



Doctoral Thesis

Error-Controlled Quantum Chemical Exploration of Reaction Networks

Author(s):

Simm, Gregor Nils Christoph

Publication Date:

2018

Permanent Link:

<https://doi.org/10.3929/ethz-b-000299881> →

This page was generated automatically upon download from the [ETH Zurich Research Collection](#). For more information please consult the [Terms of use](#).

DISS ETH NO. 25343

Error-Controlled Quantum Chemical Exploration of Reaction Networks

A THESIS SUBMITTED TO ATTAIN THE DEGREE OF
DOCTOR OF SCIENCES OF ETH ZURICH
(DR. SC. ETH ZURICH)

PRESENTED BY
GREGOR NILS CHRISTOPH SIMM

MSc ETH IN INTERDISCIPLINARY SCIENCES, ETH ZURICH

BORN ON 20.12.1991
CITIZEN OF GERMANY

ACCEPTED ON THE RECOMMENDATION OF
PROF. DR. MARKUS REIHER, EXAMINER
PROF. DR. GUNNAR JESCHKE, CO-EXAMINER

ETH ZURICH
ZURICH, SWITZERLAND
2018

Abstract

For a detailed analysis of a chemical system, all relevant intermediates and elementary reactions on the potential energy surface (PES) connecting them need to be known. An in-depth understanding of all reaction pathways would allow one to study the evolution of a system over time, given a set of initial conditions (e.g., reactants and their concentrations, temperature, and pressure) and propose derivatives of the original reactants to avoid undesired side reactions. Manual explorations of complex reaction mechanisms employing quantum-chemical methods are slow and error-prone. In addition, due to the high dimensionality of PESs exhaustive exploration is generally unfeasible. However, to rationalize, for instance, the formation of undesired side products or decomposition reactions, unexpected reaction pathways need to be uncovered.

In this thesis, we present a computational protocol that constructs reaction networks, consisting of intermediates and transition states, in a fully automated fashion. Starting from a set of initial reagents new intermediates are explored through intra- and intermolecular reactions of already explored ones. This is done by assembling reactive complexes based on heuristic rules derived from conceptual electronic-structure theory and exploring the corresponding approximate reaction path. A subsequent path refinement leads to a minimum-energy path which connects the new intermediate to the existing ones to form a connected reaction network. Tree traversal algorithms are then employed to detect reaction channels and catalytic cycles. We apply our protocol to the formose reaction to study different pathways of sugar formation and to rationalize its autocatalytic nature. Furthermore, we investigate the Schrock dinitrogen-fixation catalyst and discover alternative pathways of catalytic ammonia production.

To be able to draw reliable conclusions from the generated reaction networks, accurate relative energies between intermediates and transition states are required. To date, density functional theory (DFT) is the only method that is computationally feasible for the *ab initio* exploration in this detail. However, DFT often fails to provide sufficiently accurate results, especially for systems containing transition metals. In this thesis, we apply a framework based on Bayesian statistics that allows for error estimation of properties calculated with DFT. Instead of considering only the best-fit parameters of an approximate density functional, we assign a conditional probability distribution to the continuous set of parameters from which a confidence interval can be calculated for any observable. We assess our approach at two challenging chemical systems: catalytic nitrogen fixation and the formose reaction.

Finally, to overcome the lack of systematic improvability of approximate quantum chemical methods we apply Bayesian statistical learning. This new approach allows for the systematic, problem-oriented, and rolling improvement of quantum chemical results through the application of Gaussian processes. Due to its Bayesian nature, reliable error estimates are provided for each prediction. A reference method of high accuracy will be employed to provide a new data point if the uncertainty associated with a particular calculation is above a given threshold. This data point is then added to a growing data set in order to continuously improve the model, and as a result, all subsequent predictions. Previous predictions are validated by the updated model to ensure that uncertainties remain within the given confidence bound, which we call backtracking. We demonstrate our approach with the example of a complex chemical reaction network.

