



# Globally Optimized Molecular Embeddings for Dynamic Reaction Solvate Shell Optimization and Active Site Design

## From Abstract Catalyst Design Via Electric Field Optimization Back to the Real World

Dominik M. Behrens<sup>1</sup> · Bernd Hartke<sup>1</sup>

Accepted: 23 July 2021 / Published online: 4 August 2021  
© The Author(s) 2021

### Abstract

We demonstrate how a full QM/MM derivatization of the recently developed GOCAT model can be utilized in the global optimization of molecular embeddings. To this end, we provide two distinct examples: An  $S_N2$  reaction, and one enzymatic example of recent interest, the ketosteroid isomerase. These serve us to highlight the advantages of such an approach and sketch the roadmap for further improvements.

**Keywords** Catalysis design · Oriented external electric fields · Global optimization · Evolutionary algorithms · Enzymes · QM/MM

## 1 Introduction

Computational catalyst design has made impressive progress in recent years [1]. Since reliable ways of finding specific catalysts have been a long-standing goal in synthetic chemistry, their *in silico* design has even been called a “Holy Grail” [2]. With the ever-growing and ever-improving theoretical toolset our community has at its disposal and with the advance of sophisticated machine-learning approaches, this has certainly become a bustling field of opportunity.

Our recently developed Globally Optimized Catalyst (GOCAT) scheme [3] joins several other promising approaches to inverse, *in silico* design of catalysts [4–8] (also see the discussion in Ref. [1] and in the remainder of this contribution). In the GOCAT scheme, we employ genetic algorithms (GA) to find globally optimal embeddings for the reaction paths of arbitrary reactions, to lower their energy barrier via electric fields (which is one important influence of electric fields on chemistry [9]). These embeddings, up until now, have been exclusively made up of abstract point charges of optimized values, signs and spatial position, but

remained unchanged throughout the whole reaction. In a first iteration in the development of this idea, the catalysis of these fixed point charges was evaluated on a pre-defined minimum energy path (MEP), and the transferability to dynamic reactions was implied by introducing restrictions on the residual gradient at each point in the MEP. Following a recent extension [10], this is no longer the case. Instead, we allow the recalculation of the MEP at each step in the global optimization, yielding far more effective and certainly more transferable embeddings. With this strategy, we could confirm external electric field effects on a Diels-Alder reaction that not only lead to strong catalysis, but also to mechanistic changes from a concerted pathway to a zwitterionic, step-wise one, as predicted earlier [11–13] and confirmed by different experimental single-molecule techniques [14, 15]. Although the MEP recalculation dramatically improves the GOCAT approach, it also highlights one of its downsides: The complex electric fields generated by the point charges are often wildly inhomogeneous and thus become hard to cast into chemical counter-parts. However, since these complex electric fields constitute an abstractly defined global optimum in terms of catalytic effects, it can be expected that less complex fields that approximate them but can be realized more simply by actual chemical frameworks will still lead to substantial catalysis. While this is one way of exploiting our GOCAT scheme, which we currently follow

✉ Bernd Hartke  
hartke@pctc.uni-kiel.de

<sup>1</sup> Institute for Physical Chemistry, Kiel University,  
Max-Eyth-Strasse 2, 24118 Kiel, Germany

in our lab, another way of derivatizing the original GOCAT idea is presented here.

When thinking of a way to manipulate electric fields at sub-molecular level and guaranteeing the right orientation of the reactant during the reaction, enzymes spring to mind. An example of recent interest is the *ketosteroid isomerase* (KSI). By employing Stark spectroscopy on KSI, Wu and Boxer [16] suggested that the extraordinary catalytic proficiency of KSI is owed to the strong electric field generated within the active site of the enzyme.

Consequently, the question is: How to bridge the gap between catalytic embedding and enzymatic design? As has been shown in the past [17], the *in silico* design of new enzymes has made huge steps forward, and certainly warrants the search for ever new ways of fueling its success [18, 19]. An intuitive way of extracting enzymatic design improvement from a GOCAT result would be to start from a known catalytically active basis and then improve upon it by finding point mutations resulting in a more favourable electrostatic field. A similar approach has been recently suggested by Beker and Sokalski [20].

Given the remarkable universal applicability of the GOCAT scheme, we aspire to a different way of utilizing it. Since our point charge embeddings are handled via a simplified QM/MM approach, the transition to full QM/MM embeddings using actual atoms is both intuitive and easily realized. With enzymes in mind, this could become a tool to automatically find an optimal active site for any arbitrary reaction. The very concept of predicting catalytic structures using protein building blocks is nothing new – in fact, it has been around and refined for more than two decades, with the Houk group being a central driving force [21, 22]. But our approach could significantly refine the process of finding these theozymes, as Houk calls them, for a given reaction. Both the fact that we can easily incorporate not only amino acid side chains but also transition metal ions, water and any other entities in arbitrary numbers, as well as the fact that we globally optimize both these entities and the *reaction path* itself hold great promise to this effect.

