HELVETICA

# Expedited Screening of Active and Regioselective Catalysts for the Hydroformylation Reaction

Matthew D. Wodrich,[a] Michael Busch,[a,b,c] and Clémence Corminboeuf[a,b,*]

[a] Laboratory for Computational Molecular Design, Institute of Chemical Sciences and Engineering, Ecole Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland, clemence.corminboeuf@epfl.ch

[b] National Centre for Computational Design and Discovery of Novel Materials (MARVEL), Ecole Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland

[c] Current Address: Department of Physics, Chalmers University of Technology, Fysikgränd 3, SE-412 96 Göteborg, Sweden

Dedicated to François Diederich on the occasion of his retirement.

The discovery of new homogeneous catalysts that preferentially form one product over another in regio- or enantioselective chemical reactions has traditionally been the province of experimental chemists. Today computational based approaches have carved an increasingly important role, which, for computational catalytic designs, often rely on highly inefficient combinatorial-based screening methods. To increase the pace of discovery, tools capable of rapidly assessing large numbers of prospective species and identify those possessing desirable properties, such as activity and selectivity, are vital. Here, through the examination of the hydroformylation of 2-methylpropene, we demonstrate how a new tool built upon molecular volcano plots can be used to quickly predict the activity of molecular catalysts as well as estimate the intrinsic ability of each species to form one regioisomer over the other with striking accuracy. Following training and validation, these regioselective molecular volcanoes are employed to predict catalysts that preferentially form the branched product (2,2-dimethylpropanal) in violation of Keulemans' seventy-year-old law. Eighteen species (out of 68 predicted) were computationally predicted to have regiomeric excess (*r.e.*) values greater than 90. Overall, these tools can be used to quickly screen the activity and selectivity of potential catalysis based on two easily computed descriptor variables.

**Keywords**: homogeneous catalysis • density functional calculations • volcano plots • linear scaling relationships • hydroformylation

## Introduction

Developing protocols that result in the increasingly selective formation of a desirable product while minimizing undesirable side reactions represents a cornerstone of contemporary chemistry. Naturally, achieving new or more efficient reactions is often closely linked to the catalyst charged with accelerating the chemical transformation from reactants to products. As a result, developing new catalysts that simultaneously increase selectivity, in order to maximize the quantity of the valued product, and enhance activity, in order to form this product in a timely manner, constitutes an extremely active research avenue. While historically new catalysts where synthesized and tested directly in the laboratory, the past decade has seen an explosion in the number of computational simulations, which have made significant contributions in rationalizing the experimental behavior of existing catalysts, as well as identifying new species.[1, 2] For analyzing homogenously catalyzed reactions, the sheer number of possible metal/ligand combinations renders computational approaches based on frequently used combinatorial designs highly inefficient, specifically because considerable amounts of time and effort are devoted to species with undesirable properties. To a degree, this inefficiency can be mitigated and a subset of catalysts anticipated to have enhanced properties identified through the use of structure-activity type relationships[2-4] or computational processes that pinpoint favorable reaction pathways in an automated manner.[5-15] Machine learning models have also recently been used to predict reaction yields as demonstrated by Dreher and Doyle.[16]

Today, tools originally developed for either homogeneous or heterogeneous catalysis are increasingly finding use within the other domain.[17] One computational tool matching this description, that has become indispensible for identifying new heterogeneous and electrocatalysts,[18, 19] is the volcano plot[20, 21] that has been espoused by Nørskov.[2] This tools assists in the rapid identification of attractive catalysts by mapping their intrinsic properties onto an easily computed descriptor variable, such as the catalyst/substrate interaction energy. In this way, the volcano plot is a pictorial facsimile of Sabatier's principle,[22, 23] which states that an "ideal catalyst" is characterized by an energetic balance between reactant adsorption and product dissociation. The characteristic volcano shape aids in quickly discriminating catalysts with attractive properties, which appear near the top, from less desirable species that appear in the lower regions. Because these plots relate catalytic activity to an easily computed descriptor variable, they have become a cornerstone of computational based techniques for the high-

throughput screening for novel catalysts,[24] including when used in tandem with machine learning models.[25-28]

While applications of volcano plots are prolific in the fields of heterogeneous and electrocatalysis, analogous uses involving homogeneous catalysis are virtually absent. Aside from being used to rationalize catalytic behavior [29, 30] it was not until 2015 when we constructed the first "molecular volcano plot" that predicted the thermodynamics of catalysts for a Suzuki cross-coupling reaction.[31]

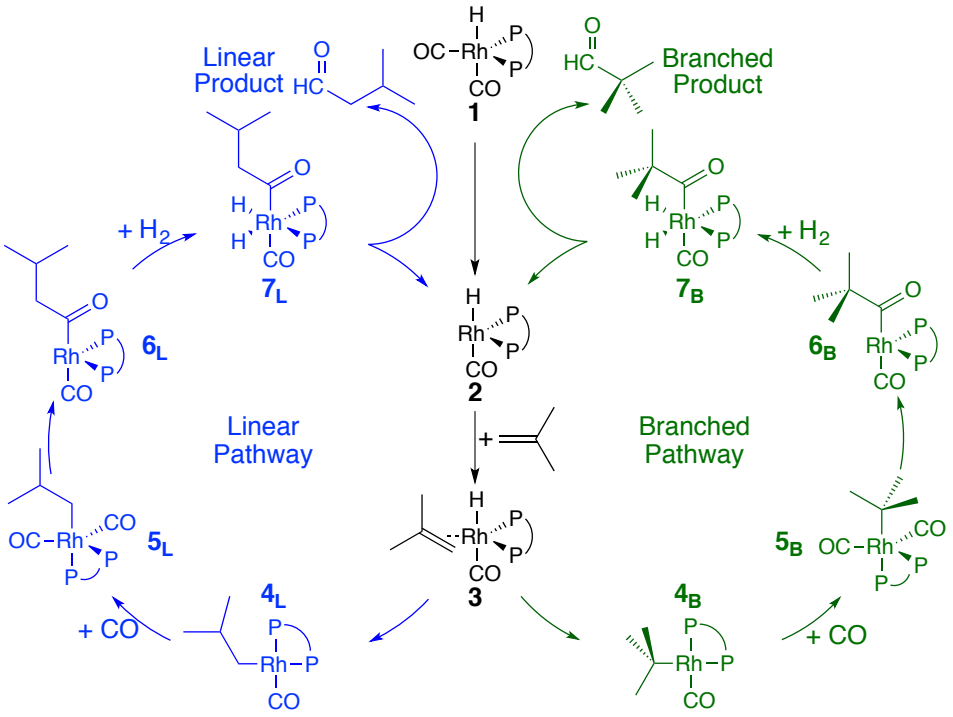
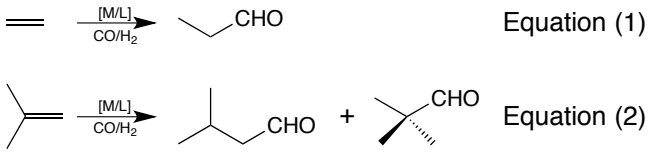



