Perturbation theory model of reactivity and enantioselectivity of palladium-catalyzed Heck–Heck cascade reactions

C. Blázquez-Barbadillo, E. Aranzamendi, E. Coya, E. Lete, N. Sotomayor and H. González-Díaz

Perturbation Theory Model of Reactivity and Enantioselectivity of Palladium-catalyzed Heck-Heck cascade reactions

/

C. Blázquez-Barbadillo,[a] E. Aranzamendi,[a] E. Coya,[a] E. Lete,[a] N. Sotomayor[a] and H. González-Díaz[a,b*]

Enantioselective intramolecular Heck-Heck cascade reactions have emerged as an excellent tool for the construction of polycyclic frameworks, such as Lycorane alkaloids, Xestoquinone and analogues. However, it is particularly difficult to rationalize the effect of simultaneous changes in both the structure of many molecular entities and experimental conditions (temperature, time, solvent, ligand, catalyst loading, *etc.*) on reactivity and enantioselectivity. In this work, a computational model to predict the enantiomeric excess and the yield of Heck-Heck cascade reactions has been developed. The model combines Perturbation Theory (PT) and Quantitative Structure-Reactivity Relationships (QSRR) ideas for the prediction of two different outputs with the same equation (% ee and % yield). This model predicted 520 experimental outcomes with a correlation coefficient R = 0.89, standard error of estimates SEE = 1.19 %, and a cross-validation correlation coefficient q^2 = 0.79. The use of the model has been illustrated with a case of study, the Heck-Heck cascade reaction of a 2,3-dialkenyl pyrrole using Pd(dba)$_2$ and (R)-BINAP. For the first time, a 2000-points simulation in a ternary phase diagrams shows the effect of the concentration of the catalyst, the base, and ligand on the enantioselectivity of this reaction. The QSRR model also predicts trends in structural outcomes, such halides vs. triflates, or the ligand structure. Therefore, the model opens the door to the design of new chiral ligands and helps to find trends to improve the experimental results in enantioselective polyene cyclisations.

Introduction

Mizoroki-Heck reaction[1] is one of the most important palladium(0)-catalyzed C-C bond-forming reactions, which has allowed the preparation of complex organic molecules. In particular, asymmetric Heck reaction[2] is a highly efficient method for obtaining optically active molecules with tertiary and quaternary stereocentres. In this context, the use of Heck-Heck cascade reactions[3] is a powerful synthetic strategy for rapidly increasing structural and stereochemical complexity. In this case, the σ-alkylpalladium intermediate, resulting from the migratory insertion of the arylpalladium to the alkene, may undergo an insertion with an alkene in inter- or intramolecular way, so carbopalladation is repeated one or several times prior to β-hydride elimination. This class of polyene cyclisation reactions has expanded the synthetic utility of enantioselective intramolecular Heck reaction, allowing the construction of polycyclic frameworks in a one-pot operation. For example, a 6-*exo* Mizoroki-Heck reaction on adequately dialkenyl substituted substrates can generate a quaternary stereocentre, giving rise to the σ-alkylpalladium intermediate that undergoes a 6-*endo* insertion to give the polycyclic system (Scheme 1). In this regard, we have recently reported that the Lycorane tetracyclic framework of *Amaryllidaceae* alkaloids can be formed in one-pot through an enantioselective palladium-catalyzed intramolecular Heck-Heck cascade process, starting from halogenated 2,3-dialkenyl (hetero)arylmethylpyrroles.[4] It should be pointed out that it has been possible to control the selectivity and switch the reaction to the alkene, avoiding competition with direct

arylation, which could arise with this type of substrates as we have demonstrated earlier.[5]

A similar strategy had been previously applied by Keay and co-workers in the total synthesis of marine natural product, (+)-Xestoquinone using aryl triflates as coupling partners.[6] The method was extended to the corresponding aryl halides by Shibasaki.[7]
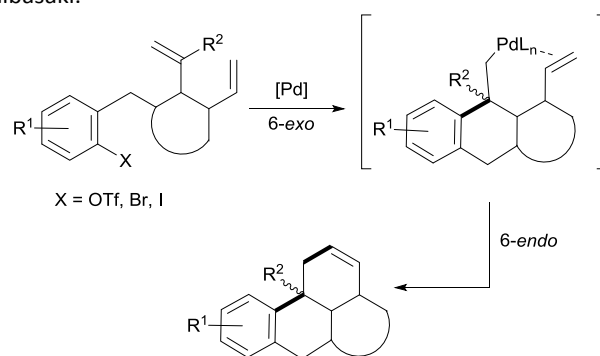
Scheme 1 General scheme of a Heck-Heck cascade reaction

The cyclisation of these aryl polyene triflates has been studied theoretically using a density functional theory (DFT) model, which demonstrated that the main factor controlling the *exo/endo* selectivity, at both thermodynamic and kinetic levels, was the relative stability of the cyclic systems resulting from the migratory step. The DFT calculations also showed that the furan ring of the substrate can play an important role in the control of the *exo/endo* selectivity of the first insertion reaction rather than in the enantioselectivity.[8]

Although computational chemistry has indeed helped to understand the mechanism of these palladium-catalyzed cross

coupling reactions,[9] it is still difficult to rationalize how the different parameters of a Heck reaction affect its stereochemical outcome.[10] Thus, one of the underlying challenges in asymmetric catalysis is the selection or the design of the adequate catalyst (or ligand) and experimental conditions, without engaging in a long term empirical investigation. Therefore, we sought to develop computational chemistry methods for the prediction of the enantioselectivity and the yield for these cascade reactions.

The development of new computational models to predict reactivity has become an important goal.[11] Sigman[12] and co-workers have demonstrated that Quantitative Structure-Reactivity Relationship (QSRR) models can be applied in asymmetric catalytic reactions. They have evaluated steric/enantioselection relationships in different types of reactions using physical parameters, such as Charton[13] and Sterimol[14] values. Subsequently, methodologies based on QSRR modelling have been applied to predict enantioselectivity of different types of reactions, *e.g.* asymmetric copper-catalyzed cyclopropanation of alkenes[15] or Henry reaction.[16]

QSRR models use molecular descriptors (numerical parameters of chemical structure) as input to predict chemical reactivity parameters (output). Many of these QSRR models using Linear Free Energy Relationships (LFER) have been reported in this area.[17] The classic Hammett's method predicts reactivity constants ε_i using as inputs different LFER potentials $^i\mu_k$ of type k for the i-th experiment.[18] Hansch extended the LFER method using an extra-thermodynamic approach. Thus, other input parameters or functions, such as molecular refractivities (MR), were used.[19] The QSRR method is also flexible for the selection of different output functions $f(\varepsilon_i)$ of the original reactivity constant.[20] Examples of Hansch's QSRR models based on the LFER method are shown in Equations 1 and 2.

$$\varepsilon_i = b_0 + b_1 \cdot \mu_1(\text{partition}) + b_2 \cdot \mu_2(\text{ionization}) = b_0 + b_1 \cdot \log P + b_2 \cdot pK_a \quad (1)$$

$$f(\varepsilon_i) = a_0 + a_1 \cdot \log P + a_2 \cdot pK_a + a_3 \cdot MR + a_4 \cdot (\log P)^2 \quad (2)$$

On the other hand, Perturbation Theory (PT) ideas, such as the Møller and Plesset (MP) perturbation theory,[21] could be used to predict chemical reactivity. For instance, Elm and Jørgensen[22] used MP2 and other QM methods to study gas phase hydrogen abstraction reaction kinetics of short chained oxygenated hydrocarbons. Jaufeerally *et al.*[23] studied the isomerisation and decomposition reactions of telluroformaldehyde analogues using MP2. Sakaki[24] carried out one MP2 study of the catalytic hydrogenation of 2,2,2-trifluoroacetophenone using a Ge(II)-hydride compound, in comparison with a Rh(I)-hydride complex, and predicted the full catalytic cycle by Ge(II)-hydride.

