

**Automated and high-throughput reactivity analysis in homogeneous catalysis**  
**The deactivation complexity of Mn(I) hydrogenation catalysts**

Hashemi, A.

**DOI**

[10.4233/uuid:610c0657-9706-4fde-90ce-6bc9ecd14620](https://doi.org/10.4233/uuid:610c0657-9706-4fde-90ce-6bc9ecd14620)

**Publication date**

2023

**Document Version**

Final published version

**Citation (APA)**

Hashemi, A. (2023). *Automated and high-throughput reactivity analysis in homogeneous catalysis: The deactivation complexity of Mn(I) hydrogenation catalysts*. [Dissertation (TU Delft), Delft University of Technology]. <https://doi.org/10.4233/uuid:610c0657-9706-4fde-90ce-6bc9ecd14620>

**Important note**

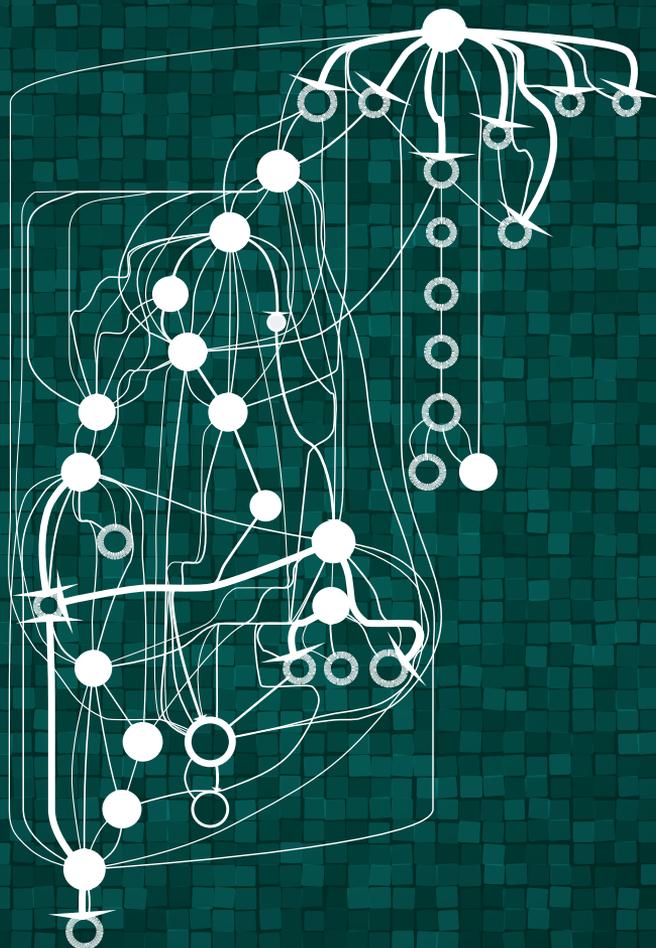
To cite this publication, please use the final published version (if applicable).  
Please check the document version above.

**Copyright**

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

**Takedown policy**

Please contact us and provide details if you believe this document breaches copyrights.  
We will remove access to the work immediately and investigate your claim.



**AUTOMATED AND HIGH-THROUGHPUT REACTIVITY  
ANALYSIS IN HOMOGENEOUS CATALYSIS**

THE DEACTIVATION COMPLEXITY OF  
 $Mn(I)$  HYDROGENATION CATALYSTS

**ALI HASHEMI**

# **AUTOMATED AND HIGH-THROUGHPUT REACTIVITY ANALYSIS IN HOMOGENEOUS CATALYSIS**

THE DEACTIVATION COMPLEXITY OF Mn(I) HYDROGENATION  
CATALYSTS



# **AUTOMATED AND HIGH-THROUGHPUT REACTIVITY ANALYSIS IN HOMOGENEOUS CATALYSIS**

**THE DEACTIVATION COMPLEXITY OF MN(I) HYDROGENATION  
CATALYSTS**

## **Dissertation**

for the purpose of obtaining the degree of doctor  
at Delft University of Technology  
by the authority of the Rector Magnificus, prof. dr. ir. T.H.J.J. van der Hagen,  
chair of the Board for Doctorates  
to be defended publicly on  
Monday 22 May 2023 at 12:30 o'clock

by

**Ali HASHEMI**

Master of Science in Chemical Engineering , University of Tehran, Tehran  
born in Esfahan, Iran

This dissertation has been approved by the promotors.

Composition of the doctoral committee:

Rector Magnificus,	chairperson
Prof. dr. E. Pidko,	Delft University of Technology, <i>promotor</i>
Prof. dr. M.P. Gaigeot,	Université d'Evry val d'Essonne, <i>promotor</i>

*Independent members:*

Prof. dr. C. Corminboeuf	École Polytechnique Fédérale de Lausanne
Prof. dr. E.J. Meijer	University of Amsterdam
Prof. dr. T.J.H. Vlugt	Delft University of Technology
Prof. dr. F. Grozema	Delft University of Technology
Dr. F. Buda	Leiden University

The work described in this dissertation was carried out in the Inorganic Systems Engineering (ISE) group, Department of Chemical Engineering, Faculty of Applied Science, Delft University of Technology. The research was supported by the European Research Council under the European Union's Horizon 2020 Research and Innovation Program (Grant Agreement No. 725686). Access to SurfSARA supercomputer resources was sponsored by the Netherlands Organization for Scientific Research (NWO).



*Printed by:* RidderPrint | Alblasterdam, the Netherlands

Copyright © 2023 by A. Hashemi

ISBN 978-94-6366-688-6

An electronic copy of this dissertation is available at  
<https://repository.tudelft.nl/>.

To my parents



تقدیم به

نگاه پدرم و دستان مادرم

که معنای بسط خاطر اند .

و تقدیم به برادرانم

که دلگرمی جاودان من هستند .



# CONTENTS

<b>1. Towards bias-free mechanistic models</b>	<b>1</b>
1.1. Introduction	2
1.2. State-of-the-Art in the Reaction Path-Finding Strategies	3
1.3. Computational Tools to Explore Catalytic Reaction Mechanisms	5
1.4. Comparison of Reaction Mechanism Exploration Tools	9
1.5. Scope of the thesis and outlook:	13
<b>2. Comparative Benchmark Study</b>	<b>21</b>
2.1. Introduction	22
2.2. TSSCDS method for accelerated reaction network analysis	24
2.3. Automated reaction explorations for acetone conversion	26
2.3.1. Homogeneous catalysis for selective carbonyl reduction	26
2.3.2. TSSCDS and ReNeGate Reaction Network Comparison Results	27
2.3.3. Challenges observed with TSSCDS as a scalable tool for reaction network analysis	31
2.3.4. Exhaustive “static” conformational search for favorable association complexes: A pre-processing step to improve the efficiency of the original strategy	32
2.4. Conclusion	34
<b>3. ReNeGate</b>	<b>41</b>
3.1. Introduction	43
3.2. Automated reaction exploration method	46
3.2.1. Reactive Space Exploration	47
3.2.2. Reaction Event Identification and Network Construction	47
3.2.3. Fingerprinting and Reaction Network Construction	49
3.2.4. Reaction network trimming	49
3.2.5. Fragment analysis	50
3.3. Validation	52
3.3.1. Case Study I: Mn(CO) <sub>5</sub> Br pre-catalyst activation	52
3.3.2. Case Study II: Base-activation of an Mn(Br)(CO) <sub>3</sub> -NN transfer hydrogenation pre-catalyst	54
3.3.3. Case Study III: The formation of multinuclear ensembles upon the base-activation of Mn(CO) <sub>5</sub> Br	55
3.4. Conclusion and outlook	58
<b>4. HiREX: High-Throughput Reactivity Exploration</b>	<b>71</b>
4.1. Introduction	73

4.2. Methods . . . . .	75
4.2.1. Functionalization Strategy with ChemSpaX . . . . .	75
4.2.2. Reactivity Explorations - the ReNeGate workflow . . . . .	76
4.2.3. Comparison with reference structures . . . . .	77
4.3. Results and discussion . . . . .	79
4.3.1. High-throughput reactivity exploration . . . . .	79
4.3.2. Correlation analysis of the structural features and reactivity . . . . .	81
4.3.3. Finding the most frequent classes of interactions: K-mode clustering . . . . .	83
4.4. New Chemistry . . . . .	85
4.4.1. Decoordination of the Mn-D (D = C, N, P) bonds . . . . .	85
4.4.2. Migrations of the CF <sub>3</sub> moieties . . . . .	91
4.5. Conclusions . . . . .	94
<b>5. Summary and Outlook</b> . . . . .	<b>105</b>
5.1. Samenvatting . . . . .	108
<b>6. Conclusion</b> . . . . .	<b>111</b>
<b>A. Supplementary Information Chapter3</b> . . . . .	<b>113</b>
A.1. Relevant Molecular Graph Theory Terminology . . . . .	113
A.2. Molecular conformation and relationship to graph theory . . . . .	114
A.3. DFT energies of structures identified by ReNeGate . . . . .	117
<b>B. Supplementary Information Chapter4</b> . . . . .	<b>121</b>
B.1. Correlation analysis of the structural features and reactivity for structures in [-40,25] kcal.mol <sup>-1</sup> range . . . . .	121
B.2. Role of Adducts in Observed Reactivities: . . . . .	122
B.3. Exploration results leading to decooordination of Mn-N ligands for catalysts with different backbones: . . . . .	122
B.4. Exploration results for attack by NHC moieties of the CNC pincer on the carbonyl ligands (Mn-C interactions): . . . . .	122
B.5. Supplementary Tables . . . . .	122
<b>Acknowledgements</b> . . . . .	<b>147</b>
<b>Curriculum Vitæ</b> . . . . .	<b>151</b>
<b>List of Publications</b> . . . . .	<b>153</b>

# 1

## TOWARDS BIAS-FREE MECHANISTIC MODELS

*Emerging Tools for Automated Reaction Path Analysis and Predictions*

CATALYTIC systems are commonly represented by complex mixtures of reactants, catalyst precursors, ligands, additives, and solvent that may give rise to the formation of a wide variety of species that may show varied catalytic activity and behavior towards other components of the reaction mixture. The primary tasks of computational catalysis are to identify among these pre-reaction complexes those that contribute most to the catalytic reaction and identify mechanisms of the main catalytic cycle and competing reaction channels giving rise to unselective conversion routes or catalyst deactivation. As we will demonstrate throughout this chapter, the development of a comprehensive molecular-level picture of a catalytic system is a very challenging task due to the enormous complexity of the associated chemical reaction. However, I firmly believe that if all these tasks are accomplished, the resulting reactivity model could be used to guide the development of more active and efficient catalysts, which is the core idea of the rational catalyst design strategy.

---

This chapter is partially based on: K. D. Vogiatzis, M. V. Polynski, J. K. Kirkland, J. Townsend, A. Hashemi, C. Liu, E. A. Pidko *Computational Approach to Molecular Catalysis by 3d Transition Metals: Challenges and Opportunities*. *Chemical Reviews* 2019, 119 (4), 2453.1 (A. Hashemi's contribution: literature analysis, draft preparation, revisions with the major contribution to the section on the automated approaches for reactivity analysis in homogeneous catalysis)

## 1.1. INTRODUCTION

Modern density functional theory (DFT)[1] and wave-function theory (WFT) methods[2] provide the necessary computational toolbox for sufficiently and accurately evaluating the structures and energetics of sequences of intermediates and transition states within an envisioned catalytic reaction mechanism[3, 4]. Conventionally, the research strategies commonly employed in computational catalysis imply that the quantum chemical methods are used to analyze pre-defined mechanistic proposals formulated based either on previous suggestions for (often vaguely) related systems, expert opinion or chemical intuition. The importance of expert input in mechanistic analysis of catalytic paths is one of the focal points of this chapter. Classical computational chemistry methods do not provide means to enable discovery of new catalytic paths, but their role is often limited to evaluating reactions within the scope of the existing chemical knowledge. Such a situation is common for many fields of sciences and is often referred to as the “streetlight effect”[5–7]. Dewyer and Zimmerman in their excellent recent perspective state that “for reaction mechanisms, where no hypotheses are available—and the researchers “just don’t know”—computation has not offered practical solutions to discover these unknown mechanisms”[8].

Recent developments in the field are transforming the ability of computational chemistry to minimize the expert bias in mechanistic analysis and even to discover reaction paths that could not be deduced based on prior knowledge or “chemical intuition”[8–10]. The basic idea behind these emerging methodologies is that with specified reactant molecules and catalyst, programs automatically determine feasible sequences of elementary reaction steps. Such approaches may give rise to practical tools for identifying unexpected reaction mechanisms at reasonable computational cost, enabling a new paradigm of research in quantum chemistry.

From the computational perspective, the construction of a reaction path even for a single elementary step is a quite demanding and non-trivial multistep procedure. Most contemporary approaches to locating reaction paths start from the approximation of a transition state (TS) followed by optimization of the stationary point and intrinsic reaction coordinate (IRC) computations to elucidate the adjacent minima states. Because many of the steps involved in this procedure are highly demanding and can easily fail, substantial efforts of the research community are currently put into the development of alternative streamlined approaches for finding reaction paths directly from mechanistic hypotheses[9]. Automated methods capable of formulating mechanistic hypotheses for elementary steps in combination with efficient approaches for reaction path and transition state optimization would drastically lower the amount of chemical intuition and expert bias involved in mechanistic research and may become the basis for the true predictive computational catalysis methodologies.

## 1.2. STATE-OF-THE-ART IN THE REACTION PATH-FINDING STRATEGIES

Catalytic reactivity in a general sense is determined by the complex network of chemical reactions taking place simultaneously or consequently between the different (transient) components of the catalytic mixture. Each of the stages of the catalytic process – that is the catalyst activation, catalytic cycle propagation, catalyst deactivation, non-selective conversion paths – may involve multiple elementary steps that proceed via mechanism and involve reaction intermediates that are not known *a priori*. Even the most advanced experimental operando techniques are not able to unravel such a high molecular-level complexity, and its computational analysis requires a much broader exploration of the chemical and configuration space to identify the minima on the potential energy surface (PES) and the pathways connecting them. Such a computational reaction discovery may be facilitated by narrowing down the reaction space by either applying pre-defined heuristic rules (e.g., bond breaking) to generate intermediates or by artificially pushing the reactants together in a simulation to induce chemical transformations.

In principle, *ab initio* molecular dynamics (AIMD) provide the direct means to probe reaction events at the molecular level; however, the major challenge here is that even the fastest chemical reactions are considered rare events making the adequate scanning of the reaction space by the direct atomistic AIMD simulations based on sufficiently accurate electronic structure methods challenging. The frequency of the reaction events can be greatly accelerated by applying bias potentials that push the system away from the free energy minima along a collective variable, which assumes some knowledge of the reaction coordinate or collective variable along which to apply the biasing potential[11–13]. A similar reactivity enhancement can be achieved by simulating reactions at extremely high-temperature or high-pressure regimes, which effectively shift the equilibrium to products with higher entropy or lower volume, respectively[14]. The determination of the reaction paths and, particularly, the transition state search with quantum chemical methods is a non-trivial task and it is commonly associated with high computational demands. The tutorial review by Schlegel[15] presents a comprehensive overview of the methodological aspects and capabilities of modern approaches for geometry optimization and transition state search. The state-of-the-art in reaction pathway finding strategies is summarized in a comprehensive review by the Zimmerman group[9]. We refer the interested reader to these works for the details on the methodologies and strategies. In this chapter, we will limit ourselves to a brief description of the main concepts underlying these strategies and highlighting the most relevant examples of their applications to the topics relevant to the field of molecular catalysis.

The reaction path exploration strategies can be categorized in four main groups summarized in Figure 1.1. All these methods require explicit definition of a designated set of reactants and catalysts, but all subsequent steps are supposed to operate without external bias or with minimal possible interference from a researcher. The first class of methods, designated in Figure 1.1 as *Concept 1*, largely follow the way chemist researchers develop mechanistic proposals. The basis for

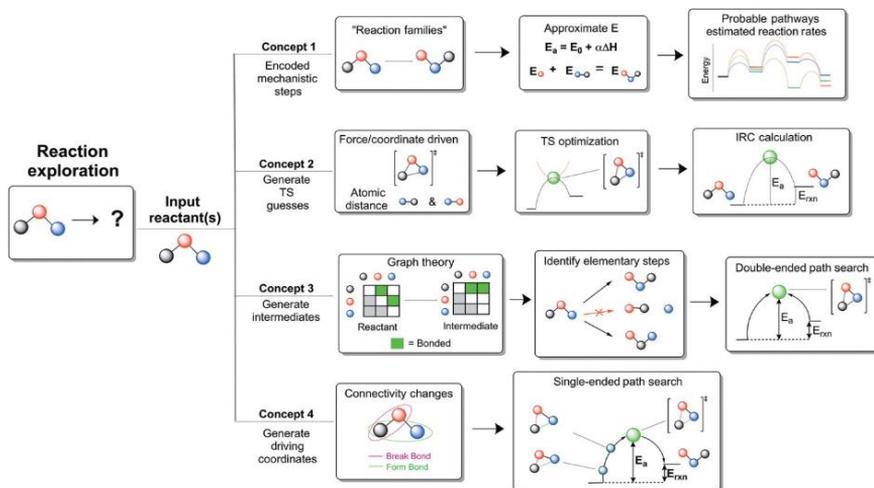


Figure 1.1.: Classification of automated reaction path exploration methods. Reprinted with permission from WIREs Comput. Mol. Sci. 2018, 8, e1354. Copyright 2017, John Wiley and Sons.

this approach is a set of chemistry rules formalized by encoded elementary step types from databases or rules of chemical heuristics. These are used to describe reaction pathways between the reacting molecules defined at the start. Methods in this category commonly use activation barriers and reaction rates estimated using approximate methods and only seldom involve the explicit TS search. In the second category (*Concept 2*) approximate TSs are first generated from the reactants, followed by local TS optimization and IRC calculations. One of the most common methods for generation of the initial TS approximation in such methods involves exposing two molecules to artificial forces that would push them together to induce a reaction. The methods within *Concept 3* start with the generation of a collection of putative elementary steps with the corresponding intermediates formed, followed by the application of double-ended methods to refine reaction paths and locate TSs. The methods within *Concept 4* involve the generation of hypothetical reaction coordinates that are generated, followed by applying single-ended methods to carry out the reaction path searches along these coordinates.

All these four methodological concepts share a common spirit but differ in the conceptual implementation and details that may be paramount to their degree of success. They are all designed to generate approximate reaction paths, estimate reaction barriers and then integrate these paths into reaction networks of elementary steps. The current practical realization and implementations of these methods are still far from perfection. They all can fail at one point or another during the reaction search and none of them can deliver an ideal balance of high accuracy and comprehensive PES analysis at a reasonable computational cost. Despite the ultimate goal to establish a bias-free reaction prediction and the promise of the

non-user interaction reaction exploration, chemical intuition in one form or another still needs to be introduced to some extent for the successful utilization of these methods. In the next sections, we will briefly discuss the available methods for automated reaction path analysis.

### 1.3. COMPUTATIONAL TOOLS TO EXPLORE CATALYTIC REACTION MECHANISMS

The strategies for automated reaction mechanism exploration tools discussed above are quite general and are designed to analyze general (most often organic chemistry) reactions occurring over a barrier[16]. The extension of such methods to transition metal catalysis requires several auxiliary algorithms designed to deal with the specific features of transition metal catalytic reactions, mainly the particular sensitivity to the 3D geometric structure of transition metal complexes and the need for comprehensive sampling of reactions within and outside the catalytic cycles.

*ZStruct2*[17, 18] is a reaction exploration tool introduced by the Zimmerman group. The method combinatorically samples driving coordinates (DC), which are bond-addition or bond-breaking vectors describing elementary reactions[8, 19]. These reactive coordinates are designed for use with the single-ended growing string method[20, 21] (GSM) that generates reaction paths, TSs, and intermediate structures for single elementary steps consistent with the DC. *ZStruct2* handles intramolecular and bimolecular reactions by aligning reactants in a way consistent with the DC. The incorporation of knowledge of the transition metal center geometry and the alignment of the reactants enabled the application of *ZStruct2* to perform studies on transition metal-catalyzed reactions. *ZStruct2* has been successfully used to explore mechanisms of such processes as Pd-catalyzed C-H arylation of piperidine[19],  $\text{FeCl}_3$ -catalyzed carbonyl-olefin metathesis[22, 23], Ni-catalyzed thiazole polymerization[24] and others[25, 26]. For example, in the case of piperidine arylation, *ZStruct2* was able to identify all major steps of the catalytic cycles including the roles that the multiple supporting reagents play in driving forward the reaction. In the study of thiazole polymerization, *ZStruct* identified an unexpected route for chain termination that prevents the controlled polymer growth[24].

The Artificial Force Induced Reaction (AFIR) Method[27] provides a more comprehensive and systematic approach to finding reaction paths to analyze mechanisms and predict selectivity of catalytic reactions. Starting from a given set of reactants and catalyst, AFIR searches all of the important (known, unknown, or unexpected) reaction pathways.

The concept of AFIR is straightforward: simply push or pull fragments A and B in the catalytic mixture together. When both A and B are atoms, they can be pushed together by adding a linear function of their distance  $r_{AB}$  to their potential energy  $E(r_{AB})$ . Figure 1.2 depicts a diatomic potential curve  $E(r_{AB})$ . A barrier separates the reactant pair A + B and the product X in this curve. This barrier can be removed by adding the term  $\alpha r_{AB}$  to  $E(r_{AB})$ , where  $\alpha$  is a constant parameter. The resulting function, depicted in blue in Figure 1.2, contains no barrier. On this function, the product region,  $\alpha r_{AB} + E(r_{AB})$ , can be efficiently reached from the reactant pair

simply by minimizing the function. In polyatomic systems, the same procedure can be carried out by minimizing the following AFIR function.

$$F(Q) = E(Q) + \rho\alpha \frac{\sum_{i \in A} \sum_{j \in B} \omega_{ij} r_{ij}}{\sum_{i \in A} \sum_{j \in B} \omega_{ij}} \quad (1.1)$$

This function consists of two terms, i.e., the Born-Oppenheimer potential energy surface (PES)  $E(Q)$  of geometrical parameters  $Q$  and the artificial force term. The parameter  $\alpha$  in the artificial force term determines the force's strength. The coefficient  $\rho$  is either 1 to push fragments together or -1 to separate them. The force term is given as a weighted sum of the distances  $r_{ij}$  between atoms  $i$  in fragment A and  $j$  in fragment B, and the weight function  $\omega_{ij}$  is as follows:

$$\omega_{ij} = \left[ \frac{(R_i + R_j)}{r_{ij}} \right]^p \quad (1.2)$$

This weight function assigns a stronger force to the closer atom pairs and a weaker force to the more distant pairs. In Eq. 1.2, the inverse distance  $1/r_{ij}$  is scaled by  $R_i + R_j$ , the sum of covalent radii of atoms  $i$  and  $j$ , to treat all elements equivalently. It was confirmed that results did not strongly depend on the choice of  $p$ , and  $p$  is usually set to 6.0[28].

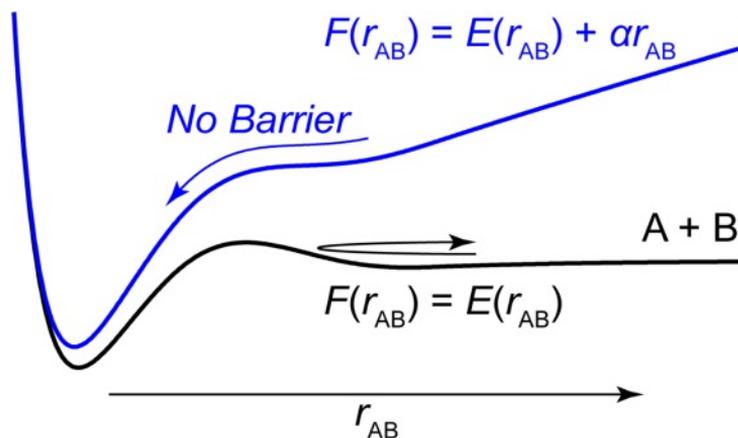


Figure 1.2.: A diatomic potential curve  $E(r_{AB})$  between atoms A and B (black curve) and the corresponding AFIR function  $E(r_{AB}) + \alpha r_{AB}$  (blue curve).  $r_{AB}$  is the distance between A and B, and  $\alpha$  is a constant parameter. (Reprinted with permission from Chem. Rec.2016,16, 2232–2248VC2016 The Chemical Society of Japan & Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim).

The AFIR method not only predicts reaction mechanisms for the desired product, but also explores the side paths resulting in the formation of byproducts. The AFIR methods can be successfully employed to explore reaction paths for relatively simple

catalytic systems (small models, limited number of participating species, etc.), but when applied to realistic molecular systems the full AFIR becomes prohibitively demanding and a restricted AFIR search is more appropriate. In this case, an artificial force should be added between selected fragments. The basic principles and application of AFIR in computational catalysis studies using the AFIR method has been explained in a comprehensive article by Morokuma and co-workers[27].

The AFIR method combined with DFT calculations was used to explore the mechanism and the selectivity of the aqueous Mukaiyama aldol reaction catalyzed by an Fe-based chiral complex[27]. More than 40 approximate TSs were located by AFIR and classified into 12 groups. It is worth mentioning that the AFIR method located 10 TSs relevant for the selectivity of the reaction, some of which may have been missed when following the traditional chemical intuition-guided approaches. This systematic study provides important mechanistic insights relevant for the development of Fe-based catalysts for carbon-carbon bond formation reactions. Besides, the AFIR method has also been successfully employed to analyze the reaction mechanism and identify factors controlling the stereoselectivity of the Kobayashi modification of the Mukaiyama aldol reaction, catalyzed by water-tolerant lanthanide-based  $\text{Ln}(\text{OTf})_3$  Lewis acid catalysts in aqueous media[29–31].

For relatively large molecular systems, the computational cost of AFIR searches can be reduced by using the hybrid model definitions within the NIOM(QM:QM) or ONIOM(QM:MM) methods. Conventionally, the AFIR analysis is carried out with the higher-level model described at the DFT level with a relatively small basis set, while the lower-level part of the system is treated using semiempirical or force field methods. After approximate local minima (LM) and TSs are identified, standard more accurate computational methods (e.g., DFT with a large basis set) are used for the full molecular system to determine the true LMs and TSs and to rationalize the reaction mechanism and selectivity of the catalytic reaction.

*Transition State Search using Chemical Dynamics Simulations (TSSCDS)* is an automated strategy to predict the reaction mechanisms and kinetics of organometallic-catalyzed reactions[32]. This method starts with the division of the catalytic system into smaller subsystems, which are sorted by order of increasing complexity. The TSSCDS method is then applied within each of the subsystems to locate the TSs and intermediates, which are subsequently merged into a single reaction network. Finally, this reaction network is used to calculate overall kinetics. The TSSCDS method is based on a procedure that combines accelerated direct dynamics with an efficient geometry-based post-processing algorithm to find transition states. The method operates with the starting geometries and concentrations of the catalyst and reagents as well as the viscosity of the solvent. The geometries provide the starting configurations for locating intermediates and transition states, while concentrations and the parameters of the solvent are used as the input for the kinetic simulations[33].

TSSCDS has been successfully tested on the cobalt-catalyzed hydroformylation of ethylene and provided a mechanistic outcome that verified the main pathway proposed by Heck and Breslow[34]. The predicted rate law reproduced the one obtained experimentally. Importantly, the TSSCDS is claimed to be able to reveal

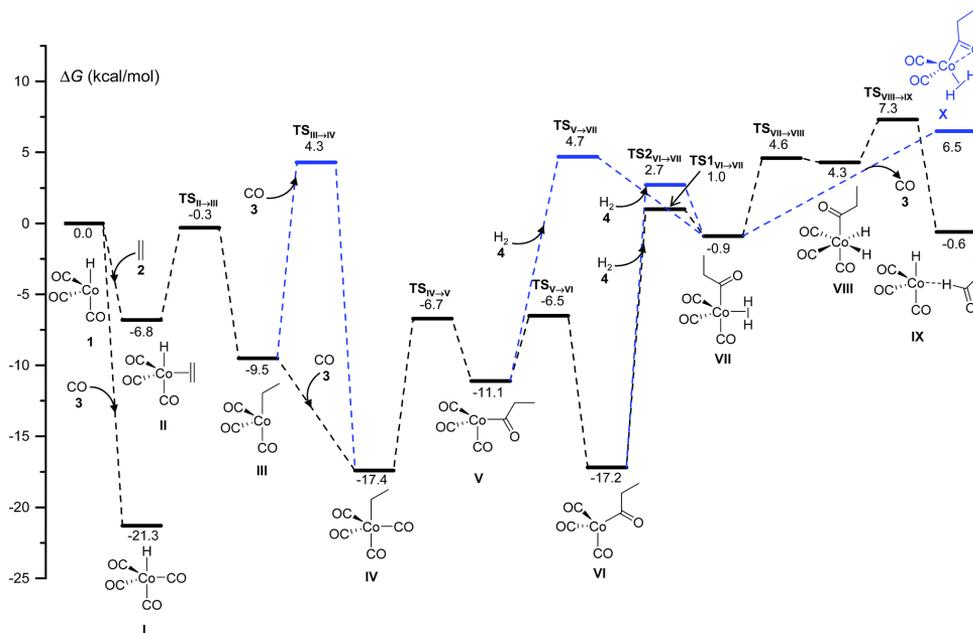


Figure 1.3.: DFT-calculated free energy profile for the Co-catalyzed hydroformylation of ethylene up to the reductive elimination affording the species IX. The blue lines are used for the less competitive pathways observed in the kinetics simulations. The relative free energy values  $\Delta G$  are calculated by subtracting the sum of the free energies of the catalyst and the starting materials.

wasteful side reactions and predict their yields – a unique feature that can be used to optimize the reaction conditions and tune selectivity of the catalytic process. For the test hydroformylation process, alkene hydrogenation was identified as the undesirable side-path and it was found to dominate the catalytic process at very low CO pressures[33].(Figure 1.3)

TSSCDS uses dynamics simulations, which can be efficiently parallelized. The dynamics module contains algorithms allowing non-uniform sampling of the phase space[35] which can accelerate the TS search or guide the dynamics to mechanisms of greater interest. Finally, the method allows using specific reaction parameters in the semiempirical Hamiltonian for systems where standard parameterization is not efficient or where the procedure needs to be speed up by skipping high-energy paths.

*Reiher and coworkers*[36] have put forward an original algorithm for finding vertices in the reaction network that makes use of conceptual electronic-structure theory to apply heuristic rules for the search of potential chemical transformations within complex reaction mechanisms. The heuristic rules guide the construction of high-energy guess-structures of supermolecules composed of the components of the

reactive systems, from which the products of transformations (intermediates in the reaction networks) are derived upon structure optimization. The structures of these intermediates enter an emerging reaction network, in which elementary reactions can be identified in an automated way. In a standard procedure, a species of interest (e.g., a catalyst complex) reacts with a reactive species (e.g., a radical or a charged particle) to produce an intermediate. A collection of all intermediates is arranged in a reaction network. This chemical reaction network can be pruned by defining a certain energy cutoff that excludes consideration of high-energy intermediates that are inaccessible under a range of reasonable physical reaction conditions and within a characteristic time scale. The heuristics-guided exploration protocol by Reiher and coworkers has been applied to the Chatt–Schrock nitrogen-fixation cycle (Figure 1.4)[36]. Its competing reaction paths were not studied in sufficient detail before. A vast number of possible elementary reactions were explored that describe protonation, proton-rearrangement, and reduction steps. The resulting network turned out to be highly complex and alternative routes that still sustain the catalytic cycle emerge.

## 1.4. COMPARISON OF REACTION MECHANISM EXPLORATION TOOLS

Table 1.1 lists the available strategies for reaction mechanism exploration. These can vary significantly in both computational demand and comprehensiveness of the description of the reaction networks that they produce. The Anharmonic downward distortion following (ADDF)[37] method appears to provide the most comprehensive analysis of the chemical space, but its use is also fundamentally limited by rapid increases in cost with growing system size. On the other side of (in)completeness is the heuristics-based approach by the group of Reiher[38] that is largely based on chemical intuition that potentially greatly limits the number of reaction pathways to be explored. Nevertheless, this method still allows identification of a great number of pathways that could not be directly envisaged with the expert knowledge only. Similar to other aspects of computational chemistry and catalysis, automated reaction path analysis faces the same problem needed for establishing a balance between accuracy and quality of the model for the investigated problem, and the associated computational burden[8].

The different approaches for automated reaction path analysis can be classified by the degree and type of human guidance required for their optimal functioning. The knowledge-based approaches such as Reaction Mechanism Generator (RMG)[39] that involve decision making based on the similarities found with the reactions from pre-defined libraries, the performance and depth of analysis is ultimately limited by the quality of the available reaction data.

The reaction exploration by Artificial force-induced reaction (AFIR)[28] and ZStruct[17, 18] strategies could also be limited by the need to impose configurational preferences on the reacting configurations. The initial implementation of ZStruct worked best for intramolecular reactions due to the requirement that reactants need to be pre-aligned. This requirement limited the applicability of the method

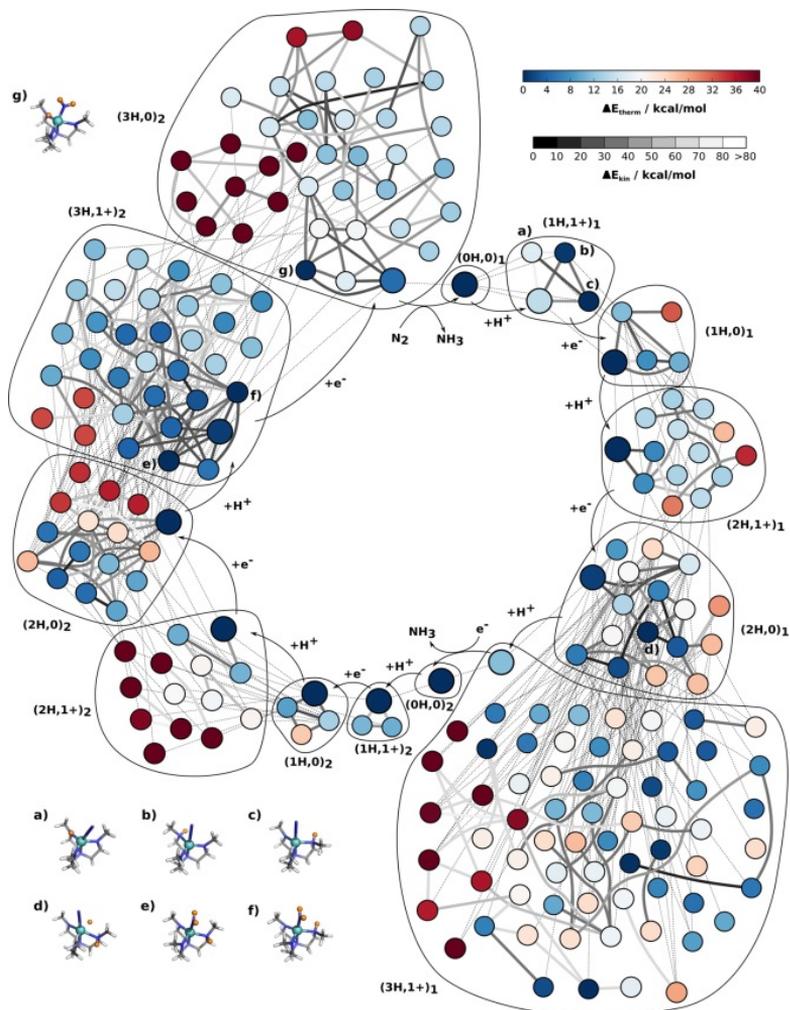


Figure 1.4.: Chatt–Schrock network of catalytic nitrogen fixation. lowest-energy intermediates of a subnetwork(Dark-blue vertices), and corresponding highest-energy intermediates(dark-red vertices). Schrock intermediates (enlarged vertices). Low-energy transition barriers between intermediates of the same subnetwork are indicated by dark-gray edges, high-energy transition barriers by light-gray edges. Internetwork connections are indicated by dashed lines. In a)–g) a selection of intermediates is shown. Element color code: gray, C; blue, N; turquoise, Mo; white, H; orange, H added to reactive sites. Reprinted with permission from *J. Chem. Theory Comput.* 2015, 11, 5712-5722. Copyright 2015 American Chemical Society

for systematic reaction exploration. Furthermore, the ZStruct approach did not guarantee that pairs of intermediates were connected by a single elementary step, causing double-ended GSM to struggle in obtaining a single representative TS for a multistep pathway.

Enhanced collision and reaction acceleration strategies within TSSCDS could also be considered as a limiting factor for large systems as vibrational mode selection for reactivity analysis will inevitably be incomplete and require manual guidance due to the large number of mode combinations that may be populated. Human input is also required in AFIR strategy in the form of selecting which pairs of molecules/catalyst react as well as intramolecular fragment selection (i.e., active atom selection). Generally speaking, all available strategies, with the exception of the Nanoreactor implementation[40, 41], have deficiencies when multiple reactants are involved, or solvent participates in the reaction. Such methods still lack capacity for the truly bias-free mechanism exploration desired for a comprehensive analysis of multistep, multicomponent reaction paths.

For all the available methods, the efficient TS search and optimization is paramount for the overall success in the mechanism discovery. The availability of robust and efficient tools that would consistently converge reaction pathways and TSs is critical for the overall convergence of the above strategies. While methods such as the single-ended GSM[21] and West's[42] TS estimator have provided some advances in this regard, there is so far no method available that would provide a failsafe TS search algorithm. A failed reaction path optimization may mean the path does not exist, it is highly unfavorable, or simply that the optimizer could not reach convergence. This uncertainty is particularly troubling, as the automated approaches would disregard any failed path, even if it were the actual major reaction pathway.

As could be seen from the overview in Table 1.1, the development of automated methods in reaction mechanism exploration is a very active field in contemporary computational chemistry. With the steady progress in the field, we will move closer to uncovering the full details of chemical reactions with less and less guidance needed from the user or user's chemical intuition. Realistic catalytic systems are multicomponent complex systems, in which myriads of potential reaction channels can in principle be found. Therefore, the practical application of the automated and comprehensive reaction network analysis tools requires establishing a balance of exploration-exploitation approaches. Currently, this is achieved either in the framework of the graph-representation of molecular systems or by pre-defining the reactive centers within the molecular ensembles[9]. The former approach is particularly attractive in terms of computational efficiency but has specific limitations when applied to systems with complex electronic structures such as transition metal complexes and clusters, where the application of the concept of valence is not straightforward.

Table 1.1.: Comparison of recent methods for reaction path discovery and TS optimization

Method*	Input Required	Exploration Strategy	TS Finding Strategy	Intermediate
<b>ADDF</b> [30]	Reactant(s)	Anharmonic Downward Distortion(ADD)	Anharmonic mode following then TS optimization	IRC
<b>AFIR</b> [28, 43]	Reactant(s)	Artificial External Force	TS optimization along biased pathway	IRC
<b>TSSCDS</b> [32, 33, 44, 45]	Reactant(s)	High Energy Dynamics	Optimize TS from where bond change occurs	Dynamics and IRC
<b>West</b> [42]	Reactants, library of TS geometries	Interatomic Distances	Reactive atom constraints followed by TS optimization	IRC
<b>ZStruct</b> [46]	Reactant, reactive atoms	Graph Rules	Double-ended reaction path optimization	Graph Rules
<b>Green</b> [47, 48]	Reactant, Reactive atoms	Graph Rules	Freezing String then local TS optimization	Graph Rules
<b>Habershon</b> [49]	Initial reactants and intermediates	Reaction/Graphical Hamiltonian	Double-ended reaction path optimization	Graph Rules
<b>Reiher</b> [36, 38, 50–52]	Reactants, Reactive sites	Reactive Sites (Heuristics)	Interpolation then local TS optimization	Heuristic rules and IRC
<b>Nanoreactor</b> [41]	Reactants	High p, T dynamics	Double-ended reaction path optimization	MD Trajectories
<b>Zstruct2</b> [9]	Reactants, Reactive atoms	Graph Rules	Single-ended growing string	Single-ended growing string
<b>MD/CD</b> [53]	Reactant(s)	Distance between reactive atoms	Interpolation then local TS optimization	Trajectories, interatomic distances
<b>ReNeGate</b> [54]	Reactants	RMSD biased metadynamics	-	Trajectories, Graph theory

The available methodologies require trimming the explored reaction networks to keep them computationally tractable at the expense of potential loss of some of the relevant pathways[9, 36]. Furthermore, to ensure an exhaustive exploration of the chemical space the completeness of the set of transformation rules is required. However, for an arbitrary, unknown chemical system this cannot be guaranteed. One will then be restricted to known or anticipated chemical transformations, which may hamper the discovery of new chemical processes.

Besides these general problems that need to be solved to progress further in this direction such as the inclusion of multicomponent reaction paths, efficient conformational screening and identification of reaction paths, there are more specific challenges related to the field of catalysis by earth-abundant 3d transition metals. This is mainly related to the failure of most of the approximate fast electronic representation methods suitable for exhaustive mechanistic analysis[55]; they may produce qualitatively incorrect results particularly when dealing with multimetallic and/or paramagnetic systems. Similarly, the less common paths involving such effects as single-electron transfer or excitation-induced reactivity may represent a particular challenge for the approximate methods used in such exploration schemes. Thus, an additional major and crucial limitation of all the current approaches for the automated reaction network analysis originates from the method accuracy of the underlying quantum chemical approaches. Because it is not known which reaction paths will emerge from the automated analysis, one cannot apply the expert knowledge to the selection of the most appropriate quantum chemical method for each specific transformation and reaction channel. The reliance on the expert bias and intuition at least at the stage of the initial selection of the computational methodology for a given system appears to be a persistent issue in modern computational chemistry and catalysis.

## 1.5. SCOPE OF THE THESIS AND OUTLOOK:

Electronic structure calculations have become an indispensable tool in catalysis research. They are currently routinely employed to rationalize experimental observations, support mechanistic proposals and even to guide the design of new catalytic systems. Nevertheless, the vast majority of computational studies in catalysis by transition metals still has an explanatory character and focus on describing only a small part of the actual catalyst system. The transition to truly predictive computational modeling requires the development of more complex chemical models that would allow an adequate description of the full reaction networks underlying the catalytic processes. Those approaches can be combined with the topics discussed in this thesis (novel electronic structure theory methods, complex modeling, computation of multiple reaction pathways, reaction network analysis) for building a comprehensive theoretical framework for predictive computational catalysis.

---

Abbreviations: **ADDF**: Anharmonic downward distortion following, **AFIR**: Artificial force-induced reaction, **CD**: Coordinate Driving, **MD**: Molecular dynamics, **IRC**: Intrinsic reaction coordinate, **TS**: Transition State, **TSSCDS**: Transition state search using chemical dynamics simulation

An extensive overview of state of the art in automated mechanistic studies and their application to catalysis by transition metals has been presented **Chapter 1**. We then identified methodological requirements for the development of robust methods suitable for our goals of understanding the complex deactivation chemistry of transition metal catalysts and exploring how those paths could be suppressed through catalyst design. In the next **Chapter 2**, we directly compare the ability of our developed ReNeGate methodology and TSSCDS as the method identified as the most efficient in the application to transition metal catalysis for our goals of the exploration of deactivation channels in Mn(I) carbonyl hydrogenation catalysis. Reaction networks were built starting from similar starting points and results were compared both in comprehensiveness of the results as well as efficiency and scalability of the methodologies. In the next **Chapter 3** we describe the details and underlying theory of the ReNeGate methodology and validate it by considering 3 representative examples relevant for the carbonyl reduction catalysis by Mn(I) homogeneous complexes. Special focus of these case studies was on exploring the rare conversion events accompanying the common catalyst activation procedures. In **Chapter 4**, we extend the application of the ReNeGate automated reactivity discovery approach to studies on large databases of catalytic structures. We combine automated procedures for generating in silico catalyst libraries with the ReNeGate automated reactivity exploration procedures and complement this workflow with the new automated reactivity analysis tool. The new reactivity patterns and insights into the catalytic chemistry are identified from these studies and are discussed in this chapter. The thesis is concluded with a summary and outlook, presenting the key challenges in the modern computational catalysis and outlining the main findings and developments described in this thesis.

## REFERENCES

- [1] O. Acevedo and W. L. Jorgensen. “Advances in quantum and molecular mechanical (QM/MM) simulations for organic and enzymatic reactions”. In: *Accounts of Chemical Research* 43.1 (2010), pp. 142–51. DOI: 10.1021/ar900171c.
- [2] K. E. Riley, M. Pitonak, P. Jurecka, and P. Hobza. “Stabilization and Structure Calculations for Noncovalent Interactions in Extended Molecular Systems Based on Wave Function and Density Functional Theories”. In: *Chemical Reviews* 110.9 (2010), pp. 5023–5063. DOI: 10.1021/cr1000173.
- [3] Y.-h. Lam, M. N. Grayson, M. C. Holland, A. Simon, and K. N. Houk. “Theory and Modeling of Asymmetric Catalytic Reactions”. In: *Accounts of Chemical Research* 49.4 (2016), pp. 750–762. DOI: 10.1021/acs.accounts.6b00006.
- [4] T. Sperger, I. A. Sanhueza, and F. Schoenebeck. “Computation and Experiment: A Powerful Combination to Understand and Predict Reactivities”. In: *Accounts of Chemical Research* 49.6 (2016), pp. 1311–1319. DOI: 10.1021/acs.accounts.6b00068.
- [5] Å. Lernmark. “The Streetlight Effect—Is There Light at the End of the Tunnel?”. In: *Diabetes* 64.4 (2015), pp. 1105–1107. DOI: 10.2337/db15-0011.
- [6] W. A. Haynes, A. Tomczak, and P. Khatri. “Gene annotation bias impedes biomedical research”. In: *Scientific Reports* 8.1 (2018), p. 1362. DOI: 10.1038/s41598-018-19333-x.
- [7] C. S. Hendrix. “Searching for climate–conflict links”. In: *Nature Climate Change* 8.3 (2018), pp. 190–191. DOI: 10.1038/s41558-018-0083-3.
- [8] A. L. Dewyer and P. M. Zimmerman. “Finding reaction mechanisms, intuitive or otherwise”. In: *Organic and Biomolecular Chemistry* 15.3 (2017), pp. 501–504. DOI: 10.1039/c6ob02183b.
- [9] A. L. Dewyer, A. J. Argüelles, and P. M. Zimmerman. “Methods for exploring reaction space in molecular systems”. In: *WIREs Computational Molecular Science* 8.2 (2018), e1354. DOI: 10.1002/wcms.1354.
- [10] I. Ismail, C. Robertson, and S. Habershon. “Successes and challenges in using machine-learned activation energies in kinetic simulations”. In: *The Journal of Chemical Physics* 157.1 (2022), p. 014109. DOI: 10.1063/5.0096027.
- [11] B. Ensing, M. De Vivo, Z. Liu, P. Moore, and M. L. Klein. “Metadynamics as a Tool for Exploring Free Energy Landscapes of Chemical Reactions”. In: *Accounts of Chemical Research* 39.2 (2006), pp. 73–81. DOI: 10.1021/ar040198i.

- [12] L. Alessandro and L. G. Francesco. “Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science”. In: *Reports on Progress in Physics* 71.12 (2008), p. 126601. DOI: 10.1088/0034-4885/71/12/126601.
- [13] V. van Speybroeck and R. J. Meier. “A recent development in computational chemistry: chemical reactions from first principles molecular dynamics simulations”. In: *Chemical Society Reviews* 32.3 (2003), pp. 151–157. DOI: 10.1039/B210410P.
- [14] L.-P. Wang, R. T. McGibbon, V. S. Pande, and T. J. Martinez. “Automated Discovery and Refinement of Reactive Molecular Dynamics Pathways”. In: *Journal of Chemical Theory and Computation* 12.2 (2016), pp. 638–649. DOI: 10.1021/acs.jctc.5b00830.
- [15] H. B. Schlegel. “Geometry optimization”. In: *WIREs Computational Molecular Science* 1.5 (2011), pp. 790–809. DOI: 10.1002/wcms.34.
- [16] E. Martinez-Nunez. “An Automated Method to Find Transition States Using Chemical Dynamics Simulations”. In: *Journal of Computational Chemistry* 36.4 (2015), pp. 222–234. DOI: 10.1002/jcc.23790.
- [17] P. M. Zimmerman. “Automated discovery of chemically reasonable elementary reaction steps”. In: *Journal of Computational Chemistry* 34.16 (2013), pp. 1385–1392. DOI: 10.1002/jcc.23271.
- [18] P. M. Zimmerman. “Navigating molecular space for reaction mechanisms: an efficient, automated procedure”. In: *Molecular Simulation* 41.1-3 (2015), pp. 43–54. DOI: 10.1080/08927022.2014.894999.
- [19] A. L. Dewyer and P. M. Zimmerman. “Simulated Mechanism for Palladium-Catalyzed, Directed gamma-Arylation of Piperidine”. In: *ACS Catalysis* 7.8 (2017), pp. 5466–5477. DOI: 10.1021/acscatal.7b01390.
- [20] P. Zimmerman. “Reliable Transition State Searches Integrated with the Growing String Method”. In: *Journal of Chemical Theory and Computation* 9.7 (2013), pp. 3043–3050. DOI: 10.1021/ct400319w.
- [21] P. M. Zimmerman. “Single-ended transition state finding with the growing string method”. In: *Journal of Computational Chemistry* 36.9 (2015), pp. 601–611. DOI: 10.1002/jcc.23833.
- [22] J. R. Ludwig, P. M. Zimmerman, J. B. Gianino, and C. S. Schindler. “Iron(III)-catalysed carbonyl-olefin metathesis”. In: *Nature* 533.7603 (2016), pp. 374–379. DOI: 10.1038/nature17432.
- [23] J. R. Ludwig, S. Phan, C. C. McAtee, P. M. Zimmerman, I. Devery James J., and C. S. Schindler. “Mechanistic Investigations of the Iron(III)-Catalyzed Carbonyl-Olefin Metathesis Reaction”. In: *Journal of the American Chemical Society* 139.31 (2017), pp. 10832–10842. DOI: 10.1021/jacs.7b05641.

- [24] M. L. Smith, A. K. Leone, P. M. Zimmerman, and A. J. McNeil. "Impact of Preferential pi-Binding in Catalyst-Transfer Polycondensation of Thiazole Derivatives". In: *ACS Macro Letters* 5.12 (2016), pp. 1411–1415. DOI: 10.1021/acsmacrolett.6b00886.
- [25] L. V. A. Hale, T. Malakar, K.-N. T. Tseng, P. M. Zimmerman, A. Paul, and N. K. Szymczak. "The Mechanism of Acceptorless Amine Double Dehydrogenation by N,N,N-Amide Ruthenium(II) Hydrides: A Combined Experimental and Computational Study". In: *ACS Catalysis* 6.8 (2016), pp. 4799–4813. DOI: 10.1021/acscatal.6b01465.
- [26] B. R. Ellington, B. Paul, D. Das, A. K. Vitek, P. M. Zimmerman, and E. N. G. Marsh. "An Unusual Iron-Dependent Oxidative Deformylation Reaction Providing Insight into Hydrocarbon Biosynthesis in Nature". In: *ACS Catalysis* 6.5 (2016), pp. 3293–3300. DOI: 10.1021/acscatal.6b00592.
- [27] W. M. C. Sameera, S. Maeda, and K. Morokuma. "Computational Catalysis Using the Artificial Force Induced Reaction Method". In: *Accounts of Chemical Research* 49.4 (2016), pp. 763–773. DOI: 10.1021/acs.accounts.6b00023.
- [28] S. Maeda, Y. Harabuchi, M. Takagi, T. Taketsugu, and K. Morokuma. "Artificial Force Induced Reaction (AFIR) Method for Exploring Quantum Chemical Potential Energy Surfaces". In: *The Chemical Record* 16.5 (2016), pp. 2232–2248. DOI: 10.1002/tcr.201600043.
- [29] M. Hatanaka, S. Maeda, and K. Morokuma. "Sampling of Transition States for Predicting Diastereoselectivity Using Automated Search Method-Aqueous Lanthanide-Catalyzed Mukaiyama Aldol Reaction". In: *Journal of Chemical Theory and Computation* 9.7 (2013), pp. 2882–2886. DOI: 10.1021/ct4002637.
- [30] M. Hatanaka and K. Morokuma. "Role of Water in Mukaiyama-Aldol Reaction Catalyzed by Lanthanide Lewis Acid: A Computational Study". In: *Journal of the American Chemical Society* 135.37 (2013), pp. 13972–13979. DOI: 10.1021/ja407357c.
- [31] M. Hatanaka and K. Morokuma. "How Can Fluctional Chiral Lanthanide (III) Complexes Achieve a High Stereoselectivity in Aqueous Mukaiyama-Aldol Reaction?" In: *ACS Catalysis* 5.6 (2015), pp. 3731–3739. DOI: 10.1021/acscatal.5b00438.
- [32] E. Martínez-Núñez. "An automated transition state search using classical trajectories initialized at multiple minima". In: *Physical Chemistry Chemical Physics* 17 (22 2015), pp. 14912–14921. DOI: 10.1039/C5CP02175H.
- [33] J. A. Varela, S. A. Vazquez, and E. Martinez-Nunez. "An automated method to find reaction mechanisms and solve the kinetics in organometallic catalysis". In: *Chemical Science* 8.5 (2017), pp. 3843–3851. DOI: 10.1039/c7sc00549k.
- [34] R. F. Heck and D. S. Breslow. "Reaction of cobalt hydrotetracarbonyl with olefins". In: *Journal of the American Chemical Society* 83.19 (1961), pp. 4023–4027. DOI: 10.1021/ja01480a017.

