# Designing bifunctional alkene isomerization catalysts using predictive modelling

Landman, I.R.; Paulson, E.R.; Rheingold, A.L.; Grotjahn, D.B.; Rothenberg, G.

**DOI**
10.1039/c7cy01106g

**Publication date**
2017

**Document Version**
Final published version

**Published in**
Catalysis Science & Technology

Catalysis Science & Technology

PAPER

View Article Online
View Journal | View Issue





Cite this: *Catal. Sci. Technol.*, 2017, 7, 4842

# Designing bifunctional alkene isomerization catalysts using predictive modelling†

Iris R. Landman, [a] Erik R. Paulson,[b] Arnold L. Rheingold,[c] Douglas B. Grotjahn*[b] and Gadi Rothenberg *[a]

Received 2nd June 2017,
Accepted 19th September 2017

DOI: 10.1039/c7cy01106g

rsc.li/catalysis

Controlling the isomerization of alkenes is important for the manufacturing of fuel additives, fine-chemicals and pharmaceuticals. But even if isomerization seems to be a simple unimolecular process, the factors that govern catalyst performance are far from clear. Here we present a set of models that describe selectivity and activity, enabling the rational design and synthesis of alkene isomerization catalysts. The models are based on simple molecular descriptors, with a low computational cost, and are tested and validated on a set of eleven known Ru-imidazol-phosphine complexes and two new ones. Despite their simplicity, these models show good predictive power, with $R^2$ values of 0.60–0.85. Using a combination of principal components analysis (PCA) and partial least squares (PLS) regression, we construct a "catalyst map", that captures trends in reactivity and selectivity as a function of electrostatic charge on the N* atom, $E_{HOMO}$, polar surface area and the optimal mass substituents on P/distance Ru–P ratio. In addition to indicating "good regions" in the catalyst space, these models also give insight into mechanistic steps. For example, we find that the electrostatic charge on N*, $E_{HOMO}$ and polar surface area are crucial in the rate-limiting step, whereas the optimal mass of substituents on P/distance Ru–P is correlated with the product selectivity.

## Introduction

Although less common than terminal alkenes, internal alkenes are highly valuable chemical feedstocks, that are especially relevant in the fine-chemicals and pharmaceuticals sectors. The industrial applications of internal alkenes include high-value fuel additives, agrochemicals, flavours, fragrances and drugs. They are typically used in metathesis, C–C coupling, amination, and polymerisation reactions.[1–4]

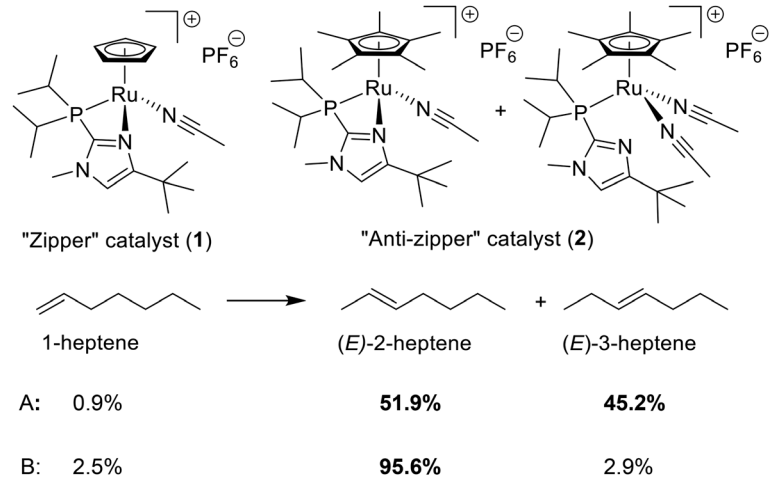
One way for making internal alkenes is by isomerising terminal alkenes. The most famous example is the isomerization of C6–C8 hydrocarbons to increase the relative octane number (RON) of gasoline.[1] A more specific example is the synthesis of eugenol from naturally occurring allylic isomers. In that isomerization process, both isomers are formed in a ratio of $E/Z$ = 20 : 1, while only the *E*-isomer is desired.[2] The ability to control both regiochemistry and stereochemistry is thus highly important in practical applications.

Over the past decade, the Grotjahn lab in San Diego has developed a set of bifunctional catalysts that show unique selectivity and activity for double bond isomerization. The first catalyst (**1**), the so-called alkene "zipper" (Scheme 1) can isomerize terminal alkenes to a mixture of internal ones, in the case of linear unbranched alkenes. Reactions are short (minutes in some cases) and conditions are mild. As a further bonus, **1** gives preferentially the *E*-isomers.[5,6] Subsequently a mixture of "anti-zipper" complexes (**2**) was developed which is even more selective, giving >95% yields of the 2-*E*-isomer.[7,8]

Many ligands were screened to try and decipher the factors that govern catalyst performance and thereby design new catalysts. But this is a complex and multi-dimensional problem that could not be solved using chemical intuition alone. We therefore turned to a combined experimental and predictive modelling workflow which was developed by the Rothenberg lab in Amsterdam (Fig. 1). This workflow is an iterative data-driven approach that enhances synthetic efforts by using descriptor models to highlight "good regions" in the catalyst space.[9] Such a data-driven approach has proven

[a] *Van 't Hoff Institute for Molecular Sciences, University of Amsterdam, Science Park 904, 1098XH, Amsterdam, The Netherlands.*
*E-mail: g.rothenberg@uva.nl; Web: http://hims.uva.nl/hcsc*
[b] *Department of Chemistry and Biochemistry, San Diego State University, 5500 Campanile Drive, San Diego, CA 92182-1030, USA.*
*E-mail: dbgrotjahn@mail.sdsu.edu; Web: http://grotjahngroup.org*
[c] *Department of Chemistry and Biochemistry, University of California San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0358, USA*

† Electronic supplementary information (ESI) available: Variable importance and mass-balance plots, fully assigned $^1$H and $^{13}$C NMR spectra for all new structures, an explanation of the modelling methods and the complete dataset of all the descriptors used in this study and their respective values. CCDC 1550064 and 1570068. For ESI and crystallographic data in CIF or other electronic format see DOI: 10.1039/c7cy01106g




**Scheme 1** The "zipper" and "anti-zipper" ruthenium catalysts for the isomerization of 1-heptene.[8] Reaction conditions A: 1 mol% **1**, RT, 4 h, B: 1 mol% **2**, 40 °C, 48 h.

