Review

Molecular Understanding and Practical In Silico Catalyst Design in Computational Organocatalysis and Phase Transfer Catalysis—Challenges and Opportunities

Choon Wee Kee

Institute of Sustainability for Chemicals, Energy and Environment (ISCE2), Agency for Science, Technology and Research (A*STAR), 1 Pesek Road, Jurong Island, Singapore 627833, Republic of Singapore; kee_choon_wee@isce2.a-star.edu.sg

Abstract: Through the lens of organocatalysis and phase transfer catalysis, we will examine the key components to calculate or predict catalysis-performance metrics, such as turnover frequency and measurement of stereoselectivity, via computational chemistry. The state-of-the-art tools available to calculate potential energy and, consequently, free energy, together with their caveats, will be discussed via examples from the literature. Through various examples from organocatalysis and phase transfer catalysis, we will highlight the challenges related to the mechanism, transition state theory, and solvation involved in translating calculated barriers to the turnover frequency or a metric of stereoselectivity. Examples in the literature that validated their theoretical models will be showcased. Lastly, the relevance and opportunity afforded by machine learning will be discussed.

Keywords: organocatalysis; phase transfer catalysis; DFT; machine learning; organic reactions; solvation; potential energy; computational chemistry; free energy; kinetics

1. Introduction

Organocatalysis's contribution to modern society has been recognized by one of the pinnacle scientific awards—the Nobel Prize. Tremendous advancement in catalyst design and activation mode has been achieved since its inception [1,2]. Phase transfer catalysis (PTC) usually involves using a lipophilic organocatalyst to facilitate the movement of reactive ions into an organic phase where the reactants are in high concentration [3]. Its merits include the potential to use extremely low catalyst loading and to provide access to the reactivity of otherwise insoluble ions. The centerpiece of asymmetric PTC is tight ion-pairing. Cationic chiral catalysts such as quaternary ammonium-based [4,5], guanidinium-related [6], and phosphonium-related [7] salts have been extensively reported. Gouverneur and co-workers reported the use of neutral urea or thiourea as PTC in the activation of inorganic fluorides [8]. Phosphate anions enable reversed-polarity phase transfer catalysis where the chirality originates from the anionic component of the catalyst [9]. PTC adds another dimension of complexity to understanding the reaction mechanism and in silico prediction, as the reactions could occur at the interface (liquid/liquid or solid/liquid) [10].

Various reviews on or with emphasis on computational organocatalysis include the works from the group of Duarte [11], Krenske [12], and Trujillo [13,14]. Modeling with computational chemistry provides microscopic insight at a level that cannot be otherwise obtained experimentally. With continued advancement in hardware and computational chemistry software, increasingly more realistic and accurate simulations can be performed. For instance, the time taken to compute the energy of a molecule with a given size in the 2008 commentary of Houk and Cheong has substantially reduced due to these advances [15]. From an organocatalysis or PTC developmental point of view, a balance between accuracy and time required is of paramount importance. Most modern organocatalytic and phase transfer (PT) catalytic systems have sufficient complexity (in terms of

mechanism and conformational degree of freedoms) that often render in silico prediction of catalytic performance (e.g., turnover frequency: TOF and level of stereoselectivity [16,17]) impractical relative to empirical or domain expertise-driven experimentation. Practical is a relative term that depends on the computational resources available to the research group. However, it is reasonable to assume that most research groups would not have exclusive access to computing resources equivalent to the one used to generate the open-catalyst project data [18,19]. As such, continued synergetic development in hardware and software is still required to reach a practical in silico catalyst design accessible to the average research group. Pertinent to this point, readers are encouraged to refer to the excellent outlooks on the future of computational chemistry in the perspectives of Grimme and Schreiner [20], Houk and Liu [21], and Sunoj [22].

In this review, we will elaborate on the main points in the succinct overview by Harvery et al. [23], and Funes-Ardoiz and Schoenebeck [24], but in the context of organocatalysis and PTC. We will highlight the challenges and opportunities in computational organocatalysis and PTC from both the reactivity and the selectivity perspective. While we will prioritize more recent work (2018 and beyond), classic examples that are representative of a particular key aspect of computational organocatalysis or PTC will be discussed even if it predates our arbitrary definition of recent. Our approach will be balanced—we will provide examples demonstrating notable advances and caveats based on the literature. Proof-of-concept calculations will be included in this review if they reinforce essential concepts.

1.1. Overview

The flow of this review will follow a typical workflow to study a reaction via the location of key stationary points on the potential energy surface or PES (Figure 1); this is currently the dominant workflow for computational organocatalysis or PTC, as the alternative—to study each elementary step via molecular dynamics (coupled with enhanced sampling) is currently not practical or even intractable in a realistic model. The numbering in Figure 1 corresponds to the sequence in which the mentioned topic will appear in this review. The readers can refer to the next Section 1.2 Content for the general structure of this review.
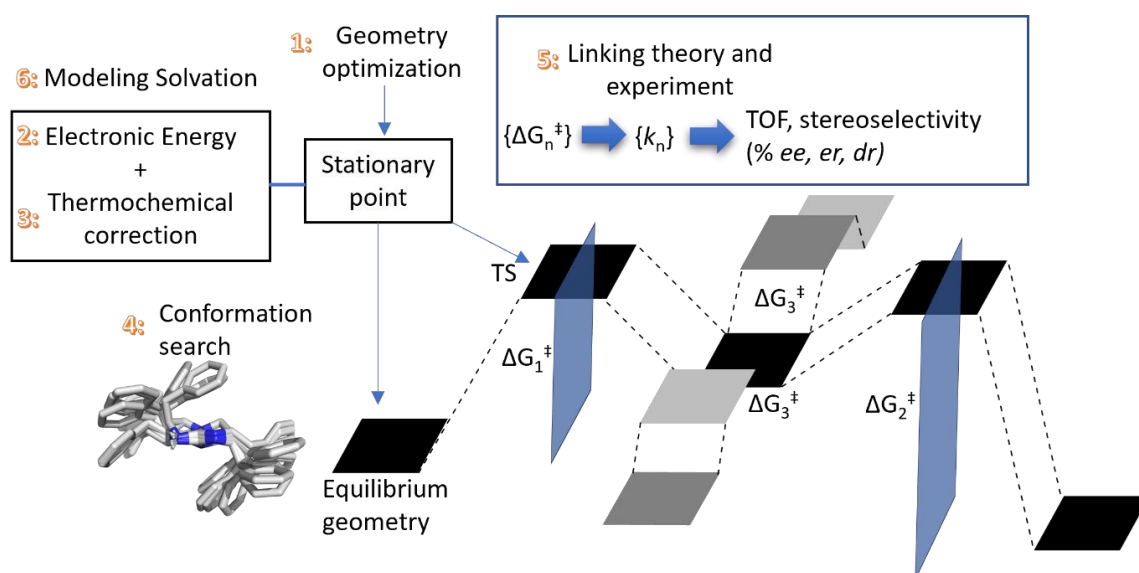
Figure 1. Summary of this review. Workflow to calculate a reaction performance metric from key stationary points on the potential energy surface. Numbers in orange corresponds to the sequence in which the mentioned topics will appear in this review.

1.2. Content

Section 2.1.1. Geometry Optimization: The location of key stationary points requires at least the capability to calculate the energy and gradient (first derivative of energy with respect to coordinates) of a given structure; therefore, we will first discuss the choice of method to obtain these quantities, and its impact on quality based on theoretical benchmarks (1 in Figure 1). Density functional theory (DFT) has emerged as the method of choice to obtain these key stationary points (equilibrium geometries and transition state structures) on the PES.

Section 2.1.2. Compound Method or "Double-Barreling": Upon obtaining these key stationary points, it is a common practice to calculate the electronic energy at a different level of theory (2 in Figure 1). Extensive theoretical benchmarks of density functional approximation (DFA) are based on this concept. We will provide the recommended DFA obtained from a selection of gas-phase benchmark studies. Works that explore the limit of current theoretical methods will be highlighted as they are important for theoretical development. In addition, they caution the applied computational chemists to potential pitfalls. The compound method provides the opportunity to include the influence of solvent via an implicit solvation model. However, due to the challenges associated with modeling the solvent, we will defer a more detailed discussion to a later section.

Section 2.2. Thermochemical Corrections: Standard statistical thermodynamics equations are usually employed to obtain key experimental observables such as Gibbs free energy (3 in Figure 1). Theoretically, the PES surface is not an experimental observable; a physically accurate comparison requires calculating the free energy.

Section 2.3. Conformational Sampling: With the capability to calculate the free energy of a given molecular structure, conformation sampling is discussed with emphasis on its importance and the software available to achieve automated sampling (4 in Figure 1).

Section 2.4. Translating from Calculated Barrier to Rate Constant and Section 2.5. Extending to Catalysis—Turnover Frequency: Once all the relevant stationary points and their respectively thermally accessible conformers are obtained, translation from the energy representation to the rate representation will be required (5 in Figure 1). Transition state theory (TST) bridges the theoretically derived barrier and experimental rate and, consequently, reaction performance metrics such as selectivity and turnover frequency (TOF).

Section 2.6. The Computational Challenge—Mechanism: We will highlight some challenges associated with investigating key stationary points on the PES. For instance, the role of essential additives or reagents may not be obvious, and the potential of aggregation of catalysts (Section 2.6.1 Non-Linear Effect in Asymmetric Catalysis) further complicates the matter. Theoretical considerations will be examined here, such as the assumption of one TS structure leading to multiple products due to dynamic effect (Section 2.6.2 Bifurcating potential energy surface).

Section 2.7. Modeling Solvation: Solvents are central to organocatalysts and PTC. Their influence on the reaction outcome could be potentially game-changing (6 in Figure 1). This section considers the various theoretical approaches to model the solvent's influence. Examples from the literature with emphasis on solvation will be discussed in Section 2.7.1 Examples Related to Solvation.

Section 2.8. Validation of Calculations: With the challenges and background discussion set, the audience would have realized the numerous assumptions in obtaining key stationary points of the catalytic reaction. We will showcase some examples where the models (combination of stationary points and level of theory) are validated by their capability to predict several experimental data points (e.g., kinetic isotope effect and enantiomeric excess).

Section 2.9. Understanding/Interpretation: An analysis to propose the origin of selectivity or reactivity is appealing to experimental chemists, as it could help augment their domain expertise and inspire the development of new catalysts based on an updated chemist's perspective toward the reaction in question.

Section 2.10. Machine Learning and its Relevance to the Field: With the advent of artificial intelligence and machine learning, many opportunities have arisen to tackle the challenge of in silico catalyst design differently. We will highlight examples that map descriptors derived from molecular structures to various performance metrics (Section 2.10.1. Supervised Learning) and 3D structures to potential energy and forces (Section 2.10.2. Machine Learning Potential).

2. Discussion

2.1. Choice of Methods

The method of choice for most computational studies that involve organocatalysis or PTC is density functional theory (DFT). The researcher is immediately faced with choosing which density functional approximation (DFA) to use. Benchmark studies in the literature offer valuable guidance on the performance of various DFA on geometry optimization, reaction barriers, and non-covalent interactions. It should be noted that these benchmarks are usually for calculations in the gas phase. Extrapolation and application to the condensed phase, where organocatalysis and PTC generally occur, requires careful validation by the computational chemist.
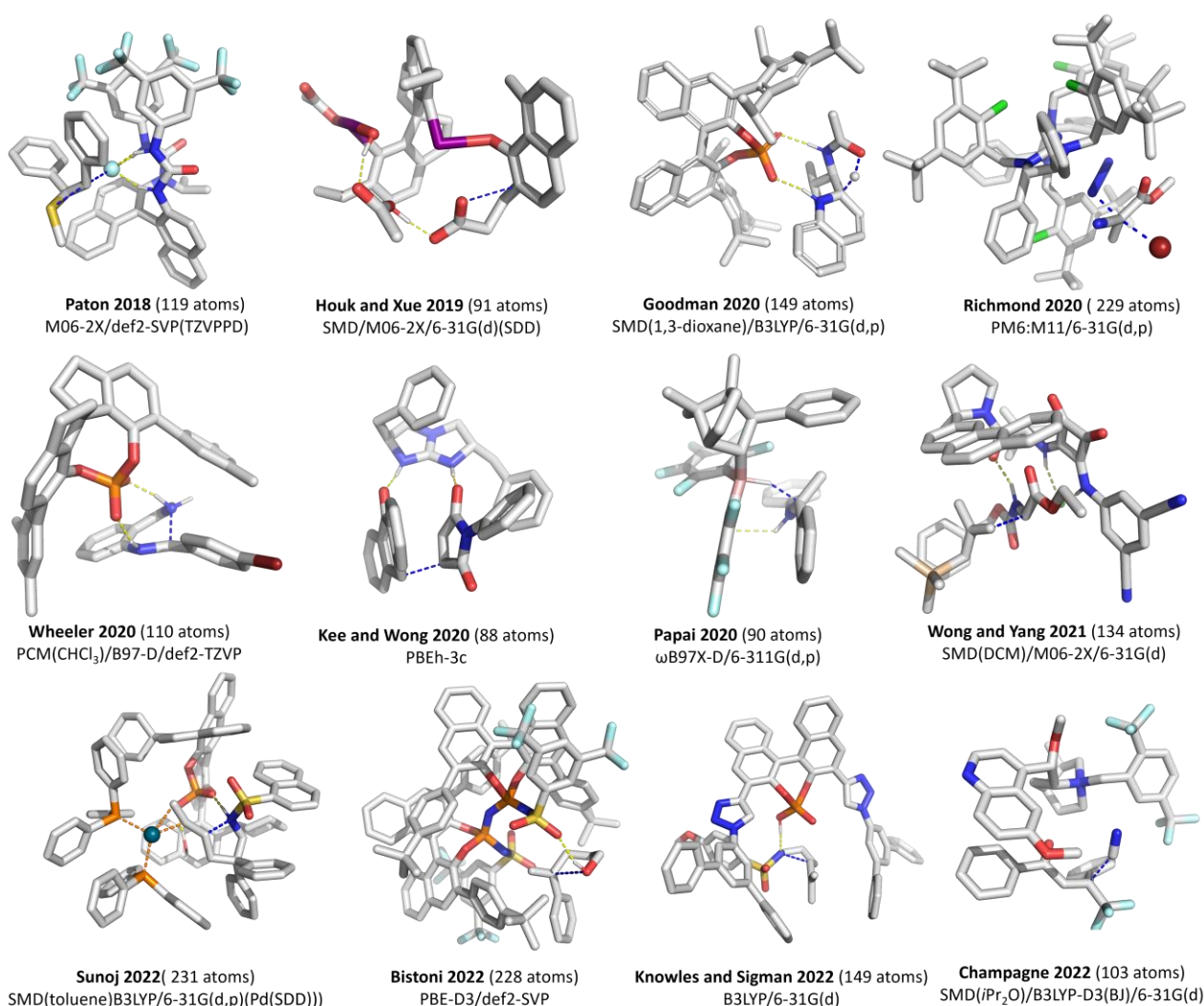
2.1.1. Geometry Optimization

Obtaining a reasonable geometry is the prerequisite for calculating many properties of interest to a chemist. A geometry optimization involves searching the potential energy surface (PES) to find a stationary point. Analysis of the Hessian matrix reveals whether this stationary point corresponds to an equilibrium geometry (no imaginary frequency, reactant molecules, stable intermediate, or complexes in a reaction) or a transition state structure (one imaginary frequency). It forms the basis for studying a reaction mechanism via the location of stationary points on the PES.

Within the single reference wavefunction limit, coupled-cluster singles, doubles, and perturbative triples or CCSD(T) [25] at complete basis set (CBS) is generally accepted as the "gold standard" which can achieve an error of less than 1 kcal/mol (for chemical reaction calculation). However, geometry optimized at this level of theory is rare. Vermeeren et al. optimized key stationary points of five pericyclic reactions at CCSD(T)/cc-pVTZ [26]. They benchmarked the BP86 gradient corrected functional with basis sets of various sizes and the use of empirical dispersion correction. They found that BP86 with a double zeta basis set gave the smallest transitional and rotational projected Root-Mean-Square Deviation (RMSD). A larger basis set worsens the RMSD relative to the CCSD(T) reference, and dispersion correction improvement is inconsistent amongst all the stationary points tested. In the benchmark study of various ionic liquid clusters by Seeger and Izgorodina, ωB97X-D/aVDZ and M06-2X were recommended by the authors based on FMO3-SRS-MP2/cc-pVTZ geometries [27]. However, they emphasized that empirical dispersion correction is crucial.

The location of TS structures is usually the costliest part of a computational mechanistic investigation via the location of stationary points. Multiple exact Hessian calculations are typically needed for the optimization to be successful. DFT is usually the method of choice due to its computational cost and the availability of analytical Hessian, which may be essential in the successful location of TS structures. For relatively large systems such as modern organocatalysts and PTC, the choice of DFA and basis set available is limited. The use of a relatively less expensive generalized gradient approximation (GGA) functional such as PBE, BP86, and B97-D is common. Theoretical benchmark against structure from a higher level of theory is generally infeasible. Instead, the quality is usually assessed by whether experimental results can be reproduced. A non-exhaustive list of DFA/basis set combinations reported in the literature to optimize TS structures for various classes of asymmetric organocatalysts (we included one example of cooperative catalysis by Sunoj and co-workers [28]) is given in Figure 2. The Minnesota functionals and B3LYP appear to be the DFA of choice for different groups. For a very fast elementary step (barriers of a few

kcal/mol), the choice of DFT and basis set can profoundly impact whether the TS could be located. For instance, Paton, Gouverneur, and co-workers found that the fluoride ion must be described with a sufficiently large and diffuse basis set, or a TS structure could not be located [29].

Figure 2. Selected examples of organocatalysts TS structures and their level of theory for geometry optimization. Yellow Dash line: hydrogen bond. Blue Dash line: key bonds in TS. Three-dimensional representations are generated from coordinates given in the cited references: Paton 2018 [29]; Houk and Xue 2019 [30]; Goodman 2020 [31]; Kee and Wong 2020 [32]; Richmond 2020 [33]; Wheeler 2020 [34]; Papai 2020 [35]; Wong and Yang 2021 [36]; Sunoj 2022 [28]; Bistoni 2022 [37]; Champagne 2022 [38]; Knowles and Sigman 2022 [39].

Bistoni and co-workers evaluated the impact of changing the DFA, basis set, and the inclusion of solvent on the RMSD of one of their TS structures (calculated at PBE-D3/def2-SVP) [40]. They found that with the def2-SVP basis set, the TPSSh-D3 and PBE-D3 gave an RMSD of 0.071 (hybrid vs. GGA DFA). The impact of increasing the basis set size—from def2-SVP to def2-TZVP(-f)—is more significant, giving an RMSD of 0.439. At the same time, the inclusion of solvent via the CPCM model only provides an RMSD of 0.043. The impact of these deviations in the predictive capability remained to be determined.

Geometric counterpoise correction by Grimme and co-workers includes an empirical correction to account for the effect of a finite basis set on geometry and vibrational frequencies. For DFT, PBEh-3c [41], R^2SCAN-3c [42], and B97-3c [43] are available for different

needs. Head-Cordon and co-workers proposed the DFT-C scheme to correct the basis set superposition error for the def2-SVPD basis set [44]. These corrections can be applied to a large system where geometry optimization and Hessian calculations with a larger basis set would be computationally too demanding.

The hybrid QMQM or the ONIOM [45] method can substantially speed up the location of key stationary points. However, their performance has not been extensively benchmarked or validated and may likely be system-dependent, although they could serve as starting points for further filtering or refinement of a large ensemble of conformers.

2.1.2. Compound Method or "Double-Barreling"

It is a common practice to perform an additional single-point calculation on the optimized structure with a larger basis set or with an altogether different method, also known as "double-barrelling" or a compound method. The motivation can be to use a larger basis set with the hope of obtaining more accurate electronic energy, to include the solvent effect via an implicit solvation model, or to use a linear-scaling CCSD(T) in the hope of approaching chemical accuracy (<1 kcal/mol).

Various comprehensive benchmark studies based on this concept have provided valuable information on popular density functional approximation (DFA) performance in predicting different properties. Goerigk and co-workers performed extensive DFA benchmarks on the GMTKN55 dataset [46] (a collection of high-quality references for main-group thermochemistry, kinetics, and noncovalent interactions) [47–49]. The comprehensive benchmark by Mardirossian and Head-Gordon is an authoritative reference for the performances of over 200 DFA over a wide range of scenarios (noncovalent interaction, thermochemical data, barrier heights) [50].

Jacob's ladder analogy [51] is used to classify DFA according to improvement from the lowest ladder rung (LDA: local density approximation). In LDA, the electron density is assumed to be uniform throughout space. Its uses in organocatalysis or PTC are expected to be limited. The addition of information on the electron density gradient (gradient Laplacian) brings us to the second rung—generalized gradient approximation (GGA). The non-separable gradient approximation (NGA), in which the exact and correlation are combined, proposed by Peverati and Truhlar, is classified under the second rung [52]. Meta-GGA/NGA at the third rung includes the kinetic energy density on top of the electron density and the density gradient. It represents the highest rung for semi-local DFA and provides good accuracy for equilibrium geometries [53]. The fourth rung is where computational organocatalysis work is generally utilized (see Figure 2). At the fourth rung, Hybrid GGA/ meta-GGA DFA replaces a specific degree of exchange in semi-local DFT with Fock exchange (occupied orbitals information). Finally, virtual orbital information is included at the fifth rung to form the double-hybrid DFA. The high cost of double-hybrid DFA restricts its use in organocatalysis or PTC to mainly single-point calculations.

We compiled a recommendation in Table 1 based on the cited references. The basis set is not listed, but they are generally at least triple-zeta quality. The reader should refer to the original literature for detailed technical information. The emphasis will be on DFAs and their performance in calculating barrier heights. However, the reader should note that accurate reaction thermodynamics is essential to predict metrics such as TOF. A lower value in Table 1 indicates better performance (within the benchmark dataset used).

The more computationally economical GGA generally produces a considerable error in calculating gas-phase barrier height (BH). In general, hybrid DFA has been consistently found to be the best performer in BH, except for B3LYP, which is a piece of common knowledge in the literature. The range-separated hybrid meta-GGA ω97M-V is generally recommended by various studies, as indicated in Table 1. Ho and co-workers' benchmark on ionic S_N2 reaction against canonical CCSD(T)/CBS demonstrates that its performance is impressive even with a small ma-def2-SVP basis set [54]. For the set of three reactions, they found that DSD-PBEP86-D3(BJ) tends to overestimate the barriers while ω97X-V tends to

underestimate them. Their results indicate that the conclusion is reasonably independent of basis set sizes.

Table 1. Selected compilation of recommended DFAs at various rungs of the DFA Jacob's ladder and their associated deviations in gas phase benchmark for barrier heights and noncovalent interactions.

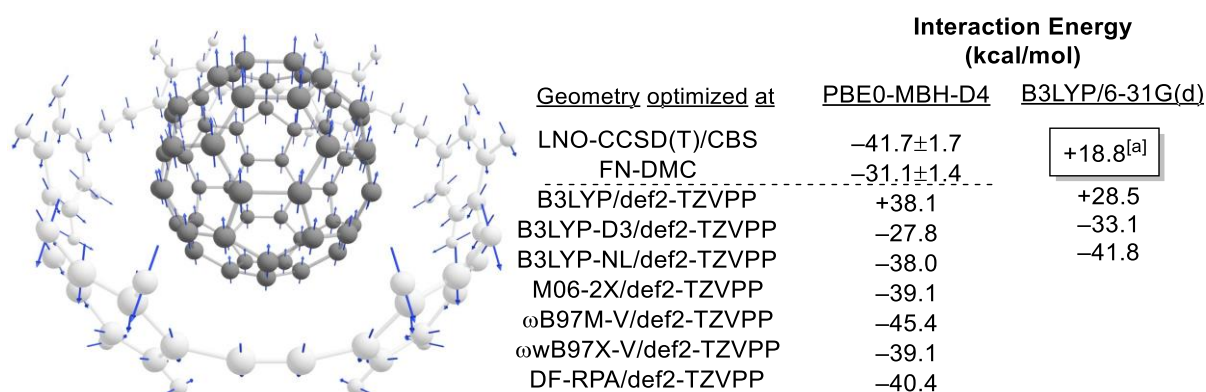| Rung on the Jacob Ladder	DFA	Correction (Recommended by the Ref) [a]	Barrier Heights			GMTKN55
			RMSD [b]	WTMAD-2 [c]	BH9 RMSD [d] (MAE [e])	WTMAD-2
2: GGA/ NGA	B97	D3(BJ) [46]	8.32	13.15	11.27	8.55 [46]
		D3(0) [50]	8.65	NA	NA	NA
	revPBE	D3(BJ) [46]	8.30	15.79	11.24	8.27 [46]
		D3(0) [50]	8.26	NA	NA	NA
	BLYP	D3(0) [27,50]	10.13	NA	NA	NA
		D3(BJ) [46]	9.91	NA	12.10	9.51
3: Meta-GGA/ NGA	SCAN	D3(BJ) [46]	7.83	14.94	NA	7.86 [46]
	r^{2}SCAN	None	NA	NA	7.90	NA
		D3BJ	NA	NA	9.25	NA
		3c [42]	NA	NA	NA	7.5 [42]
	revTPSS	D3(BJ) [46]	NA	15.78	NA	8.50 [46]
	M06L	D3(0) [46,50]	6.84	7.56	NA	8.61 [46]
	B97M	V [48,50]	4.35	7.53	6.70 (4.14)	5.46 [48]
4: Hybrid GGA/Meta-GGA	ωB97M	V [48,50,54–56]	1.68	3.40 [f]	(2.08)	3.53 [48]
	ωB97X	V [46,50,56]	2.44	4.21 [f]	(3.20)	3.98 [46]
		D3(0) [46,50]	2.28	4.67	NA	4.61 [46]
	M06-2X	None [56]	NA	NA	(2.27)	NA
		D3(0) [26,50]	2.60	5.60	NA	4.94 [46]
	B3LYP-D3	D3(0)	5.92	NA	NA	NA
		D3BJ	NA	NA	6.77	6.42 [46]
5: Double-Hybrid	DSD-PBEP86	D3(BJ) [46]	NA	3.52	4.04	3.14 [46]
		NL [48]	NA	3.25	NA	2.84 [48]
	revDSD-PBEP86	D3(BJ) [26,55]	NA	NA	2.96	2.42 [55]
	ωB97X-2	D3(BJ) [48]	NA	3.25	NA	2.97 [48]
	ωB97M(2)	None [55]	NA	NA	NA	2.19 [55]
	DSD-BLYP	D3(BJ) [46]	NA	3.04	NA	3.08 [46]
		NL [48]	NA	2.86	NA	3.05 [48]
	B2GPPLYP	D3(BJ) [46]	NA	3.24	NA	3.26 [46]

NA: Not Available. [a] Empirical correction (mostly dispersion). References that recommend this DFA + correction combination are indicated by the citation(s). [b] RMSD: Root Mean Squared Deviation in kcal/mol from ref [50]. [c] WTMAD-2: Weighted Total Mean Absolute Deviation 2 in kcal/mol. See ref [46] for the definition. [d] RMSD based on the entire BH9 dataset [56]. Values in kcal/mol from ref [57]. [e] Mean Absolute Error of the entire BH9 dataset. Values in kcal/mol from ref [56]. [f] post-SCF version of non-local correction.