In this context, we have combined PT and QSRR ideas to obtain multi-output PT-QSRR models of complex molecular systems, such as nanoparticles, biosystems, and asymmetric catalytic reactions.[25] The main advantage of multi-output PT-QSRR models is to use a single equation to predict more than one output parameter [e.g. enantiomeric excess *ee* (%) or reaction yield Yld (%)] for complex systems. Besides, these models quantify multiple external/internal reaction factors with low computational cost. Both aspects have been illustrated in the model we have developed to predict the enantioselectivity of intramolecular carbolithiation reactions.[25a,b] Herein, we describe the first multi-output PT-QSRR model to correlate and predict the enantioselectivity and the yield of the above mentioned type of Heck-Heck cascade reactions, using both our previous experimental work[4] and the literature data[6,7] for related substrates. The model is expected to become a useful tool for the design of new chiral ligands for Heck-Heck cascade reaction, as well as for finding trends to improve the experimental results.

Methods
General workflow

The general workflow used to seek the QSRR model for this problem is shown in Figure 1. The first step was the compilation of a large dataset of Heck-Heck cascade reactions from the literature.[4,6] Then, the dataset was organized as a rectangular array of raw data, in which the columns are the different important variables for the reaction, while the rows are the different reactions. For the same reaction, a new row was added if at least one parameter changed (different experimental conditions, *e.g.* temperature, time, concentrations, *etc.* (see Electronic Supplementary Information).

Fig. 1 General workflow

The second step was the calculation of the molecular descriptors $V_k(m_q)$ used to quantify the chemical structure of all the molecules involved in the reaction (see next section). The third step was the calculation of the functions $<^p\varepsilon_{ij}>$ and γ_q and the products $\gamma_q \cdot V_k(m_q)$ for each class of molecule. This calculation is simple to carry out in one Excel calculation sheet using average and multiplication operations. The last step was the processing of the data obtained to find the QSRR model using statistical analysis software (see next sections).

Calculation of molecular descriptors

The values of the following structural variables were calculated with the software DRAGON:[26] $V_1(sub) = Hy(sub)$, $V_2(sub) = TPSANO(sub)$, $V_3(prod) = logP(prod)$, $V_4(base) = logP(base)$, $V_5(lig) = TPSAtot(lig)$, $V_6(solv) = D(solv)$. These final input variables were selected both with variable selection strategy (forward stepwise) and expert knowledge. The values of Hy, TPSANO, logP, and D for the other molecules were also calculated, but not all were selected after variable selection strategies. The structure of the molecules was uploaded to the software DRAGON as SMILE codes,[27] which were obtained using ChemOffice[28] software suite. In our case, the values of logP of the five molecular entities involved in our reaction (ligand, solvent, substrate, base and product) were calculated.

Theoretical details of the PT-QSRR models

Gonzalez-Díaz *et al.*[25a] have previously formulated a general-purpose Perturbation Theory (PT) method for multiple-boundary Chemoinformatic problems. In this work, we have adapted this method to predict the yield and enantioselectivity in Heck-Heck cascade reactions. The PT-QSRR model proposed here is the following additive equation:

$$^{p}\varepsilon_{ij}(\%)_{new} = e_0 + a_0 \cdot <^{p}\varepsilon_{ij}> + a_1 \cdot \gamma_1 \cdot V_1(sub)_{new} \quad (3)$$
$$+ a_2 \cdot \gamma_2 \cdot V_2(sub)_{new} + a_3 \cdot \gamma_3 \cdot V_3(prod)_{new}$$
$$+ a_4 \cdot \gamma_4 \cdot V_4(base)_{new} + a_5 \cdot \gamma_5 \cdot V_5(lig)_{new}$$
$$+ a_6 \cdot \gamma_6 \cdot V_6(solv)_{new}$$

The output variable $^{p}\varepsilon_{ij}(\%)_{new}$ measures two types of properties, yield Yld (%) and enantiomeric excess ee (%). This variable takes the values $^{1}\varepsilon_{ij}(\%)_{new}$ = Yld(%) when p = 1, and $^{2}\varepsilon_{ij}(\%)_{new}$ = *ee(%) when p = 2. The value *ee(%) = δ·ee(%) is the enantiomeric excess of product multiplied by $\delta_0 = 1$ when referred to the ee of (R) enantiomer or by $\delta_0 = -1$ when referred to the ee of (S) enantiomer. There are two types of input variables: $<^{p}\varepsilon_{ij}(\%)>$ and $\gamma_k \cdot V_k(m_q)$. The first type $(<^{p}\varepsilon_{ij}(\%)>)$ is the expected value of the yield or enantiomeric excess in the boundary reaction conditions. The second type $[\gamma_k \cdot V_k(m_q)]$ is related to the structural properties of each molecule involved in the reaction (V_{1-6}). These variables are directly multiplied by their corresponding factor (γ_{1-6}), which quantify non-structural parameters, such as temperature, concentration, *etc*. A Multivariate Linear Regression (MLR) analysis was performed with the software STATISTICA,[29] combining forward stepwise with standard procedures of variable selections. MLR allowed us to seek the model calculating the coefficients of variables and statistical parameters (see next sections), as well as observed, predicted, residuals, and deleted residuals of all reactions (see Electronic Supplementary Information).