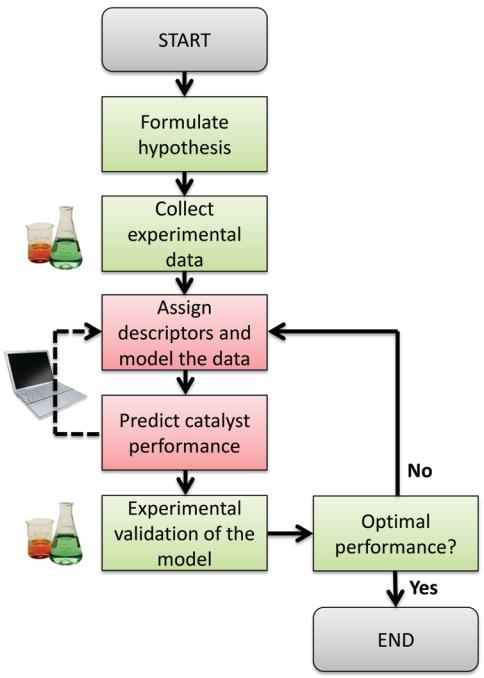



**Fig. 1** Flowchart showing the process of data collection, model building and validation.[25] The process starts with a hypothesis on the structure of a 'good' catalyst. Then, experimental data is collected for a set of catalysts, aiming at diversity in structure and performance. This first set is assigned descriptor values, and from those the model predicts the so-called figure of merit, in this case the catalyst performance. The next step is experimental validation: new catalysts whose performance was predicted by the model are synthesised and tested. If the performance criteria are reached, the process ends. Otherwise, the cycle iterates back to the descriptor assignment step, adding the new experimental data to the dataset and updating the model. Meta-modelling cycles (modelling of virtual catalysts, denoted by the dashed line) can also be used to optimise the models' performance. For a detailed step-by-step explanation of the principles and applications of this workflow please see our tutorial review.[9]

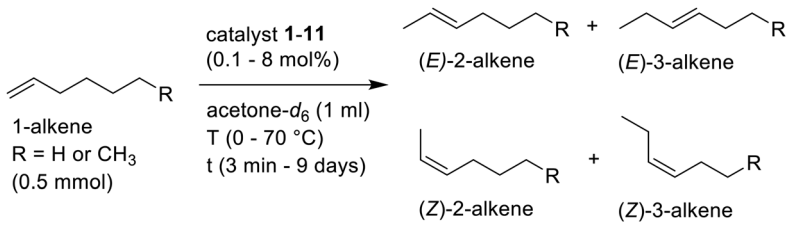
highly effective for predicting catalyst performance in a variety of reactions, including hydrocyanation,[10,11] fluoroarylation,[12] hydrogenation,[13,14] oxidation,[15,16] C–H functionalization,[17] oxygen reduction,[18] C–N coupling,[19] and C–C coupling.[20–24]

Here, in a collaboration between the two groups, we applied this workflow to predicting the activity and selectivity of ruthenium zipper catalysts for isomerising 1-hexene or 1-heptene to the respective internal 2-*E* and 3-alkenes (Scheme 2). The model's predictive power was demonstrated by synthesising and testing two new catalysts. Moreover, the structure–activity relationships generated by the model give new insights, helping to understand the reaction mechanism.

## Experimental

### Materials and instrumentation

All reactions involving organoruthenium complexes were performed under nitrogen, using standard Schlenk techniques and a glovebox. All solvents were dried by standard methods and distilled or flushed under nitrogen before use. The syntheses of catalysts **1**, **2**,[8] **3**, **4**, **5**, **6**, **7**, **8**,[26] **10**,[27] **11** and **12** are or will be reported elsewhere. Catalysts **9** and **13** are new compounds, prepared as detailed below by modifying the procedure published by Larsen *et al.*[8] Ligands and precursors were prepared according to published procedures.[8,26,27] All other chemicals were obtained from Aldrich and used without further purification. Unless stated otherwise, NMR spectra were recorded at 25 °C on a Varian instrument (400 MHz or 500 MHz, relaxation delay 10 s, observation pulse 15°) with tetramethylsilane as internal standard. $^{1}$H NMR chemical shifts are reported in parts per million and referenced to solvent resonances ($^{1}$H NMR: 2.05 ppm for acetone-$d_6$). $^{31}$P NMR chemical shifts were referenced to −145 ppm of $PF_6^-$ relative to external 85% $H_3PO_4$. X-ray crystallographic data were obtained at the UCSD Chemistry and Biochemistry X-ray facility, using Bruker single-crystal diffractometers with CCD detectors and low-temperature cryostats (100 K, wavelength of 0.71073 Å, Mo radiation).




**Scheme 2** General reaction scheme showing all the variables of the alkene isomerization reaction.

### Procedure for synthesising catalysts 9 and 13

**[CpRu(*t*-Bu$_2$PBZI)CH$_3$CN]PF$_6$ (9).** First, we synthesised the ligand 2-di-*t*-butyl-*N*-methyl-benzimidazolephosphine by reacting the benzimidazole with *n*BuLi followed by ClP(*t*-Bu)$_2$, as reported procedure for close analogues.[8,26,27] Catalyst **9** was then synthesized as follows. In a nitrogen-filled glovebox, [CpRu(NCCH$_3$)$_3$]PF$_6$ (84.3 mg, 0.194 mmol) was weighed out in a scintillation vial equipped with a magnetic stirring bar. The phosphine ligand (53.7 mg, 0.194) was weighed out in a separate scintillation vial and dissolved in 2 ml acetone. The phosphine solution was added slowly dropwise to the solution of [CpRu(NCCH$_3$)$_3$]PF$_6$ and the vial containing the ligand solution was rinsed with acetone (3 × 0.5 mL). After stirring for 20 h at room temperature, the completion of the reaction was checked by $^1$H NMR and $^{31}$P NMR spectroscopy, using J. Young NMR tubes. The solvent was evaporated and acetone (2 ml) was added, followed by evaporation. This step was then repeated two more times, resulting in yellow-orange foam, which was crushed and dried under vacuum for 1 h, leaving a yellow-orange powder (120.4 mg, 98.7% yield). The complex was characterized by NMR spectroscopy and X-ray crystallography. This is a new compound, see structure in Fig. 2 (also deposited in CCDC database, CSD-1550064).[28] $^1$H NMR (400 MHz, acetone-$d_6$) $\delta$ 7.83–7.75 (m, 2H), 7.55–7.46 (m, 2H), 4.71 (s, 5H, cyclopentadienyl), 4.15 (s, 3H), 2.36 (s, 3H, CH$_3$CN–Ru), 1.55 (broad s, 9H), 1.51 (broad s, 9H). $^{13}$C NMR (125.7 MHz, acetone-$d_6$): 158.1 (d, 10.2 Hz), 142.1 (d, 16.4 Hz) 136.7, 125.8, 124.9, 118.6, 112.6, 71.2, 33.7. $^{31}$P NMR (162 MHz, acetone-$d_6$) $\delta$ 70.15 (s, 0.08 units; possibly a small amount of corresponding bis-nitrile complex), 67.12 (s, 7.80 units, P–Ru), −145.00 (sept, $J$ = 707.5 Hz, 10 units, PF$_6$). The full 2D NMR analysis is included in the ESI.†