**Figure 1**. Catalytic cycles leading to the two regioisomers for Rh-catalyzed hydroformylation of an olefin.

Subsequent work has aimed at extending the applicability of these tools[32] and has been used to construct a generalized picture of cross-coupling reactions.[33, 34] However, each of these studies considered only the catalytic cycle's thermodynamic aspects and ignored kinetic considerations. Clearly, the inclusion of the activation barriers not only provides an enhanced picture of overall catalytic activity, but also is essential for establishing the product ratio (*e.e.* or *r.e.*) formed during a reaction. To partially address this, we examined the hydroformylation of an unsubstituted olefin (ethene), Equation (1), which established that molecular volcanoes can identify catalysts based on kinetic criteria.[35] This reaction, however, leads only a single product, and, therefore, avoided issues involving selectivity. However, the idea of introducing product selectivity into the molecular volcano plot framework has strong conceptual appeal, principally because it would open the door for use of these practical tools in important fields such as asymmetric homogeneous catalysis. With this idea in mind, we evaluated a related hydroformylation reaction involving a dialkyl olefin (2-methylpropene), equation (2), to ascertain the viability of using molecular volcano plots to predict both the activity and selectivity of homogeneous catalysts.




The conversion of an olefin, CO, and $H_2$ into an aldehyde via hydroformylation is a highly important industrial process that creates millions of tons of product annually.[36] Transition metal catalysts incorporating a cobalt atom have been known to facilitate the reaction since the early 1960s.[37] Following the seminal work of Wilkinson,[38, 39] today rhodium based species are the preferred as they operate under milder conditions. The proposed catalytic cycles leading to formation of both the linear (blue) and branched (green) products are depicted in Figure 1. Dissociation of a CO ligand from **1** leads to the entry point of the catalytic cycle, **2**. Binding of 2-methylpropene leads to formation of the π-complex, **3**. From **3**, two pathways emerge; the hydrogen atom can be transferred in either a Markovnikov type addition [*i.e.*, to the β-carbon atom of the olefin] leading to **4L** and the linear (blue) pathway, or anti-Markovnikov [*i.e.*, to the olefin's α-carbon], which yields **4B** via the green pathway. Following formation of the σ-complex **4**, CO addition (**5**) and subsequent insertion into the Rh-C bond leads to **6**. The addition of

$H_2$ to **6** gives **7**, which undergoes reductive elimination to reform **2**, closing the catalytic cycle.

Industrially, linear products are generally more desired than their branched counterparts,[36] however recent research effort has also been focused on creating more branched products.[40-43] The use of bisphosphine ligands is particularly interesting, as changes in ligand bite angle have been to rationalize for product selectivity.[44-50] Taken together, the hydroformylation of 2-methylpropene provides a particularly titillating example not only because there are two regioisomers that can be products, but also because the branched product creates a tertiary aldehyde, pivaldehyde (2,2-dimethylpropanal), in violation of Keulemans' law.[51] The purpose of this contribution is twofold: first we demonstrate that molecular volcano plots can be used to identify catalysts based both on selectivity and activity and second, we use this newly constructed tool to hunt for "rule-breaking" catalysts that are predicted to be highly selective for the formation of the branched aldehyde, equation (2).

## Results and Discussion

To assess the ability of volcano plots in identifying highly active and selective catalysts, we employed DFT computations (see Computational Details for further information) to assess the two catalytic cycles depicted in Figure 1 for a series of rhodium catalysts featuring bidentate phosphine ligands (Figure 2, Training Set). The entirety of the catalytic cycle depicted in Figure 1 was computed for the six training set catalysts whose ligands featured diverse backbone and flanking substituents. The resulting linear free energy scaling relationships (LFESRs) that arose from the training set were used to construct the volcano plots used throughout this work (see SI Figures S1 and S2).

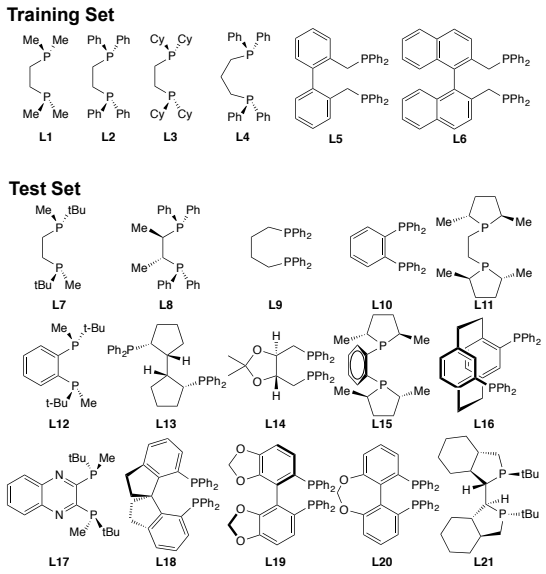
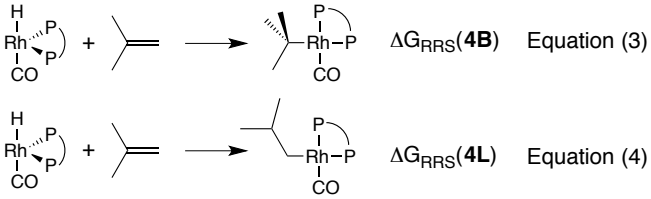



**Figure 2**. Bidentate phosphine ligands used in this study. The "Training Set" was used to identify linear free energy scaling relationship and construct volcano plots. The "Test Set" was used to test the predictions made using the previously established scaling relationships and volcano plots.