- [35] E. Martínez-Núñez and D. V. Shalashilin. “Acceleration of Classical Mechanics by Phase Space Constraints”. In: *Journal of Chemical Theory and Computation* 2.4 (2006), pp. 912–919. DOI: 10.1021/ct060042z.
- [36] M. Bergeler, G. N. Simm, J. Proppe, and M. Reiher. “Heuristics-Guided Exploration of Reaction Mechanisms”. In: *Journal of Chemical Theory and Computation* 11.12 (2015), pp. 5712–22. DOI: 10.1021/acs.jctc.5b00866.
- [37] S. Maeda, T. Taketsugu, K. Morokuma, and K. Ohno. “Anharmonic Downward Distortion Following for Automated Exploration of Quantum Chemical Potential Energy Surfaces”. In: *Bulletin of the Chemical Society of Japan* 87.12 (2014), pp. 1315–1334. DOI: 10.1246/bcsj.20140189.
- [38] G. N. Simm and M. Reiher. “Context-Driven Exploration of Complex Chemical Reaction Networks”. In: *Journal of Chemical Theory and Computation* 13.12 (2017), pp. 6108–6119. DOI: 10.1021/acs.jctc.7b00945.
- [39] C. W. Gao, J. W. Allen, W. H. Green, and R. H. West. “Reaction Mechanism Generator: Automatic construction of chemical kinetic mechanisms”. In: *Computer Physics Communications* 203 (2016), pp. 212–225. DOI: 10.1016/j.cpc.2016.02.013.
- [40] S. H. Petrosko, R. Johnson, H. White, and C. A. Mirkin. “Nanoreactors: Small Spaces, Big Implications in Chemistry”. In: *Journal of the American Chemical Society* 138.24 (2016), pp. 7443–7445. DOI: 10.1021/jacs.6b05393.
- [41] L.-P. Wang, A. Titov, R. McGibbon, F. Liu, V. S. Pande, and T. J. Martinez. “Discovering chemistry with an ab initio nanoreactor”. In: *Nature Chemistry* 6.12 (2014), pp. 1044–1048. DOI: 10.1038/nchem.2099.
- [42] P. L. Bhoorasingh and R. H. West. “Transition state geometry prediction using molecular group contributions”. In: *Physical Chemistry Chemical Physics* 17.48 (2015), pp. 32173–82. DOI: 10.1039/c5cp04706d.
- [43] S. Maeda, K. Ohno, and K. Morokuma. “Systematic exploration of the mechanism of chemical reactions: the global reaction route mapping (GRRM) strategy using the ADDF and AFIR methods”. In: *Physical Chemistry Chemical Physics* 15.11 (2013), pp. 3683–3701. DOI: 10.1039/C3CP44063J.
- [44] R. J. Shannon, E. Martínez-Núñez, D. V. Shalashilin, and D. R. Glowacki. “ChemDyME: Kinetically Steered, Automated Mechanism Generation through Combined Molecular Dynamics and Master Equation Calculations”. In: *Journal of Chemical Theory and Computation* 17.8 (2021), pp. 4901–4912. DOI: 10.1021/acs.jctc.1c00335.
- [45] S. A. Vázquez and E. Martínez-Núñez. “HCN elimination from vinyl cyanide: product energy partitioning, the role of hydrogen–deuterium exchange reactions and a new pathway”. In: *Physical Chemistry Chemical Physics* 17.10 (2015), pp. 6948–6955. DOI: 10.1039/C4CP05626D.
- [46] P. M. Zimmerman. “Growing string method with interpolation and optimization in internal coordinates: Method and examples”. In: *The Journal of Chemical Physics* 138.18 (2013). DOI: 10.1063/1.4804162.

- [47] C. A. Grambow, A. Jamal, Y.-P. Li, W. H. Green, J. Zádor, and Y. V. Suleimanov. “Unimolecular Reaction Pathways of a  $\gamma$ -Keto hydroperoxide from Combined Application of Automated Reaction Discovery Methods”. In: *Journal of the American Chemical Society* 140.3 (2018), pp. 1035–1048. DOI: 10.1021/jacs.7b11009.
- [48] Y. V. Suleimanov and W. H. Green. “Automated Discovery of Elementary Chemical Reaction Steps Using Freezing String and Berny Optimization Methods”. In: *Journal of Chemical Theory and Computation* 11.9 (2015), pp. 4248–4259. DOI: 10.1021/acs.jctc.5b00407.
- [49] S. Habershon. “Automated Prediction of Catalytic Mechanism and Rate Law Using Graph-Based Reaction Path Sampling”. In: *Journal of Chemical Theory and Computation* 12.4 (2016), pp. 1786–1798. DOI: 10.1021/acs.jctc.6b00005.
- [50] J. P. Unsleber, S. A. Grimmel, and M. Reiher. “Chemoton 2.0: Autonomous Exploration of Chemical Reaction Networks”. In: *Journal of Chemical Theory and Computation* 18.9 (2022), pp. 5393–5409. DOI: 10.1021/acs.jctc.2c00193.
- [51] M. Steiner and M. Reiher. “Autonomous Reaction Network Exploration in Homogeneous and Heterogeneous Catalysis”. In: *Topics in Catalysis* 65.1 (2022), pp. 6–39. DOI: 10.1007/s11244-021-01543-9.
- [52] J. P. Unsleber and M. Reiher. “The Exploration of Chemical Reaction Networks”. In: *Annual Review of Physical Chemistry* 71.1 (2020), pp. 121–142. DOI: 10.1146/annurev-physchem-071119-040123.
- [53] M. Yang, J. Zou, G. Wang, and S. Li. “Automatic Reaction Pathway Search via Combined Molecular Dynamics and Coordinate Driving Method”. In: *The Journal of Physical Chemistry A* 121.6 (2017), pp. 1351–1361. DOI: 10.1021/acs.jpca.6b12195.
- [54] A. Hashemi, S. Bougueroua, M.-P. Gaigeot, and E. A. Pidko. “ReNeGate: A Reaction Network Graph-Theoretical Tool for Automated Mechanistic Studies in Computational Homogeneous Catalysis”. In: *Journal of Chemical Theory and Computation* 18.12 (2022), pp. 7470–7482. DOI: 10.1021/acs.jctc.2c00404.
- [55] S. Spicher and S. Grimme. “Robust Atomistic Modeling of Materials, Organometallic, and Biochemical Systems”. In: *Angewandte Chemie-International Edition* 59.36 (2020), pp. 15665–15673. DOI: 10.1002/anie.202004239.



# 2

## COMPARATIVE BENCHMARK STUDY

*The exploration of secondary reaction paths in homogeneous Mn(I) reduction catalysis using the TS Search via Chemical Dynamic Simulations (TSSCDS) method and comparison with the ReNeGate approach*

**S**ECONDARY reaction paths in homogeneous Mn(I) reduction catalysis are explored using the TS Search via Chemical Dynamic Simulations (TSSCDS) method. The reaction networks as well as the efficiency and scalability of the TSSCDS method are evaluated and compared with our perspective ReNeGate methodology. TSSCDS has been chosen as the method identified as the most efficient in the application to transition metal catalysis for our goals of the exploration of deactivation channels in Mn(I) carbonyl hydrogenation catalysis. Reaction networks for TSSCDS and ReNeGate were built starting from similar starting points and results were compared both in comprehensiveness of the results as well as efficiency and scalability of the methodologies. TSSCDS procedure identifies many TS structures, that are merely isomerization of rotamers. Further studies with an auxiliary static exploration strategy have proven the sensitivity of the TSSCDS outcome on to the initial starting configuration. Results from the static exploration were, however, shown to lead to unphysical structures where multiple bonds were broken and formed simultaneously. Stepwise following of the reaction steps is also not possible for structures observed from static explorations. When focused on finding minima corresponding to potential deactivated state, TSSCDS is found to be excessively expensive. A comparative study with the alternative ReNeGate workflow described in detail in the following chapter showed that similar number of unique structures can be identified and reaction network can be constructed using an order of magnitude less time. Furthermore, this study revealed inefficiency of the vibrational-excitation guided MD for chemistry exploration. The development of our ReNeGate workflow therefore was done using RMSD-biased metadynamics simulations. With the scalable ReNeGate reaction network discovery tools, extension to more realistic representation of chemical reactions is envisaged through considering multiple competitive/cooperative interactions among multiple catalyst- substrate-solvent combinations. Such improvements in the scalability of the exploration algorithms has also paved the way for exploration efficiency for studies on databases of catalyst structures.

## 2.1. INTRODUCTION

**T**HE common practice of homogeneous catalysis is the use of relatively stable pre-catalyst species in the form of either the molecularly-defined transition metal (TM) complexes or a combination of a TM precursor and a ligand, which upon the exposure to the reactants or specific activating additives under the conditions of the catalytic process would undergo a chemical transformations to form the reactive species capable of engaging in and propagating the catalytic cycle. The successful catalyst would establish a smooth energy profile for the conversion of the substrates to the products with all intermediates showing moderate stability and low barriers for their interconversions. The formation of too stable intermediates during the activation step or in the course of the catalytic reaction would naturally mean the need for overcoming a higher energy barrier to propagate the catalytic cycle, whereas the formation of unstable intermediates would also inherently be associated with the need to add energy to the system to reach that state. Besides the catalytic cycle, the precatalyst as well as the activated species may engage in other (less likely, for a highly performing catalyst) transformations that would lead to catalyst deactivation, which could be attributed to the transition to a molecular state, from which it would be prohibitively expensive from the energy viewpoint to return to the catalytic cycle. If such a state has an intermediate stability, it can be considered a resting or inhibited state. Very high stability of the species would indicate a long-term deactivation of the catalytic species.

For a successful catalyst, all these side-reactions should proceed with the barriers higher than those of the elementary steps in the catalytic cycle to ensure the kinetic stabilization of the catalytic species and successful catalytic turnover[1, 2]. Rare events of side-reactions would eliminate catalytic species from the reaction mixture by forming the stable deactivated intermediates and decreasing the overall efficiency[3, 4]. We propose that the identification of such stable states can be used to optimize the deactivation behavior of the catalyst system and use it as a tool for the design and optimization of the catalyst system. The assumption that the catalyst deactivation is dictated by the formation of thermodynamically highly stable species connected to some catalytically relevant state redefines the inherently complex and difficult to solve kinetic problem into a thermodynamic one that is much more tractable from the computational chemistry standpoint.

The secondary nature, high variety and complexity of such side-reaction channels makes it particularly challenging to study them using conventional expert-guided mechanistic analysis strategies. Chapter 1 presented a concise overview of the automated reaction network analysis approaches that could potentially be used to unravel the potential secondary conversion paths that could be associated with such deactivation and non-selective catalyst conversion channels. Among the different strategies available, the dynamic screening approaches such as that implemented in TSSCDS5 provide the necessary expert-bias free framework for the exploration of the unknown conversion paths of the catalytic species under the reaction conditions.

The power and capabilities of TSSCDS methodology for mechanistic studies in homogeneous catalysis by transition metal complexes has been demonstrated with a relevant example of the cobalt-catalyzed ethylene hydroformylation[5]. The method

has revealed the main pathway proposed by Heck and Breslow[6, 7] and also identified alternative paths involving the Co-catalyzed conversion of ethylene, CO and H<sub>2</sub> substrates. A total of 230 chemical species and 448 elementary reactions were discovered. Among the 448 reactions, 57 were barrierless and the remaining 391 contained defined transition states. The reaction network produced besides the main hydroformylation product propanal also the hydrogenation side-products (ethane and propene) as well as the species resulting from decarbonylation (formaldehyde) and dehydration (water) reaction paths. This study illustrates the power and capabilities of the automated TSSCDS method for successful DFT-based exploration of reaction mechanisms in homogeneous catalysis by organometallic complexes. Importantly, the TSSCDS could predict wasteful side reactions and their relative rates, which could be used to optimize the reaction conditions.

In this chapter, TSSCDS method has been employed to explore the reaction networks underlying the transformations of a representative Mn-based non-pincer catalyst for selective carbonyl reduction previously studied computationally and experimentally in the ISE group[8, 9]. The main focus was on analyzing the potential of this method for the discovery of secondary reaction channels for 3d transition metal catalysts and evaluating the compatibility of this approach with the perspective high-throughput reactivity exploration workflows. To accelerate the TSSCDS procedure, the auxiliary “static exploration” method provided with TSSCDS has been explored to improve on the iterative approach used by TSSCDS for finding new minima structures as starting points for TSSCDS calculations. The results obtained from the static exploration have been analyzed and the efficiency of this auxiliary tool has been evaluated. When applying TSSCDS method to our Mn-based catalyst system and having in mind the extension of this approach to high-throughput computational analysis, some methodological drawbacks were identified that were circumvented in our own ReNeGate[10] automated graph theory-based reaction network analysis pipeline, which will be presented and discussed in detail in the next Chapter 3. This chapter will primarily focus on the analysis of the results obtained with the TSSCDS platform and briefly compare them with the outcome of the ReNeGate method.

It should be noted that while TSSCDS methodology is focused on searching for TS guesses, ReNeGate pipeline emphasizes the search for minima and deactivated species. While ReNeGate has been developed with a focus on finding deactivated species and is designed to be scalable for high throughput computations, TSSCDS is clearly advantageous when a complete and comprehensive reaction network needs to be constructed.

The chapter is organized as follows: Section 2.2 presents a brief overview of the TSSCDS methodology followed by Section 2.3 that compares the results of the reaction network analysis carried out with TSSCDS and ReNeGate workflows on the same molecular system. Our analysis highlights the potentially higher efficiency of ReNeGate in identifying low energy intermediates due to differences in structure identification algorithms between the two workflows. The chapter is completed with the conclusion section that summarizes the key findings and provides suggestions on further improvements to ReNeGate methodology.

## 2.2. TSSCDS METHOD FOR ACCELERATED REACTION

### NETWORK ANALYSIS

TSSCDS methodology is based on following high-temperature molecular dynamics (MD) trajectories on the defined molecular ensemble representing the reactive system of interest, during which reactive events take place involving breaking and making new bonds.<sup>5</sup> The geometries formed along the trajectories are analysed using a post-processing algorithm to find reactive pathways. Geometries with partially formed/broken bonds are gathered and investigated as guess structures for the subsequent transition state (TS) optimizations. The MD simulations produce a collection of guess TS structures, located based on changes in the connectivity matrix for multiple parallel trajectories. The geometries of these guess TS structures are optimized, and intrinsic reaction coordinate (IRC) calculations are performed to connect the TSs and reaction intermediates and thus build the reaction network.

The method can theoretically be used in an iterative fashion, in which the energetically favourable minima found in the first iteration of the MD exploration and IRC calculations, can serve as the starting geometries for the next iteration. Such an iterative strategy comes with the exponential growth in the number of species explored in the chemical space. By default the TSSCDS employs two levels of electronic structure theory for the reaction space exploration and the refinement of the nodes (minima and TSs) within the reaction networks. The Lower Level (LL) calculations are carried out using a semi-empirical method PM6,<sup>[11]</sup> whereas the Higher Level (HL) calculations are carried out at the DFT B3LYP/6-31g(d,p)<sup>[12, 13]</sup> level of theory. MOPAC2016<sup>[14]</sup> package is used in the TSSCDS workflow for the production of guess TS structures via semi-empirical reactive MD simulations as well as for the subsequent IRC calculations. The LL data is refined through the coarse graining strategy followed by energy and structure refinement at the HL (DFT) level with Gaussian 16 rev. C0.1 program<sup>[15]</sup>.

The reaction network exploration with TSSCDS starts with the definition of the initial state that, in the context of this work, is a molecular mixture model representing the relevant pre-catalyst and the key reaction components. The steps of TSSCDS strategy employed for the reaction network construction are summarized in Figure 2.1. In brief, the input coordinates are optimized, and the frequencies of the initial structure are first calculated at LL for further use in normal mode sampling and is followed by accelerated MD simulations, which explore the reaction space. A microcanonical or canonical ensemble of vibrationally excited molecules is constructed using the 3N-6 (or 3N-5 if the molecular system (MS) is linear) vibrational normal modes computed in the previous step. Because the molecular system is highly vibrationally excited, the dynamics are accelerated and allow a more comprehensive exploration of the potential energy surface (PES). The trajectories produced by the dynamic simulations are then analysed automatically to identify guess transition states.

An algorithm within TSSCDS is specifically designed to find guess TS structures that involve bond breakage/formation search (BBFS). BBFS provides the initial list of TSs (first iteration), which are used for the subsequent IRC calculations to connect

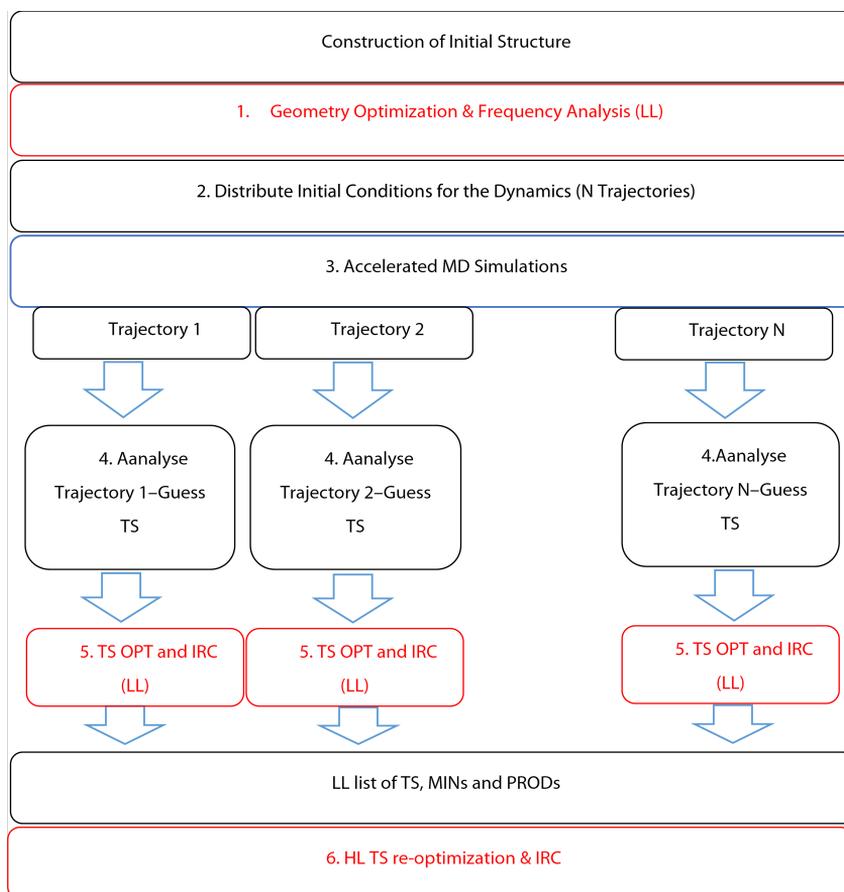


Figure 2.1.: Flowchart of the TSSCDS methodology for automated reaction network analysis.

the TSs with the adjacent intermediates. Primary “All states” Reaction networks (RXNet) are then constructed including all the identified guess TS structures and the associated IRC calculations. The intermediates identified from the IRC calculations are labelled as minima(MINs) or products (PRODs) based on the relative number of fragments compared to the starting input structure. These extended networks are next refined into Coarse Grained networks (RXNetCG). The refinement is made based on grouping of isomers with similar structures in single nodes. The intermediates found from the reactive trajectories are also refined at HL energy and coarse grained networks are formed for results obtained from similar optimization and IRC calculations performed at LL. For further details on the identification of MINs and PRODs as well as coarse graining for networks, we refer the interested reader to the TSSCDS publication[5].

## 2.3. AUTOMATED REACTION EXPLORATIONS FOR ACETONE CONVERSION

### 2.3.1. HOMOGENEOUS CATALYSIS FOR SELECTIVE CARBONYL REDUCTION

Catalytic selective reduction of carbonyl derivatives such as aldehydes, ketones and carboxylic acid derivatives with molecular hydrogen has evolved as a solution towards the synthesis of bulk and fine chemicals[16–19]. While pharmaceutical industry primarily uses sodium borohydride and lithium aluminium hydride for stoichiometric reduction processes, use of molecular hydrogen as reducing agent has proven to be more efficient and environmentally friendly[20]. The focus in recent years has shifted from application of noble metal-based catalysts[17] (Ru, Os, Ir) to early transition-metal-based catalysts (Fe[21], Co[22–24], and Mn[16, 25–27]).

In recent years, manganese has become a desirable metal for pharmaceutical applications due to its wide availability and good biocompatibility[16, 28]. Manganese's use as a catalyst has expanded quickly because of how well it reduces challenging carbonyl substrates like carboxylic acid esters and nitriles[9]. This development can partially be attributed to the use of proven pincer ligand platforms[29, 30] rather than specifically designed ligands. Manganese and noble metals have different chemical reactivity, although the causes of these differences are still not entirely understood. Manganese's role in hydrogenation catalysis was poorly understood prior to 2016[16, 27, 31], however it has since been demonstrated to be useful in the hydrogenation of nitriles, ketones, and aldehydes[32]. Numerous mechanistic investigations proposed Mn-hydride as the active species in the hydrogenation catalysis[33–36].

It has been found that during the catalytic process, such species can convert to more stable off-cycle intermediates and hinder the reactivity or undergo secondary conversion resulting in elimination of the reactive species from the catalytic reaction[8]. Understanding the diverse reaction paths constituting the catalytic mechanism as well as various potential side-reactions in transition metal catalysis is very challenging both experimentally and computationally. To investigate the ability of advanced automated procedures to reveal the known mechanistic details of representative catalyst systems as well as to suggest alternative reaction paths, we employed the TSSCDS method to explore the chemical conversion paths of the molecularly defined Noyori-type Mn hydrido tricarbonyl diamine catalyst ( $\text{H}(\text{CO})_3\text{Mn-NN}$ ). For the reactivity analysis, a highly reductionist model was used consisting of the Mn-hydrido active species and acetone reagent molecule. The metal bound hydrides are commonly proposed as the active sites for the transfer hydrogenation of ketones[9]. Here we specifically aimed at testing if the automated reaction network procedures will be able to identify favorable conversion routes and predict its higher stability compared to the reactive Mn-hydrido catalyst, as established by previous experimental and computational studies.

### 2.3.2. TSSCDS AND ReNeGate REACTION NETWORK COMPARISON RESULTS

Figure 2.2 presents the comparison of the original and processed reaction networks explored by TSSCDS and ReNeGate. For the purpose of comparing the fingerprinting and reaction network exploration efficiencies, we have provided the 3D structure of the catalytic system as xyz input to TSSCDS workflow (Figure 2.2a). The reactive trajectories produced based on the TSSCDS sampling scheme has been provided to both TSSCDS and ReNeGate workflows for further analysis. The reaction networks constructed using the TSSCDS are summarized in Figure 2.2b. The pipeline generates two types of the networks, namely, the RXNet (All states) (Figure 2.2.b.1) and RXNetCG “Coarse grained Network” (Figure 2.2.b.2) produced from RXNet and RXNetCG output files of the TSSCDS calculations. For comparison purposes, we reformulated the connections given in these files to prepare the reaction network representations. The TSSCDS pipeline[37] identifies species in the reactive MD trajectories and labels them as MIN (minima) or PROD (product) based on the number of fragments present in comparison with the reference state. In other words, PRODs are species with more fragments compared to the reference state. Each MIN and PROD state present in RXNet and RXNetCG files has been considered as node in the network with edges representing pairs connected by IRC calculations. Labels on the edges for RXNetCG network are based on TS energies connecting the two MIN-MIN or MIN-PROD pairs of structures. Edges are colored according to the mentioned relative TS energies in kcal.mol<sup>-1</sup>. The “All states” RXNet reaction network formed from TSSCDS explorations includes 270 (redundant) species connected with single edges between pairs of nodes. Similar to RXNet, in RXNetCG network, nodes represent structural isomers of the same MIN or PROD species identified in the RXNet network. Edges also represent summation of all connections previously identified through IRC calculations between different representative species present in the RXNet network. Similar to RXNet, RXNetCG edges are colored according to the mentioned lowest relative TS energies in kcal.mol<sup>-1</sup>.

Same reactive MD trajectory has been provided as input to the ReNeGate fingerprinting algorithm. Figure 2.2c shows the reference (Figure 2.2c.1) and trimmed (Figure 2.2c.2) ReNeGate reaction networks. Briefly, trimming is the process of removing physically impossible nodes and edges from the reference network based on (pre-defined) threshold values for species and transformations. Nodes represent unique species found and are labeled based on the given ID and frequency of appearance of the specific node in the trajectory. IDs for nodes are based on the order of the first appearance of unique conformer in the reactive trajectory. Directed edges connecting nodes are labeled with the frequency of the specific (directed) transformation. Further details on the definitions of reference reaction network, trimmed network and trimming process can be found in Chapter 3. As an improvement to the initial TSSCDS network, TSSCDS coarse graining algorithm has been applied to the nodes present in the initial network and has led to a grouped network of 32 structures. Trimming on the initial ReNeGate network has also led to a network of 6 unique structures. Further details on the trimming process are discussed in the next chapter where the ReNeGate workflow will be presented in

detail but are used here for the purposes of comparison of efficiency.

Unique structures observed in TSSCDS and ReNeGate networks were manually inspected and the comparison results are summarized in Figure 2.3. Figure 2.3 proves that both algorithms have found the same number of unique species in the 50 kcal.mol<sup>-1</sup> energy window from the starting structure. It is worth mentioning, however, that while the TSSCDS RXNet network includes 270 nodes, the initial ReNeGate network directly identifies the unique 13 species verified by manual inspection. Such non-redundant enumeration of transformations is of critical importance to scalability of higher level (DFT) calculations required when applied to more complex model systems. Further proof for such scalability are provided in Chapters 3 and 4 with the application to more complex model systems and databases of catalyst structures.

Figure 2.3 summarizes all unique structures identified with TSSCDS and ReNeGate algorithms based on the given input structure. Boxplots of Figure 2.3, represent the differences in energy for isomers of the same structure type. In order to differentiate between the two methods, unique labels for TSSCDS structures are shown in red. New structures within 50 kcal.mol<sup>-1</sup> range from the starting input structure have been found with both workflows. ReNeGate has categorized 24 TS structures from TSSCDS into 6 unique structures. High energy structures TS28, TS29 and TS30 have been found only with TSSCDS methodology. For all conformers identified with ReNeGate, structures for isomers of the same structure have been overlapped and are shown on top of the column.

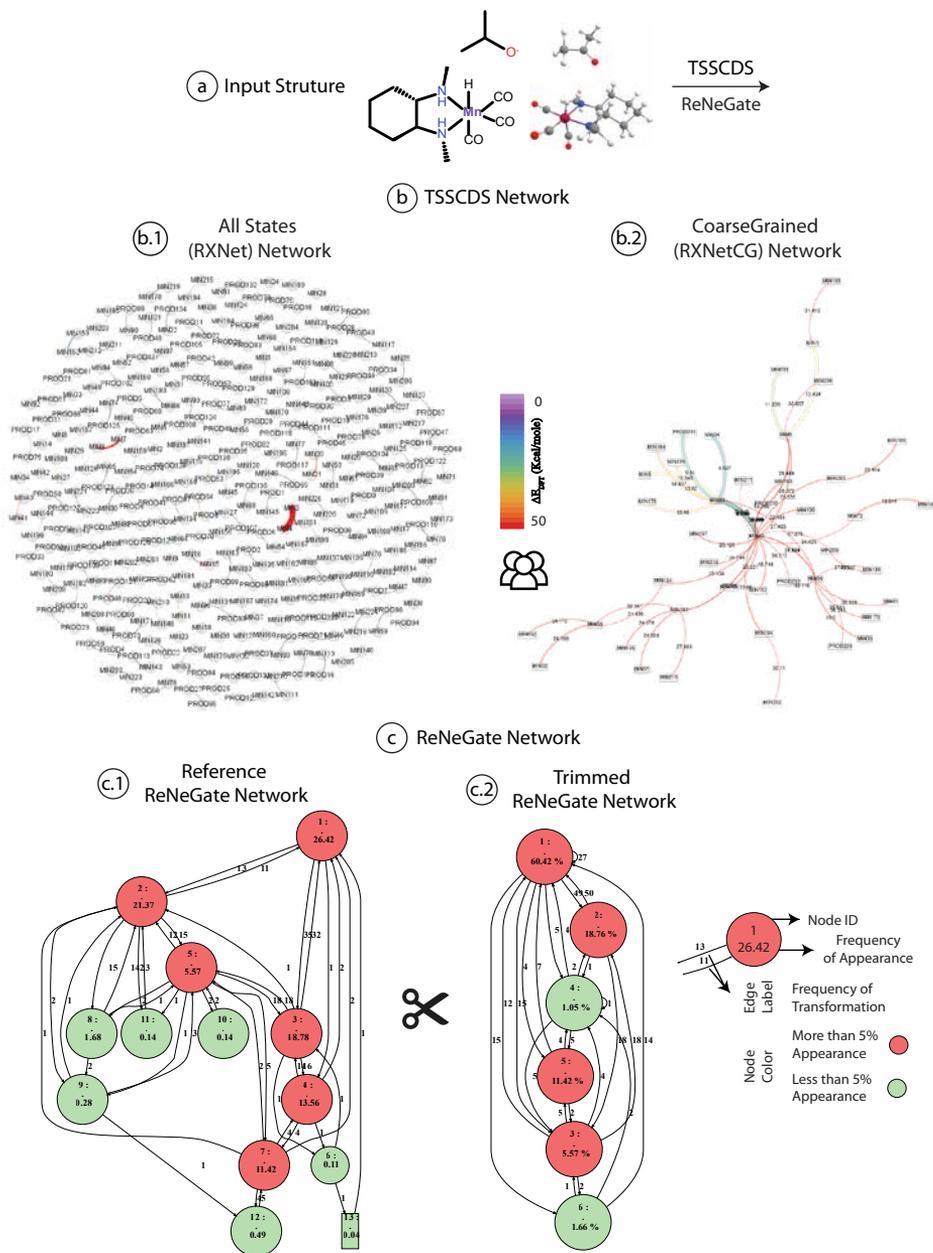


Figure 2.2.: Comparative illustrations of ReNeGate and TSSCDs Networks for conformer explorations on  $\text{H}(\text{CO})_3\text{Mn-NN}$  input structure(a): TSSCDs Networks at two different levels all states (RXNet) (b.1) and Coarse grained (RXNetCG) (b.2) as well as (b), Reference ReNeGate Network(c.1) and Trimmed ReNeGate network (c.2)

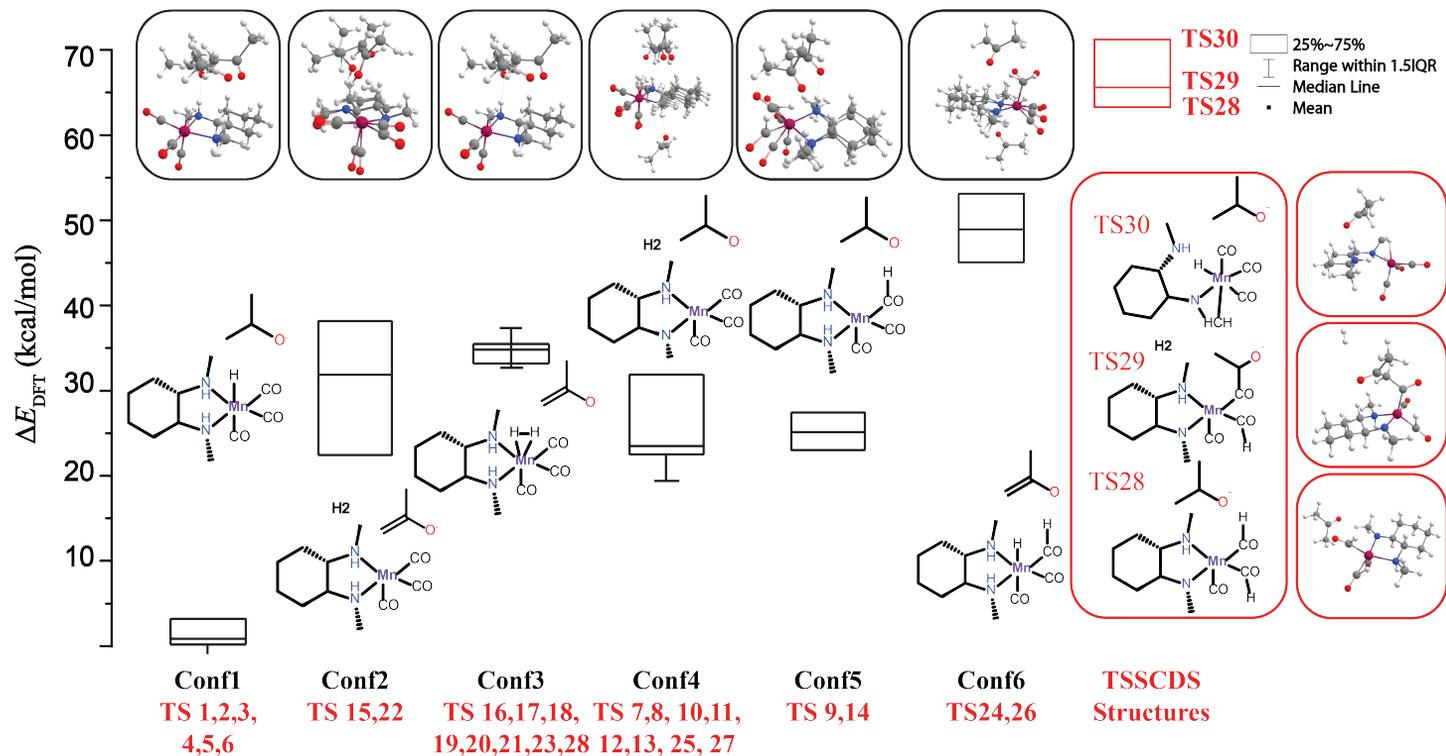


Figure 2.3.: Comparison of the stability of the unique structures found from TSSCDS and ReNeGate explorations: ReNeGate and TSSCDS labels for structures with similar bonding patterns are shown in black and red respectively. Boxplots represents energy ranges for isomers. Structures for isomers of the same structure have been overlapped and are shown on top of the column

In brief, the automated reaction network exploration identified the following new intermediates. Conf1(Configuration1) represent the manganese hydride as starting structure and the configurational isomers of the same structure. The hydride abstraction from the alkoxide led to the formation of 5-coordinated Mn-amido complex as well as non-coordinated H<sub>2</sub> and enolate species in Conf2. Similar transformations as in Conf2 but without H<sub>2</sub> release led to the formation of a  $\sigma$ -H<sub>2</sub> complexation with an octahedrally-coordinated Mn in species Conf3. As observed for Conf6 in Figure 2.3, direct insertion of the Mn-hydride into the carbonyl moiety is also observed and results in high energy (45-55 kcal.mol<sup>-1</sup>) products. For the formation of Conf 4&5, the proton for the respective formation of H<sub>2</sub> or insertion into the carbonyl moiety could also come from the NH moiety bound to the Mn center. Comparison between configurations 2 and 4 shows that abstraction of the proton from the NH moiety is energetically more favorable by ca. 25 kcal.mol<sup>-1</sup>. Double migration of protons, from Mn hydride and NH moiety to carbonyls is observed for TS28. TS29 includes simultaneous hydride migration from Mn center and enolate insertion into two different carbonyls, while H<sub>2</sub> is released. Decoordination of the N donor is observed to be accompanied by agostic Mn-H interaction with the methyl group on the nitrogen donor for TS30.

### 2.3.3. CHALLENGES OBSERVED WITH TSSCDS AS A SCALABLE TOOL FOR REACTION NETWORK ANALYSIS

TSSCDS strategies for selection of a set of guess structures along each RP is designed to improve the chances of finding TS structures. However, inspection of the proposed TS structures has shown that vast majority of TSs are not actual chemical reactions but merely isomerizations of the organic parts and do not reveal any real chemical transformation. Comparison of the MINs and PRODs with ReNeGate results have shown challenges in scalability of TSSCDS algorithms in their application to more complex systems.

Further investigations on the cost and efficiency of the calculations required to converge the studies for the same input structure are summarized in Table 2.1. For the purposes of the current investigation, TSSCDS and ReNeGate have identified similar structures in the 50 kcal.mol<sup>-1</sup> window from the input structure. However, ReNeGate calculations required an order of magnitude less computational hours than TSSCDS calculations. Considering an average 12 hour time for the convergence of TS and IRC calculations, ReNeGate has proven to be ca. 13 times more efficient for calculations on a simple system. In other words, while convergence to similar results within 50 kcal.mol<sup>-1</sup> for TSSCDS requires ca. 2000 cpu hours, ReNeGate calculations will converge in 150 CPU hours. It should be noted, however, that such extensive calculations are completely justified for the purposes of exhaustive TS search and comprehensive mechanistic explorations on instances of catalytic mixtures. While such differences are acceptable for single studies on simple systems, it renders the application of methodologies like TSSCDS either to more complex systems or high throughput calculations on large databases of catalytic structures very challenging.

Table 2.1.: Comparison of calculations required to converge the studies for the same input structure

	LL Jobs		DFT Jobs		Identified Species		
	TS	IRC	TS	IRC	Total Species	Unique Species	Unique within 50 kcal.mol <sup>-1</sup>
<b>TSSCDS</b>	54	108	54	108	162	9	6
<b>ReNeGate</b>	Trimming based on trajectory energies (0)		13(OPT)		13	6	6

### 2.3.4. EXHAUSTIVE “STATIC” CONFORMATIONAL SEARCH FOR FAVORABLE ASSOCIATION COMPLEXES: A PRE-PROCESSING STEP TO IMPROVE THE EFFICIENCY OF THE ORIGINAL STRATEGY

Despite the power of the TSSCDS methodology in dynamically exploring relative possible interactions between catalytic mixture components in transition metal catalyzed systems under study, it was shown that vast majority of TSs are not actual chemical reactions but merely isomerizations of the organic parts and do not reveal any real chemical transformation. Furthermore, calculations with catalyst-substrate pairs showed an unexpected sensitivity to the initial relative orientations of structures provided to the workflow. In line with the iterative nature of TSSCDS in growing networks from minima found at each iteration, there is need to reduce the computational cost to find the multiple minima by running a complete dynamic network cycle at both levels of theory. In order to remove the sensitivity to the initial structures as well as to reduce the computational cost to find the multiple minima, we have conducted “static explorations” using the “association” function of TSSCDS. Exhaustive “*static*” conformational search for transition states between catalyst and substrate species was explored as a pre-processing step for the TSSCDS methodology. Through such a static search, different relative orientations of substrate (being acetophenone in the following section)-catalyst are tested for low-energy association complexes at lower cost compared to the dynamic explorations and sensitivity to the initial orientation is thought to be removed. Such energetically favored pairs can in principle replace the “single” heuristically chosen input in the dynamic batches. Implementation of such association complexes as the starting minima in the iterative reaction network exploration could enhance the extent of exploration while increasing the efficiency.

As to do so, the TSSCDS association module was used with a sample catalyst-substrate pair (Figure 2.4.a). Input data needed for this sampling study include the coordinates of catalyst and fragment (ligand/solvent) structures (respectively A & B), details of rotations (rotate) and number of different association positions to be explored (NAssoc). This information is shown under chemical dynamics simulation (CDS) section in Figure 2.4.a. In addition, four different parameters should be sequentially provided to the static rotations module. The first two values are the

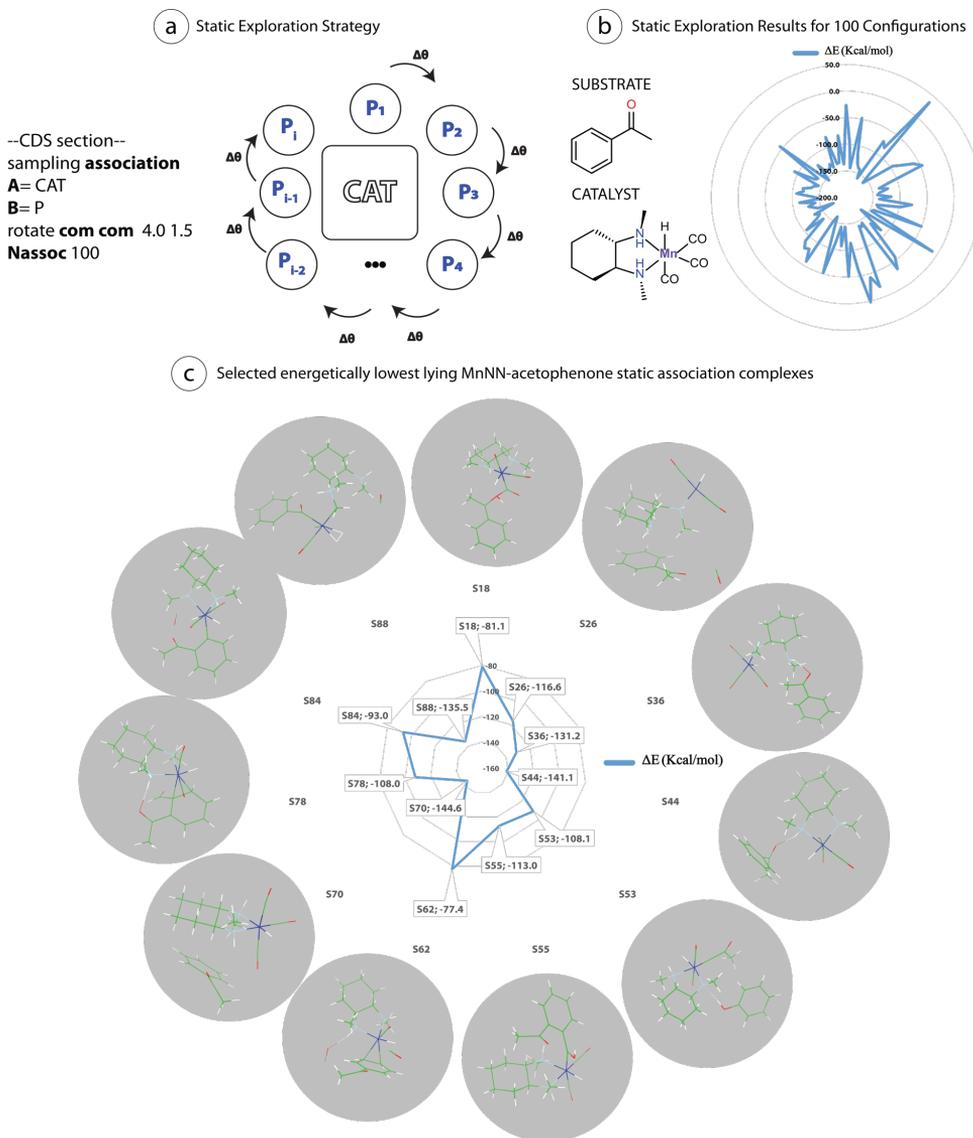


Figure 2.4.: Exhaustive “static” explorations of the MnNN catalyst with acetophenone: a) static exploration strategy b) Static exploration results for 100 configurations for sample substrate-catalyst pair c) Selected energetically low lying MnNN-acetophenone association complexes among the sequence of 100 association species are illustrated.

pivot positions of the rotations (the atomic positions to be fixed while rotations are being made): the center of mass (com) of fragment “CAT” and the center of mass of

fragment "P" in our example (these pivots could be labels of atoms and therefore integers). The last two values are the distance (in Å) between both pivots and the minimum intermolecular distance between any two atoms of both fragments, respectively.  $N_{Assoc}$  defines the number of *association* complexes or relative positions to be explored according to the rotation values. The association complexes produced based on the given input parameters are optimized at DFT level for further analysis. In this case study, static explorations were done for catalyst-acetophenone pair and energetics of the explored structures for  $N_{Assoc} = 100$  configurations are summarized in Figure 2.4.b. Energies calculated for the 100 association complexes vary in relative values in  $[-200 \text{ kcal.mol}^{-1}, 50 \text{ kcal.mol}^{-1}]$  range. Representative structures for the 10 low-lying structures are illustrated in Figure 2.4.c.

Static explorations have resulted in identification of structures with relative energies of  $\sim 150 \text{ kcal.mol}^{-1}$  compared to the reference Catalyst/Substrate combination (Figure 2.4.b) as shown in radar plots of Figure 2.4.c. However, few of the explored structures are anticipated or observed with TSSCDS and ReNeGate explorations on similar system. This incentivized our closer look on the origin of chemical transformations leading to such structures. Since the static rotations are not done based on prior dynamic forces from physical trajectories and rely simply on the pure mathematical distribution of the Cartesian space around the pivot fragment, multiple bonds have been observed to be broken at the same time for multiple instances of the identified structures (S53, S55, S62, S78, S84, S88). This could be attributed to transformations that are not physically allowed and are purely artifacts of close contacts due to the orientations in which molecules are fixed relevant to each other. Although one can choose to focus on optimizing the rotation module parameters to avoid such close contacts while simultaneously not being limited to observing only van der Waals isomers of the same structure, we have chosen to avoid using static methods further in our studies. This choice was made due to conflicts with automated analysis of results: such complexities (and/or artifacts) strictly limit the (automated) stepwise mechanistic explanations and will introduce mechanistic complexities that cannot be explained through chemical reaction networks. This has motivated the implementation of more complex RMSD biased metadynamics simulations discussed in chapter 3 as an alternative to the normal mode sampling and static exploration results provided by TSSCDS methodology.

## 2.4. CONCLUSION

The applicability of the TS Search via Chemical Dynamic Simulations (TSSCDS) method for the exploration of secondary reaction paths in homogeneous Mn(I) reduction catalysis was carried out and the results obtained were used to define and justify the development of our own ReNeGate workflow presented in detail in chapter 3. A brief benchmarking of the TSSCDS and ReNeGate automated reaction analysis tools is presented to highlight the potential of our new methodology as the basis for new approaches for high-throughput reactivity exploration in organometallic chemistry and homogeneous catalysis. TSSCDS procedure identifies many TS structures, that are merely isomerization of rotamers. Further studies with

an auxiliary static exploration strategy have proven the sensitivity of the TSSCDS outcome on to the initial starting configuration. Results from the static exploration where, however, shown to lead to unphysical structures where multiple bonds were broken and formed simultaneously. Stepwise following of the reaction steps is also not possible for structures observed from static explorations.

When focused on finding minima, this approach is excessively expensive. The comparative study with the alternative ReNeGate workflow described in detail in this chapter showed that similar number of unique structures can be identified and reaction network can be constructed using an order of magnitude less time. Furthermore, this study revealed inefficiency of the vibrational-excitation guided MD for chemistry exploration. The development of our ReNeGate workflow therefore was done using RMSD-biased metadynamics simulations. With the scalable ReNeGate reaction network discovery tools, extension to more realistic representation of chemical reactions is envisaged through considering multiple competitive/cooperative interactions among multiple catalyst- substrate-solvent combinations. Such improvements in the scalability of the exploration algorithms has also paved the way for exploration efficiency for studies on databases of catalyst structures further discussed in Chapter 4.

Once equipped with adequate model and method accuracies, it is desired to verify (question) the validity of the classically simplified views on the role of ligand modifications on catalyst efficiencies. Deeper insights from a series of such investigations will ultimately provide the researcher in the field with guidelines for “design” of better catalysts. An inversion in the trend of collaborations between experimentalists and theoreticians would be the optimal outcome of this step, that is, material discovery based on systematic comparisons of reaction networks will be tested in the lab. These proposed research questions are further discussed in the following chapters of this thesis.



## REFERENCES

- [1] A. J. Martin, S. Mitchell, C. Mondelli, S. Jaydev, and J. Perez-Ramirez. “Unifying views on catalyst deactivation”. In: *Nature Catalysis* 5.10 (2022), pp. 854–866. DOI: s41929-022-00842-y.
- [2] G. J. P. Britovsek. “Homogeneous Catalysts. Activity—Stability—Deactivation. By PietW.N.M. vanLeeuwen and JohnC. Chadwick”. In: *Angewandte Chemie-International Edition* 51.7 (2012), pp. 1518–1518. DOI: 10.1002/anie.201108293.
- [3] G. A. Filonenko, R. Van Putten, E. J. Hensen, and E. A. Pidko. “Catalytic (de) hydrogenation promoted by non-precious metals—Co, Fe and Mn: recent advances in an emerging field”. In: *Chemical Society Reviews* 47.4 (2018), pp. 1459–1483. DOI: 10.1039/C7CS00334J.
- [4] J. Pritchard, G. A. Filonenko, R. Van Putten, E. J. Hensen, and E. A. Pidko. “Heterogeneous and homogeneous catalysis for the hydrogenation of carboxylic acid derivatives: history, advances and future directions”. In: *Chemical Society Reviews* 44.11 (2015), pp. 3808–3833. DOI: 10.1039/C5CS00038F.
- [5] J. A. Varela, S. A. Vázquez, and E. Martínez-Núñez. “An automated method to find reaction mechanisms and solve the kinetics in organometallic catalysis”. In: *Chemical Science* 8.5 (2017), pp. 3843–3851. DOI: 10.1039/C7SC00549K.
- [6] R. F. Heck and D. S. Breslow. “The Reaction of Cobalt Hydrotetracarbonyl with Olefins”. In: *Journal of the American Chemical Society* 83.19 (1961), pp. 4023–4027. DOI: 10.1021/ja01480a017.
- [7] R. F. Heck. “Addition reactions of transition metal compounds”. In: *Accounts of Chemical Research* 2.1 (1969), pp. 10–16. DOI: 10.1021/ar50013a002.
- [8] R. van Putten, J. Benschop, V. J. de Munck, M. Weber, C. Müller, G. A. Filonenko, and E. A. Pidko. “Efficient and practical transfer hydrogenation of ketones catalyzed by a simple bidentate Mn-NHC complex”. In: *ChemCatChem* 11.21 (2019), pp. 5232–5235. DOI: 10.1002/cctc.201900882.
- [9] R. Van Putten, G. A. Filonenko, A. Gonzalez De Castro, C. Liu, M. Weber, C. Müller, L. Lefort, and E. Pidko. “Mechanistic Complexity of Asymmetric Transfer Hydrogenation with Simple Mn–Diamine Catalysts”. In: *Organometallics* 38.16 (2019), pp. 3187–3196. DOI: 10.1021/acs.organomet.9b00457.
- [10] A. Hashemi, S. Bougueroua, M.-P. Gageot, and E. A. Pidko. “ReNeGate: A Reaction Network Graph-Theoretical Tool for Automated Mechanistic Studies in Computational Homogeneous Catalysis”. In: *Journal of Chemical Theory and Computation* 18.12 (2022), pp. 7470–7482. DOI: 10.1021/acs.jctc.2c00404.

- 2
- [11] J. J. Stewart. "Optimization of parameters for semiempirical methods V: Modification of NDDO approximations and application to 70 elements". In: *Journal of Molecular Modeling* 13.12 (2007), pp. 1173–1213. DOI: 10.1007/s00894-007-0233-4.
- [12] A. D. Becke. "Density-functional thermochemistry. III. The role of exact exchange". In: *The Journal of Chemical Physics* 98.7 (1993), pp. 5648–5652. DOI: 10.1063/1.464913.
- [13] C. Lee, W. Yang, and R. G. Parr. "Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density". In: *Physical Review B* 37.2 (1988), pp. 785–789. DOI: 10.1103/PhysRevB.37.785.
- [14] J. Stewart. *MOPAC2016, version: 16.175 W. Stewart Computational Chemistry*.
- [15] M. e. Frisch, G. Trucks, H. B. Schlegel, G. Scuseria, M. Robb, J. Cheeseman, G. Scalmani, V. Barone, G. Petersson, H. Nakatsuji, *et al.* *Gaussian 16*. 2016.
- [16] B. Maji and M. K. Barman. "Recent Developments of Manganese Complexes for Catalytic Hydrogenation and Dehydrogenation Reactions". In: *Synthesis* 49.15 (2017), pp. 3377–3393. DOI: 10.1055/s-0036-1590818.
- [17] J. Pritchard, G. A. Filonenko, R. van Putten, E. J. M. Hensen, and E. A. Pidko. "Heterogeneous and homogeneous catalysis for the hydrogenation of carboxylic acid derivatives: history, advances and future directions". In: *Chemical Society Reviews* 44.11 (2015), pp. 3808–3833. DOI: 10.1039/c5cs00038f.
- [18] D. Wang and D. Astruc. "The Golden Age of Transfer Hydrogenation". In: *Chemical Reviews* 115.13 (2015), pp. 6621–6686. DOI: 10.1021/acs.chemrev.5b00203.
- [19] C. Q. Hu, D. Creaser, S. Siahrostami, H. Gronbeck, H. Ojagh, and M. Skoglundh. "Catalytic hydrogenation of C=C and C=O in unsaturated fatty acid methyl esters". In: *Catalysis Science & Technology* 4.8 (2014), pp. 2427–2444. DOI: 10.1039/c4cy00267a.
- [20] J. Magano and J. R. Dunetz. "Large-Scale Carbonyl Reductions in the Pharmaceutical Industry". In: *Organic Process Research & Development* 16.6 (2012), pp. 1156–1184. DOI: 10.1021/op2003826.
- [21] L. C. M. Castro, H. Q. Li, J. B. Sortais, and C. Darcel. "When iron met phosphines: a happy marriage for reduction catalysis". In: *Green Chemistry* 17.4 (2015), pp. 2283–2303. DOI: 10.1039/c4gc01866d.
- [22] J. L. Renaud and S. Gaillard. "Recent Advances in Iron- and Cobalt-Complex-Catalyzed Tandem/Consecutive Processes Involving Hydrogenation". In: *Synthesis* 48.21 (2016), pp. 3659–3683. DOI: 10.1055/s-0035-1562791.
- [23] Y. W. Jing, S. Chakraborty, W. W. Brennessel, and W. D. Jones. "Additive-Free Cobalt-Catalyzed Hydrogenation of Esters to Alcohols". In: *ACS Catalysis* 7.5 (2017), pp. 3735–3740. DOI: 10.1021/acscatal.7b00623.
- [24] T. J. Korstanje, J. I. van der Vlugt, C. J. Elsevier, and B. de Bruin. "Hydrogenation of carboxylic acids with a homogeneous cobalt catalyst". In: *Science* 350.6258 (2015), pp. 298–+. DOI: 10.1126/science.aaa8938.