**[CpRu((*t*-Bupim)$_3$P)CH$_3$CN]PF$_6$ (13).** Catalyst **13** was similarly prepared starting from [CpRu(NCCH$_3$)$_3$]PF$_6$ (49.9 mg, 0.115 mmol) and (*t*-Bupim)$_3$P (60.7 mg, 0.115 mmol).[29] A yellow-orange-powder (83.3 mg, 82.7% yield) was obtained after repeated addition and evaporation with acetone. Catalyst **13** is a new compound and was characterized by NMR spectroscopy and X-ray crystallography. See structure in Fig. 3 (also deposited in CCDC database, CSD-1570068). $^1$H NMR (400 MHz, acetone-$d_6$) $\delta$ 7.46 ppm (d, $J$ = 1.5 Hz, 1H), 7.42 (d, $J$ = 2.5 Hz, 1H), 7.33 (s, 1H), 5.28–5.16 (m, 2H), 5.14–5.05 (m, 1H), 4.78 (s, 5H, cyclopentadienyl), 2.07 (d, $J$ = 2.0 Hz, CH$_3$CN–Ru), 1.59 (d, $J$ = 6.7 Hz, 3H), 1.56 (d, $J$ = 6.8 Hz, 3H), 1.54 (d, $J$ = 6.9 Hz, 3H), 1.47 (d, $J$ = 6.6 Hz, 3H), 1.40 (s, 9H), 1.27 (s, 9H), 1.19 (s, 9H), 0.97 (d, $J$ = 6.8 Hz, 3H), 0.89 (d, $J$ = 6.6 Hz, 3H). $^{13}$C NMR (125.7 MHz, acetone-$d_6$) 155.4 ppm (d, $J$ = 12.9 Hz), 154.3 (d, $J$ = 10.0 Hz), 153.6 (d, $J$ = 16.5 Hz), 144.6 (d, $J$ = 56.6 Hz), 133.3 (d, $J$ = 91.3 Hz), 131.4 (d, $J$ = 85.9 Hz), 117.3, 116.2, 114.5, 72.7 (d, $J$ =