Volcano plots are built upon linear free energy scaling relationships (LFESRs), which allow the relative energy of each of the intermediates and transition states present in the catalytic cycle to be estimated by knowing the value of a chosen descriptor variable. Simple mathematical procedures (see SI for details) allow the corresponding energies of any reaction step [e.g., $\Delta G$(**3**→**4L**) and $\Delta G$(**3**→**4B**)] to be extracted from these LFESRs. These same relationships also provide estimations of the heights of transition state barriers [e.g., $\Delta G$(**3**→**TS3,4L**) and $\Delta G$(**3**→**TS3,4B**)], akin to the well-known Bell-Evans-Polanyi relationships,[52, 53] that can be used to assess the activity and selectivity of different catalysts. For the chemical reaction displayed equation (2), testing revealed $\Delta G_{RRS}$(**4B**) and $\Delta G_{RRS}$(**4L**) [equations (3) and (4)] to be excellent descriptor variables (see SI Figures S1 and S2 for correlation coefficients showing that these descriptor capably describe the relative stability of other catalytic cycle intermediates). Using these descriptors, the free energy associated with any reaction step can be established (see SI Figure S3). The final volcano plot is then formed by plotting only the most energetically costly reaction steps [-$\Delta G$(Rxn)] for each descriptor (x-axis) value (a detailed explanation of creating volcano plots from linear scaling relationships is provided in the SI). Using this information the activity of any prospective catalyst established.



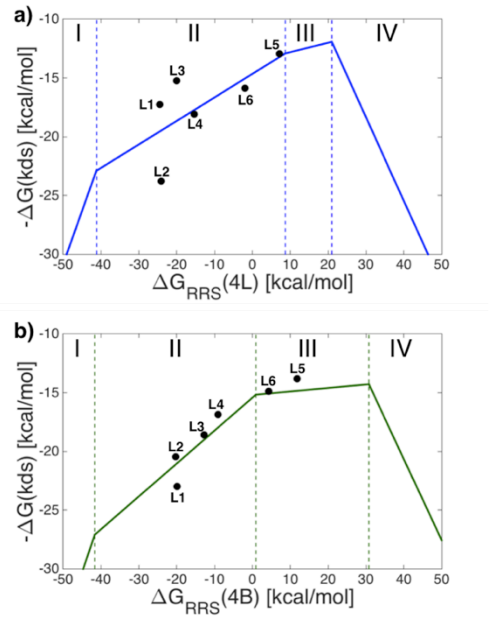



**Figure 3.** Volcano plots for the linear (a) and branched (b) pathways. Species appearing high on the volcano plot have more attractive free energy profiles since the free energy needed to complete the most difficult intermediary step in the catalytic cycle is smaller. Dashed lines indicate a change in the most energetically costly reaction step. For regions I to IV, **TS7,2→2**, **5→TS5,6**, **3→TS3,4**, and **2→3** are the kinetic determining step (y-axis), respectively.

Figure 3 shows the volcano plots for both the linear and branched regioisomers created from the catalysts in the training set (Figure 1). The volcanoes reveal that one of four reaction steps (distinguished by vertical dashed lines) is always the most energetically costly, the exact nature of the step depends on the value of descriptor plotted along the x-axis. As is typical seen in molecular volcano plots, two of the four steps correspond to oxidative addition (**2→3**, region IV), which involves association of the reactant with the catalyst, and reductive elimination (**TS7,2→2**, region I) that encompasses product release. Drawing an analogy to heterogeneous catalysis, species lying in these regions either have difficulty placing the reactant onto the catalyst (region I) or releasing the product once formed (region IV). Thus, regions I and IV correspond exactly to "weak binding" and "strong binding" situations that characterize the left and right sides of the volcano, respectively, and provide a direct connection to Sabatier's principle. On the other hand, regions II and III represent situations where the catalyst/reactant interaction energies are more balanced and, therefore, more closely match Sabatier's definition of an ideal catalyst. For these species, the most energetically costly reaction step either involves overcoming the transition state barrier associated with CO insertion into the Rh-C bond (**5→TS5,6**, region II) or hydrogen atom transfer from the metal center to the olefin (**3→TS3,4**, region III).

Because rhodium phosphine catalysts are well established for this reaction, it is not surprising that each member of the "training set" appears in either region II or III. However, not all catalysts lying in these two regions have equivalent activity: those species located most proximate to the highest point on the plot (located at the intersection of regions III and IV) are expected be the most active. Thus, Rh(L5) would be the most active of the "training set" catalysts for the formation of either regiomeric product, since it is located highest on both of the Figure 1 volcano plots. Indeed, this catalyst is experimentally known to be quite effective and has been employed by Kodak for the formation of straight chain aldehydes.[44, 54]

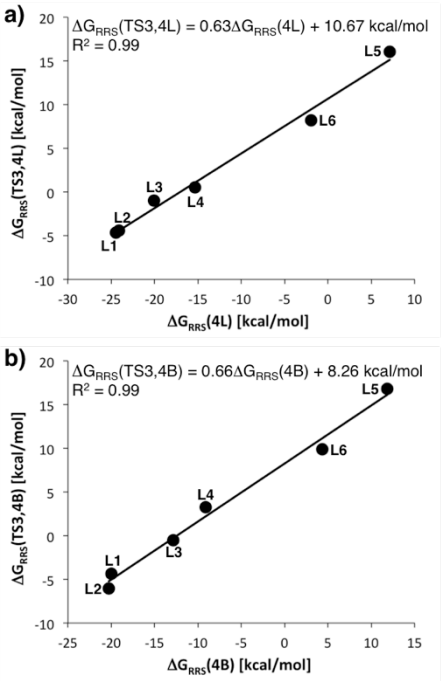



**Figure 4.** Linear free energy scaling relationships between descriptors **4L** and **4B** (x-axis) and **TS3,4L** and **TS3,4B** (y-axis). The high correlation coefficient reveals that the relative free energy of the two transition states can be predicted with high accuracy from the descriptor variables..