### 1.1 Using Actual Molecules

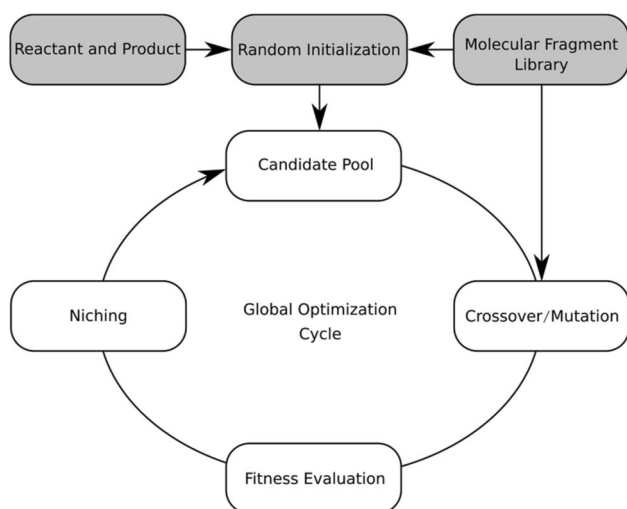
Owing to the recalculation of MEPs in every global optimization step, the required input can be as minimalistic as only two structures, reactant and product. Together with a library of molecules (e.g. amino acid side chains, solvent molecules, etc.) and an efficient genetic algorithm, this produces increasingly proficient active site candidates, with the promise of eventually obtaining the objectively best (global optimum). One of the huge benefits of this approach is that it not only automatically generates appropriate candidates, but it also provides the entire MEP, including the transition state, right out of the box. On the downside, the size

of the search space (depending on the supplied molecule library and desired amount of entities) and computational effort put heavy limits on the underlying quantum mechanical methods. However, using efficient codes such as the new semi-empirical methods of the GFN-xTB family [23] for the QM part and the very simple non-polarizable OPLS-AA force field [24] for the MM part, global optimizations with reactants of up to 30–50 QM atoms and upwards of 200 MM atoms are absolutely feasible.

## 2 Methodology

Going from an abstract concept such as point charges to actual molecules, one has to consider several challenges:

1. The point charges in the GOCAT model are fixed in space, magnitude and sign throughout the full MEP, which could be argued to be a valid approach for catalysis by abstract electric fields. With molecules, this is harder to justify. In fact, the ability for molecules to move and rearrange during the reaction can be seen as one of the vital advantages this approach has over the rigid point charges.
2. However, including the molecular degrees of freedom in the MEP optimization will necessarily increase the time spent in MEP and fitness evaluation and must be balanced cautiously.
3. Another point of consideration is the search space: While point charges only impact the electrostatic interactions of the reaction centre, molecules will also impact the dispersion interactions (in the MM picture), and thus open up an additional multitude of optimizable degrees of freedom, possibly increasing the search space radically.
4. It is thus intuitive to stick to embeddings that have a limited selection of building blocks, such as solvents, crystal- or surface structures and enzymes.
5. In superstructures such as enzymes, the overall geometry of active sites is also determined and restrained by the structures not involved in the reaction itself (i.e., the backbone). However, in our current global optimization approach, only individual fragments of the superstructure (like the side chains of amino acids) can be included, in order to preserve practicality of the global optimization operators. This is in contrast to many of the computational enzyme design projects (e.g., Refs. [17, 18, 22]), in which including the backbone often constitutes a significant part of the overall effort, while possibly limiting the predictive power due to computational protein folding not being perfect yet.



**Fig. 1** A scheme of the modified global optimization cycle used in OGOLEM

Despite these challenges, we found it worthwhile to pursue this avenue; In the following, the modifications made to our OGOLEM [25] code will be explained in more detail. The modified global optimization cycle is shown in Fig. 1.

The overall process remains largely the same as previously. The most important differences are highlighted in the following:

**Random initialization** Instead of initializing point charges with random values in Cartesian space, the initialization randomly draws  $N$  molecules from the library to form the initial candidates. A candidate thus contains the structural information of the desired reaction reactant and product, as well as the structure of  $N$  random molecular fragments. This also includes the force field parameters, as well as the bonding information.

**Crossover/Mutation** The most important crossover operator in this genetic algorithm scheme is still defined by cutting two structures along a randomly oriented plane and then exchanging the pieces to obtain two new structures. In applications of OGOLEM to heterogeneous atomic/molecular clusters consisting of two (or more) particle types A and B, we had to post-adjust this cutting plane or to convert A-particles to B-particles or vice versa, in order to avoid changing the cluster composition by a crossover operation. Here, however, a change in the nature of one or more individual molecular fragments is an allowed and possibly desired event. Along this line also comes a new mutation operator that switches individual molecular fragments in candidates for a random new one drawn from the library.