Double hybrid functional, which currently represents the highest rung of the DFT Jacob's ladder, is generally the most reliable from gas phase benchmark results. Goergik and coworkers reported a double hybrid benchmark of the GMTKN55 dataset [58]. The group of Martin has performed extensive work on double hybrid DFA [55,59,60]. They have reported the impact of basis set size for double-hybrid DFA and the use of F12 explicit correlation to accelerate basis set convergence [61]. For double-hybrid DFA, a non-local post-SCF correction was beneficial to both barrier heights and the overall noncovalent interaction dataset in the GMTKN55 benchmark [48].

The domain-based local pair natural orbital (DLPNO [62,63]) and the localized molecular orbital (LMO [64,65]) approximations, implemented in ORCA [66] and MRCC [67], respectively, have enabled single-point calculation to be performed on a molecule with around up to a few thousand atoms (or tens of thousands of basis functions) [65,68]. These generally approximate the canonical CCSD(T) result for large molecules in the benchmark dataset described above. Bistoni and co-workers have reported the DLPNO-CCSD(T)/def2-TZVP single-point calculations for two computational studies on organocatalysis [37,40].

The accuracy of DLPNO and LNO approximations are benchmarked against a small system where the corresponding canonical CCSD(T) calculation is tractable. For DLPNO-CCSD(T), its accuracy is dictated by three main cutoffs to determine the molecular orbital space that will be included. For the recommendation of cutoff choice and its impact, the reader is referred to the work of Neese and co-workers [69] and Seeger and Izgorodina (on ionic liquids) [70]. In their ionic S_N2 computational study, Ho and co-workers presented a comparison of canonical and DLPNO CCSD(T) with various basis set sizes [54]. For related considerations, and different schemes that utilized DLPNO to approximate the canonical result, the readers are encouraged to refer to the references cited in this paragraph and the references cited within them.

Brandenburg, Tkatchenko, and co-workers reported that for large molecules (up to 132 atoms), the disagreement between LNO-CCSD(T)/CBS and the quantum diffusion Monte Carlo (DMC)-calculated interaction energy can be of the order of 7.6 kcal/mol (Figure 3) [71].

	Interaction Energy (kcal/mol)	
Geometry optimized at	PBE0-MBH-D4	B3LYP/6-31G(d)
LNO-CCSD(T)/CBS	−41.7±1.7	+18.8[a]
FN-DMC	−31.1±1.4	
B3LYP/def2-TZVPP	+38.1	+28.5
B3LYP-D3/def2-TZVPP	−27.8	−33.1
B3LYP-NL/def2-TZVPP	−38.0	−41.8
M06-2X/def2-TZVPP	−39.1	
ωB97M-V/def2-TZVPP	−45.4	
ωwB97X-V/def2-TZVPP	−39.1	
DF-RPA/def2-TZVPP	−40.4	

Figure 3. All values are in kcal/mol. LNO-CCSD(T) and FN-DMC values are taken from ref [71]. Values below the dotted line are calculated with ORCA 5.0.3 [66] using the following keywords: "DFA/def2-TZVPP autoaux verytightscf rijcosx" with the geometry supplied by ref [71]. DF-RPA is calculated with MRCC [67]. Blue arrows are the displacement vectors of the imaginary frequency at B3LYP/6-31G(d). [a] Interaction energy at the same level of theory as the geometry optimization, i.e., at B3LYP/6-31G(d).

Based on the structures provided by Brandenburg, Tkatchenko, and co-workers, we evaluated the interaction energies from a selection of recommended DFA, as given in Table 1. It is reasonable to assume that dispersion interaction will dominate the interaction between C60 and 6-CPPA. The significance of dispersion is especially evident in the B3LYP/def2-TZVPP calculations in Figure 3. Without any empirical dispersion correction, the interaction between C_{60} and 6-CPPA is repulsive. The choice of dispersion correction will significantly determine if the result is closer to the LNO-CCSD(T) or the FN-DMC values. M06-2X, ωB97X-V, and DF-RPA are the closest to the LNO-CCSD(T) values. In the following paragraph, we emphasize that it is not an exhaustive study but a quick demonstration that may have significance in the double-barreling discussion. The problem of approximate solution to CCSD(T) of a large system where dispersion dominated the system (which is very well described) has been pointed out by the studies of Herbert and co-workers [72]

and Patel and Wilson [73]. Lastly, Hobza and co-workers pointed out that CCSD[T] is superior to CCSD(T) in describing noncovalent interaction [74].

The caveat on double-barrelling has been discussed by Kevin, where he analyzed the error introduced in double-barrelling by varying the bond lengths of small molecules about their equilibrium positions [75]. The C_{60} and 6-CPPA host–guest complex will be used to illustrate this point further. With B3LYP/6-31G(d) optimized geometry, the complex is a transition state (TS) structure which corresponds to the expulsion of C_{60} from the complex (ref to the blue arrows in Figure 3). Single-point calculations with empirical dispersion correction indicate that their interaction is attractive, which contradicts it being a TS that expels C_{60} from 6-CPPA.

Pertinent to organocatalysis, Chin and Krenske highlighted a similar problem of double-barrelling in their highly detailed computational study of chiral phosphoric acid (CPA)-catalyzed asymmetric Nazarov reaction. The calculated % *ee* varies from −24% to +94% with various modern DFA, basis set, and implicit solvation models [76]. Despite this and the challenges highlighted before, we wish to emphasize that a successful prediction of experimental enantioselectivity is unlikely not to have an element of error cancellation to various and often unknown extents. This brings us to testing the computational methodology against diverse experiment data points to increase confidence in the calculated results (see Section 2.8 Validation of Calculations).

2.2. Thermochemical Corrections

Under most organocatalytic reaction conditions, the most relevant quantity to predict reaction rates (consequently yield) and selectivity (enantio- or diastereo-) is Gibbs free energy, as most reactions are performed under constant pressure (usually atmospheric).

The statistical thermodynamics approach is then used to calculate the correction for Gibbs free energy, which is then added to the electronic energy to obtain the Gibbs free energy of a molecule. The assumptions in deriving the equations apply—ideal gas, rigid rotor, and standard pressure. The impact of approximating low vibrational frequencies as a rigid rotor is well documented. Various corrections are available to handle the correction of these low vibrational frequencies [77–79]. The open-source PYTHON code GoodVibes from Paton's group allows different schemes of correction to the rigid rotor to be applied as a post-processing step [80,81].

Plata and Singleton highlighted the importance of various important but widely ignored corrections to the entropy calculated with standard statistical thermodynamic equations [82]. These include residual entropy due to symmetry and entropy of mixing.

Thermochemical correction when solvation is involved is an area of extensive debate. The controversy lies in how to handle translational and rotational entropy. The approximation implemented in most electronic structure packages is the ideal gas/rigid rotor/harmonic oscillator (RRHO) approach. Some authors have shown that the RRHO can be used without correction in selected cases [83–86]. In contrast, others recommend various corrections (the reader is referred to the work of Besora et al. [87] for details of all these corrections). These corrections are evaluated by Besora et al. by comparing values derived from umbrella sampling with explicit solvation (molecules of solvation are fully represented at the atomic level) [87]. Jacobsen and co-workers tested RRHO and two quasi-RRHO schemes based on calculated levels of enantioselectivity vs. experimental ones (for more detail, please refer to Section 2.8 Validation of Calculations) [88].

2.3. Conformational Sampling

Except in the extreme cases where the molecules or catalysts have minimal flexibility (or rotatable bonds), a low-quality ensemble of conformers can result in biased results. In calculating the contribution of competing pathways (Blue vs. Red in Figure 4), the Curtin–Hammett principle is usually assumed to be valid—the relative free energies of the appropriate ensembles of TS structures can be used for prediction contribution from each competing pathway. With this assumption fulfilled, one can utilize the Boltzmann

distribution to calculate a selectivity metric. Selectivity, in this case, can be relative product distribution, diastereoselectivity, or enantioselectivity. For enantioselectivity, the tutorial review by Paton and co-workers provides the relevant details and equations to perform this calculation [16]; the same concept can be used for other forms of selectivity listed.
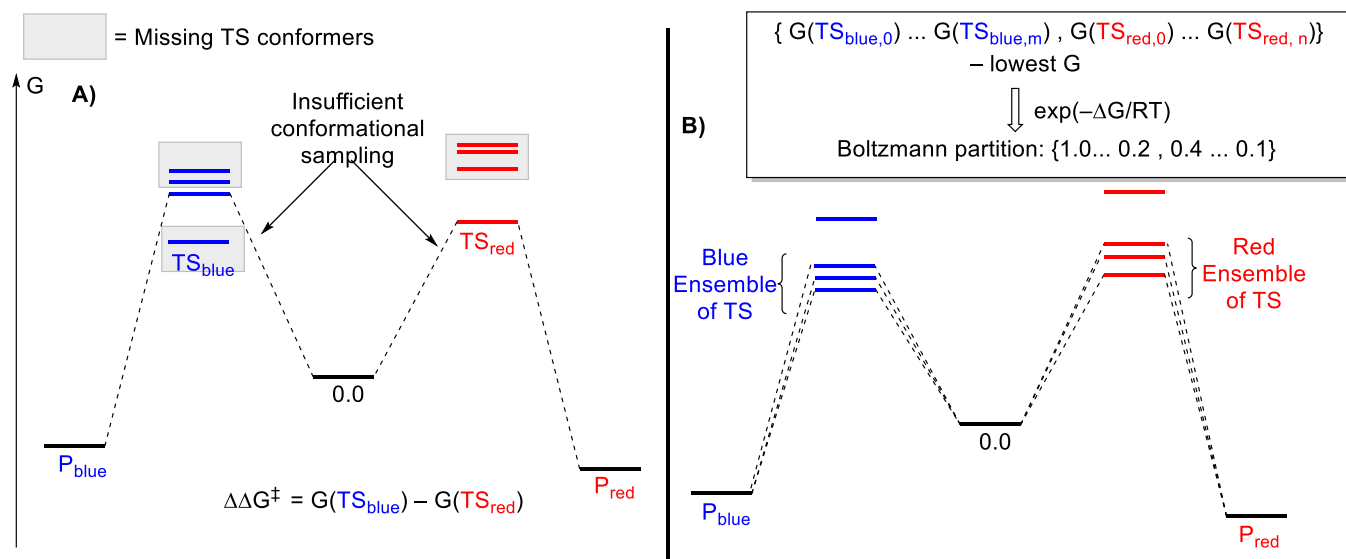
Figure 4. Hypothetical scenario (**A**) The lowest-lying TS_{blue} conformer is omitted. The opposite selectivity is thus predicted. (**B**) Many conformers with significant populations at the reaction temperature and pressure. An ensemble is required to obtain a more accurate calculated value of selectivity.

Figure 4A shows a hypothetical situation where only one TS structure for each pathway was located. This is an extreme case, in which for each pathway, it is sufficient to consider only one TS as the other conformations' contribution is negligible due to their high energies. However, in this scenario, the lowest energy TS_{blue} is not located. Therefore, P_{red} is predicted to be the primary product. There is a possibility that a lack of conformation sampling can lead to a fortuitous agreement with the experimental result. However, extensive validation with multiple experimental outcomes is expected to allude to this deficiency. In Figure 4B, an ensemble of TS structures must be considered for each pathway, as many conformations contribute significantly to product formation. The standard procedure is to calculate the product selectivity based on a Boltzmann population derived from the relative energies of the two combined ensembles.

As discussed above, critical conformations (those with significant Boltzmann populations at the reaction temperature and pressure) must be located. This step is usually done automatically by using conformational sampling software. The conformational space search algorithm and the accuracy of the Hamiltonian employed to evaluate the energies during the search are two critical aspects for a successful conformation sampling. However, many structures are usually generated during the search, and each must be optimized. Therefore, DFT-based methods are generally infeasible. The use of classical force field is common but risky, as they are usually not trained on the system of interest; it risks missing key conformations if an energy-based cut-off is employed. As a compromise, semiempirical methods or potential Machine Learning Potential (vide infra) can be used. The open-source CREST, coupled with the GFN2-XTB tight-binding Hamiltonian, is a readily available and powerful tool for this purpose [89]. Its non-covalent interaction sampling mode is particularly suited for sampling TS structures, as key internal coordinates of the TS can be restrained during the sampling.

For an evaluation of various conformational sampling methods in quaternary ammonium-based PTC and the application of unsupervised machine learning techniques for analysis of the results, the reader is referred to the work of Iribarren and Trujillo [90].

2.4. Translating from Calculated Barrier to Rate Constant

In a chemical reaction with one elementary step, such as cycloaddition, and nucleophilic substitution, the rate constant and the concentration(s) of the reactant, at any point in time, determines the rate of product formation. If a pre-equilibrium kinetic model is assumed—an equilibrium exists between the reactive conformation (or RC: reactant complex) and the reactant(s) in their most stable state—then the effective rate constant (kK$_{pre-equi}$) can be obtained from the free energies of the TS and reactants. Our work in the in silico design of halogen bond-based catalysts uses this model to compare with the predicted TOF of our proposed catalysts [91].

The conventional transition state theory (TST) links theoretical activation energy ($\Delta G^{\ddagger}$ in Figure 5) and the experimental rate constant. A more sophisticated form of the TST, such as the variational TST, is expected to improve the accuracy of the rate constants [92]. However, their required computational costs are generally intractable for system sizes found in most organocatalysis.
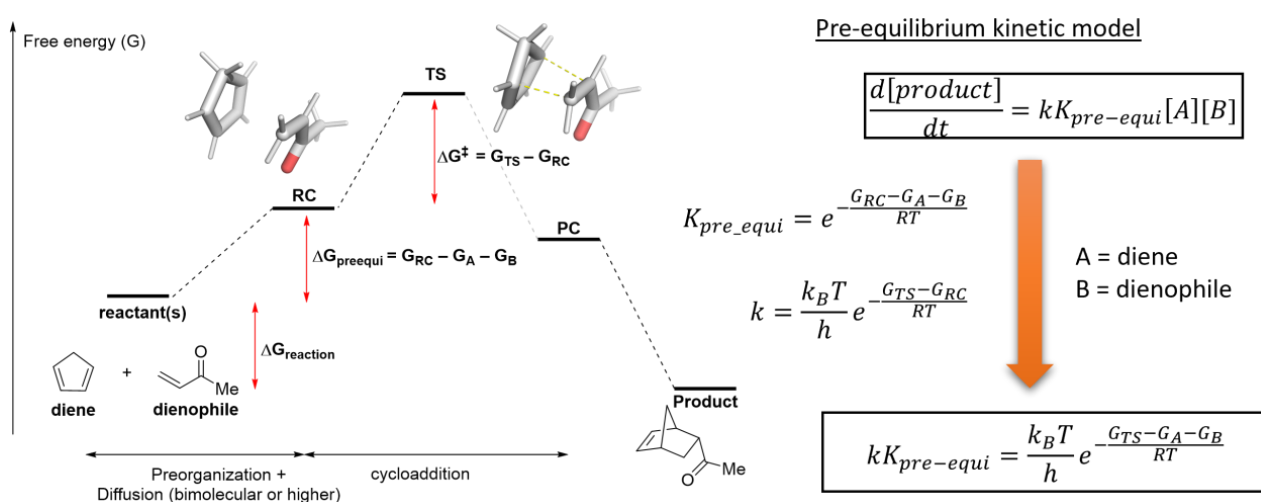
Figure 5. The reaction profile of a [4 + 2] cycloaddition. RC: Reactant Complex. PC: Product Complex. G is the free energy of the species labeled in the subscript. Reprinted with permission from ref [91]. Copyright 2016 American Chemical Society.

Tunneling, a quantum mechanical phenomenon, can result in a larger rate constant than predicted from the barrier height. The semiclassical Wigner correction provides a simple means to determine the importance of tunneling (Equation (1)) [93]. However, Truhlar and co-workers recommended using this correction if κ(T) does not exceed 1.2 due to the various approximations involved [94]. The open-source pyQuiver can calculate tunneling correction via the Wigner and Bell inverted parabola from a cartesian Hessian matrix [95].

$$\kappa(T) = 1 + \frac{1}{24}\left|\frac{h\omega^{\ddagger}}{2\pi k_B T}\right|^2 \tag{1}$$

Experimentally, the involvement of quantum tunneling effect has been reported in the seminal work of Schreiner and co-workers where they introduced the concept of a tunneling-controlled reaction [96], in addition to the commonly used kinetic- and thermodynamic-controlled reaction [97]. More sophisticated tunneling correction exists, but sophistication generally comes with an increasing computational cost [98].

Diffusion-controlled rate constant approximation, such as the equilibrium between reactant(s) and RC, can be calculated with Equation (2). r_A and r_B are radii of the reactant A and B, respectively, which can be estimated from an intrinsic solvation model calculation. The procedure to calculate W, D_A, and D_B is detailed in the work of Pliego and co-workers [99].

$$k_{diffusion} = \frac{4\pi(r_A + r_B)(D_A + D_B)W}{e^W - 1} \tag{2}$$

2.5. Extending to Catalysis—Turnover Frequency

In catalysis with a multi-elementary steps cycle with potential non-productive branches, turnover frequency (TOF) is frequently employed to measure the efficiency and stability of a catalyst. A decreasing TOF over time indicates catalyst deactivation as the reaction progresses. The prediction of TOF will be more challenging as the structure difference between the transition state and reactant(s) would translate to vastly different electronic structures and correspondingly less favorable error cancellation. Bearing in mind the exponential relationship between rate constants and barrier height, a 1 kcal/mol difference (upper bound of chemical accuracy) may translate to up to an order of magnitude difference in TOF (Table 2). For catalyst design, however, one might be more concerned with relative TOF between different catalysts than the absolute TOF of a single catalyst.
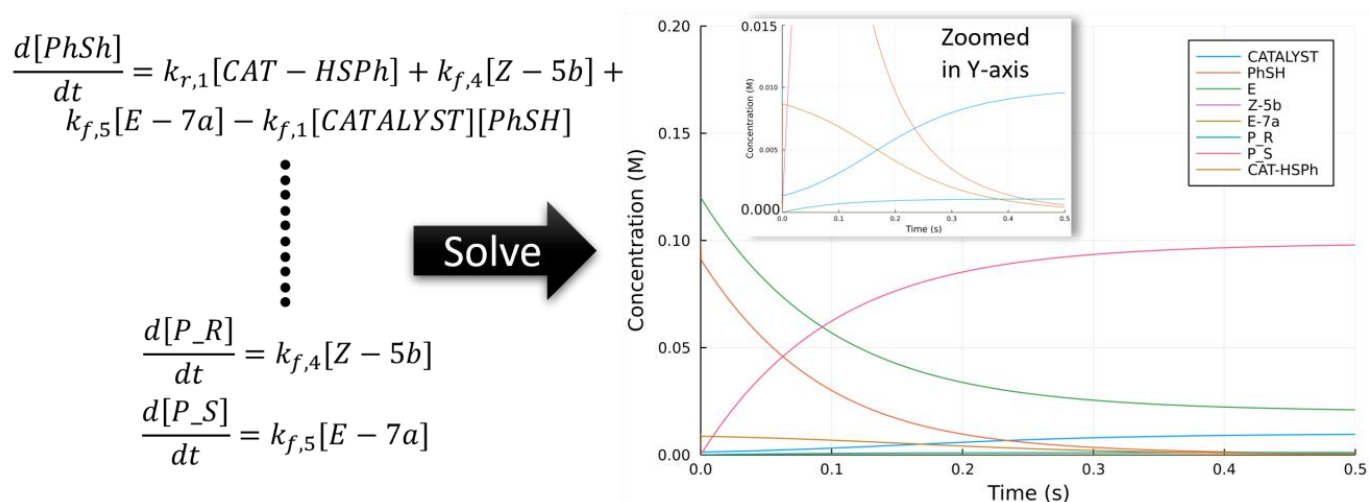
Table 2. Energetic span and TOF at 223.15 K and 298.15 K.

Temperature	Energetic Span (kcal/mol) *	TOF (h^{-1}) *
	20	49
298.15	21	9.1
	22	1.67
	15	34
223.15	16	3.58
	17	0.38

* Calculated with AUTOF_Excel_v09_1 [100].

A catalytic cycle is a sequence of interconnecting elemental steps described by a set of coupled ordinary differential equations (ODE). Obtaining an analytical expression is usually not possible for complicated catalytic cycles. These ODE can be solved numerically to high precision with various ODE solvers (Figure 6). However, when the rate constants are vastly different in magnitudes, the time step for the numerical integration must be sufficiently small to ensure the stability of the simulation. As a result, the time scale that can be transverse could be in the picosecond region or less. This is highlighted and tackled in the rate constants matrix contraction method developed by Sumiya and Maeda [101]. The energetic span model by Kozuch and Shaik uses the steady-state approximation to simplify the situation. A recent extension (gTOFfee) by Garay-Ruiz and Bo enables the energetic span to be calculated for a reaction network [102].

An illustration of the various methods to translate a reaction with two interconnected cycles (commonly occurring for asymmetric organocatalysis) with the work of Wong and co-workers is given in Figure 7 [103]. The authors used Kozuch and Shaik's energetic span model (ESM) to calculate the TOF separately for the major *R*- and *S*-pathway and proposed, based on the TOF, that the *S*-pathway is favored. We note here that the model by Kozuch and Shaik does not consider coupled reaction pathways, as depicted in Figure 7. Both pathways share a common intermediate; thus, they will complete with the shared pool of CAT-HSPh. The impact of each competing elementary step captured by solving the system of ordinary coupled differential equations (ODE) numerically is similar to that derived by calculating the TOF for each pathway separately.

$$\frac{d[PhSh]}{dt} = k_{r,1}[CAT - HSPh] + k_{f,4}[Z - 5b] + k_{f,5}[E - 7a] - k_{f,1}[CATALYST][PhSH]$$

$$\vdots$$

$$\frac{d[P\_R]}{dt} = k_{f,4}[Z - 5b]$$

$$\frac{d[P\_S]}{dt} = k_{f,5}[E - 7a]$$

Solve →

Figure 6. Solving a system of coupled ODE to obtain the concentration of all species over time (only selected equations are shown). The dotted line represented omitted equations.
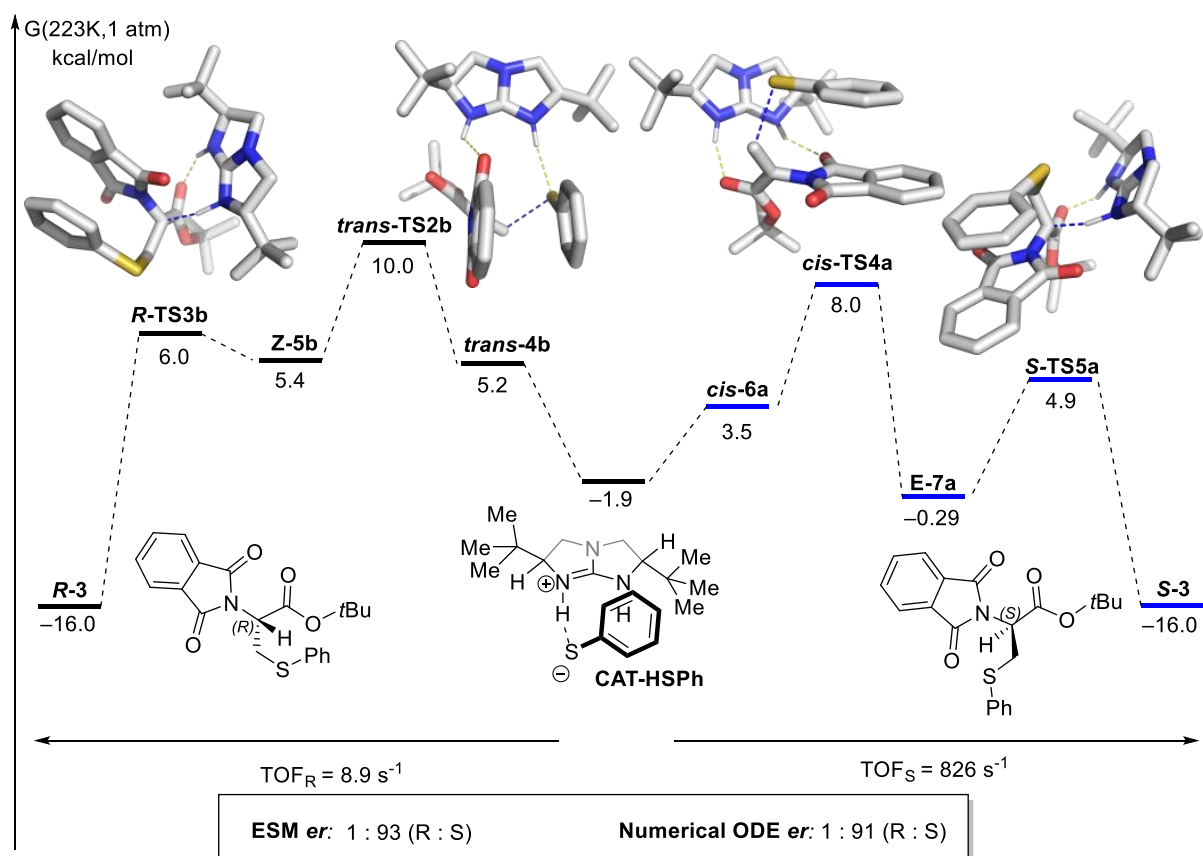
Figure 7. Reprinted with permission from ref [103]. Copyright 2012 American Chemical Society. TOF is calculated with AutoTOF from values supplied by the cited reference. Numerical ODE values are generated with DifferentialEquations.jl [104] and Arbnumerics modules in Julia [105]. The product formation is assumed to be irreversible. A pre-equilibrium kinetic model is used to obtain the rate constants from CAT−HSPh.

Translating computed quantities from the potential energy surface to experimental observables can be challenging and may lead to erroneous conclusions. This is especially important in cases where the solvent molecules are considered part of the reactive species and in PTC where insoluble solids are involved. Some examples will be highlighted below.

In the neutral hydrolysis of ester, Wang and Cao proposed the involvement of six water molecules (excluding the reactive water, a total of seven altogether) to lower the barrier of methyl formate hydrolysis [106]. Pliego and co-workers pointed out that the barrier is computed with the RC as the reference (a rigid cluster of methyl formate and seven water molecules, see Figure 5 for an explanation of RC) [99]. This neglected the penalty incurred from preorganization (Figure 5). However, in bulk water, water molecules are expected to be present as hydrogen-bonded oligomers [107]; therefore, taking the reference point as seven isolated water monomers, as proposed by Pliego and co-workers, may not be the most accurate representation. More discussion on this point will be presented in the "2.7.1. Examples Related to Solvation" Section.

One of the challenges in computing a realistic reaction profile involves obtaining the free energies of the reactants (such as G_A and G_B in Figure 5). The starting points (reactants, reagents, and catalysts) are the reference from which the relative free energies of all critical intermediates and TS are calculated. Thus, it is essential to compute key metrics such as TOF or experimental rate constants.

In PTC, inorganic salts (fluoride, carbonates, hydroxides of group 1 and 2 metals) are frequently employed. The energy calculated for an isolated molecule of these inorganic salts is unlikely to be representative of the bulk. As such, some groups have reported a hybrid approach where experimental data for these inorganic salts are combined with calculations in a thermodynamic cycle.

Pliego and Riveros corrected their crown ether-catalyzed nucleophilic substitution of alkyl bromide with KF with experimental standard free energies of formation for KF and KBr [108]. In their reaction profile, the initial steps are omitted (complexation of KF to 18-crown-6); instead, the reaction profile begins with the exchange of the crown ether complex of KBr (18-C-6-KBr) with 18-C-6-KF (Figure 8).