- [25] G. A. Filonenko, R. van Putten, E. J. M. Hensen, and E. A. Pidko. “Catalytic (de)hydrogenation promoted by non-precious metals - Co, Fe and Mn: recent advances in an emerging field”. In: *Chemical Society Reviews* 47.4 (2018), pp. 1459–1483. DOI: 10.1039/c7cs00334j.
- [26] F. Kallmeier and R. Kempe. “Manganese Complexes for (De)Hydrogenation Catalysis: A Comparison to Cobalt and Iron Catalysts”. In: *Angewandte Chemie-International Edition* 57.1 (2018), pp. 46–60. DOI: 10.1002/anie.201709010.
- [27] J. R. Carney, B. R. Dillon, and S. P. Thomas. “Recent Advances of Manganese Catalysis for Organic Synthesis”. In: *European Journal of Organic Chemistry* 23 (2016), pp. 3912–3929. DOI: 10.1002/ejoc.201600018.
- [28] J. W. Finley and C. D. Davis. “Manganese deficiency and toxicity: are high or low dietary amounts of manganese cause for concern?” In: *Biofactors* 10.1 (1999), pp. 15–24. DOI: 10.1002/biof.5520100102.
- [29] J. R. Khusnutdinova and D. Milstein. “Metal-Ligand Cooperation”. In: *Angewandte Chemie-International Edition* 54.42 (2015), pp. 12236–12273. DOI: 10.1002/anie.201503873.
- [30] T. Zell and D. Milstein. “Hydrogenation and Dehydrogenation Iron Pincer Catalysts Capable of Metal Ligand Cooperation by Aromatization/Deaomatization”. In: *Accounts of Chemical Research* 48.7 (2015), pp. 1979–1994. DOI: 10.1021/acs.accounts.5b00027.
- [31] D. A. Valyaev, G. Lavigne, and N. Lugan. “Manganese organometallic compounds in homogeneous catalysis”. In: *Coordination Chemistry Reviews* 308 (2016), p. 191. DOI: 10.1016/j.ccr.2015.06.015.
- [32] S. Elangovan, M. Garbe, H. Jiao, A. Spannenberg, K. Junge, and M. Beller. “Hydrogenation of Esters to Alcohols Catalyzed by Defined Manganese Pincer Complexes”. In: *Angewandte Chemie-International Edition* 55.49 (2016), p. 15364. DOI: 10.1002/anie.201607233.
- [33] Y. Wang, L. Zhu, Z. Shao, G. Li, Y. Lan, and Q. Liu. “Unmasking the Ligand Effect in Manganese-Catalyzed Hydrogenation: Mechanistic Insight and Catalytic Application”. In: *Journal of the American Chemical Society* 141.43 (2019), pp. 17337–17349. DOI: 10.1021/jacs.9b09038.
- [34] F. Freitag, T. Irrgang, and R. Kempe. “Mechanistic Studies of Hydride Transfer to Imines from a Highly Active and Chemoselective Manganate Catalyst”. In: *Journal of the American Chemical Society* 141.29 (2019), pp. 11677–11685. DOI: 10.1021/jacs.9b05024.
- [35] W. Yang, I. Y. Chernyshov, R. K. A. van Schendel, M. Weber, C. Müller, G. A. Filonenko, and E. A. Pidko. “Robust and efficient hydrogenation of carbonyl compounds catalysed by mixed donor Mn(I) pincer complexes”. In: *Nature Communications* 12.1 (2021), p. 12. DOI: 10.1038/s41467-020-20168-2.

- [36] W. Yang, I. Y. Chernyshov, M. Weber, E. A. Pidko, and G. A. Filonenko. “Switching between Hydrogenation and Olefin Transposition Catalysis via Silencing NH Cooperativity in Mn(I) Pincer Complexes”. In: *ACS Catalysis* 12.17 (2022), pp. 10818–10825. DOI: 10.1021/acscatal.2c02963.
- [37] E. Martínez-Núñez. “An automated method to find transition states using chemical dynamics simulations”. In: *Journal of computational chemistry* 36.4 (2015), pp. 222–234. DOI: 10.1039/C5CP02175H.

# 3

## RENeGATE

*Reaction Network Graph Theoretical tool for automated mechanistic studies in computational homogeneous catalysis*



***E**XPLORATION of the chemical reaction space of chemical transformations in multicomponent mixtures is one of the main challenges in contemporary computational chemistry. In order to remove expert bias from mechanistic studies and to discover new chemistries an automated graph-theoretical methodology is proposed, which puts forward a network formalism of homogeneous catalysis reactions and utilizes a network analysis tool for mechanistic studies. The method can be used for analyzing trajectories with single and multiple catalytic species and can provide unique conformers of catalysts including multinuclear catalyst clusters along with other catalytic mixture components. The presented three-step approach has the integrated ability to handle multicomponent catalytic systems of arbitrary complexity (mixtures of reactants, catalyst precursors, ligands, additives, and solvent). It is not*

---

This chapter is based on Hashemi, A.; Bougueroua, S.; Gageot, M.-P.; Pidko, E. A. *ReNeGate: A Reaction Network Graph-Theoretical Tool for Automated Mechanistic Studies in Computational Homogeneous Catalysis*. *Journal of Chemical Theory and Computation* 2022, 18, 7470–7482.

limited to pre-defined chemical rules, does not require pre-alignment of reaction mixture components consistent with a reaction coordinate and is not agnostic to the chemical nature of transformations. Conformer exploration, Reactive event identification and Reaction network analysis are the main steps taken for identifying the pathways in catalytic systems given the starting pre-catalytic reaction mixture as the input. Such a methodology allows to efficiently explore catalytic systems in realistic conditions for either previously observed or completely unknown reactive events in the context of a network representing different intermediates. Our workflow for the catalytic reaction space exploration exclusively focuses on the identification of the thermodynamically feasible conversion channels, representative of the (secondary) catalyst deactivation or inhibition paths, which are usually most difficult to anticipate based solely on the expert chemical knowledge. Thus, the expert bias is sought to be removed at all steps and the chemical intuition is limited to the choice of the thermodynamic constraint imposed by the applicable experimental conditions in terms of threshold energy values for allowed transformations. The capabilities of the proposed methodology have been tested by exploring reactivity of Mn complexes relevant for catalytic hydrogenation chemistry to verify previously postulated activation mechanisms and unravel unexpected reaction channels relevant to rare deactivation events.

### 3.1. INTRODUCTION

Contemporary computational chemistry has reached a stage at which massive exploration into chemical reaction space with unprecedented resolution with respect to the number of potentially relevant molecular structures is becoming a realistic task. Various algorithmic advances have shown that extensive structural screenings can nowadays be automated and carried out using modern computational chemistry protocols[1–5]. Automated computational strategies for predicting multi-step reaction mechanisms for complex chemical processes such as pyrolysis, combustion or catalytic transformations offer substantial advantages over the conventional strategy largely based on the expert-guided exploration of selected and restricted number of mechanistic alternatives.

Practical catalytic systems are represented by complex mixtures usually containing the catalyst precursor, ligands, solvents and various additives and promoters next to the substrates and the conversion products. The interactions between these components and their interconversions form large and highly interconnected reaction networks that determine the overall behaviour and the performance of the catalytic system. The experimental and computational mechanistic studies aim at identifying the state of the catalytic species and key reaction intermediates, their role in the main catalytic mechanism and the competing reaction channels towards unselective conversion routes or catalyst deactivation[6–13]. Such mechanistic insights are critically important for guiding the design and optimization of new and improved catalytic systems in a rational manner[14–18].

Catalytic reactivity is determined by complex networks of chemical transformations that take place simultaneously or consequently between the different (transient) components of the catalytic mixture. Different stages in a catalytic process, namely catalyst activation, catalytic cycle propagation, catalyst deactivation and different non-selective conversion paths may involve reaction intermediates that are not known *a priori* and will proceed through multiple elementary steps. Even most advanced experimental operando techniques are not able to capture such a high molecular-level complexity. To establish a comprehensive picture of the catalytic process, computational analysis on such systems requires a thorough exploration of the chemical and configuration space to identify the minima on the potential energy surface (PES) and the pathways connecting them.

The characterization and exploration of PES is a tedious and challenging task. A conventional workflow in applied computational catalysis studies approaches this task via manual structural explorations, which rely largely on the expert knowledge and a substantial amount of chemical intuition, limiting thus the study to the expected reactivity domains. The last decade has seen a rapid development of various computational approaches to automate the exploration and discovery of complex chemical reaction networks targeting the reconstruction of a complete atomistic representation of the mechanism of a chemical conversion process[1–5]. Strategies for the accelerated exploration of reaction networks can vary substantially in the computational costs as well as the comprehensiveness and accuracy of the chemical reaction network that they produce[19, 20].

For example, the Global Reaction Route Mapping (GRRM) approach introduced by

Maeda et al,[21] in which starting from a given “reactant” configuration, the PES is explored to discover new transition states and intermediates forming the reaction network. Multimolecular reaction paths can be successfully followed using the artificial force-induced reaction (AFIR) method,[21–23] which directs the transitions from one equilibrium structure to another by applying splitting or merging force to two interacting fragments. This approach was used to automatically construct the catalytic paths for various homogeneous catalytic reactions with transition metal complexes[24–26].

Despite being highly systematic, such curvature-based exploration strategies may be impractical to studying very large and complex catalytic systems. By introducing principles of algorithmic search, the efficiency of the path-finding for the conversion of a given substrate state to a defined product state can be substantially improved[27]. A complementary approach to streamline the exploration of the reaction mechanism is to employ the conceptual knowledge of chemistry. Chemical reactivity can often be well captured by a set of heuristic rules for the transformations that can be applied to graph representations of the molecular system, as successfully demonstrated by Zimmerman and co-workers in their mechanistic studies on organometallic systems[28–31]. Reiher and co-workers introduced a method based on system-independent heuristic rules,[32] which was successfully employed to exploit alternative mechanisms of ammonia production with the Schrock dinitrogen-fixation catalyst. Further developments of the method enabled exploration of transformations involving multiple reactive centers on the molecular fragments and/or interactions between different components of the reactive system[33].

The configuration and reaction space of a molecular system can be directly sampled by solving the nuclear equations of motion in *ab initio* molecular dynamics (AIMD) simulations[34]. However, considering that even the fastest chemical reactions are rare events, adequate scanning of the reaction space of realistic catalyst systems by the direct atomistic AIMD simulations becomes prohibitively expensive when executed using sufficiently accurate electronic structure methods. The frequency of the reaction events can be greatly accelerated by applying bias potentials that push the system away from the free energy minima along a collective variable (CV), which requires the knowledge of the reaction coordinate and therefore limits the application of this method in the exploratory studies[35–37]. Martínez-Núñez and co-workers introduced an automated procedure called TS Search using Chemical Dynamic Simulations (TSSCDS) for the global search of transition states on intermolecular potential energy surfaces based on the PES exploration via the high-energy molecular dynamic simulations[38–41]. To increase the chances for the chemical transformations to occur, the method populates vibrational modes in the system. Similar strategy has been also used to guide the exploration of the configurational space for multinuclear transition metal species in zeolite-based heterogeneous catalysts[42]. Shannon et al combined molecular dynamics and statistical rate theory within a ChemDyMe automated mechanism generation method,[43] in which the search for new reactions is constrained to only the kinetically relevant ones under the specified conditions. The various algorithmic developments in the field have

recently been integrated by the group of Reiher in a Chemoton 2.0 software that will hopefully make the autonomous mechanistic explorations of complex chemical systems accessible to the wide chemistry community[44].

The various automated reaction network analysis tools described above enable the automated transition state search and the construction of detailed mechanistic pictures for practical chemical systems. However, the computational demand for such a detailed PES analysis increase exponentially with model system complexity, which reduces the utility of these methods to exploratory search of secondary transformations (such as non-selective conversion paths, catalyst deactivation, etc) in extended realistic catalytic systems and/or their integration in high-throughput computational catalyst screening workflows. In this work, we propose a graph-based 3-step methodology for exhaustive conformer ensemble exploration and reaction event finding enabling a comprehensive analysis of complex reaction networks in large molecular ensembles at a reduced computational cost. Here we employ the CREST[45, 46] method with the systematic RMSD biases in terms of pulling factors, which drive the system away from the conformers that have already been explored. The conformer ensembles populated through such parallel metadynamics simulations are then interpreted as molecular graphs and analyzed by the proposed graph theoretical tools to find unique chemical structures in terms of bonding patterns. The graph theory and computer-based approaches for the analysis of molecular trajectories have proved its value over the last decade in computational chemistry[47–51]. The main concepts of the molecular graph theory, on which the current work is based are summarized in section A.1 with the common terminology explained in detail in section A.2 of the appendix. A reaction network of such unique chemical species is formed and the network is further analyzed through inspection of nodes and edges present in the network. The power of the introduced strategy is demonstrated through the analysis of the reaction networks generated for representative model Mn-based homogeneous ketone transfer hydrogenation systems.

The chapter is structured as follows. In the next section 3.2, we present the description of the new 3-step reaction exploration methodology. We present the detailed rationale for the conformer exploration approach introduced to simulate extended molecular systems and a new graph-based tool allowing to follow the changes in bonding patterns within reactive trajectories to identify reaction events. Section 3 illustrates the capabilities of the developed methodology on three representative case studies of the catalytic and coordination chemistry of Mn(I) compounds. A conclusion section summarizes the presented methodology and obtained results at the end of the manuscript. The additional details of the methodology and the computational results obtained in the validation studies and the full datasets are provided as the Appendix A. The ReNeGate code is publicly available at: <https://github.com/ahashemiche/ReNeGate>

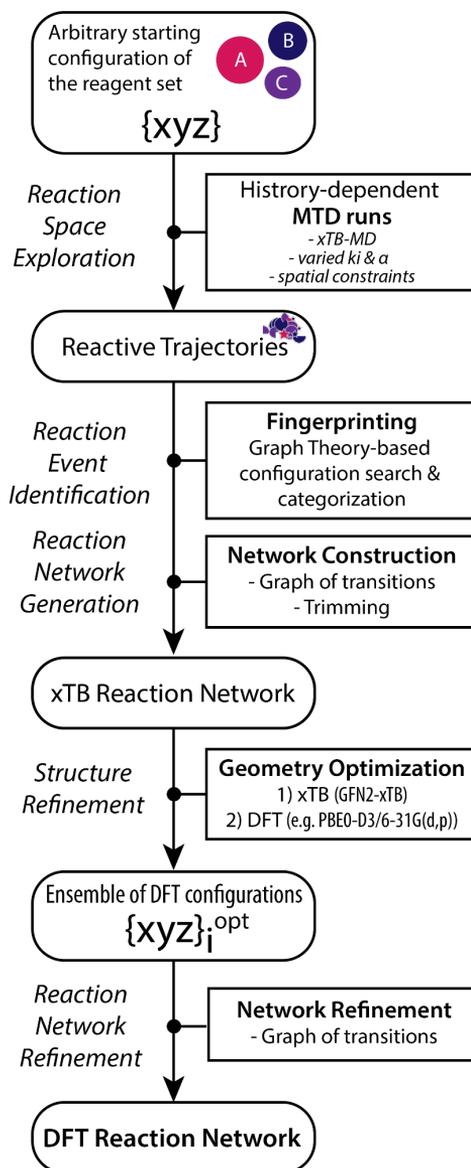


Figure 3.1.: Schematic representation of the ReNeGate workflow involving the sequential reactive space exploration, structure analysis, reaction network generation and refinement steps.

### 3.2. AUTOMATED REACTION EXPLORATION METHOD

A three step methodology denoted as ReNeGate is proposed, which is able to automatically handle catalytic systems of arbitrary complexity (multicomponent

catalytic mixtures of reactants, catalyst precursors, ligands, additives, and solvent) and is not limited to either pre-defined chemical rules or pre-defined reaction coordinates. *Reactive space exploration*, *Reactive event identification* and *Reaction network generation* are the main steps taken for understanding the underlying mechanistic pathways in catalytic mixtures. Such a methodology will then be able to comprehensively explore catalytic systems in realistic conditions for either previously observed or completely unknown reactive events. Human bias is sought to be removed in either of the 3 steps and chemical intuition is limited to the choice of thermodynamic constraints imposed by applicable experimental conditions in terms of threshold values for allowed transformations.

### 3.2.1. REACTIVE SPACE EXPLORATION

Figure 3.1 schematically presents the ReNeGate reaction exploration methodology. The starting point is the exhaustive reaction exploration carried out on a given starting set of reaction components. The identification of unique reactive configurations and reaction states is carried out by analysing the simulated reactive trajectories in the framework of the graph theory. The thus identified reactive states are then refined by geometry optimization at the DFT level appropriate for the specific chemical system explored and final accuracy targeted in the simulations[10]. Initial reaction space exploration are done using the CREST functionality[46] in the GFN-xTB[52] code where semiempirical xTB-MD calculations with root-mean-square-deviation (RMSD) based meta-dynamics simulations (MTD) to ensure that the initial reaction space exploration is exhaustive and thorough[45]. Recent investigations have demonstrated a sufficient accuracy of the xTB for high-throughput screening of transition metal complexes[53] including Mn(I)-based systems discussed herein as the representative model catalysts[54]. Imposing RMSD-based metadynamics allows for a thorough exploration of the compound space. The choice of the collective variables (CVs) in MTD is critical and distinct approaches to this challenging problem in the chemical and biomolecular simulations have been proposed including Diffusion map MD[55], Targeted MD[56], Tabu Search[57, 58] methods. Here we employ the standard root-mean-square deviation (RMSD) in Cartesian space as an unbiased metric as implemented in the CREST functionality in xTB.

Reactive trajectories populated with configurations from the collective MTDs from CREST calculations are then analyzed with our dedicated graph-based tool described below for finding unique chemical structures based on bonding patterns. The implementation of such an automated conformer exploration scheme is sought to automate mechanistic studies on catalytic system of interest and help to reveal *unconventional mechanisms* and *deactivation pathways*, which are usually hard to find using conventional expert knowledge-based strategies to mechanistic studies.

### 3.2.2. REACTION EVENT IDENTIFICATION AND NETWORK CONSTRUCTION

Next, we employed graph theory-based algorithms to analyze the ensembles of structures produced in the configuration exploration step and to categorize them

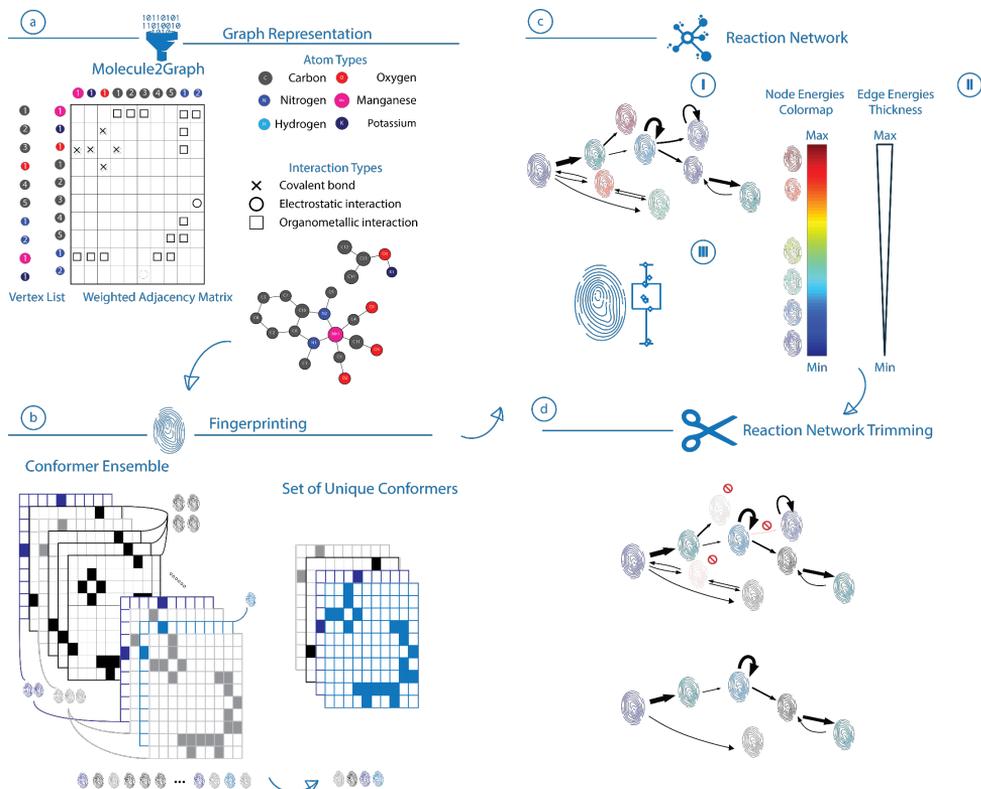


Figure 3.2.: Reaction Event Exploration scheme following the sequence of (a) the graph representation of the reaction ensemble, (b) fingerprinting of the discovered states, (c) generation and (d) trimming of the complete reaction network.

into “*experimentally relevant*” ensembles[59]. The procedure is schematically depicted in Figure 3.2. The procedure starts with the generation of molecular graph representations for the chemical structures of each given conformer in the graph representation module (Figure 3.2a)[59]. Next, the conformer ensembles populated in the exploration step are analysed for fingerprinting and isomorphism check. Further details on the definition of our molecular graphs representation and the isomorphism check are given in Section A.1 of Appendix A. A set of unique conformers is identified within the ensemble based on the molecular graphs formed for each conformer (Figure 3.2b). The combined results of the reaction event exploration and the fingerprinting analysis are then assembled together into the reaction network, in which the specific fingerprints represent non-redundant conformers and the edges represent connections between the conformers in the trajectory (Figure 3.2c).

Color coding and edge thickness are utilized to visualize energy descriptors for

nodes and edges present in the reaction network. The node colors are introduced by the color map (Figure 3.2.c.II) defined based on the lowest (MIN) and highest (MAX) node energies present in the reaction network. The colors for the species are then automatically chosen based on the mapping of the respective node energies. In cases where different isomers are found for a unique conformer, energy for the most stable conformer is used for color coding in the reference graph and variations in energies are visualized as boxplots next to the relevant nodes in energy diagrams. Thickness of the edges present in the reaction network are similarly adjusted by a separate inverse mapping based on highest and lowest transformation energies, where the transformation energy is defined as the energy difference between the nodes connected with the directed edge. The lowest (most probable) transformation is visualized with the thickest line while the highest energy transformation has the thinnest edge (Figure 3.2.c.I). Such an analysis allows to assess the structural flexibility of the specific reactive configurations and its relative stability within the reaction network. The final step of the reaction network assembly is the trimming of the network, in which nodes or edge connections with the energies exceeding a pre-defined energy threshold are removed from the network as schematically shown in Figure 3.2.d. The specific threshold energy value is pre-defined under the assumption that the states above it have only a minor (if any contribution) to the overall reactivity.

### 3.2.3. FINGERPRINTING AND REACTION NETWORK CONSTRUCTION

The graph isomorphism tools allow representing each conformer from the screening with a fingerprint molecular graph and compare it with other species along the simulation trajectory. The fingerprinting of the species within a reactive simulation trajectory proceeds through a sequence of initialization and conformation dynamic analysis steps. During the *initialization* step the first snapshot  $I_1$  of the MD trajectory is read and the first graph  $G_1$  is defined by identification of the different bonds. Next, the *configurational dynamics analysis* steps are carried out as follows:

1. Read a new snapshot  $I_i$  and define the associated graph  $G_i$
2. Test if  $G_i$  is isomorphic to  $G_{i-1}$
3. Else, assign to configurations already identified (update database)
4. Return to step (a) in order to read the subsequent snapshot.

### 3.2.4. REACTION NETWORK TRIMMING

Once a complete reaction network has been formed through the exploration and reactive event identification steps, nodes and edges present in the acquired network (respectively representing chemical species and dynamic connections) are inspected for being accessible within the energy thresholds defined for the system by the user based on the thermodynamic considerations for a given experiment and its representative condition. For all nodes in the graph, the species (nodes) with

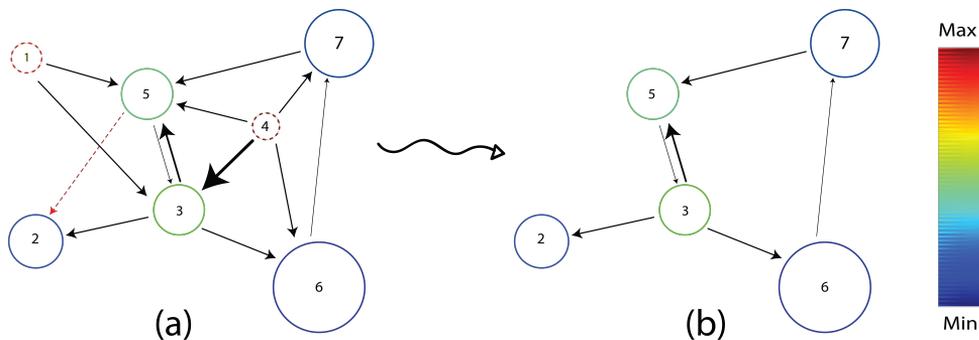


Figure 3.3.: A schematic illustration of an arbitrary reaction network trimming procedure: the original network (a) and the resulting trimmed network (b). Nodes are colored based on the mapping of energies as discussed in section 3.2 Edge widths are also adjusted based on edge mapping function based on energy differences between the nodes. Arbitrary nodes with energies in both extremes are chosen for clarity. The energies of nodes 1 and 4 (dashed circles) exceed the energy threshold defined for species in this network and are removed by the trimming procedure. The edge connecting nodes 2 and 5 (dashed red line) also exceeds the edge threshold value and is hence removed.

energies higher than a predefined “*node threshold value*” together with all respective edges going to and from these nodes are therefore discarded. For all the edges still present in the network, if the edge weight (representing the relative energy difference between the connected species) is higher than the predefined “*edge threshold value*”, then edges will be removed in the network. In short, trimming of the obtained reaction networks for the energetically possible pathways is done based on energies of different species and differences in energies for reactive events.

Figure 3.3. schematically illustrates the trimming of an arbitrary chemical reaction network. To facilitate analysis, the size of the nodes in the reaction networks is inversely adjusted by a mapping based on the lowest and highest energies for structures present in the networks. Similarly, thickness of the edges connecting nodes is adjusted by a separate mapping based on the highest and lowest value for energy differences between the connected nodes. Based on the energies calculated for the nodes present in the arbitrary network, nodes 1 and 4 (shown in dashed circles) have exceeded the predefined *node threshold value* and the edge connecting nodes 5 and 2 (shown in dashed arrow) has exceeded the *edge threshold value*. Therefore these nodes and edges are removed from the original network.

### 3.2.5. FRAGMENT ANALYSIS

Based on the developments in the reaction network exploration and trimming sections, chemical reaction networks can be built and analysed for detecting the stable (deactivated) species present on the PES. While such analysis on trajectories

including *single* instances of the catalyst molecule will result in non-redundant unique catalyst fragments, analysis of reaction networks with multinuclear catalytic ensembles is not as trivial. As an extension to the functions described earlier to be able to handle catalytic systems with more than one catalyst molecule, the fragment analysis tool has been developed to be able to provide a list of unique fragments in the cases (1) when changes in the bonding patterns happen in the non-catalyst part of the snapshot or (2) similar catalyst fragments are observed in different unique configurations. In order to be able to identify unique catalytic events in trajectories populated for systems of *arbitrary complexity* and to remove expert bias in setting the simulation scenario, the model composition should be considered as close as possible to the experimental conditions. Consideration of catalytic systems with more than one metal center introduces new levels of complexity since the algorithms explained in the previous sections should be modified to distinguish different metal centers and enable further comparisons inside a given snapshot (in addition to comparisons within different snapshots).

From a technical point of view, we use Breadth-First Search (BFS) algorithm to identify the fragments[60]. The BFS aims to travers trees in the graph. It starts at the tree root (on arbitrary vertex in the graph) and explores all the neighbour vertices at the present depth prior to moving on to the vertices at the next depth level. Each tree will represent one connected component and each connected component will represent one fragment. Figure 3.4.a,b shows examples of two graphs containing a single and two fragments, respectively. Once a given trajectory of reactive events is

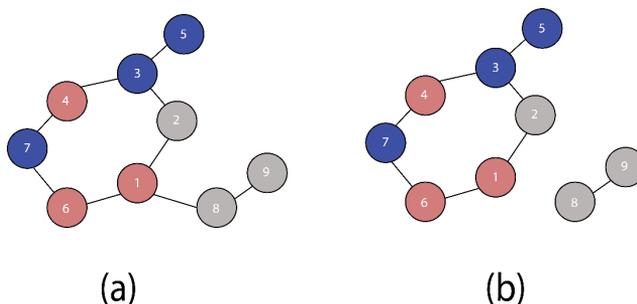


Figure 3.4.: Arbitrary molecular graphs with one (a) and two (b) connected components based on BFS algorithm.

analysed to identify unique catalyst fragments, a list of connected components is given, which represents the unique set of fragments including the transition metal (pivot) atom in the analysed trajectories. Further discussions on the application of fragment analysis tool is given based on a case study presented in Section 3.3 analysing the possibility of the formation of multinuclear Mn ensembles upon the transformation of two  $\text{Mn}(\text{CO})_5\text{Br}$  precursors in the presence of an alkoxide base.

### 3.3. VALIDATION

To assess the performance of the proposed methodology in validating previously observed and identifying unobserved chemical transformations, we applied it to selected representative multicomponent Mn(I)-based (de)hydrogenation catalytic systems. Catalytic (de)hydrogenation reactions promoted by non-precious 3d transition metal complexes represent more sustainable and environmentally benign alternatives to the established stoichiometric and noble metal-catalyzed processes[61–66]. Such reactions are of critical importance since they enable efficient transformations of amines, alcohols and their oxidized counterparts bearing imine, carbonyl or carboxylate functionalities. Commonly, the successful catalytic reactions require the *in situ* activation of the transition metal complex precatalyst by the reaction with a promotor. Common procedures of catalyst activation in the Mn(I)-based carbonyl reduction systems involve the reaction with an alkoxide base promotor or a hydride donor in the presence of a hydrogen-donating solvent or gaseous H<sub>2</sub>[66–71]. The selective transformation of the pre-catalyst complex at this stage is critical for the stability and the overall behaviour of the catalytic system[68]. The formation of undesirable intermediates during the catalyst pre-activation may initiate reaction channels giving rise to non-selective conversions and catalyst deactivation. The identification of such minor reaction paths represents a particular challenge both for experimental and computational catalysis studies.

Herein, we specifically aim at utilizing the ReNeGate methodology to get an insight into such unexpected reaction paths for representative Mn(I) pre-catalysts. Two primary case studies are selected, namely, the alkoxide-base activation of (I) manganese pentacarbonyl bromide (Mn(CO)<sub>5</sub>Br) catalyst precursor, simulating a widely used protocol for homogeneous catalyst screening with *in situ* catalyst generation[62, 72, 73] and (II) *cis*-Mn(N,N'-dimethyl-1,2-cyclohexanediamine)(CO)<sub>3</sub>Br (Mn-N,N) molecularly-defined pre-catalyst[67, 74]. In addition, to demonstrate the potential of the automated fragment analysis, a more complex model capable of capturing interactions between multiple pre-catalyst species Mn(CO)<sub>5</sub>Br in the presence of the alkoxide activator and BEt<sub>3</sub> stabilizer towards the formation of multinuclear ensembles is considered with the case study III. For these systems, the reaction networks were generated through the conformer exploration, reactive event identification and trimming steps as implemented in ReNeGate. The optimized structures and the energetics of the intermediates within the produced reaction networks were obtained at the B3LYP-D3/6-31g(d,p) level of theory with empirical GD3BJ-dispersion correction and implicit SMD model[75] with the standard parameters for THF as solvent using Gaussian 16.C01 program[76].

#### 3.3.1. CASE STUDY I: Mn(CO)<sub>5</sub>Br PRE-CATALYST ACTIVATION

For the first case study, we considered the transformations of Mn(CO)<sub>5</sub>Br complex in an alkoxide base solutions simulating a common catalyst activation procedure (Figure 3.5a). A minimal model containing Mn(CO)<sub>5</sub>Br and KOiPr species was considered here. Parallel metadynamics simulations were carried out using the CREST functionality in the GFN2-xTB method[77] where the pushing and pulling strengths ( $k$  and  $\alpha$ ) were systematically varied over the parallel simulations. The

RMSD difference between structures observed in every trajectory were used to drive simulation away from observing similar structures during the trajectory. Further analysis was done on the basis of ensembles ( $\sim 350$  structures) populated based on the metadynamics simulations. Our procedure based on the fingerprinting

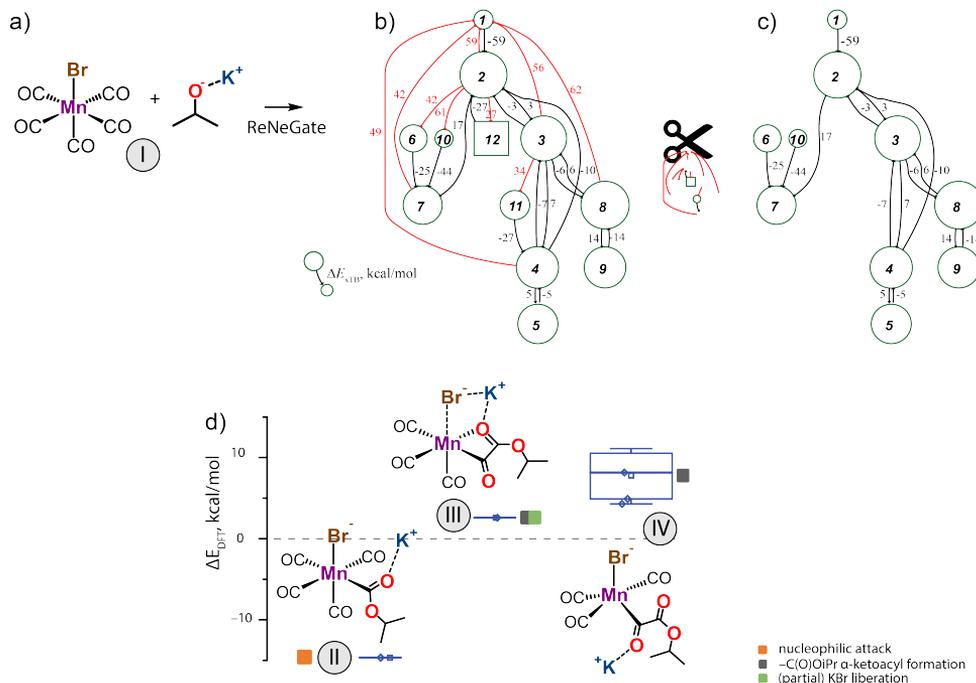


Figure 3.5.: The reaction of (a)  $\text{Mn}(\text{CO})_5\text{Br}$  and  $\text{KOiPr}$  resulting in a network of chemical transformations revealed by the ReNeGate method prior (b) and after (c) the trimming procedure. Panel (d) presents a reaction energy diagram summarizing the distinct product state configurations identified in the network.

algorithm yielded a reaction network of unique chemical structures presented in Figure 3.5b. For the trimming procedure the edges exceeding the threshold value of  $+25 \text{ kcal.mol}^{-1}$  (marked with red in Figure 3.5b) were removed to produce the trimmed reaction network shown in Figure 3.5c. The procedure also eliminated from the final network the *inaccessible nodes after edge trimming* (nodes 4 and 12 in Figure 3.5b) as well as respective *connections to prohibited nodes* (edges going out from nodes 4 and 12, Figure 3.5b). Subsequent fingerprinting of the reaction network identified 9 distinct species. The structures fingerprinted with similar covalent bonds have been grouped into ensembles of structures and further analyzed for differences in energies and non-covalent interactions. Energy values for the species found for the  $\text{Mn}(\text{CO})_5\text{Br}$  transformations network are summarized in Table A.1 of Appendix A.

Stoichiometric reaction with a strong alkoxide base is commonly employed for

the activation of halogen-containing 3d transition metal pre-catalyst in combination with an acid-base cooperative ligand to a reactive catalytic state accompanied by the liberation of KBr and ligand deprotonation[62]. Our automated procedure identified highly favorable alternative routes for the reaction of KOiPr base with the  $\text{Mn}(\text{CO})_5\text{Br}$  precursor resulting in molecular species more stable by up to  $15 \text{ kcal.mol}^{-1}$  compared to the non-activated state (I) representing separate non-interacting alkoxide base and Mn(I) precursor. The main reaction products and their relative stabilities are presented in Figure 3.5.d. Some unique configurations showed a substantial structural flexibility resulting in a range of conformers assigned to a single species and characterized by a range of relative stabilities (e.g. species IV). In all routes, the alkoxide nucleophile reacts with the Mn(I)-bound carbonyl ligand. The direct nucleophilic attack results in an Mn-acyl complex (II). This new reactivity insight has been verified experimentally and inspired the development of new Mn-mediated C-C coupling chemistry recently reported by our group[78].

In line with earlier experimental studies, the reaction event identification tool has revealed from the reactive trajectories that further migratory insertion of the CO ligand with the  $-\text{C}(\text{O})\text{OiPr}$   $\alpha$ -ketoacyl species is thermodynamically strongly unfavorable[79]. The resulting C-Mn  $\alpha$ -ketoacyl conformers (III) and (IV) are ca.  $5 \text{ kcal.mol}^{-1}$  above the energy of the separate  $\text{KO}^t\text{Bu}$  and  $\text{Mn}(\text{CO})_5\text{Br}$ . Due to the minimal size of the model and the lack of the explicit solvation, favorable paths towards the KBr liberation were not identified.

These calculations suggested that nucleophiles (e.g., hydrides and alkoxides) could react with a Mn(I)-bound carbonyl ligand, thereby resulting in the formation of Mn formyl or acyl complexes. These findings are in line with prior experimental observations[79–86] and strongly imply that such reaction paths need to be accounted for when constructing mechanistic hypotheses to rationalize catalytic results based on the in situ catalyst generation protocol employing the activation of Mn(I) carbonyl precursors in the presence of a strong nucleophile alkoxide base.

### 3.3.2. CASE STUDY II: BASE-ACTIVATION OF AN $\text{Mn}(\text{BR})(\text{CO})_3\text{-NN}$ TRANSFER HYDROGENATION PRE-CATALYST

To further evaluate the capabilities of the proposed methodology in exploring the reaction energy landscape for known and unexplored chemistries without additional input from experts, a more complex catalyst system bearing a bidentate ligand has been considered. Specifically, here we considered the activation of a  $\text{N,N}'$ -Dimethyl-1,2-cyclohexanediamino manganese tricarbonyl bromide ( $\text{MnN,N}$ ) molecularly defined Mn(I) catalyst precursor with potassium isopropoxide ( $\text{KO}^i\text{Pr}$ ) two isopropanol and one acetophenone molecules (Figure 3.6a)[67]. Conformer explorations followed by graph theoretical processing of the metadynamics based trajectories trimmed down the  $\sim 10,000$  structures observed in the exploration step at GFN2-xTB level of theory and further optimization at the DFT level reduced the set of structures to 37 species. These species are the nodes in the reaction network in Figure 3.6b. They are unique in terms of bonding patterns according to the thresholds defined for covalent and organometallic interactions (see Section A.1 in Supporting Information). The structures were further grouped in terms of the

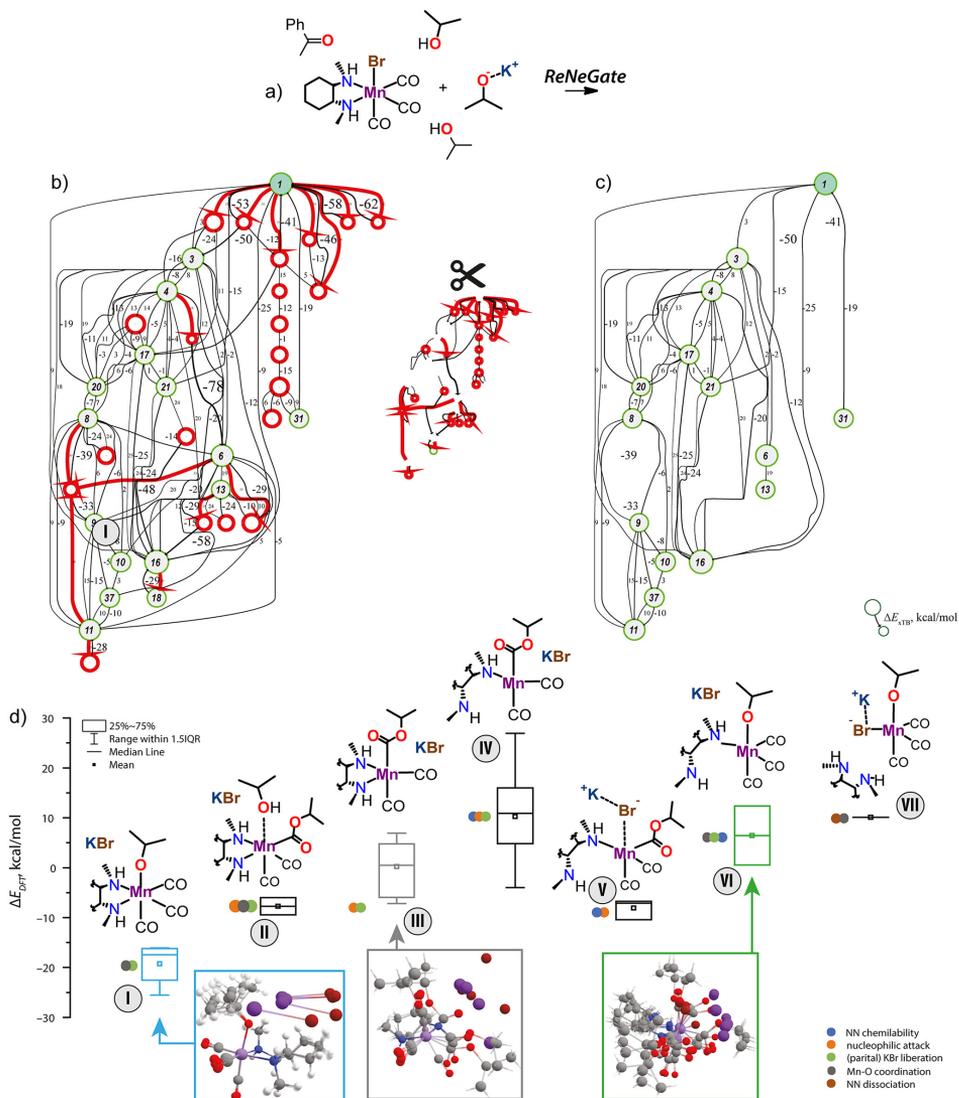
similar covalent bonds and the results are summarized in Figure 3.6.d. The distinct chemical states contain a range of isomeric structures with varied stabilities due to the differences in the relative orientations of the components in the model, while showing similar covalent bonding patterns and the nature of observed covalent and organometallic chemical changes. The associated energy variations are illustrated as boxplots. The energies of all DFT-optimized species are summarized in Table A.2 of Appendix A.

The results reveal that the most favourable path for the activation of the Mn-NN pre-catalyst with an alkoxide base is the ligand exchange reaction resulting in the rapid elimination of KBr and the formation of an Mn-alkoxide complex (**I**) in line with prior mechanistic proposals[67, 74, 87]. Such a transformation stabilizes the system by up to 25 kcal.mol<sup>-1</sup> compared to the non-interacting components (**I**, Figure 3.6d). The stability of this state featuring weakly bound KBr and the Mn-alkoxide complex depends on the relative orientation of the molecular fragment. The range of relative energies populated by the different isomers of state **I** (see blue inset in Figure 3.6d) are shown in Figure 3.6d with the box-plot. A similar representation is employed for other molecular states discovered computationally.

ReNeGate also identifies a quite unexpected thermodynamically favorable path for the KBr elimination, which is accompanied by a nucleophilic attack of the alkoxide anion by the carbonyl ligand (**II**). Simultaneously, the resulting open coordination site is taken up by an *i*PrOH solvent molecule explicitly included in the model. This path is less thermodynamically favourable by 10 kcal.mol<sup>-1</sup> than the direct ligand exchange reaction. The formation of a 5-coordinated Mn-acyl intermediate (**III**) is slightly less favourable. The reaction network analysis also identifies reaction channels resulting in a (partial) decoordination of the NN ligand from the metal center (**IV-VIII**, Figure 3.6). Whereas the ligand dissociation resulting in states **VI** and **V** is accompanied by the nucleophilic attack of <sup>-</sup>O*i*Pr by the Mn-CO moiety, the alternative paths to states **VI** and **VII** result in more conventional under-coordinated Mn-alkoxide and Mn-alcohol adducts. In **VII**, a complete ligand dissociation is observed, whereas in the other three families of intermediates only one metal-nitrogen coordination was broken. Importantly, only for the case of the Mn-acyl family intermediates (**IV**, **V**) the energy with respect to the free base and Mn pre-catalyst states is negative. This suggests that the nucleophilic attack by the carbonyl ligand facilitates the ligand decoordination, which may initiate the alkoxide-induced catalyst deactivation observed for Mn(I)-based systems[66, 68, 88].

### 3.3.3. CASE STUDY III: THE FORMATION OF MULTINUCLEAR ENSEMBLES UPON THE BASE-ACTIVATION OF Mn(CO)<sub>5</sub>Br

To additionally demonstrate the utility of the fragment analysis tool, we have expanded the case study of the base-activation of Mn pentacarbonyl bromide to a hypothetical situation including a more complex reaction mixture including two Mn(CO)<sub>5</sub>Br precursor, KO<sup>*i*</sup>Pr base and BEt<sub>3</sub> stabilizer molecules within the conformer exploration step. The proposed model was built to study the effect of catalyst-catalyst interactions in search for a comprehensive exploration of the PES in realistic reaction conditions and their contribution to stabilization of the reaction



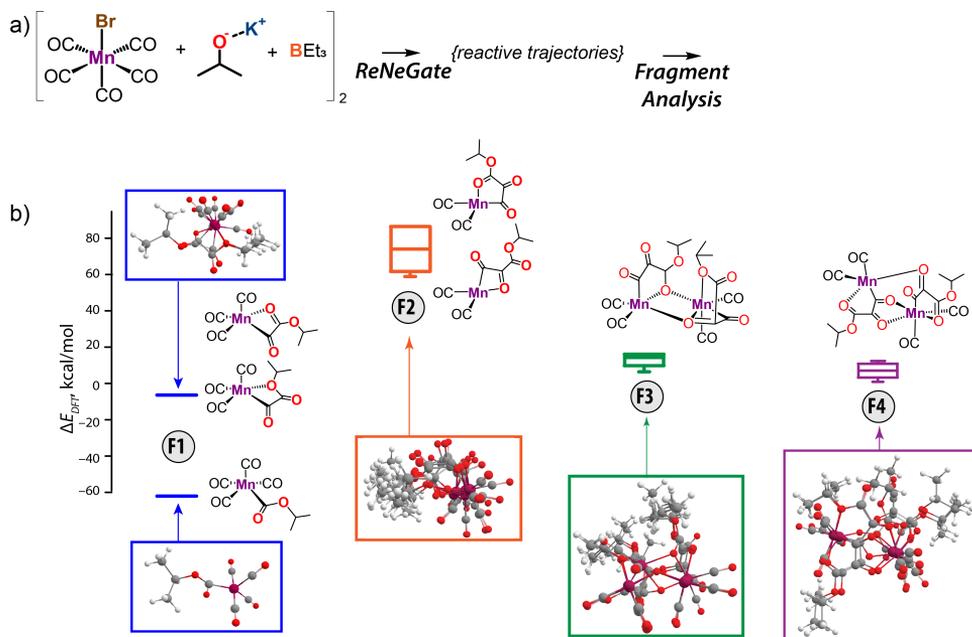


Figure 3.7.: The unique mono- and binuclear Mn complexes identified by the fragment analysis tool in the ReNeGate workflow applied to a system containing two  $\text{Mn}(\text{CO})_5\text{Br}$  and  $\text{KO}^i\text{Pr}$  species.

intermediates. We have chosen to expand the model system discussed in case study II by introducing an additional catalyst precursor molecule along with a potential  $\text{BEt}_3$  stabilizer representing a complex experimental reaction environment.<sup>88</sup> The resulting reactive trajectory was analysed for finding unique catalytic fragments using the fragment analysis tool within the ReNeGate workflow described herein, with the results summarized in Figure 3.7. In addition to mononuclear intermediates similar to those observed in case study I, we identified also multinuclear complexes, albeit much less stable than the mononuclear Mn-acyl species.

As discussed in section 3.3, detection of the fragments and consideration of connections between species is adjusted by the types of interactions (covalent/organometallic/ionic/...) considered for the graph analysis. This means that atoms are considered to be connected if only they have the specified type of interaction. Similar to the previous case studies, input structure was provided to the ReNeGate workflow, reaction networks were obtained and trimmed. The species present in the trimmed network were subject to the fragment analysis. Here, covalent and organometallic interactions have been chosen to distinguish different species when following the bonding patterns. Energy values for the species found for fragment analysis are summarized in Table A.3 of the Appendix A. The most stable configurations featured the mononuclear products of the alkoxide attack by the Mn-CO to form an acyl complex (F1, lower structure), while further migratory

insertion to form C-Mn  $\alpha$ -ketoacyl (F1, upper structure and F2 in Figure 3.7) are thermodynamically unfavorable in line with the chemistry revealed for the trivial model in case study I. In addition, di-Mn binuclear fragments have been identified representing the products of the dimerization of the very unstable  $\alpha$ -ketoacyl adducts. Such dimers were found to be much less stable than their mononuclear counterparts suggesting that clustering and aggregation of Mn centers requires redox processes not considered within the current models. The binuclear fragments F3 and F4 featured the products of dimerization C-Mn  $\alpha$ -ketoacyl adducts (F2). Although the formation of bridging ligands with the various O-atoms of the acyl moiety allowed substantially stabilizing the F2 adducts, the resulting binuclear species were still much less stable than the monodentate Mn-acyl complex. Although in the current case, the increased complexity of the model did not allow identifying new stable configurations, it clearly demonstrates the power of the automated fragment analysis tool. Such straightforward detection of all different catalytic species will become critical when dealing with complex trajectories, where multiple catalytic centers could interact to cooperatively stabilize substrates or such interactions will lead to deactivation of the active center. Changing the type of interactions for analysis (to dynamic hydrogen bonds or ionic interactions) in cases where identification of clusters of hydrogen bonded structures is of importance will be challenging and can be directly done with the help of the fragment analysis tool.

### 3.4. CONCLUSION AND OUTLOOK

We described a graph-based reaction network analysis tool for automation of explorative mechanistic studies in homogeneous catalysis. The *Conformer exploration*, *Reactive event identification* and *Reaction network analysis* are the main steps taken here for understanding the underlying mechanistic pathways in catalytic systems given the reaction mixture as the input. The configurational exploration of the catalytic system is carried out using metadynamic simulations, which results are interpreted and analysed in the framework of the graph theory to identify reactive events and key intermediates that form a reaction network. Such an initial extensive reaction network is trimmed down to *reaction-aware* networks through inspection for consistency within energetic thresholds defined for species and transformations. The resulting trimmed networks can be directly used to provide insights into experimental observations or guide the design of further experiments or in-depth computational analysis. Expert bias is sought to be removed in either of the steps and chemical intuition is limited to the choice of thermodynamic constraints imposed by the applicable experimental conditions.

The capabilities of the proposed methodology have been validated for the alkoxide base activation of manganese pentacarbonyl bromide ( $\text{Mn}(\text{CO})_5\text{Br}$ ) and  $N,N'$ -Dimethyl-1,2-cyclohexanediamino manganese tricarbonyl bromide ( $\text{MnBr}(\text{CO})_3\text{NN}$ ) organometallic complexes commonly employed as pre-catalysts for (de)hydrogenation conversions. The presented automated reaction network analysis successfully reveals the experimentally observed major reaction channels and also helps identifying the more challenging minor reaction paths, that can be initiated by the catalyst

activation procedure and open paths to long-term catalyst deactivation. Specifically, in the case of  $\text{MnBr}(\text{CO})_3\text{NN}$  catalyst precursor, the reaction with an alkoxide base, in addition to the desirable ligand exchange producing the catalytic Mn-OR intermediate, gives rise to a number of less favorable reaction channels that can be regarded as the onset of the catalyst decomposition initiated by the nucleophilic attack of the alkoxide anion by the Mn-carbonyl moiety.