**Fitness Evaluation** This is the heart of the global optimization; The fitness evaluation first locally optimizes the

candidate reactant and product structures individually. This includes optimization of the molecular fragments, which will in most cases then differ between the two end points of the reaction. Afterwards, a Nudged Elastic Band (NEB) [26] calculation is done between the two structures and finally, the fitness is calculated from the finished path. The fitness includes:

1. The residual mean square gradient (RMSG) of the reactant, transition state and product structures. This ensures that these points remain close to actual minima on the potential energy surface.
2. The sum of all energy barriers in the path. This is the main driver to produce catalysis and will favour candidates with small energy barriers.
3. The sum of the residual mean square distance (RMSD) between the individual frames. This will favour individuals that are well pre-organized and do not need extensive structural reorganization during the reaction.

In principle, the minimum energy path (MEP) thus defines the fitness exclusively and finding it in a reasonably short time is highly desirable. In our preliminary tests, we made extensive use of the GFN-xTB Hamiltonians for the QM part, as they allow for very quick electronic potential gradient evaluations. We are aware that xTB is not specifically parametrized to deliver accurate total energies. However, we are mostly interested in *relative* energies and are thus less limited in our choice of QM Hamiltonians. As long as the relative energies are accurate enough between different solution candidates, our global optimization scheme will be able to find optima that will be transferable to higher quality ab initio calculations.

**Niching** In past implementations of the GOCAT model, the atom electrostatic potential (ESP) values were used as a criterion to preserve structural diversity in the candidate pool. Although this is technically still a valid niching parameter, we have also considered more structurally-aware approaches such as SOAPs (smooth overlaps of atomic positions) [27]. The advantages of SOAPs in this application are their inherent independence on atom count and type ordering, facilitating the comparison between structures with wildly differing molecular typing. However, as of now, we lack sufficient data on whether or not this approach can benefit the global optimization process in the long run.

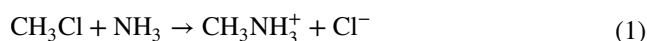
**Molecular Fragment Library** Depending on the particular system one wants to optimize, the molecular library can include any arbitrary number of different molecules. For our  $S_N2$  example, we included only the solvent molecule. For the enzyme example, we included amino acid side chains of 17 amino acids. They were all capped at the  $\alpha$ -carbon atom, removing both the carboxylate as well as

the amine group. Because of the way they are capped, we chose to omit Glycine, which would essentially become a methane molecule, as well as Leucine, because of its similarity to Isoleucine, and lastly Proline, because it can not be capped in this particular way without losing its main structural feature.

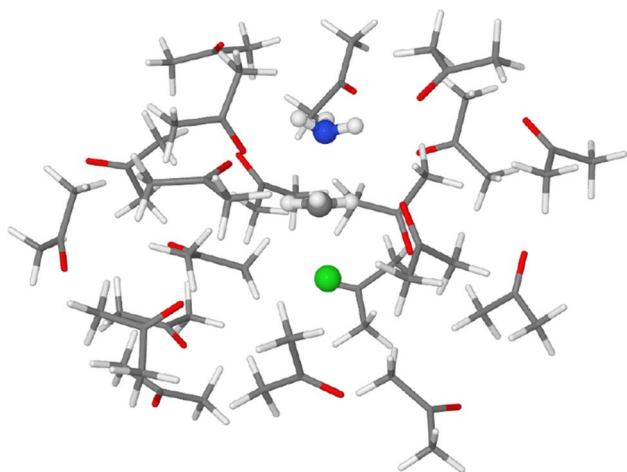
### 3 Results and Discussion

#### 3.1 Solvated $S_N2$ Reaction

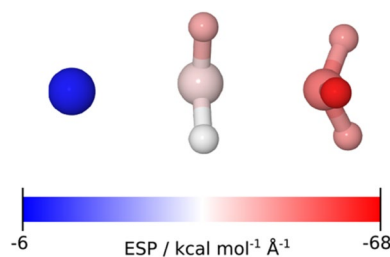
For a first foray into the use of actual molecules, we went back to the well-understood  $S_N2$  Menshutkin reaction,



which was also used in our first GOCAT publication. But instead of point charges, we added actual solvent molecules into the QM/MM approach. Aprotic, polar solvents are ideal for reactions such as this, and thus we chose to add 20 explicit acetone molecules (s. Fig. 2). This approach is very similar to a cluster optimization problem, and in an earlier publication [28] we treated it as such (with explicit water molecules as solvent). Here, the difference is in the optimization goal: It is not the overall energy of the cluster, but the energy *barrier* of the reaction taking place inside. By extension, this usage of the model also provides plenty of interesting avenues for further exploration, such as optimization of dynamic reaction solvation shell structures. For now, however, the focus is on the catalytic picture; The reaction in question is quite easily catalysed by a linear electric field pointing along the emerging bond axis. As is expected, we retrieve precisely this picture in the case of using molecules.