While the Figure 3 molecular volcano plots indicate that Rh(L5) would be the most "active" catalyst for forming both the linear or branched product, it provides no indication of the ability of each catalyst to selectively form a single regioisomer. Since this information is often equally or more important than finding an extremely active catalyst, the idea of including selectivity within a volcano plot is very appealing. Fortunately, the earlier discussed LFESRs already provide the necessary information (the heights of the transition state barriers) to determine the regiomeric excess (*r.e.*) for each catalyst. For the hydroformylation reaction, the transition state most critical to determining this quantity is **TS3,4**, which corresponds to the point at which the two catalytic cycles diverge. Figure 4 shows that the energy barriers associated with these transition states, *i.e.*, **TS3,4L**/**TS3,4B**, are predicted quite well using

the appropriate descriptor variables [$\Delta G_{RRS}$(**4L**)/$\Delta G_{RRS}$(**4B**)], which shows that BEP type relationships hold well for these catalysts. Thus, the same descriptors used to create the molecular volcano plots shown in Figure 1, also provide $\Delta G_{RRS}$(**TS3,4L**) and $\Delta G_{RRS}$(**TS3,4B**) the energy difference of which can be used to determine the *r.e.* through the following equation: $\Delta G_{RRS}(TS3,4L) - \Delta G_{RRS}(TS3,4B) = RT \ln\left(\frac{1+r.e.}{1-r.e.}\right)$, that defines a catalyst's selectivity.[55]

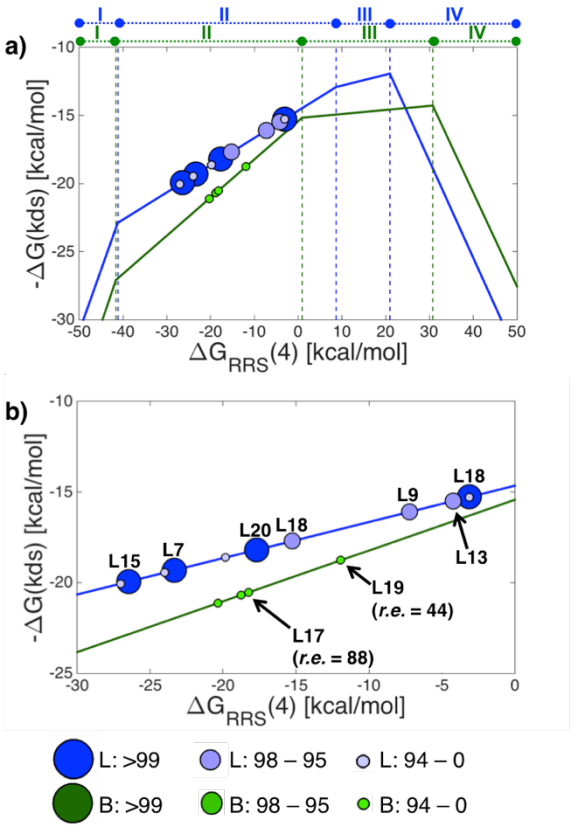



**Figure 5.** (a) Regioselective molecular volcano plot for the 15 test set catalysts depicted in Figure 2. The largest spheres indicate catalysts having *r.e.* values of > 99, while smaller spheres indicate *r.e.* values between 98 and 95 or less than 95. The blue and green colors of the volcanoes and catalyst points correspond to the preferential formation of linear and branched products, respectively. Dashed vertical lines indicate a change in the nature of the kinetic determining step (see Figure 1 for more details). Colored dotted lines above indicate changes in the nature of the kinetic determining step (see Figure 3). (b) Enlargement of (a) showing structures of species anticipated as being the most active and selective of the test set catalysts. Note that both the linear (blue) and branched (green) species are plotted on the same plot to present a comprehensive picture. The location of each of the points on x-axis correspond to using the LFESRs for the preferred product [blue points use $\Delta G_{RRS}$(**4L**) while green points use $\Delta G_{RRS}$ (**4B**)].

To assess the viability of creating molecular volcano plots that also provide information on catalytic selectivity, we utilized the LFESRs established from the "training set" on a series of 15 additional "test set" catalysts (Figure 2). *r.e.* values established using the above-mentioned formula were included into the molecular volcano plot by varying the size/color of the points to indicate the degree of regioselectivity and the preferred product formed. Figure 5 shows the regioselective molecular volcano plot for the 15 test set catalysts. Species represented by large circles are predicted to be the most selective and, thus, are of key interest. Immediately evident is that a majority of the catalysts comprising the "test set" form the linear, more industrially relevant product in excess. In particular, one species, Rh(L18) is predicted to have an *r.e.* greater than 99 while also possessing good activity. Three other catalysts, Rh(L7), Rh(L15), and Rh(L20) are predicted to have high selectivity but slightly low activities. Additionally, Rh(L9) and Rh(L13) are highly active while possessing quite good *r.e.* values between 95 and 99. Our regioselective volcano also identified one catalyst, Rh(L19), that preferentially forms the branched regioisomer with an *r.e.* of 44 and is the most active amongst those species forming the this isomer. Additionally, Rh(L17) also form the branched product, but with an *r.e.* of 88, although it is less active than Rh(L19). These two species are particularly interesting owing to the fact that they stand in violation of Keulemans' rule (*vide supra*).[51]

Having demonstrated that linear scaling relationships and volcano plots can provide quantitative values associated with catalytic selectivity (*i.e.*, the regiomeric excess), a logical question is: just how accurate are these numbers? Recall that each of the *r.e.* values depicted in the Figure 5 molecular volcano plots are obtained only by computing the descriptors [$\Delta G_{RRS}$(**4L**) and $\Delta G_{RRS}$(**4B**)] and using the corresponding LFESRs (equations in Figure 4) to estimate the energy difference between the transition states leading to the linear and branched products. To establish the accuracy of these values, we computed the two transition states (as opposed to estimating these quantities from the LFESRs) and used the free energy differences to validate our estimated *r.e.* values. Figure 6 shows very encouraging results: of the ten catalysts either predicted or computed to have *r.e.* values greater than 95, nine of the estimates correctly predict the preferred product, generally with high qualitative accuracy. Only one species, Rh(L18), was found to be a "false positive."[56] Considering the amount of computational effort needed to predict these *r.e.* values, this tool clearly will have utility in identifying new catalysts that are both highly active and selective. Moreover, we reemphasize an important message on the use of volcano plots: these tools are best used to identify a group of attractive species and not a single "best" catalyst.