Another example by Carvalho and Pliego Jr is the nucleophilic fluorination under PTC with crown ether, where they considered the solvation of KF and KBr in toluene [109]. Gouverneur, Paton, and co-workers applied the same concept to obtain a more realistic reaction profile in their asymmetric nucleophilic fluorination under PTC. However, CsF and CsBr are used instead of KF and KBr [29].

Theoretical calculations of solid-state-related quantities will add another dimension of challenge to modeling PTC via computational chemistry. Using such a thermodynamic cycle is practical when the relevant experimental data is available. However, for many salts in PTC, such information is often absent.

2.6. The Computational Challenge—Mechanism

Prediction of experimental performance metrics such as TOF or selectivity for a complicated reaction is immensely challenging. Both the significant productive (leading to desired product) and non-productive (leading to side-products) pathways must be included. Hints from experiments (product distribution, kinetics, and spectroscopy) are frequently employed to reduce the number of possibilities and render computation more tractable.

Nevertheless, strictly in silico catalyst design would require no input from experiments except for the final testing of the de novo designed catalyst. Algorithms are available to explore various reaction pathways automatically [110–114]. However, their widespread application in organocatalysis and PTC is still impeded by the availability of sufficiently accurate methods to calculate reaction barriers for realistic systems within a practical timeframe (see Table 3 for an estimate of the relative computational time for selected DFA), unless, of course, one has access to a tremendous amount of computing resources.
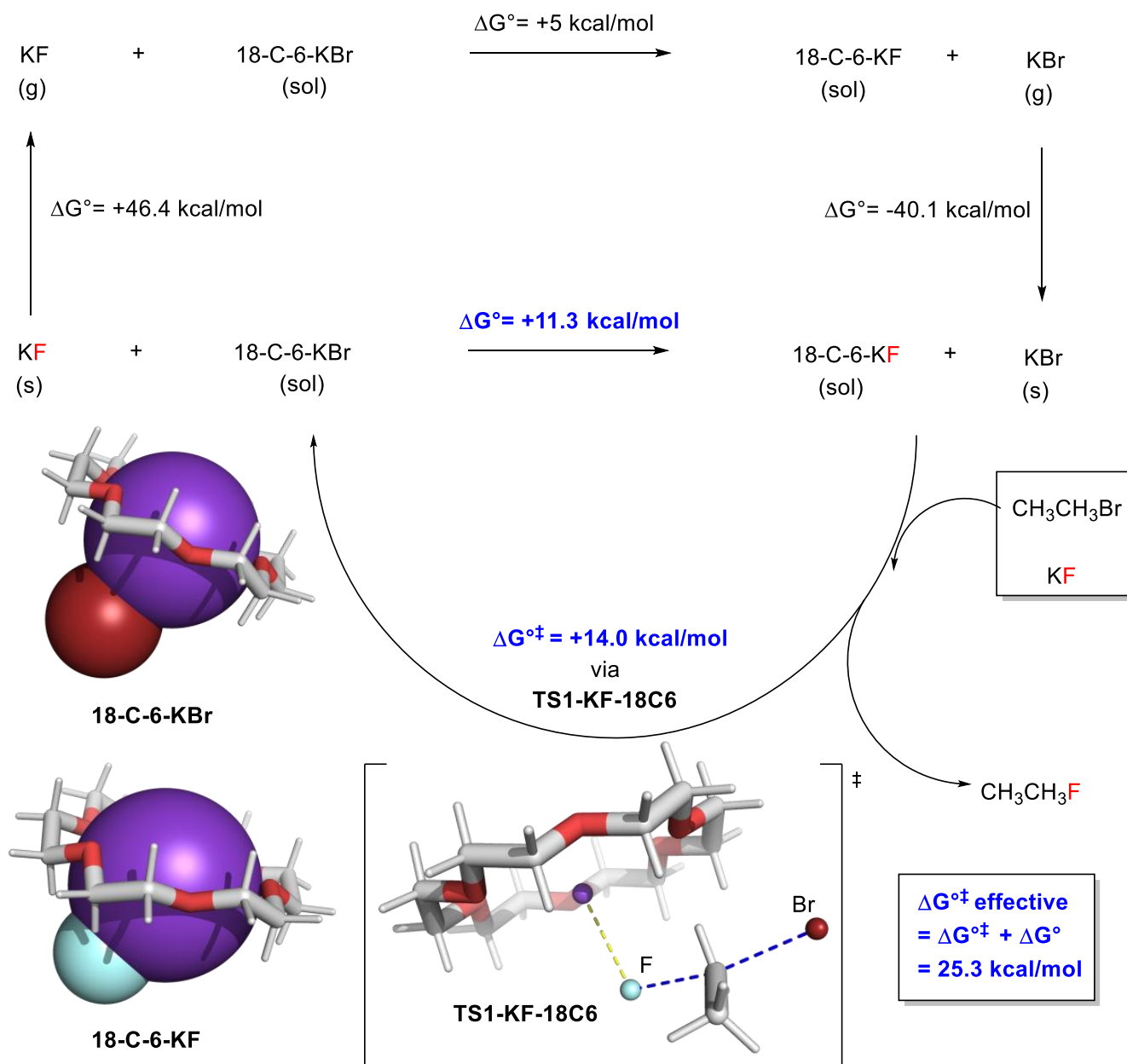
Figure 8. Use of thermodynamic cycle in 18-crown-6-catalyzed fluorination of alkyl bromide. All values are obtained from the work of Pliego Jr and Riveros in ref [108]. Fluoride is colored red to facilitate identification. (sol) refers to solution. (s) refers to solid.
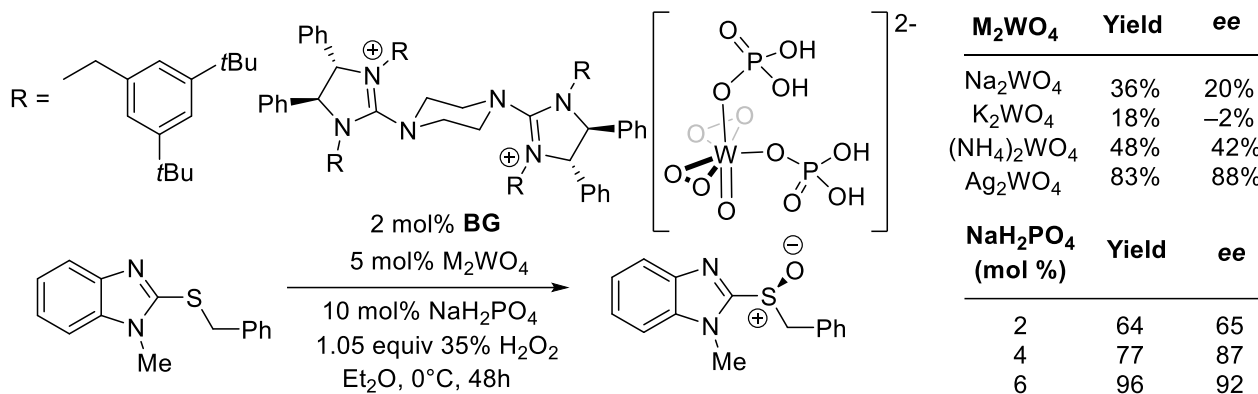
For asymmetric organocatalysis or PTC, the species involved in the critical enantioselectivity determining step must be known. Before any prediction of the level of enantioselectivity, the elementary step in which the chiral centers are formed must be known. In some cases, this might not be obvious. For instance, enantioselectivity is often improved using additives whose role is not immediately apparent. Nonetheless, it affects the level of enantioselectivity strongly suggests its involvement in the stereo-determining TS structures.

Table 3. The relative computational cost of selected DFA from Table 1.

| DFA	Relative Cost of Energy Calculation [a]*	Relative Cost of Gradient Calculation *
revPBE	1.0	1.0
r²SCAN	1.1	1.0
M06-2X	5.5	4.5
ωB97M-V	3.3	3.0 [b]
revDSD-PBEP86	360 [c]	510 [c]

TS_SS_2 from ref [32] is used. * All calculations are performed with ORCA 5.0.3 [66]. Variation is expected depending on system size and hardware setup. The following simple keywords: "def2-TZVPP autoaux verytightscf" are used. [a] Relative cost is estimated from the total SCF time reported by ORCA 5.0.3 for the SCF cycle part and divided by the number of iterations taken to converge the SCF calculation. This is normalized with the DFA with the minimum time taken—revPBE. [b] Including the time taken to calculate the non-local dispersion correction gradient. [c] Including DLPNO-MP2 calculation time.

In PTC, the counter-cation of the bases or reagents, in some cases, has a profound influence on the reactivity, enantioselectivity, or both. For instance, in the *bis*-guanidium (BG) catalyzed enantioselective oxidation of sulfide, the counter-cation (Na, K, NH$_4$, or Ag) can have a profound effect on the sulfoxide yield and *ee* (Figure 9) [115]. This strongly suggests the involvement of the cation in the stereo-center forming TS structures. The loading of the NaH$_2$PO$_4$ was found to affect both the yield and *ee*. The authors proposed that two H$_2$PO$_4$ anions coordinate to a peroxo-tungstate species. This is supported by comparing the computed key vibration modes of the purposed species with the measured Raman spectrum. The roles of the cations in the critical TS structures and their interplay with H$_2$PO$_4$ remain to be elucidated.

M$_2$WO$_4$	Yield	ee
Na$_2$WO$_4$	36%	20%
K$_2$WO$_4$	18%	−2%
(NH$_4$)$_2$WO$_4$	48%	42%
Ag$_2$WO$_4$	83%	88%

NaH$_2$PO$_4$ (mol %)	Yield	ee
2	64	65
4	77	87
6	96	92

Figure 9. Enantioselective oxidation of sulfide reported by Tan and co-workers [115]. All values are obtained from the mentioned reference.

Gouverneur, Claridge, and co-workers reported a detailed nuclear magnetic resonance (NMR) study on fluoride binding to their BINAM-derived *bis*-urea catalysts [116]. In their work, they unravel the interaction between fluoride and the catalyst. Under conditions closer to the reaction setting (CsF in CD$_2$Cl$_2$), they proposed two fluoride complexes mainly from ^{19}F chemical shifts, T$_1$ relaxation, and proton diffusion experiments. The major species, which is a 1:1 fluoride *bis*-urea complex ([*bis*-ureaCsF]), is proposed to be in rapid equilibrium relative to the NMR time scale with minor 1:2 fluoride *bis*-urea complex ([(*bis*-urea)$_2$CsF]). The mixture of these two species at equilibrium proved competent in producing the same enantioenriched product when the substrate is added. The molecular structures and the role these two species may play in the reaction provide further opportunities for computational work.

In another highly detailed mechanistic study from the same group, Wong et al. reported their work on azidation via hydrogen bonding PTC (Figure 10B) [117]. Kinetic modeling with in situ infrared measurement of the product concentration was performed. Two parallel pathways were proposed based on the measured data—a racemic uncatalyzed pathway and an enantioselective catalyzed one. Relevant constants related to the rate of these two pathways were obtained. The catalytic cycle involves the exchange of ions between different species. Solid NaCl was proposed to be an intricate part of the catalytic cycle, where it plays an inhibitory role. DFT calculations, together with extensive conformational sampling with *bis*-urea and azide, gave a calculated % *ee* of 32, which is in qualitative agreement with the experimental value (62% *ee*).
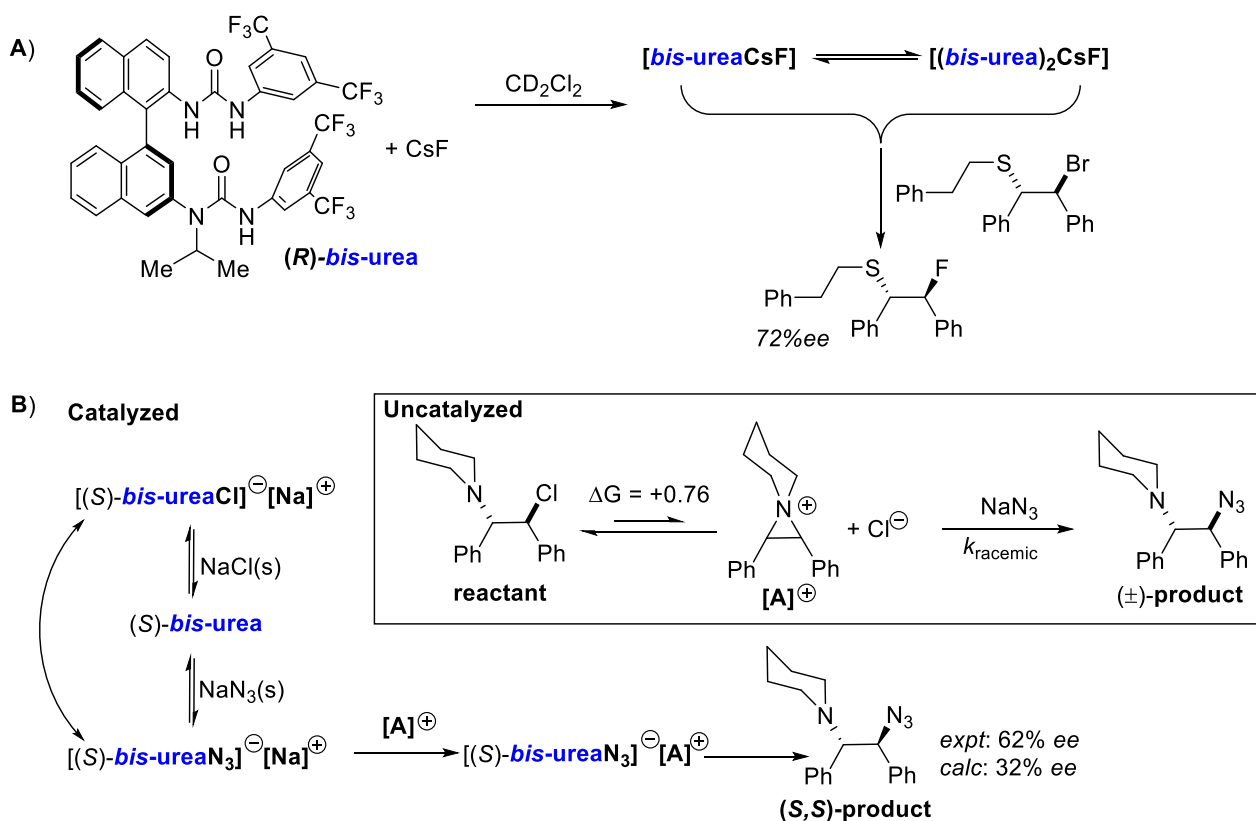
Figure 10. (**A**) Complexation of CsF and bis-urea catalyst as observed via NMR and proposed by Gouverneur, Claridge, and co-workers [116]. (**B**) Reaction pathways of *bis*-urea-catalyzed azidation were reported by Wong et al. [117].

Three potential complications can arise in studying PES via the location of key stationary points and subsequently deriving related kinetic properties from it.

1. Active species involved in catalysis: Aggregation of catalysts is well-documented, and the non-linear effect is closely associated with such a phenomenon.
2. Validity of the transition state theory:
3. Solvation: Extensive discussion on the choice of theoretical methods based on high-quality benchmarks is restricted to gas-phase modeling. Organocatalysis and PTC inevitably are solution-based chemistry. Therefore, the influence of solvents, which can be game-changing at times, often needs to be addressed.

Further elaboration on these three points will be provided in the following subsections.

2.6.1. Non-Linear Effect in Asymmetric Catalysis

A non-linear relationship between the catalyst % *ee* and the product % *ee*, known as the non-linear effect (NLE) [118,119], signals the aggregation of catalyst molecules. However,

one should note that the observation of NLE does not guarantee the involvement of catalyst aggregation [120,121]. Various classic examples are given in the review of Kagan [122]. Some recent examples in organocatalysis where NLE is observed will be discussed below.

Zhao and co-workers reported the observation of a negative NLE in their difunctionalization of allylic sulfonamides via iodination [123]. Together with observing a fractional reaction order with respect to the catalyst, the authors proposed the involvement of catalyst aggregation in the resting state (Figure 11A). However, the exact size of the aggregate is not known. Sun, Huang, and co-workers observed a positive NLE in their chiral phosphoric acid-catalyzed synthesis of α-amino esters [124]. They attributed the NLE to the catalyst's heterochiral aggregates, which have low solubility at the reaction temperature. Jørgensen and co-workers reported a positive NLE in a proline-catalyzed [12 + 2] cycloaddition [125]. They proposed that two proline molecules are involved in the stepwise conjugate addition-carbonyl addition sequence (Figure 11B). DFT calculations supported their proposal. However, Burés, Blackmond, and co-workers performed kinetic simulation on an alternative mechanistic scenario. They concluded that the observed NLE in this [10 + 2] stepwise cyclization could occur due to a complex network of reversible steps instead of involving more than one catalyst molecule [126].

A) Zhao and co-workers

Inactive resting state
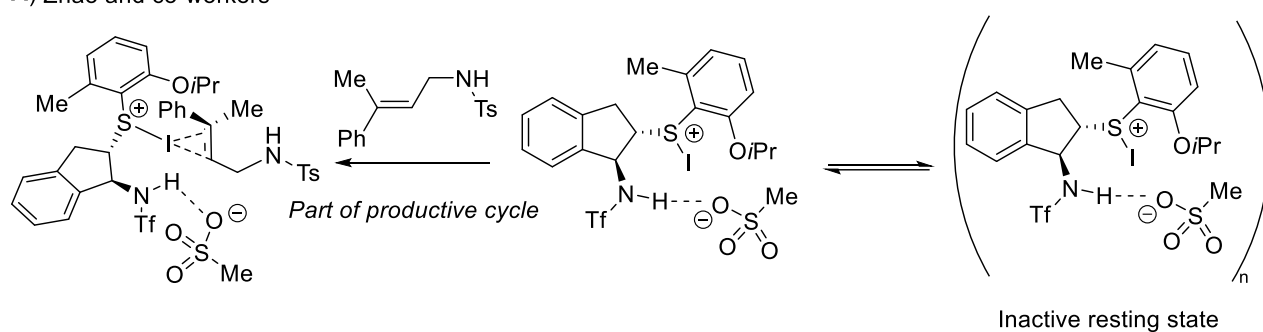
**B**) Jørgensen and co-workers

[12 + 2]
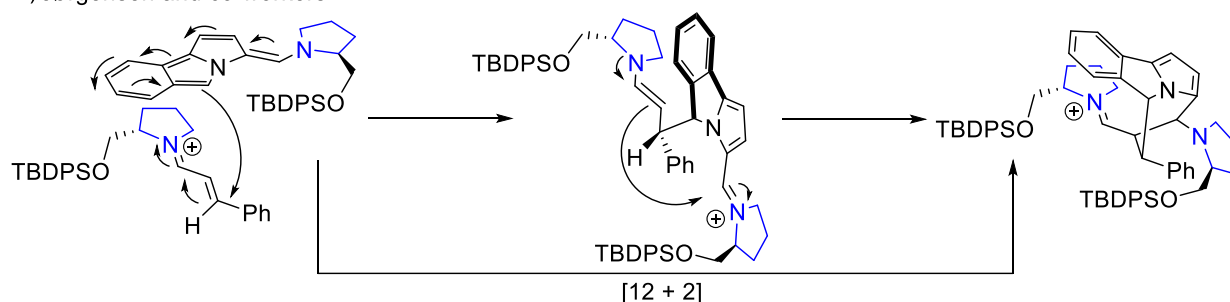
Figure 11. Examples of recent work with proposed NLE. (**A**) Difunctionalization of allylic sulfonamides via iodination as reported by Zhao and co-workers [123]. (**B**) Proline derivative-catalyzed [12+2] cycloaddition as reported by Jørgensen and co-workers [125].

In general, the aggregation of catalyst molecules could present a significant challenge to calculating key reaction performance metrics such as TOF or *ee/er* from investigating a PES for the following reasons. (1) Solubility of the aggregate generally cannot be readily obtained from the study of PES. (2) The degree of aggregate is usually not known. (3) Aggregation would render the location of TS structures or equilibrium geometries computationally impractical for large catalyst molecules. However, the calculated energy profile of a proposed mechanistic network coupled with kinetic modeling can determine the extent of NLE and, when compared with experimentally measured NLE, provide strong evidence to corroborate the proposed mechanism.

2.6.2. Bifurcating Potential Energy Surface

The TST is the cornerstone to in silico prediction of TOF and stereoselectivity, as discussed in the preceding relevant sections. The TS structure or an index-one saddle point has one imaginary frequency. The intrinsic reaction coordinate (IRC) is used to follow this imaginary frequency to determine the two equilibrium geometries (commonly referred to as the reactant/pre-TS complex and the product/post-TS complex) joined by the obtained TS structure. It is assumed that one TS structure will lead to one product via the minimum energy path (MEP) obtained from an IRC calculation. However, it has been extensively documented that many organic reactions exhibit post-TS bifurcation [127–130]. The TS structure exhibiting such behavior is called an ambimodal TS structure. Bifurcation at the "valley-ridge inflection" point allows more than one product to be accessed via one TS structure. An example of such a PES, generated from the model potential reported by Collins et al. [131], is given in Figure 12.
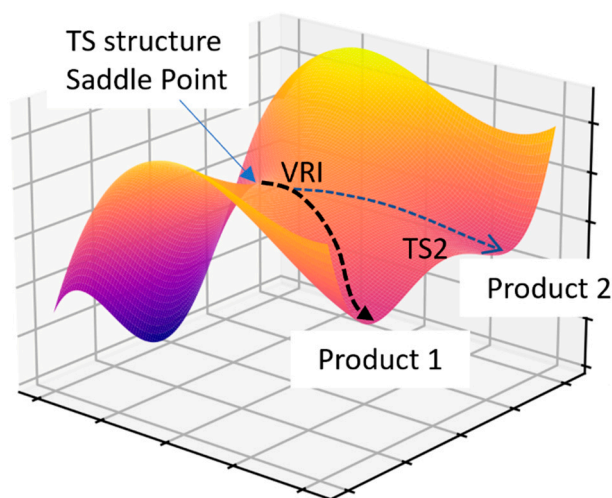
Figure 12. Bifurcation in PES. The PES is generated from the model potential provided by Collins et al. [131].

One of the key assumptions in the TST is that the redistribution of excess internal energy amongst the vibrations and internal rotation is much faster than a reactive elementary step—an assumption that may not always hold. In post-TS bifurcation, a valley-ridge point exists as one traverses the energetic landscape from the reactive complex through the TS to the product.

An example relevant to organocatalysis and PTC would be the S_N2 of alkyl fluoride (Figure 13) [132,133]. In the gas phase S_N2 of thiolate with methyl halide reported by Longo and co-workers [133], with methyl fluoride as the electrophile, the predominant trajectory is *tr*A which corresponds to a bifurcation following the IRC path (*tr*C). For methyl bromide, most of the trajectory bypasses the IRC path to form the S_N2 product direct (trN). The resulting product is HF and methyl thiolate.

Prediction of the level of enantioselectivity in asymmetric organocatalysis or PTC inevitably relies on the use of the TST. The relative energies of the ensemble of TS determine the enantioselectivity. To our knowledge, assessing the importance of bifurcating PES in organocatalysis and PTC has not been reported. Determining the level of enantioselectivity is one potential opportunity to explore post-TS bifurcation further, as the Curtin–Hammett scenario, which is central to selectivity prediction, is rendered invalid—the relative free energies of all relevance ensemble of TS structures cannot be used to predict stereoselectivity or product distribution. A recent example by Singleton, Larionov, and co-workers on the stereochemistry of a photochemically induced ring contraction is illustrated in Figure 14 [134]. While no catalyst is involved in this reaction, it is an example in which

the stereochemistry cannot be determined by the location of two sets of diastereomeric TS structures—the TS bifurcates to form both R and S chiral centers!
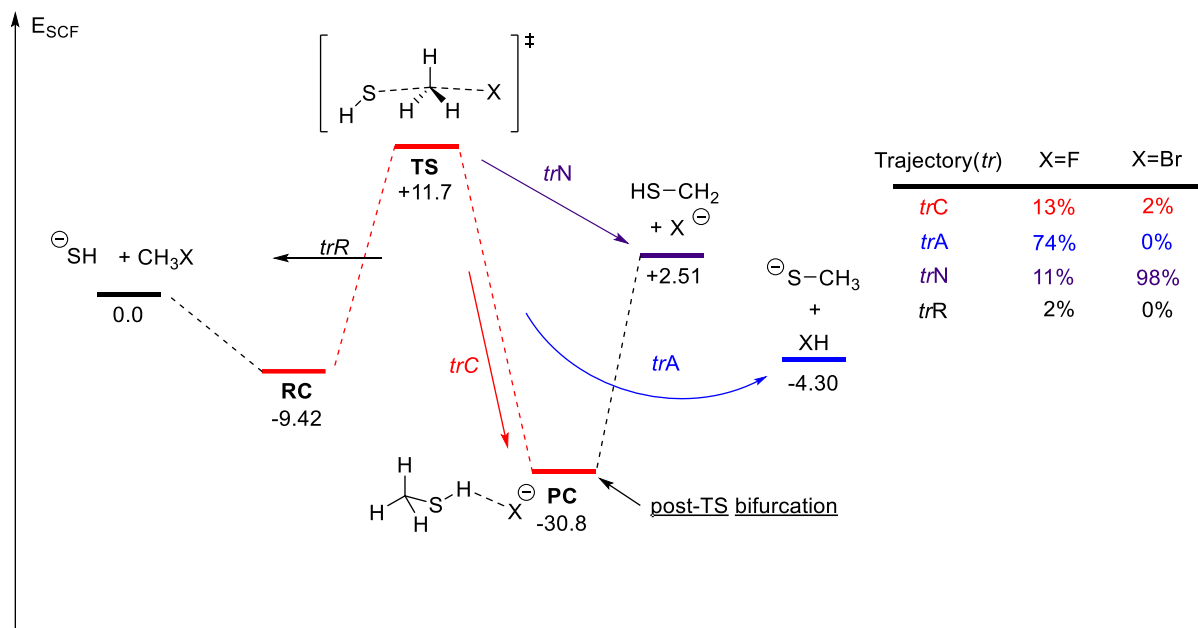
Figure 13. Values (in kcal/mol) calculated at MP2/6-311(d,p) are obtained from ref [133]. trC: approximately IRC path. trA: approximately IRC path followed by a proton transfer. trN: bypass the IRC path to form the product directly. trR: recrossing.
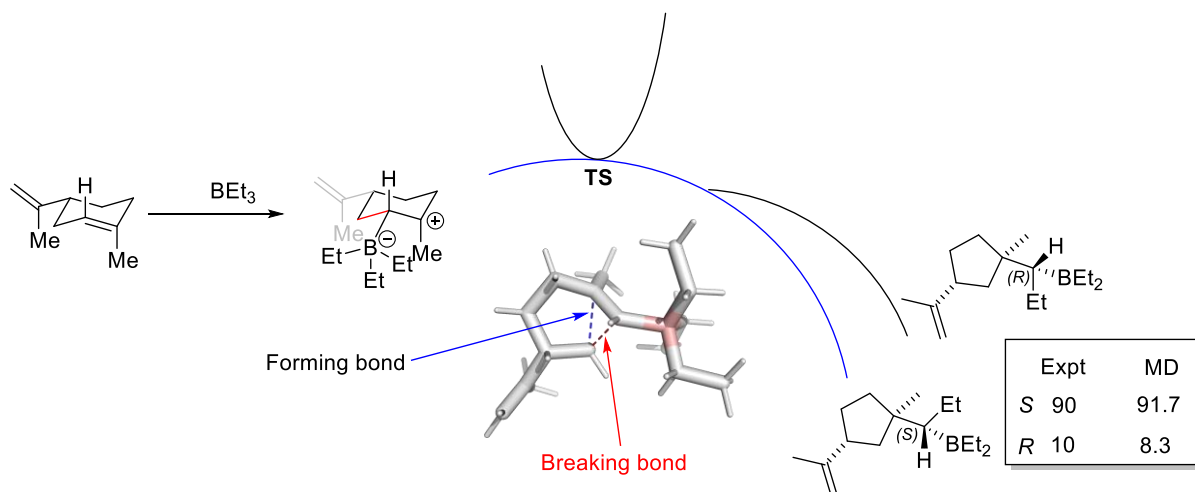

Figure 14. Singleton and co-workers reported an example of bifurcation leading to two stereochemical outcomes [134]. MD: selectivity predicted by a series of molecular dynamics simulations (all values are obtained from their work).