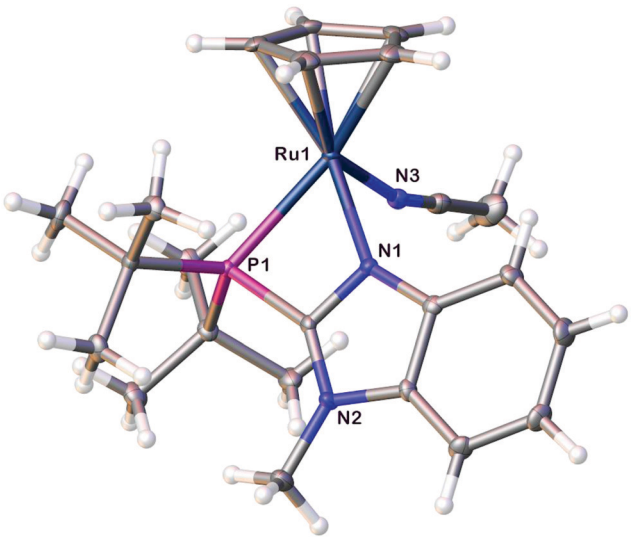



**Fig. 2** X-ray crystal structure of **9**. The counter anion is omitted for clarity.

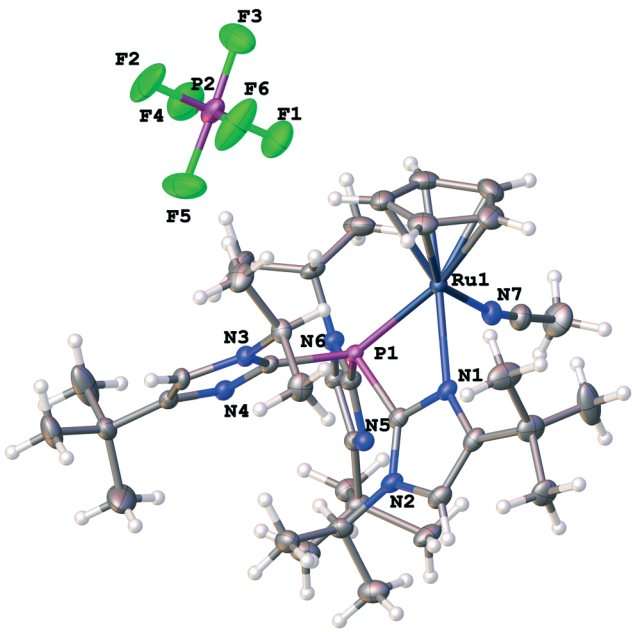



**Fig. 3** X-ray crystal structure of **13**. The counter anion is omitted for clarity.

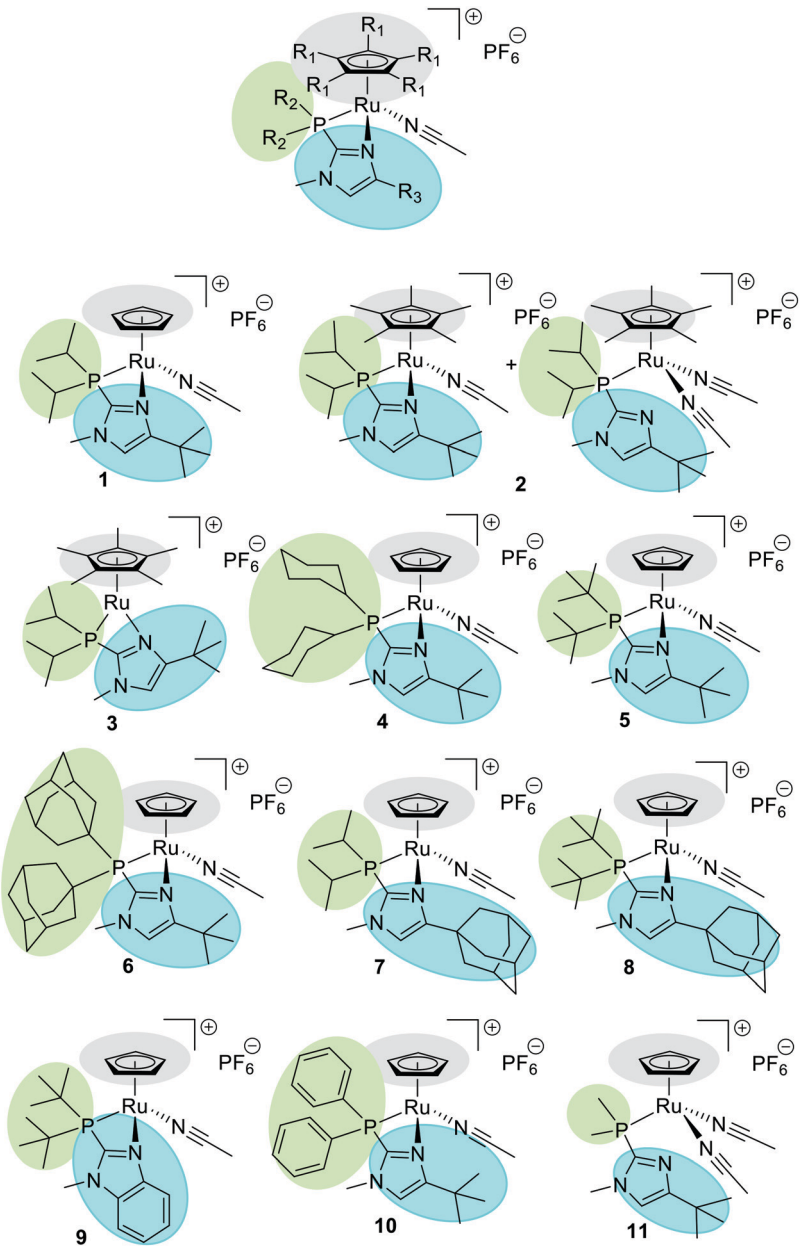
2.6 Hz), 51.3 (d, $J$ = 7.7 Hz), 50.9 (s), 50.2 (d, $J$ = 4.0 Hz), 32.9, 32.8, 32.5, 30.7, 30.4, 30.2, 24.1, 23.9, 23.7, 23.6, 23.5, 23.3, 2.8. $^{31}$P NMR (162 MHz, acetone-$d_6$) $\delta$ 8.76 (s, 0.33 units, small amount of bis-acetonitrile complex, *ca.* 3 mol%), −25.84 (s, 11.4 units, P–Ru), −142.81 (sept, $J$ = 705.97 Hz, 10 units, $PF_6$). The full 2D NMR analysis is included in the ESI.†

### General procedure for catalyst testing

In a glove box, a few grains of internal standard tetramethylsilane and substrate (0.5 mmol) were combined with acetone-$d_6$ (0.9 ml) in a J. Young NMR tube, sonicated for 10 min and an initial $^1$H NMR spectrum was recorded. Back in the glove box, the catalyst in acetone-$d_6$ (0.1 ml) was added to the substrate, $t_0$. Spectra were then recorded periodically until >99% conversion was reached (or until 6–9 days in cases of very low or no conversion). For the experiments at 40 °C and 70 °C the NMR tube was placed in a pre-heated oil bath outside the glovebox. For the low-temperature experiments, the 500 MHz Varian spectrometer was equipped with liquid nitrogen to cool the probe to the desired temperature.

### Computational methods

Descriptor calculation, analysis and data mining were performed on an Apple MacBook Pro with a 2.9 GHz Intel




**Scheme 3** General catalyst structure showing the four parts of the metal–ligand complex (above) and the corresponding structures of catalysts **1–11** (below).

Core i7 processor. Principal components analysis (PCA) and partial least squares (PLS) regression models were run using the JMP pro software.[30] The principal components were calculated by using the NIPALS algorithm,[31] which calculates the components in their order of explaining the variance in the data. All models were validated using leave-one-out cross-validation. A variable importance (VIP) analysis[32] was performed to determine the importance of the descriptors related to the FOMs (see for example Fig. S1†).

Geometry optimization was performed on the ligand–metal complexes using semi-empirical PM3 methods within the Spartan program.[33] To validate the calculations, the bond distances and angles were compared with their crystallographic data. The optimized structures were then used as the starting point for the calculation of all the descriptors. Using the programs Spartan,[33] ChemDes,[34] and Marvin Sketch[35] a total of 308 constitutional, topological, geometrical and electrostatic descriptors were computed. The entire 308 × 39 dataset matrix is included as part of the ESI.†

### Statistical data validation

First of all, the mass balance of all reactions in the dataset was checked to determine the quality and identify possible outliers. Two outliers were identified using the modified Thompson Tau test, Dixon's *Q*-test and the Grubbs test,[36,37] and were excluded from the data set. The other data points had a 5% deviation centred on 100%, and were spread randomly (see Fig. S2†), confirming that there was no systemic error in the reaction mass balance.

The models were validated using both internal validation and external validation. Internal validation was performed with the leave-one-out (LOO) algorithm within the JMP Pro program. For the external validation, the data was divided into a training set, a test set and a validation set. The catalysts for the training set and test set were used to generate the model, of which the test set provides internal validation by the leave-one-out method. The resulting model is then applied to a validation set (data that the model has never "seen"), to determine its predictive power.

## Results and discussion

### Catalyst synthesis, characterisation and testing

We started by preparing a set of catalysts (structures **1–11**, Scheme 3) following the general procedure of adding the phosphine ligand solution to the Ru-precursor solution in 1 : 1 molar ratio, after which the solvent was evaporated.[8,26,38] The catalysts were then characterised using NMR, and selected examples by X-ray crystallography.

We chose to focus on ruthenium complexes and use the predictive modelling tools for designing new ligands and thereafter synthesizing new ruthenium(II)-complexes for alkene isomerization. We have divided the catalyst structure into four regions of modification (Scheme 3): (i) $R_1$ (gray) influences the sterics and electronics on the Cp; here we used

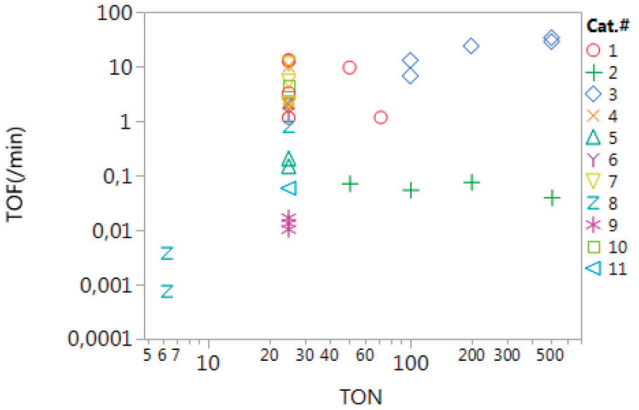



**Fig. 4** Log/log plot showing the TOF at $t_{50}$ (min) *vs.* TON at 50% conversion for the catalytic isomerization of 1-hexene and 1-heptene using catalysts **1–11**. Multiple points for each catalyst indicate different reaction conditions, such as temperature, catalyst loading or different substrate.