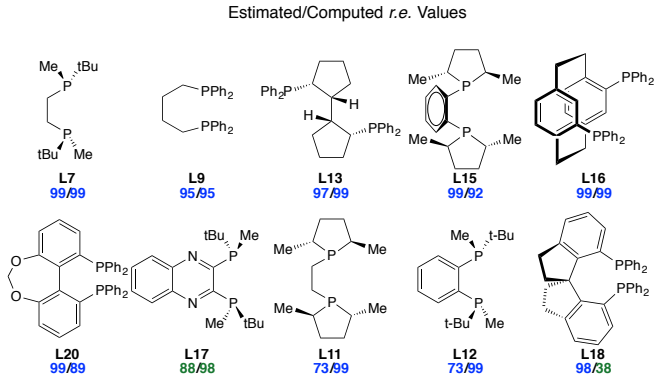



**Figure 6.** Comparisons of catalyst regioselectivities [as defined by their regiomeric excess (*r.e.*) values] for selected species. Blue numbers indicate a *r.e.* favoring the linear regioisomer while green numbers favor the branched regioisomer.

With the validity of our method established, we were particularly interested in applying it to identify the aforementioned "rule-breaking" catalysts that produce the branched product with high selectivity. With this objective in mind, we focused on the two "test set" ligands that were the most active, Rh(L19), and the most selective, Rh(L17), for forming the branched product. Alteration of the peripheral substituents as well as small changes to the ligand backbones produced a "prediction set" of 68 catalysts (L22-L89, see SI Figure S4) that were screened using the Figure 5 volcano plots for both activity and selectivity. The new regioselective molecular volcano plot is shown in Figure 7, and reveals 18 additional catalysts that form the branched product with *r.e.* values in excess of 90. Note that since our focus is on catalysts forming the branched product, only the green volcano (and not the blue volcano representing catalysts producing the linear product) is shown. The activity and selectivity of each of 68 catalysts in the "prediction set" catalysts is summarized in the SI.

The most salient features of Figure 7 are the four points (L42, L69, L76, and L61) that lie on the plateau like region (III) that corresponds to species expected to have the highest activity. All four of these catalysts are predicted to have *r.e.* values exceeding 95, with L76 anticipated to give a very high *r.e.* of 99. Catalysts in region II begin to show decreased activity as the descriptor values become more negative (moving left along the x-axis). Nonetheless, Figure 7 indicates that the most energetically costly step of the catalytic cycle (**5→TS5,6B** for region III) are still experimentally accessible (~18 kcal/mol) even for the least active catalysts (L24, L31, L38, L45, L66, and L72). Note, however, that each of these species still have *r.e.* values in excess of 99.

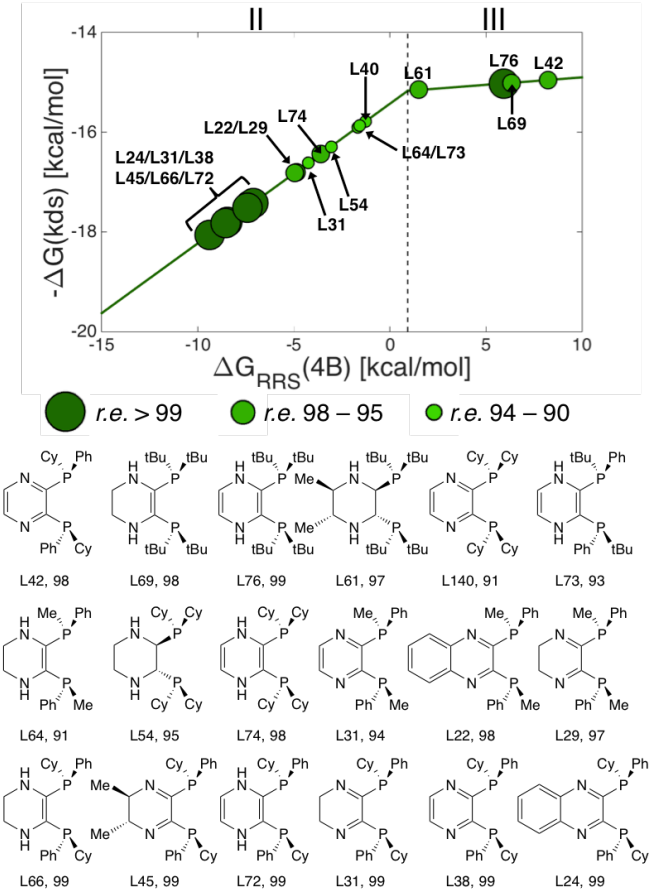



**Figure 7.** Prediction set catalysts estimated to have a *r.e.* greater than 90 for production of the branched product (pivaldehyde). Value below each structure represents *r.e.*

Structurally speaking, all of these "rule-breaking" catalysts are based on modified scaffolds of L17, whereas modifications to the L19 scaffold yielded no catalysts that resulted in high regioselectivity for the branched product. For L17, replacing the methyl groups attached to the phosphorus atoms with bulkier substituents also results in more selective catalysts. Specifically, cyclohexyl substituents, as well as phenyl/cyclohexyl combinations arranged in a $C_2$ type symmetry are estimated to lead to higher *r.e.* values. Future experimental work aimed at selectively forming more branched aldehydes during hydroformylation reactions should gauge the influence brought about by these larger substituent in place of more commonly used "R" groups.

A final note concerns comparing the estimated r.e. values from with computed rather than experimental data. It is true that differences often exist between selectivities estimated from computation and those experimentally observed. While it is not possible to confirm the accuracy of our estimated *r.e.* values without direct experimental input, it should be noted that many of the most selective catalysts have large energy differences between the transition states that dictate selectivity (on the order of ~4 kcal/mol or greater). Thus, it is expected that the selectivity of these species

should be, at least, qualitatively correct. In the future, we hope to directly address this issue by comparing our theoretical predictions directly with experimental results.

## Conclusions

In conclusion, we developed a tool that facilitates the rapid identification of highly active and highly selective homogeneous catalysts based on linear free energy scaling relationships and molecular volcano plots. Assessment of the rhodium phosphine catalyzed hydroformylation of 2-methylpropene indicates that scaling relations and volcano plots can be generated by computing the complete catalytic cycles of only a handful of "training set" catalysts. These scaling relations are quite robust and facilitate the accurate prediction of the key transition state barriers, which, in turn, can be used to deduce the selectivity of each catalyst. Validation over a "test set" of 15 catalysts showed a high success rate in accurately predicting the regiomeric excess (*r.e.*) that were established through full DFT computations of the catalytic cycle. Having validated our method, an additional 68 catalysts were screened in search of species that preferentially form the branched product, in violation of Keulemans' rule. This screening procedure yielded 18 catalysts with *r.e.* values in excess of 90, and provides a set of species that should be experimentally examined in greater detail. Owing to their ability to rapidly predict both the activity and selectivity of prospective catalysts, the regiomeric molecular volcano plots detailed within should become powerful and useful tools with applicability in diverse fields of homogeneous catalysis.