Technically, running many MD simulations to obtain a statistical distribution of potential products from a bifurcating PES is very computationally expensive. While, given the capability of modern hardware and software, it is not an impossible task, an approximate method to obtain the same outcome with less computation cost is nonetheless highly desirable. To this end, Lee and Goodman ValleyRidge PYTHON script can be used [135]. Srnec and co-workers presented the reactive mode compositive vector analysis to determine the ratio of products that arise from the bifurcation [136]. Other methods to approximate the bifurcation product distribution include the work from the groups of Carpenter [137],

Houk [138], and Schmittel [139]. The work by Gomez-Bombarelli and workers utilized a neural network potential to perform reactive MD and may potentially be applied to the study of bifurcation with reduced computational cost [140]. These tools are expected to play an important role in calculating the product distribution where post-TS bifurcation is important and can provide a preliminary indication of such behavior before one decides to invest in a more expensive MD simulation.

2.7. Modeling Solvation

A solvent scan is one of the routines in developing an organocatalyzed or a PTC reaction, as the solvent can profoundly affect both the reaction yield and selectivity. Three possible approaches to model solvation are depicted in Figure 15. The importance of solvation in modeling and its recent advances and present limitation have been highlighted in several reviews [141–144].
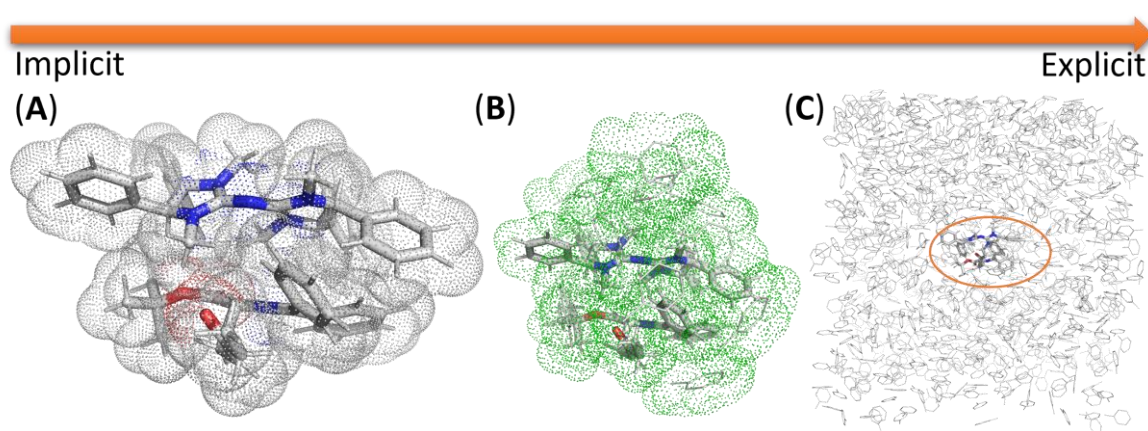
Figure 15. Various solvation models to apply to a PT catalyst that Tan and co-workers reported [145]. (**A**) Implicit solvation. Grey dots are SAS. (**B**) Hybrid approach. Solvent molecules (Toluenes) are displayed as wire. The green dots are the SAS. (**C**) Explicit solvation in a 5.6 nm cube. The catalyst is in the orange oval.

The implicit solvation model is the most common and computationally economical method to account for some of the influence of the solvent. The polarizable continuum model (PCM) is the most common and accessible method to incorporate some solvent effects. In the PCM, the solvent influence is modeled as an ensemble average at thermal equilibrium, interacting with an electric "reaction field" [146]. The surface of the solute interacting with this field can be constructed by rolling a probe of fixed radius over the solute to generate a solvent-accessible surface (SAS). For the caveats and considerations when using a PCM in the study of key points on the PES, the readers are referred to the work of Coote and co-workers [77], and Truhlar, Cramer, and co-workers [147]. The deficiency of the continuum solvation model for charged species has been extensively reported. In the current context, this will be important when overall charged species are considered (e.g., urea-fluoride complex).

The popular SMD solvation model of Truhlar and co-workers [148] is parametrized against experimental solvation data. Its usage on more exotic structures beyond its training set, such as those found in TS structures of organocatalytic or PT-catalyzed reaction, may pose a problem [149,150]. Nevertheless, many computational studies on organocatalysis utilizing this solvation model have successfully reproduced key experimental results (refer to Figure 2 and Section 2.8 Validation of Calculations).

On the other end of the spectrum is the explicit solvation model, where all solvents are included as part of the simulation. It is generally only physically meaningful when combined with calculations supporting periodic boundary conditions due to the large number of molecules required to model a condensed bulk phase—having many more molecules

implies that the conformation degree of freedom would pose an often-insurmountable computational challenge to obtain key stationary points and subsequent calculation of thermochemical correction via statistical thermodynamics, as much it is common to use molecular dynamics simulation when explicit molecules of solvent are involved. However, elementary steps corresponding to most chemical reactions are considered rare events on the MD time scale. As such, enhanced sampling algorithms have to be utilized to bias the simulation and enable free energy to be calculated [151].

In between the two extremes is a hybrid approach where a few solvent molecules could be added to the key region or sufficient solvent molecules are added to form a solvation shell (micro-solvation). A PCM-based solvent model can then be applied to model the bulk solvation. The continuum-cluster approach involves adding a cluster of solvent molecules and using the PCM model together. The thermodynamic cycle has been used extensively to calculate the solvation-free energy of ions in water. Two approaches (Figure 16) exist to construct the thermodynamic cycle—monomer-based [152,153] or cluster-based [154]. They differ by whether individual molecules of solvent or a cluster of solvent molecules are considered. The work of Goddard and co-workers provides a detailed discussion of the various consideration (especially on the choice of standard state and the associated correction) when using such a hybrid approach [155]. The thermodynamic cycle concept was extended to include the solvation energy of the TS structure (which replaces ion(H_2O)$_n$ in Figure 16) in a chemical reaction, as exemplified by the work of Ho and co-workers. However, their accuracy (on ionic S_N2 reactions) was found to be inferior to the free energy of reaction directly obtained from umbrella sampling calculations (vide infra) [55].
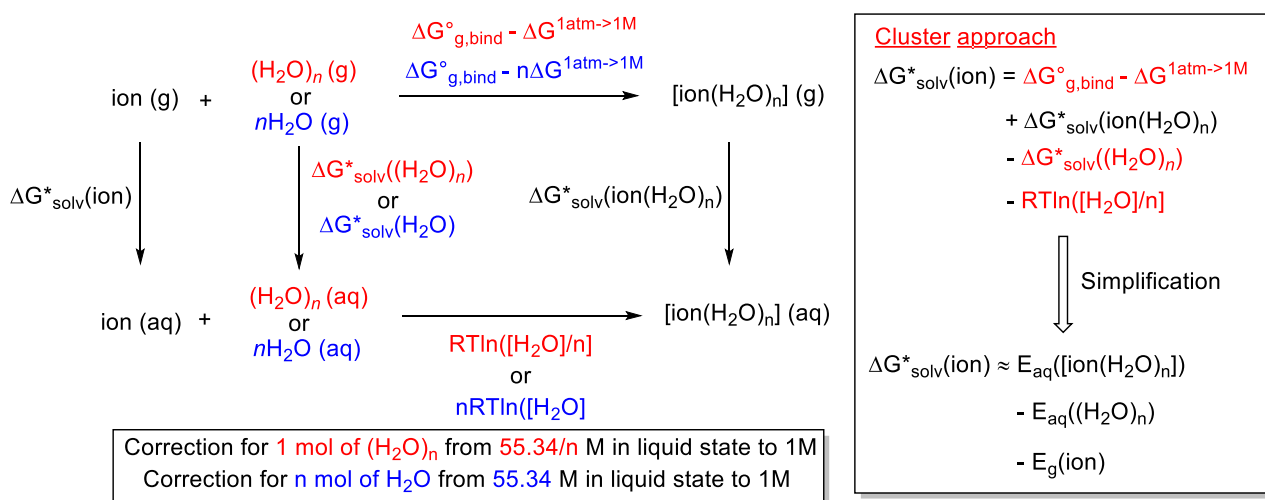
Figure 16. Thermodynamic cycle. * = standard state of 1 M, and ° = standard state of 1 atm. Red: cluster model. Blue: monomer model. $\Delta G^{1\,atm\to1\,M} = RT\ln(24.46)$ is the correction from 1 atm of an ideal gas to 1 M, where R = 0.08206 L atm K^{-1} mol^{-1} is the universal gas constant, and T is the temperature in Kelvin. E_{aq} is the electronic energy of the species calculated with implicit solvation, and E_g is the electronic energy in the gas phase. For assumptions involved in simplification, please refer to Wu and Kieffer [156].

Constructing a converged micro-solvation shell(s) in a continuum-cluster approach is challenging. The quantum cluster growth (QCG) method of Grimme and co-workers handles this task in an automated fashion [157]. ABcluster [158] by Zhang and Dolg used a bee colony algorithm (for the location of global minimum). Basdogan and Keith have used it in their computational study of the Morita–Baylis–Hillman (MBH) reaction (vide infra) [159].

2.7.1. Examples Related to Solvation

The number of atoms required in explicit solvation to produce experimentally realistic concentration is generally incommensurable with reasonable computational cost. Considering the typical concentration of an organocatalytic reaction, a concentration of 0.01 M is a reasonable estimate for the catalyst. At the atomic level, this would require a 5.5 nm cube which can contain tens of thousands of atoms (Table 4). Such a simulation is probably in the realm of linear-scaling DFT coupled with massively parallel computing resources. Another challenge will be that chemical transformation is generally considered a rare event in a molecular dynamics simulation. An enhanced sampling algorithm would need to be applied for them to be observed in a practical duration and obtain critical quantities such as free energy. Despite the increased probability of observing rare events, the number of MD steps to converge such a simulation is often prohibitively high for a DFA-based calculation suitable for modeling chemical reactions (e.g., hybrid GGA).

Table 4. The volume of a cube to contain one particle at various concentrations and the number of water molecules to fill the same cube.

Reactant Concentration (M)	Vol. of a Cube to Contain One Molecule ($Å^3$) [a]	Number of H_2O Molecules in the Volume Given on the Left [b]
1 M	1660 (11.84)	56
0.1 M	16,605 (25.512)	555
0.01 M	166,053 (54.965)	5550

[a] Conversion of M to particle/$Å^3$ = 6.022×10^{-4}. The length of the cube in $Å$ is given in parentheses. [b] Based on a density of 1 g/mL and molar mass of 18 g/mol.

An uncatalyzed reaction at a high concentration (>1 M) will be more amenable to an explicit solvation study of its reaction. Gunaydin and Houk reported using metadynamics simulation to investigate the hydrolysis of methyl formate [160]. The MD simulation with BLYP DFA was performed in a 10.1 $Å$ cube with one methyl formate and 30 water molecules ([methyl formate] ≈ 1.6 M). In their simulation, the hydrolysis proceeds via an auto-dissociation of water. The resulting proton activates the ester by coordinating with its carbonyl oxygen. The hydroxide attacks the carbonyl carbon to form a diol intermediate. The overall $\Delta G^{\ddagger}$ to include the diol intermediate was reported to be 22.8 kcal/mol. We note that in a more dilute setting, the probability of a water molecule auto-dissociating near the ester would be drastically reduced.

In the work of Pliego and co-workers discussed earlier on the hybrid approach by Wang and Cao to model ester hydrolysis [106], they argued that the reference to calculate the barrier should be taken as one methyl formate and seven isolated molecules of water (Figure 17) [99]. In their calculations, this translates to a barrier of 77.2 kcal/mol (or 60.5 kcal/mol after a correction with water concentration by RTln(55.4)). The main contribution to this large barrier was mainly from the loss of translational entropy when eight isolated molecules became one RC. The translational entropy contribution is converted to the liberation free energy in solution. It is thus unlikely to be as punishing as the authors' calculation suggested [87]. In addition, Goddard and co-workers pointed out that the formation of a water cluster can be viewed as a rearrangement of liquid water into solvated water clusters with no associated change in free energy ($\Delta G^{*}_{aq, cluster}$ in Figure 17) [155]. However, we note that the extent to which this holds for an aqueous solution of methyl formate depends on the solute's concentration. Preliminary calculations in Figure 17 give a $\Delta G^{\ddagger}$ of 32.5 kcal/mol with an aqueous cluster of 7 water molecules as a reference. While this is lower than the seven isolated water molecules case, it is still significantly higher than the reported barrier of 25.9 kcal/mol.
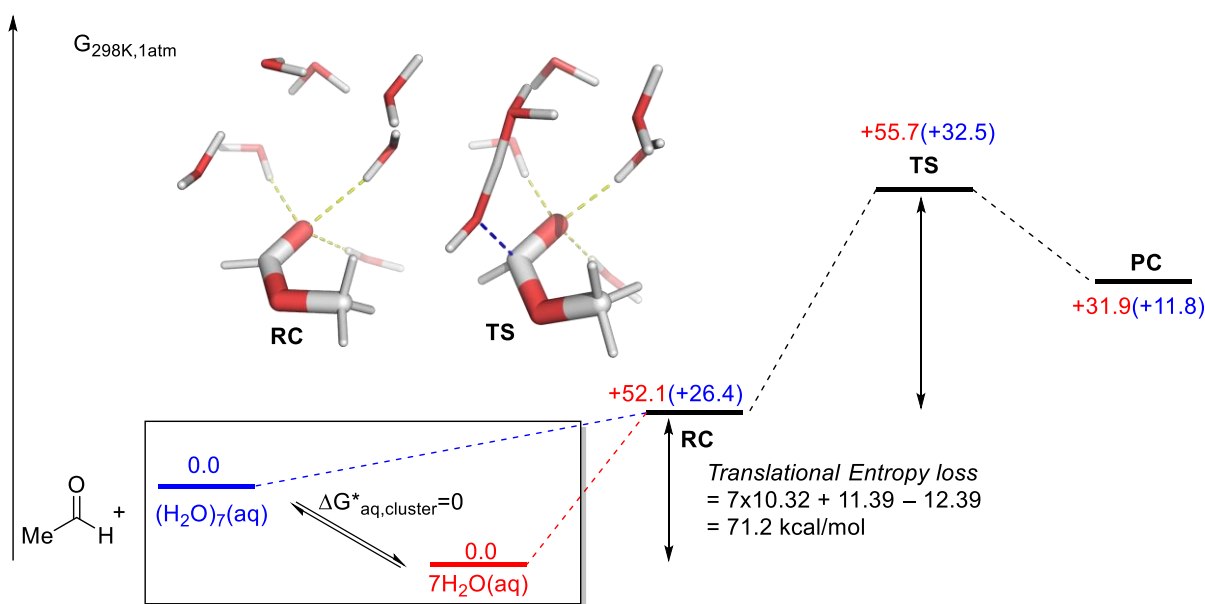
Figure 17. All values in the free energy profile are in kcal/mol and computed at CPCM(H_2O at PBEh-3c)/DLPNO-CCSD(T)/CBS//CPCM(H_2O)/PBEh-3c. CBS is based on a 2-point extrapolation with the def2-TZVPP and def2-QZVPP basis sets. For values in blue, the water cluster is the reference. For values in red, isolated water molecules are the reference. For comparison, key values from Pliego and co-workers [99] at MP4/TZVPP+diff//CPCM/X2LYP/6-31+G(d): $\Delta G^{\ddagger}$(RC to TS) = 24.5 kcal/mol and ΔG(reference to RC) = +52.7 kcal/mol. For Wang and Cao [106] at CPCM/B3LYP/6-311+G(2df,2p): $\Delta G^{\ddagger}$(RC to TS) = 25.4 kcal/mol. Correction of water concentration is not included.

One avenue to accelerate the solvation model that includes explicit solvent molecules is the QM/MM method. The solvent molecules are described with a molecular mechanics (MM) method. At the same time, the catalysts and reactants are modeled at the DFT level. This can significantly accelerate the study of organic reactions with explicit solvation. This technique has been used for some S_N2 reactions with ionic nucleophiles to compare various solvent models' performance [55,161]. Ho and co-workers found that umbrella sampling combined with the ωB97M-V/ma-def2-SVP level of theory as the QM level provides superior results relative to thermodynamic cycle-based calculations when validated against experimentally reported S_N2 reaction in various organic solvents [55]. Keith and co-workers reported a study of $NaBH_4$ reduction of CO_2 in water with various solvation models, including the more sophisticated form of PCM. They found that the PBE DFA overestimates the solvation energies when the ion pairs are in proximity. In conclusion, they recommended, based on their result, a hybrid approach or the use of a conductor-like screening model for realistic solvation(COSMO-RS) [162] and reference interaction site method (ESM-RISM) [163] to account for specific interactions in ionic species [164].

In the context of organocatalysis, the Morita–Baylis–Hillman reaction, with its zwitterionic species and strong solvent dependence, remains challenging to model. Following the detailed experimental measurement reported by Plata and Singleton [82], $\Delta G^{*\ddagger}$ and ΔG^* reactions for several elementary steps were obtained. Sunoj and co-workers attempted to reconcile the difference between values calculated from key stationary points and the experimentally reported values [86]. They used DLPNO-CCSD(T) values to minimize gas phase electronic energy errors. Solvation energies of key stationary points calculated by the free energy perturbation (FEP) method with explicit solvent molecules (modeled with classical forcefield) provide an improvement over the implicit solvation model in some cases, but stepwise proton transfer to form the enolate remains in disagreement (Figure 18). We emphasize that due to the experimental limitation (e.g., unable to observe the change in concentration of all proposed intermediates under realistic experimental conditions), some assumptions need to be made in calculating and assigning values to each point in the
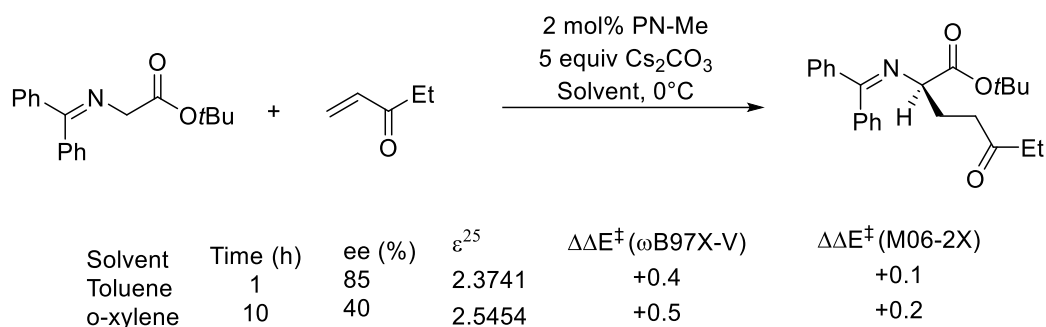
free energy profile. Plata and Singleton pointed this out in their work [82]. Basdogan and Keith, in an extensive study with a diverse set of computational techniques for the same reaction, found that with five explicit MeOH molecules, they obtained the best agreement with experimental values [159].
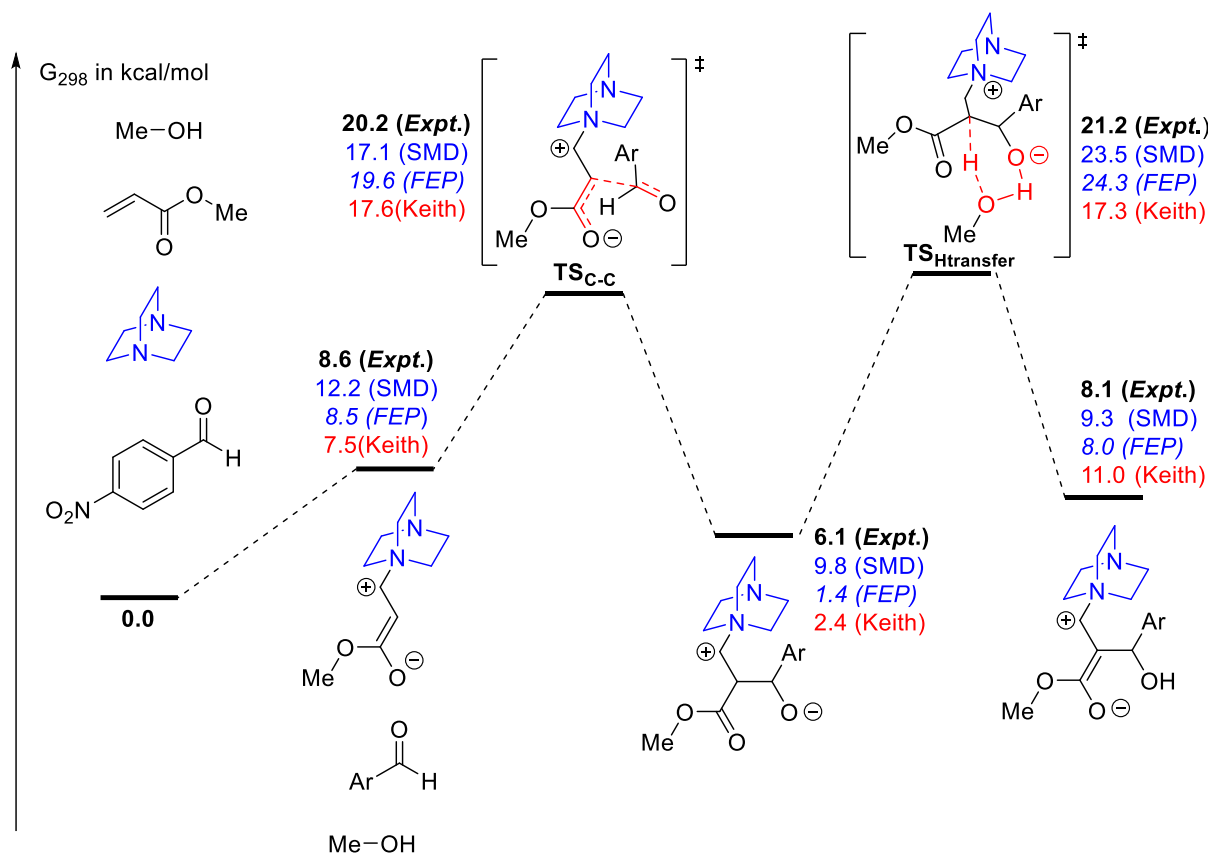
Figure 18. Partial free energy profile of MBH reaction. Experimental values are obtained from the work of Plata and Singleton [82]. Blue calculated values are from the work of Sunoj and co-workers [86] at DLPNO-CCSD(T)/CBS/B3LYP-D3/6-31+G(d,p) with solvation via SMD or FEP as indicated in parentheses. Red calculated values are from the work of Basdogan and Keith at DLPNO-CCSD(T)/def2-TZVP/BP86-D3/def2-SVP [159]. Explicit solvent molecules are not shown.

The inclusion of solvent effects is challenging in PTC, as the key elementary steps could occur at the interface where the solvents' properties differ significantly relative to their respective bulk properties. For instance, solvents with very similar dielectric constants (the basic parameters in the PCM solvation model) can vary substantially in % *ee* in asymmetric PTC. The work of Tan and co-workers on the pentanidium-catalyzed conjugate addition reveals such a phenomenon (Figure 19) [145]. The product yield and % *ee* decreased when the solvent was switched from toluene to *o*-xylene. Both solvents are aromatic hydrocarbons that differ only by a methyl group in the ortho position. Single-point calculations with the SMD solvation model indicated no significant difference in the relative energy of the two key TS structures (as reported by us [165]) in toluene and *o*-xylene. Chin and Krenske highlighted the limitation of the implicit solvation model in describing ion pairs in their computational study of CPA-catalyzed asymmetric Nazarov cyclization, which could be significant in the case presented here [76].

Solvent	Time (h)	ee (%)	ε^{25}	$\Delta\Delta E^{\ddagger}$ (ωB97X-V)	$\Delta\Delta E^{\ddagger}$ (M06-2X)
Toluene	1	85	2.3741	+0.4	+0.1
o-xylene	10	40	2.5454	+0.5	+0.2

Figure 19. Pentanidium-catalyzed asymmetric conjugate addition. Experimental values are from the work of Tan and his co-workers [145]. ωB97X-V/def2-TZVPP and M06-2X/6-31+G(d,p) values included SMD solvation calculated with structures from Kee and Wong [165].

β-cyclodextrin(CD) is an inverse PTC, as it transfers organic molecules into the aqueous layer. Elk and Benjamin studied the phase transfer process of β-CD in a biphasic layer of water and 1-bromobutane [166] or 1-bromooctane [167]. Classical forcefields were employed in these studies. The same group also reported an empirical valence bond (EVB) approach to study the β-CD catalyzed S_N2 reaction of CH_3Cl and chloride via umbrella sampling with explicit solvation [168]. Their calculations indicate that the reaction barrier with β-CD complexation and in its absence is similar in bulk water. However, at the liquid/liquid interface, β-CD complexation reduces the reaction barrier by 2.4 or 3.4 kcal/mol relative to the uncatalyzed reaction.

2.8. Validation of Calculations

As in the case of machine learning, validation and testing lend credence to the predictive capability of computational methodologies employed in organocatalytic calculations. This is especially important due to the many layers of approximations and assumptions involved in arriving at predicted values.

Validation with experimental TOF is more challenging as most organocatalysis or PTC works do not measure the TOF of their reported reaction but only the isolated yield at the end of a specific duration. Measurement of TOF experimentally is cumbersome, as several points must be sampled accurately and fitted to a function where the gradient at the desired point in time can be calculated. For such purposes, the isolated yield after a purification process (such as column chromatography) is unlikely to reflect the catalyst efficiency accurately due to losses. In situ spectroscopic measurement (to obtain the spectroscopic yield) is the preferred method for such a purpose. Calculation of the TOF of a product from key stationary points is highly challenging, as it requires the critical productive and non-productive pathways to be known. In addition, modeling the thermodynamic and kinetic of phase transfer processes in PTC, which is highly challenging to model accurately, will add to the challenge. Advances in automated reaction pathway exploration algorithms coupled with a fast and reliable means to evaluate the energy and forces of a molecular system will be imperative to systematically obtain all significant pathways which the reactants and catalysts can access under a given reaction condition. In silico prediction of TOF in organocatalysis and PTC remains a challenge to be surmounted.