$R_1$ = H and more sterically hindered $R_1$ = $CH_3$. (ii) $R_2$ (green) controls the sterics and electronics of the phosphine; here we applied a range of substituents to study the effect of changing steric hindrance. (iii) $R_3$ (blue) provides steric control of the basic nitrogen environment. As in the case of $R_2$, we studied the effect of changing steric hindrance at $R_3$. (iv) Each catalyst can have zero, one or two acetonitrile molecules, a parameter we used to analyse the rate differences between the different catalysts.

Subsequently, we tested catalysts **1–11** in the isomerization of 1-hexene and 1-heptene (Scheme 2), following the reaction progress using NMR spectroscopy. In a typical reaction, 0.5 mmol of 1-hexene or 1-heptene were mixed with 0.1–8.0 mol% of catalyst in a sealed NMR tube and held at various temperatures (0–70 °C; see details in Experimental section).

With this setup, we generated our initial experimental dataset using catalysts **1–11**. We chose the TON and TOF (calculated based on $t_{50}$, the time at which 50% conversion is

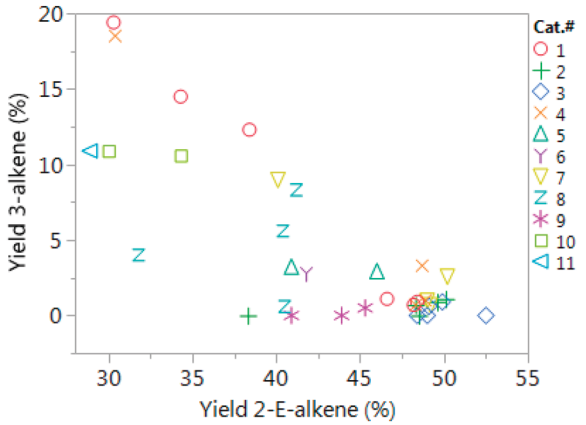



**Fig. 5** Yield of 3-alkene *vs.* yield of 2-*E*-alkene at 50% conversion using catalysts **1–11**. Multiple points for each catalyst indicate different reaction conditions, such as temperature, catalyst loading or different substrate.

reached), as the figures of merit (FOMs) to give direct information on the catalytic activity and stability. Fig. 4 compares the catalysts' performance. The values vary considerably: three orders of magnitude for the TONs and up to six orders of magnitude for the TOFs (due to the large range, the plot is shown on a double logarithmic scale). This shows that our catalysts span a large part of the performance space. Note that because the TON values are based on 50% conversion, there are multiple data points of exactly 25, but with a different TOF.

To evaluate the regioselectivity, we used the relative yields of the 3-alkene and 2-*E*-alkene. In Fig. 5, some catalysts give only 2-*E*-alkene and no 3-alkene, such as catalyst **9**, whereas catalyst **1** gives also 3-alkene. Note that in some cases the yields in Fig. 5 do not add up to 50% conversion, reflecting the formation of some *Z*-product.

### Selecting catalyst descriptors and building predictive models

With the dataset at hand, we built a descriptor model, correlating the reaction data with descriptors containing structural and/or physical information. A detailed discussion on descriptor selection is published elsewhere,[39] but in general descriptors should be chemically meaningful, representative for the catalytic system, and simple and fast to calculate. To find the most relevant descriptors, we started from a large set of 308 constitutional, topological, geometrical and electrostatic parameters that were screened and related to the FOMs using the JMP program.[15] Naturally, such a large number of descriptors must be reduced to avoid the risk of

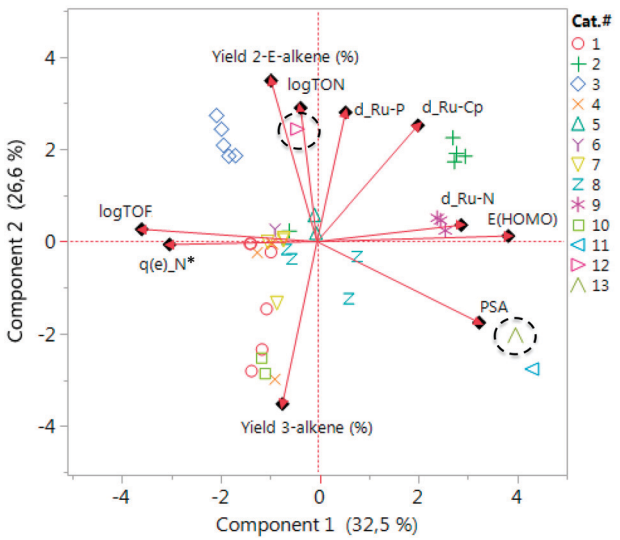
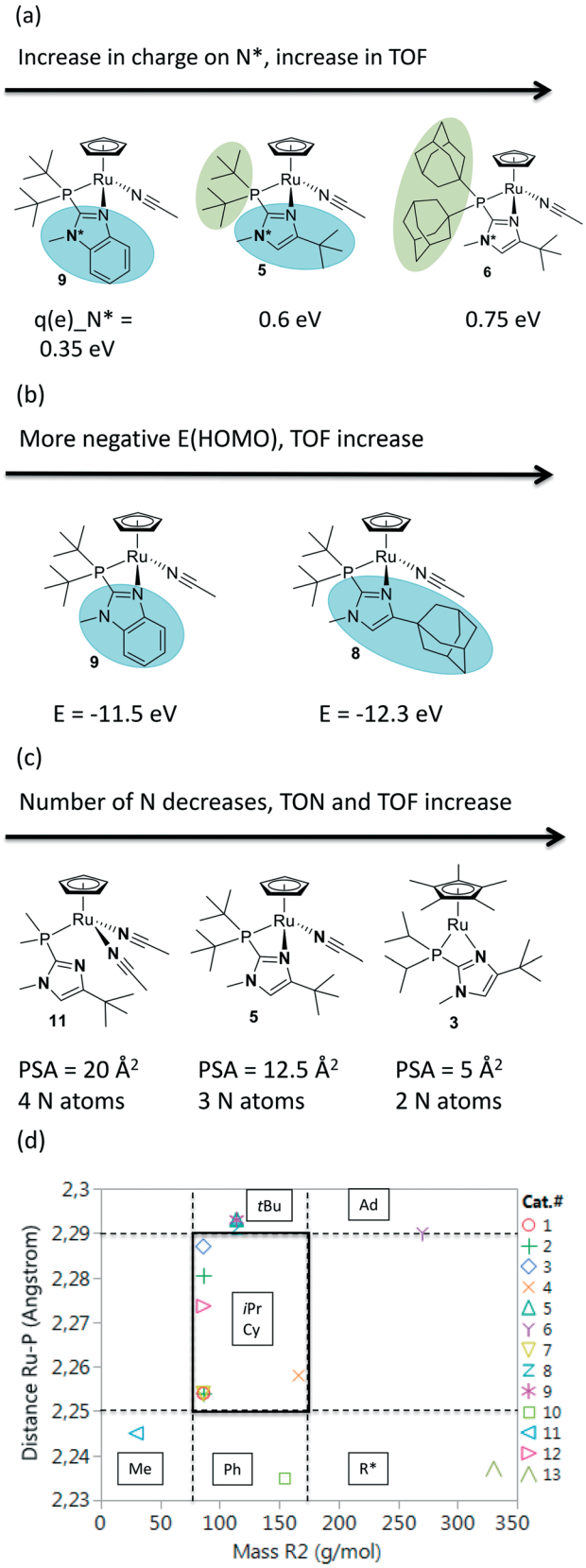
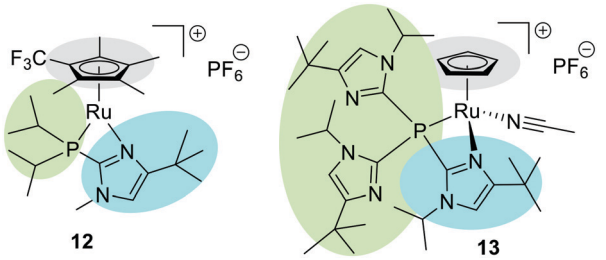



