cleavage steps, with the latter being found to be the rate- and selectivity-determining step for both substrates. Further, the calculated barriers attest to the reversibility of the reactions, and suggest that HPP formation would be faster than benzoin formation. Both predictions are in line with experimental evidence.[21]

Importantly, the calculations satisfactorily reproduce the experimentally observed change of enantioselectivity in the formation of benzoin *vs.* 2-HPP products. Namely, for benzoin formation, the calculated energy difference of 9.3 kcal/mol in favor of the *R*-enantiomer is consistent with the observed complete stereocontrol,[23] while for the formation of 2-HPP, the calculated energy difference of 0.3 kcal/mol in favor of the *S*-enantiomer is consistent with the observation of *S*-HPP with 90% *ee*.[17]

The obtained energy differences are overestimated in the case of benzoin formation and underestimated in the case of HPP formation. The reasons for this could be the size and the

rigidity of the adopted active site model. Although the model is more than 300 atoms large (which is a considerable size in this kind of modeling studies), it still represents only a small part of the enzyme. A number of atoms are kept fixed at the periphery of the model to maintain the overall structure of the active site. This scheme grants some flexibility to the model, but it is likely that a larger model could potentially allow better flexibility of the various groups to adopt to the changes that take place during the reaction thereby yielding better quantitative agreement. However, we do not expect this to qualitatively change the obtained energy profiles nor alter the conclusions. In addition, a much larger model would be associated with other shortcomings, such as multiple minima problems that would require extensive sampling.[28-31]

Previous experimental and modeling studies indicated that residues Phe464 and Leu461 would play a major role in stereochemical control of carboligation.[21,27] Inspection of the optimized geometries of the TSs for the C-C bond formation (**TS1**) and C-C bond cleavage (**TS3**) steps confirms the relevance of those residues. Further, it suggests an important role for Ser26 in overall reactivity as well as in controlling the stereochemical outcome. This residue forms hydrogen bonds to the oxygen of the acceptor through its hydroxyl group and its backbone NH. These interactions are clearly important, since the initial C-C bond formation could not take place when one or both were lost. Importantly for the enantioselectivity, the hydrogen bonds to Ser26 hinder the mobility of the acceptor and set restrictions as to how the substituent of the acceptor can point in the active site. In the TS leading to the *R*-enantiomer, the substituent points into the large cavity, while in the TS leading to the *S*-enantiomer the substituent points into the small cavity (see Figure 1).

In the case of benzoin formation, the phenyl group of the benzaldehyde acceptor points into the large cavity, since pointing into the small cavity results in numerous steric clashes that lead to a higher energy intermediate. As a result, the *R*-enantiomer of benzoin is the major product. On the other hand, the methyl group of the acetaldehyde acceptor readily fits into the small cavity. This turns out to be energetically favorable and leads to the preferred formation of the *S*-enantiomer of 2-HPP. A plausible reason for this result is that the attractive dispersion interactions between the methyl group of acetaldehyde and the side-chains of the residues lining the small pocket are stronger than any unfavorable steric clashes. Similar observations were recently reported in the study of the enantioselectivity of secondary alcohol dehydrogenase from *Thermoanaerobacter brockii*.[37] Conceivably, the methyl group has the

perfect size to fit into the small cavity. If true, there would be a trend for acceptors with increasingly larger substituents to show increasing preference for the large cavity and, concomitantly, the ratio of the *R*-product would increase. This analysis is consistent with the observation made when propanal is used as acceptor. This aldehyde only has one extra carbon compared to acetaldehyde, but the selectivity of the reaction is already reversed (*ee* = 28% *R*).[27]

We believe that the insights developed in the present work will be valuable for future applications of BFDC (and other ThDP-dependent enzymes) in asymmetric biocatalysis. It should also be stressed here that calculation of enantioselective mechanisms requires inordinately high accuracy, and the fact that the current calculations account for the experimentally observed trends and selectivity reversal for the two substrates is remarkable and is certainly a testament to the strength of the adopted quantum chemical methodology in this field.

Acknowledgements

FH acknowledges financial support from the Swedish Research Council and MJM thanks the National Science Foundation (CHE 1306877).

Supporting Information

Superposition of structures of different benzaldehyde and acetaldehyde binding modes, tables with calculated energies and energy corrections, and Cartesian coordinates of all optimized structures.