**Fig. 2** Globally optimized structure for the acetone solvated  $S_N2$  system at the transition state. The MM part is displayed as sticks-only

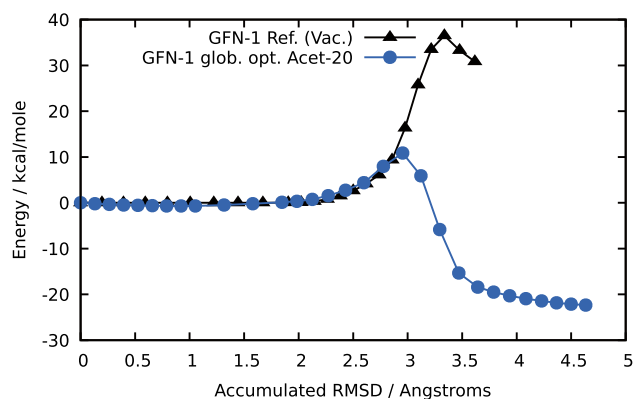


**Fig. 3** Central quantum-mechanical part of the transition state for the globally optimized 20-acetone  $S_N2$  reaction system. The atoms have been colorized according to their ESP values (cf. main text)

Fig. 3 shows the electrostatic potential (ESP) values on the individual atoms in the transition state generated by the optimized solvent cluster. They are, at each reactant atom  $i$ , calculated from the point charges  $q_j$  assigned to the atoms  $j$  in the OPLSAA Force Field (with  $e$  being the elementary charge) by:

$$\text{ESP}(r_i) = e^2 \sum_j \frac{q_j}{r_{ij}} \quad (2)$$

These values allow a visualization of how the electric field generated by the embedding influences the individual atoms. Greater negative values result from more negatively charged entities while less negative or positive values correspond to more positively charged entities in the atom's environment. The difference between the ESP values at the individual atoms then corresponds to how the electronic structure is polarized in consequence. The energy path for the optimized individual shown here can be found in Fig. 4. As expected, the linear polarization resulting from the solvent embedding catalyses this reaction significantly



**Fig. 4** Shown are the energies at the individual frames along the MEP of the (vacuum) reference and the globally optimized acetone-20  $S_N2$  reaction system

in comparison to the reaction in vacuum, its energy barrier is almost quartered.

Under the hood, the optimization problem thus remains very similar to the original GOCAT approach; it is still the electrostatics that are the driving force of catalysis (as they remain the only thing that enters via the QM/MM ansatz). The difference lies in the constraints that are put on the positions of the charged entities, as they now have become chemically intuitive molecules surrounding the reaction.

Of course, the implication of this example for the Menshutkin reaction in acetone is neither that one should attempt to manipulate solvent molecules into such globally optimal arrangements around the reaction centre, nor that freely moving solvent molecules will arrange themselves in this way each and every time a reactive event is about to occur. Instead, this example shows that surrounding acetone molecules can be arranged in a way that a catalytically relevant electric field results at the reaction centre, that a global optimization provides the qualitatively expected result, and what the maximally possible effect of this kind can be in quantitative terms. When this reaction actually happens in this solvent, a distribution of solvent molecule arrangements will occur, between this maximally favourable one and others that may actually disfavour the reaction, as shown before [28]. A large-scale, long-time molecular dynamics study could reveal the probabilities for such arrangements to occur, allowing for an estimation of the true catalytic effect obtainable in this way, which will likely be lower than the one shown in Fig. 4. However, other groups [29, 30] have already investigated how external electric fields could order solvent molecules around a reaction center or could lead to catalysis despite solvent molecules counteracting the external field.

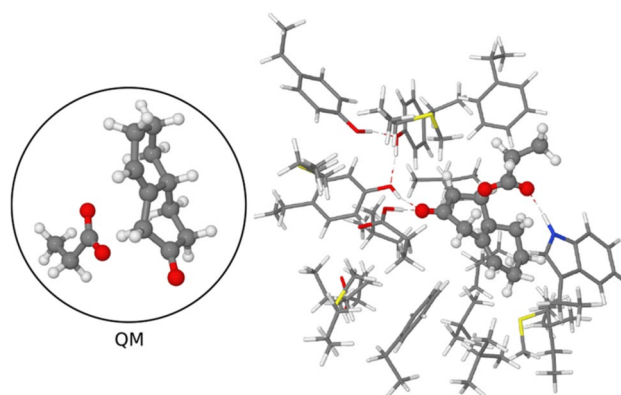
A different way to exploit the result shown here for catalysis design would be to design a hollow molecular framework

decorated with functional groups similar to acetone molecules (e.g., with similar dipole moments) and approximating the spatial distribution obtained here.

