Toward in silico Catalyst Optimization

Matthew D. Wodrich1, Rubén Laplaza2, Nicolai Cramer3, Markus Reiher4, and Clemence Corminboeuf5

Abstract: In this minireview, we overview a computational pipeline developed within the framework of NCCR Catalysis that can be used to successfully reproduce the enantiomeric ratios of homogeneous catalytic reactions. At the core of this pipeline is the SCINE Molassembler module, a graph-based software that provides algorithms for molecular construction of all periodic table elements. With this pipeline, we are able to simultaneously functionalize and generate ensembles of transition state conformers, which permits facile exploration of the influence of various substituents on the overall enantiomeric ratio. This allows preconceived back-of-the-envelope design models to be tested and subsequently refined by providing quick and reliable access to energetically low-lying transition states, which represents a key step in undertaking in silico catalyst optimization.

Keywords: Computational chemistry · Enantioselectivity · Homogeneous catalysis · Transition metals

Matthew D. Wodrich earned BS and PhD degrees from the University of Arizona and University of Georgia, respectively. Following several postdoctoral positions in Switzerland he was appointed to a permanent scientist position at EPFL in 2015 and promoted to senior scientist in 2019. His current research interests lie in developing and applying computational tools to better understand and identify new homogeneous catalysts.

Rubén Laplaza obtained his bachelor degree in chemistry from the University of Zaragoza. In 2020, he completed his PhD in theoretical chemistry in Sorbonne Université, under the supervision of Profs. V. Polo and J. Contreras-Garcia. He has since been a postdoctoral researcher in LCMD. His current work involves using machine learning and automated modelling pipelines to investigate catalytic processes.

Nicolai Cramer was born in Stuttgart, Germany. From 1998–2003, he studied chemistry at the University of Stuttgart and earned his PhD in 2005 under the guidance of Prof. Sabine Laschat. After a research stage at Osaka University, Japan, he joined the group of Prof. Barry M. Trost at Stanford University as a postdoctoral fellow in 2006. From 2007 on, he worked on his habilitation at the ETH Zurich associated to the chair of Prof. Erick M. Carreira and received the venia legendi in 2010. In 2010, he started as Assistant Professor at the EPF Lausanne and was promoted to Associate Professor in 2013 and to Full Professor in 2015. His main research program encompasses enantioselective metal-catalyzed transformations and their implementation for the synthesis of biologically active molecules.

Markus Reiher received his PhD in theoretical chemistry from the University of Bielefeld in 1998, working with Juergen Hinze. After his habilitation in the group of Bernd Artur Hess at the University of Erlangen, he worked as a private docent in Erlangen in 2003 and at the University of Bonn in 2004. In 2005, he accepted an offer for a professorship in physical chemistry from the University of Jena. In 2006, he moved to ETH Zurich. His research covers many different areas in theoretical chemistry which range from relativistic quantum chemistry to the development of new electron-correlation theories and smart algorithms for the automated exploration of complex chemical reaction networks with automated rolling benchmarking and uncertainty quantification. As a dedicated teacher, he is interested in conveying new ways to help students grasp physical and chemical concepts.

Clemence Corminboeuf started her independent career at the EPFL as an assistant professor and Sandoz Family Foundation Chair. She was promoted to associate (2014) and full professor (2019). She was awarded two European Research Council (ERC) grants (2012/2018), received the Werner Prize of the Swiss Chemical Society in 2014, the Theoretical Chemistry Award from the ACS Physical Chemistry Division in 2018. In 2021, she received the Heilbronner-Hückel Lecture Award from the Swiss and German Chemical Societies and the Per-Olov Löwdin (Uppsala) lecture in 2022. Her research on electronic structure theory exploits the interplay of deterministic and statistical approaches applied to the area of homogeneous catalysts and molecular organic materials.

1. Introduction

The tailored production of chiral molecules possessing one or more stereocenters is a backbone of modern synthetic chemistry. Often, such molecules are created through processes that rely on homogeneous catalysts that promote the production of...
one enantiomer over the other. Through a combination of intuition and experience, synthetic chemists have become adept at developing ‘catalyst design’ strategies and creating back-of-the-envelope models that drive reaction enantioselectivity through the strategic placement of various electronic and steric elements. This may, for example, involve hindering access to the reaction center for an incoming substrate in one orientation while allowing (or promoting) an alternative orientation. Such conjectural models are simple and elegant, yet may not always correspond to a completely accurate picture at the molecular level. In this sense, computational chemistry has a significant role to play, not only in rationalizing experimental results but also in improving catalyst design by providing more detailed atomistic pictures.\[1-5\] Despite the availability of more sophisticated and intricate treatments, computational studies of homogeneous catalysis still most often rely on the use of density functional theory. One of the principal problems, however, is that these traditionally used ‘static’ DFT computations often take a single structure as representative of the complete system (\textit{pars pro toto}). While this can lead to insightful mechanistic information,\[6-9\] predictions of more subtle properties that are governed by very small free energy differences, such as enantioselectivity, are much more challenging. If accurate predictions of selectivity are desired, it is of vital importance that all energetically low-lying transition states of an enantiodetermining step (\textit{i.e.}, those that contribute to the enantiomeric ratio, \(er\)) be considered.