Zusammenfassung

Um die Reaktivität eines chemischen Systems detailliert verstehen zu können, ist eine ausführliche Analyse der Potenzialhyperfläche unabdingbar. Für die vollständige Aufklärung eines chemischen Prozesses muss diese Analyse alle relevanten Intermediate und Übergangszustände enthalten. Damit wäre es möglich (gegeben die Anfangskonzentration aller chemischen Spezies, die Temperatur und den Druck), den zeitlichen Verlauf einer Reaktion vorherzusagen. Zudem wäre man in der Lage, alternative Reaktanden und Reaktionsbedingungen vorzuschlagen, um unerwünschte Nebenreaktionen zu vermeiden. Die manuelle Untersuchung komplexer Reaktionsmechanismen mit Hilfe von quantenchemischen Methoden ist fehleranfällig und sehr zeitaufwendig. Außerdem ist eine vollständige Analyse in der Regel nicht möglich, da die zu untersuchenden Hyperflächen meist hochdimensional sind. Um jedoch zum Beispiel Nebenreaktionen vorhersagen zu können, müssen auch unerwartete Reaktionspfade untersucht werden.

In dieser Doktorarbeit erarbeiten wir ein Verfahren, das die vollautomatische Exploration eines Reaktionsnetzwerks, bestehend aus Intermediaten und Übergangszuständen, erlaubt. Durch intra- und intermolekulare Reaktionen zwischen bereits entdeckten Intermediaten werden neue Intermediate dem Netzwerk hinzugefügt. Mit Hilfe von heuristischen Regeln, welche auf Konzepten der Elektronenstrukturtheorie basieren, werden sogenannte *reaktive Komplexe* erstellt. Wird diesen Komplexen Energie (beispielsweise in Form von kinetischer Energie) zugeführt, können genäherte Reaktionspfade erforscht werden. Die Verfeinerung dieser Pfade führt zu Elementarreaktionen, welche neue Intermediate mit bereits Abgebildeten verbinden, sodass sich ein zusammenhängendes Reaktionsnetzwerk ergibt. Algorithmen zur Analyse von Netzwerken erlauben einem dann, Reaktionskanäle und katalytische Zyklen zu entdecken. Am Beispiel einer präbiotischen Polymerisierungsreaktion, der Formosereaktion, zeigen wir, dass es mit unserem Verfahren möglich ist, auf automatische Weise große Reaktionsnetzwerke zu erstellen. Insbesondere werden auch autokatalytische Eigenschaften dieser Reaktion reproduziert. Zudem untersuchen wir die Reaktivität eines Stickstoff fixierenden Katalysators, dem Yandulov-Schrock-Katalysators, und finden dabei Nebenreaktionen, die seine niedrige Wechselzahl erklären.

Genauere (freie) Energien sind nötig, um zuverlässige Schlüsse aus der Analyse von Reaktionsnetzwerken ziehen zu können. Momentan ist Dichtefunktionaltheorie (DFT) die einzige Elektronenstrukturmethode, die eine Untersuchung solcher detaillierter Netzwerke erlaubt. Mittels DFT ist es jedoch nicht immer möglich, genaue thermodynamische Grö-

ßen zu berechnen, insbesondere bei Systemen, die Übergangsmetalle enthalten. Deshalb entwickeln wir in dieser Dissertation einen Ansatz, der auf bayesscher Statistik beruht und für Observablen, die mit DFT berechnet wurden, eine Fehlerabschätzung ermöglicht. Typischerweise werden für empirische Parameter in approximativen Dichtefunktionalen lediglich vordefinierte optimierte Werte verwendet. Im Gegensatz dazu wird in unserem statistischen Ansatz eine Wahrscheinlichkeitsverteilung für die Parameter bestimmt, mit der Vertrauensintervalle für beliebige Observablen ermittelt werden können. Wir wenden die neu entwickelte Methode auf die Formosereaktion und den Katalysezyklus des Yandulov–Schrock-Katalysators an.

Im letzten Kapitel dieser Dissertation nutzen wir bayessches statistisches Lernen, um approximative quantenchemische Methoden systematisch zu verbessern. Mit Hilfe von Gaussprozessen können nun zuverlässige Abschätzungen zu statistischen Unsicherheiten von berechneten Resultaten gemacht werden. Ist diese Unsicherheit zu groß, wird eine hochgenaue Referenzberechnung durchgeführt. Das Resultat dieser Berechnung wird einem wachsenden Referenzdatensatz hinzugefügt, um Vorhersagen des Gaussprozesses zu verbessern. Anschließend werden bisherige Vorhersagen überprüft, um sicherzustellen, dass die Unsicherheiten innerhalb des gegebenen Konfidenzintervalls liegen. Wir demonstrieren die Nützlichkeit dieses Ansatzes zur Fehlerabschätzung an einem komplexen Reaktionsnetzwerk.