### 3.2 Ketosteroid Isomerase

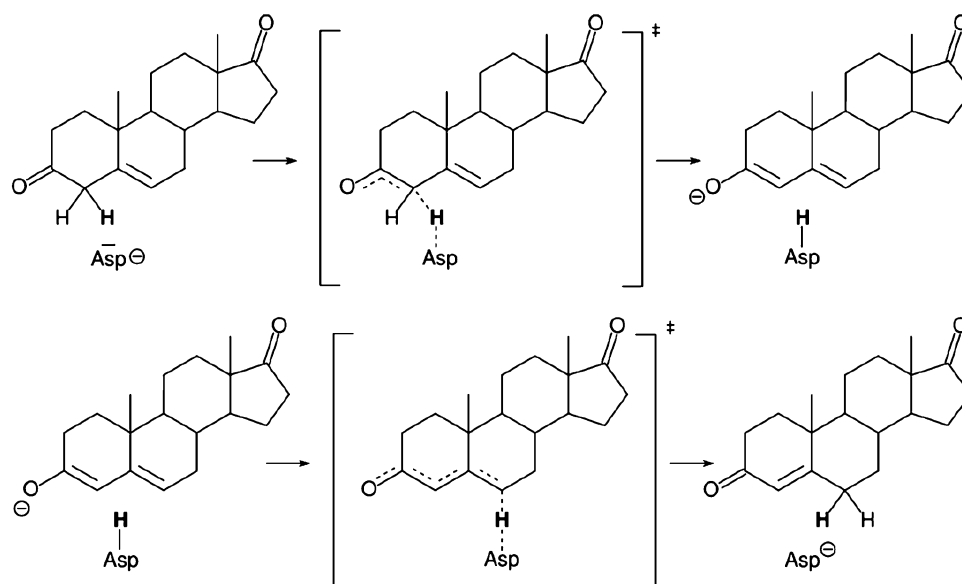
The second reaction we are addressing in an ongoing study is that of KSI. A reaction catalysed by this enzyme is shown on the example of 5-androstene-3,17-dione in two separate reaction steps [31] in Scheme 1.

For a reference, we started with the crystal structure of 1OHO [32]. We removed all the water molecules, replaced the inhibitor with a simplified reactant (cf. Fig. 5) and then manually selected 20 side chains in close proximity of the active site. These were extracted, capped and then, using

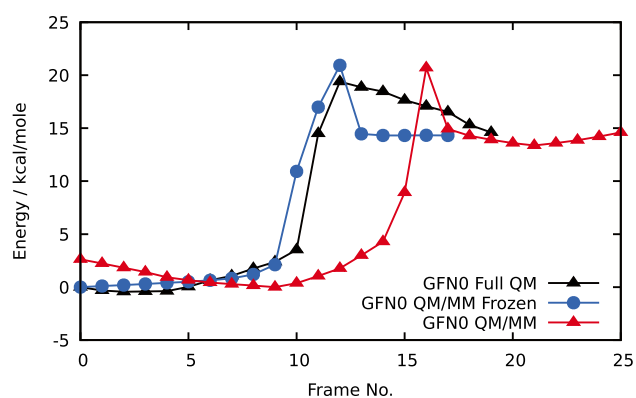


**Fig. 5** The inset on the left shows only the QM region, which is made up of the simplified reactant as well as the capped Asp. The right structure shows the entire QM/MM system, with the MM molecules shown as sticks-only

**Scheme 1** Step I (top) and II (bottom) for the reaction catalysed by KSI





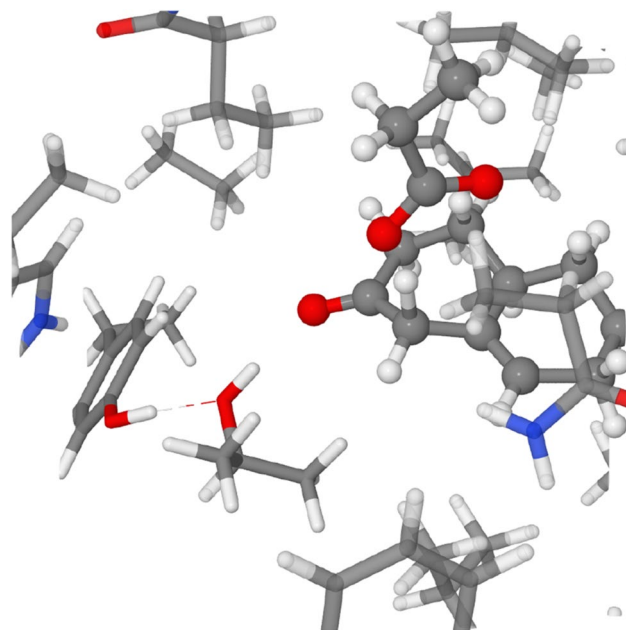


**Fig. 6** This graph shows the MEP profiles of three distinct variants of the KSI active site reference: (1) The all-QM system where *all* atoms of the system are optimized in xTB during the MEP calculation, with the  $\alpha$  carbon atoms fixed. (2) A QM/MM system where, with the exception of the Asp involved in the proton abstraction, all side chains enter as MM atoms instead, with their  $\alpha$ -carbon atoms fixed. (3) A QM/MM system with the  $\alpha$ -carbons not fixed. More info in the main text