Recently, we showed that the \(er\) of \(\text{C–H}\) functionalization reactions of benzohydroxamates to form dihydroisoquinolones using chiral rhodium-cyclopentadienyl catalysts\[10,11\] (Scheme 1a) could be accurately predicted from a newly developed computational pipeline.\[12\] Reproducing quantitatively accurate \(er\) values required on-the-fly production of a conformer library that was adaptable to the specific catalyst and substrate steric moieties present in the system of interest. In other words, we needed a piece of software that reliably and quickly produced numerous conformers covering the complete conformational space of the catalytic system that we were interested in studying, while simultaneously being sufficiently adaptable that various components of the catalyst and substrate could be modified with minimal human intervention and with no degradation in coverage of the complete conformational space. While automation toolkits for quantum chemistry are increasingly available, including autodE,\[13\] CatVS,\[14\] and QCASIM,\[15\] they do not fully satisfy all of our demands for conformational exploration. As such, we employed the SCINE Molassembler library, which is a graph-based computer program developed by Sobez and Reiher that provides algorithms for molecular construction of molecules from elements of the whole periodic table,\[16\] as the centerpiece of a computational pipeline (Fig. 1) aimed at predicting enantioselectivity in homogeneous catalysis. The versatility of Molassembler was particularly poignant here, principally because haptic \(\eta^1\)-coordination of the Cp ring with the metal atom allows the catalyst to have a large degree of conformational freedom (Scheme 1b). To ensure accurate \(er\) values, numerous conformers of the enantiodetermining transition state must be considered.

### 2. Computational Protocol

To reproduce and predict \(er\) values for the reaction in Scheme 1, we began by rapidly generating 50 conformers for each of the four possible orientations of the styrene substrate with the catalyst (Fig. 2a–d) using Molassembler. This process was done individually for each of the 12 mannitol-Cp/R-group derivates given in Scheme 1a,c (1A–4C). Owing to the speed and flexibility of Molassembler, this process could be completed in under 1 minute per catalyst/R-group combination. Structures were subsequently refined at the GFN2-xTB level\[17\] in implicit ethanol solvent using the ALPB formalism\[18\] and Gaussian16\[19\] to drive the transition state optimization. Converged TS structures were then further optimized at the B3PW91\[20-22\]-D3(BJ)\[23,24\]/def2-SVP\[25\] level followed by single point computations at the B3PW91-D3(BJ)/def-TZVP\[26\] level with the SMD implicit solvation model\[27\] (ethanol). Finally, the solvent-corrected free energies were grouped by selectivity (\(R\) and \(S\)) and Boltzmann weighted at 296.15 K. Enantiomeric ratios (\(er\)) were subsequently computed from theoretical kinetic constants

\[k_{R/S} = \exp \left( \frac{-\Delta G_{TS}^{\text{eff}}}{RT} \right) \text{ as } er_{R/S} = 100 \times \frac{k_{R/S}}{k_R + k_S}\]

---

**Scheme 1.** (a) \(\text{C–H}\) functionalization of benzohydroxamates to form dihydroisoquinolones. (b) Main conformational degrees of freedom in the rhodium-cyclopentadienyl catalysts. (c) Pool of chiral Cp\(^\text{\textdegree}\) ligands studied. Reproduced from ref. [12] with permission from the Royal Society of Chemistry.
Fig. 2. (a–d) Substrate pathways leading to the pro-R and pro-S enantiomers. (e) Back-of-the-envelope ligand design model to induce enantioselectivity where the sidewall and backwall act as steric shields that direct the substrate into the DR orientation. Reproduced from ref. [12] with permission from the Royal Society of Chemistry.

where identical pre-exponential factors and concentrations are assumed. Note that the computational procedure outlined was employed in our recent publication[12] but represents only one of many possibilities. The computational pipeline featuring Molassembler is entirely adaptable to other theoretical levels and should be adapted as needed by the user.