There exist several open questions and challenges for further research and expansion of the presented methodology. The entropic effects could be reconstructed based on the conformation energies stored as node attribute for each node present in the reaction network. Threshold values for trimming the reaction networks are decided based on system specific “reasonable” values by the users. Although this system specificity can be mediated by correlating the trimming values with temperature at which the experiments are usually done for the catalyst not to decompose, it stays as an open question to automatically determine the trimming values based on the chemical nature of the catalytic system. Opportunities exist for using network operations, based on the reaction network formalism discussed in the manuscript to describe chemical reactions. This can include but is not limited to: pathway finding operations (starting from/including/leading to specific species present in the network), finding (weighted) shortest paths to identify mechanisms, finding critical steps (important intermediates) in the network (nodes with large degrees). Such operations and further analysis on networks using machine learning algorithms on the attributes of nodes and edges is also possible for larger reaction networks and is subject of our on-going studies on automated generation of extended databases of catalytic ensembles.

### **Supplementary Information**

Supplementary theoretical details on molecular graph theory for reaction identification; molecular graph theory terminology; supplementary DFT results are provided in Appendix A at the end of the thesis.

## REFERENCES

- [1] A. L. Dewyer, A. J. Arguelles, and P. M. Zimmerman. “Methods for exploring reaction space in molecular systems”. In: *WIREs Computational Molecular Science* 8.2 (2018), e1354. DOI: ARTNe135410.1002/wcms.1354.
- [2] S. Vernuccio and L. J. Broadbelt. “Discerning complex reaction networks using automated generators”. In: *AIChE Journal* 65.8 (2019), e16663. DOI: 10.1002/aic.16663.
- [3] J. P. Unsleber and M. Reiher. “The Exploration of Chemical Reaction Networks”. In: *Annual Review of Physical Chemistry* 71.1 (2020), pp. 121–142. DOI: 10.1146/annurev-physchem-071119-040123.
- [4] A. Puliyaanda, K. Srinivasan, K. Sivaramakrishnan, and V. Prasad. “A review of automated and data-driven approaches for pathway determination and reaction monitoring in complex chemical systems”. In: *Digital Chemical Engineering* 2 (2022), p. 100009. DOI: 10.1016/j.dche.2021.100009.
- [5] M. Steiner and M. Reiher. “Autonomous Reaction Network Exploration in Homogeneous and Heterogeneous Catalysis”. In: *Topics in Catalysis* 65.1 (2022), pp. 6–39. DOI: 10.1007/s11244-021-01543-9.
- [6] S. Santoro, M. Kalek, G. Huang, and F. Himo. “Elucidation of Mechanisms and Selectivities of Metal-Catalyzed Reactions using Quantum Chemical Methodology”. In: *Accounts of Chemical Research* 49.5 (2016), pp. 1006–1018. DOI: 10.1021/acs.accounts.6b00050.
- [7] P. A. Dub and J. C. Gordon. “The role of the metal-bound N–H functionality in Noyori-type molecular catalysts”. In: *Nature Reviews Chemistry* 2.12 (2018), pp. 396–408. DOI: 10.1038/s41570-018-0049-z.
- [8] L. Grajciar, C. J. Heard, A. A. Bondarenko, M. V. Polynski, J. Meeprasert, E. A. Pidko, and P. Nachtigall. “Towards operando computational modeling in heterogeneous catalysis”. In: *Chemical Society Reviews* 47.22 (2018), pp. 8307–8348. DOI: 10.1039/c8cs00398j.
- [9] A. S. Kashin and V. P. Ananikov. “Monitoring chemical reactions in liquid media using electron microscopy”. In: *Nature Reviews Chemistry* 3.11 (2019), pp. 624–637. DOI: 10.1038/s41570-019-0133-z.
- [10] K. D. Vogiatzis, M. V. Polynski, J. K. Kirkland, J. Townsend, A. Hashemi, C. Liu, and E. A. Pidko. “Computational Approach to Molecular Catalysis by 3d Transition Metals: Challenges and Opportunities”. In: *Chemical Reviews* 119.4 (2019), pp. 2453–2523. DOI: 10.1021/acs.chemrev.8b00361.

- [11] A. J. Martín, S. Mitchell, C. Mondelli, S. Jaydev, and J. Pérez-Ramírez. “Unifying views on catalyst deactivation”. In: *Nature Catalysis* (2022). DOI: 10.1038/s41929-022-00842-y.
- [12] L. Artús Suárez, D. Balcells, and A. Nova. “Computational Studies on the Mechanisms for Deaminative Amide Hydrogenation by Homogeneous Bifunctional Catalysts”. In: *Topics in Catalysis* 65.1 (2022), pp. 82–95. DOI: 10.1007/s11244-021-01542-w.
- [13] J. Lan, X. Li, Y. Yang, X. Zhang, and L. W. Chung. “New Insights and Predictions into Complex Homogeneous Reactions Enabled by Computational Chemistry in Synergy with Experiments: Isotopes and Mechanisms”. In: *Accounts of Chemical Research* 55.8 (2022), pp. 1109–1123. DOI: 10.1021/acs.accounts.1c00774.
- [14] Z. Lin. “Interplay between Theory and Experiment: Computational Organometallic and Transition Metal Chemistry”. In: *Accounts of Chemical Research* 43.5 (2010), pp. 602–611. DOI: 10.1021/ar9002027.
- [15] J. Jover and N. Fey. “The Computational Road to Better Catalysts”. In: *Chemistry – An Asian Journal* 9.7 (2014), pp. 1714–1723. DOI: 10.1002/asia.201301696.
- [16] S. Ahn, M. Hong, M. Sundararajan, D. H. Ess, and M.-H. Baik. “Design and Optimization of Catalysts Based on Mechanistic Insights Derived from Quantum Chemical Reaction Modeling”. In: *Chemical Reviews* 119.11 (2019), pp. 6509–6560. DOI: 10.1021/acs.chemrev.9b00073.
- [17] I. Funes-Ardoiz and F. Schoenebeck. “Established and Emerging Computational Tools to Study Homogeneous Catalysis—From Quantum Mechanics to Machine Learning”. In: *Chem* 6.8 (2020), pp. 1904–1913. DOI: 10.1016/j.chempr.2020.07.008.
- [18] N. Fey and J. M. Lynam. “Computational mechanistic study in organometallic catalysis: Why prediction is still a challenge”. In: *WIREs Computational Molecular Science* 12.4 (2022), e1590. DOI: 10.1002/wcms.1590.
- [19] S. Maeda and Y. Harabuchi. “On Benchmarking of Automated Methods for Performing Exhaustive Reaction Path Search”. In: *Journal of Chemical Theory and Computation* 15.4 (2019), pp. 2111–2115. DOI: 10.1021/acs.jctc.8b01182.
- [20] C. A. Grambow, A. Jamal, Y.-P. Li, W. H. Green, J. Zádor, and Y. V. Suleimanov. “Unimolecular Reaction Pathways of a  $\gamma$ -Keto hydroperoxide from Combined Application of Automated Reaction Discovery Methods”. In: *Journal of the American Chemical Society* 140.3 (2018), pp. 1035–1048. DOI: 10.1021/jacs.7b11009.
- [21] S. Maeda, K. Ohno, and K. Morokuma. “Systematic exploration of the mechanism of chemical reactions: the global reaction route mapping (GRRM) strategy using the ADDF and AFIR methods”. In: *Physical Chemistry Chemical Physics* 15.11 (2013), pp. 3683–3701. DOI: 10.1039/C3CP44063J.

- [22] S. Maeda, Y. Harabuchi, M. Takagi, T. Taketsugu, and K. Morokuma. “Artificial Force Induced Reaction (AFIR) Method for Exploring Quantum Chemical Potential Energy Surfaces”. In: *The Chemical Record* 16.5 (2016), pp. 2232–2248. DOI: 10.1002/tcr.201600043.
- [23] S. Maeda, Y. Harabuchi, M. Takagi, K. Saita, K. Suzuki, T. Ichino, Y. Sumiya, K. Sugiyama, and Y. Ono. “Implementation and performance of the artificial force induced reaction method in the GRRM17 program”. In: *Journal of Computational Chemistry* 39.4 (2018), pp. 233–251. DOI: 10.1002/jcc.25106.
- [24] T. Shibata, N. Shiozawa, S. Nishibe, H. Takano, and S. Maeda. “Pt(ii)-Chiral diene-catalyzed enantioselective formal [4 + 2] cycloaddition initiated by C–C bond cleavage and elucidation of a Pt(ii)/(iv) cycle by DFT calculations”. In: *Organic Chemistry Frontiers* 8.24 (2021), pp. 6985–6991. DOI: 10.1039/D1Q001467F.
- [25] F. Yu, Z. Zhou, J. Song, and Y. Zhao. “DFT and AFIR study on the copper(i)-catalyzed mechanism of 5-enamine-trisubstituted-1,2,3-triazole synthesis via C–N cross-coupling and the origin of ring-opening of 2H-azirines”. In: *RSC Advances* 11.5 (2021), pp. 2744–2755. DOI: 10.1039/D0RA07498E.
- [26] T. Yoshimura, S. Maeda, T. Taketsugu, M. Sawamura, K. Morokuma, and S. Mori. “Exploring the full catalytic cycle of rhodium(i)–BINAP-catalysed isomerisation of allylic amines: a graph theory approach for path optimisation”. In: *Chemical Science* 8.6 (2017), pp. 4475–4488. DOI: 10.1039/C7SC00401J.
- [27] A. Nakao, Y. Harabuchi, S. Maeda, and K. Tsuda. “Leveraging algorithmic search in quantum chemical reaction path finding”. In: *Physical Chemistry Chemical Physics* 24.17 (2022), pp. 10305–10310. DOI: 10.1039/D2CP01079H.
- [28] I. M. Pendleton, M. H. Pérez-Temprano, M. S. Sanford, and P. M. Zimmerman. “Experimental and Computational Assessment of Reactivity and Mechanism in C(sp<sup>3</sup>)–N Bond-Forming Reductive Elimination from Palladium(IV)”. In: *Journal of the American Chemical Society* 138.18 (2016), pp. 6049–6060. DOI: 10.1021/jacs.6b02714.
- [29] J. R. Ludwig, P. M. Zimmerman, J. B. Gianino, and C. S. Schindler. “Iron(III)-catalysed carbonyl–olefin metathesis”. In: *Nature* 533.7603 (2016), pp. 374–379. DOI: 10.1038/nature17432.
- [30] A. L. Dewyer and P. M. Zimmerman. “Simulated Mechanism for Palladium-Catalyzed, Directed  $\gamma$ -Arylation of Piperidine”. In: *ACS Catalysis* 7.8 (2017), pp. 5466–5477. DOI: 10.1021/acscatal.7b01390.
- [31] J. R. Ludwig, S. Phan, C. C. McAtee, P. M. Zimmerman, J. J. Devery, and C. S. Schindler. “Mechanistic Investigations of the Iron(III)-Catalyzed Carbonyl–Olefin Metathesis Reaction”. In: *Journal of the American Chemical Society* 139.31 (2017), pp. 10832–10842. DOI: 10.1021/jacs.7b05641.
- [32] M. Bergeler, G. N. Simm, J. Proppe, and M. Reiher. “Heuristics-Guided Exploration of Reaction Mechanisms”. In: *Journal of Chemical Theory and Computation* 11.12 (2015), pp. 5712–5722. DOI: 10.1021/acs.jctc.5b00866.

- [33] G. N. Simm and M. Reiher. “Context-Driven Exploration of Complex Chemical Reaction Networks”. In: *Journal of Chemical Theory and Computation* 13.12 (2017), pp. 6108–6119. DOI: 10.1021/acs.jctc.7b00945.
- [34] S. H. Petrosko, R. Johnson, H. White, and C. A. Mirkin. “Nanoreactors: Small Spaces, Big Implications in Chemistry”. In: *Journal of the American Chemical Society* 138.24 (2016), pp. 7443–7445. DOI: 10.1021/jacs.6b05393.
- [35] B. Ensing, M. De Vivo, Z. Liu, P. Moore, and M. L. Klein. “Metadynamics as a Tool for Exploring Free Energy Landscapes of Chemical Reactions”. In: *Accounts of Chemical Research* 39.2 (2006), pp. 73–81. DOI: 10.1021/ar040198i.
- [36] L. Alessandro and L. G. Francesco. “Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science”. In: *Reports on Progress in Physics* 71.12 (2008), p. 126601.
- [37] V. van Speybroeck and R. J. Meier. “A recent development in computational chemistry: chemical reactions from first principles molecular dynamics simulations”. In: *Chemical Society Reviews* 32.3 (2003), pp. 151–157. DOI: 10.1039/B210410P.
- [38] S. A. Vázquez and E. Martínez-Núñez. “HCN elimination from vinyl cyanide: product energy partitioning, the role of hydrogen–deuterium exchange reactions and a new pathway”. In: *Physical Chemistry Chemical Physics* 17.10 (2015), pp. 6948–6955. DOI: 10.1039/C4CP05626D.
- [39] E. Martínez-Núñez. “An automated transition state search using classical trajectories initialized at multiple minima”. In: *Physical Chemistry Chemical Physics* 17.22 (2015), pp. 14912–14921. DOI: 10.1039/C5CP02175H.
- [40] J. A. Varela, S. A. Vázquez, and E. Martínez-Núñez. “An automated method to find reaction mechanisms and solve the kinetics in organometallic catalysis”. In: *Chemical Science* 8.5 (2017), pp. 3843–3851. DOI: 10.1039/C7SC00549K.
- [41] A. Rodríguez, R. Rodríguez-Fernández, S. A. Vázquez, G. L. Barnes, J. J. P. Stewart, and E. Martínez-Núñez. “TSSCDS2018: A code for automated discovery of chemical reaction mechanisms and solving the kinetics”. In: *Journal of Computational Chemistry* 39.23 (2018), pp. 1922–1930. DOI: 10.1002/jcc.25370.
- [42] E. V. Khramenkova, M. G. Medvedev, G. Li, and E. A. Pidko. “Unraveling the Nature of Extraframework Catalytic Ensembles in Zeolites: Flexibility and Dynamics of the Copper-Oxo Trimers in Mordenite”. In: *The Journal of Physical Chemistry Letters* 12.44 (2021), pp. 10906–10913. DOI: 10.1021/acs.jpcclett.1c03288.
- [43] R. J. Shannon, E. Martínez-Núñez, D. V. Shalashilin, and D. R. Glowacki. “ChemDyME: Kinetically Steered, Automated Mechanism Generation through Combined Molecular Dynamics and Master Equation Calculations”. In: *Journal of Chemical Theory and Computation* 17.8 (2021), pp. 4901–4912. DOI: 10.1021/acs.jctc.1c00335.

- [44] J. P. Unsleber, S. A. Grimmel, and M. Reiher. “Chemoton 2.0: Autonomous Exploration of Chemical Reaction Networks”. In: *Journal of Chemical Theory and Computation* 18.9 (2022), pp. 5393–5409. DOI: 10.1021/acs.jctc.2c00193.
- [45] P. Pracht, F. Bohle, and S. Grimme. “Automated exploration of the low-energy chemical space with fast quantum chemical methods”. In: *Physical Chemistry Chemical Physics* 22 (14 2020), pp. 7169–7192. DOI: 10.1039/C9CP06869D.
- [46] S. Grimme. “Exploration of Chemical Compound, Conformer, and Reaction Space with Meta-Dynamics Simulations Based on Tight-Binding Quantum Chemical Calculations”. In: *Journal of Chemical Theory and Computation* 15.5 (2019), pp. 2847–2862. DOI: 10.1021/acs.jctc.9b00143.
- [47] B. L. Mooney, L. Corrales, and A. E. Clark. “MolecularNetworks: An integrated graph theoretic and data mining tool to explore solvent organization in molecular simulation”. In: *Journal of Computational Chemistry* 33.8 (2012), pp. 853–860. DOI: 10.1002/jcc.22917.
- [48] A. Ozkanlar and A. E. Clark. “ChemNetworks: A complex network analysis tool for chemical systems”. In: *Journal of Computational Chemistry* 35.6 (2014), pp. 495–505. DOI: 10.1002/jcc.23506.
- [49] B. L. Mooney, L. R. Corrales, and A. E. Clark. “Novel Analysis of Cation Solvation Using a Graph Theoretic Approach”. In: *The Journal of Physical Chemistry B* 116.14 (2012), pp. 4263–4275. DOI: 10.1021/jp300193j.
- [50] C. M. Tenney and R. T. Cygan. “Analysis of Molecular Clusters in Simulations of Lithium-Ion Battery Electrolytes”. In: *The Journal of Physical Chemistry C* 117.47 (2013), pp. 24673–24684. DOI: 10.1021/jp4039122.
- [51] J.-H. Choi, H. Lee, H. R. Choi, and M. Cho. “Graph Theory and Ion and Molecular Aggregation in Aqueous Solutions”. In: *Annual Review of Physical Chemistry* 69.1 (2018), pp. 125–149. DOI: 10.1146/annurev-physchem-050317-020915.
- [52] S. Grimme, C. Bannwarth, and P. Shushkov. “A Robust and Accurate Tight-Binding Quantum Chemical Method for Structures, Vibrational Frequencies, and Noncovalent Interactions of Large Molecular Systems Parametrized for All spd-Block Elements ( $Z = 1-86$ )”. In: *Journal of Chemical Theory and Computation* 13.5 (2017), pp. 1989–2009. DOI: 10.1021/acs.jctc.7b00118.
- [53] T. Gensch, G. dos Passos Gomes, P. Friederich, E. Peters, T. Gaudin, R. Pollice, K. Jorner, A. Nigam, M. Lindner-D’Addario, M. S. Sigman, and A. Aspuru-Guzik. “A Comprehensive Discovery Platform for Organophosphorus Ligands for Catalysis”. In: *Journal of the American Chemical Society* 144.3 (2022), pp. 1205–1217. DOI: 10.1021/jacs.1c09718.
- [54] V. Sinha, J. J. Laan, and E. A. Pidko. “Accurate and rapid prediction of pKa of transition metal complexes: semiempirical quantum chemistry with a data-augmented approach”. In: *Physical Chemistry Chemical Physics* 23 (4 2021), pp. 2557–2567. DOI: 10.1039/D0CP05281G.

- [55] J. Preto and C. Clementi. “Fast recovery of free energy landscapes via diffusion-map-directed molecular dynamics”. In: *Physical Chemistry Chemical Physics* 16 (36 2014), pp. 19181–19191. DOI: 10.1039/C3CP54520B.
- [56] S. Wolf and G. Stock. “Targeted Molecular Dynamics Calculations of Free Energy Profiles Using a Nonequilibrium Friction Correction”. In: *Journal of Chemical Theory and Computation* 14.12 (2018), pp. 6175–6182. DOI: 10.1021/acs.jctc.8b00835.
- [57] F. Glover and M. Laguna. “Tabu Search”. In: *Handbook of Combinatorial Optimization: Volume 1–3*. Ed. by D.-Z. Du and P. M. Pardalos. Boston, MA: Springer US, 1998, pp. 2093–2229. ISBN: 978-1-4613-0303-9. DOI: 10.1007/978-1-4613-0303-9\_33.
- [58] S. Nandi, S. R. McAnanama-Brereton, M. P. Waller, and A. Anoop. “A tabu-search based strategy for modeling molecular aggregates and binary reactions”. In: *Computational and Theoretical Chemistry* 1111 (2017), pp. 69–81. DOI: 10.1016/j.comptc.2017.03.040.
- [59] S. Bougueroua, R. Spezia, S. Pezzotti, S. Vial, F. Quessette, D. Barth, and M.-P. Gaigeot. “Graph theory for automatic structural recognition in molecular dynamics simulations”. In: *The Journal of Chemical Physics* 149.18 (2018), p. 184102. DOI: 10.1063/1.5045818.
- [60] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein. *Introduction to Algorithms, 3rd Edition*. MIT Press, 2009. ISBN: 978-0-262-03384-8.
- [61] B. Maji and M. K. Barman. “Recent Developments of Manganese Complexes for Catalytic Hydrogenation and Dehydrogenation Reactions”. In: *Synthesis* 49.15 (2017), pp. 3377–3393. DOI: 10.1055/s-0036-1590818.
- [62] G. A. Filonenko, R. van Putten, E. J. M. Hensen, and E. A. Pidko. “Catalytic (de)hydrogenation promoted by non-precious metals - Co, Fe and Mn: recent advances in an emerging field”. In: *Chemical Society Reviews* 47.4 (2018), pp. 1459–1483. DOI: 10.1039/c7cs00334j.
- [63] J. Teichert. *Homogeneous Hydrogenation with Non-Precious Catalysts*. Wiley, 2020. ISBN: 9783527344390.
- [64] F. Kallmeier and R. Kempe. “Manganese Complexes for (De)Hydrogenation Catalysis: A Comparison to Cobalt and Iron Catalysts”. In: *Angewandte Chemie-International Edition* 57.1 (2018), pp. 46–60. DOI: 10.1002/anie.201709010.
- [65] P. Chandra, T. Ghosh, N. Choudhary, A. Mohammad, and S. M. Mobin. “Recent advancement in oxidation or acceptorless dehydrogenation of alcohols to valorised products using manganese based catalysts”. In: *Coordination Chemistry Reviews* 411 (2020), p. 213241. DOI: 10.1016/j.ccr.2020.213241.
- [66] R. van Putten, J. Benschop, V. J. de Munck, M. Weber, C. Müller, G. A. Filonenko, and E. A. Pidko. “Efficient and Practical Transfer Hydrogenation of Ketones Catalyzed by a Simple Bidentate Mn-NHC Complex”. In: *ChemCatChem* 11.21 (2019), pp. 5232–5235. DOI: 10.1002/cctc.201900882.

- [67] R. van Putten, G. A. Filonenko, A. Gonzalez de Castro, C. Liu, M. Weber, C. Müller, L. Lefort, and E. Pidko. “Mechanistic Complexity of Asymmetric Transfer Hydrogenation with Simple Mn–Diamine Catalysts”. In: *Organometallics* 38.16 (2019), pp. 3187–3196. DOI: 10.1021/acs.organomet.9b00457.
- [68] W. Yang, I. Y. Chernyshov, R. K. van Schendel, M. Weber, C. Müller, G. A. Filonenko, and E. A. Pidko. “Robust and efficient hydrogenation of carbonyl compounds catalysed by mixed donor Mn (I) pincer complexes”. In: *Nature Communications* 12.1 (2021), pp. 1–8. DOI: 10.1038/s41467-020-20168-2.
- [69] W. Yang, T. Y. Kalavalapalli, A. M. Krieger, T. A. Khvorost, I. Y. Chernyshov, M. Weber, E. A. Uslamin, E. A. Pidko, and G. A. Filonenko. “Basic Promoters Impact Thermodynamics and Catalyst Speciation in Homogeneous Carbonyl Hydrogenation”. In: *Journal of the American Chemical Society* 144.18 (2022), pp. 8129–8137. DOI: 10.1021/jacs.2c00548.
- [70] J. M. Pérez, R. Postolache, M. Castiñeira Reis, E. G. Sinnema, D. Vargová, F. de Vries, E. Otten, L. Ge, and S. R. Harutyunyan. “Manganese(I)-Catalyzed H–P Bond Activation via Metal–Ligand Cooperation”. In: *Journal of the American Chemical Society* 143.48 (2021), pp. 20071–20076. DOI: 10.1021/jacs.1c10756.
- [71] Y. Wang, S. Liu, H. Yang, H. Li, Y. Lan, and Q. Liu. “Structure, reactivity and catalytic properties of manganese-hydride amidate complexes”. In: *Nature Chemistry* (2022). DOI: 10.1038/s41557-022-01036-6.
- [72] R. van Putten, E. A. Uslamin, M. Garbe, C. Liu, A. Gonzalez-de-Castro, M. Lutz, K. Junge, E. J. M. Hensen, M. Beller, L. Lefort, and E. A. Pidko. “Non-Pincer-Type Manganese Complexes as Efficient Catalysts for the Hydrogenation of Esters”. In: *Angewandte Chemie-International Edition* 56.26 (2017), pp. 7531–7534. DOI: 10.1002/anie.201701365.
- [73] T. González and J. J. García. “Catalytic CO<sub>2</sub> hydrosilylation with [Mn(CO)<sub>5</sub>Br] under mild reaction conditions”. In: *Polyhedron* 203 (2021), p. 115242. DOI: 10.1016/j.poly.2021.115242.
- [74] A. M. Krieger and E. A. Pidko. “The Impact of Computational Uncertainties on the Enantioselectivity Predictions: A Microkinetic Modeling of Ketone Transfer Hydrogenation with a Noyori-type Mn-diamine Catalyst”. In: *ChemCatChem* 13.15 (2021), pp. 3517–3524. DOI: 10.1002/cctc.202100341.
- [75] A. V. Marenich, C. J. Cramer, and D. G. Truhlar. “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”. In: *The Journal of Physical Chemistry B* 113.18 (2009), pp. 6378–6396. DOI: 10.1021/jp810292n.
- [76] M. e. Frisch, G. Trucks, H. B. Schlegel, G. Scuseria, M. Robb, J. Cheeseman, G. Scalmani, V. Barone, G. Petersson, H. Nakatsuji, *et al.* *Gaussian* 16. 2016.

- [77] C. Bannwarth, S. Ehlert, and S. Grimme. “GFN2-xTB—An Accurate and Broadly Parametrized Self-Consistent Tight-Binding Quantum Chemical Method with Multipole Electrostatics and Density-Dependent Dispersion Contributions”. In: *Journal of Chemical Theory and Computation* 15.3 (2019), pp. 1652–1671. DOI: 10.1021/acs.jctc.8b01176.
- [78] R. van Putten, G. A. Filonenko, A. M. Krieger, M. Lutz, and E. A. Pidko. “Manganese-Mediated C–C Bond Formation: Alkoxyacylation of Organoboranes”. In: *Organometallics* 40.6 (2021), pp. 674–681. DOI: 10.1021/acs.organomet.0c00781.
- [79] G. L. Geoffroy, J. B. Sheridan, S. L. Bassner, and C. Kelley. “Migratory-insertion of carbon monoxide into metal-acyl bonds”. In: *Pure and Applied Chemistry* 61.10 (1989), pp. 1723–1729. DOI: doi:10.1351/pac198961101723.
- [80] F. De Angelis, A. Sgamellotti, and N. Re. “A dynamical density functional study of CO insertion into the metal–alkyl bond in  $\text{Ti}(\text{Cp})_2(\text{CH}_3)_2$ ”. In: *Journal of the Chemical Society, Dalton Transactions* (7 2001), pp. 1023–1028. DOI: 10.1039/B009833G.
- [81] P. Walker and R. Mawby. “Patterns of nucleophilic attack on tricarbonyl  $\pi$ -arene complexes of manganese(I)”. In: *Inorganica Chimica Acta* 7 (1973), pp. 621–625. DOI: 10.1016/S0020-1693(00)94897-7.
- [82] R. W. Johnson and R. G. Pearson. “Kinetics and mechanism of the cleavage reactions of acylmanganese pentacarbonyl and methylmanganese pentacarbonyl”. In: *Inorganic Chemistry* 10.10 (1971), pp. 2091–2095. DOI: 10.1021/ic50104a001.
- [83] C. M. Lukehart, G. P. Torrence, and J. V. Zeile. “Reactions on coordinated molecules. IV. Preparation of tris(cis-diacetyltetracarbonylmanganate)aluminum. Metalloacetylacetonate complex”. In: *Journal of the American Chemical Society* 97.23 (1975), pp. 6903–6904. DOI: 10.1021/ja00856a071.
- [84] C. P. Casey and C. A. Bunnell. “Site of nucleophilic attack on acylpentacarbonylmanganese (I) compounds”. In: *Journal of the American Chemical Society* 98.2 (1976), pp. 436–441.
- [85] J. Gladysz, G. Williams, W. Tam, and D. L. Johnson. “A convenient preparation of metal carbonyl monoanions by trialkylborohydride cleavage of metal carbonyl dimers; observation and reactions of a bimetallic manganese-formyl intermediate”. In: *Journal of Organometallic Chemistry* 140.1 (1977), pp. C1–C6. DOI: 10.1016/S0022-328X(00)84402-2.
- [86] J. Selover, M. Marsi, D. W. Parker, and J. Gladysz. “Mononuclear anionic formyl complexes; synthesis and properties”. In: *Journal of Organometallic Chemistry* 206.3 (1981), pp. 317–329. DOI: 10.1016/S0022-328X(00)83029-6.
- [87] A. M. Krieger, V. Sinha, G. Li, and E. A. Pidko. “Solvent-Assisted Ketone Reduction by a Homogeneous Mn Catalyst”. In: *Organometallics* 41.14 (2022), pp. 1829–1835. DOI: 10.1021/acs.organomet.2c00077.

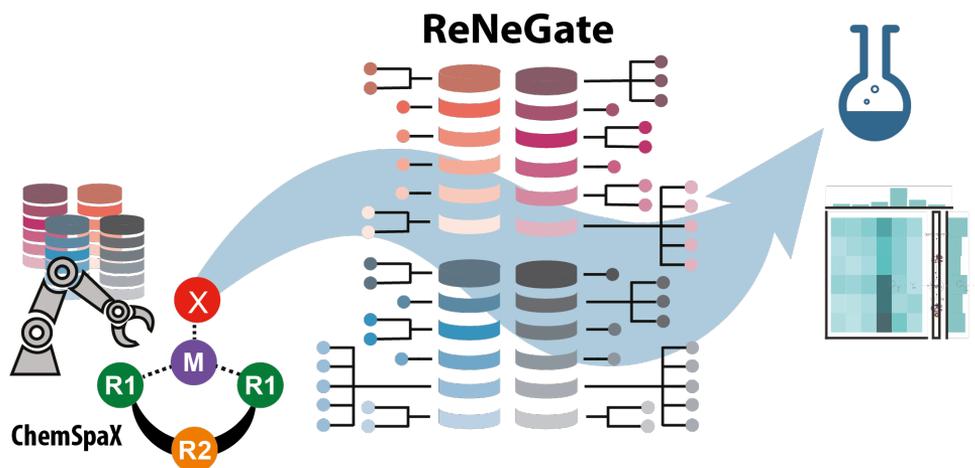
- [88] R. van Putten. "Catalysis, chemistry, and automation: Addressing complexity to explore practical limits of homogeneous Mn catalysis". Thesis. 2021. DOI: 10.4233/uuid:e94a4c47-b3f7-45a7-a38c-9c319c38a51f.



# 4

## HiREX: HIGH-THROUGHPUT REACTIVITY EXPLORATION

*for Extended Databases of Transition Metal Catalysts*



**A** METHOD is introduced for the automated reactivity exploration of extended *in silico* databases of transition metal catalysts. The proposed workflow is designed to tackle two key challenges for bias-free mechanistic explorations on large databases

This chapter is based on A. Hashemi, S. Bougueroua, M.-P. Gageot, and E. A. Pidko. "High-Throughput Reactivity Exploration for Extended Databases of Transition Metal Catalysts." In: *ChemRxiv* (2023). DOI: 10.26434/chemrxiv-2023-f76nv

*of catalysts: 1. automated exploration of the chemical space around each catalyst with unique structural and chemical features and 2. automated analysis of the resulting large chemical datasets. To address these challenges we have extended the application of our previously developed ReNeGate method for bias-free reactivity exploration and implemented an automated analysis procedure to identify the classes of reactivity patterns within specific catalyst groups. Our procedure applied to an extended series of representative Mn(I) pincer complexes revealed correlations between structural and reactive features pointing to new channels for catalyst transformation under the reaction conditions. Such an automated high-throughput virtual screening of systematically generated hypothetical catalyst datasets opens new opportunities for the design of high performance catalysts as well as an accelerated method for expert bias-free high-throughput in silico reactivity exploration.*

## 4.1. INTRODUCTION

THE ability and the need to create molecular structures with tailored (bio)chemical functions has been driving chemical research. In order to address this need, traditional experimental chemistry, conventionally guided by intuition, chemical knowledge, and serendipity has been successful in discovering functional molecular frameworks and improving their characteristics towards desired properties. For example, molecular catalysts are decorated with diverse functional groups to explore their activity and stability and devise possible strategies for the activity improvement[1–3]. However, such a design strategy is always limited by the synthetic and physical availability of the particular chemicals to the experimentalists, which substantially limits the scope of the exploration of the theoretically accessible catalysis space. While *in vitro* functionalization can provide insights into the chemical design principles behind high activity, selectivity, and stability, it can also be demanding in terms of time and resources. As a result, computational (*in silico*) molecular design is becoming a practical and promising alternative, due to recent advancements in quantum chemical methods and high-performance computing[1–10]. These high-throughput computational methods can help create highly effective functionalization strategies by exploring geometries within the local chemical space of a given molecular framework[5].

Organometallic chemistry space presented for development of new catalysts for useful chemical transformations is very large: it can be viewed as a combinatorics of the (i) transition metal (TM) centers with (ii) varied oxidation states in (iii) different coordination environments established by the organic ligands. The common practice is to assume that the variations in the chemistry of the ligand environment (ligand functionalization) does not affect the main mechanistic and reactivity properties, but only the energetics of the associated paths. Therefore, one can first investigate in detail the mechanism and reactivity of a particular selected catalyst, followed by high-throughput screening using the descriptors or targets identified for a specific complex. However, one can expect that the chemical modification of a catalyst can open new mechanistic possibilities. The catalytic properties of the organometallic complexes are governed by a much wider reactivity space. Therefore, to enable the high-throughput *in silico* catalyst screening, one ideally has to explore the reactivity of each member of the catalyst library, with the associated problem of the combinatorial explosion resulting in an extremely large and complex datasets of results that need to be analyzed. Furthermore, the featurization and labeling of homogeneous catalysts based on their structural features as well as distinct reactivities remains a challenge when exploring unexpected chemistries. In the context of expert-bias free reactivity exploration, methods for correlating structural and reactive features and extraction of reaction classes would be required. The collection of comprehensive results on the reactivities of structures with systematic modifications to tune the properties in databases would then facilitate the design of functional catalysts for specific targets. In this work, we introduce a workflow that potentially addresses these challenges by combining automated functionalization with automated reactivity screening towards catalyst deactivation and the subsequent automated analysis of the resulting large dataset. The workflow is developed for

the analysis of organometallic manganese pincer complexes as representative and relevant catalyst family.

Organometallic pincer complexes are homogeneous catalysts used in various reactions and have been successful in (de)hydrogenation of various substrates[11–13]. Pincer complexes have been successfully applied in energy, pharmaceutical, and fine chemistry[14, 15]. Examples include Ru[16, 17], Ir[18], and earth-abundant 3d transition metals like Fe and Mn[19, 20]. Manganese is especially appealing due to its high biocompatibility in food and pharma applications.

Experimental investigations on finding catalysts with optimal properties for defined functions are limited to a few accessible variants of the functionalized pincer backbones. Theoretical investigations, on the other hand, can be much broader by design and can navigate through arbitrary regions of the chemical space in exploratory search for defined properties[21–24]. Such chemical space exploration can be guided by systematic functionalization of the backbone to find highly stable and active catalysts. Modern computational chemistry methods are instrumental for such a task and have been used successfully in the past to screen through large databases of functionalized TM complexes[25, 26], including pincer complexes, for activity, regioselectivity, and ligand effects[27–29]. Recently, Krieger et al presented a computational stability study on a virtual library of Mn pincer complexes within the constraints of the pre-defined deactivation chemistry[30].

Here we aim at removing expert bias from the reactivity analysis by integrating the automated procedures for organometallic complexes generation, reactivity exploration and analysis into a unified workflow (Figure 4.1). The ChemSpaX[31] fully automated procedure was used for ligand functionalization to generate 576 Mn pincer complexes[31] that form the input for the dynamic reactivity explorations with the ReNeGate[32] procedure. The reactivity exploration results are organized in a database with structural and reactive properties observed for every record as features in the database. The database is then analysed to find correlations between structural and reactive features followed by the extraction of reaction classes identified for each catalyst entry and family. Such systematic, automated and bias-free exploration and analysis has provided insights in defining reaction classes and deactivated states correlated with specific combinations of backbone and ligand modifications.

The chapter is organized as follows: first, we provide an extensive description of the methodological aspects of the presented workflow. We start with the introduction of the functionalization strategy ChemSpaX used to construct the extended synthetic pincer catalyst set, followed by the description of the automated dynamic exploration workflow ReNeGate applied to this dataset. The methodology section is concluded by the presentation of the automated analysis method applied to multiple reactive trajectories generated for the extended synthetic catalyst dataset. The results and discussion section presents the application of this workflow on the selected families of Mn pincer catalysts. The manuscript is completed with a conclusion section. *HiREX* code is publicly available at: <https://github.com/ahashemiche/HiREX>

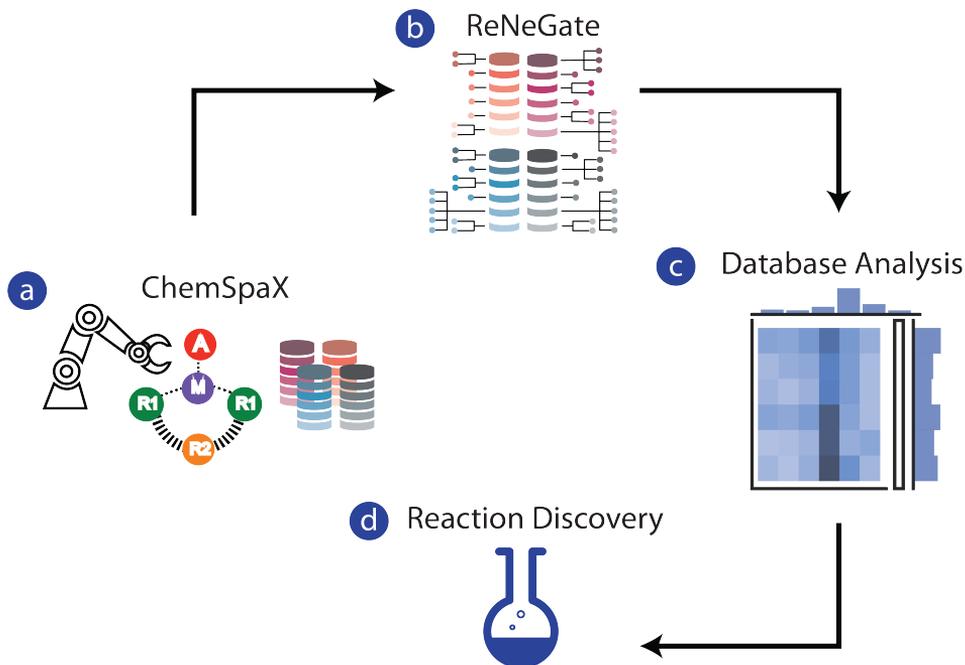


Figure 4.1.: Computational workflow for (a) generation, (b) chemical space exploration and (c) analysis of extended in silico catalyst datasets for (d) the discovery of new reactivities and particularly those giving rise to the thermodynamically driven catalyst degradation/deactivation.

## 4.2. METHODS

### 4.2.1. FUNCTIONALIZATION STRATEGY WITH CHEMSPA<sub>X</sub>

In silico catalyst screening aims at analyzing the effect of functionalization type and site (backbone/donor) on the reactivity behavior. The ChemSpax virtual Mn catalyst library used for this study contained complexes with four representative pincer scaffolds, namely, PNP- (bis(3-phosphaneylpropyl)amine)-, SNS- (azanediylbis(ethane-1-thiol))-, CNC- (bis(2-(1H-3λ 4-imidazol-3-yl)ethyl)amine)- and PNN- (N1-(2-phosphaneylethyl)ethane-1,2-diamine)- backbones coordinated to a Mn(I) center stabilized by CO ligands and an anionic (X) group (Figure 4.2). The effects of first and second coordination spheres are respectively analysed by varying  $R_1$  (directly coordinated to Mn center) and  $R_2$  (ligand backbone) substituents. To simplify analysis, ChemSpa<sub>X</sub> virtual library has been limited to symmetric functionalizations for each site group, although this does not represent the fundamental limitation of the ChemSpa<sub>X</sub> procedure, which is described in detail elsewhere[31].

For each pincer ligand, four different Mn-adducts were considered with Br-, OH-, OMe-, and  $O^t$ Bu- anions as the anionic ligands representing the common pre-catalyst

and/or resting state species (Figure 4.2). The different combinations of  $R_1$ ,  $R_2$  and X functionalization with the four pincer scaffolds gave rise to a virtual library containing 576 complexes that were further used for the reactivity exploration and analysis. Pincer complexes based on these ligands have been reported for various transition metals, including manganese[19, 33, 34]. SNS-[35–39], CNC-[40–44], and PCP-[45–49] ligands are primarily known for their use in 4d and 5d transition metal catalysis. Very efficient catalyst systems for a wide range of catalytic transformations have been established using 4d and 5d transition metal pincer complexes[36, 40, 48, 50]. The literature on their 3d-transition metal counterparts often reports lower catalytic efficiencies[33, 49], due to the tendency of such systems to deactivate and form highly stable resting states under the reaction conditions, limiting their catalytic performance[51, 52].

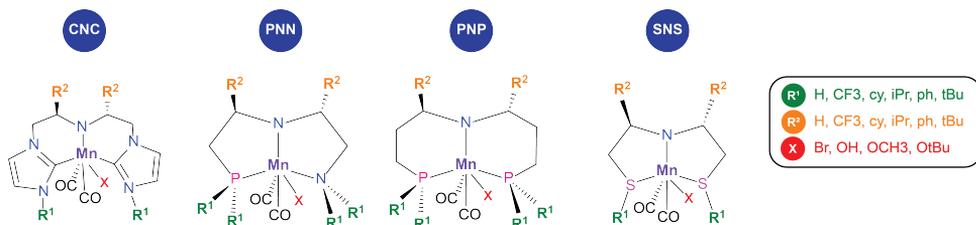


Figure 4.2.: Representative 3d transition metal (Mn) pincer scaffolds with the ligand modification  $R_1$ ,  $R_2$  and TM adduct (X) within the scope of this work. The functionalizations  $R_1$  and  $R_2$  include proton (H), trifluoromethyl ( $\text{CF}_3$ ), cyclohexyl(cy), isopropyl( $i\text{Pr}$ ), phenyl(Ph), and tert-butyl( $t\text{Bu}$ ); and adducts (X) can be represented by bromide (Br), hydroxyl(OH), methoxide ( $\text{OCH}_3$ ), tert-butoxide ( $\text{O}^t\text{Bu}$ ) anions.

The design of the virtual library is guided by our focus on the application of Mn(I)-pincers as dehydrogenation catalysts[53–55]. A set of ligands are chosen for functionalizations at  $R_1$  and  $R_2$  positions. The ligands are representative of the choices made by experimentalists for screening purposes and cover a broad range of steric and electron withdrawing properties. Among the set of X adducts, Br is a common precursor to the active form of manganese pincer catalysts[56]. Such precursors often go through activation with a strong base to produce a highly active 5 coordinated complex. OR-adducts (OH,  $\text{OCH}_3$  and  $\text{O}^t\text{Bu}$ ) are usually formed upon addition of alcohol/water/base via metal ligand cooperative addition[57]. Formation of such adducts might form slow down or even deactivated the catalyst depending on their stability[52, 58–62].

#### 4.2.2. REACTIVITY EXPLORATIONS - THE ReNeGate WORKFLOW

The virtual organometallic pre-catalyst library was fed as the input for the dynamic mechanistic explorations. In order to remove expert bias from mechanistic studies and to discover new chemistries, our automated graph-theoretical methodology implemented in the ReNeGate[32] workflow is used to explore the potential energy

surface around each starting structure. The individual entities in the database are automatically provided in parallel as input to the exploration step, where the chemical space around each structure is exhaustively explored for alternative chemical structures and new reactions. The detailed description of the ReNeGate workflow is provided elsewhere[32], while below we shortly summarize the key elements of the procedure.

The procedure involves an exhaustive reaction space through root-mean-square-deviation (RMSD) biased meta-dynamics at the semiempirical xTB level of theory using the CREST functionality in the GFN-xTB code[63]. Recent studies demonstrate the sufficient accuracy of the xTB for high-throughput screening of TM complexes including Mn(I)-based systems discussed herein as the representative model catalysts[64]. Implementation of the RMSD bias in the metadynamics simulations helps with the exhaustiveness of the exploration. CREST workflow sets a penalty on the configurations that have already been visited by calculating relative structural RMSD values for new configurations. Reactive trajectories from all metadynamics runs for all input structures from the ChemSpaX library are provided in parallel to the ReNeGate graph-theoretical analysis tool to identify unique structures and compare the observed configurations with the respective reference structures (Section 2.3). Alternative explored structures for all starting geometries in the database are then collected in a global database. Each element in the database is labelled based on the known structural features of the reference structure (backbone scaffold, ligands at  $R_1$  and  $R_2$  position, adduct) as well as the calculated features including relative energies ( $\Delta E$ , kcal.mol<sup>-1</sup>) and specific reactivity observed for species compared to the starting reference structure (broken or formed bonds). The database is initially analyzed for correlations between structural features and observed reactivities.

### 4.2.3. COMPARISON WITH REFERENCE STRUCTURES

Exploration results for each starting structure are analyzed to provide insights into possible reaction (or deactivation) channels for given catalyst structures. Herein we used as starting structure a high-throughput computational database designed to investigate the thermodynamic stability of different adducts in various functionalized Mn(I) based pincer complexes[30]. While the database was originally designed to investigate metal-ligand cooperative activation of HX (X=Br, OH, OMe, O<sup>t</sup>Bu) bonds, one can expect that the chemical modification of a catalyst can open new mechanistic possibilities. Therefore, to enable the high-throughput in silico catalyst screening, we have explored the reactivity of each member of the catalyst library. Compared to the protocol used in the original ReNeGate[32] workflow, uniquely identified structures are compared against the respective reference structure, which allows the transfer of the extracted reaction labels between different backbone,  $R_1$ ,  $R_2$  and adduct classes. This provides global insights into the role of individual features on the energetics and reactivity patterns. Pipeline for populating databases of reactivities for chemical structures is illustrated in Figure 4.3. Indexing trajectories, identification of conformers with unique fingerprints, population of database and featurization and analysis are the key steps for extracting insights from high throughput virtual screenings of the current study. The algorithm proposed for

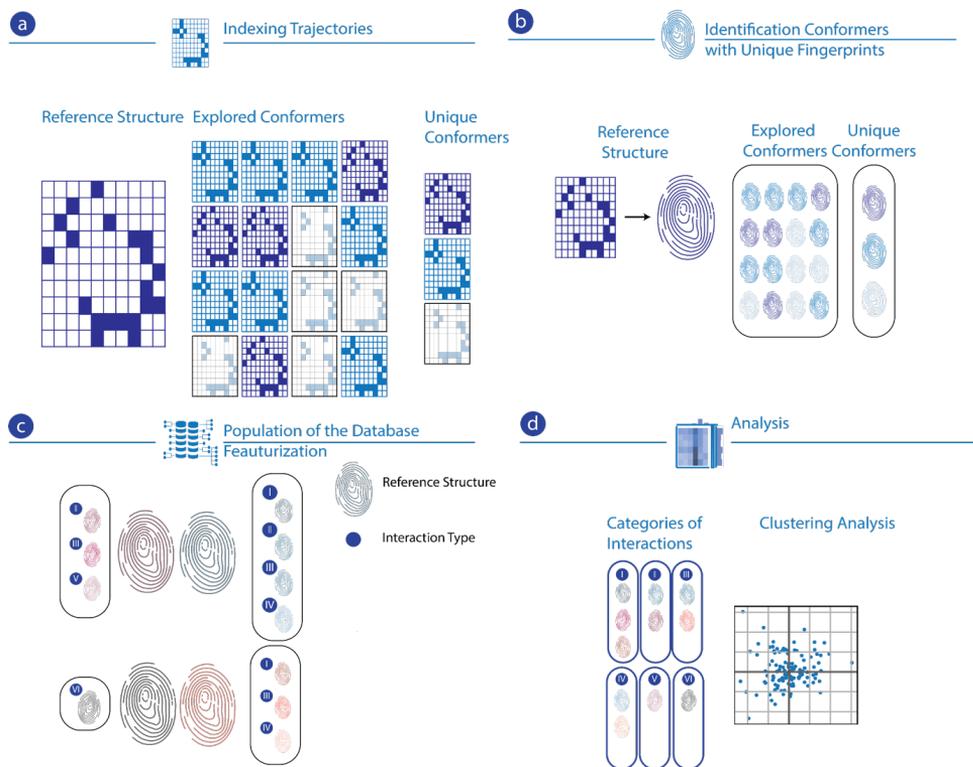


Figure 4.3.: Pipeline for populating the reactivity databases of chemical structure based on comparisons with reference structures involving a) indexing trajectories, b) identification of conformers with unique fingerprints, c) population of the database, and d) analysis through categorization and clustering.

making the comparison is discussed below.

Given a sequence of structures  $S_1, S_2, S_3, \dots, S_n$  where  $S_i$  represents unique configurations observed based on conformer exploration and  $S_1$  is the reference structure, the algorithm, first analyzes the conformer changes based on the dynamics of interactions and bonds by applying isomorphism tests. Once the conformers are identified (noted  $C_1, C_2, \dots, C_m$ ), we compare these conformers to the reference structure represented by the first conformer  $C_1$ .

The algorithm follows the following steps:

1. Read the xyz file, at each step:
  - a) Construct the mixed graph[32, 65]
  - b) Apply the isomorphism test
  - c) Identify conformers that are not isomorphic

- d) Compare bonding pattern between new conformer and reference structure
2. Read the xyz file, at each step:
    - a) Extract the energy values from the trajectory analyzed at the level of theory (either DFTB or DFT) and assign it to the correspond conformer
    - b) Identify the minimum and maximum value of energy for each conformer
  3. Construct the graph of evolution of structures

The bonding patterns of the identified conformer and the respective reference are compared. For the bonds present in the new conformer, but not in the reference, we consider that the bond is formed (marked with a '+') and, respectively, for the bonds present in the reference and not in the new conformer, the bond is considered broken (and marked with a '-'). Based on the changes in the bonding patterns, the differences between each conformer and the reference structure are stored in the database. In addition, we extract the energy values from the input file (the xyz file that contains the structures  $S_1, S_2, S_3, \dots, S_n$ ), and calculate the relative energy difference of the conformer through comparison with the reference. For the purposes of the current study, we consider the minimum energy value from the isomers' list as the representative energy value assigned to the unique configuration for further analysis on the database. We then construct the graph of evolution of the reference structures. The vertices in the graph represent the identified conformers and edges connect the conformers and the reference structure. The representation includes the 2D image of the conformer, the list of structures that belong to this conformer and the energy values extracted from the trajectory at the analysis level of theory (either DFTB or DFT). Similar to the original ReNeGate implementation, the reaction network calculations are carried out at two different levels of theory. Initial dynamic explorations are done at the GFN2-xTB[63] level of theory and then the identified unique conformer structures are refined at B3LYP-D3/6-31g(d,p) level of theory with a GD3BJ dispersion correction[66] and an implicit SMD model[67] with the standard parameters for THF solvent.

## 4.3. RESULTS AND DISCUSSION

### 4.3.1. HIGH-THROUGHPUT REACTIVITY EXPLORATION

The primary results of the application of our automated reactivity analysis procedure to the virtual library of Mn pincer complexes are summarized in Figure B.4 and demonstrate that the explored structures vary in terms of relative energy in [-600, +100] kcal.mol<sup>-1</sup> range. Few species have been observed with the relative energies as low as -600 kcal.mol<sup>-1</sup>, which can be attributed to the fact that the structures systematically built by substitution of ligands could be unphysical and provides clues to the real (un)synthesizability of the reference structure. Therefore, we have limited further analysis in classification of reaction classes to species observed in [-40, +25] kcal.mol<sup>-1</sup> range. The refined results in terms of the relative stabilities of the discovered species for each catalyst class/group are summarized in Figure 4.4. The introduced energy constraint is expected to improve the reliability of the

defined reaction classes. This will also help make sure that low energies obtained for some states are not artifacts coming from automated design of structures and not synthesizable species. Exploration results on virtual library of Mn catalyst is

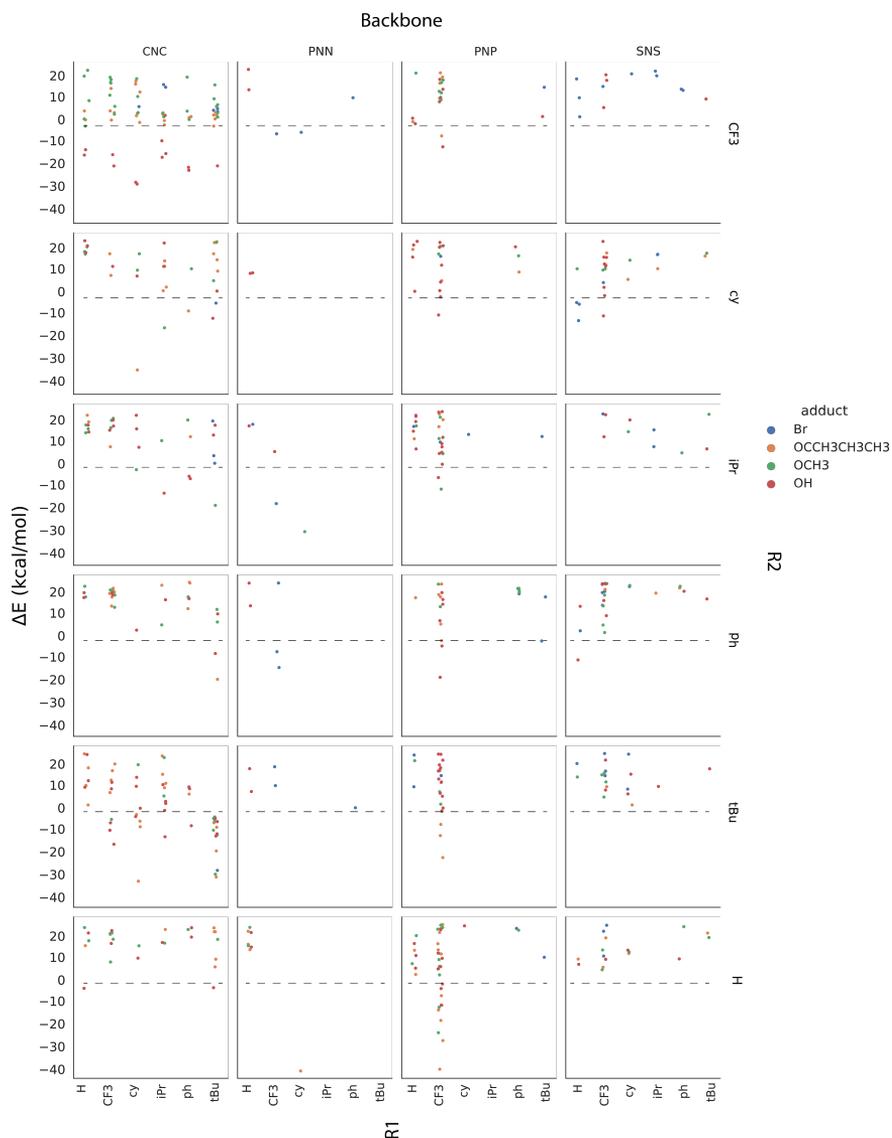


Figure 4.4.: Exploration results for the evolution of the Mn-CNC, PNN, PNP and SNS pre-catalysts in  $[-40, 25]$  kcal.mol<sup>-1</sup> stability range. The configurations are analyzed based on the backbone type and moieties at  $R_1$  and  $R_2$  positions. Data points are colored according to the adduct X at the Mn center.

summarized in Figure 4.4 Relative energies of the explored species are illustrated as a function of backbones, modifications on  $R_1$  and  $R_2$  positions. Data points are colored based on the adduct X at the Mn center. Our results highlight varied trends in reactivity as a function of pincer scaffold and functionalization. For example, the data in Figure 4.4 suggest that pincers featuring CNC backbone can isomerize into more stable alternative species for all  $R_1$  and  $R_2$  functionalization and Mn-X adduct types. On the other hand, for the PNP family, the favorable transformation to stable alternative species is mainly limited to complexes with a  $CF_3$  substituent at  $R_1$ . Systems with the PNN backbone are the least likely to form alternative configurations. These apparent trends can be readily deduced from the current representations of stability diagrams. However, further and deeper correlation analysis is necessary to gain more insight into the observed variations.