Dataset of Heck-Heck cascade reaction

For this work, a large dataset of enantioselective palladium-catalyzed Heck-Heck cascade reactions was collected, including the polyene cyclisation reactions developed in our group[4] and related reactions reported previously in the literature by Keay and co-workers.[6] Figure 2 shows the structures of selected chiral ligands used.
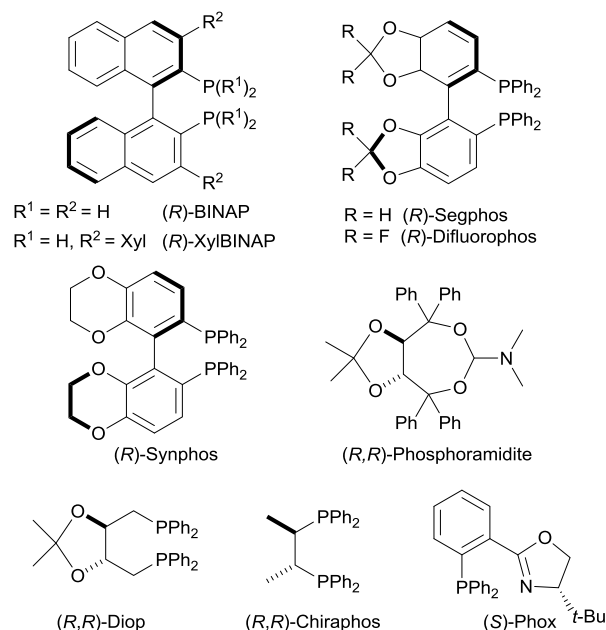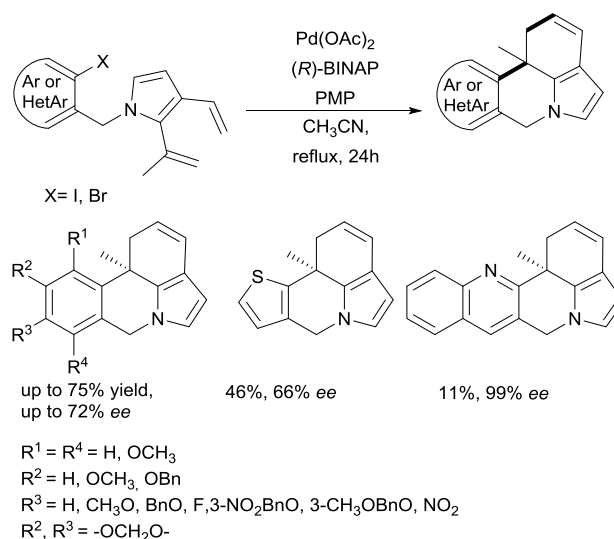

R[1] = R[2] = H   (R)-BINAP
R[1] = H, R[2] = Xyl   (R)-XylBINAP

R = H (R)-Segphos
R = F (R)-Difluorophos

(R)-Synphos

(R,R)-Phosphoramidite

(R,R)-Diop

(R,R)-Chiraphos

(S)-Phox   *t*-Bu

Fig. 2 Selected ligands for the Heck-Heck cascade reactions.

X= I, Br

up to 75% yield, up to 72% ee

46%, 66% ee

11%, 99% ee

R[1] = R[4] = H, OCH₃
R[2] = H, OCH₃, OBn
R[3] = H, CH₃O, BnO, F,3-NO₂BnO, 3-CH₃OBnO, NO₂
R[2], R[3] = -OCH₂O-

Scheme 2 Enantioselective Heck-Heck cascade reaction of 2,3-dialkenyl *N*-benzylpyrroles.

A summary of the results obtained by our group, and included in the data set, on the asymmetric Heck-Heck cascade reaction of halogenated 2,3-dialkenyl (hetero)arylmethyl-pyrroles is shown in Scheme 2.[4] The benchmark reaction was the polyene cyclisation of *N*-(2-iodo-4,5-dimethoxybenzyl)-2-(prop-1-en-2-yl)-3-vinyl-1*H*-pyrrole catalysed by Pd(OAc)₂ using (R)-BINAP as ligand. After the optimization of experimental conditions (base, additives, solvent, temperature), it was found that the choice of the PMP as base and CH₃CN as solvent was crucial to lead the sequence to completion, avoiding hydride transfer. After screening of the chiral ligands, (R)-BINAP was shown to be the most efficient phosphane ligand. The reaction was then extended to a series of substrates with

different substitution patterns on the aromatic ring, and also to heteroaromatic rings. This procedure allows a rapid and efficient access to a wide variety (16 examples) of enantiomerically enriched C-11b substituted Lycorane analogues.

On the other hand, Keay *et al*.[6] had previously applied this type of asymmetric cascade reactions to dialkenyl furans with a tethered naphthoyl triflate using $Pd_2(dba)_3$ as catalyst and (*S*) or (*R*)-BINAP as chiral ligands in the presence of PMP as base and toluene as solvent, obtaining the corresponding polycyclic ketones in moderate to good enantiomeric excesses (Scheme 3). The procedure was improved either by tuning the experimental conditions (base, solvent)[6c] or by changing the chiral ligands (Figure 2).[6d-f] Naphthoyl bromides can also be used instead of the corresponding triflates.[7] All these reactions have also been included in the dataset.

Thus, the new dataset combining our group´s work and literature data includes 27 different substrates, 2 different palladium sources, 23 ligands, and 12 solvents. Besides, entries for reactions experimentally carried out with a ligand of a given configuration have been duplicated for the enantiomeric ligand, assuming that the product of opposite configuration should be obtained in the same yield and the same enantiomeric excess, but opposite in sign. In electronic supplementary information (Tables S1 and S2), a detailed list of all reactions, with reactants, conditions, references, molecular descriptors, etc. is given.
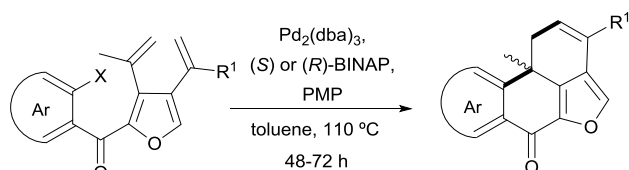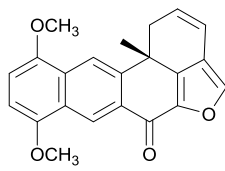

X = OTf, I
R^1 = CH_3, H
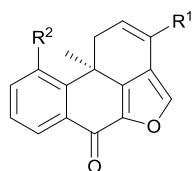
(*R*)-BINAP
$R^1 = R^2$ = H   83 %, 71 % *ee*
$R^1 = CH_3$, R^2 = H 78 %, 90 % *ee*
R^1 = H, R^2= CH_3, 71 %, 96 % *ee*
$R^1 = R^2 = CH_3$ 68 %, 71 % *ee*

(*S*)-BINAP
82 %, 62 % *ee*

Scheme 3 Enantioselective Heck-Heck cascade reactions developed by Keay´s group.

Results and Discussion
PT-QSRR theoretical model for Heck-Heck reactions

Once the dataset of enantioselective palladium-catalyzed Heck-Heck cascade reactions was collected, we focus on the creation of a general QSRR model to predict the enantioselectivity and yield of these and related reactions. As stated above, although DFT calculations had been used to explain the regioselectivity in the Heck-Heck cascade process

developed by Keay's group,[8] the enantioselectivity is a still a remaining challenge. To our knowledge, there are no general models to predict the effect of structural changes and experimental conditions (temperature, solvent, etc.) over enantioselectivity. Therefore, we sought to develop a general QSRR model for the prediction of the enantioselectivity and yield for these Heck-Heck cascade reactions under many different experimental conditions. The new QSRR model reported here allows both a computational screening of very large series of reactions with different substrates and a scanning of temperature, reaction time, solvent or dipole moment, as well as concentration of catalyst, base, and ligand. Therefore, this QSRR model could help to select the best catalysts, ligands and experimental conditions, without engaging in a long term empirical investigation.

The general picture considered as starting point for this theory is the following. Let be a set of molecules (m_q) that play different roles (q) in a Heck-Heck cascade reaction. The specific roles considered in this work are: q = 1 and q = 2 => substrate, q = 3 => product, q = 4 => base, q = 5 => ligand, q = 6 => solvent. Consider also that these molecules undergo one Heck-Heck cascade reaction in the reaction media under a set of boundary conditions or factors $f_n \equiv (f_1, f_2, f_3... f_n)$. These factors may quantify information about chemical symmetry, physical variables, and/or operational boundary conditions or factors that apply to all the reactions in the dataset. In fact, the problem involves two consecutive intramolecular Heck-Heck reactions, which take place in one pot; given that the structure of the substrate changes (transformed into a product), but the catalyst and chiral ligand are recovered. We constructed two sets of functions γ_q and $V_k(m_q)$ to quantify all the information related to the structure of all the molecules involved and the external factors. The external factors have been taken into consideration into the γ_k functions (Table 1).