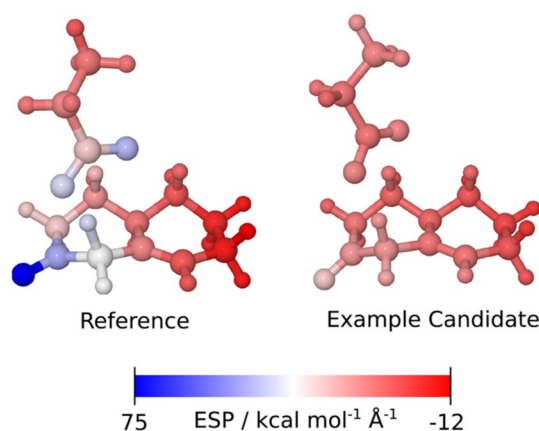
GFNO-xTB, relaxed with their  $\alpha$ -carbon atoms fixed. This was repeated for the intermediate structure (product of reaction step I). Then, a MEP was calculated between those two structures. The reaction energy profile is shown in Fig. 6. The same Figure also includes the energy profiles of two QM/MM variants of this system, where all side chains except the Asp residue involved in the proton exchange enter as MM atoms. Both of these QM/MM profiles were obtained by passing the reference structure into the fitness function that is used in the global optimization later, in order to obtain a comparable reference to optimize against. The difference between the two systems is that in one, the  $\alpha$ -carbons of the side chains were held fixed (as in the full QM system), whereas in the other, they were allowed to move freely. This is important, because during the global optimization, we do not constrain or fix the alpha carbon atoms. This turns out to be non-detrimental, since while the unfrozen QM/MM energy path features energy wells before and after the reaction (these are exclusively rearrangements in the MM subsystem, see SI for more details), the overall barrier remains almost exactly identical when measured from lowest point to highest. Both of the QM/MM systems (the structure of the frozen one is shown in Fig. 5) thus adequately reproduce the full QM energetics well enough for our purposes.

We want to finally present some of our early findings when globally optimizing this system with freely interchangeable side chains as described in the Methodology section. The input for this optimization is the Asp residue together with the simplified reactant as the QM subsystem (same as in Fig. 5) and our molecular library of side chains for the MM embedding. As of now, we can only base our analysis on very early individuals (generated in the first

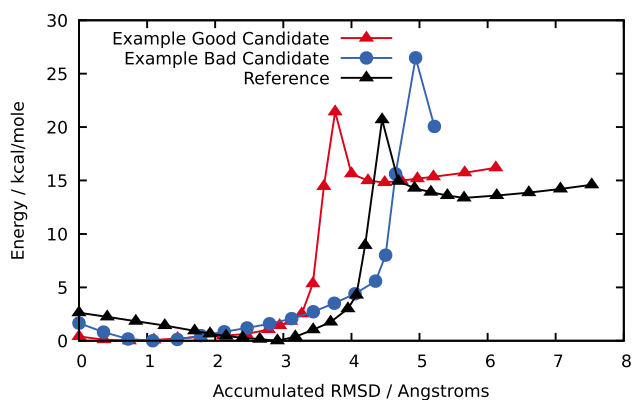
10000 global optimization cycles, which, considering the size of the search space, is certainly not enough to be even close to convergence), but even from very early on, our fitness function performs quite admirably. One good (or fit in terms of the fitness function) example from our pool of candidates is shown in Fig. 7. In this zoomed-in view, one of the main features is visible: without any bias in the population or distribution of side chains, a hydrogen-bonding network somewhat similar to the original enzymatic structure has emerged. It is only *somewhat* similar in the sense that if the ESP values are compared to the reference (Fig. 8), one



**Fig. 7** The region around the reactive QM site in an example (good) candidate produced by our global optimization algorithm. MM atoms are shown as sticks-only



**Fig. 8** Shown are the QM regions at the transition state for the (not-frozen) reference and an example (good) candidate. Atoms are colored according to the respective ESP values at their positions



**Fig. 9** Shown are the energy profiles of the (not frozen) reference of the enzymatic active site and of the example (good) candidate produced by our global optimization algorithm

can see that there is still a long ways to go; While the general trend is very similar, the reference boasts an intricate and distinct polarization scheme, with a strong stabilization of the emerging negative charge at the oxygen atom and a tryptophan hydrogen bond stabilizing the position of the reactive Asp. On the other hand, the energy profile (Fig. 9) for this (good) candidate already looks astonishingly similar to the reference and the energy barriers themselves are only  $0.74 \text{ kcal} \cdot \text{mol}^{-1}$  apart. While there are certainly individuals in the candidate pool that have a lot higher energy barriers, the ease by which (almost) equally good embeddings are found tells us that we could be missing an important ingredient to how the catalytic activity is achieved; This could certainly be tied to the backbone we have omitted, or the lack of thermodynamic considerations, which will definitely have to be addressed in future studies.