#### 4.3.2. CORRELATION ANALYSIS OF THE STRUCTURAL FEATURES AND REACTIVITY

Histogram plots in Figure 4.5 were made to illustrate how the combinations of different structural features are associated with the possibility of the formation of alternative stable configurations from a given pre-catalyst. These plots summarize the frequency of appearance of species with relative stabilities in the  $[0, -40]$  kcal.mol<sup>-1</sup> with respect to the given structural/compositional features. Correlations between (a) the  $R_1$  groups and other structural features including catalyst backbone, (b)  $R_2$  groups, (c) coordination number (CN) of the metal (Mn) center, as well as (d) the energetics of the alternative species ( $\Delta E$ ) and (e) the adducts (X) type on the Mn center are summarized in Figure 4.5. The number of explored species in the database having specific features with different combinations of  $R_1$  with other parameters (backbone,  $R_2$ , CN,  $\Delta E$ , X) are counted on the upper and right side of each histogram. Specific combination cells are colored according to the frequency of the observations with the color coding shown next to each plot. The frequency of observations of identified structures with specific feature combinations has been considered as a measure of correlation between structural and reactive features. A similar correlation analysis for configurations with energies within +25 kcal.mol<sup>-1</sup> from their respective reference state are collected in Figure B.1 of the supporting information.

Figure 4.5.a reveals a correlation between the presence of <sup>t</sup>Bu and  $CF_3$  groups at the  $R_1$  positions for CNC and PNP backbones. As opposed to catalysts with PNN, PNP and SNS backbones, stable species with CNC backbones have been found with all variations of ligands at  $R_1$  positions. Figure 4.5.a shows that explorations on CNC catalysts have resulted in the formation of alternative stable species for all ligand substituents at  $R_1$  positions. The  $CF_3$  group at  $R_1$  leads to the formation of more stable structures for PNN, PNP and SNS backbones. Regarding the  $R_1$ - $R_2$  combinations (Figure 4.5.b), the presence of <sup>t</sup>Bu and  $CF_3$  at  $R_1$  always leads to the formation of alternative structures, indicating the direct involvement of  $R_1$  ligands in the formation of stable intermediates.

The formation of stable alternative species with 5- and 6-coordinated Mn centers is respectively correlated with the presence of <sup>t</sup>Bu and  $CF_3$  moieties at  $R_1$  (Figure

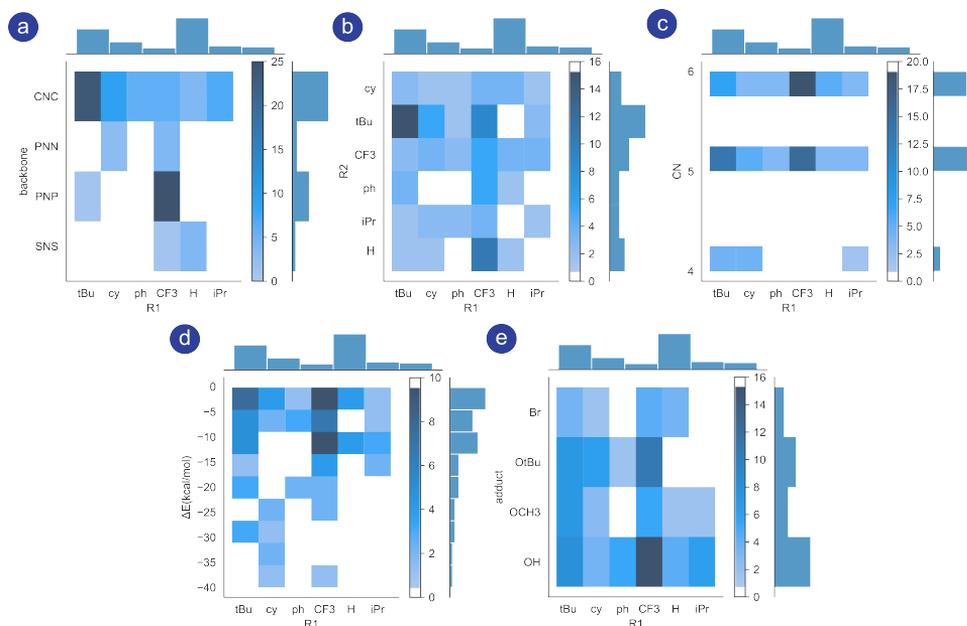


Figure 4.5.: Histogram plots for the frequency of appearance of alternative stable species based on modified structures for CNC, PNP, PNN and SNS catalysts (lower energies than the respective reference structure) as a function of feature combination (a)  $R_1$ -backbone, (b)  $R_1$ - $R_2$ , (c)  $R_1$ -CN (coordination number of the metal center) and (d)  $R_1$ - $\Delta E$  (relative energy), (e)  $R_1$ -X.

4.5.c). 4-coordinated Mn complexes were observed only for the pincer complexes with most bulky  $t$ Bu, Cy and  $i$ Pr substituents at the donor atoms. Interestingly, the  $R_1$  substitutions with the bulky  $t$ Bu and Cy as well as the most reactive  $CF_3$  groups is found to be correlated with the more pronounced stabilization of the alternative configurations ( $R_1$ - $\Delta E$  correlations in Figure 4.5.d). The analysis of the  $R_1$ -X correlations presented in Figure 4.5.e suggests that the OH adducts universally tend to convert to other more stable configurations for all  $R_1$  functionalizations. This emphasizes the role of OH- ligand in forming stable structures and this will be discussed in detail below. The diversity of the species explored based on the type of the adduct follows the order  $OH > O^tBu > OCH_3 > Br$ . Energetics of the observed structures colored based on different adducts are further illustrated in Figures B.2, B.3 and B.4 for  $[-40,0]$ ,  $[-40,25]$  and  $[-600,100]$  kcal.mol $^{-1}$  ranges, respectively. These insights from correlations observed between the structural and reactive features of the catalysts provides clues for a more detailed analysis of the observed species and new chemistries in Section 4. In the next section, we present clustering analysis for finding better insights into the catalysts' chemical reactivity features. The nature of the observed chemistries are discussed in detail Section 4.

### 4.3.3. FINDING THE MOST FREQUENT CLASSES OF INTERACTIONS: K-MODE CLUSTERING

In the next step towards implementing data-driven analysis of large databases of homogeneous catalyst structures, we have leveraged on labeled data organized based on automated exploration on Mn-based virtual library described previously. We have enumerated the global distinct types of interactions leading to alternative stable species and subsequently done clustering to find the most frequent modes. Clustering, in general, is an unsupervised learning method whose task is to divide the population or data points into a certain number of groups, such that data points belonging to the same group are similar to each other and dissimilar to the data points in the other groups. It is basically a collection of objects based on similarity and dissimilarity between them. K-mode clustering<sup>[68]</sup> is one of the unsupervised machine learning algorithms that is used to cluster categorical variables. Here we used the `kmodes 0.12.2`<sup>[68]</sup> library for categorical clustering based on the reactivities of the observed species. The top 10 clusters shown in Figure 4.6 represent the most frequent types of interactions observed for different  $R_1$ ,  $R_2$  ligand, adduct and backbone combinations will be discussed.

The results of the clustering analysis is summarized in Figure 4.6 by presenting the relative energy diagrams of the alternative configurations discovered for each of the backbone type and classified according to the interaction type realized in them. Analyzing the nature of interactions identified in Figure 4.6, we can draw insights into the new reactivity of Mn(I) pincers and their correlations with the type of backbone and modifications to the ligand backbones at  $R_1$  and  $R_2$  positions. Filtering the new transformations in the database for events have led to three different sets of reactions:

- Decoordination of the Mn-D (D = C, N, P) bonds
- Nucleophilic attack on the carbonyl ligands
- Migrations of the  $CF_3$  moieties

Decoordination of the Mn donor ligand bonds for catalysts with CNC, PNP, PNN and SNS backbones has been observed to produce alternative stable structures and can be followed to identify hemilabile donor ligands. Hemilability can occur with polydentate ligands, which have at least two coordinating groups often with different electronic properties. For complexes where hemilabile ligands are present, one coordinating group is easily displaced from the metal center while the other group remains firmly bound. In addition to the general observation on the cleavage of the Mn donor bonds, we identify that the starting complex can undergo nucleophilic attack by the CO ligand (NuA), persistent for all catalyst types. Migration of  $CF_3$  moieties to the Mn center has been observed for catalysts with PNP backbone. Besides, we also find that coordinative unsaturation of the Mn center can be stabilized by the formation of defined agostic interactions with C-H (Mn-H) or C-F moieties (Mn-F) on the distant ligand substituents. Such additional interactions are commonly observed as accompanying the former isomerization types that commonly

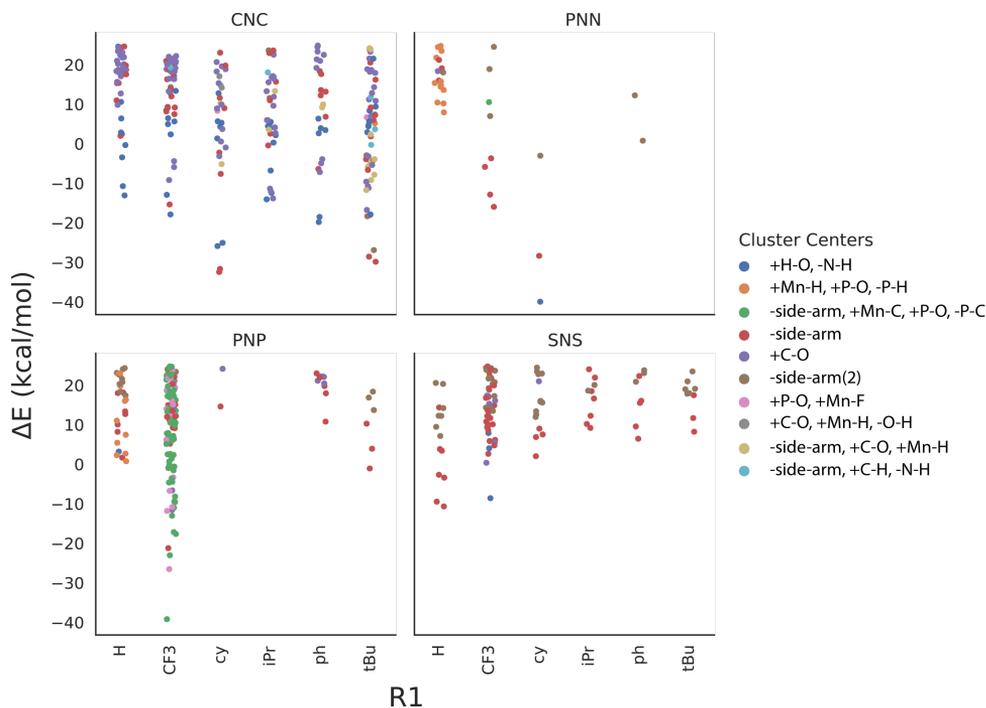


Figure 4.6.: The top 10 clusters representing the most frequent types of interactions observed for different  $R_1$ ,  $R_2$  ligand and backbone combinations. The energies for species explored species in the  $[-40, +25 \text{ kcal.mol}^{-1}]$  energy window are plotted as a function of ligand modification at  $R_1$  position. Data points are colored by the type of interactions defined by the cluster centers (10 centers).

give rise to the change of the coordination environment of the Mn center. Further details on the specific cases where such interactions happen are discussed below.

The cluster centers identified and used to classify species in Figure 4.6 are explained in terms of the chemistry they represent:

- **+H-O, -N-H:** for catalysts with CNC backbone and OH (or OCH<sub>3</sub>) adducts with  $R_2 = CF_3$ , the hydrogen atom on the backbone nitrogen can migrate to the adduct and further be followed by the detachment of the adduct from the Mn center (Table B.2).
- **+Mn-H, +P-O, -P-H:** for catalyst with PNN and PNP backbones with  $R_1 = H$ , manganese hydride species is formed via the transfer of H from  $R_1$  to Mn (+Mn-H, -P-H) followed by bridging of the alkoxide species between Mn and the phosphorus atom (+P-O).
- **-side-arm, +Mn-C, +P-O, -P-C:** For catalysts with PNP backbone (also a single

case for PNN catalyst), the sidearm (-side-arm) P-donor decoordinates from the Mn center, the  $CF_3$  moiety migrates from  $R_1$  to Mn (+Mn-C, +P-O, -P-C).

- **-side-arm(-Mn-N, -Mn-C, -Mn-P, -Mn-S):** decoordination of the sidearm observed for catalyst with CNC, PNN, PNP and SNS backbones resulting in new stable configurations.
- **'+C-O', '+C-O, +Mn-H, -O-H', '-side-arm, +C-O, +Mn-H':** For catalysts with CNC backbone, nucleophilic attack from adducts (+C-O) into carbonyls is observed for all different modifications to the moieties at  $R_1$  position. Such interactions have been observed to be accompanied by agostic hydride interactions (+Mn-H) or the dissociation of the side-arm.
- **-side-arm(2):** decoordination of two side-arms observed mostly for catalysts with SNS backbone with all types of  $R_1$  substitution. Such reactivity is also observed for catalysts with PNN backbone with  $R_1 = CF_3$  and for PNP backbones with  $R_1 = H$  or  $tBu$ .
- **+P-O, +Mn-F:** for PNP catalysts with  $R_1 = CF_3$ , the alkoxide adduct ( $X = OR$ ) is transferred into the bridging position between Mn and P (+P-O), the new state is stabilized by an agostic Mn-F interactions.
- **-side-arm, +C-H, -N-H:** decoordination of the side-arm accompanied by the insertion of the hydride from the backbone amine has been observed for catalysts with CNC backbone when either bulky ( $tPr$  or  $tBu$ ) or  $CF_3$  moieties are present at  $R_1$  position.

Analysis based on the nature of the bond changes shows that the cleavage of the Mn donor bonds is most frequently observed and is accompanied by other reactive events. As shown in Figure 4.6, the de-coordination of the donor ligands allows the formation of stable structures for catalysts with CNC, PNN, PNP and SNS backbones. De-coordination for catalysts with CNC backbone is observed when bulky substituents ( $tBu$  and  $cy$ ) are present at the  $R_1$  position of the N-heterocyclic carbene donor groups. While the formation of stable structures for catalysts with PNN and SNS backbones was always accompanied by dissociation of the donor ligands, stable structures were observed with and without the dissociation of the donor ligands for the CNC and PNP backbones. These results are discussed in detail in the next Section.

## 4.4. NEW CHEMISTRY

### 4.4.1. DECOORDINATION OF THE MN-D (D = C, N, P) BONDS

In coordination chemistry and catalysis, the reactivity and stability of catalytic species can be tuned through hemilability of the pre-catalytic species[69]. The partial and reversible displacement of the ligand may provide an additional kinetic stabilization of the reactive catalytic complexes and serve as the point of entry into the catalytic cycle or facilitate the regeneration of the catalytically active species at

the stage of the product release. The facile interconversion between the fully and hemi-coordinated states may be critical for the fast catalytic turn-over rate[70]. On the other hand, the increased reactivity of the complex due to the (partial) ligand dissociation may give rise to further conversion resulting in the long-term catalyst deactivation. We therefore first focused the more in-depth reactivity analysis on the phenomenon of labile donor atoms of the pincer scaffold. The reactivity patterns of structures were queried for the decooordination of either or a combination of side arm donor ligands (-Mn-C, -Mn-N, -Mn-P or -Mn-S). Presence of one or more of these interactions for a structure has then been marked as an indicator for the probable presence of hemilabile Mn-donor bonds in the reference structures. Figure 4.8 summarizes the results of the automated hemilability detection by showing the relative stabilities of the different structures in the database labeled as either intact (not exhibiting decooordination of side arm donor ligands) or having the -Mn-D( $D = C, N, P, S$ ) feature indicating hemilability.

4

Our analysis reveals that the decooordination of Mn-N and Mn-S bonds are the most frequent side-arm decooordination events (Figure 4.7). For clarity we have clustered structures including Mn-N decooordination with a similar “-Mn-N Set” label. Further discussions on different combinations of Mn-N decooordination behavior has been discussed in Figure 4.8. Further analysis on the different  $R_1$ - $R_2$ -adduct combinations leading to specific reactivities including -Mn-N has been summarized in Section B.3 of the Supporting Information. Figures B.5, B.6, B.7 and B.8 summarize the detailed “-Mn-N set” interaction for CNC, PNN, PNP and SNS backbones, respectively. Phosphine dissociation in PNN and PNP catalyst families is rare with most cases found for the bulky  $R_1 = Ph$ , cy and  $tBu$ . The detailed plots including expanded lability interactions are included in the supporting information and will be discussed here.

For CNC catalysts, bulky  $tBu$  and cy groups at the backbone  $R_2$  sites favor the rather unexpected decooordination of the central amino donor(-Mn-N) of the pincer ligand. A similar behavior (-Mn-N) is also observed when  $CF_3$  or no functional groups are present at  $R_2$  position along with H or  $CF_3$  groups at  $R_1$  is observed. Ligand dissociation was not observed for all configurations of the Br adduct of Mn-CNC. (Figure B.9). For PNN complexes, the presence of electron withdrawing  $CF_3$  groups at either of  $R_1$  or  $R_2$  position generally promotes the N-donor hemilability (-Mn-N) that results in a substantial stabilization (by  $-200 \text{ kcal.mol}^{-1}$ ) of the Mn complexes for all  $R_1$ - $R_2$ -X Mn-PNN combinations. Decoordination of both Mn-N bonds (“-Mn-N(2)”), “-Mn-N, -Mn-P” or complete decooordination of the ligands are uniquely observed for the species with  $R_1 = CF_3$  (Figure B.6).

The decooordination behavior of Mn-PNP catalysts includes the dissociation of both the central amino (-Mn-N) and phosphine side-arm (-Mn-P) donors. While decooordination of the Mn-N bond is observed for all  $R_1$ - $R_2$  combinations, decooordination of one Mn-P is observed only for  $R_1 = H$  or  $CF_3$ . Furthermore, a complete dissociation of all donor atoms of the backbone is observed for  $R_1 = CF_3$  and  $R_2 = H$  (Figure B.7).

For the catalysts with the SNS pincer scaffolds, the ligand lability was also detected for all donor atoms of the pincer (-Mn-N, -Mn-S and -Mn-S(2), where the latter

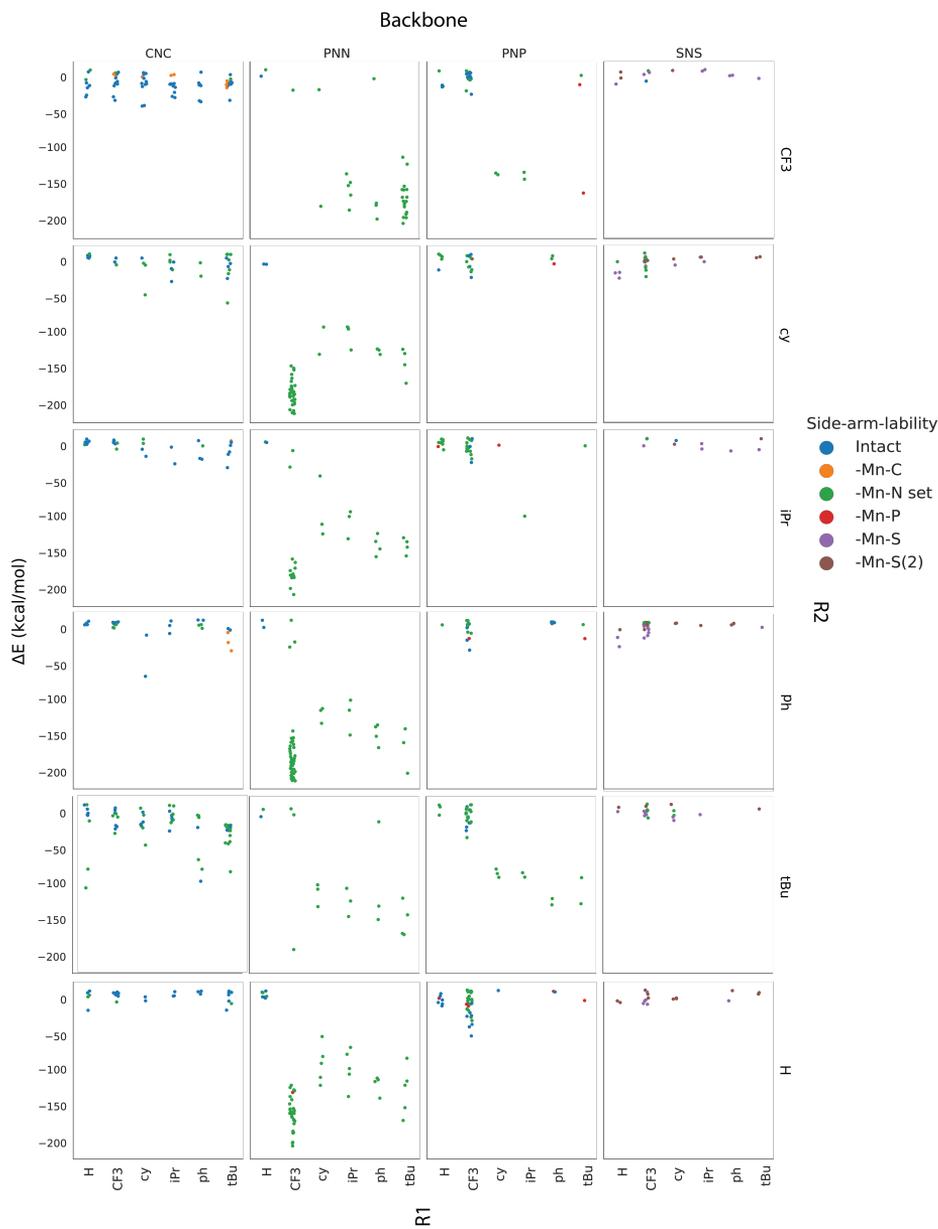


Figure 4.7.: De-coordination of the donor atoms in the Mn(I) complexes with CNC, PNN, PNP and SNS backbones within  $[-200 \text{ kcal.mol}^{-1}, 25 \text{ kcal.mol}^{-1}]$   $\Delta E$  from reference structures. Structures are colored based on the type of the donor atom dissociated as presented in the legend.

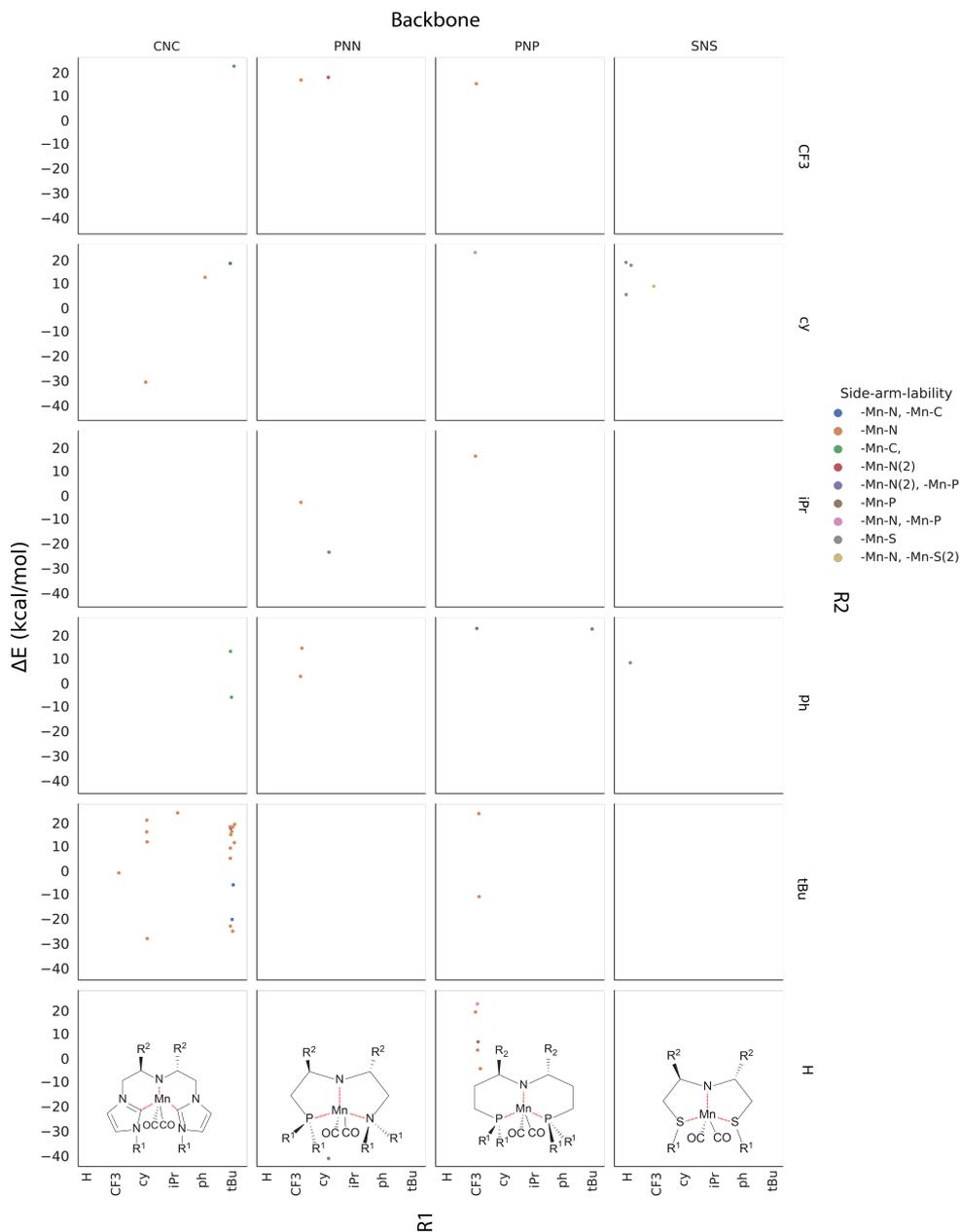


Figure 4.8.: Decoordination behavior for catalysts with CNC, PNN, PNP and SNS scaffolds stabilized in  $[-40, 0]$  kcal.mol<sup>-1</sup> range with respect to the reference structures.

indicates the decoordination of both S donor atoms). The presence of  $CF_3$  groups at  $R_1$  again promotes strongly the dissociation of both the N and S donors. In this catalyst family, the dissociation of the central Mn-N is less common than the cleavage of the weaker Mn-S bonds and it was observed for the  $R_1 = CF_3$  and  $R_2 = cy$  combination (Figure B.8).

In order to identify possible hemilabile states and to distinguish between the side-arm ligand decoordination and hemilability, we further classified the decoordination reactivities to only events leading to more stable structures compared to the reference structure (Figure 4.8). Limiting the analysis to species within a narrow energy window provides a better understanding of the discrimination between reversible and irreversible decoordination of sidearm donor ligands.

As shown in Figure 4.8 and also Figure B.9 in the supporting information, bulky ( $tBu$  and  $cy$ ) groups at  $R_1$  position of CNC, induce decoordination of the central N donor (-Mn-N) and may also promote the dissociation of the NHC side-arms (-Mn-C). Decoordination of the NHC moieties of CNC (-Mn-C bond) is also observed with  $tBu$ - $cy$  combination of the  $R_1$ - $R_2$  groups. Presence of  $CF_3$  on the phosphine donors of the PNP backbone also promotes the cleavage of the side-arm (Mn-P or Mn-N) ligands, whereas in the case of the SNS pincers, the H atoms at  $R_1$  sites generally promote the cleavage of one or both of the Mn-S side-arm donor ligands. (Figures B.7 and B.8) Most frequently, the energetically favorable dissociation of the central nitrogen donor of the pincer scaffold (-Mn-N bond) has been observed for CNC, PNN or PNP backbones. The dissociation of the Mn-C bond (either as '-Mn-C' or '-Mn-N, -Mn-C') is unique for the CNC backbone and will be discussed in section 3.2 in detail. Other interactions leading to the complete dissociation of the backbone ligands, including ('(-Mn-N) $_2$ ' for PNN backbones, '(-Mn-N) $_2$ , -Mn-P' for catalysts with PNP backbone, '-Mn-N, (-Mn-S) $_2$ ' for catalysts with SNS backbone) have been also observed, but they led to structures with generally higher energies than those where only a single side-arm has dissociated. Presence of  $tBu$  groups at  $R_2$  position of the backbone promotes the hemilability of the side arm donor atoms for catalysts with all 4 backbones in this study. In the meantime, for PNN, PNP and SNS catalysts, the favorable dissociation of the side-arm is not affected with the  $R_2$  functionalization of the backbone, but is solely controlled by the nature of the  $R_1$  group at the donor moiety. The different correlations between the nature of the adduct and hemilability of side-arm donor ligands are summarized in Figures B.5-B.9 in the supplementary information. Based on the number of explored species for which side-arm decoordination was observed, the diversity of the hemilabile species explored based on the type of the adduct follows the order  $OH > O^tBu > OCH_3 > Br$ .

Nucleophilic attack and insertion into the carbonyls have been observed for all of the basic adducts (OH,  $OCH_3$  and  $O^tBu$ ) in the database. This is in line with our previous hypothesis on the potential relevance of such a reactivity for the base-assisted deactivation paths of Mn catalysts stabilized by carbonyl CO ligands[32, 71, 72]. Let us consider in more details the revealed reactivity patterns by focusing on the Mn-CNC complex family. Figure 4.9 summarizes the results of the automated analysis and shows the optimized structures of the most stable species identified for each class (denoted with the roman numbers in Figure 4.9). Energetic of the

observed structures is plotted on the Y-axis against the set of reactivities observed for catalyst structures with the CNC backbone shown on the X-axis. Symbols representing modifications at  $R_1$  and  $R_2$  positions as well as color coding for adduct present on the Mn center are shown in the legend.

Besides the migratory insertion of the basic hydroxide or alkoxide adduct, the NHC moieties of the CNC pincer have also been observed to attack the carbonyl ligands as illustrated with structures VII, IX and X (Figure 4.9). It should be noted that for the very bulky substituents where  $t$ Bu and cy groups are respectively present at  $R_1$  and  $R_2$  positions, the central nitrogen ligand can also decoordinate as observed in structure VII. Such transformations are accompanied by the dissociation of the amine donor of the pincer ligand and the associated energy losses are compensated partially by either the conformational changes to the backbone (VIII and IX) or the formation of an additional agostic interactions with Mn (X). Presence of the bulky groups at  $R_2$  position causes decoordination of nitrogen and provides the NHC to act as a base and leads to carbene-carbonyl coupling. Carbene-carbonyl coupling was always precluded by decoordination of the nitrogen from the Mn center and dissociation of the Mn-C bond before the coupling took place. Such (-Mn-N, +C-O) series of events lead to stabilization by as much as  $-14 \text{ kcal.mol}^{-1}$ . The complete set of CNC structures including -Mn-C reactivity are further analyzed in Section B.4 of the Supporting Information. Figure B.9 and Table B.1 summarize all observed -Mn-C interactions (with no trimming based on relative energies). The stabilization by agostic interactions were detected with the H atoms from (one or two)  $t$ Bu groups at  $R_1$  positions or even the C-H moieties from the pincer backbone (II, III and IV). For the complexes bearing less bulky (OH and  $\text{OCH}_3$ ) adducts, their attack on the CO ligands was not accompanied by the pincer decoordination ('+C-O', '+C-O, +Mn-H') (V and VI). The favorable NHC-carbonyl coupling has been observed for the most bulky CNC complexes featuring the combination of  $R_1 = t$ Bu and  $R_2 = cy$  or  $t$ Bu and  $X = Br$  or  $\text{O}^t\text{Bu}$  (VIII, IX and X, 9). Such chemistry has been reported earlier in several experimental studies. For example, Ruiz et al. explored the production of N-metalated NHC generated by the deprotonation of 1-phenylimidazole (L) in a cationic fac-[Mn(L)(CO)<sub>3</sub>(bipy)]<sup>+</sup> complex and described an similar reactivity for the transient formation of carbene[73]. It was established that the deprotonated imidazole and an auxiliary carbonyl ligand engaged in the mechanism converting carbene into the more stable imidazolyl tautomer, which was responsible for the production of the acyl intermediate. It has also been shown that lithiated azoles can be added nucleophilically to  $[\text{M}(\text{CO})_6]$  ( $M = Cr, Mo, W$ ) to yield acyl intermediates that can then be alkylated to generate azolyl alkoxy carbene complexes[74]. Huertos et al. reported on the intramolecular nucleophilic attack of deprotonated imidazoles to coordinated bipyridine and imidazole ligands[75]. The new pathways for nucleophilic additions observed through high throughput screening are of interest due to the high utility of catalytic transformations incorporating CO into organic substrates to create higher value products. Exemplary cases of such interactions include the hydroaminomethylation of simple vinylic arenes to produce a variety of useful pharmaceuticals in a one-pot reaction and the phosgene-free carbonylation of amino and phenolic compounds. The extent M-NHC catalysts will promote

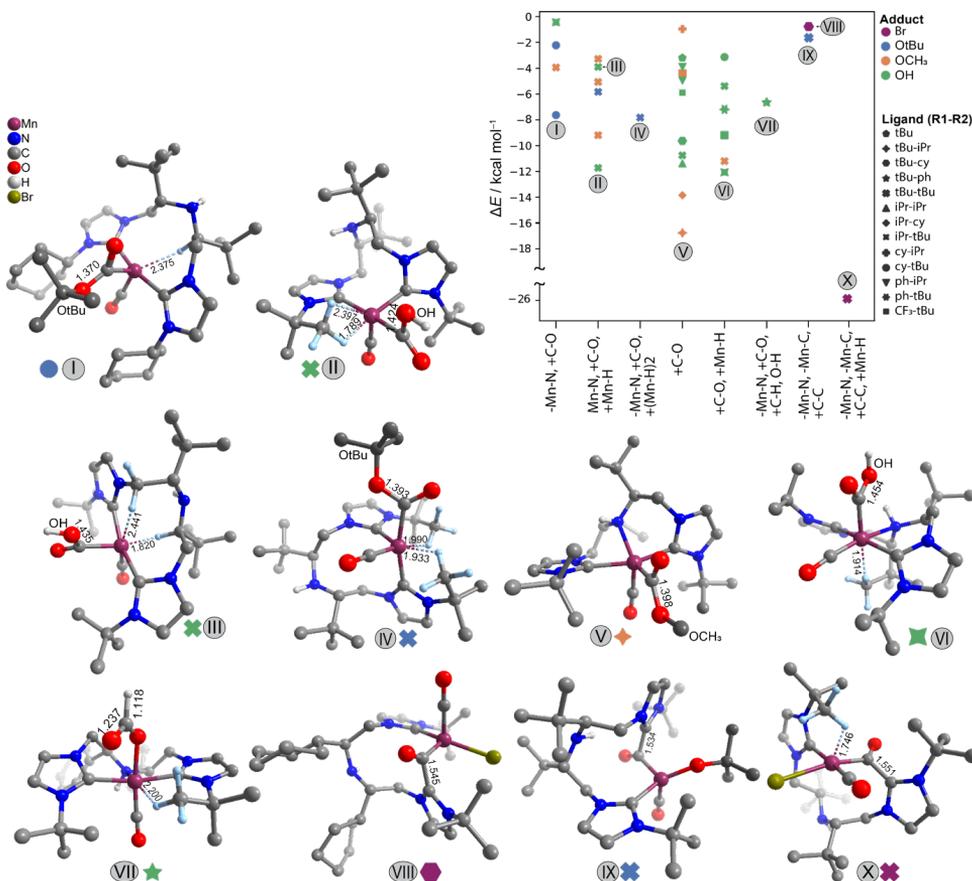


Figure 4.9.: The optimized structures of the most stable species explored for CNC family with different R<sub>1</sub>-R<sub>2</sub> combinations (hydrogen atoms are omitted for clarity, only the H-atoms involved in agostic interactions with the Mn center are shown with as light blue spheres).

reactions that use CO as a C1-carbon source has been extensively discussed in the literature[76, 77].

#### 4.4.2. MIGRATIONS OF THE CF<sub>3</sub> MOIETIES

Another new reactivity has been identified for the complexes featuring CF<sub>3</sub>-functionalized ligand scaffolds. The spontaneous migration of the originally P-bound CF<sub>3</sub> moiety to the Mn center with the concomitant exchange of the original alkoxide species has been identified. The most stable respective structures are summarized in Figure 4.10.

Figure 4.10 shows the new chemical interactions observed including the migration of the CF<sub>3</sub> groups. Energetic of the observed structures is plotted on the Y-axis

against the set of reactivities observed for catalyst structures with the PNP backbone shown on the X-axis. Symbols representing modifications at R<sub>1</sub> and R<sub>2</sub> positions as well as color coding for adduct present on the Mn center are shown in the legend. The set of reactivities summarized on the X-axis are briefly discussed below.

- \* : summarizes a series of three events leading to the migratory insertion of the CF<sub>3</sub> group to the Mn center including: I) the cleavage of the P-CCF<sub>3</sub> bond (-P-C), II) formation of the new P-O bond between P and the alkoxide (OH or O<sup>t</sup>Bu) adduct on the Mn center (+P-O(OH or O<sup>t</sup>Bu)) and III) formation of the new Mn-CCF<sub>3</sub> bond (+Mn-C).

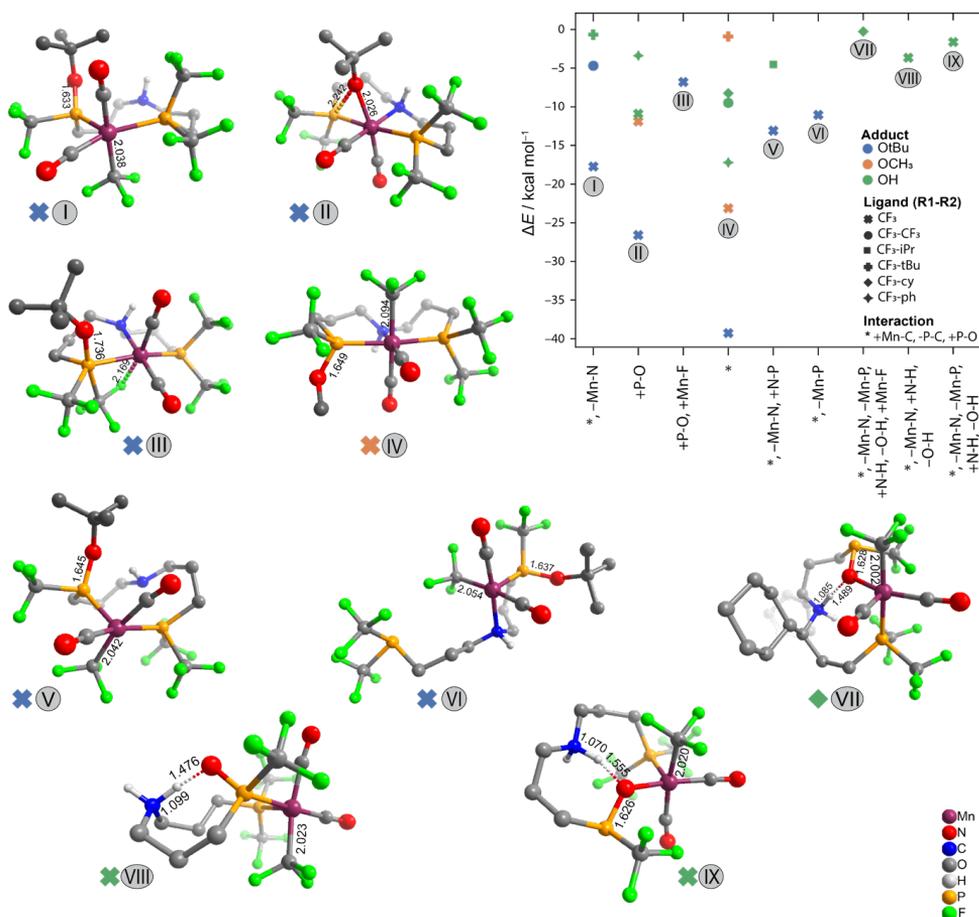


Figure 4.10.: Formation of manganese carbonyl trifluoromethyl complexes for catalysts with PNP backbone. The optimized structures of the most stable species from each class of interactions are presented (hydrogen atoms are omitted for clarity).

Other reactivities described in Figure 4.10 along with the migratory insertion of  $\text{CF}_3$  include:

- \*, **-Mn-N** Structure I: For structures with OH and  $\text{O}^t\text{Bu}$  adducts, decoordination of the nitrogen in the backbone has been observed along with the  $\text{CF}_3$  migration and leads to structures stabilized by  $\text{ca.}18 \text{ kcal.mol}^{-1}$ .
- **+P-O**: Structure II : Irrespective of the type of the adduct on the Mn center, it has been observed that the oxygen atom on the adduct can form a bridge with the adjacent phosphorus and can be an initial step for migration of  $\text{CF}_3$  groups.
- **+P-O, +Mn-F**: Structure III: Migration of the adduct to the phosphorus and the undercoordination of the Mn center is compensated by coordination with the F atom (Structure III)
- \*, **-Mn-P**: (Structure VI) the migration of the  $\text{CF}_3$  moiety to the Mn center has been observed to be accompanied by the dissociation of the Mn-P bond when there are no ligands present at  $\text{R}_2$  position (Structure VI)

Figure 4.10. shows that structures formed via the migratory insertion are more stable than their reference structures within the  $[-39.2, -23.1] \text{ kcal.mol}^{-1}$  range of energies when no substitutions are present at  $\text{R}_2$  and within  $[-17.1 \text{ kcal.mol}^{-1}, -0.8 \text{ kcal.mol}^{-1}]$  when ph,  $\text{CF}_3$ , cy and  $^t\text{Bu}$  groups are present. (Figure 4.10 and, Table B.3). Migratory insertion of  $\text{CF}_3$  onto the Mn center can be accompanied by the cleavage of the Mn-N bond for  $\text{O}^t\text{Bu}$  and OH adducts. The adduct can be bridged between the Mn and P centers (initial step for migration of  $\text{CF}_3$  groups, Structure II, Figure 4.10). Migration of the adduct to the phosphorus is additionally stabilized by the formation of a short contact between the Mn and fluoride ion (Structure III). The migration of  $\text{CF}_3$  onto Mn center with concomitant exchange of the original alkoxide species was observed for OH and  $\text{OCH}_3$  adducts.

This chemistry that we have identified purely from the expert-bias-free high-throughput computational reactivity analysis is in line with the previous studies chemical systems with metal perfluoroalkyl bonds (M-RF) and the respective catalytic applications. Indeed, the more conventional organometallic compounds and especially metal alkyls (M-R) are immensely important players in catalysis[78]. Catalysis utilizing metal fluoroalkyl complexes, however, is less common due to the inherent stability of M-RF[79]. Such compounds, on the other hand, receive increasing attention for their utility for the field of fluoro-organic synthesis[80, 81]. For example, [Cu]-RF compounds are utilized as the stoichiometric reagents for perfluoroalkyl transfer to organic substrates[82–85]. Recently, there is much attention to transition-metal-catalyzed (with such metals as Cu, Ni, Pd) C-RF (where RF is usually  $\text{CF}_3$ ) bond-forming reactions[86, 87] providing a route to various valuable fluorinated pharmaceuticals and agrochemicals[80, 81, 88]. Daniels et al. discussed the synthesis, characterization and reactivity of several bi- and tridentate, N-ligated manganese carbonyl trifluoromethyl complexes[79]. All these complexes feature elongated Mn- $\text{CCF}_3$  bonds suggesting the lability of the moiety, which could

potentially be exploited for the transfer or insertion of the  $\text{CF}_3$  group into organic substrates. Poli and co-workers investigated in detail the thermal decarbonylation of the acyl compounds  $[\text{Mn}(\text{CO})_5(\text{CORF})]$  ( $\text{RF} = \text{CF}_3, \text{CHF}_2, \text{CH}_2\text{CF}_3, \text{CF}_2\text{CH}_3$ ) and the formation of  $[\text{Mn}(\text{CO})_5(\text{RF})]$  species containing M-alkyl moieties[89]. The Mn-RF moiety is highly labile and can undergo homolytic dissociation upon moderate heating or when subject to photochemical (UV or visible light). For example, such activation procedures allow  $[\text{Mn}(\text{CO})_5(\text{CF}_3)]$  compound with the strongest Mn-RF bond initiate the radical polymerization of vinylidene fluoride ( $\text{CH}_2 = \text{CF}_2$ ) to produce poly(vinylidene fluoride)[89]. The migration of the  $\text{CF}_3$  moiety to the Mn center identified by the current automated algorithm is thus in line with the previous experimental investigations on the related chemistries and suggests new avenues to expand this field by utilizing  $\text{CF}_3$ -modified ligand scaffolds and secondary transformations of the respective Mn catalysts.

## 4.5. CONCLUSIONS

A method is introduced for the automated exploration of reactivities of extended databases of transition metal catalysts. The proposed workflow is designed to tackle the key challenges for bias-free mechanistic explorations on large databases of catalysts, namely: I) the automated exploration of the chemical space around each catalyst given specific structural features and (II) the automated analysis of results from such chemical datasets and provision of design rules for catalyst with improved performances or new reactivity. To address these challenges, we have extended the application of our previously developed ReNeGate method for bias-free chemical space exploration to databases of synthetic organometallic catalysts. We implemented an analysis procedure to identify the classes of reactivity patterns within specific catalyst groups in the large organometallic datasets. Our procedure applied to an extended series of representative Mn(I) pincer complexes revealed new correlations between the structural and reactive features pointing to new channels for catalyst transformation under the reaction conditions. Specifically, we have identified different hemilability behavior for catalyst with CNC, PNN, PNP and SNS backbones with a data-driven approach. Understanding hemilability is important because it affects the energy changes associated with pre-activation and regeneration steps in the catalytic process, and in turn, influences the coordination sphere and geometry of the complex. In addition, two new classes of reactivities, namely, nucleophilic attack on carbonyl ligands and migration of  $\text{CF}_3$  moieties have been identified through high-throughput virtual screening on the databases. Such a bias-free high-throughput virtual screening on the systematically designed structures opens new opportunities for the design of high-performance catalysts as well as an accelerated method for exploring new reactivity patterns.

## REFERENCES

- [1] C. Y. Cheng, J. E. Campbell, and G. M. Day. “Evolutionary chemical space exploration for functional materials: computational organic semiconductor discovery”. In: *Chemical Science* 11.19 (2020), pp. 4922–4933. DOI: 10.1039/D0SC00554A.
- [2] M. Renom-Carrasco and L. Lefort. “Ligand libraries for high throughput screening of homogeneous catalysts”. In: *Chemical Society Reviews* 47.13 (2018), pp. 5038–5060. DOI: 10.1039/C7CS00844A.
- [3] S. Hiroto, Y. Miyake, and H. Shinokubo. “Synthesis and functionalization of porphyrins through organometallic methodologies”. In: *Chemical Reviews* 117.4 (2017), pp. 2910–3043. DOI: 10.1021/acs.chemrev.6b00427.
- [4] J. H. Van Drie. “Computer-aided drug design: the next 20 years”. In: *Journal of Computer-Aided Molecular Design* 21.10 (2007), pp. 591–601. DOI: 10.1007/s10822-007-9142-y.
- [5] N. Fey. “Lost in chemical space? Maps to support organometallic catalysis”. In: *Chemistry Central Journal* 9.1 (2015), pp. 1–10. DOI: 10.1186/s13065-015-0104-5.
- [6] D. E. Clark. “What has virtual screening ever done for drug discovery?”. In: *Expert Opinion on Drug Discovery* 3.8 (2008), pp. 841–851. DOI: 10.1517/17460441.3.8.841.
- [7] A. Jain, Y. Shin, and K. A. Persson. “Computational predictions of energy materials using density functional theory”. In: *Nature Reviews Materials* 1.1 (2016), pp. 1–13. DOI: 10.1038/natrevmats.2015.4.
- [8] L. Cheng, R. S. Assary, X. Qu, A. Jain, S. P. Ong, N. N. Rajput, K. Persson, and L. A. Curtiss. “Accelerating electrolyte discovery for energy storage with high-throughput screening”. In: *The Journal of Physical Chemistry Letters* 6.2 (2015), pp. 283–291. DOI: 10.1021/jz502319n.
- [9] H. B. Dizaji and H. Hosseini. “A review of material screening in pure and mixed-metal oxide thermochemical energy storage (TCES) systems for concentrated solar power (CSP) applications”. In: *Renewable and Sustainable Energy Reviews* 98 (2018), pp. 9–26. DOI: 10.1016/j.rser.2018.09.004.
- [10] J. Hachmann, R. Olivares-Amaya, S. Atahan-Evrenk, C. Amador-Bedolla, R. S. Sánchez-Carrera, A. Gold-Parker, L. Vogt, A. M. Brockway, and A. Aspuru-Guzik. “The Harvard clean energy project: large-scale computational screening and design of organic photovoltaics on the world community grid”. In: *The Journal of Physical Chemistry Letters* 2.17 (2011), pp. 2241–2251. DOI: 10.1021/jz200866s.

- [11] S. Werkmeister, K. Junge, and M. Beller. "Catalytic hydrogenation of carboxylic acid esters, amides, and nitriles with homogeneous catalysts". In: *Organic process research & development* 18.2 (2014), pp. 289–302. DOI: 10.1021/op4003278.
- [12] L. Maser, L. Vondung, and R. Langer. "The ABC in pincer chemistry—From amine-to borylene-and carbon-based pincer-ligands". In: *Polyhedron* 143 (2018), pp. 28–42. DOI: 10.1016/j.poly.2017.09.009.
- [13] S. Werkmeister, J. Neumann, K. Junge, and M. Beller. "Pincer-Type Complexes for Catalytic (De)Hydrogenation and Transfer (De)Hydrogenation Reactions: Recent Progress". In: *Chemistry – A European Journal* 21.35 (2015), pp. 12226–12250. DOI: 10.1002/chem.201500937.
- [14] A. Mohanty, R. Sharma, and P. Daw. "Chapter 1 - Application of pincer metal complexes in catalytic transformations". In: *Pincer-Metal Complexes*. Ed. by A. Kumar. Elsevier, 2022, pp. 1–68. ISBN: 978-0-12-822091-7. DOI: 10.1016/B978-0-12-822091-7.00003-8.
- [15] X. Yang and X. Chen. "Chapter 5 - Mechanistic Insights and Computational Prediction of Base Metal Pincer Complexes for Catalytic Hydrogenation and Dehydrogenation Reactions". In: *Pincer Compounds*. Ed. by D. Morales-Morales. Elsevier, 2018, pp. 101–110. ISBN: 978-0-12-812931-9. DOI: 10.1016/B978-0-12-812931-9.00005-0.
- [16] W. Kuriyama, T. Matsumoto, O. Ogata, Y. Ino, K. Aoki, S. Tanaka, K. Ishida, T. Kobayashi, N. Sayo, and T. Saito. "Catalytic Hydrogenation of Esters. Development of an Efficient Catalyst and Processes for Synthesising (R)-1,2-Propanediol and 2-(1-Menthoxy)ethanol". In: *Organic Process Research & Development* 16.1 (2011), pp. 166–171. DOI: 10.1021/op200234j.
- [17] T. Otsuka, A. Ishii, P. A. Dub, and T. Ikariya. "Practical selective hydrogenation of  $\alpha$ -fluorinated esters with bifunctional pincer-type ruthenium(II) catalysts leading to fluorinated alcohols or fluoral hemiacetals". In: *Journal of the American Chemical Society* 135.26 (2013), pp. 9600–9603. DOI: 10.1021/ja403852e.
- [18] W. Aoki, N. Wattanavinin, S. Kusumoto, and K. Nozaki. "Development of Highly Active Ir–PNP Catalysts for Hydrogenation of Carbon Dioxide with Organic Bases". In: *Bulletin of the Chemical Society of Japan* 89.1 (2016), pp. 113–124. DOI: 10.1246/bcsj.20150311.
- [19] M. Garbe, K. Junge, S. Walker, Z. Wei, H. Jiao, A. Spannenberg, S. Bachmann, M. Scalone, and M. Beller. "Manganese (I)-Catalyzed Enantioselective Hydrogenation of Ketones Using a Defined Chiral PNP Pincer Ligand". In: *Angewandte Chemie-International Edition* 129.37 (2017), pp. 11389–11393. DOI: 10.1002/anie.201705471.
- [20] F. Kallmeier and R. Kempe. "Manganese complexes for (de) hydrogenation catalysis: a comparison to cobalt and iron catalysts". In: *Angewandte Chemie-International Edition* 57.1 (2018), pp. 46–60. DOI: 10.1002/anie.201709010.