Table 1 The formula and definition of f_q and γ_q

| k | Factors | | No structural factor | Structural Variables (V_k) | | |
	γ_q	f_j		V_k (m_q)	DRAGON symbol	Name of $^iV_k(q)$
		-	-	-	e_0	Independent term
				1	1	Expected value of yield(%) or *ee*(%)
1				$V_1(s)$	Hy (sub)	Hydrophobicity
2				$V_2(s)$	$TPSA_{NO}$ (sub)	Topological polar surface area
3				$V_3(p)$	logP (prod)	
4	$\gamma_4 = f_1 \cdot f_2$	$\gamma_4 = <^p\varepsilon_{ij}> \cdot B(\%)$	$<^p\varepsilon_{ij}>$ and amount of base	$V_4(b)$	logP (base)	
5	$\gamma_5 = f_1 \cdot f_3 \cdot f_4$	$\gamma_5 = =<^p\varepsilon_{ij}> \cdot L(\%) \cdot Pd(\%)$	$<^p\varepsilon_{ij}>$ and amount of Palladium and ligand	$V_5(l)$	$TPSA_{tot}$ (lig)	Topological polar surface area
6	$\gamma_6 = f_1 \cdot f_5 \cdot f_6$	$f_6 = <^p\varepsilon_{ij}> \cdot T \cdot t$	$<^p\varepsilon_{ij}>$ and temperature and time of the reaction	$V_6(v)$	μ(solv)	Dipolar moment

We included a total of six factors $f_1 = <{}^p\varepsilon_{ij}>$, $f_2 = B(\%)$, $f_3 = L(\%)$, $f_4 = Pd(\%)$, $f_5 = T(K)$, and $f_6 = t(h)$. These functions of type γ_q quantify information about all the factors f_n. The values γ_q have been calculated as a product $\gamma_q = f_1 \cdot f_2 ... f_{max}$ of all the factors (f_n) considered to affect the molecules of class q^{th}. In the case of the substrate and the product, we used $\gamma_1 = \gamma_2 = \gamma_3 = f_1 = <{}^p\varepsilon_{ij}>$. However, in the case of the base $\gamma_4 = f_1 \cdot f_2 = <{}^p\varepsilon_{ij}> \cdot B(\%)$, because it is important to take into account the amount of base in the reaction media. In the same way, we used a similar formula for the ligands $\gamma_5 = f_1 \cdot f_3 \cdot f_4 = <{}^p\varepsilon_{ij}> \cdot L(\%) \cdot Pd(\%)$. Finally, the time $t(h)$ and the temperature of the reaction $T(K)$ were included in the solvent term $\gamma_6 = f_1 \cdot f_5 \cdot f_6 = <{}^p\varepsilon_{ij}> \cdot T \cdot t$. The factor f_1 is one of the most important factors, because it quantifies the chirality of the ligand and the product. Thus, when $f_1 < 0$ the configuration of the ligand is (*S*) and when $f_1 > 0$ its configuration is (*R*) (see Table 1). It should be highlighted that f_1 multiplies all the variables of the designed equation. Consequently, when $f_1 = <{}^p\varepsilon_{ij}>$ changes its sign, the value of the predicted $^*ee(\%)$ also changes. In this way, it is possible to predict the configuration of the new product depending on the sign of $^2\varepsilon_{ij} = {}^*ee(\%)$. The formulae of f_q and γ_q for all classes of molecules appear in Table 1.

In our general hypothesis the functions γ_q and $V_k(m_q)$ for each class of molecule m_q can be combined to construct a QSRR model for Heck-Heck cascade reactions. Thus, different schemes (multiplicative, additive, perturbation like, non-linear) can be used in the construction of the QSRR functions. As it is not known what kind of model would perform better, different initial models (H_0 hypothesis) were selected and tested. Here we selected one additive-perturbation model that uses as starting point the expected value of efficiency of the reaction $<{}^p\varepsilon_{ij}>$ and adds successive perturbation terms or corrections. The equation for a simple linear multiplicative-additive PT-QSRR scheme is the following:

$$
{}^p\varepsilon_{ij}(\%)_{new} = e_0 + a_0 \cdot <{}^p\varepsilon_{ij}> + a_1 \cdot \gamma_1 \cdot V_1(sub)_{new} \qquad (4)
$$
$$
+ a_2 \cdot \gamma_2 \cdot V_2(sub)_{new} + a_3 \cdot \gamma_3 \cdot V_3(prod)_{new}
$$
$$
+ a_4 \cdot \gamma_4 \cdot V_4(base)_{new} + a_5 \cdot \gamma_5 \cdot V_5(lig)_{new}
$$
$$
+ a_6 \cdot \gamma_6 \cdot V_6(solv)_{new}
$$

$$
{}^p\varepsilon_{ij}(\%)_{new} = e_0 + a_0 \cdot <{}^p\varepsilon_{ij}> + a_1 \cdot f_1 \cdot V_1(sub)_{new} \qquad (5)
$$
$$
+ a_2 \cdot f_1 \cdot V_2(sub)_{new} + a_3 \cdot f_1 \cdot V_3(prod)_{new}
$$
$$
+ a_4 \cdot f_1 \cdot f_2 \cdot V_4(base)_{new} + a_5 \cdot f_1 \cdot f_3 \cdot f_4 \cdot V_5(lig)_{new}
$$
$$
+ a_6 \cdot f_1 \cdot f_5 \cdot f_6 \cdot V_6(solv)_{new}
$$

$$
{}^p\varepsilon_{ij}(\%)_{new} = e_0 + a_0 \cdot <{}^p\varepsilon_{ij}> + a_1 \cdot <{}^p\varepsilon_{ij}> \cdot V_1(sub)_{new} \qquad (6)
$$
$$
+ a_2 \cdot <{}^p\varepsilon_{ij}> \cdot V_2(sub)_{new} + a_3 \cdot <{}^p\varepsilon_{ij}> \cdot V_3(prod)_{new}
$$
$$
+ a_4 \cdot <{}^p\varepsilon_{ij}> \cdot base(\%) \cdot V_4(base)_{new}
$$
$$
+ a_5 \cdot <{}^p\varepsilon_{ij}> \cdot L(\%) \cdot Pd(\%) \cdot V_5(lig)_{new}
$$
$$
+ a_6 \cdot <{}^p\varepsilon_{ij}> \cdot T \cdot t \cdot V_6(solv)_{new}
$$

PT-QSRR model using a dataset of Heck-Heck cascade reactions

The best PT-QSRR model found with this algorithm is shown in Equation 7. Table 2 shows the values of training and error level.