3. Results and Discussion

Using the computational procedure outlined above, we derived theoretical er values for the 12 catalyst/R-group combinations shown in Scheme 1. This not only allowed us to establish the accuracy of our computational pipeline, but also to analyze the accuracy of the back-of-the-envelope catalyst design model originally developed by Cramer and coworkers.[10,11] The proposed design model involves using steric elements to orient the incoming styrene substrate into a single orientation. The blue ‘backwall’ ensures that the substrate cannot access the reaction center from the opposite (back) side which would destroy catalyst selectivity, while the role of the downward pointing methyl/isopropyl group located on the six-membered ring (i.e., the ‘sidewall’, Fig. 2e) along with the Cp5 ligand is to sterically interfere with the phenyl group of the styrene. Finally, the purple ceiling should promote association of the styrene substrate in which the phenyl group is pointed in a ‘downward’ orientation. Taken together, the back-of-the-envelope design model should energetically favor the DR catalyst/substrate orientation while disfavoring the DS, UR, and US pathways (Figs 2a–d).

Employing this design strategy, Cramer and coworkers synthesized seven catalysts and examined the corresponding er values (italic values in colored squares, Fig. 3), finding many that resulted in good enantioselectivity (e.g., 1A, 1C, 2A, 2C). Using the computational protocol outlined above, we aimed to reproduce the er values for systems for which reference experimental data was available and predict values for systems that remained experimentally untested. Fig. 3 shows our predictions, which are color coded to illustrate the error of computation relative to experiment (where green corresponds to better agreement with available experimental data). The two sets of results shown in Fig. 3 aid in creating a greater understanding of the accuracy of the proposed back-of-the-envelope catalyst design model. As initially proposed, the bulk of the Cp5 ligand should preclude coordination of the styrene substrate in an orientation where the phenyl group is pointed upward (e.g., UR and US conformers should not be present). If only the DR and DS conformers are Boltzmann-weighted, the er values shown in Fig. 3b would be predicted. The large amount of yellow, orange and red in Fig. 3b (indicative of divergence from available experimental values) shows that considering only the ‘down’ conformers does not lead to an accurate picture of reality. However, if all conformers are considered (e.g., inclusion of UR and US) we arrive at the much more accurate er values shown in Fig. 3a. To examine the accuracy of the proposed back-of-the-envelope design strategy more closely, below we present three brief case studies illustrating important elements for designing highly enantioselective catalysts.

3.1 Case Study 1: Loss of Enantioselectivity by Access to one ‘up’ Orientation

Catalyst 4B (Scheme 1) was experimentally found to impart no enantioselectivity with a 50:50 er value that was closely matched by our computation (61:39). Fig. 4 shows the lowest energy structures for each of the four possible catalyst/substrate orientations. 4B is most notably characterized by having a relatively small backwall and larger (iPr) sidewall components as part of the Cp5 ligand. The ‘ceiling’ is comprised of three C–H groups as components of the five-membered cyclopentadienyl ring. In accordance with expectations from the design model, the DR orientation was found to be lowest in energy, lying well below the relative energy
of the lowest energy DS orientation (+3.77 kcal/mol) which points to the general efficacy of both the back- and sidewalls. Thus, if only structures presumed to exist in the original design model were to be considered (i.e., DR and DS only), then the catalyst would be expected to demonstrate high enantioselectivity. However, the lack of steric interactions between the styrene substrate and the Cp ligand ceiling complicates matters, as sufficient space exists for the substrate to associate with the catalyst with the phenyl group being in the ‘up’ (US) configuration. Interestingly, the UR configuration retains a very high relative energy, despite the lack of any clear steric interactions. Regardless, an energy difference of only +0.42 kcal/mol between the US and DR configurations results in a computed R:S ee ratio of 61:39. Overall, a loss of enantioselectivity is seen here as the back-of-the-envelope design model does not consider the existence of a single energetically low-lying up configuration of the substrate. Presumably inclusion of more steric interaction associated with the Cp ligand ceiling would further disfavor the US configuration and lead to improved enantioselectivity.

3.2 Case Study 2: Gain of Enantioselectivity by Access to one ‘up’ Orientation

Catalyst 1A was found to be the most selective species with an experimental ee of 96:4. Relative to 4B, the design elements of this catalyst include a much more significant sized backwall, slightly smaller sidewall elements (Me vs iPr), and a ceiling that is unchanged. Fig. 5 shows the lowest energy conformers for each of the aforementioned catalyst substrate orientations. While the increased size of the backwall seems to retain efficacy by disfavoring a backside attack (i.e., the US configuration lies higher in energy), as in the previous example, the lack of steric bulk in the Cp ligand ceiling allows for the styrene substrate to orient in the ‘up’ configuration. In fact, steric interactions between the Cp ring and styrene phenyl group are sufficiently lacking that the UR orientation is lower in energy (+2.83 kcal/mol) than either of the ‘down’ orientations. Here, the UR orientation was found to be the most energetically favorable, with the US orientation also lying lower in energy (+2.83 kcal/mol) than either of the ‘down’ orientations (DR: +2.89 kcal/mol, DS: +4.28 kcal/mol). Furthermore, the reduction in steric bulk of the leaving group compared to 4B (Case Study 1) allows the UR conformer to become further stabilized relative to both ‘S’ conformations. As such, 4C represents a particular case where the lack of steric bulk in the ceiling and leaving group appears to be beneficial for the overall enantioselectivity.