## Experimental Section

### *Computational Details*

The structures of all catalytic cycle intermediates and transition state, as well as relevant reference species, were optimized using the M06[57, 58] density functional and the def2-SVP basis set[59] in implicit benzene solvent (SMD model[60]) using Gaussian09.[61] The "ultrafine" integration grid was uniformly employed to remove known problems related to grid size with the Minnesota family of density functionals.[62] Refined energy estimates were obtained using a density dependent dispersion correction, -dDsC,[63-66] appended to the PBE0[67, 68] functional with the TZ2P basis set as implemented in ADF.[69, 70] Reported free energies include COSMO-RS[71] solvation corrections (in benzene) determined at the PBE0-dDsC/TZ2P level. Estimations of the free energy were obtained at the M06/def2-SVP level using the rigid-rotor harmonic oscillator model along with a correction for translational entropy in solution as proposed by Grimme[72] within the GoodVibes[73] program developed by Paton and Funes-Ardoiz. Note that we have previously found good correlations between linear scaling relationships computed using the PBE0-dDsC and M06 functionals for a similar hydroformylation reaction (equation 1).[35] As such, it is anticipated that the regiomeric molecular volcano plots should not change significantly based on the theoretical level used.

## Supplementary Material

Linear scaling relationships, mathematical derivations of volcano plots, simulated reaction profiles, schematic structures of relevant compounds, relevant computed energies, and Cartesian coordinates of optimized structures are provided. Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/MS-number.

## Acknowledgements

The National Centre of Competence in Research (NCCR) "Materials' Revolution: Computational Design and Discovery of Novel Materials (MARVEL)" of the Swiss National Science Foundation (SNSF) and the EPFL are acknowledged for financial support.

## Author Contribution Statement

MDW, MB, and CC conceived the research; MDW ran the computations; MDW, MB, and CC wrote the manuscript.

## References

[1] K. N. Houk, P. H.-Y. Cheong, 'Computational Prediction of Small-Molecule Catalysts', *Nature* **2008**, *455*, 309-313.

[2] J. K. Nørskov, T. Bligaard, J. Rossmeisl, C. H. Christensen, 'Towards the computational design of solid catalysts', *Nat. Chem.* **2009**, *1*, 37-46.

[3] M. S. Sigman, K. C. Harper, E. N. Bess, A. Milo, 'The Development of Multidimensional Analysis Tools for Asymmetric Catalysis and Beyond', *Acc. Chem. Res.* **2016**, *49*, 1292-1301.

[4] N. Fey, A. C. Tsipis, S. E. Harris, J. N. Harvey, A. G. Orpen, R. A. Mansson, 'Development of a ligand knowledge base: 1. Computational descriptors for phosphorus donor ligands', *Chem. Eur. J.* **2006**, *12*, 291-302.

[5] B. J. Rooks, M. R. Haas, D. Sepúlveda, T. Lu, S. E. Wheeler, 'Prospects for the Computational Design of Bipyridine N,N′-Dioxide Catalysts for Asymmetric Propargylation Reactions', *ACS Catal.* **2015**, *5*, 272-280.

[6] A. C. Doney, B. J. Rooks, T. Lu, S. E. Wheeler, 'Design of Organocatalysts for Asymmetric Propargylations through Computational Screening', *ACS Catal.* **2016**, *6*, 7948-7955.

[7] Y. Guan, S. E. Wheeler, 'Automated Quantum Mechanical Predictions of Enantioselectivity in a Rhodium-Catalyzed Asymmetric Hydrogenation', *Angew. Chem., Int. Ed.* **2017**, *56*, 9101-9105.

[8] S. Habershon, 'Sampling reactive pathways with random walks in chemical space: Applications to molecular dissociation and catalysis', *J. Chem. Phys.* **2015**, *143*, 094106.

[9] S. Habershon, 'Automated Prediction of Catalytic Mechanism and Rate Law Using Graph-Based Reaction Path Sampling', *J. Chem. Theory Comput.* **2016**, *12*, 1786-1798.

[10] E. Martínez-Núñez, 'An automated method to find transition states using chemical dynamics simulations', *J. Comput. Chem.* **2015**, *36*, 222-234.

[11] J. A. Varela, S. A. Vázquez, E. Martínez-Núñez, 'An automated method to find reaction mechanisms and solve the kinetics in organometallic catalysis', *Chem. Sci.* **2017**, *8*, 3843-3851.

[12] S. Maeda, K. Morokuma, 'A systematic method for locating transition structures of A+B→X type reactions', *J. Chem. Phys.* **2010**, *132*, 241102.

[13] S. Maeda, K. Ohno, K. Morokuma, 'Systematic exploration of the mechanism of chemical reactions: the global reaction route mapping (GRRM) strategy using the ADDF and AFIR methods', *Phys. Chem. Chem. Phys.* **2013**, *15*, 3683-3701.

[14] S. Maeda, T. Taketsugu, K. Morokuma, 'Exploring transition state structures for intramolecular pathways by the artificial force induced reaction method', *J. Comput. Chem.* **2014**, *35*, 166-173.

[15] T. Yoshimura, S. Maeda, T. Taketsugu, M. Sawamura, K. Morokuma, S. Mori, 'Exploring the full catalytic cycle of rhodium(I)–BINAP-catalysed isomerisation of allylic amines: a graph theory approach for path optimisation', *Chem. Sci.* **2017**, *8*, 4475-4488.

[16] D. T. Ahneman, J. G. Estrada, S. Lin, S. D. Dreher, A. G. Doyle, 'Predicting reaction performance in C–N cross-coupling using machine learning', *Science* **2018**, *360*, 186-190.

[17] L. Falivene, S. M. Kozlov, L. Cavallo, 'Constructing Bridges between Computational Tools in Heterogeneous and Homogeneous Catalysis', *ACS Catal.* **2018**, *8*, 5637-5656.

[18] H. Dau, C. Limberg, T. Reier, M. Risch, S. Roggan, P. Strasser, 'The Mechanism of Water Oxidation: From Electrolysis via Homogeneous to Biological Catalysis', *ChemCatChem* **2010**, *2*, 724-761.