## 4 Outlook

We have demonstrated that the extension of our GOCAT model with full QM/MM molecular entities already holds some interesting potential and opportunity. The most important parameter in our modified scheme remains the fitness function: As the entire success depends on its general validity, it becomes very important to manufacture it just right, so that it does not produce artefacts, systematic errors or overfitting – which, in a complex setting such as this, turns out to be intricate. We are, however, determined and optimistic that it is, indeed, possible. Admittedly, the fitness function itself is not the only thing that can still be readily improved upon. As discussed earlier, we must certainly explore other force field options, as well as more sophisticated QM/MM embedding schemes (such as linking atoms) to allow for optimization of reactive molecular fragments, which is a feature that we are already currently working on. Closely

related to this is the facilitation of proton transfers between individual molecular fragments, as well as the inclusion of explicit water molecules, as recent studies [33] have suggested and reiterated their importance for catalytic pathways in, e.g., KSI.

Once our initial setup produces satisfactory results, we are looking to apply the model to cases like the Kemp Isomerase [17, 34], where we think lies its eventual strength: Finding optimal active sites for reactions that do not yet have specialized enzymes provided by biological evolution. As indicated above, linking our isolated side chains with a protein backbone that folds deterministically in the desired fashion is a major issue, but we view it as a major strength of our approach that we need not do that. In fact, in our lab we are currently developing an automated assembly of molecular frameworks that can be combined with an optimization towards placing catalytically active functional groups close to the spatial positions and orientations determined by the methods presented here. As discussed in the introduction of Ref. [10], this breaks down the hard task of de novo catalyst design into two more accessible steps: (1) Optimization of the immediate, catalytically active surrounding of the reaction center, done with abstract point charges in Ref. [10] or with concrete molecular fragments here; and (2) construction of a support structure for these charges or fragments, to be reported in future work.

**Supplementary Information** The online version of this article contains supplementary material available (<https://doi.org/10.1007/s11244-021-01486-1>).

**Acknowledgements** Financial support by DFG project Ha2498/16-2 is gratefully acknowledged.

**Funding** Open Access funding enabled and organized by Projekt DEAL.

**Data availability** Structures (xyz-files), partial charges and NEB log files that support the findings of this study are available in a figshare repository [35].