Some examples of detailed kinetic studies in organocatalysis include guanidine and its derivatives [169,170], proline [171–173], and thiourea/urea [174]. More recent examples in organocatalysis or PTC include the work by Jacobsen and co-workers [88] and Gouverneur and co-workers [117]. However, performing a PES study to locate all the required stationary points required to reproduce the experimentally measured rate constants is going to be highly challenging. As such, they are not attempted even in these highly detailed studies. In addition, the intrinsic error associated with absolute rate measurement will impede its effectiveness in validating against theoretically derived ones.

2.8.1. Validation: Heavy-Atom Kinetic Isotope Effect

When judiciously applied, the kinetic isotope effect (KIE) can provide insight into the nature of the TS [175]. Coupled with the calculated TS structure, KIE can provide a multi-point validation for theoretical calculations. The synergy between heavy atom KIE and theoretical modeling via the location of TS structures, together with their associated challenges, are presented in detail by Lloyd-Jones and co-workers [176]. Proline catalysis will be a classic example where ^{13}C KIE was utilized in conjunction with the theoretical calculation [177]. Two selected recent examples in organocatalysis will be highlighted below.

Kwan, Jacobsen, and co-workers utilized a distortionless enhancement by polarization transfer (DEPT) technique to measure the $^{13}C/^{12}C$ KIE in a thiourea-catalyzed glycosylation (Figure 20). Conventionally, an S_N1-based mechanism gives rise to an inverse or small normal KIE. In contrast, an S_N2-based mechanism gives rise to a large normal KIE [178]. However, in thiourea-catalyzed glycosylation where an asynchronous S_N2, which involves the formation of a loose ion pair by dissociation of the leaving group followed by a rate-determining nucleophilic addition, can occur [179]. In conjunction with TS structure calculations, the authors provide evidence for an asynchronous S_N2 by comparing calculated S_N2 KIE at C1 and S_N1 equilibrium isotope effect (an approximation to S_N1 KIE) at C1 with experimentally measured values.
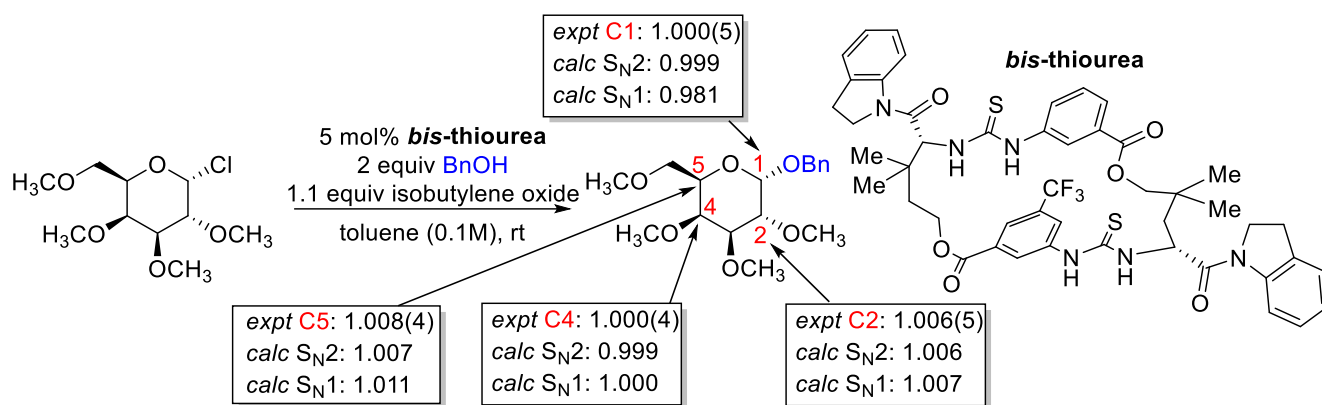
Figure 20. *bis*-Thiourea catalyzed glycosylation. All values are obtained from the work of Kwan, Jacobsen, and co-workers [178]. "*Expt*" refers to experimental ^{13}C KIE relative to the C2 methoxyl group. "*Calc S_N2*" refers to the calculated KIE value at PCM/PBE0-D3(BJ)/6-31G(d) and the Biegeleisen–Mayer method [180] adjusted with the Bell infinite parabola tunneling correction. "*Calc S_N1*" refers to the equilibrium isotope effect.

Hirschi, Mukherjee, and co-workers used ^{13}C KIE as evidence to differentiate between various mechanistic scenarios. Experimental and calculated ^{13}C KIE synergistically provide support for an E1cB mechanism [181]. They observed a significant inverse ^{13}C KIE, usually associated with increased steric congestion at the carbon center on C2 (Figure 21). Based on this, they eliminated the C-C bond formation as the first irreversible step. Instead, they proposed nitro group elimination as the candidate. The E1 and E2 pathways were eliminated by their insurmountable barriers under the reaction condition. The ^{13}C KIE work further reinforced this point.
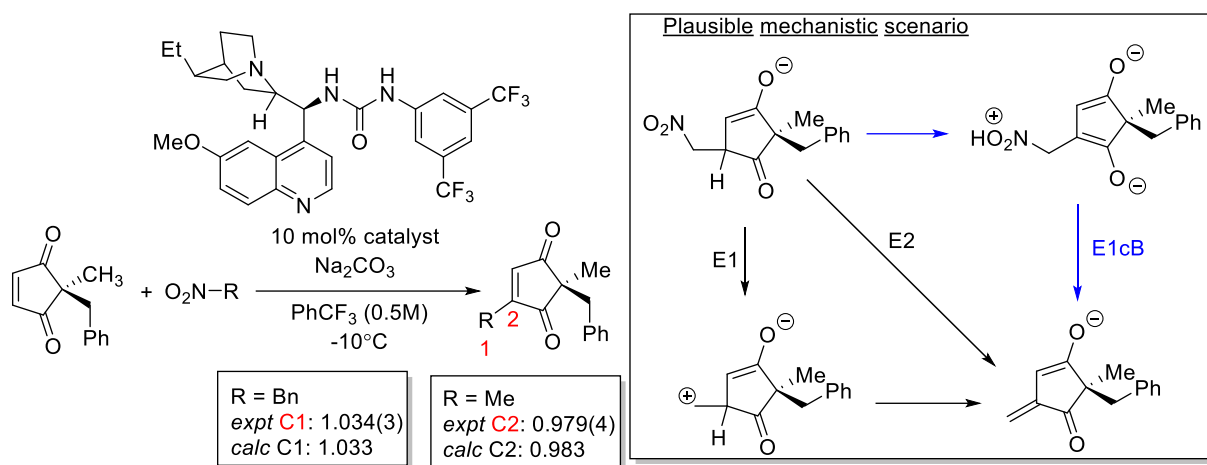
Figure 21. Urea-catalyzed desymmetrization [182]. All values are obtained from the work of Hirschi, Mukherjee, and co-workers [181]. "*Expt*" refers to the experimental ^{13}C KIE relative (please refer to the cited reference for details. *Calc*" refers to the calculated ^{13}C KIE at IEFPCM/B3LYP/def2-SVP and the Biegeleisen–Mayer method [180].

2.8.2. Validation: Enantioselectivity

Asymmetric organocatalysis provides a testing ground for modern computational chemistry. A metric of enantioselectivities such as *ee* or *er* can be measured reasonably accurately if sufficient care is taken. Theoretical calculation via the stationary point approaches does not usually require a highly detailed map of the reaction energy profile but only the key TS structures of the stereocenter-determining steps. The Curtin–Hammet principle is then invoked to determine the level of enantioselectivity by using a Boltzmann distribution on the ensemble of TS structures [16,183].

Due to the non-linear relationship between % *ee* or *er* with $\Delta\Delta G^{\ddagger}$, an error of 0.5 kcal/mol can translate to a very different magnitude of error in % *ee* (Figure 22). For instance, at 253.15K, 0 to 0.5 kcal/mol of $\Delta\Delta G^{\ddagger}$ corresponds to 0 to 46% *ee*, but from 2.0 to 2.5 kcal/mol, the difference in % *ee* is only 2.3%. This implies that at high % *ee*, the room for error in prediction is more forgiving than at low % *ee*. Therefore, a computational study that successfully predicts high enantioselectivity is often insufficient to showcase its predictive capabilities. Based on this analysis, we will showcase some selected examples in organocatalysis that demonstrate such predictive capabilities.
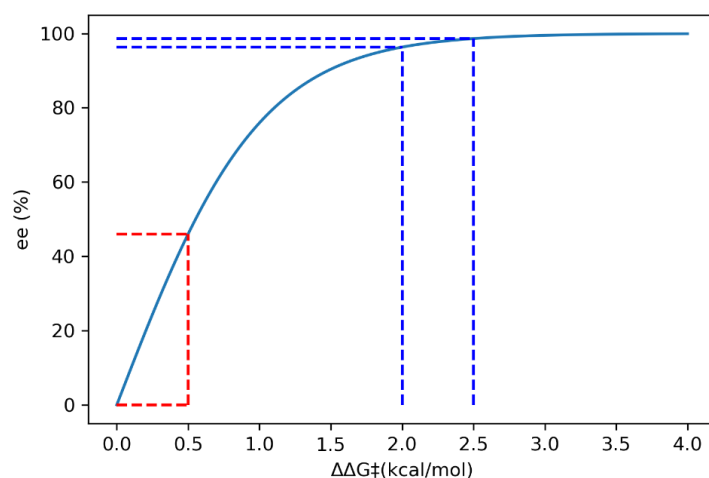

Figure 22. % *ee* vs. $\Delta\Delta G^{\ddagger}$(kcal/mol) at 253 K. Two intervals of 0.5 kcal/mol are included to illustrate the difference in calculated % *ee*.

Ermanis, Phipps, Goodman, and co-workers amply demonstrated this in their work on the chiral phosphoric acid-catalyzed enantioselective Minisci reaction (Figure 23). They observed a positive NLE, which is attributed to the formation of the insoluble heterogeneous dimer that is not directly involved in the main productive pathway [31]. The authors' calculations can reproduce the variation in experimental % *ee* with various substrates. In this case, no dispersion correction is applied in the geometry optimization. However, their choice of method is sufficiently accurate to reproduce key experimental trends in % *ee* with selected examples from (70% *ee* and beyond).

Expt:	73% ee	92% ee	94% ee	69% ee
Calc:	75% ee	86% ee	99% ee	47% ee
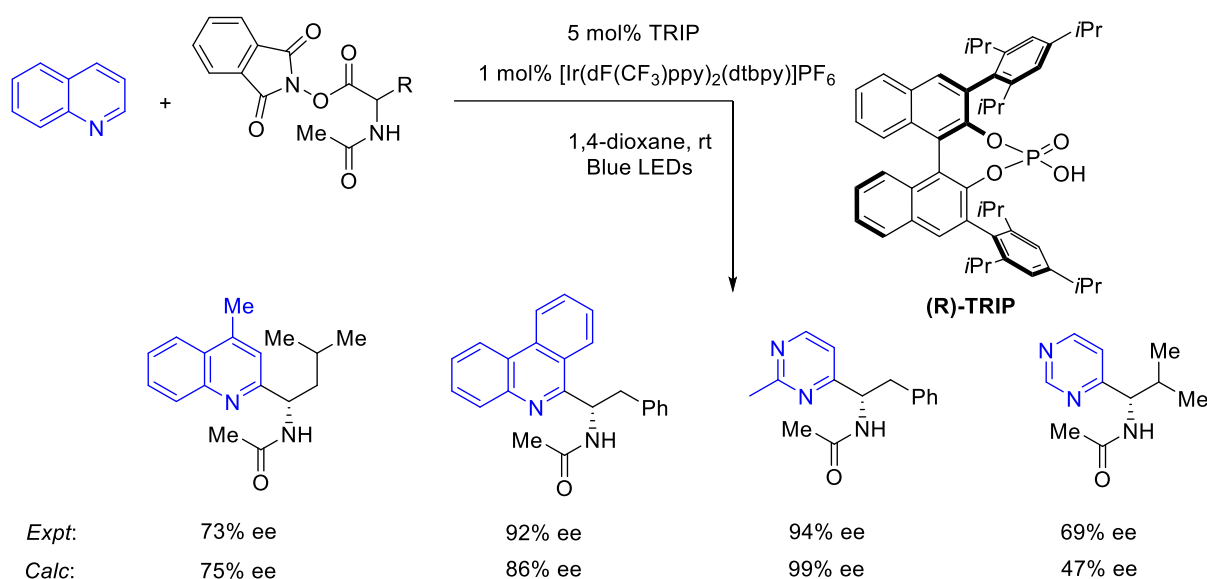
Figure 23. Enantioselective chiral CPA catalyzed Minisci reaction [184]. All values are from the work of Goodman and co-workers [31]. % *ee* are calculated from $\Delta\Delta G^{\ddagger}_{298.15\,K,\,1\,atm}$ at SMD(1,4-dioxane)/M06-2X/def2-TZVPD//B3LYP/6-31G(d,p) from the cited reference.

We have demonstrated in the bicyclic guanidine catalyzed reaction between *N*-phenyl-malemide and anthracen-9(10H)-one that our calculations can reproduce the difference in % *ee* between the *tert*-butyl and phenyl bicyclic guanidine (Figure 24) [32]. Our calculated % *ee* indicates that 4,5-dichloroanthracen-9(10H)-one has a lower level of enantioselectivity than reported by the authors [185]. Upon closer inspection of their supporting information, the reported 99% *ee* is likely overestimated. This highlights the experimental errors associated with high % *ee*. The minor enantiomer peaks might have significant noise from the baseline if the sample used for analysis is too dilute or the integration of peaks is done incorrectly.

Guandine			
Expt:	81% ee (253K)	30% ee (298K)	94% ee (253K)
Calc:	79% ee	25% ee	80% ee
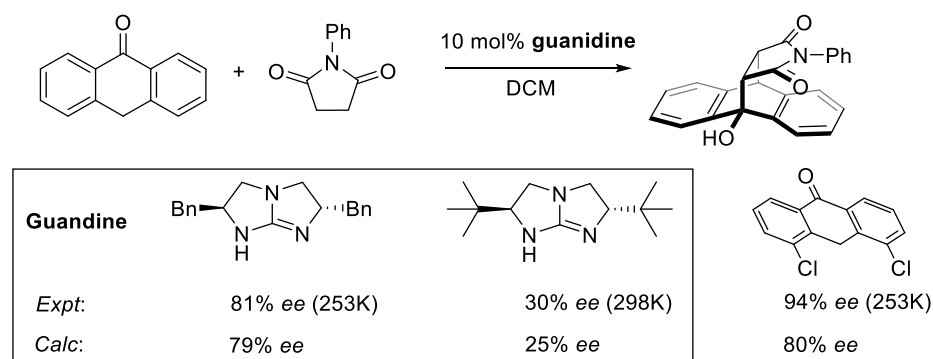
Figure 24. Bicyclic guanidine catalyzed cycloaddition [185]. Values are from the work of Kee and Wong [32]. % *ee* are calculated from $\Delta\Delta G^{\ddagger}$ at CPCM(DCM)/DSD-PBE95-D3(BJ)/def2-QZVP//PBEh-3c as reported by the cited reference.
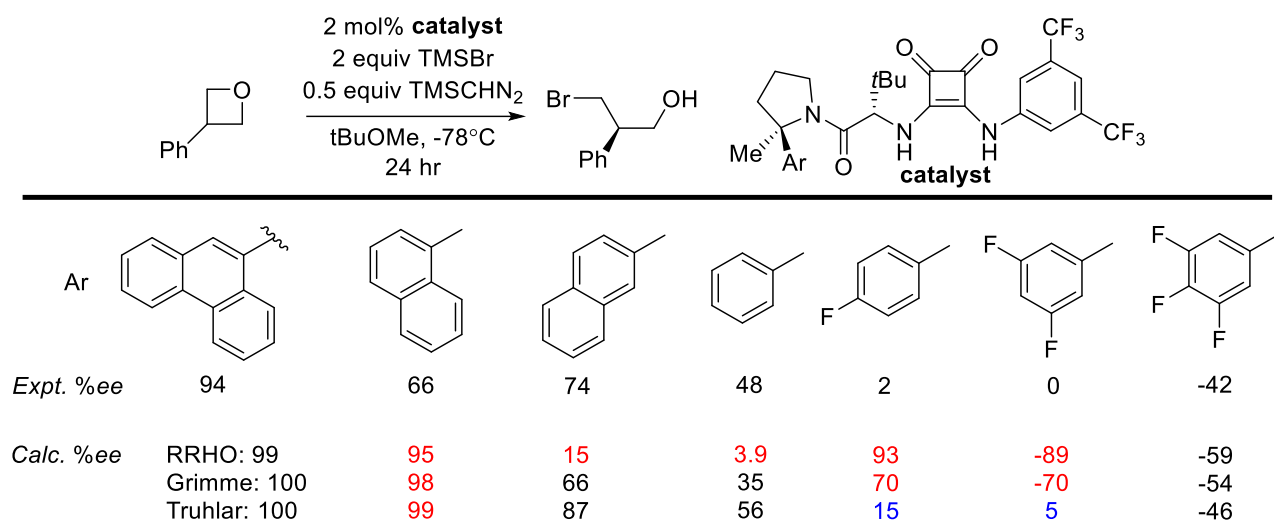
In their study of SPINOL-phosphoric acid (SPA) catalyzed, Wheelers and co-workers show that the improvement in the level of enantioselectivity by varying the SPA's side arm can be reproduced by their modeling (Figure 25) [34]. While the difference in % *ee* between calculated and experimental can be 10% or more in magnitude, the reader should note that such differences, when translated to $\Delta\Delta G^{\ddagger}$, can range from a difference of 0.14 kcal/mol to 1.1 kcal/mol due to the non-linear relationship between % *ee* and $\Delta\Delta G^{\ddagger}$ (see Figures 22 and 25). Since chemical accuracy is usually defined to be 1 kcal/mol or less, good agreement with the experimental % *ee* is generally observed (near the limit of chemical accuracy, refer to values in parentheses in Figure 25), ranging from the low % *ee* of the 2,6-dimethylphenyl group to the drastic improvement in phenyl and 9-anthracyl group.
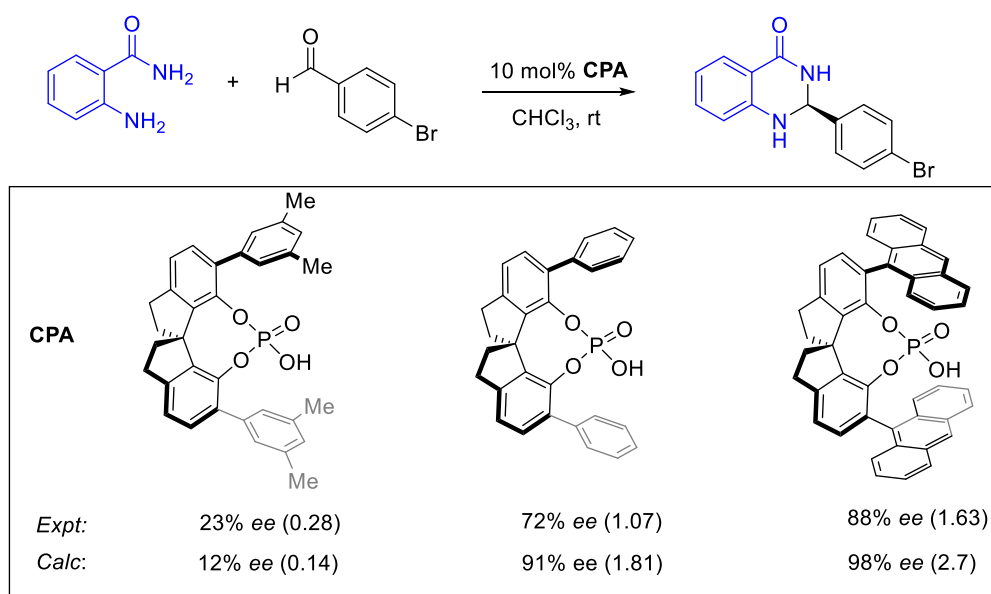
Figure 25. SPINOL-CPA catalyzed asymmetric synthesis of dihydroquinazolinones [186]. All values are obtained from the work of Wheeler and co-workers. % *ee* are calculated from $\Delta\Delta G^{\ddagger}_{298.15\ K,\ 1\ atm}$ at PCM(CHCl$_3$)/B97-D3/def2-TZVP//PCM(CHCl$_3$)/B97-D/def2-TZVP from the cited reference. Numbers in parentheses are the corresponding $\Delta\Delta G^{\ddagger}$ at 298.15 K in kcal/mol.

Jacobsen and co-workers reported a detailed mechanistic study on the ring-opening of oxetane with TMSBr (Lewis acid mechanism) or HBr (Brønsted acid mechanism) catalyzed by chiral squaramide (Figure 26). The TS structures for the critical enantioselective steps were located at SMD(Et$_2$O)/B97D/def2-SVP [88]. For a set of seven catalysts variation for the Lewis acid mechanism, they explored the use of two different quasi-RRHO corrections to low vibrational frequencies and RRHO, where they found that the use of quasi-RRHO correction is generally beneficial. Some of the enantioselectivity appears to arise mainly from vibrational entropy, as they are highly sensitive to the correction used (see also the work of Chin and Krenske [76]). The quasi-RRHO correction by Truhlar reproduces the low level of enantioselectivity for the 4-fluoro- and 3,5-difluoro-phenyl groups. The decrease in % *ee* from the 9-phenanthryl to 1-naphthyl catalyst cannot be reproduced by applying a quasi-rotor correction.

These selected, but not exhaustive, examples demonstrate the potential of DFT calculations for in silico design of asymmetric reaction. However, from a practical point of view, given current technological limitations and costs, one would still need access to a considerable amount of computing power to perform these calculations within a competitive timeframe with experimental screening.

Ar						
Expt. %ee	94	66	74	48	2	0	-42

Calc. %ee						
RRHO: 99	95	15	3.9	93	-89	-59
Grimme: 100	98	66	35	70	-70	-54
Truhlar: 100	99	87	56	15	5	-46

Figure 26. Squaramide-catalyzed enantioselective oxetane opening. Values are from the work of Jacobsen and co-workers [88]. Only the Lewis acid mechanism result is shown here. % *ee* are calculated from the $\Delta\Delta G^{\ddagger}_{195.15\,K,\,1\,atm}$ calculated at SMD(Et$_2$O)/B97D3/def2-TZVP//SMD(Et$_2$O)/B97D/def2-SVP as reported by the cited reference.

2.9. Understanding/Interpretation

The activation strain model (ASM) model decomposes the activation barrier into two contributions: (1) the energetic penalty imposed to distort the reactants and catalyst from their lowest energy conformers to adopt the reactive conformation and (2) the interaction between distorted components [187,188]. The second part requires an energy decomposition analysis (EDA) which will be described in more detail below. The open-source PyFrag 2019 code can be used to perform the ASM calculations. Readers are encouraged to refer to the work of Hamlin, Bickelhaupt, and co-workers for more detail [189]. ASM can be applied to asymmetric catalysis instead of a reaction pathway, where the difference in geometrical strain/distortion between the pair of stereo-differentiating TS structures can be compared.

In organocatalysis or PTC, energy decomposition analysis (EDA) allows the interaction energies between the catalyst and substrates in the crucial TS structures to be decomposed into different contributions (electrostatic, exchange-repulsion, and dispersion). Neese and co-workers developed the local energy decomposition (LED), which uses a supermolecular approach to decompose the DLPNO-CCSD(T) interaction energy [190]. The LED decomposes the interaction energy into electronic promotion, electrostatic, exchange, dynamic charge polarization, and dispersion contributions. An application of LED is discussed below.

List and co-workers' silyated C-H acid catalyzed enantioselective Diels–Alder [191] was studied in detail by Bistoni and co-workers [40]. Both LED and ASM were applied in their work. Through LED of DLPNO-CCSD(T) interaction energy, the authors showed that dispersion interaction is the primary source of energetic difference between the two diastereomeric TS structures (Figure 27). ASM was used to quantify the geometry distortion between the two TS structures.

The non-covalent interaction Index (NCI) is a convenient way to visualize the non-covalent interactions between fragments in a molecular system [192]. It is based on the dimensionless reduced density gradient (RDG). The sign and magnitude of the RDG can be used to classify the type of interaction ranging from strongly attractive to strongly repulsive. NCI isosurfaces can be generated with NCIplot [193] or Multiwfn [194]. It can be used in organocatalysis to illustrate the key noncovalent interactions in critical TS structures. We will demonstrate this with the work of Bistoni and co-workers, as discussed above. The NCI isosurfaces indicate the difference in *van der Waals* type of interaction between

the TS that leads to the major product—the *S, S* enantiomer, and that which leads to the minor enantiomer. The interaction between the naphthyl group of the catalyst and the cyclopentadiene is missing (Figure 27), which is consistent with the LED result.
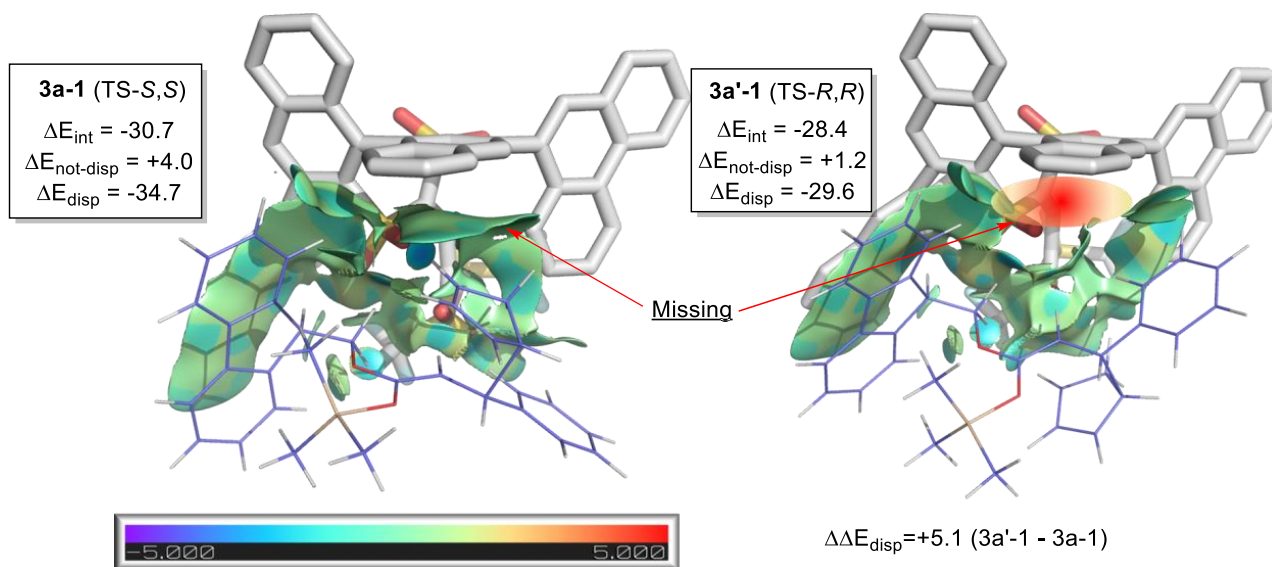
Figure 27. Geometries optimized at PBE-D3/def2-SVP were taken from Bistoni and co-workers [40]. NCI surface was generated by NCIplot with promolecular density. ΔE_{int} is the interaction energy between the catalyst and substrates calculated at DLPNO-CCSD(T)/def2-TZVP. $\Delta E_{not\text{-}disp}$ and ΔE_{disp} are the non-dispersion and dispersion contributions to the interaction energy, respectively, from LED analysis. All values are in kcal/mol and taken from Bistoni and co-workers [40].