**Fig. 6** PCA biplot showing the catalysts, descriptors and figures of merit. Each descriptor or FOM is shown as an arrow. The length of each arrow corresponds to the size of the contribution coefficient of this parameter, and the direction shows the correlation between different parameters. The first two PCs explain 59% of the variance in the data. The two points in circles are the catalysts of the prediction set.




**Scheme 4** The four trends inferred from the PCA biplot in Fig. 5. (a) Electrostatic charge on N*, *q*(e)_N*, (b) energy of the HOMO, *E*(HOMO), (c) polar surface area (Å$^2$), PSA, (d) distance Ru–P (Å)/mass $R_2$ (g mol$^{-1}$), R* = (*t*-Bupim)$_3$P, *d*_Ru–P. Note that the complexes are $PF_6$ salts.




**Scheme 5** External validation set of catalysts.

over-fitting and the likelihood of correlation between descriptors. In our approach, we start with an inclusive set (as many descriptors as possible) and then rank them and reduce their numbers drastically. In this way, you avoid "missing" important descriptors, and the ranking is done by a simple analytical formula. For more details on descriptor selection and ranking in homogeneous catalysis, see our tutorial review.[9]

The descriptors were ordered based on their contribution to the FOMs using the variable importance (VIP) technique.[40,41] For each FOM, we then chose those two descriptors with the highest contribution coefficients, giving a total of six important descriptors. Subsequently, we used principal component analysis (PCA) and partial least squares (PLS) regression, for building a correlation model between the descriptors and the figures of merit. The models were validated using both internal and external validation[32] (see Experimental section for details).

### Creating a "catalyst map" for alkene isomerization

The set of 11 catalysts for alkene isomerization can be described by six descriptors and four FOMs. Using PCA, we can create a "catalyst map", showing the relationship between the catalysts, descriptors and FOMs. Fig. 6 shows the PCA biplot, a combined plot of the loadings (descriptors) and the scores (experimental points). The various descriptors and FOMs are indicated as arrows. In each case, the correlation between each two parameters depends on the angle between the respective arrows. The arrow lengths indicate how much each descriptor or FOM contributes to a given PC. Descriptors that appear closer to the centre have small loadings and are less important, whereas important descriptors appear far from the centre. When two arrows point in the same direction, the descriptors are correlated, and *vice versa*. We can also see in this plot the spread of catalysts. Catalysts that are close together have similar performance.

From the catalyst map in Fig. 6 we can infer four trends (Scheme 4). (a) An increase in electrostatic charge on N* is correlated with an increase in logTOF. Both $R_2$ and $R_3$ influence the charge on N*, where increases in performance based on N* charge may be due to a resonance effect, *i.e.* increasing charge on N* could decrease the charge on the basic N, enhancing the basicity. (b) The energy of the HOMO is anti-correlated with logTOF. In other words, a more negative HOMO energy is correlated with higher TOF values. (c) The polar surface area is a measure of the area of polar atoms in the molecule. The number of nitrogens is