References

[1] Hegeman, G. D. Synthesis of the Enzymes of the Mandelate Pathway by *Pseudomonas putida*. I. Synthesis of Enzymes by the Wild Type. *J. Bacteriol.* **1966**, *91*, 1140–1154.
[2] Hegeman, G. D. Benzoylformate Decarboxylase (*Pseudomonas putida*). *Methods Enzymol.* **1970**, *17*, 674–678.

[3] Andrews, F. H.; McLeish, M. J. Substrate Specificity in Thiamin Diphosphate-Dependent Decarboxylases. *Bioorganic Chemistry* **2012**, *43*, 26-36.
[4] Candy, J. M.; Duggleby, R. G. Structure and Properties of Pyruvate Decarboxylase and Site-Directed Mutagenesis of the *Zymomonas mobilis* Enzyme. *Biochim. Biophys. Acta - Protein Struct. Mol. Enzymol.* **1998**, *1385*, 323–338.
[5] Liu, M.; Sergienko, E. A.; Guo, F.; Wang, J.; Tittmann, K.; Hübner, G.; Furey, W.; Jordan, F. Catalytic Acid−Base Groups in Yeast Pyruvate Decarboxylase. 1. Site-Directed Mutagenesis and Steady-State Kinetic Studies on the Enzyme with the D28A, H114F, H115F, and E477Q Substitutions. *Biochemistry* **2001**, *40*, 7355–7368.
[6] Sergienko, E. A.; Jordan, F. Catalytic Acid-Base Groups in Yeast Pyruvate Decarboxylase. 2. Insights into the Specific Roles of D28 and E477 from the Rates and Stereospecificity of Formation of Carboligase Side Products. *Biochemistry* **2001**, *40*, 7369–7381.
[7] Schütz, A.; Sandalova, T.; Ricagno, S.; Hübner, G.; König, S.; Schneider, G. Crystal Structure of Thiamindiphosphate-Dependent Indolepyruvate Decarboxylase from *Enterobacter cloacae*, an Enzyme Involved in the Biosynthesis of the Plant Hormone Indole-3-Acetic Acid. *Eur. J. Biochem.* **2003**, *270*, 2312–2321.
[8] Berthold, C. L.; Gocke, D.; Wood, M. D.; Leeper, F. J.; Pohl, M.; Schneider, G. Structure of the Branched-Chain Keto Acid Decarboxylase (KdcA) from *Lactococcus lactis* Provides Insights into the Structural Basis for the Chemoselective and Enantioselective Carboligation Reaction. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2007**, *63*, 1217–1224.
[9] Versées, W.; Spaepen, S.; Vanderleyden, J.; Steyaert, J. The Crystal Structure of Phenylpyruvate Decarboxylase from *Azospirillum brasilense* at 1.5 Å Resolution: Implications for Its Catalytic and Regulatory Mechanism. *FEBS J.* **2007**, *274*, 2363–2375.
[10] Hasson, M. S.; Muscate, A.; McLeish, M. J.; Polovnikova, L. S.; Gerlt, J. A.; Kenyon, G. L.; Petsko, G. A.; Ringe, D. The Crystal Structure of Benzoylformate Decarboxylase at 1.6 Å Resolution: Diversity of Catalytic Residues in Thiamin Diphosphate-Dependent Enzymes. *Biochemistry* **1998**, *37*, 9918–9930.
[11] Sergienko, E. A.; Wang, J.; Polovnikova, L.; Hasson, M. S.; McLeish, M. J.; Kenyon, G. L.; Jordan, F. Spectroscopic Detection of Transient Thiamin Diphosphate-Bound Intermediates on Benzoylformate Decarboxylase. *Biochemistry* **2000**, *39*, 13862–13869.
[12] Polovnikova, E. S.; McLeish, M. J.; Sergienko, E. A.; Burgner, J. T.; Anderson, N. L.; Bera, A. K.; Jordan, F.; Kenyon, G. L.; Hasson, M. S. Structural and Kinetic Analysis of Catalysis by a Thiamin Diphosphate-Dependent Enzyme, Benzoylformate Decarboxylase. *Biochemistry* **2003**, *42*, 1820–1830.
[13] Yep, A.; Kenyon, G. L.; McLeish, M. J. Saturation Mutagenesis of Putative Catalytic Residues of Benzoylformate Decarboxylase Provides a Challenge to the Accepted Mechanism. *Proc. Natl. Acad. Sci.* **2008**, *105*, 5733–5738.
[14] Bruning, M.; Berheide, M.; Meyer, D.; Golbik, R.; Bartunik, H.; Liese, A.; Tittmann, K. Structural and Kinetic Studies on Native Intermediates and an Intermediate Analogue in Benzoylformate Decarboxylase Reveal a Least Motion Mechanism with an Unprecedented Short-Lived Predecarboxylation Intermediate. *Biochemistry* **2009**, *48*, 3258–3268.