The independent gradient model (IGM) developed by Hénon and co-workers is based on electron density contragradience [195]. The NCI and IGM isosurfaces of C60@6-CPPA based on DFT densities are shown in Figure 28. Interestingly, despite the repulsive interaction between C60 and 6-CPPA at B3LYP/6-31G(d) calculated at +18 kcal/mol, NCI and IGM isosurfaces indicate that their interaction is weakly attractive.
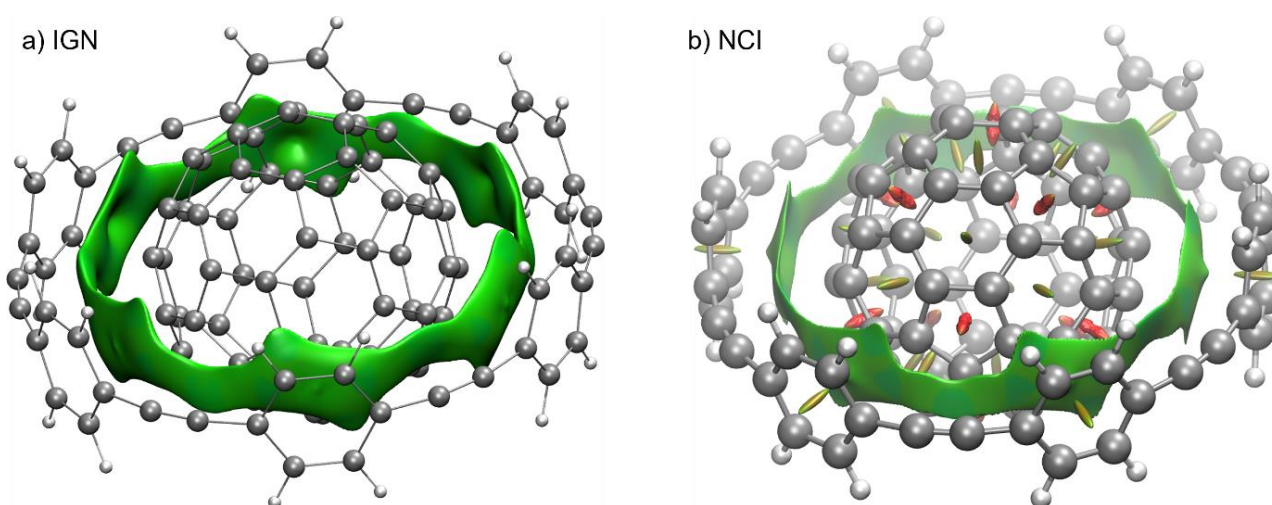
Figure 28. (**a**) IGM isosurface and (**b**) NCI isosurface generated with densities at B3LYP/6-31G(d). Visualization was done with VMD [196].

2.10. Machine Learning and its Relevance to the Field

Machine learning (ML) is a ubiquitous topic in contemporary research and is currently a very active area. It is beyond the scope of this review to provide details into the multitude of applications that ML has impacted chemistry. Our emphasis will be on aspects of ML that can be utilized in in silico prediction of catalysts' performance (yield, TOF, and stereoselectivity). Therefore, unsupervised learning-based works will not be discussed, although they are of great importance [107,197–199]. Readers are encouraged to read authoritative reviews and perspectives on these topics and the references cited to better appreciate ML's depth and impact in chemistry [200–207].

2.10.1. Supervised Learning

As discussed in the preceding section, in silico calculation of reaction efficiency metrics such as TOF or TON is challenging. In supervised ML, reactants (including catalyst and solvent) can be transformed into features or descriptors which can be mapped to the reaction yield via a model. Doyle's group used results from high-throughput experiments of C-N cross-coupling with ML to predict reaction yields [208]. The use of ML to predict yields or turnover numbers in synthetic chemistry has been reported for transition metal-catalyzed asymmetric reactions [209–211], Heck reaction [212], multi-component reactions [213], and Michael addition [214].

Pertinent to asymmetric organocatalysis, Denmark and co-workers developed the average steric occupancy (ASO) descriptor to predict the level of enantioselectivity [215,216]. In later work by the same group, the use of random numbers to replace the features has pointed to a potential problem of bias in the splitting of dataset training randomly [217]. This problem was reported earlier by Chuang and Keiser, who employed two classical tests of generalization to Doyle and co-workers' data: random number as features (straw model) and one-hot vector, which encodes the presence or absence of reaction components [218].

The number of data points required to train a meaningful model is generally incommensurate with organocatalysis methodology development. For instance, 1000 data points would be considered a tremendous amount of experimental work. However, from an ML perspective, it may be wholly insufficient (depending on the diversity of the data points). A methodology-developing chemist (without access to a high-throughput facility) would be unlikely to generate this many data points throughout the optimization process. The diversity of the data points is crucial for ML. Skewed data points would result in the problems discussed previously. In Doyle and co-workers' reply to Chuang and Keiser, they employed domain expertise (knowledge of the roles of reactants) to partition their reported dataset and showed that the one-hot vector performance is no longer comparable to the chemical-based features model [219].

Feature engineering is an intrinsic part of ML. In the quest to develop generally applicable and computationally economical descriptors for general modeling reaction yield and selectivity, Glorius and co-workers engineered features or descriptors from multiple molecular fingerprints [220]. Luo, Zhang, and co-workers used an ensemble of structural and physical organic chemistry (SPOC) descriptors [221].

The features employed in ML reaction studies are continuous or discrete numbers derived from either calculated properties or experimental observables. In the case of organocatalysis/PTC or general homogeneous catalysis, the catalysts have well-defined connectivity between their atoms that will be lost when translated into a scalar descriptor/feature. These features/descriptors cannot precisely be reversed and mapped to a catalyst's molecular structure, impeding their use to design novel and more efficient catalysts. Ideally, one would use an optimization algorithm to search the features' space for an improved catalyst version, but this is usually not possible if the features cannot be reverse-mapped to a valid experimental condition (including catalysts, additives, or reactants). The inverse design of reaction from a supervised learning model based on these features remains challenging.

2.10.2. Machine Learning Potential

At the heart of an in silico calculation of catalytic performance is the capability to calculate the energy and force of a system with various workflows. The ability to compute these quantities accurately and in a timely fashion is imperative in driving in silico reaction/catalyst design. In a machine learning potential (MLP), atom-based descriptors are mapped to energy, forces, and stress tensors. The model can be kernel-based or neural network. It can reproduce the accuracy of the dataset it is trained in a fraction of the time. For a more in-depth appreciation of MLP, the reader is referred to the review by Behler [222], Muller, Tkatchenko, and co-workers [223], and Pinherio Jr et al. [224]. Nevertheless, the "no-free-lunch" principle applies; significant effort has to be invested in generating the dataset for training MLP, and the quality of MLP is limited by the dataset on which it is trained.

Several general-purpose MLPs suitable for organocatalysis have been reported in the literature, for instance, ANI-2x [225], DimeNet++ [226], GemNet [227], and SchNet [228]. The Open Catalyst project also provides state-of-the-art MLPs that can be used out-of-the-box, although these MLPs are targeted toward heterogeneous and electro-catalysis [18,19].

A limited example is provided in Figure 29. We evaluated two MLPs on the relative potential energies of three critical TS structures in our work on the guanidine-catalyzed cycloaddition of Anthrone [32]. The time required to calculate potential energy with these MLPs is insignificant relative to the double hybrid DFA reference. GemNet-T can predict TS_SS_2 as the lowest energy TS structure, while ANI-2x predicts that TS_RR should be the most stable TS structure. The difference of less than 2 kcal/mol is noteworthy as these MLPs have not been trained on these TS structures. We use ANI-2x to perform a constrained optimization of TS_SS_2. The RMSD between the PBEh-3c and ANI-2x optimized geometries is 0.41, which is similar to that reported by Bistoni and co-workers when going from a double-zeta to a triple-zeta basis set as discussed above [40].

	TS_SS_1	TS_SS_2	TS_RR_16
ANI-2x	+2.1	0.0	−1.5
GemNet-T	+0.6	0.0	+0.5
DSD-PBEB95	+2.0	0.0	+2.2
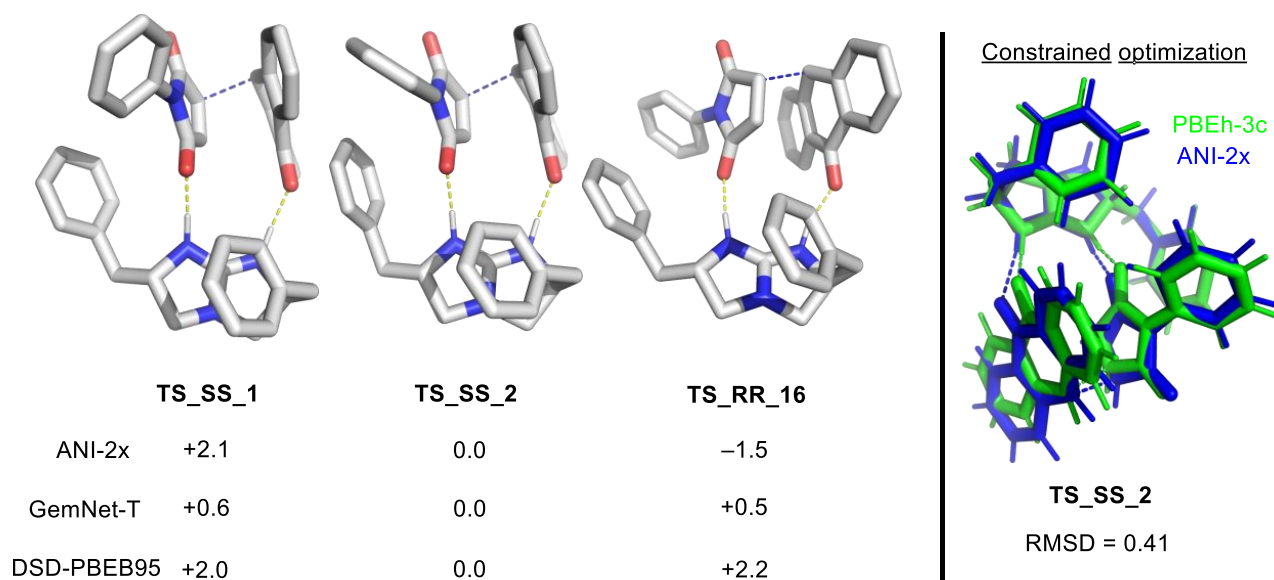
Figure 29. DSD-PBEB95: DSD-PBEB95-D3(BJ)/def2-QZVP. Values are taken from ref [32]. Geometries are generated from coordinates supplied by ref [32]. RMSD calculated by alignment function of PYMOL. Constrained optimization is done with ASE. In the constrained optimization, the two C atoms involved in the C-C bond formation are fixed (atom constraint).

Here, we list some examples that could be of relevance to PTC. MD simulations with MLP on condensed phases have provided molecular insights into phenomena such as the dynamics of proton transfer in NaOH [229], dissolving salts in water [230], permeation of water into nanocapillaries [231], and zeolite hydrolysis [232]. Studies on aqueous hydroxide,

commonly used in PTC, can potentially be extended to investigating relevant phase-transfer catalyzed reactions.

With NN potential, MD simulation with tens of thousands of atoms is generally within reach via a decent Graphic Processing Unit (GPU) accelerator. With sufficient computation resources, GPU accelerated MD simulations with deepMD-kit [233] have allowed the study of ice nucleation for a 300,000 atoms system on a SCAN DFA trained MLP [234], thus providing a super slow motion movie of physical phenomena resolved at the level of atoms.

Challenges remain on whether the GGA and meta-GGA DFA typically used to train MLP have the appropriate accuracy in modeling the key interactions in organocatalysis and PTC. The compilation of DFA in Table 1 shows that computationally economical GGA falls short of the standard chemical accuracy requirement. Nevertheless, as the benchmark is referenced against calculations in a vacuum, it remains to be seen if such errors will translate into the condensed or the interphase. Careful validation with experimental properties is imperative to ensure that the trained MLP is physically accurate enough for the production simulation. Encouraging results from the group of Keith indicate that with sufficiently solvated intermediates and TS structures, the difference between GGA DFA, hybrid DFA, and DLPNO-CCSD(T) is generally slight [159,235].

Despite the shortcoming of GGA and meta-GGA DFA, generating hundreds of thousands of high-quality potential energy and forces with hybrid GGA and beyond will be an astronomical task. While there have been reports of MLP built on CCSD(T) [236] or DMC data [237], these works represent the exception rather than the norm.

Technical challenges in generating a dataset for the training of MLP include k-points and plane-wave cut-off convergence. The use of empirical dispersion, critical to modeling noncovalent interaction, requires careful evaluation of k-points convergence due to the long-range nature of these corrections. Training of MLP is generally done with a smaller cell, and the resulting MLP is applied to a much larger cell. Care has to be taken to ensure that the smaller cell is large enough to contain the phonons or collective vibrations.

The selection of a representative dataset to train an MLP is challenging. Typically, MLP is trained with data generated from an ab initio MD simulation. However, only a small percentage of data from the simulation is used. One possible way to reduce the number of expensive DFT calculations is to utilize active learning. Active learning determines on the fly if a reference(usually DFT) calculation should be performed and whether this point has to be added to the training set [238–240].

Despite the challenges that remain, these works are encouraging, as with more advances in MLP design and training, scientists can couple these ML to conformational search algorithms, reaction path discovery algorithms, and molecular dynamics, thus enabling chemists to simulate system size and timescale that are closer to experiments. To end our discussion, we referred the reader to the comprehensive roadmap on ML in various subfields of chemistry for more detail [241].

3. Conclusions and Outlook

As noted by various authors, computational chemistry has advanced over the years as applied computational chemists can scale larger systems with improved hardware and software. Algorithmic advances in linear approximation to the CCSD(T) have provided the computational chemist with an additional method besides DFT to approach the problem. Nevertheless, the performance of such approximation on larger systems might not be equivalent to the smaller system on which they are benchmarked. Despite the advances in gas phase calculations, organocatalysis and phase transfer catalysis inevitably occur in the condensed phase. Solvation remains a challenge, and this aspect has been extensively discussed in this review and the references cited.

Regarding achieving a practical in silico catalyst design, while algorithms to achieve automated exploration or mapping of key points on the PES is available, these remain highly expensive in terms of computational effort, thus impeding their widespread uses. Achieving a DFT or even CCSD(T) level of accuracy with forcefield-like computational

effort will be imperative in driving practical in silico catalyst design. Houk and Liu predicted that an accurate polarizable force field could be attained by 2040 [21], while Grimme and Schreiner predicted that low-cost DFT capable of scaling to 10000 atoms would be available by 2043 [20]. Thus, if their prediction comes to fruition, and together with continued advanced in hardware, computational chemistry could be an intricate part of the methodology-developing chemists. This review ends with a discussion on ML potential, as we expect that MLP could assume the role of accurate polarizable FF or low-cost DFT.

Besides the pragmatic aspect of in silico catalyst design, the capability to perform molecular dynamics simulation with the number of atoms that approach the macroscopic scale will allow chemists access to atomic motions resolved at femtosecond resolution. The capability to perform MD simulation at such a scale can potentially provide us with the capacity to model solvents explicitly. Thus, obtaining accurate free energy of solvation can be used to validate and develop a more advanced implicit solvation model, which can drive catalyst development. The challenge of bifurcation PES can also be addressed. Together, they will bring us towards understanding and practical in silico catalyst design.

References

1. Xiang, S.-H.; Tan, B. Advances in Asymmetric Organocatalysis over the Last 10 Years. *Nat. Commun.* **2020**, *11*, 3786. [CrossRef] [PubMed]
2. Han, B.; He, X.-H.; Liu, Y.-Q.; He, G.; Peng, C.; Li, J.-L. Asymmetric Organocatalysis: An Enabling Technology for Medicinal Chemistry. *Chem. Soc. Rev.* **2021**, *50*, 1522–1586. [CrossRef]
3. Hashimoto, T.; Maruoka, K. Recent Development and Application of Chiral Phase-Transfer Catalysts. *Chem. Rev.* **2007**, *107*, 5656–5682. [CrossRef] [PubMed]
4. Qian, D.; Sun, J. Recent Progress in Asymmetric Ion-Pairing Catalysis with Ammonium Salts. *Chem.—A Eur. J.* **2019**, *25*, 3740–3751. [CrossRef]
5. Nakamura, T.; Okuno, K.; Nishiyori, R.; Shirakawa, S. Hydrogen-Bonding Catalysis of Alkyl-Onium Salts. *Chem.—An Asian J.* **2020**, *15*, 463–472. [CrossRef]
6. Zong, L.; Tan, C.H. Phase-Transfer and Ion-Pairing Catalysis of Pentanidiums and Bisguanidiniums. *Acc. Chem. Res.* **2017**, *50*, 842–856. [CrossRef]
7. Li, H.; Liu, H.; Guo, H. Recent Advances in Phosphonium Salt Catalysis. *Adv. Synth. Catal.* **2021**, *363*, 2023–2036. [CrossRef]
8. Pupo, G.; Gouverneur, V. Hydrogen Bonding Phase-Transfer Catalysis with Alkali Metal Fluorides and Beyond. *J. Am. Chem. Soc.* **2022**, *144*, 5200–5213. [CrossRef]
9. Phipps, R.J.; Hamilton, G.L.; Toste, F.D. The Progression of Chiral Anions from Concepts to Applications in Asymmetric Catalysis. *Nat. Chem.* **2012**, *4*, 603–614. [CrossRef]
10. Mąkosza, M.; Fedoryński, M. Interfacial Processes—The Key Steps of Phase Transfer Catalyzed Reactions. *Catalysts* **2020**, *10*, 1436. [CrossRef]
11. Sterling, A.J.; Zavitsanou, S.; Ford, J.; Duarte, F. Selectivity in Organocatalysis—From Qualitative to Quantitative Predictive Models. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2021**, *11*, e1518. [CrossRef]
12. Chin, Y.P.; See, N.W.; Jenkins, I.D.; Krenske, E.H. Computational Discoveries of Reaction Mechanisms: Recent Highlights and Emerging Challenges. *Org. Biomol. Chem.* **2022**, *20*, 2028–2042. [CrossRef]

13. Melnyk, N.; Iribarren, I.; Mates-Torres, E.; Trujillo, C. Theoretical Perspectives in Organocatalysis. *Chem.—A Eur. J.* **2022**, *28*, e202201570. [CrossRef] [PubMed]
14. Iribarren, I.; Garcia, M.R.; Trujillo, C. Catalyst Design within Asymmetric Organocatalysis. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2022**, *12*, e1616. [CrossRef]
15. Houk, K.N.; Cheong, P.H.Y. Computational Prediction of Small-Molecule Catalysts. *Nature* **2008**, *455*, 309–313. [CrossRef] [PubMed]
16. Peng, Q.; Duarte, F.; Paton, R.S. Computing Organic Stereoselectivity—From Concepts to Quantitative Calculations and Predictions. *Chem. Soc. Rev.* **2016**, *45*, 6093–6107. [CrossRef] [PubMed]
17. Krenske, E.H. Challenges in Predicting Stereoselectivity. In *Applied Theoretical Organic Chemistry*; WORLD SCIENTIFIC (EUROPE): Singapore, 2018; pp. 583–604. ISBN 9781786344090.
18. Chanussot, L.; Das, A.; Goyal, S.; Lavril, T.; Shuaibi, M.; Riviere, M.; Tran, K.; Heras-Domingo, J.; Ho, C.; Hu, W.; et al. Open Catalyst 2020 (OC20) Dataset and Community Challenges. *ACS Catal.* **2021**, *11*, 6059–6072. [CrossRef]
19. Tran, R.; Lan, J.; Shuaibi, M.; Wood, B.M.; Goyal, S.; Das, A.; Heras-Domingo, J.; Kolluru, A.; Rizvi, A.; Shoghi, N.; et al. The Open Catalyst 2022 (OC22) Dataset and Challenges for Oxide Electrocatalysts. *arXiv* **2022**, arXiv:2206.08917. [CrossRef]
20. Grimme, S.; Schreiner, P.R. Computational Chemistry: The Fate of Current Methods and Future Challenges. *Angew. Chem.—Int. Ed.* **2018**, *57*, 4170–4176. [CrossRef] [PubMed]
21. Houk, K.N.; Liu, F. Holy Grails for Computational Organic Chemistry and Biochemistry. *Acc. Chem. Res.* **2017**, *50*, 539–543. [CrossRef]
22. Sunoj, R.B. Coming of Age of Computational Chemistry from a Resilient Past to a Promising Future. *Isr. J. Chem.* **2022**, *62*, e202100106. [CrossRef]
23. Harvey, J.N.; Himo, F.; Maseras, F.; Perrin, L. Scope and Challenge of Computational Methods for Studying Mechanism and Reactivity in Homogeneous Catalysis. *ACS Catal.* **2019**, *9*, 6803–6813. [CrossRef]
24. Funes-Ardoiz, I.; Schoenebeck, F. Established and Emerging Computational Tools to Study Homogeneous Catalysis—From Quantum Mechanics to Machine Learning. *Chem* **2020**, *6*, 1904–1913. [CrossRef]
25. Bartlett, R.J. How and Why Coupled-Cluster Theory Became the Pre-Eminent Method in an Ab into Quantum Chemistry. In *Theory and Applications of Computational Chemistry*; Elsevier: Amsterdam, The Netherlands, 2005; pp. 1191–1221. ISBN 9780444517197.
26. Vermeeren, P.; Dalla Tiezza, M.; Wolf, M.E.; Lahm, M.E.; Allen, W.D.; Schaefer, H.F.; Hamlin, T.A.; Bickelhaupt, F.M. Pericyclic Reaction Benchmarks: Hierarchical Computations Targeting CCSDT(Q)/CBS and Analysis of DFT Performance. *Phys. Chem. Chem. Phys.* **2022**, *24*, 18028–18042. [CrossRef]
27. Seeger, Z.L.; Izgorodina, E.I. A Systematic Study of DFT Performance for Geometry Optimizations of Ionic Liquid Clusters. *J. Chem. Theory Comput.* **2020**, *16*, 6735–6753. [CrossRef]
28. Tribedi, S.; Sunoj, R.B. Molecular Insights into Chirality Transfer from Double Axially Chiral Phosphoric Acid in a Synergistic Enantioselective Intramolecular Amination. *Chem. Sci.* **2022**, *13*, 1323–1334. [CrossRef]
29. Pupo, G.; Ibba, F.; Ascough, D.M.H.; Vicini, A.C.; Ricci, P.; Christensen, K.E.; Pfeifer, L.; Morphy, J.R.; Brown, J.M.; Paton, R.S.; et al. Asymmetric Nucleophilic Fluorination under Hydrogen Bonding Phase-Transfer Catalysis. *Science* **2018**, *360*, 638–642. [CrossRef] [PubMed]
30. Zheng, H.; Sang, Y.; Houk, K.N.; Xue, X.S.; Cheng, J.P. Mechanism and Origins of Enantioselectivities in Spirobiindane-Based Hypervalent Iodine(III)-Induced Asymmetric Dearomatizing Spirolactonizations. *J. Am. Chem. Soc.* **2019**, *141*, 16046–16056. [CrossRef] [PubMed]
31. Ermanis, K.; Colgan, A.C.; Proctor, R.S.J.; Hadrys, B.W.; Phipps, R.J.; Goodman, J.M. A Computational and Experimental Investigation of the Origin of Selectivity in the Chiral Phosphoric Acid Catalyzed Enantioselective Minisci Reaction. *J. Am. Chem. Soc.* **2020**, *142*, 21091–21101. [CrossRef]
32. Kee, C.W.; Wong, M.W. Bicyclic Guanidine-Catalyzed Asymmetric Cycloaddition Reaction of Anthrones—Bifunctional Binding Modes and Origin of Stereoselectivity. *J. Org. Chem.* **2020**, *85*, 15139–15153. [CrossRef] [PubMed]
33. Ren, J.; Ban, X.; Zhang, X.; Tan, S.M.; Lee, R.; Tan, C. Kinetic and Dynamic Kinetic Resolution of Racemic Tertiary Bromides by Pentanidium-Catalyzed Phase-Transfer Azidation. *Angew. Chem. Int. Ed.* **2020**, *59*, 9055–9058. [CrossRef]
34. Laconsay, C.J.; Seguin, T.J.; Wheeler, S.E. Modulating Stereoselectivity through Electrostatic Interactions in a SPINOL-Phosphoric Acid-Catalyzed Synthesis of 2,3-Dihydroquinazolinones. *ACS Catal.* **2020**, *10*, 12292–12299. [CrossRef]
35. Hamza, A.; Sorochkina, K.; Kótai, B.; Chernichenko, K.; Berta, D.; Bolte, M.; Nieger, M.; Repo, T.; Pápai, I. Origin of Stereoselectivity in FLP-Catalyzed Asymmetric Hydrogenation of Imines. *ACS Catal.* **2020**, *10*, 14290–14301. [CrossRef]
36. Zhu, L.; Yang, H.; Wong, M.W. Asymmetric Nucleophilic Allylation of α-Chloro Glycinate via Squaramide Anion-Abstraction Catalysis: SN1 or SN2 Mechanism, or Both? *J. Org. Chem.* **2021**, *86*, 8414–8424. [CrossRef]
37. Harden, I.; Neese, F.; Bistoni, G. An Induced-Fit Model for Asymmetric Organocatalytic Reactions: A Case Study of the Activation of Olefins via Chiral Brønsted Acid Catalysts. *Chem. Sci.* **2022**, *13*, 8848–8859. [CrossRef]
38. Buttard, F.; Champagne, P.A. Binding Modes and Origins of Enantioselectivity in the Phase-Transfer-Catalyzed Conjugate Cyanation of β-Trifluoromethylated Chalcones. *ACS Catal.* **2022**, *12*, 8185–8194. [CrossRef]
39. Xu, E.Y.; Werth, J.; Roos, C.B.; Bendelsmith, A.J.; Sigman, M.S.; Knowles, R.R. Noncovalent Stabilization of Radical Intermediates in the Enantioselective Hydroamination of Alkenes with Sulfonamides. *J. Am. Chem. Soc.* **2022**, *144*, 18948–18958. [CrossRef] [PubMed]

40. Yepes, D.; Neese, F.; List, B.; Bistoni, G. Unveiling the Delicate Balance of Steric and Dispersion Interactions in Organocatalysis Using High-Level Computational Methods. *J. Am. Chem. Soc.* **2020**, *142*, 3613–3625. [CrossRef]

41. Grimme, S.; Brandenburg, J.G.; Bannwarth, C.; Hansen, A. Consistent Structures and Interactions by Density Functional Theory with Small Atomic Orbital Basis Sets. *J. Chem. Phys.* **2015**, *143*, 054107. [CrossRef]

42. Grimme, S.; Hansen, A.; Ehlert, S.; Mewes, J.M. R2SCAN-3c: A "Swiss Army Knife" Composite Electronic-Structure Method. *J. Chem. Phys.* **2021**, *154*, 064103. [CrossRef]

43. Brandenburg, J.G.; Bannwarth, C.; Hansen, A.; Grimme, S. B97-3c: A Revised Low-Cost Variant of the B97-D Density Functional Method. *J. Chem. Phys.* **2018**, *148*, 064104. [CrossRef] [PubMed]

44. Witte, J.; Neaton, J.B.; Head-Gordon, M. Effective Empirical Corrections for Basis Set Superposition Error in the Def2-SVPD Basis: GCP and DFT-C. *J. Chem. Phys.* **2017**, *146*, 234105. [CrossRef]

45. Chung, L.W.; Sameera, W.M.C.; Ramozzi, R.; Page, A.J.; Hatanaka, M.; Petrova, G.P.; Harris, T.V.; Li, X.; Ke, Z.; Liu, F.; et al. The ONIOM Method and Its Applications. *Chem. Rev.* **2015**, *115*, 5678–5796. [CrossRef]

46. Goerigk, L.; Hansen, A.; Bauer, C.; Ehrlich, S.; Najibi, A.; Grimme, S. A Look at the Density Functional Theory Zoo with the Advanced GMTKN55 Database for General Main Group Thermochemistry, Kinetics and Noncovalent Interactions. *Phys. Chem. Chem. Phys.* **2017**, *19*, 32184–32215. [CrossRef]

47. Goerigk, L.; Mehta, N. A Trip to the Density Functional Theory Zoo: Warnings and Recommendations for the User. *Aust. J. Chem.* **2019**, *72*, 563. [CrossRef]

48. Najibi, A.; Goerigk, L. The Nonlocal Kernel in van Der Waals Density Functionals as an Additive Correction: An Extensive Analysis with Special Emphasis on the B97M-V and Ωb97M-V Approaches. *J. Chem. Theory Comput.* **2018**, *14*, 5725–5738. [CrossRef]

49. Mehta, N.; Casanova-Páez, M.; Goerigk, L. Semi-Empirical or Non-Empirical Double-Hybrid Density Functionals: Which Are More Robust? *Phys. Chem. Chem. Phys.* **2018**, *20*, 23175–23194. [CrossRef]

50. Mardirossian, N.; Head-Gordon, M. Thirty Years of Density Functional Theory in Computational Chemistry: An Overview and Extensive Assessment of 200 Density Functionals. *Mol. Phys.* **2017**, *115*, 2315–2372. [CrossRef]

51. Perdew, J.P. Jacob's Ladder of Density Functional Approximations for the Exchange-Correlation Energy. In Proceedings of the AIP Conference Proceedings, Oxnard, CA, USA, 14 August 2001; AIP: College Park, MD, USA, 2001; Volume 577, pp. 1–20.