[19] J. Greeley, N. M. Markovic, 'The road from animal electricity to green energy: combining experiment and theory in electrocatalysis', *Energy Environ. Sci.* **2012**, *5*, 9246-9256.

[20] H. Gerischer, 'Mechanismus der Elektrolytischen Wasserstoffabscheidung und Adsorptionsenergie von Atomarem Wasserstoff', *Bull. Soc. Chim. Belg.* **1958**, *67*, 506-527.

[21] R. Parsons, 'The rate of electrolytic hydrogen evolution and the heat of adsorption of hydrogen', *Trans. Faraday Soc.* **1958**, *54*, 1053-1063.

[22] P. Sabatier, 'Hydrogénations et déhydrogénations par catalyse', *Ber. Deutsch. Chem. Gesellshaft* **1911**, *44*, 1984-2001.

[23] P. Sabatier, *La Catalyse en Chimie Organique*, Librarie Polytechnique, Paris, 1913.

[24] J. Greeley, T. F. Jaramillo, J. Bonde, I. Chorkendorff, J. K. Nørskov, 'Computational high-throughput screening of electrocatalytic materials for hydrogen evolution', *Nat. Mater.* **2006**, *5*, 909-913.

[25] X. Ma, Z. Li, L. E. K. Achenie, H. Xin, 'Machine-Learning-Augmented Chemisorption Model for CO2 Electroreduction Catalyst Screening', *J. Phys. Chem. Lett.* **2015**, *6*, 3528-3533.

[26] Z. W. Ulissi, A. J. Medford, T. Bligaard, J. K. Nørskov, 'To Address Surface Reaction Network Complexity using Scaling Relations Machine Learning and DFT Calculations', *Nat. Commun.* **2017**, *8*, 14621.

[27] R. B. Wexler, J. M. P. Martirez, A. M. Rappe, 'Chemical Pressure-Driven Enhancement of the Hydrogen Evolving Activity of Ni2P from Nonmetal Surface Doping Interpreted via Machine Learning', *J. Am. Chem. Soc.* **2018**, *140*, 4678-4683.

[28] B. Meyer, B. Sawatlon, S. Heinen, O. A. von Lilienfeld, C. Corminboeuf, 'Machine Learning meets Volcano Plots: Computational Discovery of Cross-Coupling Catalysts', *Chem. Sci.* **2018**, DOI:10.1039/C1038SC01949E.

[29] G. F. Swiegers, *Mechanical Catalysis: Methods of Enzymatic, Homogeneous, and Hetereogeneous Catalysis*, John Wiley & Sons, Hoboken, NJ, 2008.

[30] S. Kozuch, in *Understanding Organometallic Reaction Mechanisms and Catalysis: Computational and Experimental Tools* (Ed.: V. P. Ananikov), Wiley-VCH, Weinheim, 2015.

[31] M. Busch, M. D. Wodrich, C. Corminboeuf, 'Linear Scaling Relationships and Volcano Plots in Homogeneous Catalysis - Revisiting the Suzuki Reaction', *Chem. Sci.* **2015**, *6*, 6754-6761.

[32] M. D. Wodrich, B. Sawatlon, M. Busch, C. Corminboeuf, 'On the Generality of Molecular Volcano Plots', *ChemCatChem* **2018**, *10*, 1586-1591.

[33] M. Busch, M. D. Wodrich, C. Corminboeuf, 'A Generalized Picture of C-C Cross-Coupling', *ACS Catal.* **2017**, *7*, 5643-5653.

[34] M. Busch, M. D. Wodrich, C. Corminboeuf, 'Improving the Thermodynamic Profiles of Potential Suzuki-Miyaura Cross-Coupling Catalysts by Altering the Electrophilic Coupling Component', *ChemCatChem* **2018**, *10*, 1592-1597.

[35] M. D. Wodrich, M. Busch, C. Corminboeuf, 'Accessing and Predicting the Kinetic Profiles of Homogeneous Catalysts from Volcano Plots', *Chem. Sci.* **2016**, *7*, 5723-5735.

[36] R. Franke, D. Selent, A. Börner, 'Applied Hydroformylation', *Chem. Rev.* **2012**, *112*, 5675-5732.

[37] R. F. Heck, D. S. Breslow, 'The Reaction of Cobalt Hydrotetracarbonyl with Olefins', *J. Am. Chem. Soc.* **1961**, *83*, 4023-4027.

[38] J. A. Osborn, G. Wilkinson, J. F. Young, 'Mild hydroformylation of olefins using rhodium catalysts', *Chem. Commun.* **1965**, 17.

[39] D. Evans, J. A. Osborn, G. Wilkinson, 'Hydroformylation of alkenes by use of rhodium complex catalysts', *J. Chem. Soc. A* **1968**, 3133-3142.

[40] G. M. Noonan, J. A. Fuentes, C. J. Cobley, M. L. Clarke, 'An Asymmetric Hydroformylation Catalyst that Delivers Branched Aldehydes from Alkyl Alkenes', *Angew. Chem., Int. Ed.* **2012**, *51*, 2477-2480.

[41] G. M. Noonan, C. J. Cobley, T. Mahoney, M. L. Clarke, 'Rhodium/phospholane–phosphite catalysts give unusually high regioselectivity in the enantioselective hydroformylation of vinyl arenes', *Chem. Commun.* **2014**, *50*, 1475-1477.

[42] P. Dingwall, J. A. Fuentes, L. Crawford, A. M. Z. Slawin, M. Bühl, M. L. Clarke, 'Understanding a Hydroformylation Catalyst that Produces Branched Aldehydes from Alkyl Alkenes', *J. Am. Chem. Soc.* **2017**, *139*, 15921-15932.

[43] A. Phanopoulos, K. Nozaki, 'Branched-Selective Hydroformylation of Nonactivated Olefins Using an N-Triphos/Rh Catalyst', *ACS Catal.* **2018**, *8*, 5799-5809.

[44] C. P. Casey, G. T. Whiteker, M. G. Melville, L. M. Petrovich, J. A. Gavney Jr., D. R. Powell, 'Diphosphines with Natural Bite Angles near 120" Increase Selectivity for n-Aldehyde Formation in Rhodium-Catalyzed Hydroformylation', *J. Am. Chem. Soc.* **1992**, *114*, 5535-5543.

[45] P. W. N. M. van Leeuwen, P. C. J. Kamer, J. N. H. Reek, P. Dierkes, 'Ligand Bite Angle Effects in Metal-catalyzed C−C Bond Formation', *Chem. Rev.* **2000**, *100*, 2741-2770.