3.3 Case Study 3: Retention of Enantioselectivity by Access to two ‘up’ Orientations

The final case study represents a complete departure from the original back-of-the-envelope design model. Catalyst 4C was found to have a high R:S ee ratio (92:8, experiment), but this results not from any significant energetic contributions from the ‘down’ orientations. As seen in Fig. 6 the Cp ligand is unchanged from 4A, with the difference between the two systems coming from the less bulky ‘R’ in the protecting group (OrBu vs Me). As in 4A, the backwall capably serves to prevent backside attack of the catalyst by the substrate while the sidewall succeeds in preferentially orienting the styrene phenyl group in the DR over the DS orientation. Thus, the back-of-the-envelope design model would lead to an expected favoring of R over the S orientation by 1.29 kcal/mol (DR vs DS). However, as in the other cases the lack of a sterically bulky ceiling allows the styrene to most easily associate in the ‘up’ orientation. Here, the UR orientation was found to be the most energetically favorable, with the US orientation also lying lower in energy (+2.83 kcal/mol) than either of the ‘down’ orientations (DR: +2.89 kcal/mol, DS: +4.28 kcal/mol). Furthermore, the reduction in steric bulk of the leaving group compared to 4B (Case Study 1) allows the UR conformer to become further stabilized relative to both ‘S’ conformations. As such, 4C represents a particular case where the lack of steric bulk in the ceiling and leaving group appears to be beneficial for the overall enantioselectivity.

3.4 Toward Improved Catalyst Design

Ye and Cramer succeeded in developing highly enantioselective reactions to transform benzoxycarbonylates and alkenes to dihydroxyquinolines using chiral rhodium-cyclopentadienyl catalysts,[10] however, their back-of-the-envelope design model used to develop enantioselective ligands was found by our computational analysis to be imprecise. Analyzing the shortcomings of the model through the examination of low-lying transition state leading to both the R- and S-products provides routes to further modify ligand structures to better fit the model. For example, we demonstrated that the lack of steric bulk in the Cp ligand ceiling allowed the styrene phenyl group to adopt energetically favorable “up” orientations which complicate the simple and intuitive route...
to imparting selective by the design model. Indeed, Cramer has more recently begun using the Cp6 ligand to include more steric bulk at all positions to catalyze this27 and other reactions28–30 using both rhodium and cobalt catalysts. This is a key illustration of how the synergy between experiment and computation, specifically how the development of computational tools and pipelines can serve as a valuable instrument for refining and developing improved experimental work.

4. Conclusion
We provided an overview of a computational pipeline featuring Molassembler at its core that aims to reproduce the enantioselectivity of homogeneous catalytic reactions. The efficacy of the protocol was demonstrated by examining the R:S enantiomeric ratios of the conversion of benzohydroxamates and alkenes to dihydrosinoquinolones catalyzed by various chiral rhodium-cyclopentadienyl catalysts. As illustrated within, the computational pipeline not only accurately reproduces experimental $\text{er}$ values, but further facilitates the development of catalyst design models by providing the structures of the key low-lying transition states. Examination of these structures allows the designer to more accurately determine the efficacy of the sterically used to induce enantioselectivity. In turn, back-of-the-envelope design models can be refined to arrive at highly selective catalysts using this computational pipeline. Overall, this work represents a prime example of not only collaborations between groups of theoretical/computational chemists possessing various expertise (e.g., the Cerninboeuf and Reicher groups) but also illustrates the synergy between theory and experiment that is present within the NCCR Catalysis.

Acknowledgements
The authors are grateful to the EPFL and ETHZ for financial support and computational resources. This publication was created as part of NCC Catalysis (grant number 180544), a National Centre for Competence in Research funded by the Swiss National Science Foundation.

Received: December 16, 2022

License and Terms
This is an Open Access article under the terms of the Creative Commons Attribution License CC BY 4.0. The material may not be used for commercial purposes. The license is subject to the CHIMIA terms and conditions. The definitive version of this article is the electronic one that can be found at https://doi.org/10.2533/chimia.2023.139