- [21] J. R. Carney, B. R. Dillon, and S. P. Thomas. "Recent Advances of Manganese Catalysis for Organic Synthesis". In: *European Journal of Organic Chemistry* 23 (2016), pp. 3912–3929. DOI: 10.1002/ejoc.201600018.
- [22] B. Maji and M. K. Barman. "Recent Developments of Manganese Complexes for Catalytic Hydrogenation and Dehydrogenation Reactions". In: *Synthesis* 49.15 (2017), pp. 3377–3393. DOI: 10.1055/s-0036-1590818.
- [23] E. A. Bielinski, P. O. Lagaditis, Y. Zhang, B. Q. Mercado, C. Würtele, W. H. Bernskoetter, N. Hazari, and S. Schneider. "Lewis acid-assisted formic acid dehydrogenation using a pincer-supported iron catalyst". In: *Journal of the American Chemical Society* 136.29 (2014), pp. 10234–10237. DOI: 10.1021/ja505241x.
- [24] S. Schneider, J. Meiners, and B. Askevold. "Cooperative aliphatic PNP amido pincer ligands—versatile building blocks for coordination chemistry and catalysis". In: *European Journal of Inorganic Chemistry* 2012.3 (2012), pp. 412–429. DOI: 10.1002/ejic.201100880.
- [25] P. Friederich, G. dos Passos Gomes, R. De Bin, A. Aspuru-Guzik, and D. Balcells. "Machine learning dihydrogen activation in the chemical space surrounding Vaska's complex". In: *Chemical Science* 11 (18 2020), pp. 4584–4601. DOI: 10.1039/D0SC00445F.
- [26] J. P. Janet, F. Liu, A. Nandy, C. Duan, T. Yang, S. Lin, and H. J. Kulik. "Designing in the Face of Uncertainty: Exploiting Electronic Structure and Machine Learning Models for Discovery in Inorganic Chemistry". In: *Inorganic Chemistry* 58.16 (2019), pp. 10592–10606. DOI: 10.1021/acs.inorgchem.9b00109.
- [27] M. D. Wodrich, B. Sawatlon, E. Solel, S. Kozuch, and C. Corminboeuf. "Activity-based screening of homogeneous catalysts through the rapid assessment of theoretically derived turnover frequencies". In: *ACS Catalysis* 9.6 (2019), pp. 5716–5725. DOI: 10.1021/acscatal.9b00717.
- [28] B. Sawatlon, M. D. Wodrich, and C. Corminboeuf. "Unraveling metal/pincer ligand effects in the catalytic hydrogenation of carbon dioxide to formate". In: *Organometallics* 37.24 (2018), pp. 4568–4575. DOI: 10.1021/acs.organomet.8b00490.
- [29] M. D. Wodrich, M. Busch, and C. Corminboeuf. "Expedited screening of active and regioselective catalysts for the hydroformylation reaction". In: *Helvetica Chimica Acta* 101.9 (2018), e1800107. DOI: 10.1002/hlca.201800107.
- [30] A. M. Krieger, V. Sinha, A. V. Kalikadien, and E. A. Pidko. "Metal-ligand cooperative activation of HX (X=H, Br, OR) bond on Mn based pincer complexes". In: *Journal of Inorganic and General Chemistry* 647.14 (2021), pp. 1486–1494. DOI: 10.1002/zaac.202100078.
- [31] A. V. Kalikadien, E. A. Pidko, and V. Sinha. "ChemSpaX: exploration of chemical space by automated functionalization of molecular scaffold". In: *Digital Discovery* 1 (1 2022), pp. 8–25. DOI: 10.1039/D1DD00017A.

- [32] A. Hashemi, S. Bougueroua, M.-P. Gaigeot, and E. A. Pidko. “ReNeGate: A Reaction Network Graph-Theoretical Tool for Automated Mechanistic Studies in Computational Homogeneous Catalysis”. In: *Journal of Chemical Theory and Computation* 18.12 (2022), pp. 7470–7482. DOI: 10.1021/acs.jctc.2c00404.
- [33] F. Schneck, M. Finger, M. Tromp, and S. Schneider. “Chemical Non-Innocence of an Aliphatic PNP Pincer Ligand”. In: *Chemistry – A European Journal* 23.1 (2016), pp. 33–37. DOI: 10.1002/chem.201604407.
- [34] J. Zhang, E. Balaraman, G. Leitus, and D. Milstein. “Electron-Rich PNP- and PNN-Type Ruthenium(II) Hydrido Borohydride Pincer Complexes. Synthesis, Structure, and Catalytic Dehydrogenation of Alcohols and Hydrogenation of Esters”. In: *Organometallics* 30.21 (2011), pp. 5716–5724. DOI: 10.1021/om200595m.
- [35] N. Biswas, R. Sharma, and D. Srimani. “Ruthenium Pincer Complex Catalyzed Selective Synthesis of C-3 Alkylated Indoles and Bisindolylmethanes Directly from Indoles and Alcohols”. In: *Advanced Synthesis & Catalysis* 362.14 (2020), pp. 2902–2910. DOI: 10.1002/adsc.202000326.
- [36] J. Schörghener, A. Zimmermann, and M. Waser. “SNS-Ligands for Ru-Catalyzed Homogeneous Hydrogenation and Dehydrogenation Reactions”. In: *Organic Process Research & Development* 22.7 (2018), pp. 862–870. DOI: 10.1021/acs.oprd.8b00142.
- [37] X. Chen, Y. Jing, and X. Yang. “Unexpected Direct Hydride Transfer Mechanism for the Hydrogenation of Ethyl Acetate to Ethanol Catalyzed by SNS Pincer Ruthenium Complexes”. In: *Chemistry – A European Journal* 22.6 (2016), pp. 1950–1957. DOI: 10.1002/chem.201504058.
- [38] D. Spasyuk, S. Smith, and D. G. Gusev. “Replacing Phosphorus with Sulfur for the Efficient Hydrogenation of Esters”. In: *Angewandte Chemie-International Edition* 52.9 (2013), pp. 2538–2542. DOI: 10.1002/anie.201209218.
- [39] T.-a. Koizumi, T. Teratani, K. Okamoto, T. Yamamoto, Y. Shimoi, and T. Kanbara. “Nickel(II) complexes bearing a pincer ligand containing thioamide units: Comparison between SNS- and SCS-pincer ligands”. In: *Inorganica Chimica Acta* 363.11 (2010), pp. 2474–2480. DOI: 10.1016/j.ica.2010.04.012.
- [40] X. Wu, L. Ji, Y. Ji, E. H. M. Elageed, and G. Gao. “Hydrogenation of ethylene carbonate catalyzed by lutidine-bridged N-heterocyclic carbene ligands and ruthenium precursors”. In: *Catalysis Communications* 85.NA (2016), pp. 57–60. DOI: 10.1016/j.catcom.2016.07.015.
- [41] M. Hernández-Juárez, J. López-Serrano, P. Lara, J. P. Morales-Cerón, M. Vaquero, E. Álvarez, V. Salazar, and A. Suárez. “Ruthenium(II) Complexes Containing Lutidine-Derived Pincer CNC Ligands: Synthesis, Structure, and Catalytic Hydrogenation of C-N bonds”. In: *Chemistry – A European Journal* 21.20 (2015), pp. 7540–7555. DOI: 10.1002/chem.201406040.

- [42] G. A. Filonenko, E. Cosimi, L. Lefort, M. P. Conley, C. Copéret, M. Lutz, E. J. M. Hensen, and E. A. Pidko. "Lutidine-Derived Ru-CNC Hydrogenation Pincer Catalysts with Versatile Coordination Properties". In: *ACS Catalysis* 4.8 (2014), pp. 2667–2671. DOI: 10.1021/cs500720y.
- [43] A. R. Naziruddin, C.-L. Kuo, W.-J. Lin, W.-H. Lo, C.-S. Lee, B.-J. Sun, A. H. H. Chang, and W.-S. Hwang. "Ruthenium Complexes Bearing Unsymmetric CNC Pincer Ligands: Molecular Structures and Electronic Properties". In: *Organometallics* 33.10 (2014), pp. 2575–2582. DOI: 10.1021/om500205p.
- [44] A. A. Danopoulos, N. Tsoureas, J. A. Wright, and M. E. Light. "N-Heterocyclic Pincer Dicarbene Complexes of Iron(II): C-2 and C-5 Metalated Carbenes on the Same Metal Center". In: *Organometallics* 23.2 (2003), pp. 166–168. DOI: 10.1021/om0341911.
- [45] M. Gagliardo, P. A. Chase, S. Brouwer, G. P. M. van Klink, and G. van Koten. "Electronic Effects in PCP-Pincer Ru(II)-Based Hydrogen Transfer Catalysis". In: *Organometallics* 26.9 (2007), pp. 2219–2227. DOI: 10.1021/om060874f.
- [46] B. C. Gruver, J. J. Adams, S. J. Warner, N. Arulsamy, and D. M. Roddick. "Acceptor Pincer Chemistry of Ruthenium: Catalytic Alkane Dehydrogenation by (CF<sub>3</sub>PCP)Ru(cod)(H)". In: *Organometallics* 30.19 (2011), pp. 5133–5140. DOI: 10.1021/om200354y.
- [47] S. Tang, N. von Wolff, Y. Diskin-Posner, G. Leitus, Y. Ben-David, and D. Milstein. "Pyridine-Based PCP-Ruthenium Complexes: Unusual Structures and Metal–Ligand Cooperation". In: *Journal of the American Chemical Society* 141.18 (2019), pp. 7554–7561. DOI: 10.1021/jacs.9b02669.
- [48] J. J. Adams, N. Arulsamy, and D. M. Roddick. "Acceptor PCP Pincer Iridium(I) Chemistry: Stabilization of Nonmeridional PCP Coordination Geometries". In: *Organometallics* 30.4 (2011), pp. 697–711. DOI: 10.1021/om100864g.
- [49] S. Murugesan and K. Kirchner. "Non-precious metal complexes with an anionic PCP pincer architecture". In: *Dalton transactions* 45.2 (2016), pp. 416–439. DOI: 10.1039/c5dt03778f.
- [50] E. A. Bielinski, P. O. Lagaditis, Y. Zhang, B. Q. Mercado, C. Würtele, W. H. Bernskoetter, N. Hazari, and S. Schneider. "Lewis Acid-Assisted Formic Acid Dehydrogenation Using a Pincer-Supported Iron Catalyst". In: *Journal of the American Chemical Society* 136.29 (2014), pp. 10234–10237. DOI: 10.1021/ja505241x.
- [51] P. A. Dub and J. C. Gordon. "The role of the metal-bound N–H functionality in Noyori-type molecular catalysts". In: *Nature Reviews Chemistry* 2.12 (2018), pp. 396–408. DOI: 10.1038/s41570-018-0049-z.
- [52] C. Liu, R. van Putten, P. O. Kulyaev, G. A. Filonenko, and E. A. Pidko. "Computational insights into the catalytic role of the base promoters in ester hydrogenation with homogeneous non-pincer-based Mn-PN catalyst". In: *Journal of Catalysis* 363.NA (2018), pp. 136–143. DOI: 10.1016/j.jcat.2018.04.018.

- [53] K. Larmier, W.-C. Liao, S. Tada, E. Lam, R. Verel, A. Bansode, A. Urakawa, A. Comas-Vives, and C. Copéret. “CO<sub>2</sub>-to-Methanol Hydrogenation on Zirconia-Supported Copper Nanoparticles: Reaction Intermediates and the Role of the Metal-Support Interface”. In: *Angewandte Chemie-International Edition* 56.9 (2017), pp. 2318–2323. DOI: 10.1002/anie.201610166.
- [54] W.-H. Wang, Y. Himeda, J. T. Muckerman, G. F. Manbeck, and E. Fujita. “CO<sub>2</sub> Hydrogenation to Formate and Methanol as an Alternative to Photo- and Electrochemical CO<sub>2</sub> Reduction”. In: *Chemical Reviews* 115.23 (2015), pp. 12936–12973. DOI: 10.1021/acs.chemrev.5b00197.
- [55] Y. Haiyan, C. Zhang, P. Gao, H. Wang, X. Li, L. Zhong, W. Wei, and Y. Sun. “A review of the catalytic hydrogenation of carbon dioxide into value-added hydrocarbons”. In: *Catalysis Science & Technology* 7.20 (2017), pp. 4580–4598. DOI: 10.1039/c7cy01403a.
- [56] G. A. Filonenko, R. van Putten, E. J. Hensen, and E. A. Pidko. “Catalytic (de)hydrogenation promoted by non-precious metals - Co, Fe and Mn”. In: *Chemical Society Reviews* 47.4 (2018), p. 1459. DOI: 10.1039/C7CS00334J.
- [57] M. Andérez-Fernández, L. K. Vogt, S. Fischer, W. Zhou, H. Jiao, M. Garbe, S. Elangovan, K. Junge, H. Junge, R. Ludwig, and M. Beller. “A Stable Manganese Pincer Catalyst for the Selective Dehydrogenation of Methanol”. In: *Angewandte Chemie-International Edition* 56.2 (2016), pp. 559–562. DOI: 10.1002/anie.201610182.
- [58] C. J. A. Daley and S. H. Bergens. “The first complete identification of a diastereomeric catalyst-substrate (alkoxide) species in an enantioselective ketone hydrogenation. Mechanistic investigations”. In: *Journal of the American Chemical Society* 124.14 (2002), pp. 3680–3691. DOI: 10.1021/ja0102991.
- [59] R. J. Hamilton and S. H. Bergens. “An unexpected possible role of base in asymmetric catalytic hydrogenations of ketones. Synthesis and characterization of several key catalytic intermediates”. In: *Journal of the American Chemical Society* 128.42 (2006), pp. 13700–13701. DOI: 10.1021/ja065460s.
- [60] R. J. Hamilton and S. H. Bergens. “Direct Observations of the Metal-Ligand Bifunctional Addition Step in an Enantioselective Ketone Hydrogenation”. In: *Journal of the American Chemical Society* 130.36 (2008), pp. 11979–11987. DOI: 10.1021/ja8034812.
- [61] M. Z.-D. Iuliis and R. H. Morris. “Kinetic Hydrogen/Deuterium Effects in the Direct Hydrogenation of Ketones Catalyzed by a Well-Defined Ruthenium Diphosphine Diamine Complex”. In: *Journal of the American Chemical Society* 131.31 (2009), pp. 11263–11269. DOI: 10.1021/ja9043104.
- [62] A. Passera and A. Mezzetti. “Mn(I) and Fe(II)/PN(H)P Catalysts for the Hydrogenation of Ketones: A Comparison by Experiment and Calculation”. In: *Advanced Synthesis & Catalysis* 361.20 (2019), pp. 4691–4706. DOI: 10.1002/adsc.201900671.

- [63] P. Pracht, F. Bohle, and S. Grimme. “Automated exploration of the low-energy chemical space with fast quantum chemical methods”. In: *Physical Chemistry Chemical Physics* 22.14 (2020), pp. 7169–7192. DOI: 10.1039/C9CP06869D.
- [64] V. Sinha, J. J. Laan, and E. A. Pidko. “Accurate and rapid prediction of  $pK_a$  of transition metal complexes: semiempirical quantum chemistry with a data-augmented approach”. In: *Physical Chemistry Chemical Physics* 23.4 (2021), pp. 2557–2567. DOI: 10.1039/D0CP05281G.
- [65] S. Bougueroua, R. Spezia, S. Pezzotti, S. Vial, F. Quessette, D. Barth, and M. P. Gaigeot. “Graph theory for automatic structural recognition in molecular dynamics simulations”. In: *The Journal of Chemical Physics* 149.18 (2018), p. 184102. DOI: 10.1063/1.5045818.
- [66] S. Grimme, S. Ehrlich, and L. Goerigk. “Effect of the damping function in dispersion corrected density functional theory”. In: *Journal of Computational Chemistry* 32.7 (2011), pp. 1456–1465. DOI: 10.1002/jcc.21759.
- [67] A. V. Marenich, C. J. Cramer, and D. G. Truhlar. “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”. In: *The Journal of Physical Chemistry. B* 113.18 (2009), pp. 6378–6396. DOI: 10.1021/jp810292n.
- [68] N. J. d. Vos. “kmodes categorical clustering library”. In: (2015). URL: <https://github.com/nicodv/kmodes>.
- [69] P. Braunstein and F. Naud. “Hemilability of hybrid ligands and the coordination chemistry of oxazoline-based systems”. In: *Angewandte Chemie-International Edition* 40.4 (2001), pp. 680–699. DOI: 10.1002/1521-3773(20010216)40:4<680::AID-ANIE6800>3.0.CO;2-0.
- [70] C. S. Slone, D. A. Weinberger, and C. A. Mirkin. “The transition metal coordination chemistry of hemilabile ligands”. In: *Progress in Inorganic Chemistry* 48 (1999), pp. 233–350. DOI: 10.1002/9780470166499.ch3.
- [71] R. van Putten, G. A. Filonenko, A. M. Krieger, M. Lutz, and E. A. Pidko. “Manganese-Mediated C–C Bond Formation: Alkoxyacylation of Organoboranes”. In: *Organometallics* 40.6 (2021), pp. 674–681. DOI: 10.1021/acs.organomet.0c00781.
- [72] R. van Putten. “Catalysis, chemistry, and automation: Addressing complexity to explore practical limits of homogeneous Mn catalysis”. English. PhD thesis. Delft University of Technology, 2021. ISBN: 978-94-6384-246-4. DOI: 10.4233/uuid:e94a4c47-b3f7-45a7-a38c-9c319c38a51f.
- [73] J. Ruiz, B. F. Perandones, J. F. Van der Maelen, and S. García-Granda. “On the existence of an N-metalated N-heterocyclic carbene: a theoretical study”. In: *Organometallics* 29.20 (2010), pp. 4639–4642. DOI: 10.1021/om100469m.

- [74] H. G. Raubenheimer, Y. Stander, E. K. Marais, C. Thompson, G. J. Kruger, S. Cronje, and M. Deetlefs. "Group 6 carbene complexes derived from lithiated azoles and the crystal structure of a molybdenum thiazolynylidene complex". In: *Journal of Organometallic Chemistry* 590.2 (1999), pp. 158–168. DOI: 10.1016/S0022-328X(99)00445-3.
- [75] M. A. Huertos, J. Pérez, and L. Riera. "Pyridine ring opening at room temperature at a rhenium tricarbonyl bipyridine complex". In: *Journal of the American Chemical Society* 130.17 (2008), pp. 5662–5663. DOI: 10.1021/ja801655d.
- [76] A. S. Veige. "Carbon monoxide as a reagent: A report on the role of N-heterocyclic carbene (NHC) ligands in metal-catalyzed carbonylation reactions". In: *Polyhedron* 27.15 (2008), pp. 3177–3189. DOI: 10.1016/j.poly.2008.07.019.
- [77] J.-B. Peng, H.-Q. Geng, and X.-F. Wu. "The Chemistry of CO: Carbonylation". In: *Chem* 5.3 (2019), pp. 526–552. DOI: 10.1016/j.chempr.2018.11.006.
- [78] D. Steinborn. *Fundamentals of organometallic catalysis*. John Wiley & Sons, 2011. ISBN: 3527327169.
- [79] A. L. Daniels, J. G. Da Gama, R. Edjoc, B. M. Gabidullin, and R. T. Baker. "Synthesis and Reactivity of Mn-CF<sub>3</sub> Complexes". In: *Inorganics* 7.1 (2019), p. 3. DOI: 10.3390/inorganics7010003.
- [80] I. Ojima. *Fluorine in medicinal chemistry and chemical biology*. John Wiley & Sons, 2009. ISBN: 0470744081. DOI: 10.1002/9781444312096.
- [81] P. Kirsch. *Modern fluoroorganic chemistry: synthesis, reactivity, applications*. John Wiley & Sons, 2013. ISBN: 3527651373. DOI: 10.1002/352760393X.
- [82] A. I. Konovalov, A. Lishchynskiy, and V. V. Grushin. "Mechanism of trifluoromethylation of aryl halides with CuCF<sub>3</sub> and the ortho effect". In: *Journal of the American Chemical Society* 136.38 (2014), pp. 13410–13425. DOI: 10.1021/ja507564p.
- [83] G. G. Dubinina, H. Furutachi, and D. A. Vicić. "Active trifluoromethylating agents from well-defined copper (I)-CF<sub>3</sub> complexes". In: *Journal of the American Chemical Society* 130.27 (2008), pp. 8600–8601. DOI: 10.1021/ja802946s.
- [84] D. M. Wiemers and D. J. Burton. "Pregeneration, spectroscopic detection and chemical reactivity of (trifluoromethyl) copper, an elusive and complex species". In: *Journal of the American Chemical Society* 108.4 (1986), pp. 832–834. DOI: 10.1021/ja00264a043.
- [85] V. McLoughlin and J. Thrower. "A route to fluoroalkyl-substituted aromatic compounds involving fluoroalkylcopper intermediates". In: *Tetrahedron* 25.24 (1969), pp. 5921–5940. DOI: 10.1016/S0040-4020(01)83100-8.
- [86] W. J. Choi, S. Choi, K. Ohkubo, S. Fukuzumi, E. J. Cho, and Y. You. "Mechanisms and applications of cyclometalated Pt(II) complexes in photoredox catalytic trifluoromethylation". In: *Chemical Science* 6.2 (2015), pp. 1454–1464. DOI: 10.1039/c4sc02537g.

- [87] D. A. Nagib and D. W. C. Macmillan. “Trifluoromethylation of arenes and heteroarenes by means of photoredox catalysis”. In: *Nature* 480.7376 (2011), pp. 224–228. DOI: 10.1038/nature10647.
- [88] R. P. Hughes. “Organo-transition metal compounds containing perfluorinated ligands”. In: *Advances in organometallic chemistry*. Vol. 31. Elsevier, 1990, pp. 183–267. ISBN: 0065-3055. DOI: 10.1016/S0065-3055(08)60511-0.
- [89] R. Morales-Cerrada, C. Fliegel, J.-C. Daran, F. Gayet, V. Ladmiral, B. Ameduri, and R. Poli. “Fluoroalkyl radical generation by homolytic bond dissociation in pentacarbonylmanganese derivatives”. In: *Chemistry—A European Journal* 25.1 (2019), pp. 296–308. DOI: 10.1002/chem.201804007.



# 5

## SUMMARY AND OUTLOOK

CATALYTIC systems are commonly represented by complex mixtures of reactants, catalyst precursors, ligands, additives, and solvent that may give rise to the formation of a wide variety of species that may show varied catalytic activity and behavior towards other components of the reaction mixture. The primary tasks of computational catalysis are to identify among these pre-reaction complexes those that contribute most to the catalytic reaction and identify mechanisms of the main catalytic cycle and competing reaction channels giving rise to unselective conversion routes or catalyst deactivation. As we will demonstrate throughout this thesis, the development of a comprehensive molecular-level picture of a catalytic system is a very challenging task due to the enormous complexity of the associated chemical reaction. However, I firmly believe that if all these tasks are accomplished, the resulting reactivity model could be used to guide the development of more active and efficient catalysts, which is the core idea of the rational catalyst design strategy.

Electronic structure calculations have become an indispensable tool in catalysis research. They are currently routinely employed to rationalize experimental observations, support mechanistic proposals and even to guide the design of new catalytic systems. Nevertheless, the vast majority of computational studies in catalysis by transition metals still has an explanatory character and focus on describing only a small part of the actual catalyst system. The transition to truly predictive computational modeling requires the development of more complex chemical models that would allow an adequate description of the full reaction networks underlying the catalytic processes. Those approaches can be combined with the topics discussed in this thesis (novel electronic structure theory methods, complex modeling, computation of multiple reaction pathways, reaction network analysis) for building a comprehensive theoretical framework for computational catalysis. The elimination of expert bias in mechanistic studies is one of the important targets and here the emergence of automated methods for reaction network analysis is particularly exciting and holds a great promise of delivering a paradigm shift in catalysis research. Particularly important is the extension of mechanistic studies from the descriptions of catalytic cycles to other reaction paths resulting in selectivity losses and catalyst deactivation as these processes actually determine the efficiency of the catalysts and their durability.

Similarly, the adequate description of the complex chemistry of 3d transition metals

necessitates the introduction of newer, more accurate, fast and expert-bias free methodologies suitable for dealing with the multiconfigurational effects in such catalytic systems. The development of methods to understand reactivity as defined by complicated reaction mechanisms in practical 3d metal-based catalysts would provide one more gear to optimize activity and selectivity. Additionally, there is a need for accurate automated reaction discovery tools; however, this necessitates fast and reasonably accurate electronic structure calculations to make them applicable to realistic catalytic systems. The development of increasingly accurate and efficient reaction discovery tools and electronic structure methods with broader applicability will aid in the development of practical, earth abundant metal catalysts for many vital reactions.

An alternative promising approach is the application of machine learning for the fast exploration of the chemical space. Data-driven methodologies are expected to accelerate the identification of key properties that can be used as descriptors of the catalytic and contribute to computational catalyst design.[1–3] Application of such bias-free mechanistic studies to high throughput virtual screening of comprehensive libraries of catalysts will provide data-driven approaches for the rational design of catalysts with improved functions.

This thesis deals with the development of automated catalyst screening methodologies based on a combination of accelerated semiempirical methods to explore the chemical space and ab initio quantum chemical approaches to refine the mechanistic and reactivity insights obtained at the exploration stage. Special focus is laid on identifying reactivity channels resulting in the long-term catalyst deactivation. The key hypothesis here is that the deactivation is a thermodynamically favorable phenomenon allowing thus to reduce the problem of the automated mechanistic analysis to exploring the thermodynamically favorable conversion channels only. An important goal of this work was to arrive at the reactivity exploration methodology that could be combined with a high-throughput screening of extended 3d transition metal catalyst spaces. In **Chapter 1**, I present an overview of the state of the art in the automated mechanistic studies with an emphasis on their application to catalysis by transition metals. These studies have guided us in identifying methodological requirements for the development of the more robust methods suitable. Following up in **Chapter 2**, we have done a comparative study between our developed ReNeGate methodology and a representative method we identified as the most efficient in the application to transition metal catalysis. Reaction networks were built starting from similar starting points and results were compared both in comprehensiveness of the results as well as efficiency and scalability of the methodologies with respect to our specific goals of discovering new paths relevant for the catalyst deactivation chemistry. **Chapter 3** summarizes the detailed explanations of our ReNeGate methodology its implementation and capabilities. The development of the method is accompanied by illustrative case studies relevant to catalytic reduction by homogeneous Mn(I) catalysts. Specifically, 3 case studies are discussed, starting with the analysis of the conversion paths of the common  $\text{Mn}(\text{CO})_5\text{Br}$  precursor in the presence of an alkoxide base activator, followed by a more detailed analysis of the possible reaction paths between the activator and a more relevant Mn(I)-based catalyst bearing a representative diamine (NN) ligand. We next extended our analysis to the investigation of multinuclear catalytic species by considering the transformations of a model system

containing multiple Mn(I) complexes. The last case study was used to introduce a specifically designed fragment analysis tool capable of extracting structural information and identifying unique catalyst fragments upon analysis of complex catalytic systems. Our analysis revealed new chemical conversion paths and inspired the experimental studies that validated our experimental predictions. In the next **Chapter 4**, we have extended the application of ReNeGate methodology to studies on large datasets of catalytic structures. We combined an automated workflow for catalyst modification ChemSpaX to generate an in silico database of hypothetical Mn(I) pincer catalysts with the ReNeGate reactivity exploration workflow. We introduced a workflow for the automated analysis of the very large reactivity datasets to identify new reactivity patterns and correlate the structural features with the new reactivity predictions. This chapter is finalized by a detailed discussion of new catalytic insights resulting from this high throughput study. The workflows presented in this thesis can be utilized for exploring various catalytic systems well beyond the limited set of Mn(I) catalysts discussed herein. On the other hand, the insights into the deactivation channels and catalytic chemistry obtained by using the developed workflow could not be obtained by using the conventional expert-driven exploration approaches.

## 5.1. SAMENVATTING

**K**ATALYTISCHE systemen worden gewoonlijk voorgesteld door complexe mengsels van reactanten, katalysatorprecursors, liganden, additieven en oplosmiddelen die aanleiding kunnen geven tot de vorming van een grote verscheidenheid aan soorten die een gevarieerde katalytische activiteit en gedrag ten opzichte van andere componenten van het reactiemengsel kunnen vertonen. De primaire taken van computationele katalyse zijn het identificeren van deze pre-reactiecomplexen die het meest bijdragen aan de katalytische reactie en het identificeren van mechanismen van de belangrijkste katalytische cyclus en concurrerende reactiekanalen die aanleiding geven tot niet-selectieve conversieroutes of katalysatordeactivering. Zoals we in dit proefschrift zullen aantonen, is de ontwikkeling van een alomvattend beeld op moleculair niveau van een katalytisch systeem een zeer uitdagende taak vanwege de enorme complexiteit van de bijbehorende chemische reactie. Ik ben er echter vast van overtuigd dat als al deze taken zijn volbracht, het resulterende activiteitsmodel kan worden gebruikt om de ontwikkeling van actievere en efficiëntere katalysatoren te begeleiden, wat het kernidee is van de strategie voor rationeel katalysatorontwerp.

Elektronische structuurberekeningen zijn een onmisbaar hulpmiddel geworden in het katalyseonderzoek. Ze worden momenteel routinematig gebruikt om experimentele waarnemingen te rationaliseren, mechanistische voorstellen te ondersteunen en zelfs om het ontwerp van nieuwe katalytische systemen te begeleiden. Desalniettemin heeft de overgrote meerderheid van computationele studies naar katalyse door overgangsmetalen nog steeds een verklarend karakter en richten ze zich op het beschrijven van slechts een klein deel van het eigenlijke katalysatorsysteem. De overgang naar echt voorspellende computationele modellering vereist de ontwikkeling van complexere chemische modellen die een adequate beschrijving mogelijk maken van de volledige reactienetwerken die ten grondslag liggen aan de katalytische processen. Deze benaderingen kunnen worden gecombineerd met de onderwerpen die in dit proefschrift worden besproken (nieuwe methoden voor de theorie van elektronische structuren, complexe modellering, berekening van meerdere reactiepaden, analyse van reactienetwerken) voor het bouwen van een alomvattend theoretisch raamwerk voor computationele katalyse.

De eliminatie van expertbias in mechanistische studies is een van de belangrijke doelen en hier is de opkomst van geautomatiseerde methoden voor reactienetwerkanalyse bijzonder opwindend en houdt een grote belofte in voor een paradigmaverschuiving in katalyseonderzoek. Bijzonder belangrijk is de uitbreiding van mechanistische studies van de beschrijvingen van katalytische cycli naar andere reactiepaden die resulteren in selectiviteitsverliezen en deactivering van de katalysator, aangezien deze processen feitelijk de efficiëntie van de katalysatoren en hun duurzaamheid bepalen.

Evenzo vereist de adequate beschrijving van de complexe chemie van 3d overgangsmetalen de introductie van nieuwere, nauwkeurigere, snellere en expert-bias-vrije methodologieën die geschikt zijn voor het omgaan met de multiconfiguratie-effecten in dergelijke katalytische systemen. De ontwikkeling van methoden om reactiviteit te begrijpen, zoals gedefinieerd door gecompliceerde reactiemechanismen in praktische 3d metaalgebaseerde katalysatoren, zou een extra versnelling bieden om activiteit en selectiviteit te optimaliseren. Bovendien is er behoefte aan nauwkeurige geautomatiseerde hulpmiddelen voor het ontdekken van reacties; dit vereist echter snelle en redelijk nauw-

keurige berekeningen van de elektronische structuur om ze toepasbaar te maken voor realistische katalytische systemen. De ontwikkeling van steeds nauwkeurigere en efficiëntere hulpmiddelen voor het ontdekken van reacties en methoden voor elektronische structuur met een bredere toepasbaarheid zal helpen bij de ontwikkeling van praktische, in de aarde overvloedige metaalkatalysatoren voor veel vitale reacties.

Een alternatieve, veelbelovende benadering is de toepassing van machine learning voor het snel verkennen van de chemische ruimte. Van gegevensgestuurde methodologieën wordt verwacht dat ze de identificatie versnellen van sleuteleigenschappen die kunnen worden gebruikt als descriptoren van de katalysator en bijdragen aan het ontwerpen van computationele katalysatoren. [1–3] Toepassing van dergelijke bias-vrije mechanistische studies op virtuele screening met hoge doorvoer van uitgebreide bibliotheken van katalysatoren zal gegevensgestuurde benaderingen opleveren voor het rationele ontwerp van katalysatoren met verbeterde functies.

Dit proefschrift behandelt de ontwikkeling van geautomatiseerde methodologieën voor het screenen van katalysatoren, gebaseerd op een combinatie van versnelde semi empirische methoden om de chemische ruimte te verkennen en ab initio kwantumchemische benaderingen om de mechanistische en reactiviteitsinzichten verkregen in de verkenningfase te verfijnen. Speciale aandacht wordt besteed aan het identificeren van reactiviteitskanalen die resulteren in de deactivering van de katalysator op lange termijn. De sleutelhypothese hier is dat de deactivering een thermodynamisch gunstig fenomeen is, waardoor het probleem van de geautomatiseerde mechanistische analyse dus kan worden teruggebracht tot het verkennen van alleen de thermodynamisch gunstige conversiekanalen. Een belangrijk doel van dit werk was om te komen tot de reactiviteitsverkenningmethodiek die gecombineerd zou kunnen worden met een high-throughput screening van uitgebreide 3d overgangsmetaalkatalysatorruimten. In **Hoofdstuk 1** presenteer ik een overzicht van de stand van zaken in de geautomatiseerde mechanistische studies met de nadruk op hun toepassing op katalyse door overgangsmetalen. Deze studies hebben ons geleid bij het identificeren van methodologische vereisten voor de ontwikkeling van de meer robuuste methoden die geschikt zijn. Als vervolg op **hoofdstuk 2** hebben we een vergelijkende studie gedaan tussen onze ontwikkelde ReNeGate-methodiek en een representatieve methode die we hebben geïdentificeerd als de meest efficiënte in de toepassing op overgangsmetaalkatalyse. Er werden reactienetwerken gebouwd uitgaande van vergelijkbare uitgangspunten en de resultaten werden vergeleken, zowel wat betreft volledigheid van de resultaten als efficiëntie en schaalbaarheid van de methodologieën met betrekking tot onze specifieke doelen om nieuwe paden te ontdekken die relevant zijn voor de katalysatordeactiveringschemie. **Hoofdstuk 3** geeft een samenvatting van de gedetailleerde uitleg van onze ReNeGate-methodologie, de implementatie en mogelijkheden ervan. De ontwikkeling van de methode gaat gepaard met illustratieve case-study's die relevant zijn voor katalytische reductie door homogene Mn(I)-katalysatoren. Specifiek worden 3 casestudies besproken, beginnend met de analyse van de conversiepaden van de gemeenschappelijke Mn(CO)<sub>5</sub>Br precursor in aanwezigheid van een alkoxidebase-activator, gevolgd door een meer gedetailleerde analyse van de mogelijke reactiepaden tussen de activator en een relevantere op Mn(I) gebaseerde katalysator met een representatieve diamine (NN)-ligand. Vervolgens hebben we onze analyse uitgebreid naar het onderzoek van meerkernige katalytische soorten door de transformaties

van een modelsysteem met meerdere Mn (I) -complexen te beschouwen. De laatste case study werd gebruikt om een specifiek ontworpen tool voor fragmentanalyse te introduceren die structurele informatie kan extraheren en unieke katalysatorfragmenten kan identificeren bij analyse van complexe katalytische systemen. Onze analyse omhulde nieuwe chemische conversiepaden en inspireerde de experimentele studies die onze experimentele voorspellingen valideerden. In het volgende **hoofdstuk 4** hebben we de toepassing van de ReNeGate-methode uitgebreid naar studies over grote datasets van katalytische structuren. We combineerden een geautomatiseerde workflow voor katalysatormodificatie ChemSpaX om een in silico-database van hypothetische Mn(I)-tangkatalysatoren te genereren met de ReNeGate-workflow voor reactiviteitsonderzoek. We hebben een workflow geïntroduceerd voor de geautomatiseerde analyse van de zeer grote reactiviteitsdatasets om nieuwe reactiviteitspatronen te identificeren en de structurele kenmerken te correleren met de nieuwe reactiviteitsvoorspellingen. Dit hoofdstuk wordt afgesloten met een gedetailleerde bespreking van nieuwe katalytische inzichten die voortkomen uit deze high throughput studie. De workflows die in dit proefschrift worden gepresenteerd, kunnen worden gebruikt voor het verkennen van verschillende katalytische systemen die veel verder gaan dan de beperkte set Mn(I)-katalysatoren die hierin worden besproken. Aan de andere kant konden de inzichten in de deactiveringskanalen en katalytische chemie verkregen door gebruik te maken van de ontwikkelde workflow niet worden verkregen door gebruik te maken van de conventionele door experts aangestuurde exploratiebenaderingen.

# 6

## CONCLUSION

**I**N this thesis, I highlighted a number of projects aimed at developing and testing new simulation methods for studying complex reactive systems, with a particular emphasis on simulation strategies based on the concept of bonding graphs. These mathematical structures provide useful tools for a variety of algorithms developed over the last few decades. Through automated analysis of exhaustive exploration trajectories, I have been able to capture serendipities that could escape the expert heuristics or otherwise needed expertise in different disciplines to be interpreted correctly. Such discoveries could range from very obvious one-step reactions that were just not “normally” considered to multi-step complex reactions that were not imaginable to the expert. With automated reactivity screenings on *in silico* catalyst libraries, we have taken a big step towards “rational” catalyst design.

Increasingly affordable high-performance computing hardware to enable *ab initio* electronic structure calculations as well as the advancement of AI/ML methods, have resulted in new opportunities for automatic reaction discovery methods in the last decade. Such automatic reaction discovery schemes are becoming more widely available for studying complex chemical reactions. It should, however, be considered that machine learning approaches towards *in silico* catalyst research and development carry the inaccuracies of the *ab initio* methods used during the training. With the new understandings from this thesis, we now know that we might have only touched the tip of the iceberg for in depth understanding of the catalytic conversions and there is room for applying the automated explorations to this classic fields. While the routine catalysis research efforts are implicitly dedicated to optimizing the “known” or “assumed” conditions/mechanisms, exploring and understanding the “unknown” (deactivation) mechanisms should be regarded as one of the primary challenge for the future.

From the development point of view, the pipelines introduced in this thesis can foster for more reliable and fast methods for unbiased exploration of the chemical space in the future. In the application of chemical reaction networks for understanding catalytic behaviour, there should be measures to distinguish when our autogenerated chemical reaction networks are satisfactorily complete and accurate to answer the specific physical or chemical questions. In other words, we should know “when to stop” the exploration of the vast catalytic space. With the advent of efficient and reliable methods for exhaustively exploring different options, we are able to produce big data and should be ready to

address challenges in providing guidance or even replacing the expert via the automate analysis of big data from (virtual) high throughput screening.

# A

## SUPPLEMENTARY INFORMATION CHAPTER 3

### A.1. RELEVANT MOLECULAR GRAPH THEORY TERMINOLOGY

- **Graphs:** A graph  $G$  is defined as  $G = (V, E)$ , where  $V$  represents the set of **vertices** (also called nodes or points) and  $E$  represents the set of all **edges** (also called links) in the graph. We distinguish many types of graphs:
  1. **Directed graph:** a graph  $G = (V, A)$ , where all the edges are **directed** from one vertex to another. The edges are in general called **arcs**.
  2. **Undirected graph:** a graph  $G = (V, E)$ , where all the edges are **bidirectional**.
  3. **Mixed graph:** a graph  $G = (V, E, A)$  consisting of a set of **undirected** edges  $E$ , and a set of **directed** edges (arcs)  $A$ .

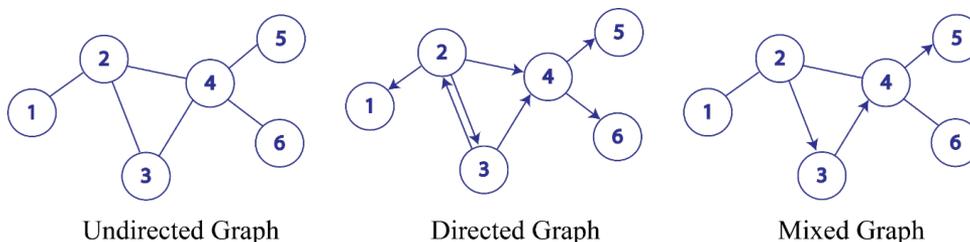


Figure A.1.: Graph Types based on directionality of edges

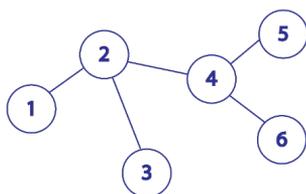
- **Subgraph:** a subgraph  $G' = (V', E')$  of a graph  $G = (V, E)$ , is a subset of vertices of  $V$  and a subset of edges  $E$ . An **induced subgraph**, is subgraph where the subset of edges contains only the vertices
- **coloured graph:** it is a graph in which each vertex is assigned a colour. In our case, colours on the vertices represents the chemical types of the atoms.

- **Adjacency Matrices**

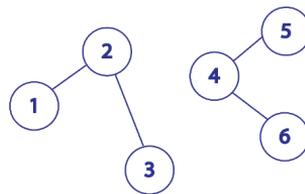
The adjacency matrix is a mathematical way to represent graph. It is an  $n \times n$  matrix where  $n$  represents the number of vertices of the graph and the values in the matrix represents the edges between these vertices. It is important to note that the matrix representation of an undirected graph is symmetric which is not always the case for a directed or mixed graph.

- **Connected component**

In graph theory, a connected component of an undirected graph is an induced subgraph in which each pair of vertices is connected to each other via a path.



Graph with one connected component



Graph with two connected components

Figure A.2.: Connected components in graphs

## A.2. MOLECULAR CONFORMATION AND RELATIONSHIP TO GRAPH THEORY

- **Molecular Graph** : One crucial step of the algorithm is to set up a model that defines a molecular conformation with the right level of granularity. In our case, we have chosen to define the Configuration in terms of covalent bonds and organometallic and ionic interactions formed between the atoms. The definitions are based on *Euclidian distances* as described below:
- **A bond** is formed between a pair of atoms  $[a, b]$  with respective Cartesian coordinates  $(x_a, y_a, z_a)$  and  $(x_b, y_b, z_b)$ , if the Euclidean distance

$$\sqrt{(x_a - x_b)^2 + (y_a - y_b)^2 + (z_a - z_b)^2} \quad (\text{A.1})$$

is less than a cut-off distance  $D_r$ . For **covalent bonds**, the algorithm defines the  $D_r$  distance by the sum of covalent radii of atoms  $a$  and  $b$  with an additional margin of 2% of this sum. For the **organometallic and ionic interactions**, expert user is given the freedom to set case specific  $D_r$  distances, for example, the distance between manganese and oxygen atoms (2.44 Angstrom), etc.

By defining the different bonds, graphs corresponding to different conformations are constructed. By definition, a conformation is translated into a **mixed graph** [1]

$$G = (V, E_C, A_H, E_I, E_O) \quad (\text{A.2})$$

where:

- $V$ : is the set of all atoms present in the conformation. Each atom represents one vertex in the graph  $G$ .
- $E_C$ : the set of **covalent** bonds. Each covalent bond represents **undirected edge** in the graph  $G$ .
- $A_H$ : the set of **hydrogen** bonds. Each hydrogen bond represents **directed edge** in the graph  $G$ .
- $E_I$ : the set of **intermolecular/ionic** interactions. Each intermolecular/ionic interaction represents **undirected edge** in the graph  $G$ .
- $E_O$ : the set of **organometallic** interactions. Each organometallic interaction represents **undirected edge** in the graph  $G$ .

The Cartesian atomic positions taken from the trajectory are used *only* for forming the mixed graphs. Once the graphs are obtained, the changes in the conformations are analysed through the comparison of these graphs using isomorphism check.

- **Molecular Graph isomorphism**

Molecular configurations are translated into unique graphs based on threshold values (Equation A1). Once conceived as graphs, isomorphism check is done to track possible changes along the trajectory. To reduce the cost of the isomorphism tests, optimisations were used as defining orbits around the atoms [2]. For the present work, the algorithm is developed for analysing molecular dynamics trajectories in terms of arbitrary bonding types which represent the conformational change.

Two conformations are different if and only if they are not isomorphic. The isomorphism between two graphs is defined by a **bijection** between them: Two graphs  $G_a$  and  $G_b$  are **isomorphic** if and only if there exists a **bijection**  $\theta_{a,b} : V_a \rightarrow V_b$  such that :

- $\forall v \in V_a, \theta(v) = \theta(\theta_{a,b}(v)), \theta_{a,b}(v) \in V_b$
- $[v, u] \in E_{c_a} \Leftrightarrow [\theta_{a,b}(v), \theta_{a,b}(u)] \in E_b$
- $(v, u) \in A_{H_a} \Leftrightarrow (\theta_{a,b}(v), \theta_{a,b}(u)) \in A_{H_b}$
- $[v, u] \in E_{I_a} \Leftrightarrow [\theta_{a,b}(v), \theta_{a,b}(u)] \in E_{I_b}$
- $[v, u] \in E_{O_a} \Leftrightarrow [\theta_{a,b}(v), \theta_{a,b}(u)] \in E_{O_b}$

In our case, we apply an isomorphism test considering the atom chemical type as an attribute for different nodes. Consequently, each atom will be given a *colour*. Two atoms can be exchanged if and only if have the same colour. Using this partitioning will not only allow a coherent comparison between graphs but also reduce the number of possible permutations to decide if the graphs are identical or not and

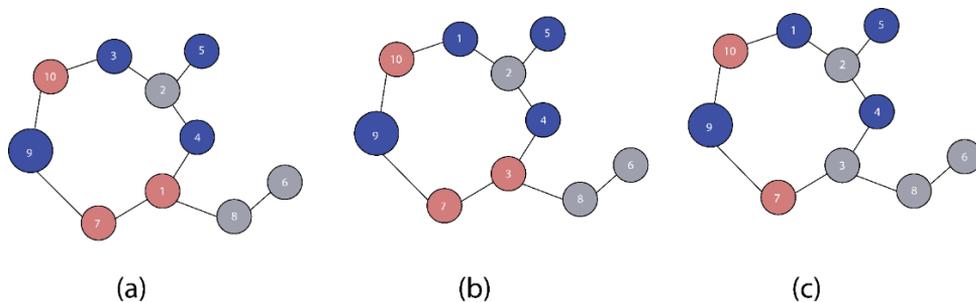


Figure A.3.: using the isomorphism conditions, graphs a and b are considered isomorphic while graph c is not isomorphic to them

thus the algorithm can perform faster. Figure A3.1. illustrates three example graphs where graphs (a) and (b) are isomorphic, while graph (c) is not isomorphic to a and b.

A molecular system is defined as a graph such that vertices represent atoms of the molecular system and edges represent the bonds formed between these atoms (covalent bond, hydrogen bonds, intermolecular electrostatic interactions, etc., depending on the system). Conformational dynamics of molecular system do occur (evidently depending on internal energies and energy barriers on the potential energy surface), with "fastest motions" being hydrogen/ionic bond dynamics (forming/breaking) along time while larger amplitude motions can induce large structural modifications such as torsional movements. Chemical reactions are defined as the occurrences of formation/breakage (covalent) bonds, leading to numerous changes in the chemical entities over time. Bond dynamics (forming/breaking) represent, in the graph terminology, a change in the edge sets. The exploration of different configurations can be seen as an exploration of different graph topologies, tracked using graph theory methods via checking for isomorphic graphs in the configuration ensemble [1]. An isomorphism between two graphs is a bijection between their vertex sets that preserves adjacency<sup>3</sup>, in other words, it is a function between the elements of two sets, where each element of one set is paired with exactly one element of the other set, and each element of the other set is paired with exactly one element of the first set.

Isomorphism checks are the key components of the *reaction event exploration* step and is used to identify unique conformations from the reactive trajectory. Based on the time evolution of the unique conformers in the reactive trajectory, a graph of transition (showing how the conformations are related one to another and the time sequence) for species present in the reference network is formed. The changes in conformations are followed over time evolution of trajectories for changes in bonding patterns of choice (among hydrogen bond(s), proton transfer(s), coordination number(s), covalent bond(s) and organometallic interaction(s)).