$$
{}^p\varepsilon_{ij}(\%)_{new} = 0.020653 + 5.658742 \cdot <{}^p\varepsilon_{ij}> + 7.115341 \cdot \gamma_1 \cdot V_1(sub)_{new} \quad (7)
$$
$$
- 0.026377 \cdot \gamma_2 \cdot V_2(sub)_{new} + 0.427418 \cdot \gamma_3 \cdot V_3(prod)_{new}
$$
$$
+ 0.032456 \cdot \gamma_4 \cdot V_4(base)_{new} + 0.000020 \cdot \gamma_5 \cdot V_5(lig)_{new}
$$
$$
- 0.001770 \cdot \gamma_6 \cdot V_6(solv)_{new}
$$

$$
n = 520 \quad R = 0.89 \quad R^2 = 0.79 \quad SEE = 1.19
$$
$$
F = 289.83 \quad p < 0.005 \quad q^2 = 0.79
$$

Table 2 Values of training and error level

Variable	b	Std.Err.	t(750)	p-level
e_0	0.020653	1.192266	0.01732	0.986186
$<e>$	5.658742	1.519959	3.72296	0.000219
$\gamma_1 \cdot V_1(sub)$	7.115341	2.033165	3.49964	0.000507
$\gamma_2 \cdot V_2(sub)$	-0.026377	0.005896	-4.47372	0.000009
$\gamma_3 \cdot V_3(prod)$	0.427418	0.079480	5.37765	0.000000
$\gamma_4 \cdot V_4(base)$	0.032456	0.009861	3.29137	0.001066
$\gamma_5 \cdot V_5(lig)$	0.000020	0.000008	2.32867	0.020269
$\gamma_6 \cdot V_6(solv)$	-0.000017	0.000006	-2.86564	0.004335

This model includes seven variables and the statistical parameters are n, R, R^2, SEE, F, p, and q^2. The parameter n = 520 is the number of cases in the dataset (the reactions carried out in our group[4] combined with the literature reactions[6]). On the other hand, R is the multivariate linear correlation coefficient; which measures the association between two variables, while R^2 measures the proportion variability of the data explained by the model (in this case 79%). The Standard Error of Estimates (SEE) has a value of 1.19%. The p-level < 0.05 comes from the Fisher test and shows that correlation parameter (R) is significant in this case. The model was validated using Jack-knife cross-validation that is based on Leave-One-Out (LOO) procedure.[29] The q^2 value is the cross validation coefficient, which measures the quality of the validation in this procedure. Notably, $q^2 = R^2 = 0.79$, which may be indicative of a good predictive value of the model. The statistical parameters of each variable of the model are shown in Table 3.

The multiplying coefficients a_k are expected to measure the contribution of the structural (V_k) and non-structural factors (γ_k) to the reaction. For instance, the model has two variables to quantify the effect of the substrate $\gamma_1 \cdot V_1(sub)$ and $\gamma_2 \cdot V_2(sub)$. Nevertheless, the term $\gamma_1 \cdot V_1(sub)$ has a high positive contribution (coefficient $a_1 = 7.1$), but also a relatively high error 2.03. On the other hand, the term $\gamma_2 \cdot V_2(sub)$ has a lower negative contribution (coefficient $a_2 = -0.02$), but also a relatively lower error 0.006.

However, the existence of co-linearity among variables may hide the real contribution of a given effect.[27] In Table 3 we depict the correlation matrices for all the variables used before and after pre-multiplication by the γ_k. In fact, we should not forget that $f_1 = <{}^p\varepsilon_{ij}(\%)>$, expected yield or enantiomeric excess, multiplies all the variables of the equation. To explain this, it is important to see that both variables $\gamma_1 \cdot V_1(sub)$ and $\gamma_2 \cdot V_2(sub)$ (Table 3) are co-linear (r = -0.86). Then, we can conclude that both variables are correlated and may be hiding

the real contribution of the substrate. However, it was not possible to eliminate one of the variables due to the lack of linear correlation (in terms of R).

**Table 3** Correlation matrix for input terms.

Variable	$\langle^p\varepsilon_{ij}\rangle$ (%)	V_1 (sub)	V_2 (sub)	V_5 (lig)	V_4 (base)	V_3 (prod)	V_6 (solv)
$\langle^p\varepsilon_{ij}\rangle$(%)	1.00	0.08	0.08	0.02	0.01	-0.02	-0.08
V_1(sub)		1.00	0.97	0.38	0.18	-0.50	-0.76
V_2(sub)			1.00	0.33	0.18	-0.31	-0.76
V_5(lig)				1.00	0.06	-0.28	-0.37
V_4(base)					1.00	-0.05	-0.02
V_3(prod)						1.00	0.31
V_6(solv)							1.00
Variable	$\langle^p\varepsilon_{ij}\rangle$ (%)	V_1 (sub)	V_2 (sub)	V_5 (lig)	V_4 (base)	V_3 (prod)	V_6 (solv)
$\langle^p\varepsilon_{ij}\rangle$(%)	1.00	-0.99	0.92	0.84	0.90	0.99	0.57
V_1(sub)		1.00	-0.86	-0.87	-0.85	-0.99	-0.65
V_2(sub)			1.00	0.64	0.97	0.88	0.26
V_5(lig)				1.00	0.65	0.84	0.64
V_4(base)					1.00	0.87	0.25
V_3(prod)						1.00	0.62
V_6(solv)							1.00

Experimental study and simulation of a case of study

In this section, the use of our model through one practical case of study is illustrated. First, the case is presented, including the experimental results of one unpublished reaction. Next, the model is used to carry out a simulation of the general space of reaction. Last, the model is used to predict the specific susceptibility of the reaction to changes on the substrates and/or ligands of the catalyst.

We selected the reaction of the substrate depicted in Scheme 4 with Pd(dba)$_2$ in the presence of (R)-BINAP. The details of the experimental conditions, spectroscopic characterisation of the product, and the determination of the yield and enantiomeric excess (Chiral Stationary Phase HPLC) for this reaction are given in the Electronic Supplementary Information. This specific reaction, which had not been previously published, was included in the dataset due to its promising experimental results. The change of the catalyst from Pd(OAc)$_2$ to Pd(dba)$_2$, using ethanol as solvent instead of acetonitrile, has allowed to increase the *ee* to 70% with 66 % yield. These values are interesting but susceptible of improvement, being desirables values over 90%. In addition, the model makes a good prediction of these values for this reaction [Yld(%) = 53.9 and *ee*(%) = 79.1] taking into account the complexity of the problem and the error of the experimental method. For these reasons, this reaction is a good candidate for a computational simulation study.
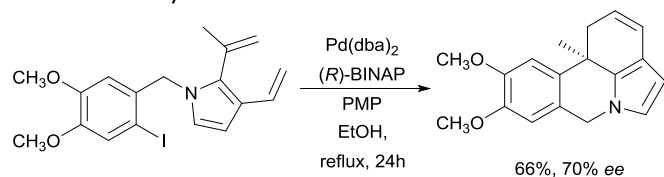
Scheme 4 Enantioselective Heck-Heck cascade reaction of 2,3-dialkenyl pyrrole using Pd(dba)$_2$ and (R)-BINAP.