52. Peverati, R.; Truhlar, D.G. Exchange-Correlation Functional with Good Accuracy for Both Structural and Energetic Properties While Depending Only on the Density and Its Gradient. *J. Chem. Theory Comput.* **2012**, *8*, 2310–2319. [CrossRef]

53. Hao, P.; Sun, J.; Xiao, B.; Ruzsinszky, A.; Csonka, G.I.; Tao, J.; Glindmeyer, S.; Perdew, J.P. Performance of Meta-GGA Functionals on General Main Group Thermochemistry, Kinetics, and Noncovalent Interactions. *J. Chem. Theory Comput.* **2013**, *9*, 355–363. [CrossRef]

54. Taylor, M.; Yu, H.; Ho, J. Predicting Solvent Effects on S N 2 Reaction Rates: Comparison of QM/MM, Implicit, and MM Explicit Solvent Models. *J. Phys. Chem. B* **2022**, *126*, 9047–9058. [CrossRef]

55. Santra, G.; Sylvetsky, N.; Martin, J.M.L. Minimally Empirical Double-Hybrid Functionals Trained against the GMTKN55 Database: RevDSD-PBEP86-D4, RevDOD-PBE-D4, and DOD-SCAN-D4. *J. Phys. Chem. A* **2019**, *123*, 5129–5143. [CrossRef]

56. Prasad, V.K.; Pei, Z.; Edelmann, S.; Otero-De-La-Roza, A.; DiLabio, G.A. BH9, a New Comprehensive Benchmark Data Set for Barrier Heights and Reaction Energies: Assessment of Density Functional Approximations and Basis Set Incompleteness Potentials. *J. Chem. Theory Comput.* **2022**, *18*, 151–166, Erratum in *J. Chem. Theory Comput.* **2022**, *18*, 4041–4044. [CrossRef]

57. Santra, G.; Calinsky, R.; Martin, J.M.L. Benefits of Range-Separated Hybrid and Double-Hybrid Functionals for a Large and Diverse Data Set of Reaction Energies and Barrier Heights. *J. Phys. Chem. A* **2022**, *126*, 5492–5505. [CrossRef]

58. Najibi, A.; Casanova-Páez, M.; Goerigk, L. Analysis of Recent BLYP- And PBE-Based Range-Separated Double-Hybrid Density Functional Approximations for Main-Group Thermochemistry, Kinetics, and Noncovalent Interactions. *J. Phys. Chem. A* **2021**, *125*, 4026–4035. [CrossRef] [PubMed]

59. Santra, G.; Cho, M.; Martin, J.M.L. Exploring Avenues beyond Revised DSD Functionals: I. Range Separation, with XDSD as a Special Case. *J. Phys. Chem. A* **2021**, *125*, 4614–4627. [CrossRef]

60. Santra, G.; Semidalas, E.; Martin, J.M.L. Exploring Avenues beyond Revised DSD Functionals: II. Random-Phase Approximation and Scaled MP3 Corrections. *J. Phys. Chem. A* **2021**, *125*, 4628–4638. [CrossRef]

61. Mehta, N.; Martin, J.M.L. Explicitly Correlated Double-Hybrid DFT: A Comprehensive Analysis of the Basis Set Convergence on the GMTKN55 Database. *J. Chem. Theory Comput.* **2022**, *18*, 5978–5991. [CrossRef]

62. Riplinger, C.; Neese, F. An Efficient and near Linear Scaling Pair Natural Orbital Based Local Coupled Cluster Method. *J. Chem. Phys.* **2013**, *138*, 034106. [CrossRef]

63. Guo, Y.; Riplinger, C.; Becker, U.; Liakos, D.G.; Minenkov, Y.; Cavallo, L.; Neese, F. Communication: An Improved Linear Scaling Perturbative Triples Correction for the Domain Based Local Pair-Natural Orbital Based Singles and Doubles Coupled Cluster Method [DLPNO-CCSD(T)]. *J. Chem. Phys.* **2018**, *148*, 011101. [CrossRef]

64. Nagy, P.R.; Kállay, M. Optimization of the Linear-Scaling Local Natural Orbital CCSD(T) Method: Redundancy-Free Triples Correction Using Laplace Transform. *J. Chem. Phys.* **2017**, *146*, 214106. [CrossRef]

65. Nagy, P.R.; Samu, G.; Kállay, M. Optimization of the Linear-Scaling Local Natural Orbital CCSD(T) Method: Improved Algorithm and Benchmark Applications. *J. Chem. Theory Comput.* **2018**, *14*, 4193–4215. [CrossRef]

66. Neese, F. Software Update: The ORCA Program System—Version 5.0. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2022**, *12*, e1606. [CrossRef]

67. Kállay, M.; Nagy, P.R.; Mester, D.; Rolik, Z.; Samu, G.; Csontos, J.; Csóka, J.; Szabó, P.B.; Gyevi-Nagy, L.; Hégely, B.; et al. The MRCC Program System: Accurate Quantum Chemistry from Water to Proteins. *J. Chem. Phys.* **2020**, *152*, 074107. [CrossRef]

68. Riplinger, C.; Pinski, P.; Becker, U.; Valeev, E.F.; Neese, F. Sparse Maps—A Systematic Infrastructure for Reduced-Scaling Electronic Structure Methods. II. Linear Scaling Domain Based Pair Natural Orbital Coupled Cluster Theory. *J. Chem. Phys.* **2016**, *144*, 024109. [CrossRef]

69. Liakos, D.G.; Sparta, M.; Kesharwani, M.K.; Martin, J.M.L.; Neese, F. Exploring the Accuracy Limits of Local Pair Natural Orbital Coupled-Cluster Theory. *J. Chem. Theory Comput.* **2015**, *11*, 1525–1539. [CrossRef]

70. Seeger, Z.L.; Izgorodina, E.I. A DLPNO-CCSD(T) Benchmarking Study of Intermolecular Interactions of Ionic Liquids. *J. Comput. Chem.* **2022**, *43*, 106–120. [CrossRef]

71. Al-Hamdani, Y.S.; Nagy, P.R.; Zen, A.; Barton, D.; Kállay, M.; Brandenburg, J.G.; Tkatchenko, A. Interactions between Large Molecules Pose a Puzzle for Reference Quantum Mechanical Methods. *Nat. Commun.* **2021**, *12*, 3927. [CrossRef]

72. Carter-Fenk, K.; Lao, K.U.; Liu, K.Y.; Herbert, J.M. Accurate and Efficient Ab Initio Calculations for Supramolecular Complexes: Symmetry-Adapted Perturbation Theory with Many-Body Dispersion. *J. Phys. Chem. Lett.* **2019**, *10*, 2706–2714. [CrossRef]

73. Patel, P.; Wilson, A.K. Domain-Based Local Pair Natural Orbital Methods within the Correlation Consistent Composite Approach. *J. Comput. Chem.* **2020**, *41*, 800–813. [CrossRef]

74. Řezáč, J.; Šimová, L.; Hobza, P. CCSD[T] Describes Noncovalent Interactions Better than the CCSD(T), CCSD(TQ), and CCSDT Methods. *J. Chem. Theory Comput.* **2013**, *9*, 364–369. [CrossRef]

75. Klein, R.A. Single Point Calculations Using Geometries Derived at a Lower Level of Theory: Caveats to Be Observed in Using Compound Levels of Theory or 'Double-Barrelling'. *Chem. Phys. Lett.* **2006**, *423*, 413–416. [CrossRef]

76. Chin, Y.P.; Krenske, E.H. Nazarov Cyclizations Catalyzed by BINOL Phosphoric Acid Derivatives: Quantum Chemistry Struggles to Predict the Enantioselectivity. *J. Org. Chem.* **2022**, *87*, 1710–1722. [CrossRef]

77. Ribeiro, R.F.; Marenich, A.V.; Cramer, C.J.; Truhlar, D.G. Use of Solution-Phase Vibrational Frequencies in Continuum Models for the Free Energy of Solvation. *J. Phys. Chem. B* **2011**, *115*, 14556–14562. [CrossRef]

78. Grimme, S. Supramolecular Binding Thermodynamics by Dispersion-Corrected Density Functional Theory. *Chem.—A Eur. J.* **2012**, *18*, 9955–9964. [CrossRef]

79. Li, Y.P.; Gomes, J.; Sharada, S.M.; Bell, A.T.; Head-Gordon, M. Improved Force-Field Parameters for QM/MM Simulations of the Energies of Adsorption for Molecules in Zeolites and a Free Rotor Correction to the Rigid Rotor Harmonic Oscillator Model for Adsorption Enthalpies. *J. Phys. Chem. C* **2015**, *119*, 1840–1850. [CrossRef]

80. Luchini, G.; Alegre-Requena, J.V.; Funes-Ardoiz, I.; Paton, R.S. GoodVibes: Automated Thermochemistry for Heterogeneous Computational Chemistry Data. *F1000Research* **2020**, *9*, 291. [CrossRef]

81. Funes-Ardoiz, I.; Paton, R.S. GoodVibes: Version 2.0. *Zenodo* **2018**, *3*, 595246.

82. Plata, R.E.; Singleton, D.A. A Case Study of the Mechanism of Alcohol-Mediated Morita Baylis-Hillman Reactions. The Importance of Experimental Observations. *J. Am. Chem. Soc.* **2015**, *137*, 3811–3826. [CrossRef]

83. Leung, B.O.; Reid, D.L.; Armstrong, D.A.; Rauk, A. Entropies in Solution from Entropies in the Gas Phase. *J. Phys. Chem. A* **2004**, *108*, 2720–2725. [CrossRef]

84. Dub, P.A.; Poli, R. A Computational Study of Solution Equilibria of Platinum-Based Ethylene Hydroamination Catalytic Species Including Solvation and Counterion Effects: Proper Treatment of the Free Energy of Solvation. *J. Mol. Catal. A Chem.* **2010**, *324*, 89–96. [CrossRef]

85. McMullin, C.L.; Jover, J.; Harvey, J.N.; Fey, N. Accurate Modelling of Pd(0) + PhX Oxidative Addition Kinetics. *Dalt. Trans.* **2010**, *39*, 10833–10836. [CrossRef]

86. Liu, Z.; Patel, C.; Harvey, J.N.; Sunoj, R.B. Mechanism and Reactivity in the Morita-Baylis-Hillman Reaction: The Challenge of Accurate Computations. *Phys. Chem. Chem. Phys.* **2017**, *19*, 30647–30657. [CrossRef]

87. Besora, M.; Vidossich, P.; Lledós, A.; Ujaque, G.; Maseras, F. Calculation of Reaction Free Energies in Solution: A Comparison of Current Approaches. *J. Phys. Chem. A* **2018**, *122*, 1392–1399. [CrossRef]

88. Strassfeld, D.A.; Algera, R.F.; Wickens, Z.K.; Jacobsen, E.N. A Case Study in Catalyst Generality: Simultaneous, Highly-Enantioselective Brønsted- And Lewis-Acid Mechanisms in Hydrogen-Bond-Donor Catalyzed Oxetane Openings. *J. Am. Chem. Soc.* **2021**, *143*, 9585–9594. [CrossRef]

89. Pracht, P.; Bohle, F.; Grimme, S. Automated Exploration of the Low-Energy Chemical Space with Fast Quantum Chemical Methods. *Phys. Chem. Chem. Phys.* **2020**, *22*, 7169–7192. [CrossRef]

90. Iribarren, I.; Trujillo, C. Efficiency and Suitability When Exploring the Conformational Space of Phase-Transfer Catalysts. *J. Chem. Inf. Model.* **2022**, *62*, 5568–5580. [CrossRef]

91. Kee, C.W.; Wong, M.W. In Silico Design of Halogen-Bonding-Based Organocatalyst for Diels-Alder Reaction, Claisen Rearrangement, and Cope-Type Hydroamination. *J. Org. Chem.* **2016**, *81*, 7459–7470. [CrossRef]

92. Bao, J.L.; Truhlar, D.G. Variational Transition State Theory: Theoretical Framework and Recent Developments. *Chem. Soc. Rev.* **2017**, *46*, 7548–7596. [CrossRef]

93. Wigner, E. On the Quantum Correction For Thermodynamic Equilibrium. *Phys. Rev.* **1932**, *40*, 749. [CrossRef]

94. Fernández-Ramos, A.; Miller, J.A.; Klippenstein, S.J.; Truhlar, D.G. Modeling the Kinetics of Bimolecular Reactions. *Chem. Rev.* **2006**, *106*, 4518–4584. [CrossRef]

95. Anderson, T.; Kwan, E.E. Ekwan/PyQuiver: Kinetic Isotope Effect Prediction with Gaussian. Available online: https://github.com/ekwan/PyQuiver (accessed on 1 February 2023).

96. Schreiner, P.R.; Reisenauer, H.P.; Ley, D.; Gerbig, D.; Wu, C.H.; Allen, W.D. Methylhydroxycarbene: Tunneling Control of a Chemical Reaction. *Science* **2011**, *332*, 1300–1303. [CrossRef]

97. Schreiner, P.R. Tunneling Control of Chemical Reactions: The Third Reactivity Paradigm. *J. Am. Chem. Soc.* **2017**, *139*, 15276–15283. [CrossRef]

98. Pu, J.; Gao, J.; Truhlar, D.G. Multidimensional Tunneling, Recrossing, and the Transmission Coefficient for Enzymatic Reactions. *Chem. Rev.* **2006**, *106*, 3140–3169. [CrossRef]

99. Da Silva, P.L.; Guimarães, L.; Pliego, J.R. Revisiting the Mechanism of Neutral Hydrolysis of Esters: Water Autoionization Mechanisms with Acid or Base Initiation Pathways. *J. Phys. Chem. B* **2013**, *117*, 6487–6497. [CrossRef]

100. Kozuch, S. A Refinement of Everyday Thinking: The Energetic Span Model for Kinetic Assessment of Catalytic Cycles. *Wiley Interdisc. Rev. Comput. Mol. Sci.* **2012**, *2*, 795–815. [CrossRef]

101. Sumiya, Y.; Maeda, S. Rate Constant Matrix Contraction Method for Systematic Analysis of Reaction Path Networks. *Chem. Lett.* **2020**, *49*, 553–564. [CrossRef]

102. Garay-Ruiz, D.; Bo, C. Revisiting Catalytic Cycles: A Broader View through the Energy Span Model. *ACS Catal.* **2020**, *10*, 12627–12635. [CrossRef]

103. Cho, B.; Tan, C.-H.; Wong, M.W. Origin of Asymmetric Induction in Bicyclic Guanidine-Catalyzed Thio-Michael Reaction: A Bifunctional Mode of Lewis Acid-Brønsted Acid Activation. *J. Org. Chem.* **2012**, *77*, 6553–6562. [CrossRef]

104. Rackauckas, C.; Nie, Q. DifferentialEquations.Jl—A Performant and Feature-Rich Ecosystem for Solving Differential Equations in Julia. *J. Open Res. Softw.* **2017**, *5*, 15. [CrossRef]

105. Bezanson, J.; Edelman, A.; Karpinski, S.; Shah, V.B. Julia: A Fresh Approach to Numerical Computing. *SIAM Rev.* **2017**, *59*, 65–98. [CrossRef]

106. Wang, B.; Cao, Z. Hydration of Carbonyl Groups: The Labile H3O+ Ion as an Intermediate Modulated by the Surrounding Water Molecules. *Angew. Chem.—Int. Ed.* **2011**, *50*, 3266–3270. [CrossRef]

107. Offei-Danso, A.; Hassanali, A.; Rodriguez, A. High-Dimensional Fluctuations in Liquid Water: Combining Chemical Intuition with Unsupervised Learning. *J. Chem. Theory Comput.* **2022**, *18*, 3136–3150. [CrossRef]

108. Pliego, J.R.; Riveros, J.M. New Insights on Reaction Pathway Selectivity Promoted by Crown Ether Phase-Transfer Catalysis: Model Ab Initio Calculations of Nucleophilic Fluorination. *J. Mol. Catal. A Chem.* **2012**, *363*, 489–494. [CrossRef]

109. Carvalho, N.F.; Pliego, J.R. Theoretical Design and Calculation of a Crown Ether Phase-Transfer-Catalyst Scaffold for Nucleophilic Fluorination Merging Two Catalytic Concepts. *J. Org. Chem.* **2016**, *81*, 8455–8463. [CrossRef]

110. Sameera, W.M.C.; Maeda, S.; Morokuma, K. Computational Catalysis Using the Artificial Force Induced Reaction Method. *Acc. Chem. Res.* **2016**, *49*, 763–773. [CrossRef]

111. Dewyer, A.L.; Argüelles, A.J.; Zimmerman, P.M. Methods for Exploring Reaction Space in Molecular Systems. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2018**, *8*, e1354. [CrossRef]

112. Simm, G.N.; Vaucher, A.C.; Reiher, M. Exploration of Reaction Pathways and Chemical Transformation Networks. *J. Phys. Chem. A* **2019**, *123*, 385–399. [CrossRef]

113. Unsleber, J.P.; Reiher, M. The Exploration of Chemical Reaction Networks. *Annu. Rev. Phys. Chem.* **2020**, *71*, 121–142. [CrossRef]

114. Miller, J.A.; Sivaramakrishnan, R.; Tao, Y.; Goldsmith, C.F.; Burke, M.P.; Jasper, A.W.; Hansen, N.; Labbe, N.J.; Glarborg, P.; Zádor, J. Combustion Chemistry in the Twenty-First Century: Developing Theory-Informed Chemical Kinetics Models. *Prog. Energy Combust. Sci.* **2021**, *83*, 100886. [CrossRef]

115. Ye, X.; Moeljadi, A.M.P.; Chin, K.F.; Hirao, H.; Zong, L.; Tan, C.H. Enantioselective Sulfoxidation Catalyzed by a Bisguanidinium Diphosphatobisperoxotungstate Ion Pair. *Angew. Chem.—Int. Ed.* **2016**, *55*, 7101–7105. [CrossRef]

116. Ibba, F.; Pupo, G.; Thompson, A.L.; Brown, J.M.; Claridge, T.D.W.; Gouverneur, V. Impact of Multiple Hydrogen Bonds with Fluoride on Catalysis: Insight from NMR Spectroscopy. *J. Am. Chem. Soc.* **2020**, *142*, jacs.0c09832. [CrossRef]

117. Wang, J.; Horwitz, M.A.; Dürr, A.B.; Ibba, F.; Pupo, G.; Gao, Y.; Ricci, P.; Christensen, K.E.; Pathak, T.P.; Claridge, T.D.W.; et al. Asymmetric Azidation under Hydrogen Bonding Phase-Transfer Catalysis: A Combined Experimental and Computational Study. *J. Am. Chem. Soc.* **2022**, *144*, 4572–4584. [CrossRef]

118. Bryliakov, K.P. Dynamic Nonlinear Effects in Asymmetric Catalysis. *ACS Catal.* **2019**, *9*, 5418–5438. [CrossRef]

119. Mayer, L.C.; Heitsch, S.; Trapp, O. Nonlinear Effects in Asymmetric Catalysis by Design: Concept, Synthesis, and Applications. *Acc. Chem. Res.* **2022**, *55*, 3345–3361. [CrossRef]

120. Kalek, M.; Fu, G.C. Caution in the Use of Nonlinear Effects as a Mechanistic Tool for Catalytic Enantioconvergent Reactions: Intrinsic Negative Nonlinear Effects in the Absence of Higher-Order Species. *J. Am. Chem. Soc.* **2017**, *139*, 4225–4229. [CrossRef]

121. Geiger, Y.; Achard, T.; Maisse-François, A.; Bellemin-Laponnaz, S. Absence of Non-Linear Effects Despite Evidence for Catalyst Aggregation. *Eur. J. Org. Chem.* **2021**, *2021*, 2916–2922. [CrossRef]

122. Satyanarayana, T.; Abraham, S.; Kagan, H.B. Nonlinear Effects in Asymmetric Catalysis. *Angew. Chem.—Int. Ed.* **2009**, *48*, 456–494. [CrossRef]

123. Liao, L.; Xu, X.; Ji, J.; Zhao, X. Asymmetric Intermolecular Iodinative Difunctionalization of Allylic Sulfonamides Enabled by Organosulfide Catalysis: Modular Entry to Iodinated Chiral Molecules. *J. Am. Chem. Soc.* **2022**, *144*, 16490–16501. [CrossRef]

124. Guo, W.; Wang, M.; Han, Z.; Huang, H.; Sun, J. Organocatalytic Asymmetric Synthesis of α-Amino Esters from Sulfoxonium Ylides. *Chem. Sci.* **2021**, *12*, 11191–11196. [CrossRef]

125. Jessen, N.I.; Bertuzzi, G.; Bura, M.; Skipper, M.L.; Jørgensen, K.A. Enantioselective Construction of the Cycl[3.2.2]Azine Core via Organocatalytic [12 + 2] Cycloadditions. *J. Am. Chem. Soc.* **2021**, *143*, 6140–6151. [CrossRef]

126. Ali, C.; Blackmond, D.G.; Burés, J. Kinetic Rationalization of Nonlinear Effects in Asymmetric Catalytic Cascade Reactions under Curtin-Hammett Conditions. *ACS Catal.* **2022**, *16*, 5776–5785. [CrossRef]

127. Hase, W.L.; Song, K.; Gordon, M.S. Direct Dynamics Simulations. *Comput. Sci. Eng.* **2003**, *5*, 36–44. [CrossRef]

128. Ess, D.H.; Wheeler, S.E.; Iafe, R.G.; Xu, L.; Çelebi-Ölçüm, N.; Houk, K.N. Bifurcations on Potential Energy Surfaces of Organic Reactions. *Angew. Chem.—Int. Ed.* **2008**, *47*, 7592–7601. [CrossRef]

129. Hare, S.R.; Tantillo, D.J. Post-Transition State Bifurcations Gain Momentum-Current State of the Field. *Pure Appl. Chem.* **2017**, *89*, 679–698. [CrossRef]

130. Carpenter, B.K. Energy Disposition in Reactive Intermediates. *Chem. Rev.* **2013**, *113*, 7265–7286. [CrossRef]

131. Collins, P.; Carpenter, B.K.; Ezra, G.S.; Wiggins, S. Nonstatistical Dynamics on Potentials Exhibiting Reaction Path Bifurcations and Valley-Ridge Inflection Points. *J. Chem. Phys.* **2013**, *139*, 154108. [CrossRef]

132. Sun, L.; Song, K.; Hase, W.L. A SN2 Reaction That Avoids Its Deep Potential Energy Minimum. *Science* **2002**, *296*, 875–878. [CrossRef]

133. Proenza, Y.G.; de Souza, M.A.F.; Longo, R.L. Dynamical Bifurcation in Gas-Phase XH−+ CH3Y SN2 Reactions: The Role of Energy Flow and Redistribution in Avoiding the Minimum Energy Path. *Chem.—A Eur. J.* **2016**, *22*, 16220–16229. [CrossRef]

134. Roytman, V.A.; Jin, S.; Nguyen, V.T.; Nguyen, V.D.; Haug, G.C.; Larionov, O.V.; Singleton, D.A. Bond Memory in Dynamically Determined Stereoselectivity. *J. Am. Chem. Soc.* **2020**, *142*, 85–88. [CrossRef]

135. Lee, S.; Goodman, J.M. Rapid Route-Finding for Bifurcating Organic Reactions. *J. Am. Chem. Soc.* **2020**, *142*, 9210–9219. [CrossRef]

136. Bharadwaz, P.; Maldonado-Domínguez, M.; Srnec, M. Bifurcating Reactions: Distribution of Products from Energy Distribution in a Shared Reactive Mode. *Chem. Sci.* **2021**, *12*, 12682–12694. [CrossRef]

137. Peterson, T.H.; Carpenter, B.K. Estimation of Dynamic Effects on Product Ratios by Vectorial Decomposition of a Reaction Coordinate. Application to Thermal Nitrogen Loss from Bicyclic Azo Compounds. *J. Am. Chem. Soc.* **1992**, *114*, 766–767. [CrossRef]

138. Yang, Z.; Dong, X.; Yu, Y.; Yu, P.; Li, Y.; Jamieson, C.; Houk, K.N. Relationships between Product Ratios in Ambimodal Pericyclic Reactions and Bond Lengths in Transition Structures. *J. Am. Chem. Soc.* **2018**, *140*, 3061–3067. [CrossRef]

139. Samanta, D.; Rana, A.; Schmittel, M. Quantification of Nonstatistical Dynamics in an Intramolecular Diels-Alder Cyclization without Trajectory Computation. *J. Org. Chem.* **2014**, *79*, 2368–2376. [CrossRef]

140. Ang, S.J.; Wang, W.; Schwalbe-Koda, D.; Axelrod, S.; Gómez-Bombarelli, R. Active Learning Accelerates Ab Initio Molecular Dynamics on Reactive Energy Surfaces. *Chem* **2021**, *7*, 738–751. [CrossRef]

141. Maldonado, A.M.; Basdogan, Y.; Berryman, J.T.; Rempe, S.B.; Keith, J.A. First-Principles Modeling of Chemistry in Mixed Solvents: Where to Go from Here? *J. Chem. Phys.* **2020**, *152*, 130902. [CrossRef]