### A.3. DFT ENERGIES OF STRUCTURES IDENTIFIED BY ReNeGate

A

Table A.1.: DFT optimised structures for species present in the ReNeGate -computed reaction network of a model  $\text{Mn}(\text{CO})_5\text{Br}$  pre-catalyst by KOiPr base (Figure.3.5)

Configuration	ID	Energy	Interaction	$\Delta E(\text{kcal.mol}^{-1})$
Configuration2	II	-5083.214368	Nucleophilic attack	-15.2
Configuration3	II	-5083.214189	Nucleophilic attack	-15.2
Configuration4	III	-5083.180776	C(O)Oipr, $\alpha$ -ketoacyl+KBr	5.9
Configuration6	II	-5083.214183	Nucleophilic attack	-15.1
Configuration7	II	-5083.213168	Nucleophilic attack	-14.5
Configuration8	IV	-5083.197949	C(O)Oipr, $\alpha$ -ketoacyl	-4.9
Configuration9	III	-5083.19795	C(O)Oipr, $\alpha$ -ketoacyl+KBr	-4.9
Configuration10	II	-5083.218841	Nucleophilic attack	-18

Table A.2.: DFT optimised structures for species present in the ReNeGate-computed reaction network of a model MnBr(CO)<sub>3</sub>NN pre-catalyst by KOiPr base in isopropanol solvent (Figure 3.6)

Configuration	Energy(Hartrees)	Interaction	$\Delta E(\text{kcal.mol}^{-1})$
Configuration43	-6055.632111	Mn-OiPr_KBr	0
Configuration4	-6055.630418	Mn-OiPr_KBr	1.1
Configuration31	-6055.624409	Mn-OiPr_KBr	4.8
Configuration39	-6055.620297	Mn-OiPr_KBr	7.4
Configuration32	-6055.618161	Mn-OiPr_KBr	8.8
Configuration40	-6055.617387	Mn-OiPr_KBr	9.2
Configuration62	-6055.61718	Mn-OiPr_KBr	9.4
Configuration33	-6055.617092	Mn-OiPr_KBr	9.4
Configuration48	-6055.606542	PrOH-Mn-OCOR_KBr	16.1
Configuration50	-6055.601041	PrOH-Mn-OCOR_KBr	19.5
Configuration44	-6055.594436	Mn-OPr_N-diss_KBr	23.7
Configuration83	-6055.602944	Mn-OCOR_KBr	18.3
Configuration77	-6055.60094	Mn-OCOR_KBr	19.6
Configuration11	-6055.594466	Mn-OCOR_KBr	23.6
Configuration81	-6055.590699	Mn-OCOR_KBr	26
Configuration85	-6055.584681	Mn-OCOR_KBr	29.8
Configuration49	-6055.583848	Mn-OCOR_KBr	30.3
Configuration6	-6055.580477	Mn-OCOR_KBr	32.4
Configuration28	-6055.5978	Mn-Br_N-diss_ROCO	21.5
Configuration88	-6055.596875	Mn-Br_N-diss_ROCO	22.1
Configuration18	-6055.583781	Mn-Br_N-diss_ROCO	30.3
Configuration26	-6055.580434	Mn-Br_N-diss_ROCO	32.4
Configuration87	-6055.57916	Mn-Br_N-diss_ROCO	33.3
Configuration27	-6055.574144	Mn-Br_N-diss_ROCO	36.4
Configuration19	-6055.573722	Mn-Br_N-diss_ROCO	36.7
Configuration23	-6055.571235	Mn-Br_N-diss_ROCO	38.2
Configuration30	-6055.566041	Mn-Br_N-diss_ROCO	41.5
Configuration70	-6055.555067	Mn-Br_N-diss_ROCO	48.4
Configuration17	-6055.548645	Mn-Br_N-diss_ROCO	52.4
Configuration75	-6055.608089	Mn-Br_OCOR_NN-coord	15.1
Configuration54	-6055.602775	Mn-Br_OCOR_NN-coord	18.4
Configuration52	-6055.602211	Mn-Br_OCOR_NN-coord	18.8
Configuration74	-6055.590595	K-Br-Mn-OR_NN-diss	26.1
Configuration45	-6055.571787	K-Br-Mn-OR_NN-diss	37.9
Configuration37	-6055.560253	NN-diss_Mn-Opr_KBr	45.1
Configuration36	-6055.5754	NN-diss_Br-Mn-OiPr	35.6

Table A.3.: DFT optimised structures for fragments in the ReNeGate -computed reaction network for cluster formation upon base-activation of  $\text{Mn}(\text{CO})_5\text{Br}$  pre-catalyst by KOiPr base (Figure 3.7)

<b>Fragment</b>	<b>Energy(Hartree)</b>	<b>Fragment ID</b>	<b><math>\Delta E(\text{kcal.mol}^{-1})</math></b>
Fragment1	-1911.439826	F1(Lower)	-61.5
Fragment7	-1911.394654	F1(Upper)	-4.6
Fragment6	-1911.394306	F1(Upper)	-4.1
Fragment5	-1911.394306	F1(Upper)	-4.1
Fragment42	-3822.782028	F4	0
Fragment15	-3822.773267	F3	5.5
Fragment34	-3822.771475	F4	6.6
Fragment35	-3822.771475	F4	6.6
Fragment43	-3822.764056	F4	11.3
Fragment24	-3822.761704	F3	12.8
Fragment29	-3822.76073	F3	13.4
Fragment25	-3822.759259	F3	14.3
Fragment27	-3822.758379	F3	14.9
Fragment17	-1911.345496	F2	57.3
Fragment18	-1911.345226	F2	57.7
Fragment16	-1911.34329	F2	60.1
Fragment20	-1911.343249	F2	60.2
Fragment8	-1911.339866	F2	64.4
Fragment9	-1911.328319	F2	80
Fragment40	-1911.324295	F2	84
Fragment21	-1911.324019	F2	84.4
Fragment39	-1911.324019	F2	84.4
Fragment38	-1911.324019	F2	84.4



# B

## SUPPLEMENTARY INFORMATION CHAPTER 4

### B.1. CORRELATION ANALYSIS OF THE STRUCTURAL FEATURES AND REACTIVITY FOR STRUCTURES IN $[-40,25]$ KCAL.MOL<sup>-1</sup> RANGE

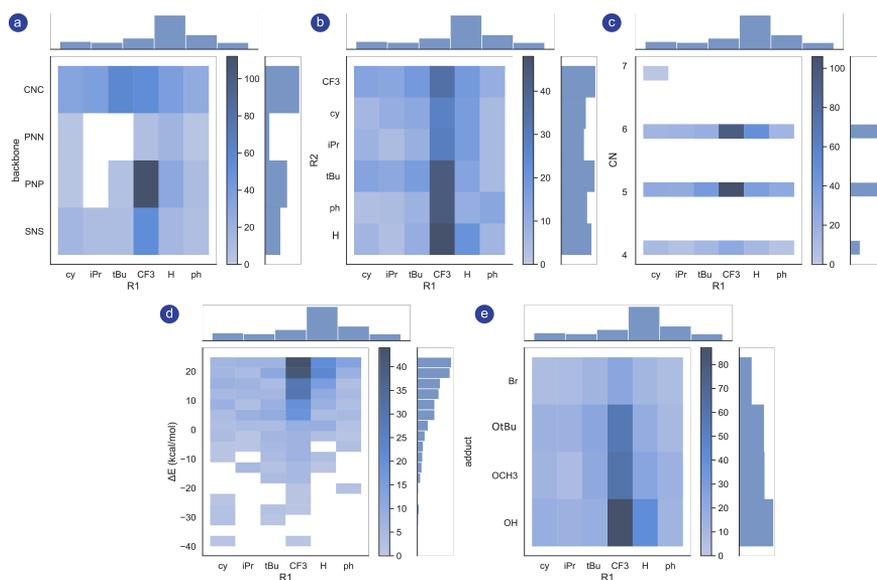


Figure B.1.: Histogram plots for the frequency of appearance of alternative stable species in the  $[-40,25]$  kcal.mol<sup>-1</sup> range based on modified structures for CNC, PNP, PNN and SNS catalysts as a function of feature combination (a)  $R_1$  - backbone, (b)  $R_1$  -  $R_2$ , (c)  $R_1$  - CN (coordination number of the metal center) (d)  $R_1$  -  $\Delta E$  (relative energy) and (e)  $R_1$ -X.

The trends observed for  $R_1$ - (backbone,  $R_2$ , CN,  $\Delta E$ , X) combinations in the [-40, 0] kcal.mol<sup>-1</sup> range are also generally observed in the [-40,25] kcal.mol<sup>-1</sup> range. Additional observations limited to only [0,25] kcal.mol<sup>-1</sup> for  $R_1$ -backbone combinations include (H, Ph)-PNN, (H,cy,Ph)-PNP, (<sup>t</sup>Bu,cy,Ph,<sup>i</sup>Pr)-SNS. Presence of bulky (Ph,cy,<sup>t</sup>Bu and <sup>i</sup>Pr) moieties at  $R_1$  can give rise to unstable alternative structures.  $R_1$ - $R_2$ : (ph,<sup>i</sup>Pr)(H,ph), cy-ph, H-(<sup>i</sup>Pr,<sup>t</sup>Bu) combinations are also limited to structures within [0,25] kcal.mol<sup>-1</sup>. 7 coordinated Mn is observed also only in [0,25] kcal.mol<sup>-1</sup> when cy is present at  $R_1$ . For different  $R_1$ -X combinations Br-(ph,<sup>i</sup>Pr), O<sup>t</sup>Bu-(H,<sup>i</sup>Pr), OCH<sub>3</sub>-ph combinations only lead to structures with higher energies than the respective reference.

## **B.2. ROLE OF ADDUCTS IN OBSERVED REACTIVITIES:**

## **B.3. EXPLORATION RESULTS LEADING TO DECOORDINATION OF MN-N LIGANDS FOR CATALYSTS WITH DIFFERENT BACKBONES:**

## **B.4. EXPLORATION RESULTS FOR ATTACK BY NHC MOIETIES OF THE CNC PINCER ON THE CARBONYL LIGANDS (MN-C INTERACTIONS):**

## **B.5. SUPPLEMENTARY TABLES**

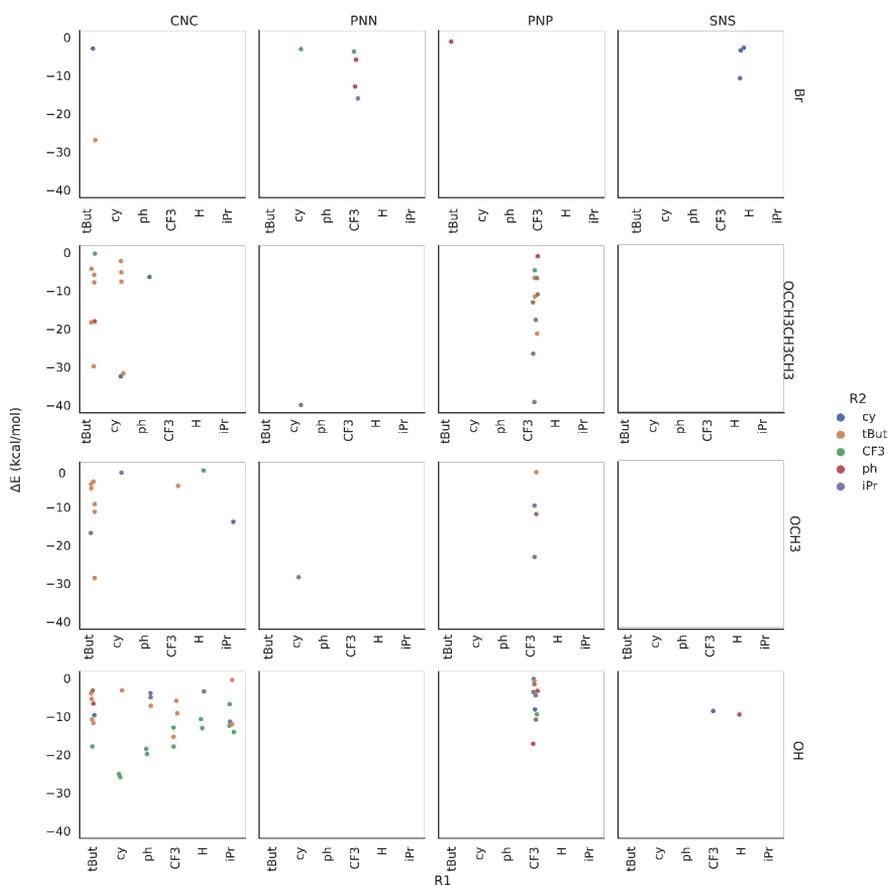


Figure B.2.: Exploration results for deactivated species. species with lower energies compared to respective reference structures are categorized based on the ligand substituent on the  $R_1$  position on the backbone of the catalyst for different backbones and adducts on the Mn center, nodes are colored based on the ligand present at  $R_2$  position on the catalyst backbone.

B

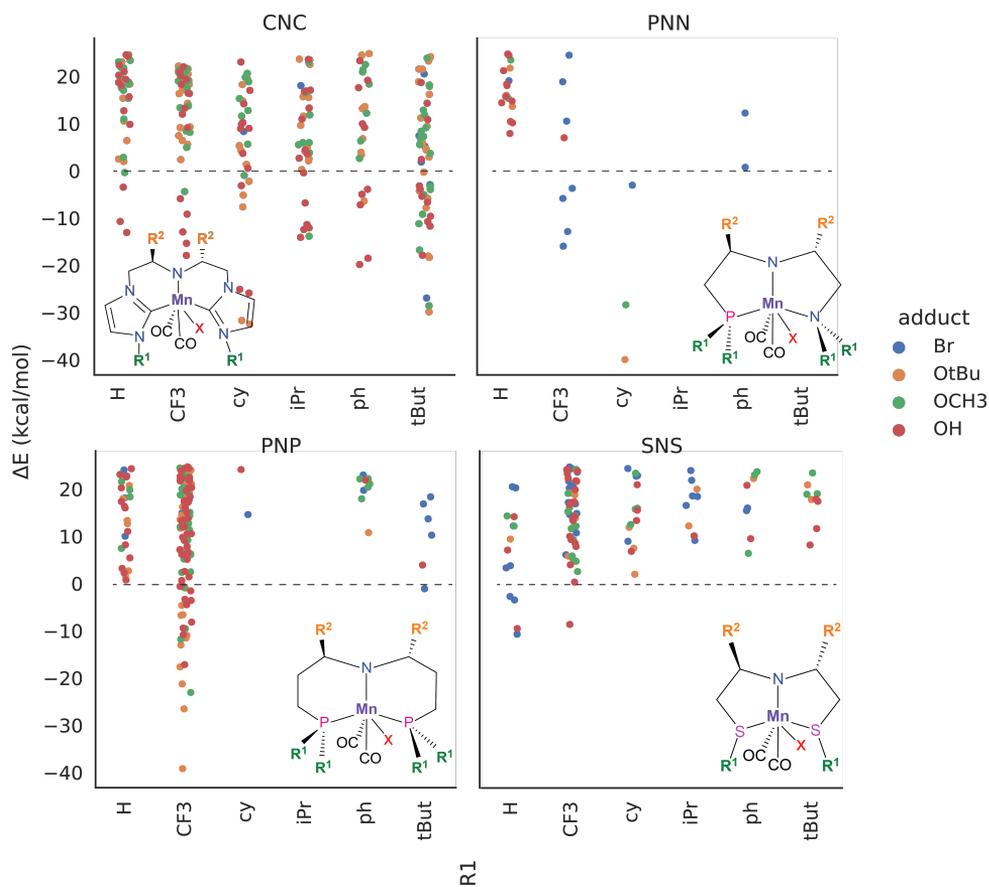


Figure B.3.: Explored results for species less than  $25 \text{ kcal}\cdot\text{mol}^{-1}$  different in energy from the respective reference structures. Species are categorized based on the ligand substituent on the  $R_1$  position on the backbone of the catalyst for different backbones and adducts on the Mn center.

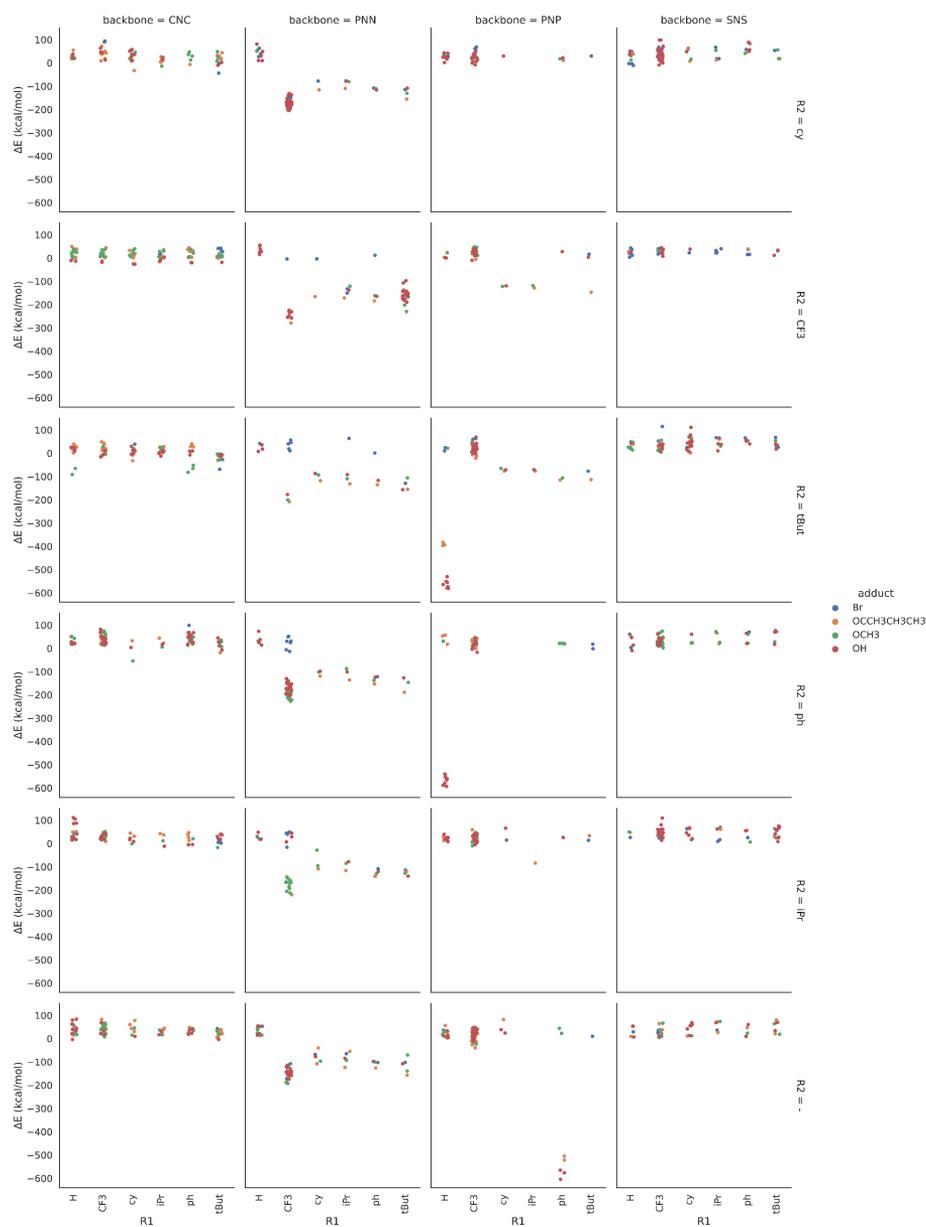


Figure B.4.: Exploration results for all explored species. species are categorized based on the ligand substituent on the  $R_1$  position on the backbone of the catalyst and adducts on the Mn center, nodes are colored based on the adduct present on the Mn.

B

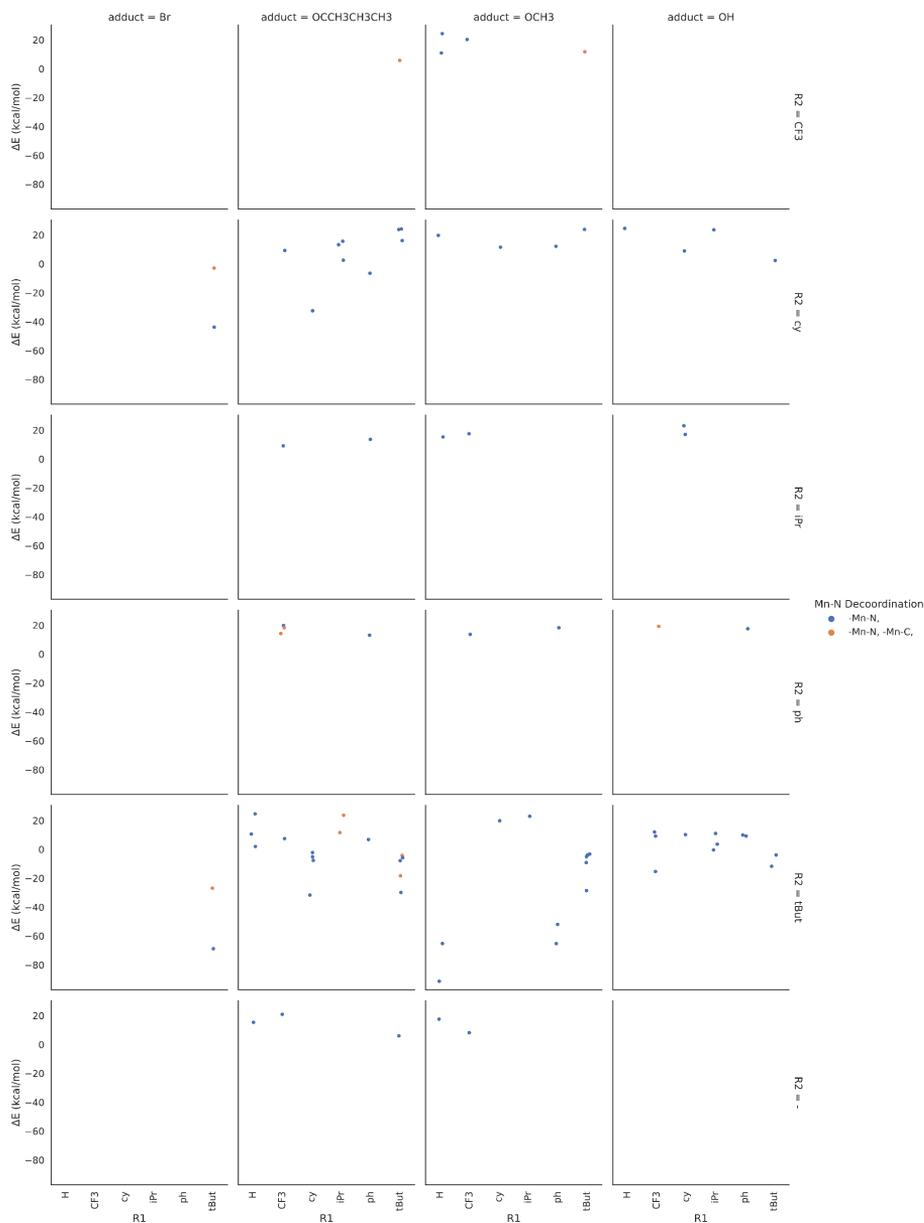


Figure B.5.: Mn-N decoordination CNC presence of bulky *t*Bu and cy functional groups at R<sub>2</sub> position leads in the decoordination of the Mn-N bond. When CF<sub>3</sub> or no functional groups are present at R<sub>2</sub> position along with H or CF<sub>3</sub> groups at R<sub>1</sub> position -Mn-N is observed. Mn decoordination is not observed when Br is coordinated to the Mn center.

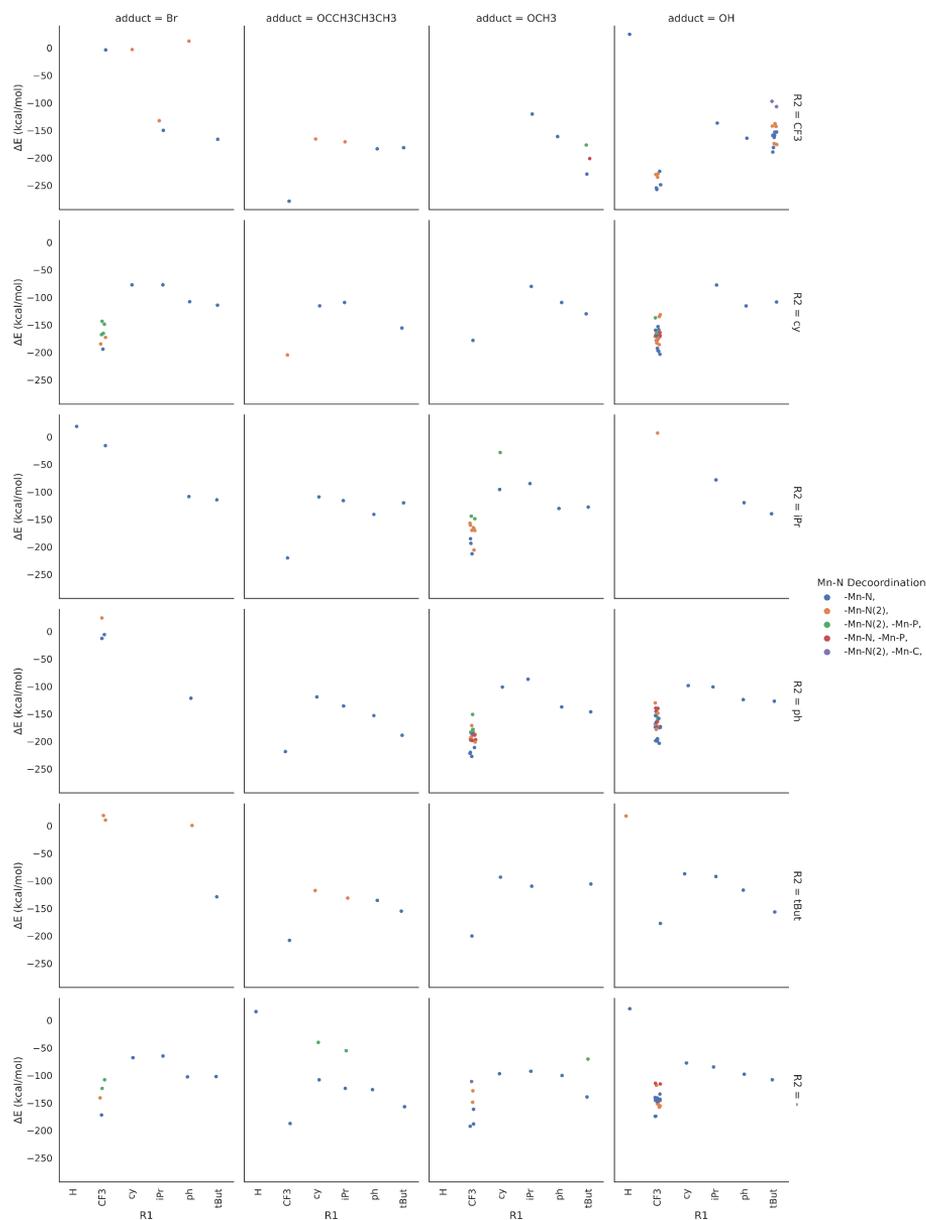


Figure B.6.: Mn-N decooordination for catalysts with PNN backbone: presence of  $\text{CF}_3$  groups at either of  $R_1$  or  $R_2$  position has a direct impact on the decooordination of the Mn-N bond. Mn-N bond very easily decoordinates with all  $R_1$ - $R_2$ -Adduct combinations. Decoordination of both Mn-N bonds, -Mn-N, -Mn-P or complete decooordination of the ligands are specific to the presence of  $\text{CF}_3$  groups at  $R_1$  position.

B

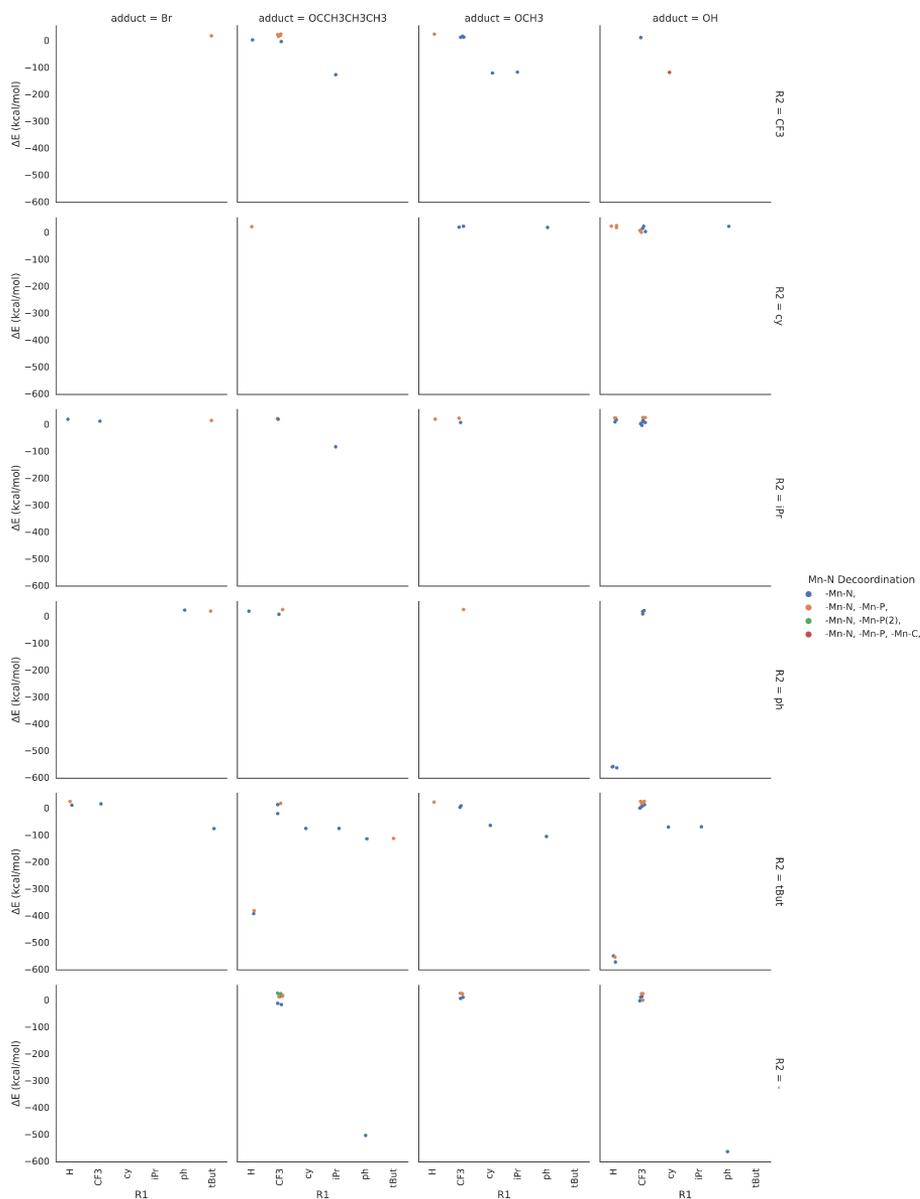


Figure B.7.: Mn-N decooordination for catalysts with PNP backbone: decooordination behaviors observed include decooordination of the -Mn-N bond along with either or both of Mn-P ligands donors. While decooordination of the Mn-N bond is observed for all  $R_1$  - $R_2$  combinations, decooordination of one Mn-P is observed when H or CF<sub>3</sub> groups are present at  $R_1$  position. Complete dissociation of all backbone ligands is observed when CF<sub>3</sub> groups are present at  $R_1$  and no  $R_2$  (-) modifications are made.

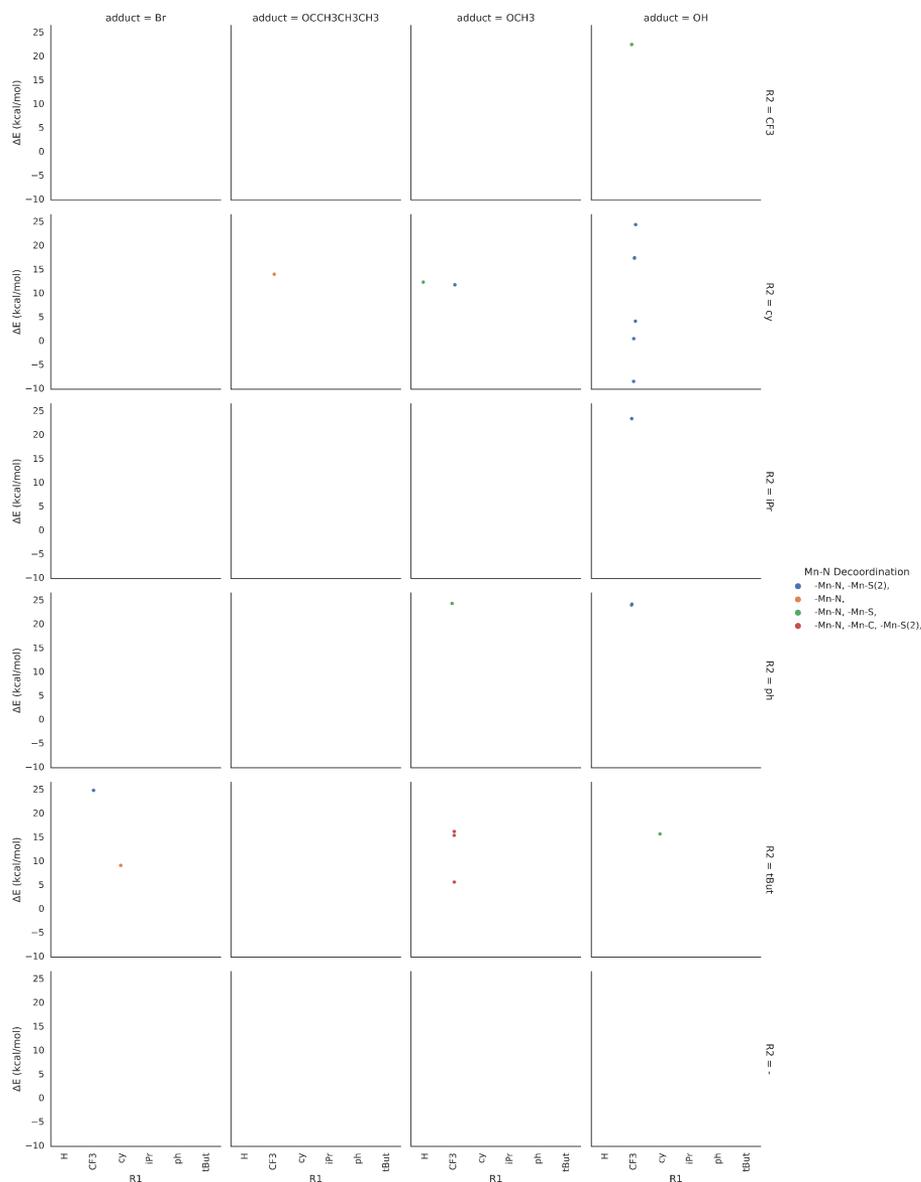


Figure B.8.: Mn-N decoordination for catalyst with SNS backbone: For catalysts with SNS scaffolds, decoordination behaviors observed include decoordination of the -Mn-N bond along with either or both of Mn-S ligands donors. Presence of CF<sub>3</sub> groups at R<sub>1</sub> position has a clear impact on the decoordination of both Mn-N and Mn-S donor ligands. Decoordination of the central Mn-N with R<sub>1</sub>: CF<sub>3</sub> and R<sub>2</sub>:cy.

B

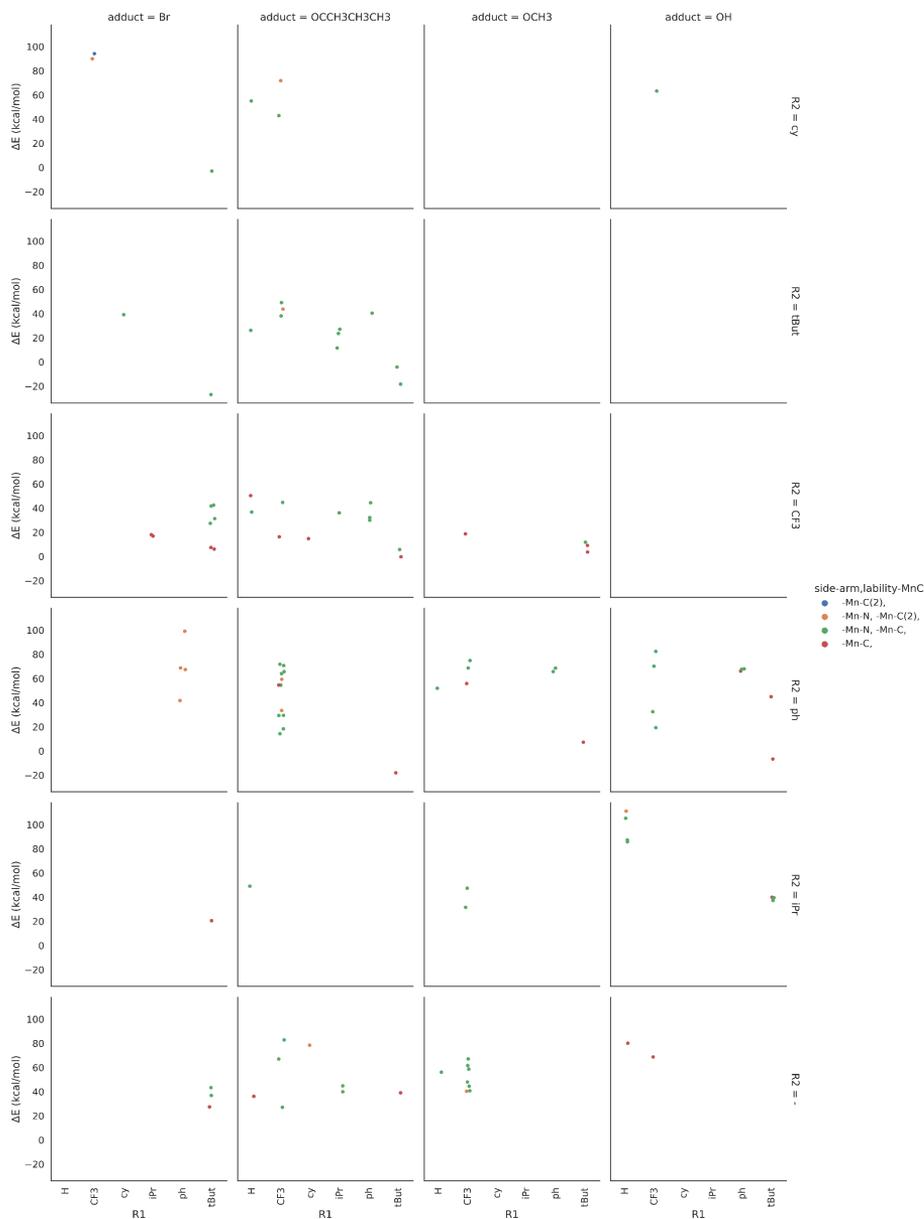


Figure B.9.: Mn-C decooordination for catalyst with CNC backbone: Presence of bulky <sup>t</sup>Bu groups at R<sub>1</sub> positions leads to the decooordination of the Mn-C along with decooordination of the MnN bond. Presence of electron withdrawing CF<sub>3</sub> groups at R<sub>2</sub> position also causes the decooordination of the Mn-C bond.

Table B.1.: Exploration data for structures where decoordination of the central Mn-C bond has been observed: Structural ( $R_1$ ,  $R_2$ , adduct, backbone) and reactive ( $\Delta E(\text{kcal.mol}^{-1})$ ), '-Mn-C' side-arm lability, coordination number of Mn (CN) features are listed for explored structures.

ID	CN	$\Delta E(\text{kcal.mol}^{-1})$	Adduct	$R_1$	$R_2$	backbone	side-arm lability
1	6	94.16	Br	$\text{CF}_3$	cy	CNC	-Mn-C(2)
2	5	89.96	Br	$\text{CF}_3$	cy	CNC	-Mn-N, -Mn-C(2)
3	5	39.12	Br	cy	$t\text{Bu}$	CNC	-Mn-N, -Mn-C
4	5	16.89	Br	$i\text{Pr}$	$\text{CF}_3$	CNC	-Mn-C
5	6	18.04	Br	$i\text{Pr}$	$\text{CF}_3$	CNC	-Mn-C
6	5	41.79	Br	ph	ph	CNC	-Mn-N, -Mn-C(2)
7	5	68.8	Br	ph	ph	CNC	-Mn-N, -Mn-C(2)
8	4	67.38	Br	ph	ph	CNC	-Mn-N, -Mn-C(2)
9	5	99.08	Br	ph	ph	CNC	-Mn-N, -Mn-C(2)
10	6	6.11	Br	$t\text{Bu}$	$\text{CF}_3$	CNC	-Mn-C
11	5	7.44	Br	$t\text{Bu}$	$\text{CF}_3$	CNC	-Mn-C
12	5	41.77	Br	$t\text{Bu}$	$\text{CF}_3$	CNC	-Mn-N, -Mn-C
13	4	42.5	Br	$t\text{Bu}$	$\text{CF}_3$	CNC	-Mn-N, -Mn-C
14	5	31.35	Br	$t\text{Bu}$	$\text{CF}_3$	CNC	-Mn-N, -Mn-C
15	6	27.43	Br	$t\text{Bu}$	$\text{CF}_3$	CNC	-Mn-N, -Mn-C
16	4	-2.93	Br	$t\text{Bu}$	cy	CNC	-Mn-N, -Mn-C
17	5	20.53	Br	$t\text{Bu}$	$i\text{Pr}$	CNC	-Mn-C
18	5	-26.92	Br	$t\text{Bu}$	$t\text{Bu}$	CNC	-Mn-N, -Mn-C
19	5	27.43	Br	$t\text{Bu}$	H	CNC	-Mn-C
20	5	43.31	Br	$t\text{Bu}$	H	CNC	-Mn-N, -Mn-C
21	5	36.82	Br	$t\text{Bu}$	H	CNC	-Mn-N, -Mn-C
22	5	16.33	$O^t\text{Bu}$	$\text{CF}_3$	$\text{CF}_3$	CNC	-Mn-C
23	4	44.73	$O^t\text{Bu}$	$\text{CF}_3$	$\text{CF}_3$	CNC	-Mn-N, -Mn-C
24	4	42.97	$O^t\text{Bu}$	$\text{CF}_3$	cy	CNC	-Mn-N, -Mn-C
25	4	71.86	$O^t\text{Bu}$	$\text{CF}_3$	cy	CNC	-Mn-N, -Mn-C(2)

26	O <sup>t</sup> Bu-CNC-CF <sub>3</sub> -ph-5	5	65.67	O <sup>t</sup> Bu	CF <sub>3</sub>	ph	CNC	-Mn-N, -Mn-C
27	O <sup>t</sup> Bu-CNC-CF <sub>3</sub> -ph-6	6	70.68	O <sup>t</sup> Bu	CF <sub>3</sub>	ph	CNC	-Mn-N, -Mn-C
28	O <sup>t</sup> Bu-CNC-CF <sub>3</sub> -ph-10	6	64.03	O <sup>t</sup> Bu	CF <sub>3</sub>	ph	CNC	-Mn-N, -Mn-C
29	O <sup>t</sup> Bu-CNC-CF <sub>3</sub> -ph-11	5	71.77	O <sup>t</sup> Bu	CF <sub>3</sub>	ph	CNC	-Mn-N, -Mn-C
30	O <sup>t</sup> Bu-CNC-CF <sub>3</sub> -ph-12	5	59.34	O <sup>t</sup> Bu	CF <sub>3</sub>	ph	CNC	-Mn-N, -Mn-C(2)
31	O <sup>t</sup> Bu-CNC-CF <sub>3</sub> -ph-13	5	54.52	O <sup>t</sup> Bu	CF <sub>3</sub>	ph	CNC	-Mn-N, -Mn-C
32	O <sup>t</sup> Bu-CNC-CF <sub>3</sub> -ph-14	5	14.3	O <sup>t</sup> Bu	CF <sub>3</sub>	ph	CNC	-Mn-N, -Mn-C
33	O <sup>t</sup> Bu-CNC-CF <sub>3</sub> -ph-15	4	29.41	O <sup>t</sup> Bu	CF <sub>3</sub>	ph	CNC	-Mn-N, -Mn-C
34	O <sup>t</sup> Bu-CNC-CF <sub>3</sub> -ph-16	5	29.48	O <sup>t</sup> Bu	CF <sub>3</sub>	ph	CNC	-Mn-N, -Mn-C
35	O <sup>t</sup> Bu-CNC-CF <sub>3</sub> -ph-17	5	18.32	O <sup>t</sup> Bu	CF <sub>3</sub>	ph	CNC	-Mn-N, -Mn-C
36	O <sup>t</sup> Bu-CNC-CF <sub>3</sub> -ph-18	4	33.52	O <sup>t</sup> Bu	CF <sub>3</sub>	ph	CNC	-Mn-N, -Mn-C(2)
37	O <sup>t</sup> Bu-CNC-CF <sub>3</sub> -ph-19	6	54.55	O <sup>t</sup> Bu	CF <sub>3</sub>	ph	CNC	-Mn-C
38	O <sup>t</sup> Bu-CNC-CF <sub>3</sub> - <sup>t</sup> Bu-6	4	38.07	O <sup>t</sup> Bu	CF <sub>3</sub>	<sup>t</sup> Bu	CNC	-Mn-N, -Mn-C
39	O <sup>t</sup> Bu-CNC-CF <sub>3</sub> - <sup>t</sup> Bu-7	5	49.16	O <sup>t</sup> Bu	CF <sub>3</sub>	<sup>t</sup> Bu	CNC	-Mn-N, -Mn-C
40	O <sup>t</sup> Bu-CNC-CF <sub>3</sub> - <sup>t</sup> Bu-8	4	43.77	O <sup>t</sup> Bu	CF <sub>3</sub>	<sup>t</sup> Bu	CNC	-Mn-N, -Mn-C(2)
41	O <sup>t</sup> Bu-CNC-CF <sub>3</sub> -3	4	27.08	O <sup>t</sup> Bu	CF <sub>3</sub>	H	CNC	-Mn-N, -Mn-C
42	O <sup>t</sup> Bu-CNC-CF <sub>3</sub> -9	6	82.83	O <sup>t</sup> Bu	CF <sub>3</sub>	H	CNC	-Mn-N, -Mn-C
43	O <sup>t</sup> Bu-CNC-CF <sub>3</sub> -10	5	66.98	O <sup>t</sup> Bu	CF <sub>3</sub>	H	CNC	-Mn-N, -Mn-C
44	O <sup>t</sup> Bu-CNC-cy-CF <sub>3</sub> -8	5	14.78	O <sup>t</sup> Bu	Cy	CF <sub>3</sub>	CNC	-Mn-C
45	O <sup>t</sup> Bu-CNC-cy-5	5	78.44	O <sup>t</sup> Bu	Cy	H	CNC	-Mn-N, -Mn-C(2)
46	O <sup>t</sup> Bu-CNC-H-CF <sub>3</sub> -5	5	50.35	O <sup>t</sup> Bu	H	CF <sub>3</sub>	CNC	-Mn-C
47	O <sup>t</sup> Bu-CNC-H-CF <sub>3</sub> -6	4	36.77	O <sup>t</sup> Bu	H	CF <sub>3</sub>	CNC	-Mn-N, -Mn-C
48	O <sup>t</sup> Bu-CNC-H-cy-3	4	55.07	O <sup>t</sup> Bu	H	Cy	CNC	-Mn-N, -Mn-C
49	O <sup>t</sup> Bu-CNC-H- <sup>i</sup> Pr-5	4	49.11	O <sup>t</sup> Bu	H	<sup>i</sup> Pr	CNC	-Mn-N, -Mn-C
50	O <sup>t</sup> Bu-CNC-H- <sup>t</sup> Bu-2	4	26.18	O <sup>t</sup> Bu	H	<sup>t</sup> Bu	CNC	-Mn-N, -Mn-C
51	O <sup>t</sup> Bu-CNC-H-2	5	36.18	O <sup>t</sup> Bu	H	H	CNC	-Mn-C
52	O <sup>t</sup> Bu-CNC- <sup>i</sup> Pr-CF <sub>3</sub> -5	4	36.12	O <sup>t</sup> Bu	<sup>i</sup> Pr	CF <sub>3</sub>	CNC	-Mn-N, -Mn-C
53	O <sup>t</sup> Bu-CNC- <sup>i</sup> Pr- <sup>t</sup> Bu-2	4	11.57	O <sup>t</sup> Bu	<sup>i</sup> Pr	<sup>t</sup> Bu	CNC	-Mn-N, -Mn-C
54	O <sup>t</sup> Bu-CNC- <sup>i</sup> Pr- <sup>t</sup> Bu-5	4	23.63	O <sup>t</sup> Bu	<sup>i</sup> Pr	<sup>t</sup> Bu	CNC	-Mn-N, -Mn-C

55	O <sup>t</sup> Bu-CNC- <sup>i</sup> Pr- <sup>t</sup> Bu-8	5	27.07	O <sup>t</sup> Bu	<sup>i</sup> Pr	<sup>t</sup> Bu	CNC	-Mn-N, -Mn-C
56	O <sup>t</sup> Bu-CNC- <sup>i</sup> Pr-4	5	44.79	O <sup>t</sup> Bu	<sup>i</sup> Pr	H	CNC	-Mn-N, -Mn-C
57	O <sup>t</sup> Bu-CNC- <sup>i</sup> Pr-5	4	39.94	O <sup>t</sup> Bu	<sup>i</sup> Pr	H	CNC	-Mn-N, -Mn-C
58	O <sup>t</sup> Bu-CNC-ph-CF <sub>3</sub> -4	4	29.98	O <sup>t</sup> Bu	Ph	CF <sub>3</sub>	CNC	-Mn-N, -Mn-C
59	O <sup>t</sup> Bu-CNC-ph-CF <sub>3</sub> -5	4	32.19	O <sup>t</sup> Bu	Ph	CF <sub>3</sub>	CNC	-Mn-N, -Mn-C
60	O <sup>t</sup> Bu-CNC-ph-CF <sub>3</sub> -6	4	44.46	O <sup>t</sup> Bu	Ph	CF <sub>3</sub>	CNC	-Mn-N, -Mn-C
61	O <sup>t</sup> Bu-CNC-ph- <sup>t</sup> Bu-4	4	40.41	O <sup>t</sup> Bu	Ph	<sup>t</sup> Bu	CNC	-Mn-N, -Mn-C
62	O <sup>t</sup> Bu-CNC- <sup>t</sup> Bu-CF <sub>3</sub> -2	4	5.81	O <sup>t</sup> Bu	<sup>t</sup> Bu	CF <sub>3</sub>	CNC	-Mn-N, -Mn-C
63	O <sup>t</sup> Bu-CNC- <sup>t</sup> Bu-CF <sub>3</sub> -3	5	-0.28	O <sup>t</sup> Bu	<sup>t</sup> Bu	CF <sub>3</sub>	CNC	-Mn-C
64	O <sup>t</sup> Bu-CNC- <sup>t</sup> Bu-ph-2	5	-18.05	O <sup>t</sup> Bu	<sup>t</sup> Bu	Ph	CNC	-Mn-C
65	O <sup>t</sup> Bu-CNC- <sup>t</sup> Bu- <sup>t</sup> Bu-3	4	-18.33	O <sup>t</sup> Bu	<sup>t</sup> Bu	<sup>t</sup> Bu	CNC	-Mn-N, -Mn-C
66	O <sup>t</sup> Bu-CNC- <sup>t</sup> Bu- <sup>t</sup> Bu-4	4	-4.24	O <sup>t</sup> Bu	<sup>t</sup> Bu	<sup>t</sup> Bu	CNC	-Mn-N, -Mn-C
67	O <sup>t</sup> Bu-CNC- <sup>t</sup> Bu-5	5	39.06	O <sup>t</sup> Bu	<sup>t</sup> Bu	H	CNC	-Mn-C
68	OCH <sub>3</sub> -CNC-CF <sub>3</sub> -CF <sub>3</sub> -11	5	18.71	OCH <sub>3</sub>	CF <sub>3</sub>	CF <sub>3</sub>	CNC	-Mn-C
69	OCH <sub>3</sub> -CNC-CF <sub>3</sub> - <sup>i</sup> Pr-3	4	31.61	OCH <sub>3</sub>	CF <sub>3</sub>	<sup>i</sup> Pr	CNC	-Mn-N, -Mn-C
70	OCH <sub>3</sub> -CNC-CF <sub>3</sub> - <sup>i</sup> Pr-9	4	47.4	OCH <sub>3</sub>	CF <sub>3</sub>	<sup>i</sup> Pr	CNC	-Mn-N, -Mn-C
71	OCH <sub>3</sub> -CNC-CF <sub>3</sub> -ph-8	6	68.73	OCH <sub>3</sub>	CF <sub>3</sub>	ph	CNC	-Mn-N, -Mn-C
72	OCH <sub>3</sub> -CNC-CF <sub>3</sub> -ph-9	5	74.91	OCH <sub>3</sub>	CF <sub>3</sub>	ph	CNC	-Mn-N, -Mn-C
73	OCH <sub>3</sub> -CNC-CF <sub>3</sub> -ph-10	5	55.84	OCH <sub>3</sub>	CF <sub>3</sub>	ph	CNC	-Mn-C
74	OCH <sub>3</sub> -CNC-CF <sub>3</sub> -6	6	58.67	OCH <sub>3</sub>	CF <sub>3</sub>	H	CNC	-Mn-N, -Mn-C
75	OCH <sub>3</sub> -CNC-CF <sub>3</sub> -7	5	61.59	OCH <sub>3</sub>	CF <sub>3</sub>	H	CNC	-Mn-N, -Mn-C
76	OCH <sub>3</sub> -CNC-CF <sub>3</sub> -8	4	67.13	OCH <sub>3</sub>	CF <sub>3</sub>	H	CNC	-Mn-N, -Mn-C
77	OCH <sub>3</sub> -CNC-CF <sub>3</sub> -9	6	40.72	OCH <sub>3</sub>	CF <sub>3</sub>	H	CNC	-Mn-N, -Mn-C
78	OCH <sub>3</sub> -CNC-CF <sub>3</sub> -10	6	47.97	OCH <sub>3</sub>	CF <sub>3</sub>	H	CNC	-Mn-N, -Mn-C
79	OCH <sub>3</sub> -CNC-CF <sub>3</sub> -11	5	40.32	OCH <sub>3</sub>	CF <sub>3</sub>	H	CNC	-Mn-N, -Mn-C(2)
80	OCH <sub>3</sub> -CNC-CF <sub>3</sub> -12	6	44.51	OCH <sub>3</sub>	CF <sub>3</sub>	H	CNC	-Mn-N, -Mn-C
81	OCH <sub>3</sub> -CNC-H-ph-5	5	51.96	OCH <sub>3</sub>	H	ph	CNC	-Mn-N, -Mn-C
82	OCH <sub>3</sub> -CNC-H-5	4	56.13	OCH <sub>3</sub>	H	H	CNC	-Mn-N, -Mn-C
83	OCH <sub>3</sub> -CNC-ph-ph-6	6	65.67	OCH <sub>3</sub>	Ph	ph	CNC	-Mn-N, -Mn-C