Firstly, we focused on the computational simulation of the effect of the concentration of the catalyst, ligand, and base on reactivity and enantioselectivity of the Heck-Heck cascade reaction shown on Scheme 4, using ternary phase diagrams. A 2000-data-points simulation was carried out. These 2000 different data points are theoretical entries (reactions) with the same substrate, catalyst, ligand, base, solvent, and product. It means, that we substituted in the model above 2000 different entries with the same values of the molecular descriptors V_1(sub), V_2(sub), V_5(lig), V_4(base), V_3(prod), and V_6(solv). However, each entry has a different set of values of the concentrations of catalyst, ligand or base. Thus, these values were varied at random in the ranges Base (eq.) = 1-10 eq., Pd(dba)$_2$ = 1-10 mol%, and Ligand (mol%) = 1-30 mol%. In this way, above 2000 values of *ee*(%) and Yld(%) were predicted by the model. Then, these values were plotted in a ternary diagram to visualize the results more clearly. Ternary diagrams have been often used to study the composition of complex systems in physical chemistry, materials, and Nanosciences. As a result, the computational models may be difficult to obtain and often apply to specific systems.[30] In addition, there are not many models of general use to predict ternary diagrams with different components. Recently, Messina et al.[31] published a general model for self-aggregating systems. However, ternary diagrams are less common in organic synthesis. We decided to use ternary phase diagrams here because these reactions have three different components (catalyst, ligand, and base) with variable concentration. In this type of plot, the ratios of the three variables Base (eq.), Pd(dba)$_2$ (mol%), and Ligand (mol%) have to sum to a constant, usually 1.0 or 100%. Consequently, we transformed the original variables into adimensional proportions as follows:

$$\text{Base(\%)} = 100 \cdot \frac{\text{B(eq.)}}{\text{B(eq.)}_{max}} \cdot \frac{1}{K} \tag{8}$$

$$\text{Ligand (\%)} = 100 \cdot \frac{\text{L(mol\%)}}{\text{L(mol\%)}_{max}} \cdot \frac{1}{K} \tag{9}$$

$$\text{Pd(dba)}_2 \ (\%) = 100 \cdot \frac{\text{Pd(mol\%)}}{\text{Pd(mol\%)}_{max}} \cdot \frac{1}{K} \tag{10}$$

$$K = \frac{\text{B(eq.)}}{\text{B(eq.)}_{max}} + \frac{\text{Pd(mol\%)}}{\text{Pd(mol\%)}_{max}} + \frac{\text{L(mol\%)}}{\text{L(mol\%)}_{max}} \tag{11}$$

To our knowledge, this model is the first computational model able to predict the enantioselectivity of a Heck-Heck cascade reaction using a ternary diagram of phases. A visual inspection of the diagram shows two different phases in the ternary plot (Figure 3). The model predicts one enantioselectivity phase in the centre-to-top area of the diagram. These results show a similar tendency with essentially the same phases even when the magnitude in the increment of the yield is lower. According to the ternary diagram predicted by the model, the enantioselectivity phase encircles certain lower-upper limits. These limits are, approximately, the following: 2.5 − 7.5 mol% of the Pd(dba)$_2$ catalyst, 2.5 − 7.5 eq. of base and 7.5 − 20 mol% of ligand.
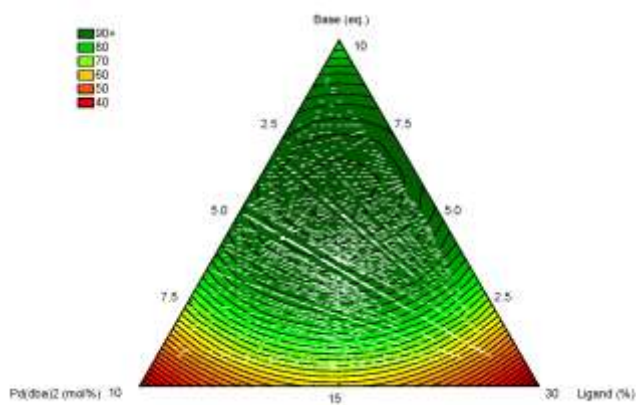
Fig. 3 Ternary diagram of phases for the enantioselectivity of the Heck-Heck cascade reaction studied here. White dashes represent data points predicted by the model.

Next, we focused on the computational study of the effect of the structure of the ligand and the substrate over the *ee*(%). Specifically, we were interested on the comparison of different ligands *vs.* different substrates when the halides used in our experimental work[4] were changed by the corresponding triflates. In the latter case, the values of *ee*(%) for a total of n = 19 ligands x 17 substrates = x reactions were calculated. In Figure 4, the results of this study are shown using an image with gradient colour, which is related to higher (green) or lower (red) *ee* (%), in order to obtain the best visual result. According to the model, the nature of the substrate (halide *vs.* triflates) has an important influence on the *ee* (%) achieved. Best results would be expected with halides under these experimental conditions. Interestingly, the screening revealed that the best *ee* (%) would be obtained employing chiral ligands as (*R*)-BINAP (L01 and related bidentate phosphane ligands with a binaphthyl unit and an oxygenated substituent in the *ortho* position to the phosphorus (i.e. L11, R = OPiv), as well as the bipyridine L17, for all the substrates (see ESI for the structures and numbering of all ligands and substrates). Besides, the predictions indicated that high enantioselectivities would be obtained for the reaction of 2,3-dialkenyl *N*-(o-iodobenzylpyrroles with almost all catalyst if there is a fluorine in *para* to the iodine atom (S14).
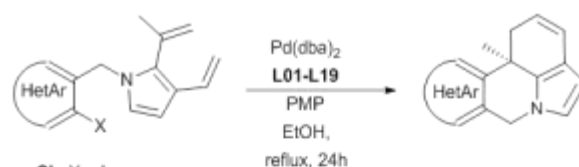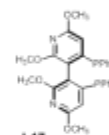
ee(%)	L01	L03	L08	L09	L10	L11	L12	L13	L14	L02	L04	L05	L06	L07	L15	L16	L17	L18	L19
S01Tf	61	43	50	50	50	63	50	50	43	57	50	43	53	46	57	57	67	50	50
S02Tf	59	47	54	54	54	67	54	54	47	61	54	47	57	50	61	61	71	54	54
S03Tf	56	37	45	45	45	58	45	45	37	52	45	37	48	40	52	52	62	45	45
S04Tf	61	42	49	49	49	63	49	49	42	57	49	42	53	45	57	57	67	49	49
S05Tf	91	27	34	34	34	48	34	34	27	42	34	27	38	30	42	42	52	34	34
S06Tf	61	43	50	50	50	63	50	50	41	57	50	43	53	46	57	57	67	50	50
S07Tf	61	43	50	50	50	63	50	50	43	57	50	43	53	46	57	57	67	50	50
S08Tf	58	40	47	47	47	60	47	47	40	54	47	40	50	43	54	54	64	47	47
S09Tf	91	25	32	32	32	46	32	32	25	40	32	25	36	28	40	40	49	32	32
S10Tf	100	14	21	21	21	35	21	21	14	29	21	14	25	17	29	29	39	21	21
S11Tf	51	0	0	0	0	6	0	0	0	6	0	0	0	0	0	0	8	0	0
S12Tf	86	21	28	28	28	42	28	28	21	35	28	21	31	24	35	35	45	28	28
S13Tf	41	21	28	28	28	41	28	28	21	35	28	21	31	24	35	35	45	28	28
S14Tf	87	62	69	69	69	81	69	69	62	77	69	62	73	65	77	77	86	69	69
S15Tf	40	42	50	50	50	63	50	50	42	57	50	42	53	46	57	57	67	50	50
S16Tf	67	30	37	37	37	50	37	37	30	44	37	30	40	33	44	44	54	37	37
S17Tf	71	71	78	78	78	81	78	78	71	85	78	71	81	74	85	85	85	78	78
ee(%)	L01	L03	L08	L09	L10	L11	L12	L13	L14	L02	L04	L05	L06	L07	L15	L16	L17	L18	L19
S01I	79	61	68	68	68	81	68	68	61	75	68	61	71	64	75	75	85	68	68
S02I	78	66	73	73	73	86	73	73	66	80	73	66	76	69	80	80	90	73	73
S03I	81	62	69	69	69	83	69	69	62	77	69	62	73	65	77	77	86	69	69
S04I	82	64	71	71	71	84	71	71	64	78	71	64	74	67	78	78	88	71	71
S05I	100	45	52	52	52	65	52	52	45	58	52	45	55	48	58	58	68	52	52
S06I	79	61	68	68	68	81	68	68	61	75	68	61	71	64	75	75	85	68	68
S07I	79	61	68	68	68	81	68	68	61	75	68	61	71	64	75	75	85	68	68
S08I	74	56	64	64	64	77	64	64	56	71	64	56	67	60	71	71	81	64	64
S09I	100	41	48	48	48	61	48	48	41	55	48	41	51	44	55	55	65	48	48
S10I	100	27	34	34	34	48	34	34	27	42	34	27	38	30	42	42	51	34	34
S11I	63	0	6	4	4	17	4	4	0	11	4	0	7	0	11	11	21	4	4
S12I	100	35	42	42	42	55	42	42	35	49	42	35	45	38	49	49	59	42	42
S13I	99	39	46	46	46	59	46	46	39	53	46	39	49	42	53	53	66	46	46
S14I	100	85	92	92	92	100	92	92	85	99	92	85	95	88	99	99	100	92	92
S15I	63	66	73	73	73	86	73	73	66	80	73	66	76	69	80	80	90	73	73
S16I	87	49	57	57	57	70	57	57	49	64	57	49	60	53	64	64	74	57	57
S17I	43	43	50	50	50	63	50	50	43	57	50	43	53	46	57	57	67	50	50