142. Norjmaa, G.; Ujaque, G.; Lledós, A. Beyond Continuum Solvent Models in Computational Homogeneous Catalysis. *Top. Catal.* **2022**, *65*, 118–140. [CrossRef]

143. Das, M.; Gogoi, A.R.; Sunoj, R.B. Molecular Insights on Solvent Effects in Organic Reactions as Obtained through Computational Chemistry Tools. *J. Org. Chem.* **2022**, *87*, 1630–1640. [CrossRef]

144. Burrows, C.J.; Harper, J.B.; Sander, W.; Tantillo, D.J. Solvation Effects in Organic Chemistry. *J. Org. Chem.* **2022**, *87*, 1599–1601. [CrossRef]

145. Ma, T.; Fu, X.; Kee, C.W.; Zong, L.; Pan, Y.; Huang, K.W.; Tan, C.H. Pentanidium-Catalyzed Enantioselective Phase-Transfer Conjugate Addition Reactions. *J. Am. Chem. Soc.* **2011**, *133*, 2828–2831. [CrossRef]

146. Tomasi, J.; Mennucci, B.; Cammi, R. Quantum Mechanical Continuum Solvation Models. *Chem. Rev.* **2005**, *105*, 2999–3093. [CrossRef]

147. Ho, J.; Klamt, A.; Coote, M.L. Comment on the Correct Use of Continuum Solvent Models. *J. Phys. Chem. A* **2010**, *114*, 13442–13444. [CrossRef]

148. 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. [CrossRef]

149. Klamt, A.; Mennucci, B.; Tomasi, J.; Barone, V.; Curutchet, C.; Orozco, M.; Luque, F.J. On the Performance of Continuum Solvation Methods. A Comment on "Universal Approaches to Solvation Modeling". *Acc. Chem. Res.* **2009**, *42*, 489–492. [CrossRef]

150. Xu, L.; Coote, M.L. Methods to Improve the Calculations of Solvation Model Density Solvation Free Energies and Associated Aqueous p Ka Values: Comparison between Choosing an Optimal Theoretical Level, Solute Cavity Scaling, and Using Explicit Solvent Molecules. *J. Phys. Chem. A* **2019**, *123*, 7430–7438. [CrossRef]

151. Hénin, J.; Lelièvre, T.; Shirts, M.R.; Valsson, O.; Delemotte, L. Enhanced Sampling Methods for Molecular Dynamics Simulations. *arXiv* **2022**, arXiv:2202.04164. [CrossRef]

152. Tawa, G.J.; Martin, R.L.; Pratt, L.R.; Russo, T.V. Solvation Free Energy Calculations Using a Continuum Dielectric Model for the Solvent and Gradient-Corrected Density Functional Theory for the Solute. *J. Phys. Chem.* **1996**, *100*, 1515–1523. [CrossRef]
153. Pliego, J.R.; Riveros, J.M. The Cluster-Continuum Model for the Calculation of the Solvation Free Energy of Ionic Species. *J. Phys. Chem. A* **2001**, *105*, 7241–7247. [CrossRef]
154. Tawa, G.J.; Topol, I.A.; Burt, S.K.; Caldwell, R.A.; Rashin, A.A. Calculation of the Aqueous Solvation Free Energy of the Proton. *J. Chem. Phys.* **1998**, *109*, 4852. [CrossRef]
155. Bryantsev, V.S.; Diallo, M.S.; Goddard, W.A. Calculation of Solvation Free Energies of Charged Solutes Using Mixed Cluster/Continuum Models. *J. Phys. Chem. B* **2008**, *112*, 9709–9719. [CrossRef]
156. Wu, W.; Kieffer, J. New Hybrid Method for the Calculation of the Solvation Free Energy of Small Molecules in Aqueous Solutions. *J. Chem. Theory Comput.* **2019**, *15*, 371–381. [CrossRef]
157. Spicher, S.; Plett, C.; Pracht, P.; Hansen, A.; Grimme, S. Automated Molecular Cluster Growing for Explicit Solvation by Efficient Force Field and Tight Binding Methods. *J. Chem. Theory Comput.* **2022**, *18*, 3189. [CrossRef]
158. Zhang, J.; Dolg, M. ABCluster: The Artificial Bee Colony Algorithm for Cluster Global Optimization. *Phys. Chem. Chem. Phys.* **2015**, *17*, 24173–24181. [CrossRef]
159. Basdogan, Y.; Keith, J.A. A Paramedic Treatment for Modeling Explicitly Solvated Chemical Reaction Mechanisms. *Chem. Sci.* **2018**, *9*, 5341–5346. [CrossRef]
160. Gunaydin, H.; Houk, K.N. Molecular Dynamics Prediction of the Mechanism of Ester Hydrolysis in Water. *J. Am. Chem. Soc.* **2008**, *130*, 15232–15233. [CrossRef]
161. Ali, H.S.; Higham, J.; De Visser, S.P.; Henchman, R.H. Comparison of Free-Energy Methods to Calculate the Barriers for the Nucleophilic Substitution of Alkyl Halides by Hydroxide. *J. Phys. Chem. B* **2020**, *124*, 6835–6842. [CrossRef]
162. Klamt, A.; Eckert, F.; Arlt, W. COSMO-RS: An Alternative to Simulation for Calculating Thermodynamic Properties of Liquid Mixtures. *Annu. Rev. Chem. Biomol. Eng.* **2010**, *1*, 101–122. [CrossRef]
163. Nishihara, S.; Otani, M. Hybrid Solvation Models for Bulk, Interface, and Membrane: Reference Interaction Site Methods Coupled with Density Functional Theory. *Phys. Rev. B* **2017**, *96*, 115429. [CrossRef]
164. Maldonado, A.M.; Hagiwara, S.; Choi, T.H.; Eckert, F.; Schwarz, K.; Sundararaman, R.; Otani, M.; Keith, J.A. Quantifying Uncertainties in Solvation Procedures for Modeling Aqueous Phase Reaction Mechanisms. *J. Phys. Chem. A* **2021**, *125*, 154–164. [CrossRef]
165. Kee, C.W.; Wong, M.W. Pentanidium-Catalyzed Asymmetric Phase-Transfer Conjugate Addition: Prediction of Stereoselectivity via DFT Calculations and Docking Sampling of Transition States, and Origin of Stereoselectivity. *Aust. J. Chem.* **2016**, *69*, 983. [CrossRef]
166. Chief Elk, J.; Benjamin, I. β-Cyclodextrin at the Water/1-Bromobutane Interface: Molecular Insight into Reverse Phase Transfer Catalysis. *Langmuir* **2015**, *31*, 5086–5092. [CrossRef]
167. Karnes, J.J.; Benjamin, I. Structure and Dynamics of Host/Guest Complexation at the Liquid/Liquid Interface: Implications for Inverse Phase Transfer Catalysis. *J. Phys. Chem. C* **2017**, *121*, 4999–5011. [CrossRef]
168. Karnes, J.J.; Benjamin, I. SN2 Reaction Rate Enhancement by β-Cyclodextrin at the Liquid/Liquid Interface. *J. Phys. Chem. C* **2017**, *121*, 19209–19217. [CrossRef]
169. Uyeda, C.; Jacobsen, E.N. Transition-State Charge Stabilization through Multiple Non-Covalent Interactions in the Guanidinium-Catalyzed Enantioselective Claisen Rearrangement. *J. Am. Chem. Soc.* **2011**, *133*, 5062–5075. [CrossRef]
170. Xue, H.; Jiang, D.; Jiang, H.; Kee, C.W.; Hirao, H.; Nishimura, T.; Wong, M.W.; Tan, C.H. Mechanistic Insights into Bicyclic Guanidine-Catalyzed Reactions from Microscopic and Macroscopic Perspectives. *J. Org. Chem.* **2015**, *80*, 5745–5752. [CrossRef]
171. Zotova, N.; Franzke, A.; Armstrong, A.; Blackmond, D.G. Clarification of the Role of Water in Proline-Mediated Aldol Reactions. *J. Am. Chem. Soc.* **2007**, *129*, 15100–15101. [CrossRef]
172. Zotova, N.; Broadbelt, L.J.; Armstrong, A.; Blackmond, D.G. Kinetic and Mechanistic Studies of Proline-Mediated Direct Intermolecular Aldol Reactions. *Bioorg. Med. Chem. Lett.* **2009**, *19*, 3934–3937. [CrossRef]
173. Orlandi, M.; Ceotto, M.; Benaglia, M. Kinetics versus Thermodynamics in the Proline Catalyzed Aldol Reaction. *Chem. Sci.* **2016**, *7*, 5421–5427. [CrossRef]
174. Zuend, S.J.; Jacobsen, E.N. Mechanism of Amido-Thiourea Catalyzed Enantioselective Imine Hydrocyanation: Transition State Stabilization via Multiple Non-Covalent Interactions. *J. Am. Chem. Soc.* **2009**, *131*, 15358–15374. [CrossRef]
175. Meyer, M.P. New Applications of Isotope Effects in the Determination of Organic Reaction Mechanisms. In *Advances in Physical Organic Chemistry*; Academic Press: Cambridge, MA, USA, 2012; Volume 46, pp. 57–120. ISBN 9780123984845.
176. Dale, H.J.A.; Leach, A.G.; Lloyd-Jones, G.C. Heavy-Atom Kinetic Isotope Effects: Primary Interest or Zero Point? *J. Am. Chem. Soc.* **2021**, *143*, 21079–21099. [CrossRef]
177. Zhu, H.; Clemente, F.R.; Houk, K.N.; Meyer, M.P. Rate Limiting Step Precedes C-C Bond Formation in the Archetypical Proline-Catalyzed Intramolecular Aldol Reaction. *J. Am. Chem. Soc.* **2009**, *131*, 1632–1633. [CrossRef]
178. Kwan, E.E.; Park, Y.; Besser, H.A.; Anderson, T.L.; Jacobsen, E.N. Sensitive and Accurate 13c Kinetic Isotope Effect Measurements Enabled by Polarization Transfer. *J. Am. Chem. Soc.* **2017**, *139*, 43–46. [CrossRef]
179. Park, Y.; Harper, K.C.; Kuhl, N.; Kwan, E.E.; Liu, R.Y.; Jacobsen, E.N. Macrocyclic Bis-Thioureas Catalyze Stereospecific Glycosylation Reactions. *Science* **2017**, *355*, 162–166. [CrossRef]

180. Bigeleisen, J.; Mayer, M.G. Calculation of Equilibrium Constants for Isotopic Exchange Reactions. *J. Chem. Phys.* **1947**, *15*, 261–267. [CrossRef]

181. Chandra Mallojjala, S.; Sarkar, R.; Karugu, R.W.; Manna, M.S.; Ray, S.; Mukherjee, S.; Hirschi, J.S. Mechanism and Origin of Remote Stereocontrol in the Organocatalytic Enantioselective Formal C(Sp2)-H Alkylation Using Nitroalkanes as Alkylating Agents. *J. Am. Chem. Soc.* **2022**, *144*, 17399–17406. [CrossRef]

182. Manna, M.S.; Mukherjee, S. Organocatalytic Enantioselective Formal C(Sp2)-H Alkylation. *J. Am. Chem. Soc.* **2015**, *137*, 130–133. [CrossRef]

183. Ryu, H.; Park, J.Y.; Kim, H.K.; Park, J.Y.; Kim, S.T.; Baik, M.H. Pitfalls in Computational Modeling of Chemical Reactions and How to Avoid Them. *Organometallics* **2018**, *37*, 3228–3239. [CrossRef]

184. Proctor, R.S.J.; Davis, H.J.; Phipps, R.J. Catalytic Enantioselective Minisci-Type Addition to Heteroarenes. *Science* **2018**, *360*, 419–422. [CrossRef]

185. Shen, J.; Nguyen, T.T.; Goh, Y.-P.; Ye, W.; Fu, X.; Xu, J.; Tan, C.-H. Chiral Bicyclic Guanidine-Catalyzed Enantioselective Reactions of Anthrones. *J. Am. Chem. Soc.* **2006**, *128*, 13692–13693. [CrossRef]

186. Huang, D.; Li, X.; Xu, F.; Li, L.; Lin, X. Highly Enantioselective Synthesis of Dihydroquinazolinones Catalyzed by SPINOL-Phosphoric Acids. *ACS Catal.* **2013**, *3*, 2244–2247. [CrossRef]

187. Van Zeist, W.J.; Bickelhaupt, F.M. The Activation Strain Model of Chemical Reactivity. *Org. Biomol. Chem.* **2010**, *8*, 3118–3127. [CrossRef]

188. Bickelhaupt, F.M.; Houk, K.N. Analyzing Reaction Rates with the Distortion/Interaction-Activation Strain Model. *Angew. Chem.—Int. Ed.* **2017**, *56*, 10070–10086. [CrossRef]

189. Vermeeren, P.; van der Lubbe, S.C.C.; Fonseca Guerra, C.; Bickelhaupt, F.M.; Hamlin, T.A. Understanding Chemical Reactivity Using the Activation Strain Model. *Nat. Protoc.* **2020**, *15*, 649–667. [CrossRef] [PubMed]

190. Schneider, W.B.; Bistoni, G.; Sparta, M.; Saitow, M.; Riplinger, C.; Auer, A.A.; Neese, F. Decomposition of Intermolecular Interaction Energies within the Local Pair Natural Orbital Coupled Cluster Framework. *J. Chem. Theory Comput.* **2016**, *12*, 4778–4792. [CrossRef]

191. Gatzenmeier, T.; Van Gemmeren, M.; Xie, Y.; Höfler, D.; Leutzsch, M.; List, B. Asymmetric Lewis Acid Organocatalysis of the Diels-Alder Reaction by a Silylated C-H Acid. *Science* **2016**, *351*, 949–952. [CrossRef]

192. Johnson, E.R.; Keinan, S.; Mori-Sánchez, P.; Contreras-García, J.; Cohen, A.J.; Yang, W.; Mori-Sánchez, P.; Contreras-García, J.; Cohen, A.J.; Yang, W. Revealing Noncovalent Interactions. *J. Am. Chem. Soc.* **2010**, *132*, 6498–6506. [CrossRef]

193. Boto, R.A.; Peccati, F.; Laplaza, R.; Quan, C.; Carbone, A.; Piquemal, J.P.; Maday, Y.; Contreras-GarcÍa, J. NCIPLOT4: Fast, Robust, and Quantitative Analysis of Noncovalent Interactions. *J. Chem. Theory Comput.* **2020**, *16*, 4150–4158. [CrossRef]

194. Lu, T.; Chen, F. Multiwfn: A Multifunctional Wavefunction Analyzer. *J. Comput. Chem.* **2012**, *33*, 580–592. [CrossRef]

195. Lefebvre, C.; Rubez, G.; Khartabil, H.; Boisson, J.C.; Contreras-García, J.; Hénon, E. Accurately Extracting the Signature of Intermolecular Interactions Present in the NCI Plot of the Reduced Density Gradient: Versus Electron Density. *Phys. Chem. Chem. Phys.* **2017**, *19*, 17928–17936. [CrossRef]

196. Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graph.* **1996**, *14*, 33–38. [CrossRef]

197. Facco, E.; D'Errico, M.; Rodriguez, A.; Laio, A. Estimating the Intrinsic Dimension of Datasets by a Minimal Neighborhood Information. *Sci. Rep.* **2017**, *7*, 12140. [CrossRef] [PubMed]

198. Melville, J.; Hargis, C.; Davenport, M.T.; Hamilton, R.S.; Ess, D.H. Machine Learning Analysis of Dynamic-Dependent Bond Formation in Trajectories with Consecutive Transition States. *J. Phys. Org. Chem.* **2022**, *35*, e4405. [CrossRef]

199. Hueffel, J.A.; Sperger, T.; Funes-Ardoiz, I.; Ward, J.S.; Rissanen, K.; Schoenebeck, F. Accelerated Dinuclear Palladium Catalyst Identification through Unsupervised Machine Learning. *Science* **2021**, *374*, 1134–1140. [CrossRef]

200. Strieth-Kalthoff, F.; Sandfort, F.; Segler, M.H.S.; Glorius, F. Machine Learning the Ropes: Principles, Applications and Directions in Synthetic Chemistry. *Chem. Soc. Rev.* **2020**, *49*, 6154–6168. [CrossRef]

201. Liu, Y.; Esan, O.C.; Pan, Z.; An, L. Machine Learning for Advanced Energy Materials. *Energy AI* **2021**, *3*, 100049. [CrossRef]

202. Crawford, J.M.; Kingston, C.; Toste, F.D.; Sigman, M.S. Data Science Meets Physical Organic Chemistry. *Acc. Chem. Res.* **2021**, *54*, 3136–3148. [CrossRef]

203. Karthikeyan, A.; Priyakumar, U.D. Artificial Intelligence: Machine Learning for Chemical Sciences. *J. Chem. Sci.* **2022**, *134*, 1–20. [CrossRef]

204. Bender, A.; Schneider, N.; Segler, M.; Patrick Walters, W.; Engkvist, O.; Rodrigues, T. Evaluation Guidelines for Machine Learning Tools in the Chemical Sciences. *Nat. Rev. Chem.* **2022**, *6*, 428–442. [CrossRef]

205. Żurański, A.M.; Martinez Alvarado, J.I.; Shields, B.J.; Doyle, A.G. Predicting Reaction Yields via Supervised Learning. *Acc. Chem. Res.* **2021**, *54*, 1856–1865. [CrossRef]

206. Jorner, K.; Tomberg, A.; Bauer, C.; Sköld, C.; Norrby, P.O. Organic Reactivity from Mechanism to Machine Learning. *Nat. Rev. Chem.* **2021**, *5*, 240–255. [CrossRef]

207. Williams, W.L.; Zeng, L.; Gensch, T.; Sigman, M.S.; Doyle, A.G.; Anslyn, E.V. The Evolution of Data-Driven Modeling in Organic Chemistry. *ACS Cent. Sci.* **2021**, *7*, 1622–1637. [CrossRef]

208. Ahneman, D.T.; Estrada, J.G.; Lin, S.; Dreher, S.D.; Doyle, A.G. Predicting Reaction Performance in C–N Cross-Coupling Using Machine Learning. *Science* **2018**, *360*, 186–190. [CrossRef]

209. Santiago, C.B.; Guo, J.Y.; Sigman, M.S. Predictive and Mechanistic Multivariate Linear Regression Models for Reaction Development. *Chem. Sci.* **2018**, *9*, 2398–2412. [CrossRef]

210. Singh, S.; Pareek, M.; Changotra, A.; Banerjee, S.; Bhaskararao, B.; Balamurugan, P.; Sunoj, R.B. A Unified Machine-Learning Protocol for Asymmetric Catalysis as a Proof of Concept Demonstration Using Asymmetric Hydrogenation. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 1339–1345. [CrossRef]

211. Xu, L.C.; Zhang, S.Q.; Li, X.; Tang, M.J.; Xie, P.P.; Hong, X. Towards Data-Driven Design of Asymmetric Hydrogenation of Olefins: Database and Hierarchical Learning. *Angew. Chem.—Int. Ed.* **2021**, *60*, 22804–22811. [CrossRef]

212. Singh, S.; Sunoj, R.B. A Transfer Learning Protocol for Chemical Catalysis Using a Recurrent Neural Network Adapted from Natural Language Processing. *Digit. Discov.* **2022**, *1*, 303–312. [CrossRef]

213. Zhu, X.Y.; Ran, C.K.; Wen, M.; Guo, G.L.; Liu, Y.; Liao, L.L.; Li, Y.Z.; Li, M.L.; Yu, D.G. Prediction of Multicomponent Reaction Yields Using Machine Learning. *Chin. J. Chem.* **2021**, *39*, 3231–3237. [CrossRef]

214. Owen, B.; Wheelhouse, K.; Figueredo, G.; Özcan, E.; Woodward, S. Machine Learnt Patterns in Rhodium-Catalysed Asymmetric Michael Addition Using Chiral Diene Ligands. *Results Chem.* **2022**, *4*, 100379. [CrossRef]

215. Zahrt, A.F.; Henle, J.J.; Rose, B.T.; Wang, Y.; Darrow, W.T.; Denmark, S.E. Prediction of Higher-Selectivity Catalysts by Computer-Driven Workflow and Machine Learning. *Science* **2019**, *363*, eaau5631. [CrossRef]

216. Rinehart, N.I.; Zahrt, A.F.; Henle, J.J.; Denmark, S.E. Dreams, False Starts, Dead Ends, and Redemption: A Chronicle of the Evolution of a Chemoinformatic Workflow for the Optimization of Enantioselective Catalysts. *Acc. Chem. Res.* **2021**, *54*, 2041–2054. [CrossRef]

217. Zahrt, A.F.; Henle, J.J.; Denmark, S.E. Cautionary Guidelines for Machine Learning Studies with Combinatorial Datasets. *ACS Comb. Sci.* **2020**, *22*, 586–591. [CrossRef]

218. Chuang, K.V.; Keiser, M.J. Comment on "Predicting Reaction Performance in C–N Cross-Coupling Using Machine Learning". *Science* **2018**, *362*, eaat8603. [CrossRef]

219. Estrada, J.G.; Ahneman, D.T.; Sheridan, R.P.; Dreher, S.D.; Doyle, A.G. Response to Comment on "Predicting Reaction Performance in C–N Cross-Coupling Using Machine Learning". *Science* **2018**, *362*, 8763. [CrossRef]

220. Sandfort, F.; Strieth-Kalthoff, F.; Kühnemund, M.; Beecks, C.; Glorius, F. A Structure-Based Platform for Predicting Chemical Reactivity. *Chem* **2020**, *6*, 1379–1390. [CrossRef]

221. Yang, Q.; Liu, Y.; Cheng, J.; Li, Y.; Liu, S.; Duan, Y.; Zhang, L.; Luo, S. An Ensemble Structure and Physicochemical (SPOC) Descriptor for Machine-Learning Prediction of Chemical Reaction and Molecular Properties. *ChemPhysChem* **2022**, *23*, e202200255. [CrossRef]

222. Behler, J. Four Generations of High-Dimensional Neural Network Potentials. *Chem. Rev.* **2021**, *121*, 10037–10072. [CrossRef]

223. Unke, O.T.; Chmiela, S.; Sauceda, H.E.; Gastegger, M.; Poltavsky, I.; Schütt, K.T.; Tkatchenko, A.; Müller, K.R. Machine Learning Force Fields. *Chem. Rev.* **2021**, *121*, 10142–10186. [CrossRef]

224. Pinheiro, M.; Ge, F.; Ferré, N.; Dral, P.O.; Barbatti, M. Choosing the Right Molecular Machine Learning Potential. *Chem. Sci.* **2021**, *12*, 14396–14413. [CrossRef]

225. Devereux, C.; Smith, J.S.; Davis, K.K.; Barros, K.; Zubatyuk, R.; Isayev, O.; Roitberg, A.E. Extending the Applicability of the ANI Deep Learning Molecular Potential to Sulfur and Halogens. *J. Chem. Theory Comput.* **2020**, *16*, 4192–4202. [CrossRef]

226. Gasteiger, J.; Groß, J.; Günnemann, S. Directional Message Passing for Molecular Graphs. *arXiv* **2020**, arXiv:2003.03123. [CrossRef]

227. Klicpera, J.; Becker, F.; Günnemann, S. GemNet: Universal Directional Graph Neural Networks for Molecules. *Adv. Neural Inf. Process. Syst.* **2021**, *9*, 6790–6802. [CrossRef]

228. Schütt, K.T.; Kindermans, P.J.; Sauceda, H.E.; Chmiela, S.; Tkatchenko, A.; Müller, K.R. SchNet: A Continuous-Filter Convolutional Neural Network for Modeling Quantum Interactions. *Adv. Neural Inf. Process. Syst.* **2017**, *2017*, 992–1002. [CrossRef]

229. Hellström, M.; Behler, J. Concentration-Dependent Proton Transfer Mechanisms in Aqueous NaOH Solutions: From Acceptor-Driven to Donor-Driven and Back. *J. Phys. Chem. Lett.* **2016**, *7*, 3302–3306. [CrossRef]

230. Zhang, C.; Yue, S.; Panagiotopoulos, A.Z.; Klein, M.L.; Wu, X. Dissolving Salt Is Not Equivalent to Applying a Pressure on Water. *Nat. Commun.* **2022**, *13*, 822. [CrossRef]

231. Ghorbanfekr, H.; Behler, J.; Peeters, F.M. Insights into Water Permeation through HBN Nanocapillaries by Ab Initio Machine Learning Molecular Dynamics Simulations. *J. Phys. Chem. Lett.* **2020**, *11*, 7363–7370. [CrossRef]

232. Jin, M.; Liu, M.; Nachtigall, P.; Grajciar, L.; Heard, C.J. Mechanism of Zeolite Hydrolysis under Basic Conditions. *Chem. Mater.* **2021**, *33*, 9202–9212. [CrossRef]

233. Zhang, Y.; Wang, H.; Chen, W.; Zeng, J.; Zhang, L.; Wang, H.; Weinan, E. DP-GEN: A Concurrent Learning Platform for the Generation of Reliable Deep Learning Based Potential Energy Models. *Comput. Phys. Commun.* **2020**, *253*, 107206. [CrossRef]

234. Piaggi, P.M.; Weis, J.; Panagiotopoulos, A.Z.; Debenedetti, P.G.; Car, R. Homogeneous Ice Nucleation in an Ab Initio Machine-Learning Model of Water. *Proc. Natl. Acad. Sci. USA* **2022**, *119*, e2207294119. [CrossRef]

235. Groenenboom, M.C.; Keith, J.A. Explicitly Unraveling the Roles of Counterions, Solvent Molecules, and Electron Correlation in Solution Phase Reaction Pathways. *J. Phys. Chem. B* **2016**, *120*, 10797–10807. [CrossRef]

236. Chmiela, S.; Sauceda, H.E.; Müller, K.R.; Tkatchenko, A. Towards Exact Molecular Dynamics Simulations with Machine-Learned Force Fields. *Nat. Commun.* **2018**, *9*, 3887. [CrossRef]

237. Ryczko, K.; Krogel, J.T.; Tamblyn, I. Machine Learning Diffusion Monte Carlo Energies. *J. Chem. Theory Comput.* **2022**, *18*, 7695–7701. [CrossRef]

238. Li, Z.; Kermode, J.R.; De Vita, A. Molecular Dynamics with On-the-Fly Machine Learning of Quantum-Mechanical Forces. *Phys. Rev. Lett.* **2015**, *114*, 096405. [CrossRef]
239. Huang, B.; von Lilienfeld, O.A. Quantum Machine Learning Using Atom-in-Molecule-Based Fragments Selected on the Fly. *Nat. Chem.* **2020**, *12*, 945–951. [CrossRef]
240. Vandermause, J.; Torrisi, S.B.; Batzner, S.; Xie, Y.; Sun, L.; Kolpak, A.M.; Kozinsky, B. On-the-Fly Active Learning of Interpretable Bayesian Force Fields for Atomistic Rare Events. *Npj Comput. Mater.* **2020**, *6*, 20. [CrossRef]
241. Kulik, H.J.; Hammerschmidt, T.; Schmidt, J.; Botti, S.; Marques, M.A.L.; Boley, M.; Scheffler, M.; Todorović, M.; Rinke, P.; Oses, C.; et al. Roadmap on Machine Learning in Electronic Structure. *Electron. Struct.* **2022**, *4*, 023004. [CrossRef]