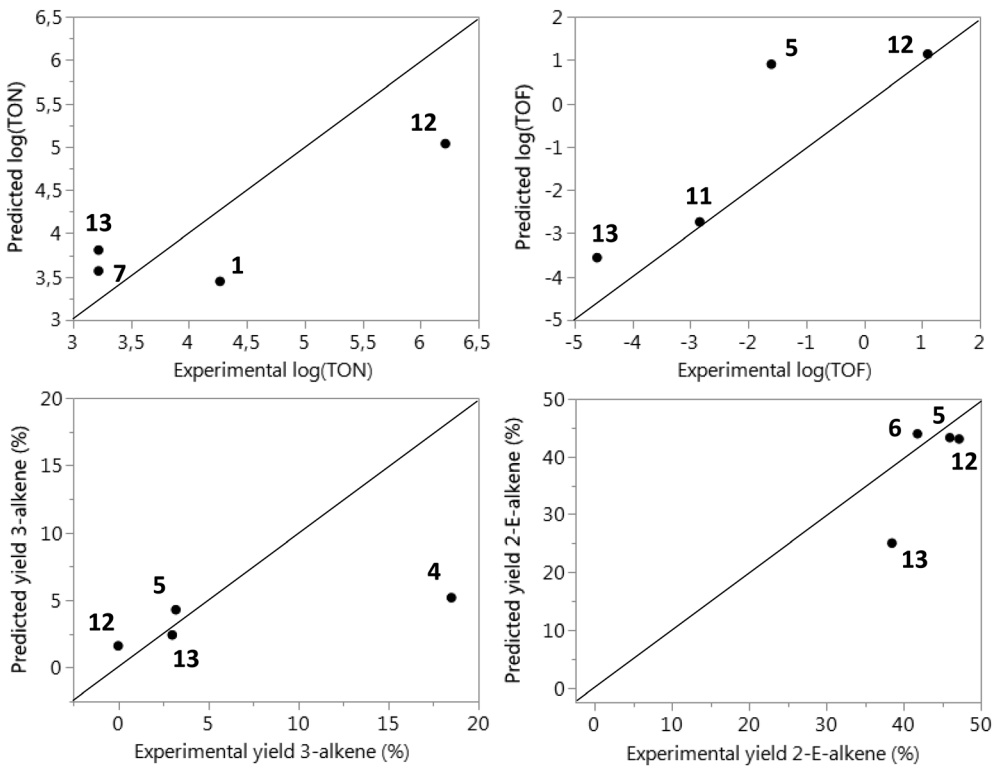
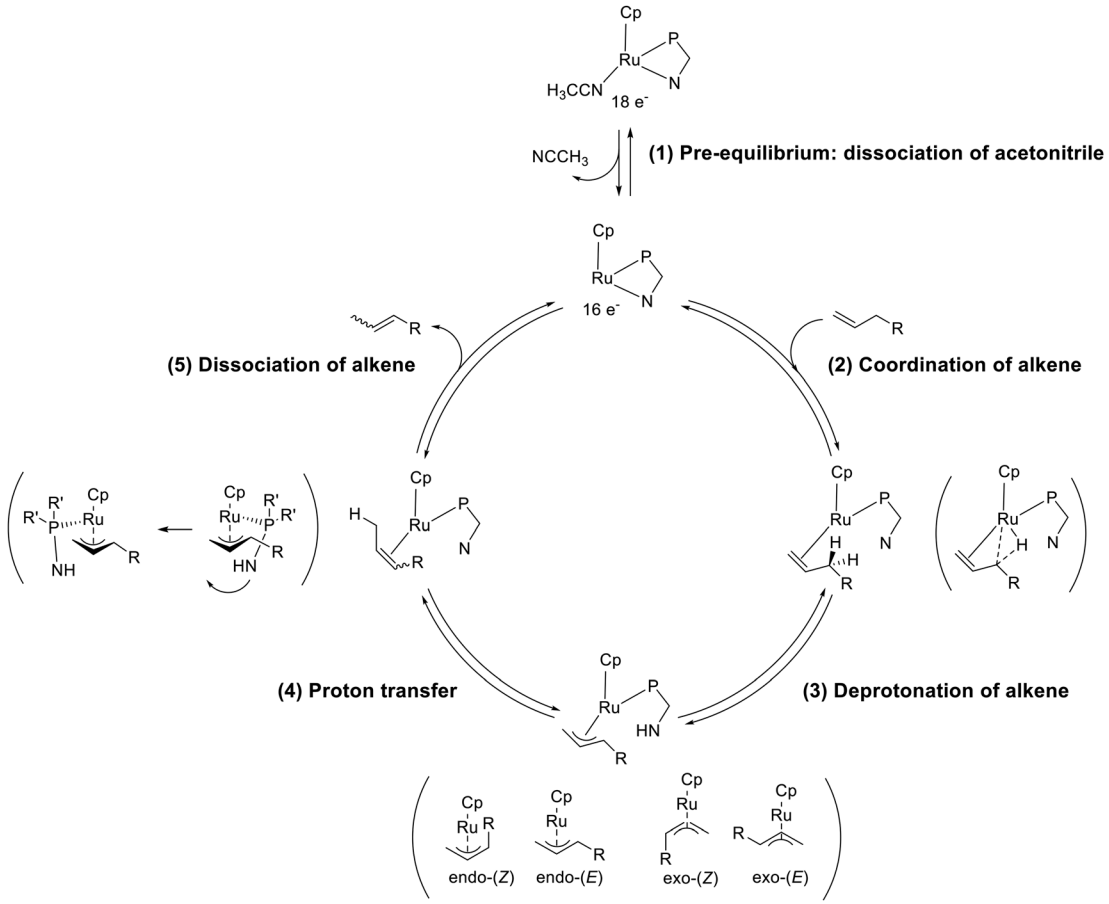



**Fig. 7** Predicted values compared to experimental values. Log(TON): $R^2$ = 0.74, RMSE = 0.5, log(TOF): $R^2$ = 0.79, RMSE = 1.3, yield 2-*E*-alkene: $R^2$ = 0.62, RMSE = 6.9, yield 3-alkene: $R^2$ = 0.66, RMSE = 1.2.




**Scheme 6** Proposed catalytic cycle.

anti-correlated with the TON and TOF. (d) The Ru–P distance, combined with the mass of $R_2$ (C + H, in g mol$^{-1}$), gives an optimal mass/distance region in which all the FOMs are maximised, a region that includes the substituents iPr and Cy on $R_2$.

### Predicting the performance of new catalysts

The real usefulness of a predictive model is in forecasting the performance of totally new catalysts. Based on the trends in the catalyst map, we designed two new catalysts. First, **12** was designed to add novel variety in the type of cyclopentadiene by replacing a methyl by a $CF_3$ substituent. We expected that the descriptor values of catalyst **12** would be similar to those of catalyst **3** (see biplot) and thus it should appear in the upper left quadrant of the plot. Second, **13** was designed to decrease the electrostatic charge on N* and increase the PSA by using a bulkier substituent than adamantyl (Scheme 5). The two catalysts were then synthesized and used to determine the predictive power of the models (see Experimental section for details).

Fig. 7 compares the results of the predicted and experimental values for each FOM. The $R^2$ values are in the 0.60–0.85 range, comparable to other descriptor models for homogeneous catalysts.[40] The TON and TOF models both show good predictions for the new catalysts **12** and **13** (note that the scales in Fig. 7 are logarithmic). We can distinguish catalysts with low, moderate or high TON/TOF within the catalyst space. The models of the yield of 2-*E*-alkene and 3-alkene resulted in lower $R^2$ values, which may reflect the larger number of data points at higher yields of 2-*E*-alkene and low yields of 3-alkene.

There are several ways to increase the models' predictive power, for example performing more accurate calculations using DFT methods and/or adding more experimental data. However, there is always a trade-off between a model's cost and its predictive power. We focus here on descriptors that are simple and quick to calculate, and the resulting range of $R^2$ = 0.60–0.85 is near optimal in trade-off.

Finally, the connection between the trends and the proposed mechanistic cycle was made (Scheme 6). For example, the number of nitrile ligands is related to proposed pre-equilibrium and with the modelling it was possible to quantify the effect of the amount of nitrogens by using PSA as descriptor. Next, during the proposed deprotonation of the alkene, an increased positive charge on N* could lead to a more basic nitrogen, enhancing the deprotonation. In the PCA biplot, any descriptors that are correlated with the TOF are suggested to be important parameters for discussing the overall rate of the reaction and hence the turnover-limiting

step. Thus PSA, electrostatic charge on N*, $E_{HOMO}$ and distance Ru–N appear to be important in the turnover-limiting step.

Conversely, the distances of Ru–P and Ru–Cp are not correlated with the TON and TOF in Fig. 6, which means that they are not important in the turnover-limiting step. Instead, these distance descriptors are correlated with the yields of 2-*E*-alkene and 3-alkene, meaning that they are important for product selectivity. For example, the size of $R_2$ and the Ru–P bond distance could be important in the proton transfer.

## Conclusions

Alkene isomerization by the zipper and anti-zipper catalysts can be described by figures of merit on activity and selectivity and the corresponding descriptors. Analysis leads to data-driven models that can predict the performance of new structurally and electronically different catalysts, with $R^2$ values between 0.60–0.85. Examining the correlation between the descriptors and the catalysts' performance gives subtle mechanistic information, which cannot be obtained by analysing the catalyst structures using chemical intuition alone.