Fig. 4 Predicted *ee* (%) for the reactions of the 2,3-dialkenyl pyrroles (upper part: triflates, lower part: iodides) using Pd(dba)₂ in ethanol with different chiral ligands (see ESI for the structures and numbering of all ligands and substrates).

It should be pointed out that this QSRR model is very easy to apply to new reactions. The users only have to calculate the structural variables of their new substrates, ligands, products, etc. with the software DRAGON and substitute these values in the equation, setting the desired values of time of reaction, temperature etc. desired. The calculation is simple to make, even in one Excel sheet. In this sense, the general application of the model is very easy.

Conclusions

In this work we developed a new computational method for predicting the enantioselectivity and the yield of enantioselective Heck-Heck cascade reactions of adequately functionalized 2,3-dialkenylpyrroles and 3,4-dialkenylfurans. The model combines Perturbation Theory (PT) and Quantitative Structure-Reactivity Relationships (QSRR) ideas to predict two different outputs with the same equation [*ee* (%)

and yield (%)]. Besides, the use of ternary phase diagrams shows the computational simulation of the effect of the concentration of the catalyst, ligand, and base, on enantioselectivity of these reactions. The QSRR model also predicts trends in structural outcomes, such the effect of the use of halides *vs*. the triflates, or the ligand structure. Thus, best *ee* would be expected with bidentate phosphane ligands with a binaphthyl unit and an oxygenated substituent in the *ortho* position to the phosphorus atom. Therefore, the developed theoretical model is expected to be useful as a reference to choose the adequate catalyst or experimental conditions for this type of enantioselective polyene cyclisation reactions

Acknowledgements

References

1   a) For recent reviews, see: (a) M. Oestreich, *The Mizoroki-Heck Reaction*, Wiley, Chichester, 2009. (b) M. Larhed, *Science of Synthesis-Cross Coupling and Heck-Type Reactions,* **Vol. 3**, Thieme, Stuttgart, 2013.

2   For selected reviews on the asymmetric Heck reaction, see: (a) L. F. Tietze, F. Lotz in *Asymmetric Synthesis*, *2nd ed.*, ed. M. Christmann, S. Braese, Wiley-VCH, Weinheim, 2008, pp 155-160. (b) A. B. Dounay, L. E. Overman in *The Mizoroki-Heck reaction*, ed. M. Oestreich, Wiley, Chichester, 2009, pp. 533-568. (e) D. McCartney, P. J. Guiry, *Chem. Soc. Rev.* 2011, 40, 5122.

3   For selected reviews, see: (a) L. F. Tietze, L. M. Levy, in *The Mizoroki-Heck Reaction*, ed. M. Oestreich, Wiley, Chichester, 2009, pp. 281-344. (b) J. T. Link, C. K. Wada, in *The Mizoroki-Heck Reaction*, ed. M. Oestreich, Wiley, Chichester, 2009, pp. 433-462; (c) H. Clavier, H. Pellisier, *Adv. Synth. Catal.* 2012, **354**, 3347. (d) H. Ohno, *Asian J. Org. Chem.* 2013, **2**, 18.

4   E. Coya, N. Sotomayor, E. Lete, *Adv. Synt. Catal.* 2015, **357**, 3206-3214 (Highlighted in: *SYNFACTS*, 2016, **12**, 0067-0067).

5   (a) S. Lage, U. Martínez-Estibalez, N. Sotomayor, E. Lete, *Adv. Synth. Catal.* 2009, **351**, 2460. (Highlighted in: *SYNFACTS*, 2010, 0023-0023); (b) E. Coya, N. Sotomayor, E. Lete, *Adv. Synth. Catal.* 2014, **356**, 1853.

6   (a) S. P. Maddaford, N. G. Andersen, W. A. Cristofoli, B. A. Keay, *J. Am. Chem. Soc.* 1996, **118**, 10766. (b) S. Y. W. Lau, B. A. Keay, *Synlett* 1999, 605. (c) S. Y. W. Lau, N. G. Andersen, B. A. Keay, *Org. Lett.* 2001, **3**, 181. (d) E. Gorobets, G.-R. Sun, B. M. M. Wheatley, M. Parvez, B. A. Keay, *Tetrahedron Lett.* 2004, **45**, 3597. (e) D. Rankic, D. Lucciola, B. A. Keay, *Tetrahedron Lett.* 2010, **51**, 5724. (f) D. Lucciola, B. A. Keay, *Synlett* 2011, 1618.

7   F. Miyazaki, K. Uotsu, M. Shibasaki, *Tetrahedron* 1998, **54**, 13073.

8   D. Balcells, F. Maseras, B. A. Keay, T. Ziegler, *Organometallics* 2004, **23**, 2784.

9   For a review, see: M. García-Melchor, A. A. C. Braga, A. Lledós, G. Ujaque, F. Maseras, *Acc. Chem. Res.* 2013, **46**, 2626.

10 See, for example: a) S. T. Henriksen, P.-O. Norrby, P. Kaukoranta, P. G. Andersson, *J. Am. Chem. Soc.* 2008, **130**, 10414. b) W.-Q. Wu, Q. Peng, D.-X. Dong, X.-L. Hou, Y.-D. Wu, *J. Am. Chem. Soc.* 2008, **130**, 9717.