[15] Planas, F.; Sheng, X.; McLeish, M. J.; Himo, F. A Theoretical Study of the Benzoylformate Decarboxylase Reaction Mechanism. *Front. Chem*. **2018**, 6: 205.
[16] Planas, F.; McLeish, M. J.; Himo, F. Computational Characterization of Enzyme-Bound Thiamin Diphosphate Reveals a Surprisingly Stable Tricyclic State: Implications for Catalysis. *Beilstein J. Org. Chem.* **2019**, *15*, 145–159.
[17] Wilcocks, R.; Ward, O. P.; Collins, S.; Dewdney, N. J.; Hong, Y.; Prosen, E. Acyloin Formation by Benzoylformate Decarboxylase from *Pseudomonas putida*. *Appl. Environ. Microbiol.* **1992**, *58*, 1699–1704.
[18] Iding, H.; Dünnwald, T.; Greiner, L.; Liese, A.; Müller, M.; Siegert, P.; Grötzinger, J.; Demir, A. S.; Pohl, M. Benzoylformate Decarboxylase from *Pseudomonas putida* as Stable Catalyst for the Synthesis of Chiral 2-Hydroxy Ketones. *Chem. - A Eur. J.* **2000**, *6*, 1483–1495.
[19] Siegert, P.; McLeish, M. J.; Baumann, M.; Iding, H.; Kneen, M. M.; Kenyon, G. L.; Pohl, M. Exchanging the Substrate Specificities of Pyruvate Decarboxylase from *Zymomonas mobilis* and Benzoylformate Decarboxylase from *Pseudomonas putida*. *Protein Eng. Des. Sel.* **2005**, *18*, 345–357.
[20] Janzen, E.; Müller, M.; Kolter-Jung, D.; Kneen, M. M.; McLeish, M. J.; Pohl, M. Characterization of Benzaldehyde Lyase from *Pseudomonas fluorescens*: A Versatile Enzyme for Asymmetric C-C Bond Formation. *Bioorg. Chem.* **2006**, *34*, 345–361.
[21] Knoll, M.; Müller, M.; Pleiss, J.; Pohl, M. Factors Mediating Activity, Selectivity, and Substrate Specificity for the Thiamin Diphosphate-Dependent Enzymes Benzaldehyde Lyase and Benzoylformate Decarboxylase. *ChemBioChem* **2006**, *7*, 1928–1934.
[22] Dünkelmann, P.; Kolter-Jung, D.; Nitsche, A.; Demir, A. S.; Siegert, P.; Lingen, B.; Baumann, M.; Pohl, M.; Müller, M. Development of a Donor-Acceptor Concept for Enzymatic Cross-Coupling Reactions of Aldehydes: The First Asymmetric Cross-Benzoin Condensation. *J. Am. Chem. Soc.* **2002**, *124*, 12084–12085.
[23] Demir, A. S.; Dünnwald, T.; Iding, H.; Pohl, M.; Müller, M. Asymmetric Benzoin Reaction Catalyzed by Benzoylformate Decarboxylase. *Tetrahedron Asymmetry* **1999**, *10*, 4769–4774.
[24] Lingen, B.; Grötzinger, J.; Kolter, D.; Kula, M.-R.; Pohl, M. Improving the Carboligase Activity of Benzoylformate Decarboxylase from *Pseudomonas putida* by a Combination of Directed Evolution and Site-Directed Mutagenesis. *Protein Eng. Des. Sel.* **2002**, *15*, 585–593.
[25] Kneen, M. M.; Pogozheva, I. D.; Kenyon, G. L.; McLeish, M. J. Exploring the Active Site of Benzaldehyde Lyase by Modeling and Mutagenesis. *Biochim. Biophys. Acta - Proteins Proteomics* **2005**, *1753*, 263–271.
[26] Siegert, P.; Vergleichende Charakterisierung der Decarboxylase- und Carboligasereaktion der Benzoylformiatdecarboxylase aus *Pseudomonas putida* und der Pyruvatdecarboxylase aus *Zymomonas mobilis* mittels gerichteter Mutagenese. Doctoral thesis, Heinrich-Heine University (Düsseldorf), **2000**.
[27] Gocke, D.; Walter, L.; Gauchenova, E.; Kolter, G.; Knoll, M.; Berthold, C. L.; Schneider, G.; Pleiss, J.; Müller, M.; Pohl, M. Rational Protein Design of ThDP-Dependent Enzymes-Engineering Stereoselectivity. *Chembiochem* **2008**, *9*, 406–412