## Conflicts of interest

There are no conflicts of interest to declare.

## Acknowledgements

We thank Dr. LeRoy Lafferty for assisting with NMR experiments. This work was co-financed by the National Science Foundation (ERP, DBG), the Holland Research School of Molecular Chemistry, the Royal Netherlands Chemical Society and the Research Priority Area Sustainable Chemistry of the University of Amsterdam (IRL, GR). http://suschem.uva.nl.

## References

1 G. W. Parshall, W. A. Nugent, D. M.-T. Chan and W. Tam, *Pure Appl. Chem.*, 1985, **57**, 1809.
2 S. K. Sharma, V. K. Srivastava and R. V. Jasra, *J. Mol. Catal. A: Chem.*, 2006, **245**, 200.
3 T. E. Müller, K. C. Hultzsch, M. Yus, F. Foubelo and M. Tada, *Chem. Rev.*, 2008, **108**, 3795–3892.
4 C.-J. Li, *Chem. Rev.*, 2005, **105**, 3095–3165.
5 D. B. Grotjahn, C. R. Larsen, J. L. Gustafson, R. Nair and A. Sharma, *J. Am. Chem. Soc.*, 2007, **129**, 9592–9593.
6 C. R. Larsen and D. B. Grotjahn, *J. Am. Chem. Soc.*, 2012, **134**, 10357–10360.
7 D. B. Grotjahn, C. R. Larsen and G. Erdogan, *Top. Catal.*, 2014, **57**, 1483–1489.
8 C. R. Larsen, G. Erdogan and D. B. Grotjahn, *J. Am. Chem. Soc.*, 2014, **136**, 1226–1229.
9 A. G. Maldonado and G. Rothenberg, *Chem. Soc. Rev.*, 2010, **39**, 1891.
10 E. Burello, P. Marion, J. C. Galland, A. Chamard and G. Rothenberg, *Adv. Synth. Catal.*, 2005, **347**, 803–810.
11 N. Fey, M. Garland, J. P. Hopewell, C. L. McMullin, S. Mastroianni, A. G. Orpen and P. G. Pringle, *Angew. Chem., Int. Ed.*, 2012, **51**, 118–122.
12 R. T. Thornbury, V. Saini, T. D. A. Fernandes, C. B. Santiago, E. P. A. Talbot, M. S. Sigman, J. M. McKenna and F. D. Toste, *Chem. Sci.*, 2017, **8**, 2890–2897.
13 E.-J. Ras, M. J. Louwerse and G. Rothenberg, *Catal. Sci. Technol.*, 2012, **2**, 2456–2464.
14 P. J. Donoghue, P. Helquist, P. O. Norrby and O. Wiest, *J. Am. Chem. Soc.*, 2009, **131**, 410–411.
15 N. Madaan, N. R. Shiju and G. Rothenberg, *Catal. Sci. Technol.*, 2016, **6**, 125–133.
16 P. E. Gormisky and M. C. White, *J. Am. Chem. Soc.*, 2013, **135**, 1–40.
17 E. N. Bess, D. M. Guptill, H. M. L. Davies and M. S. Sigman, *Chem. Sci.*, 2015, **6**, 3057–3062.
18 Y. Jiao, Y. Zheng, M. Jaroniec and S. Zhang Qiao, *J. Am. Chem. Soc.*, 2014, **136**, 4394–4403.
19 A. Milo, A. J. Neel, F. D. Toste and M. S. Sigman, *Science*, 2015, **347**, 737–743.
20 E. Burello, D. Farrusseng and G. Rothenberg, *Adv. Synth. Catal.*, 2004, **346**, 1844–1853.
21 E. Burello and G. Rothenberg, *Adv. Synth. Catal.*, 2003, **345**, 1334–1340.
22 Z. L. Niemeyer, A. Milo, D. P. Hickey and M. S. Sigman, *Nat. Chem.*, 2016, **8**, 610–617.
23 W. Smit, V. Koudriavtsev, G. Occhipinti, K. W. Törnroos and V. R. Jensen, *Organometallics*, 2016, **35**, 1825–1837.
24 O. J. S. Pickup, I. Khazal, E. J. Smith, A. C. Whitwood, J. M. Lynam, K. Bolaky, T. C. King, B. W. Rawe and N. Fey, *Organometallics*, 2014, **33**, 1751–1761.
25 A. G. Maldonado and G. Rothenberg, *Chem. Eng. Prog.*, 2009, **105**, 26–32.
26 G. Erdogan, Metal complexes with bifunctional imidazolyl phosphines for catalytic organic transformations: applications in homogeneous and polymer supported alkene isomerization, and hydrogen deuterium exchange, *Ph.D. thesis*, San Diego State University, 2012.
27 C. R. Larsen, A novel bifunctional catalyst for alkene isomerization: development, scope and limitations, and applications in organic transformations, *Ph.D. thesis*, San Diego State University, 2012.
28 C. R. Groom, I. J. Bruno, M. P. Lightfoot and S. C. Ward, *Acta Crystallogr., Sect. B: Struct. Sci., Cryst. Eng. Mater.*, 2016, **72**, 171–179.
29 T. N. Sorrell, W. E. Allen and P. S. White, *Inorg. Chem.*, 1995, **34**, 952–960.
30 *JMP pro, Version 12*, SAS Institute Inc., Cary, NC, 2015–2016.
31 S. Wold, K. Esbensen and P. Geladi, *Chemom. Intell. Lab. Syst.*, 1987, **2**, 37–52.
32 G. Rothenberg, *Catal. Today*, 2008, **137**, 2–10.
33 Spartan is distributed by Wavefunction Inc., 18401 Von Karman Ave., #370, Irvine, CA 92612 USA.
34 *ChemDes*, Comutational Biology & Drug Design Group at the School of Pharmaceutical Sciences, Central South University, 2012–2017.
35 *Marvin Sketch version 16.7.4.0*, ChemAxon, 2016.

36 F. E. Grubbs, *Technometrics*, 1969, **11**, 1–21.
37 D. B. Rorabacher, *Anal. Chem.*, 1991, **63**, 139–146.
38 E. R. Paulson, *Ph.D. thesis*, San Diego State University, 2017.
39 E. Burello and G. Rothenberg, *Adv. Synth. Catal.*, 2005, **347**, 1969–1977.
40 E. Burello and G. Rothenberg, *Int. J. Mol. Sci.*, 2006, **7**, 375–404.
41 Z. Strassberger, M. Mooijman, E. Ruijter, A. H. Alberts, A. G. Maldonado, R. V. A. Orru and G. Rothenberg, *Adv. Synth. Catal.*, 2010, **352**, 2201–2210.