84	OCH <sub>3</sub> -CNC-ph-ph-7	6	68.51	OCH <sub>3</sub>	ph	ph	CNC	-Mn-N, -Mn-C
85	OCH <sub>3</sub> -CNC- <sup>t</sup> Bu-CF <sub>3</sub> -4	5	9.15	OCH <sub>3</sub>	<sup>t</sup> Bu	CF <sub>3</sub>	CNC	-Mn-C
86	OCH <sub>3</sub> -CNC- <sup>t</sup> Bu-CF <sub>3</sub> -5	5	3.69	OCH <sub>3</sub>	<sup>t</sup> Bu	CF <sub>3</sub>	CNC	-Mn-C
87	OCH <sub>3</sub> -CNC- <sup>t</sup> Bu-CF <sub>3</sub> -6	4	11.77	OCH <sub>3</sub>	<sup>t</sup> Bu	CF <sub>3</sub>	CNC	-Mn-N, -Mn-C
88	OCH <sub>3</sub> -CNC- <sup>t</sup> Bu-ph-2	5	7.27	OCH <sub>3</sub>	<sup>t</sup> Bu	ph	CNC	-Mn-C
89	OH-CNC-CF <sub>3</sub> -cy-5	5	63.25	OH	CF <sub>3</sub>	cy	CNC	-Mn-N, -Mn-C
90	OH-CNC-CF <sub>3</sub> -ph-3	5	19.31	OH	CF <sub>3</sub>	ph	CNC	-Mn-N, -Mn-C
91	OH-CNC-CF <sub>3</sub> -ph-4	5	32.51	OH	CF <sub>3</sub>	ph	CNC	-Mn-N, -Mn-C
92	OH-CNC-CF <sub>3</sub> -ph-11	5	70.2	OH	CF <sub>3</sub>	ph	CNC	-Mn-N, -Mn-C
93	OH-CNC-CF <sub>3</sub> -ph-12	6	82.39	OH	CF <sub>3</sub>	ph	CNC	-Mn-N, -Mn-C
94	OH-CNC-CF <sub>3</sub> -5	6	68.73	OH	CF <sub>3</sub>	H	CNC	-Mn-C
95	OH-CNC-H- <sup>i</sup> Pr-6	6	87.29	OH	H	<sup>i</sup> Pr	CNC	-Mn-N, -Mn-C
96	OH-CNC-H- <sup>i</sup> Pr-7	5	111.13	OH	H	<sup>i</sup> Pr	CNC	-Mn-N, -Mn-C(2)
97	OH-CNC-H- <sup>i</sup> Pr-8	7	85.81	OH	H	<sup>i</sup> Pr	CNC	-Mn-N, -Mn-C
98	OH-CNC-H- <sup>i</sup> Pr-9	6	105.24	OH	H	<sup>i</sup> Pr	CNC	-Mn-N, -Mn-C
99	OH-CNC-H-6	5	80.06	OH	H	H	CNC	-Mn-C
100	OH-CNC-ph-ph-9	5	66.12	OH	ph	ph	CNC	-Mn-C
101	OH-CNC-ph-ph-10	5	67.68	OH	ph	ph	CNC	-Mn-N, -Mn-C
102	OH-CNC-ph-ph-11	4	67.99	OH	ph	ph	CNC	-Mn-N, -Mn-C
103	OH-CNC- <sup>t</sup> Bu- <sup>i</sup> Pr-5	6	39.97	OH	<sup>t</sup> Bu	<sup>i</sup> Pr	CNC	-Mn-C
104	OH-CNC- <sup>t</sup> Bu- <sup>i</sup> Pr-6	4	39.58	OH	<sup>t</sup> Bu	<sup>i</sup> Pr	CNC	-Mn-N, -Mn-C
105	OH-CNC- <sup>t</sup> Bu- <sup>i</sup> Pr-7	5	37.25	OH	<sup>t</sup> Bu	<sup>i</sup> Pr	CNC	-Mn-N, -Mn-C
106	OH-CNC- <sup>t</sup> Bu-ph-5	5	44.95	OH	<sup>t</sup> Bu	ph	CNC	-Mn-C
107	OH-CNC- <sup>t</sup> Bu-ph-6	5	-6.63	OH	<sup>t</sup> Bu	ph	CNC	-Mn-C

Table B.2.: Exploration results for structures with “-N-H, +H-O interaction”

ID	CN	$\Delta E(\text{kcal.mol}^{-1})$	Adduct	R <sub>1</sub>	R <sub>2</sub>	backbone	Interaction	
1	OCH <sub>3</sub> -CNC-H-CF <sub>3</sub> -3	5	-0.36	OCH <sub>3</sub>	H	CF <sub>3</sub>	CNC	-N-H, +H-O
2	OH-CNC-CF <sub>3</sub> -CF <sub>3</sub> -2	6	-12.93	OH	CF <sub>3</sub>	CF <sub>3</sub>	CNC	-N-H, +H-O
3	OH-CNC-CF <sub>3</sub> -CF <sub>3</sub> -3	6	-17.89	OH	CF <sub>3</sub>	CF <sub>3</sub>	CNC	-N-H, +H-O, +Mn-F
4	OH-CNC-cy-CF <sub>3</sub> -2	5	-25.09	OH	cy	CF <sub>3</sub>	CNC	-N-H, +H-O
5	OH-CNC-cy-CF <sub>3</sub> -3	6	-25.89	OH	cy	CF <sub>3</sub>	CNC	-N-H, +H-O, +Mn-F
6	OH-CNC-H-CF <sub>3</sub> -2	6	-13.07	OH	H	CF <sub>3</sub>	CNC	-N-H, +H-O
7	OH-CNC-H-CF <sub>3</sub> -3	5	-10.72	OH	H	CF <sub>3</sub>	CNC	-N-H, +H-O
8	OH-CNC-H-10	6	-3.46	OH	H	H	CNC	-N-H, +H-O
9	OH-CNC- <i>i</i> Pr-CF <sub>3</sub> -2	6	-14.08	OH	<i>i</i> Pr	CF <sub>3</sub>	CNC	-N-H, +H-O
10	OH-CNC- <i>i</i> Pr-CF <sub>3</sub> -3	6	-12.42	OH	<i>i</i> Pr	CF <sub>3</sub>	CNC	-N-H, +H-O, +Mn-F
11	OH-CNC- <i>i</i> Pr-CF <sub>3</sub> -5	5	-6.79	OH	<i>i</i> Pr	CF <sub>3</sub>	CNC	-N-H, +H-O
12	OH-CNC-ph-CF <sub>3</sub> -2	6	-18.51	OH	ph	CF <sub>3</sub>	CNC	-N-H, +H-O
13	OH-CNC-ph-CF <sub>3</sub> -3	5	-19.82	OH	ph	CF <sub>3</sub>	CNC	-N-H, +H-O
14	OH-CNC- <i>t</i> Bu-CF <sub>3</sub> -2	6	-17.88	OH	<i>t</i> Bu	CF <sub>3</sub>	CNC	-N-H, +H-O

Table B.3.: Nucleophilic (+C-O) interactions observed for CNC structures with different R<sub>1</sub>-R<sub>2</sub>-Adduct combinations

ID	CN	$\Delta E(\text{kcal.mol}^{-1})$	Adduct	R <sub>1</sub>	R <sub>2</sub>	backbone	Interaction	
1	OCH <sub>3</sub> -CNC-CF <sub>3</sub> - <i>t</i> Bu-2	5	-4.4	OCH <sub>3</sub>	CNC	CF <sub>3</sub>	<i>t</i> Bu	+C-O
2	OCH <sub>3</sub> -CNC-cy- <i>i</i> Pr-2	5	-0.96	OCH <sub>3</sub>	CNC	cy	<i>i</i> Pr	+C-O
3	OCH <sub>3</sub> -CNC- <i>i</i> Pr-cy-2	5	-13.84	OCH <sub>3</sub>	CNC	<i>i</i> Pr	Cy	+C-O
4	OCH <sub>3</sub> -CNC- <i>t</i> Bu- <i>i</i> Pr-2	5	-16.76	OCH <sub>3</sub>	CNC	<i>t</i> Bu	<i>i</i> Pr	+C-O
5	OH-CNC-CF <sub>3</sub> - <i>t</i> Bu-6	5	-5.9	OH	CNC	CF <sub>3</sub>	<i>t</i> Bu	+C-O
6	OH-CNC- <i>i</i> Pr- <i>i</i> Pr-2	5	-11.37	OH	CNC	<i>i</i> Pr	<i>i</i> Pr	+C-O
7	OH-CNC-ph- <i>i</i> Pr-2	5	-3.9	OH	CNC	Ph	<i>i</i> Pr	+C-O
8	OH-CNC-ph- <i>i</i> Pr-3	6	-4.96	OH	CNC	Ph	<i>i</i> Pr	+C-O

9	OH-CNC- <sup>t</sup> Bu-cy-2	6	-9.63	OH	CNC	<sup>t</sup> Bu	Cy	+C-O
10	OH-CNC- <sup>t</sup> Bu- <sup>t</sup> Bu-3	5	-10.75	OH	CNC	<sup>t</sup> Bu	<sup>t</sup> Bu	+C-O
11	OH-CNC- <sup>t</sup> Bu-2	6	-3.19	OH	CNC	<sup>t</sup> Bu	H	+C-O
12	O <sup>t</sup> Bu-PNP-CF <sub>3</sub> - <sup>t</sup> Bu-3	5	-6.6	O <sup>t</sup> Bu	PNP	CF <sub>3</sub>	<sup>t</sup> Bu	+C-O
13	OCH <sub>3</sub> -CNC- <sup>t</sup> Bu- <sup>t</sup> Bu-2	6	-11.2	OCH <sub>3</sub>	CNC	<sup>t</sup> Bu	<sup>t</sup> Bu	+C-O, +Mn-H
14	OH-CNC-CF <sub>3</sub> - <sup>t</sup> Bu-2	6	-9.19	OH	CNC	CF <sub>3</sub>	<sup>t</sup> Bu	+C-O, +Mn-H
15	OH-CNC-cy- <sup>t</sup> Bu-3	6	-3.14	OH	CNC	cy	<sup>t</sup> Bu	+C-O, +Mn-H
16	OH-CNC- <sup>i</sup> Pr- <sup>t</sup> Bu-2	6	-12.06	OH	CNC	<sup>i</sup> Pr	<sup>t</sup> Bu	+C-O, +Mn-H
17	OH-CNC-ph- <sup>t</sup> Bu-2	6	-7.18	OH	CNC	ph	<sup>t</sup> Bu	+C-O, +Mn-H
18	OH-CNC- <sup>t</sup> Bu- <sup>t</sup> Bu-2	6	-5.38	OH	CNC	<sup>t</sup> Bu	<sup>t</sup> Bu	+C-O, +Mn-H
19	O <sup>t</sup> Bu-PNP-CF <sub>3</sub> - <sup>t</sup> Bu-4	6	-11.54	O <sup>t</sup> Bu	PNP	CF <sub>3</sub>	<sup>t</sup> Bu	+C-O, +Mn-H
20	OH-CNC- <sup>t</sup> Bu-ph-6	5	-6.63	OH	CNC	<sup>t</sup> Bu	Ph	-Mn-C, +C-O, +C-H, -O-H
21	O <sup>t</sup> Bu-CNC-cy- <sup>t</sup> Bu-3	5	-2.23	O <sup>t</sup> Bu	CNC	Cy	<sup>t</sup> Bu	-Mn-N, +C-O
22	O <sup>t</sup> Bu-CNC-cy- <sup>t</sup> Bu-5	4	-7.64	O <sup>t</sup> Bu	CNC	Cy	<sup>t</sup> Bu	-Mn-N, +C-O
23	OCH <sub>3</sub> -CNC- <sup>t</sup> Bu- <sup>t</sup> Bu-5	4	-3.94	OCH <sub>3</sub>	CNC	<sup>t</sup> Bu	<sup>t</sup> Bu	-Mn-N, +C-O
24	OH-CNC- <sup>i</sup> Pr- <sup>t</sup> Bu-6	4	-0.43	OH	CNC	<sup>i</sup> Pr	<sup>t</sup> Bu	-Mn-N, +C-O
25	O <sup>t</sup> Bu-CNC-cy- <sup>t</sup> Bu-4	6	-5.15	O <sup>t</sup> Bu	CNC	Cy	<sup>t</sup> Bu	-Mn-N, +C-O, +Mn-H
26	O <sup>t</sup> Bu-CNC- <sup>t</sup> Bu- <sup>t</sup> Bu-5	5	-5.83	O <sup>t</sup> Bu	CNC	<sup>t</sup> Bu	<sup>t</sup> Bu	-Mn-N, +C-O, +Mn-H
27	OCH <sub>3</sub> -CNC- <sup>t</sup> Bu- <sup>t</sup> Bu-3	5	-9.18	OCH <sub>3</sub>	CNC	<sup>t</sup> Bu	<sup>t</sup> Bu	-Mn-N, +C-O, +Mn-H
28	OCH <sub>3</sub> -CNC- <sup>t</sup> Bu- <sup>t</sup> Bu-4	5	-3.28	OCH <sub>3</sub>	CNC	<sup>t</sup> Bu	<sup>t</sup> Bu	-Mn-N, +C-O, +Mn-H
29	OCH <sub>3</sub> -CNC- <sup>t</sup> Bu- <sup>t</sup> Bu-7	6	-5.06	OCH <sub>3</sub>	CNC	<sup>t</sup> Bu	<sup>t</sup> Bu	-Mn-N, +C-O, +Mn-H
30	OH-CNC- <sup>t</sup> Bu- <sup>t</sup> Bu-4	5	-3.9	OH	CNC	<sup>t</sup> Bu	<sup>t</sup> Bu	-Mn-N, +C-O, +Mn-H
31	OH-CNC- <sup>t</sup> Bu- <sup>t</sup> Bu-5	5	-11.72	OH	CNC	<sup>t</sup> Bu	<sup>t</sup> Bu	-Mn-N, +C-O, +Mn-H
32	O <sup>t</sup> Bu-CNC- <sup>t</sup> Bu- <sup>t</sup> Bu-6	6	-7.83	O <sup>t</sup> Bu	CNC	<sup>t</sup> Bu	<sup>t</sup> Bu	-Mn-N, +C-O, +Mn-H, +Mn-H

Table B.4.: Nucleophilic (+C-C) interactions observed for CNC structures with different R<sub>1</sub>-R<sub>2</sub>-Adduct combinations"

ID	CN	$\Delta E(\text{kcal.mol}^{-1})$	Adduct	R <sub>1</sub>	R <sub>2</sub>	backbone	Interaction
1	Br-CNC- <sup>t</sup> Bu-cy-3	4	-2.93	Br	CNC	<sup>t</sup> Bu cy	-Mn-N, -Mn-C, +C-C
2	Br-CNC- <sup>t</sup> Bu- <sup>t</sup> Bu-3	5	-26.92	Br	CNC	<sup>t</sup> Bu <sup>t</sup> Bu	-Mn-N, -Mn-C, +C-C, +Mn-H
3	O <sup>t</sup> Bu-CNC- <sup>t</sup> Bu- <sup>t</sup> Bu-4	4	-4.24	O <sup>t</sup> Bu	CNC	<sup>t</sup> Bu <sup>t</sup> Bu	-Mn-N, -Mn-C, +C-C

Table B.5.: Formation of Manganese carbonyl trifluoromethyl complexes with PNP

ID	CN	$\Delta E(\text{kcal.mol}^{-1})$	Adduct	R <sub>1</sub>	R <sub>2</sub>	backbone	Interaction
1	O <sup>t</sup> Bu-PNP-CF <sub>3</sub> -4	6	-39.2	O <sup>t</sup> Bu	PNP	CF <sub>3</sub> H	+Mn-C, -P-C, +P-O
2	OCH <sub>3</sub> -PNP-CF <sub>3</sub> - <sup>i</sup> Pr-4	6	-9.56	OCH <sub>3</sub>	PNP	CF <sub>3</sub> <sup>i</sup> Pr	+Mn-C, -P-C, +P-O
3	OCH <sub>3</sub> -PNP-CF <sub>3</sub> - <sup>t</sup> Bu-3	6	-0.83	OCH <sub>3</sub>	PNP	CF <sub>3</sub> <sup>t</sup> Bu	+Mn-C, -P-C, +P-O
4	OCH <sub>3</sub> -PNP-CF <sub>3</sub> -4	6	-23.05	OCH <sub>3</sub>	PNP	CF <sub>3</sub> H	+Mn-C, -P-C, +P-O
5	OH-PNP-CF <sub>3</sub> -CF <sub>3</sub> -7	6	-9.43	OH	PNP	CF <sub>3</sub> CF <sub>3</sub>	+Mn-C, -P-C, +P-O
6	OH-PNP-CF <sub>3</sub> -cy-5	6	-8.18	OH	PNP	CF <sub>3</sub> Cy	+Mn-C, -P-C, +P-O
7	OH-PNP-CF <sub>3</sub> -ph-7	6	-17.16	OH	PNP	CF <sub>3</sub> ph	+Mn-C, -P-C, +P-O
8	O <sup>t</sup> Bu-PNP-CF <sub>3</sub> -2	6	-26.52	O <sup>t</sup> Bu	PNP	CF <sub>3</sub> H	+P-O
9	OCH <sub>3</sub> -PNP-CF <sub>3</sub> -2	6	-11.81	OCH <sub>3</sub>	PNP	CF <sub>3</sub> H	+P-O
10	OH-PNP-CF <sub>3</sub> -ph-3	6	-3.32	OH	PNP	CF <sub>3</sub> ph	+P-O
11	OH-PNP-CF <sub>3</sub> -2	6	-10.84	OH	PNP	CF <sub>3</sub> H	+P-O
12	O <sup>t</sup> Bu-PNP-CF <sub>3</sub> -3	6	-6.72	O <sup>t</sup> Bu	PNP	CF <sub>3</sub> H	+P-O, +Mn-F
13	OH-PNP-CF <sub>3</sub> -15	5	-3.58	OH	PNP	CF <sub>3</sub> H	-Mn-N, +Mn-C, +N-H, -P-C, +P-O, -O-H
14	O <sup>t</sup> Bu-PNP-CF <sub>3</sub> -5	5	-13.03	O <sup>t</sup> Bu	PNP	CF <sub>3</sub> H	-Mn-N, +Mn-C, +N-P, -P-C, +P-O
15	OH-PNP-CF <sub>3</sub> - <sup>i</sup> Pr-9	6	-4.46	OH	PNP	CF <sub>3</sub> <sup>i</sup> Pr	-Mn-N, +Mn-C, +N-P, -P-C, +P-O
16	O <sup>t</sup> Bu-PNP-CF <sub>3</sub> -CF <sub>3</sub> -3	5	-4.63	O <sup>t</sup> Bu	PNP	CF <sub>3</sub> CF <sub>3</sub>	-Mn-N, +Mn-C, -P-C, +P-O
17	O <sup>t</sup> Bu-PNP-CF <sub>3</sub> -8	5	-17.65	O <sup>t</sup> Bu	PNP	CF <sub>3</sub> H	-Mn-N, +Mn-C, -P-C, +P-O
18	OH-PNP-CF <sub>3</sub> - <sup>t</sup> Bu-13	5	-0.6	OH	PNP	CF <sub>3</sub> <sup>t</sup> Bu	-Mn-N, +Mn-C, -P-C, +P-O

19	OH-PNP-CF <sub>3</sub> -21	5	-1.55	OH	PNP	CF <sub>3</sub>	H	-Mn-N, -Mn-P, +Mn-C, +N-H, -P-C, +P-O, -O-H
20	OH-PNP-CF <sub>3</sub> -cy-2	6	-0.19	OH	PNP	CF <sub>3</sub>	cy	-Mn-N, -Mn-P, +Mn-C, +N-H, -P-C, +P-O, -O-H, +Mn-F
21	O <sup>t</sup> Bu-PNP-CF <sub>3</sub> -6	5	-10.97	O <sup>t</sup> Bu	PNP	CF <sub>3</sub>	H	-Mn-P, +Mn-C, -P-C, +P-O

Table B.6.: Summary of data on reactivity exploration for species lower energies than their respective reference structures

ID	CN	$\Delta E(\text{kcal.mol}^{-1})$	Adduct	R <sub>1</sub>	R <sub>2</sub>	backbone	Interaction
1	5	-43.77	Br	<sup>t</sup> Bu	cy	CNC	-Mn-N
2	4	-2.93	Br	<sup>t</sup> Bu	cy	CNC	-Mn-N, -Mn-C, +C-C
3	5	-68.79	Br	<sup>t</sup> Bu	<sup>t</sup> Bu	CNC	-Mn-N
4	5	-26.92	Br	<sup>t</sup> Bu	<sup>t</sup> Bu	CNC	-Mn-N, -Mn-C, +C-C, +Mn-H
5	5	-32.46	O <sup>t</sup> Bu	cy	cy	CNC	-Mn-N
6	5	-31.68	O <sup>t</sup> Bu	cy	<sup>t</sup> Bu	CNC	-Mn-N
7	5	-2.23	O <sup>t</sup> Bu	cy	<sup>t</sup> Bu	CNC	-Mn-N, +C-O, +Mn-O, +Mn-O
8	6	-5.15	O <sup>t</sup> Bu	cy	<sup>t</sup> Bu	CNC	-Mn-N, +C-O, +Mn-H
9	4	-7.64	O <sup>t</sup> Bu	cy	<sup>t</sup> Bu	CNC	-Mn-N, +C-O, +Mn-O
10	5	-6.4	O <sup>t</sup> Bu	ph	cy	CNC	-Mn-N
11	5	-0.28	O <sup>t</sup> Bu	<sup>t</sup> Bu	CF <sub>3</sub>	CNC	-Mn-C, -N-H, +C-H
12	5	-18.05	O <sup>t</sup> Bu	<sup>t</sup> Bu	ph	CNC	-Mn-C
13	5	-29.85	O <sup>t</sup> Bu	<sup>t</sup> Bu	<sup>t</sup> Bu	CNC	-Mn-N
14	4	-18.33	O <sup>t</sup> Bu	<sup>t</sup> Bu	<sup>t</sup> Bu	CNC	-Mn-N, -Mn-C
15	4	-4.24	O <sup>t</sup> Bu	<sup>t</sup> Bu	<sup>t</sup> Bu	CNC	-Mn-N, -Mn-C, +C-C
16	5	-5.83	O <sup>t</sup> Bu	<sup>t</sup> Bu	<sup>t</sup> Bu	CNC	-Mn-N, +C-O, +Mn-H, +Mn-O
17	6	-7.83	O <sup>t</sup> Bu	<sup>t</sup> Bu	<sup>t</sup> Bu	CNC	-Mn-N, +C-O, +Mn-H, +Mn-H, +Mn-O
18	5	-4.4	OCH <sub>3</sub>	CF <sub>3</sub>	<sup>t</sup> Bu	CNC	+C-O, +Mn-O

19	OCH <sub>3</sub> -CNC-cy- <sup>i</sup> Pr-2	5	-0.96	OCH <sub>3</sub>	cy	<sup>i</sup> Pr	CNC	+C-O, +Mn-O
20	OCH <sub>3</sub> -CNC-cy-ph-2	5	-53.74	OCH <sub>3</sub>	cy	ph	CNC	+C-O, +Mn-O
21	OCH <sub>3</sub> -CNC-H-CF <sub>3</sub> -3	5	-0.36	OCH <sub>3</sub>	H	CF <sub>3</sub>	CNC	-N-H, +H-O, +Mn-O
22	OCH <sub>3</sub> -CNC-H- <sup>t</sup> Bu-2	5	-91.33	OCH <sub>3</sub>	H	<sup>t</sup> Bu	CNC	-Mn-N
23	OCH <sub>3</sub> -CNC-H- <sup>t</sup> Bu-3	5	-65.19	OCH <sub>3</sub>	H	<sup>t</sup> Bu	CNC	-Mn-N, +C-O, +Mn-O, +Mn-O
24	OCH <sub>3</sub> -CNC- <sup>i</sup> Pr-cy-2	5	-13.84	OCH <sub>3</sub>	<sup>i</sup> Pr	cy	CNC	+C-O, +Mn-O
25	OCH <sub>3</sub> -CNC-ph- <sup>t</sup> Bu-2	6	-82.1	OCH <sub>3</sub>	ph	<sup>t</sup> Bu	CNC	+C-O, +Mn-H, +Mn-O
26	OCH <sub>3</sub> -CNC-ph- <sup>t</sup> Bu-3	5	-51.97	OCH <sub>3</sub>	ph	<sup>t</sup> Bu	CNC	-Mn-N, +Mn-C, +C-O, +Mn-O
27	OCH <sub>3</sub> -CNC-ph- <sup>t</sup> Bu-4	6	-65.3	OCH <sub>3</sub>	ph	<sup>t</sup> Bu	CNC	-Mn-N, +C-O, +Mn-O, +Mn-H, +Mn-O
28	OCH <sub>3</sub> -CNC- <sup>t</sup> Bu- <sup>i</sup> Pr-2	5	-16.76	OCH <sub>3</sub>	<sup>t</sup> Bu	<sup>i</sup> Pr	CNC	+C-O, +Mn-O
29	OCH <sub>3</sub> -CNC- <sup>t</sup> Bu- <sup>t</sup> Bu-2	6	-11.2	OCH <sub>3</sub>	<sup>t</sup> Bu	<sup>t</sup> Bu	CNC	+C-O, +Mn-H, +Mn-O
30	OCH <sub>3</sub> -CNC- <sup>t</sup> Bu- <sup>t</sup> Bu-3	5	-9.18	OCH <sub>3</sub>	<sup>t</sup> Bu	<sup>t</sup> Bu	CNC	-Mn-N, +C-O, +Mn-H, +Mn-O
31	OCH <sub>3</sub> -CNC- <sup>t</sup> Bu- <sup>t</sup> Bu-4	5	-3.28	OCH <sub>3</sub>	<sup>t</sup> Bu	<sup>t</sup> Bu	CNC	-Mn-N, +C-O, +Mn-H, +Mn-O
32	OCH <sub>3</sub> -CNC- <sup>t</sup> Bu- <sup>t</sup> Bu-5	4	-3.94	OCH <sub>3</sub>	<sup>t</sup> Bu	<sup>t</sup> Bu	CNC	-Mn-N, +C-O, +Mn-O
33	OCH <sub>3</sub> -CNC- <sup>t</sup> Bu- <sup>t</sup> Bu-6	5	-28.57	OCH <sub>3</sub>	<sup>t</sup> Bu	<sup>t</sup> Bu	CNC	-Mn-N
34	OCH <sub>3</sub> -CNC- <sup>t</sup> Bu- <sup>t</sup> Bu-7	6	-5.06	OCH <sub>3</sub>	<sup>t</sup> Bu	<sup>t</sup> Bu	CNC	-Mn-N, +C-O, +Mn-O, +Mn-H, +Mn-O
35	OH-CNC-CF <sub>3</sub> -CF <sub>3</sub> -2	6	-12.93	OH	CF <sub>3</sub>	CF <sub>3</sub>	CNC	-N-H, +H-O
36	OH-CNC-CF <sub>3</sub> -CF <sub>3</sub> -3	6	-17.89	OH	CF <sub>3</sub>	CF <sub>3</sub>	CNC	-N-H, +H-O, +Mn-F, +Mn-O
37	OH-CNC-CF <sub>3</sub> - <sup>t</sup> Bu-2	6	-9.19	OH	CF <sub>3</sub>	<sup>t</sup> Bu	CNC	+C-O, +Mn-H, +Mn-O
38	OH-CNC-CF <sub>3</sub> - <sup>t</sup> Bu-3	5	-15.35	OH	CF <sub>3</sub>	<sup>t</sup> Bu	CNC	-Mn-N
39	OH-CNC-CF <sub>3</sub> - <sup>t</sup> Bu-6	5	-5.9	OH	CF <sub>3</sub>	<sup>t</sup> Bu	CNC	+C-O, +Mn-O
40	OH-CNC-cy-CF <sub>3</sub> -2	5	-25.09	OH	cy	CF <sub>3</sub>	CNC	-N-H, +H-O, +Mn-O
41	OH-CNC-cy-CF <sub>3</sub> -3	6	-25.89	OH	cy	CF <sub>3</sub>	CNC	-N-H, +H-O, +Mn-F, +Mn-O
42	OH-CNC-cy- <sup>t</sup> Bu-3	6	-3.14	OH	cy	<sup>t</sup> Bu	CNC	+C-O, +Mn-H, +Mn-O
43	OH-CNC-H-CF <sub>3</sub> -2	6	-13.07	OH	H	CF <sub>3</sub>	CNC	-N-H, +H-O
44	OH-CNC-H-CF <sub>3</sub> -3	5	-10.72	OH	H	CF <sub>3</sub>	CNC	-N-H, +H-O, +Mn-O
45	OH-CNC-H-10	6	-3.46	OH	H	H	CNC	-N-H, +H-O

46	OH-CNC- <i>i</i> Pr-CF <sub>3</sub> -2	6	-14.08	OH	<i>i</i> Pr	CF <sub>3</sub>	CNC	-N-H, +H-O
47	OH-CNC- <i>i</i> Pr-CF <sub>3</sub> -3	6	-12.42	OH	<i>i</i> Pr	CF <sub>3</sub>	CNC	-N-H, +H-O, +Mn-F, +Mn-O
48	OH-CNC- <i>i</i> Pr-CF <sub>3</sub> -5	5	-6.79	OH	<i>i</i> Pr	CF <sub>3</sub>	CNC	-N-H, +H-O, +Mn-O
49	OH-CNC- <i>i</i> Pr- <i>i</i> Pr-2	5	-11.37	OH	<i>i</i> Pr	<i>i</i> Pr	CNC	+C-O, +Mn-O
50	OH-CNC- <i>i</i> Pr- <i>t</i> Bu-2	6	-12.06	OH	<i>i</i> Pr	<i>t</i> Bu	CNC	+C-O, +Mn-H, +Mn-O
51	OH-CNC- <i>i</i> Pr- <i>t</i> Bu-6	4	-0.43	OH	<i>i</i> Pr	<i>t</i> Bu	CNC	-Mn-N, +C-O, +Mn-O
52	OH-CNC-ph-CF <sub>3</sub> -2	6	-18.51	OH	ph	CF <sub>3</sub>	CNC	-N-H, +H-O
53	OH-CNC-ph-CF <sub>3</sub> -3	5	-19.82	OH	ph	CF <sub>3</sub>	CNC	-N-H, +H-O, +Mn-O
54	OH-CNC-ph- <i>i</i> Pr-2	5	-3.9	OH	ph	<i>i</i> Pr	CNC	+C-O, +Mn-O
55	OH-CNC-ph- <i>i</i> Pr-3	6	-4.96	OH	ph	<i>i</i> Pr	CNC	+C-O, +Mn-O, +Mn-O
56	OH-CNC-ph- <i>t</i> Bu-2	6	-7.18	OH	ph	<i>t</i> Bu	CNC	+C-O, +Mn-H, +Mn-O
57	OH-CNC- <i>t</i> Bu-CF <sub>3</sub> -2	6	-17.88	OH	<i>t</i> Bu	CF <sub>3</sub>	CNC	-N-H, +H-O
58	OH-CNC- <i>t</i> Bu-cy-2	6	-9.63	OH	<i>t</i> Bu	cy	CNC	+C-O, +Mn-O, +Mn-O
59	OH-CNC- <i>t</i> Bu-ph-6	5	-6.63	OH	<i>t</i> Bu	ph	CNC	-Mn-C, +C-O, +C-H, -O-H
60	OH-CNC- <i>t</i> Bu- <i>t</i> Bu-2	6	-5.38	OH	<i>t</i> Bu	<i>t</i> Bu	CNC	+C-O, +Mn-H, +Mn-O
61	OH-CNC- <i>t</i> Bu- <i>t</i> Bu-3	5	-10.75	OH	<i>t</i> Bu	<i>t</i> Bu	CNC	+C-O, +Mn-O
62	OH-CNC- <i>t</i> Bu- <i>t</i> Bu-4	5	-3.9	OH	<i>t</i> Bu	<i>t</i> Bu	CNC	-Mn-N, +C-O, +Mn-H, +Mn-O
63	OH-CNC- <i>t</i> Bu- <i>t</i> Bu-5	5	-11.72	OH	<i>t</i> Bu	<i>t</i> Bu	CNC	-Mn-N, +C-O, +Mn-H, +Mn-O
64	OH-CNC- <i>t</i> Bu-2	6	-3.19	OH	<i>t</i> Bu	H	CNC	+C-O, +Mn-O, +Mn-O
65	Br-PNN-CF <sub>3</sub> -CF <sub>3</sub> -2	6	-3.7	Br	CF <sub>3</sub>	CF <sub>3</sub>	PNN	-Mn-N, +Mn-F
66	Br-PNN-CF <sub>3</sub> - <i>i</i> Pr-2	5	-15.95	Br	CF <sub>3</sub>	<i>i</i> Pr	PNN	-Mn-N
67	Br-PNN-CF <sub>3</sub> -ph-2	6	-12.85	Br	CF <sub>3</sub>	ph	PNN	-Mn-N, +Mn-C
68	Br-PNN-CF <sub>3</sub> -ph-8	5	-5.84	Br	CF <sub>3</sub>	ph	PNN	-Mn-N
69	Br-PNN-cy-CF <sub>3</sub> -2	4	-3.04	Br	cy	CF <sub>3</sub>	PNN	-Mn-N, -Mn-N
70	Br-PNN-cy-cy-2	5	-77.33	Br	cy	cy	PNN	-Mn-N
71	Br-PNN-cy-2	5	-67.76	Br	cy	H	PNN	-Mn-N
72	Br-PNN- <i>i</i> Pr-cy-2	5	-77.21	Br	<i>i</i> Pr	cy	PNN	-Mn-N
73	Br-PNN- <i>i</i> Pr-2	5	-64.71	Br	<i>i</i> Pr	H	PNN	-Mn-N
74	O <sup><i>t</i></sup> Bu-PNN-cy-3	4	-39.97	O <sup><i>t</i></sup> Bu	cy	H	PNN	-Mn-N, -Mn-P, +Mn-H,

75	O <sup>t</sup> Bu-PNN- <i>i</i> Pr-3	4	-55.1	O <sup>t</sup> Bu	<i>i</i> Pr	H	PNN	-Mn-N, -N-H, +H-P, -C-H -Mn-N, -Mn-P, +Mn-H, -Mn-N, +C-P, -C-H
76	OCH <sub>3</sub> -PNN-cy- <i>i</i> Pr-2	5	-95.61	OCH <sub>3</sub>	cy	<i>i</i> Pr	PNN	-Mn-N
77	OCH <sub>3</sub> -PNN-cy- <i>i</i> Pr-3	4	-28.36	OCH <sub>3</sub>	cy	<i>i</i> Pr	PNN	-Mn-N, -Mn-P, +Mn-H, -Mn-N, -C-H
78	OCH <sub>3</sub> -PNN-cy- <sup>t</sup> Bu-2	5	-93.18	OCH <sub>3</sub>	cy	<sup>t</sup> Bu	PNN	-Mn-N
79	OCH <sub>3</sub> -PNN-cy-2	5	-96.9	OCH <sub>3</sub>	cy	H	PNN	-Mn-N
80	OCH <sub>3</sub> -PNN- <i>i</i> Pr-cy-2	5	-80.04	OCH <sub>3</sub>	<i>i</i> Pr	cy	PNN	-Mn-N
81	OCH <sub>3</sub> -PNN- <i>i</i> Pr- <i>i</i> Pr-2	5	-84.9	OCH <sub>3</sub>	<i>i</i> Pr	<i>i</i> Pr	PNN	-Mn-N
82	OCH <sub>3</sub> -PNN- <i>i</i> Pr-ph-2	5	-86.96	OCH <sub>3</sub>	<i>i</i> Pr	ph	PNN	-Mn-N
83	OCH <sub>3</sub> -PNN- <i>i</i> Pr-2	5	-92.45	OCH <sub>3</sub>	<i>i</i> Pr	H	PNN	-Mn-N
84	OCH <sub>3</sub> -PNN- <sup>t</sup> Bu-3	4	-70.24	OCH <sub>3</sub>	<sup>t</sup> Bu	H	PNN	-Mn-N, -Mn-P, +Mn-H, -Mn-N, +C-P, -C-H
85	OH-PNN-cy-ph-2	5	-98.58	OH	cy	ph	PNN	-Mn-N
86	OH-PNN-cy- <sup>t</sup> Bu-2	5	-87.09	OH	cy	<sup>t</sup> Bu	PNN	-Mn-N
87	OH-PNN-cy-2	5	-77.36	OH	cy	H	PNN	-Mn-N
88	OH-PNN- <i>i</i> Pr-cy-2	5	-77.57	OH	<i>i</i> Pr	cy	PNN	-Mn-N
89	OH-PNN- <i>i</i> Pr- <i>i</i> Pr-2	5	-78.23	OH	<i>i</i> Pr	<i>i</i> Pr	PNN	-Mn-N
90	OH-PNN- <i>i</i> Pr- <sup>t</sup> Bu-2	5	-91.97	OH	<i>i</i> Pr	<sup>t</sup> Bu	PNN	-Mn-N
91	OH-PNN- <i>i</i> Pr-2	5	-84.61	OH	<i>i</i> Pr	H	PNN	-Mn-N
92	OH-PNN-ph-2	5	-97.91	OH	ph	H	PNN	-Mn-N
93	OH-PNN- <sup>t</sup> Bu-CF <sub>3</sub> -16	6	-97.18	OH	<sup>t</sup> Bu	CF <sub>3</sub>	PNN	-Mn-N, -Mn-C, +Mn-C, +Mn-H, -Mn-N, +C-C,+C-O, -C-H, +Mn-O, +Mn-F, +Mn-O
94	Br-PNP- <sup>t</sup> Bu-ph-2	5	-1.09	Br	<sup>t</sup> Bu	ph	PNP	-Mn-P
95	Br-PNP- <sup>t</sup> Bu- <sup>t</sup> Bu-2	5	-77.09	Br	<sup>t</sup> Bu	<sup>t</sup> Bu	PNP	-Mn-N
96	O <sup>t</sup> Bu-PNP-CF <sub>3</sub> -CF <sub>3</sub> -3	5	-4.63	O <sup>t</sup> Bu	CF <sub>3</sub>	CF <sub>3</sub>	PNP	-Mn-N, +Mn-C, -P-C, +P-O, +Mn-O
97	O <sup>t</sup> Bu-PNP-CF <sub>3</sub> -ph-8	5	-0.96	O <sup>t</sup> Bu	CF <sub>3</sub>	ph	PNP	-Mn-P

98	O <sup>t</sup> Bu-PNP-CF <sub>3</sub> - <sup>t</sup> Bu-2	5	-21.25	O <sup>t</sup> Bu	CF <sub>3</sub>	<sup>t</sup> Bu	PNP	-Mn-N
99	O <sup>t</sup> Bu-PNP-CF <sub>3</sub> - <sup>t</sup> Bu-3	5	-6.6	O <sup>t</sup> Bu	CF <sub>3</sub>	<sup>t</sup> Bu	PNP	+C-O, +Mn-O
100	O <sup>t</sup> Bu-PNP-CF <sub>3</sub> - <sup>t</sup> Bu-4	6	-11.54	O <sup>t</sup> Bu	CF <sub>3</sub>	<sup>t</sup> Bu	PNP	+C-O, +Mn-H, +Mn-O
101	O <sup>t</sup> Bu-PNP-CF <sub>3</sub> -2	6	-26.52	O <sup>t</sup> Bu	CF <sub>3</sub>	H	PNP	+P-O
102	O <sup>t</sup> Bu-PNP-CF <sub>3</sub> -3	6	-6.72	O <sup>t</sup> Bu	CF <sub>3</sub>	H	PNP	+P-O, +Mn-F, +Mn-O
103	O <sup>t</sup> Bu-PNP-CF <sub>3</sub> -4	6	-39.2	O <sup>t</sup> Bu	CF <sub>3</sub>	H	PNP	+Mn-C, -P-C, +P-O, +Mn-O
104	O <sup>t</sup> Bu-PNP-CF <sub>3</sub> -5	5	-13.03	O <sup>t</sup> Bu	CF <sub>3</sub>	H	PNP	-Mn-N, +Mn-C, +N-P, -P-C, +P-O, +Mn-O
105	O <sup>t</sup> Bu-PNP-CF <sub>3</sub> -6	5	-10.97	O <sup>t</sup> Bu	CF <sub>3</sub>	H	PNP	-Mn-P, +Mn-C, -P-C, +P-O, +Mn-O
106	O <sup>t</sup> Bu-PNP-CF <sub>3</sub> -8	5	-17.65	O <sup>t</sup> Bu	CF <sub>3</sub>	H	PNP	-Mn-N, +Mn-C, -P-C, +P-O, +Mn-O
107	O <sup>t</sup> Bu-PNP-cy- <sup>t</sup> Bu-2	5	-76.6	O <sup>t</sup> Bu	cy	<sup>t</sup> Bu	PNP	-Mn-N
108	O <sup>t</sup> Bu-PNP- <sup>i</sup> Pr- <sup>i</sup> Pr-2	5	-84.38	O <sup>t</sup> Bu	<sup>i</sup> Pr	<sup>i</sup> Pr	PNP	-Mn-N
109	O <sup>t</sup> Bu-PNP- <sup>i</sup> Pr- <sup>t</sup> Bu-2	5	-76.39	O <sup>t</sup> Bu	<sup>i</sup> Pr	<sup>t</sup> Bu	PNP	-Mn-N
110	OCH <sub>3</sub> -PNP-CF <sub>3</sub> - <sup>i</sup> Pr-4	6	-9.56	OCH <sub>3</sub>	CF <sub>3</sub>	<sup>i</sup> Pr	PNP	+Mn-C, -P-C, +P-O, +Mn-O
111	OCH <sub>3</sub> -PNP-CF <sub>3</sub> - <sup>t</sup> Bu-3	6	-0.83	OCH <sub>3</sub>	CF <sub>3</sub>	<sup>t</sup> Bu	PNP	+Mn-C, -P-C, +P-O, +Mn-O
112	OCH <sub>3</sub> -PNP-CF <sub>3</sub> -2	6	-11.81	OCH <sub>3</sub>	CF <sub>3</sub>	H	PNP	+P-O
113	OCH <sub>3</sub> -PNP-CF <sub>3</sub> -4	6	-23.05	OCH <sub>3</sub>	CF <sub>3</sub>	H	PNP	+Mn-C, -P-C, +P-O, +Mn-O
114	OCH <sub>3</sub> -PNP-cy- <sup>t</sup> Bu-2	5	-65.1	OCH <sub>3</sub>	cy	<sup>t</sup> Bu	PNP	-Mn-N
115	OH-PNP-CF <sub>3</sub> -CF <sub>3</sub> -7	6	-9.43	OH	CF <sub>3</sub>	CF <sub>3</sub>	PNP	+Mn-C, -P-C, +P-O, +Mn-O
116	OH-PNP-CF <sub>3</sub> -cy-2	6	-0.19	OH	CF <sub>3</sub>	cy	PNP	-Mn-N, -Mn-P, +Mn-C, +N-H, -P-C, +P-O, -O-H, +Mn-F
117	OH-PNP-CF <sub>3</sub> -cy-5	6	-8.18	OH	CF <sub>3</sub>	cy	PNP	+Mn-C, -P-C, +P-O, +Mn-O
118	OH-PNP-CF <sub>3</sub> - <sup>i</sup> Pr-9	6	-4.46	OH	CF <sub>3</sub>	<sup>i</sup> Pr	PNP	-Mn-N, +Mn-C, +N-P, -P-C, +P-O
119	OH-PNP-CF <sub>3</sub> -ph-3	6	-3.32	OH	CF <sub>3</sub>	ph	PNP	+P-O
120	OH-PNP-CF <sub>3</sub> -ph-7	6	-17.16	OH	CF <sub>3</sub>	ph	PNP	+Mn-C, -P-C, +P-O, +Mn-O
121	OH-PNP-CF <sub>3</sub> - <sup>t</sup> Bu-13	5	-0.6	OH	CF <sub>3</sub>	<sup>t</sup> Bu	PNP	-Mn-N, +Mn-C, -P-C, +P-O, +Mn-O
122	OH-PNP-CF <sub>3</sub> -2	6	-10.84	OH	CF <sub>3</sub>	H	PNP	+P-O
123	OH-PNP-CF <sub>3</sub> -15	5	-3.58	OH	CF <sub>3</sub>	H	PNP	-Mn-N, +Mn-C, +N-H, -P-C, +P-O, -O-H, +Mn-O

124	OH-PNP-CF <sub>3</sub> -21	5	-1.55	OH	CF <sub>3</sub>	H	PNP	-Mn-N, -Mn-P, +Mn-C, +N-H, -P-C, +P-O, -O-H
125	OH-PNP-cy- <sup>t</sup> Bu-2	5	-71.34	OH	cy	<sup>t</sup> Bu	PNP	-Mn-N
126	OH-PNP- <sup>i</sup> Pr- <sup>t</sup> Bu-2	5	-70.23	OH	<sup>i</sup> Pr	<sup>t</sup> Bu	PNP	-Mn-N
127	Br-SNS-H-cy-2	6	-3.41	Br	H	cy	SNS	-Mn-S, +Mn-H
128	Br-SNS-H-cy-3	6	-2.68	Br	H	cy	SNS	-Mn-S, +Mn-H
129	Br-SNS-H-cy-5	6	-10.68	Br	H	cy	SNS	-Mn-S, +Mn-H
130	OH-SNS-CF <sub>3</sub> -cy-5	5	-8.59	OH	CF <sub>3</sub>	cy	SNS	-Mn-S, -Mn-S, -Mn-N, +Mn-F, +Mn-F, +H-O, +N-C, -C-F, -C-F
131	OH-SNS-H-ph-3	5	-9.48	OH	H	ph	SNS	-Mn-S, -H-S, +H-O



## REFERENCES

- [1] B. D. McKay and A. Piperno. “Practical graph isomorphism, II”. In: *Journal of Symbolic Computation* 60 (2014), pp. 94–112. DOI: 10.1016/j.jsc.2013.09.003.
- [2] S. Bougueroua, R. Spezia, S. Pezzotti, S. Vial, F. Quessette, D. Barth, and M.-P. Gaigeot. “Graph theory for automatic structural recognition in molecular dynamics simulations”. In: *The Journal of Chemical Physics* 149.18 (2018), p. 184102. DOI: 10.1063/1.5045818.



# ACKNOWLEDGEMENTS

*Aachen, April 2023*

Thinking back of the time I spent in the past few years, I am extremely proud of the work and appreciate the opportunity to learn as much as I could. This thesis is the result of continued learning, persistence and countless interactions and meetings.

First and foremost, I want to thank my supervisor Evgeny. Thanks for keeping an open mind and letting me explore the best fit for my talents and experiences. It was hell of a ride! I would say it paid back! I learned a lot during this period.

Marie-Pierre et Sana ! Je vous ai contacté à un moment où je me sentais vraiment coincé ! C'était très agréable dès la première réponse à mon (long) message linkedIn et notre première conversation Skype. Merci beaucoup Marie-Pierre d'avoir accepté de faire ce parcours stimulant avec moi et d'avoir suivi les activités des dernières années! J'ai vraiment apprécié le temps que j'ai passé à Evry dans votre groupe pour démarrer notre collaboration. Sana ! merci pour les innombrables rencontres que nous avons eues qui ont été vraiment exemplaires pour moi en tant que collaboration interdisciplinaire. Nous avons fait un travail incroyable en nous écoutant mutuellement en tant que personnes issues de deux disciplines (à l'époque) différentes ! Je suis fier et j'évoque notre collaboration comme une belle réussite dans ma proposition n°10 !

I also want to thank the committee members for their commitment to reading this thesis! I hope you have enjoyed this! Thank you Clémence for accepting to be a committee member! I enjoyed following your research in the last years and I am honored to have you on the team! Dear Thijs! You mentioned in my GO-NO-GO meeting that this research will change the field! That simple statement kept me motivated for quite a long time! Thank you! Evert Jan! I really enjoyed your lectures at MolSim and going alongside you in the practice sessions. It is a great feeling to have you on the committee. Ferdinand! We always passed on the hallways and I am happy that you have read my thesis. Dear Dr. Buda, thanks for being on the team! I am sure your views on the subject will be enlightening for me!

Next, I want to thank my colleagues in the ISE group. Starting from my roommates Jittima, Elena, Chong and Dapeng. Jittima and Elena! The daily Ph.D. life struggles went by much easier with you! I am very thankful for the vibrant working environment in the office! Prof. Chong! I hope you read this one day! I really appreciate your efforts in my early Ph.D. days and you teaching me the basics of what I learnt about quantum calculations! Maybe I never became the *classic* computational chemist that you trained me on! However, I can now say I put those experiences in great use! Dapeng, thank you for being there for my endless coding questions. I wish you, Guanna and the twins a lot of fun! Evgeny Jr., Robbert and Annika thanks for being good colleagues and companions in the trips and group outings. Dr. Robert Franz! Thank you for being a great friend. Always

on time to get us out of our offices to have our doses of social life and experience of the diverse foods that you made. I wish you the best with your career and life! Choncheng! Thank you very much for being another engineer in the group! I enjoyed teaching with you and your view on subjects!

Georgy! I really appreciate your commitment to listening to my wild ideas in chemistry and always believing in keeping an open mind in science. You are a true hero in that respect! If it wasn't for you, I would have struggled more, not only with convincing people, but also in making nice figures! Christophe, thanks a lot for accommodating me as a neighbor when I felt like migrating to your office. I wish you and the new family the bests! Wenjun! My only colleague after formal hours (which were indeed my working hours)! I appreciate a lot the science talks we had (from which I learned a lot about the chemistry I made automatically!). I really enjoyed the talks we had and the Chinese food that you ordered sometimes! You are truly amazing!

Next, I want to express my gratitude to my neighbours in the department, to the *buurmans*! (and the *buurvrouw* of course! The order is simply chronological, since the *buurvrouw* joined only in my last year!) Ben, Bart and Marcel! *Elke dag was uitstekend voor mij met jullie!* Ben, I loved our cheerful interactions. It meant a lot to me to receive messages and letters from you on my birthday and the Persian events! You are truly a passionate and lovely human being! Marcel! Looking back at the years I worked in your next office, Although I always said everything was *uitstekend!*, there were many ups and downs and it always cheered me up to talk to you. You patiently taught me about the dutch language and culture! Thanks a lot! Bart, I was always impressed with your scientific curiosity and developments. If it was not for you and the other *buurmans*, I would not have survived this! Beste Xiaohui! Thanks for being such a nice substitute for Ben! I really enjoyed our little conversations! It always cheered me up! And last but not least, I want to thank Joyce! *Ik ben nu niet bijna, maar echt!!* klaar! Heel erg bedankt voor al je goede humeur en opgewektheid!

A special thank you to Karin! Thanks for being such a kind-hearted and welcoming person! Your love for Iran and the memories of your trips to Iran, I had a really hard time separating this acknowledgment from that of my persian friends :) Behrooz, Richard, Maryam, Afshin! I cannot thank you enough for your support and the good time! It was fun and it felt like home with you around! My hallway friends in the department! Mark, Divya, Kailun! Thanks a lot for great conversations! It was always refreshing to discuss stuff with you guys and see how we all went through similar struggles. I wish you the bests!

Last but not least, I want to thank my family for their unconditional love and support during this challenging period. Mom and Dad, I owe you everything! I love you! I thought English could not accommodate the words I needed to express my feelings for you and my brothers! That's why I dedicated this thesis to you in persian at the beginning of the thesis! I wish I could visit home more often during my Ph.D.! Mohammad and Mohsen, it is always heartwarming to talk to you and nothing can replace this for me regardless of how far we live from each other! Good news is that I am not going to miss Mohammad and Narges's wedding! Narges and Samaneh, I always have a lot of fun talking to you! I miss you all!

Finally, a great thank you to my lovely Dina! Your lively attitude and splendid view

on life has changed me for ever! Thanks for always being there for me no matter what! This is just the beginning! I could not have imagined better days than the time we have! Good luck with finishing your thesis soon as well. I am eagerly waiting to listen to your performance in Saarbrücken! I look forward to what the future brings for us!



## CURRICULUM VITÆ

Ali Hashemi was born on 8 December 1988 in Esfahan, Iran. He completed his secondary school education in physics and mathematics and started his BSc studies in chemical engineering at Sharif university of technology. He was awarded the Singapore International Pre-Graduate Award (SIPGA) to do his internship in the institute of chemical and engineering sciences on mineral carbonation. Ali continued his MSc. education in pharmaceutical engineering at University of Tehran where he followed his interest in molecular simulations. He did his master thesis on molecular dynamics simulations on block-co-polymeric controlled release systems. In November 2017, he joined the group of prof.dr. Evgeny Pidko to start his Ph.D. research. The most important results of this work are described in this dissertation.



# LIST OF PUBLICATIONS

3. K. D. Vogiatzis, M. V. Polynski, J. K. Kirkland, J. Townsend, A. Hashemi, C. Liu, and E. A. Pidko. "Computational Approach to Molecular Catalysis by 3d Transition Metals: Challenges and Opportunities". In: *Chemical Reviews* 119.4 (2019), pp. 2453–2523. DOI: 10.1021/acs.chemrev.8b00361
2. A. Hashemi, S. Bougueroua, M.-P. Gageot, and E. A. Pidko. "ReNeGate: A Reaction Network Graph-Theoretical Tool for Automated Mechanistic Studies in Computational Homogeneous Catalysis". In: *Journal of Chemical Theory and Computation* 18.12 (2022), pp. 7470–7482. DOI: 10.1021/acs.jctc.2c00404
1. A. Hashemi, S. Bougueroua, M.-P. Gageot, and E. A. Pidko. "High-Throughput Reactivity Exploration for Extended Databases of Transition Metal Catalysts." In: *ChemRxiv* (2023). DOI: 10.26434/chemrxiv-2023-f76nv

# AUTOMATED AND HIGH-THROUGHPUT REACTIVITY ANALYSIS IN HOMOGENEOUS CATALYSIS

## THE DEACTIVATION COMPLEXITY OF Mn(I) HYDROGENATION CATALYSTS

In this thesis, a number of projects aimed at developing new simulation methods for studying complex reactive systems were highlighted, with a particular emphasis on simulation strategies based on the concept of bonding graphs. These mathematical structures provide useful tools for a variety of algorithms developed over the last few decades. Through automated analysis of exhaustive exploration trajectories, serendipities were captured that could escape the expert heuristics or otherwise needed expertise in different disciplines to be interpreted correctly. Chemical discoveries using the methodologies introduced in this thesis could range from very obvious one-step reactions that were just not “normally” considered to multi-step complex reactions that were not imaginable to the expert. With high-throughput reactivity screenings on in silico catalyst libraries, a big step was taken towards “rational” catalyst design.

Ali Hashemi started his BSc studies in chemical engineering at Sharif university of technology. He was awarded the Singapore International Pre-Graduate Award (SIPGA) to do his internship in the institute of chemical and engineering sciences. Ali continued his M.Sc. education in pharmaceutical engineering at University of Tehran where he followed his interest in molecular dynamics simulations. In November 2017, he joined the group of prof.dr. Evgeny Pidko to start his doctoral studies. His Ph.D. research was focused on automated and high-throughput reactivity analysis in homogeneous catalysis. The most important results of this work are described in this dissertation.