**Declaration**

**Conflict of interest** The authors declare that they have no conflict of interest.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Freeze JG, Kelly HR, Batista VS (2019) Search for catalysts by inverse design: artificial intelligence, mountain climbers, and alchemists. *Chem Rev* 119:6595–6612
- Houk KN, Liu F (2017) Holy grails for computational organic chemistry and biochemistry. *Acc Chem Res* 50:539–543
- Dittner M, Hartke B (2018) Globally optimal catalytic fields. Inverse design of abstract embeddings for maximum reaction rate acceleration. *J Chem Theory Comput* 14:3547–3564
- Sokalski WA (1981) Theoretical model for exploration of catalytic activity of enzymes and design of new catalysts: CO<sub>2</sub> hydration reaction. *Int J Quantum Chem* 20:231–240
- Sokalski WA (1986) Nonempirical modeling of the static and dynamic properties of the optimum environment for chemical reactions. *J Mol Struct THEOCHEM* 138:77–87
- Weymuth T, Reiher M (2014) Gradient-driven molecule construction: An inverse approach applied to the design of small-molecule fixating catalysts. *Int J Quantum Chem* 114:838–850
- Saravanan K, Kitchin JR, von Lilienfeld OA, Keith JA (2017) Alchemical predictions for computational catalysis: potential and limitations. *J Phys Chem Lett* 8:5002–5007
- Chang AM, Rudshiteyn B, Warnke I, Batista VS (2018) Inverse design of a catalyst for aqueous CO/CO<sub>2</sub> conversion informed by the Ni<sup>II</sup>-Iminothiolate complex. *Inorg Chem* 57:15474–15480
- Shaik S, Stuyver T (eds) (2021) Effects of electric fields on structure and reactivity. Theoretical and computational chemistry series. The Royal Society of Chemistry, London
- Dittner M, Hartke B (2020) Globally optimal catalytic fields for a Diels-Alder reaction. *J Chem Phys* 152(11):114106
- Meir R, Chen H, Lai W, Shaik S (2010) Oriented electric fields accelerate Diels-Alder reactions and control the endo/exo selectivity. *ChemPhysChem* 11:301–310
- Shaik S, Mandal D, Ramanan R (2016) Oriented electric fields as future smart reagents in chemistry. *Nat Chem* 8:1091–1098
- Shaik S, Ramanan R, Danovich D, Mandal D (2018) Structure and reactivity/selectivity control by oriented-external electric fields. *Chem Soc Rev* 47:5125–5145
- Aragonès AC, Haworth NL, Darwish N, Ciampi S, Bloomfield NJ, Wallace GG, Diez-Perez I, Coote ML (2016) Electrostatic catalysis of a Diels-Alder reaction. *Nature* 531:88–91
- Yang C, Liu Z, Li Y, Zhou S, Lu C, Guo Y, Ramirez M, Zhang Q, Li Y, Liu Z, Houk KN, Zhang D, Guo X (2021) Electric field-catalyzed single-molecule Diels-Alder reaction dynamics. *Sci Adv* 7:eabf0689
- Wu Y, Boxer SG (2016) A critical test of the electrostatic contribution to catalysis with noncanonical amino acids in ketosteroid isomerase. *J Am Chem Soc* 138:11890–11895
- Röthlisberger D, Khersonsky O, Wollacott AM, Jiang L, DeChancie J, Betker J, Gallaher JL, Althoff EA, Zanghellini A, Dym O, Albeck S, Houk KN, Tawfik DS, Baker D (2008) Kemp elimination catalysts by computational enzyme design. *Nature* 453(7192):190–195
- Vaissier Welborn V, Head-Gordon T (2018) Computational design of synthetic enzymes. *Chem Rev* 119:6613–6630
- Li W-L, Head-Gordon T (2021) Catalytic principles from natural enzymes and translational design strategies for synthetic catalysts. *ACS Central Sci* 7:72–80
- Beker W, Sokalski WA (2020) Bottom-up nonempirical approach to reducing search space in enzyme design guided by catalytic fields. *J Chem Theory Comput* 16:3420–3429
- Tantillo DJ, Jiangang C, Houk KN (1998) Theozymes and compuzymes: theoretical models for biological catalysis. *Curr Opin Chem Biol* 2:743–750
- Kiss G, Çelebi Ölçüm N, Moretti R, Baker D, Houk KN (2013) Computational enzyme design. *Angew Chem Int Ed* 52:5700–5725
- Grimme S, Bannwarth C, Shushkov P (2017) A robust and accurate tight-binding quantum chemical method for structures, vibrational frequencies, and noncovalent interactions of large molecular systems parametrized for all spd-block elements (Z = 186). *J Chem Theory Comput* 13:1989–2009
- Jorgensen WL, Maxwell DS, Tirado-Rives J (1996) Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J Am Chem Soc* 118:11225–11236
- Dieterich JM, Hartke B (2010) OGOLEM: Global cluster structure optimisation for arbitrary mixtures of flexible molecules. A multiscaling, object-oriented approach. *Mol Phys* 108:279–291
- Henkelman G, Jónsson H (2000) Improved tangent estimate in the nudged elastic band method for finding minimum energy paths and saddle points. *J Chem Phys* 113:9978–9985
- Himanen L, Jäger MOJ, Morooka EV, Federici Canova F, Ranawat YS, Gao DZ, Rinke P, Foster AS (2020) Dscribe: library of descriptors for machine learning in materials science. *Comput Phys Commun* 247:106949
- Li Y, Hartke B (2013) Assessing solvation effects on chemical reactions with globally optimized solvent clusters. *Chem Phys Chem* 14:2678–2686
- Dutta Dubey K, Stuyver T, Kalita S, Shaik S (2020) Solvent organization and rate regulation of a menshutkin reaction by oriented external electric fields are revealed by combined MD and QM/MM calculations. *J Am Chem Soc* 142:9955–9965
- Xu L, Izgorodina EI, Coote ML (2020) Ordered solvents and ionic liquids can be harnessed for electrostatic catalysis. *J Am Chem Soc* 142:12826–12833
- Pollack RM (2004) Enzymatic mechanisms for catalysis of enolization: ketosteroid isomerase. *Bioorg Chem* 32:341–353
- Kim SW, Cha S-S, Cho H-S, Kim J-S, Ha N-C, Cho M-J, Joo S, Kim KK, Choi KY, Oh B-H (1997) High-resolution crystal structures of Δ<sup>5</sup>-3-ketosteroid isomerase with and without a reaction intermediate analogue. *Biochemistry* 36:14030–14036
- Kędzierski P, Zaczekowska M, Sokalski WA (2020) Extreme catalytic power of ketosteroid isomerase related to the reversal of proton dislocations in hydrogen-bond network. *J Phys Chem B* 124:3661–3666
- Khersonsky O, Röthlisberger D, Wollacott AM, Murphy P, Dym O, Albeck S, Kiss G, Houk KN, Baker D, Tawfik DS (2011) Optimization of the in-silico-designed kemp eliminase KE70 by computational design and directed evolution. *J Mol Biol* 407:391–412
- Data available at. <https://figshare.com/s/80224d933ec151b74681>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.