[46] J. J. Carbó, F. Maseras, C. Bo, P. W. N. M. van Leeuwen, 'Unraveling the Origin of Regioselectivity in Rhodium Diphosphine Catalyzed

Hydroformylation. A DFT QM/MM Study', *J. Am. Chem. Soc.* **2001**, *123*, 7630-7637.

[47] J. Klosin, C. R. Landis, 'Ligands for Practical Rhodium-Catalyzed Asymmetric Hydroformylation', *Acc. Chem. Res.* **2007**, *40*, 1251-1259.

[48] M. Kumar, R. V. Chaudhari, B. Subramaniam, T. A. Jackson, 'Ligand Effects on the Regioselectivity of Rhodium-Catalyzed Hydroformylation: Density Functional Calculations Illuminate the Role of Long-Range Noncovalent Interactions', *Organometallics* **2014**, *33*, 4183-4191.

[49] M. Kumar, R. V. Chaudhari, B. Subramaniam, T. A. Jackson, 'Importance of Long-Range Noncovalent Interactions in the Regioselectivity of Rhodium-Xantphos-Catalyzed Hydroformylation', *Organometallics* **2015**, *34*, 1062-1073.

[50] S. M. Mansell, 'Catalytic Applications of Small Bite-Angle Diphosphorus Ligands with Single-Atom Linkers', *Dalton Trans.* **2017**, *46*, 15157-15174.

[51] A. I. M. Keulemans, A. Kwantes, T. van Bavel, 'The structure of the formylation (OXO) products obtained from olefines and watergas', *Recl. Trav. Chim. Pays-Bas* **1948**, *67*, 298-308.

[52] R. P. Bell, 'The Theory of Reactions Involving Proton Transfers', *Proc. R. Soc. London, Ser. A* **1936**, *154*, 414-429.

[53] D. J. Evans, M. Polanyi, 'Inertia and driving force of chemical reactions', *Trans. Faraday Soc.* **1938**, *34*, 11-24.

[54] T. J. Devon, G. W. Phillips, T. A. Puckette, J. L. Stavinoha, J. J. Vanderbilt, *Vol. 4,694,109* (Ed.: U. S. Patent), Eastman Kodak Company, USA, 1986.

[55] Note that, in principle, these species could undergo beta-H elimination from 4 to reform 3. However, we estimate the rate of this back reaction should be very slow, as there is significant thermodynamic stablilization associated with CO addition (i.e., 4->5) for all of the catalysts we tested. Indeed for each of the six test set catalysts, TS3,4 represents the highest point on reaction coordinate. As such, the energy difference between TS3,4B and TS3,4L should serve as a good approximation for the reaction regioselectivity.

[56] While some outliers are to be expected using this method, an examination of Rh(L18) did not reveal any obvious reason for the poor r.e. prediction.

[57] Y. Zhao, D. G. Truhlar, 'Density Functionals with Broad Applicability in Chemistry', *Acc. Chem. Res.* **2008**, *41*, 157-167.

[58] Y. Zhao, D. G. Truhlar, 'The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals', *Theor. Chem. Acc.* **2008**, *120*, 215-241.

[59] F. Weigend, R. Ahlrichs, 'Balanced basis sets of split valence, triple zeta valence and quadruple valence quality for H to Rn: Design and assesment of accuracy', *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297-3305.

[60] A. V. Marenich, C. J. Cramer, D. G. Truhlar, 'Universal Solvation Model Based on Solute Electron Density and a Continuum Model of the Solvent Defined by the Bulk Dielectric Constant and Atomic Surface Tensions', *J. Phys. Chem. B* **2009**, *113*, 6378-6396.

[61] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. Montgomery, J. A., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, M. J. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, O. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, D. J. Fox, Gaussian, Inc., Wallingford, CT, 2009.

[62] S. E. Wheeler, K. N. Houk, 'Integration Grid Errors for Meta-GGA-Predicted Reaction Energies: Origin of Grid Errors for the M06 Suite of Functionals', *J. Chem. Theory Comput.* **2010**, *6*, 395-404.

[63] S. N. Steinmann, C. Corminboeuf, 'A System-Dependent Density-Based Dispersion Correction', *J. Chem. Theory Comput.* **2010**, *6*, 1990-2001.

[64] S. N. Steinmann, C. Corminboeuf, 'A Density Dependent Dispersion Correction', *Chimia* **2011**, *65*, 240-244.

[65] S. N. Steinmann, C. Corminboeuf, 'A Generalized-Gradient Approximation Exchange Hole Model for Dispersion Coefficients', *J. Chem. Phys.* **2011**, *134*, 044117.

[66] S. N. Steinmann, C. Corminboeuf, 'Comprehensive Benchmarking of a Density-Dependent Dispersion Correction', *J. Chem. Theory Comput.* **2011**, *7*, 3567-3577.

[67] J. P. Perdew, K. Burke, M. Ernzerhof, 'Generalized Gradient Approximation Made Simple', *Phys. Rev. Lett.* **1996**, *77*, 3865-3868.

[68] C. Adamo, V. Barone, 'Toward reliable density functional methods without adjustable parameters: The PBE0 model', *J. Chem. Phys.* **1999**, *110*, 6158-6170.

[69] C. Fonseca Guerra, J. G. Snijders, G. te Velde, E. J. Baerends, 'Towards an order-N DFT method', *Theor. Chem. Acc.* **1998**, *99*, 391-403.

[70] G. te Velde, F. M. Bickelhaupt, S. J. A. van Gisbergen, C. Fonseca Guerra, E. J. Baerends, J. G. Snijders, T. Ziegler, 'Chemistry with ADF', *J. Comput. Chem.* **2001**, *22*, 931-967.

[71] A. Klamt, 'The COSMO and COSMO-RS Solvation Models', *WIREs Comp. Mol. Sci.* **2011**, *1*, 699-709.

[72] S. Grimme, 'Supramolecular Binding Thermodynamics by Dispersion-Corrected Density Functional Theory', *Chem. Eur. J.* **2012**, *18*, 9955-9964.

[73] I. Funes-Ardoiz, R. S. Paton. 2016. GoodVibes: GoodVibes v1.0.2. http://doi.org/10.5281/zenodo.595246

HELVETICA

**Entry for the Table of Contents**