[28] Himo, F. Recent Trends in Quantum Chemical Modeling of Enzymatic Reactions. *J. Am. Chem. Soc.* **2017***, 139*, 6780–6786.
[29] Siegbahn, P. E. M.; Himo, F. The Quantum Chemical Cluster Approach for Modeling Enzyme Reactions. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2011**, *1*, 323–336.
[30] Blomberg, M.R.A.; Borowski, T.; Himo, F.; Liao, R.-Z.; Siegbahn, P.E.M. Quantum Chemical Studies of Mechanisms for Metalloenzymes. *Chem. Rev.* **2014**, *114*, 3601−3658.
[31] Lind, M. E. S.; Himo, F. Quantum Chemistry as a Tool in Asymmetric Biocatalysis: Limonene Epoxide Hydrolase Test Case. *Angew. Chemie - Int. Ed.* **2013**, *52*, 4563–4567.
[32] Lind, M. E. S.; Himo, F. Quantum Chemical Modeling of Enantioconvergency in Soluble Epoxide Hydrolase. *ACS Catal.* **2016**, *6*, 8145–8155.
[33] Lind, M. E. S.; Himo, F. Theoretical Study of Reaction Mechanism and Stereoselectivity of Arylmalonate Decarboxylase. *ACS Catal.* **2014**, *4*, 4153–4160.
[34] Payer, S. E.; Sheng, X.; Pollak, H.; Wuensch, C.; Steinkellner, G.; Himo, F.; Glueck, S. M.; Faber, K. Exploring the Catalytic Promiscuity of Phenolic Acid Decarboxylases: Asymmetric, 1,6-Conjugate Addition of Nucleophiles Across 4-Hydroxystyrene. *Adv. Synth. Catal.* **2017**, *359*, 2066-2075.
[35] Sheng, X.; Himo, F. Theoretical Study of Enzyme Promiscuity: Mechanisms of Hydration and Carboxylation Activities of Phenolic Acid Decarboxylase. *ACS Catal.* **2017**, *7*, 1733-1741.
[36] Moa, S.; Himo, F. Quantum Chemical Study of Mechanism and Stereoselectivity of Secondary Alcohol Dehydrogenase. *J. Inorg. Biochem.* **2017**, *175*, 259-266.
[37] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, and D. J. Fox, *Gaussian 09, Revision A.02*, Gaussian, Inc., Wallingford CT, 2016.
[38] Becke, A. D. Density-Functional Thermochemistry. III. The Role of Exact Exchange. *J. Chem. Phys.* **1993**, *98*, 5648–5652.
[39] Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron Density. *Phys. Rev. B* **1988**, *37*, 785–789.
[40] Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A Consistent and Accurate Ab Initio Parametrization of Density Functional Dispersion Correction (DFT-D) for the 94 Elements H-Pu. *J. Chem. Phys.* **2010**, *132*, 154104.
[41] Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the Damping Function in Dispersion Corrected Density Functional Theory. *J. Comput. Chem.* **2011**, *32*, 1456–1465.

[42] Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Universal Solvation Model Based on Solute Electron Density and on a Continuum Model of the Solvent Defined by the Bulk Dielectric Constant and Atomic Surface Tensions. *J. Phys. Chem. B* **2009**, *113*, 6378–6396.

TOC Graphic