11 (a) H. Satoh, O. Sacher, T. Nakata, L. Chem, J. Gasteiger, K. Funatsu, *J. Chem. Inf. Comput. Sci.* 1998, **38**, 210. (b) S. Patel, J. Rabone, S. Russel, J. Tissen, W. Klaffke, *J. Chem. Inf. Comput. Sci.* 2001, **41**, 926.

12 For reviews, see: (a) K. C. Harper, E. N. Bess, M. S. Sigman, *Nature Chem.* 2012, **4**, 366. (b) K. C. Harper, M. S. Sigman, *J. Org. Chem*. 2013, **78**, 2813, and references therein. For recent examples of their work in this area: (c) E. N. Bess, A. J. Bischoff, M. S. Sigman, *Proc. Nat. Acad. Sci.* 2014, **111**, 14698. (d) A. Milo, A. J. Neel, F. D. Toste, M. S. Sigman, *Science*, 2015, **347**, 737. (e) C. Zhang, C. B. Santiago, L. Kou, M. S. Sigman, *J. Am. Chem. Soc.* 2015, **137**, 7290. (f) C. Zhang, C. B. Santiago, J. M. Crawford, M. S. Sigman, *J. Am. Chem. Soc.* 2015, **137**, 15668.

13 (a) M. Charton, *J. Am. Chem. Soc.* 1975, **97**, 1552. (b) M. Charton, *J. Am. Chem. Soc.* 1975, **97**, 3691. (c) M. Charton, *J. Am. Chem. Soc.* 1975, **97**, 3694. (d) M. Charton, *J. Org. Chem.* 1976, **41**, 2217.

14 A. Verloop, In *Drug Design*, Academic Press, New York, 1976

15 S. Aguado-Ullate, M. Urbano-Cuadrado, I. Villalba, E. Pires, J. I. García, C. Bo, J. J. Carbó, *Chem. Eur. J.* 2012, **18,** 14026.

16 H. Huang, H. Zong, G. Bian, H. Yue, L. Song, *J. Org. Chem.* 2014, **79**, 9455.

17 (a) K. Roy, J. T. Leonard, *J. Chem. Inf. Model.* 2005, **45**, 1352. (b) E. E. Anslyn, D. A. Sougherty, *Modern Physical Organic Chemistry*, University Science Books, Sausalito, 2006. (c) W. Greiner, D. Rischke, L. Neiser, H. Stöcker, *Thermodynamics and Statistical Mechanics*, Springer, New York, 2000.

18 (a) C. Hansch, A. R. Steward, J. Iwasa, *Mol. Pharmacol.* 1965, **1**, 87. (b) H. Kubinyi, *QSAR: Hansch analysis and related approaches*, VCH Publishers, Weinheim, New York, Basel, Cambridge, Tokyo, 1993. (c) M. J. Hilton, B. Cheng, B. R. Buckley, L. Xu, O. Wiest, M. S. Sigman, *Tetrahedron* 2015, **71**, 6513.

19 C. Corwin, J. E. Quinlan, G. L. Lawrence, *J. Org. Chem.* 1968, **33**, 347.

20 For selected reviews, see: (a) C. Hansch, A. Kurup, R. Garg, H. Gao, *Chem. Rev.* 2001, **101**, 619. (b) C. Hansch, D. Hoekman, A. Leo, D. Weininger, C. D. Selassie, *Chem. Rev.* 2002, **102**, 783. For selected examples, see: (c) C. Hansch, W. E. Steinmetz, A. J. Leo, S. B. Mekapati, A. Kurup, D. Hoekman, *J. Chem. Inf. Comput. Sci.* 2003, **43**, 120.

21 C. Møller, M. S. Plesset, *Phys. Rev.* 1934, **46**, 618.

22 J. Elm, S. Jorgensen, M. Bilde, K. V. Mikkelsen, *Phys. Chem. Chem. Phys.* 2013, **15**, 9636.

23 N. B. Jaufeerally, H. H. Abdallah, P. Ramasami, H. F. Schaefer, *J. Phys. Chem. A* 2013, **27**, 5567.

24 N. Takagi, S. Sakaki, *J. Am. Chem. Soc.* 2013, **135**, 8955.

25 (a) H. Gonzalez-Díaz, S. Arrasate, A. Gomez-SanJuan, N. Sotomayor, E. Lete, L. Besada-Porto, J. M. Ruso, *Curr. Top. Med. Chem.* 2013, **13**, 1713. (b) H. Gonzalez-Díaz, S. Arrasate, A. Gomez-SanJuan, N. Sotomayor, E. Lete, A. Speck-Planche, J. M. Ruso, F. Luan, M. N. D. S. Cordeiro, *Curr. Drug Metabol.* 2014, **15**, 470. (c) V. V. Kleandrova, F. Luan, H. González-Díaz, J. M. Ruso, A. Speck-Planche, M. N. Cordeiro, *Environ. Sci. Technol.*

2014, **48**, 14686. (d) F. Luan, V. V. Kleandrova, H. González-Díaz, J. M. Ruso, A. Melo, A. Speck-Planche, M. N. Cordeiro. *Nanoscale.* 2014, **6**, 10623. (e) F. J. Romero-Durán, N. Alonso, M. Yañez, O. Caamaño, X. García-Mera, H. González-Díaz *Neuropharmacology* 2015, **103**, 270.

26 (a) E. Papa, F. Villa, P. Gramatica, *J. Chem. Inf. Mod.* 2005, **45**, 1256. (b) A. Mauri, V. Consonni, M. Pavan, R. Todeschini, *Match. Commun. Math. Co.* 2006, **56**, 237. (c) M. Helguera, R. D. Combes, M. PérezGonzález, M. N. D. S. Cordeiro, *Curr. Top. Med. Chem.* 2008, **8**, 1628.

27 (a) M. A. Siani, D. Weininger, J. M. Blaney, *J. Chem. Inf. Comput. Sci.* 1994, **34**, 588. (b) M. A. Siani, D. Weininger, C. A. James, J. M. Blaney, *J. Chem. Inf. Comput. Sci.* 1995, **35**, 1026. (c) A. Karwath, L. De Raedt, *J. Chem. Inf. Model.* 2006, **46**, 2432. (d) A. A. Toropov, E. Benfenati, *Curr. Drug. Discov. Technol.* 2007, **4**, 77. (e) A. A. Toropov, A. P. Toropova, E. Benfenati, D. Leszczynska, J. Leszczynski, *J. Comput. Chem.* 2010, **31**, 381.

28 ChemDraw Professional 15.0, PerkinElmer, www.cambridgesoft.com, 2015.

29 T. Hill, P. Lewicki, *Statistics, Methods and Applications*, Stat Soft Inc, Tulsa OK, 2006.

30 K. A. See, M. Leskes, J. M. Griffin, S. Britto, P. D. Matthews, A. Emly, A. Van der Ven, D. S. Wright, A. J. Morris, C. P. Grey, R. Seshadri. *J. Am. Chem. Soc.* 2014, **136**, 16368.

31 P. V. Messina, J. M. Besada-Porto, H. González-Díaz, J. M. Ruso. *Langmuir* 2015, **31**, 12009.