

# Application of molecular topology for the prediction of the reaction times and yields under solvent-free conditions

Jorge Gálvez,\* María Gálvez-Llompart and Ramón García-Domenech

Received 10th December 2009, Accepted 25th March 2010

First published as an Advance Article on the web 6th May 2010

DOI: 10.1039/b926047a

Ball milling and conventional magnetic stirring can be used to support different laboratory techniques with a highly efficient mixing of reagents under solvent-free conditions. By using multilinear regression and linear discriminant analysis, topological-mathematical models have been built to predict the yield and the reaction time for organocatalytic reactions, Suzuki reactions and reactions of synthesis of heterocyclic compounds. The results from the *in silico* predictions confirm the usefulness of the approach followed.

## 1. Introduction

Development of solvent-free reactions can lead to new environmentally benign procedures that save resources and energy. These kinds of reactions promise to be an essential facet of green chemistry and are of high interest from both financial and environmental aspects.<sup>1,2</sup> These approaches can help to reduce the amounts of undesired hazardous chemicals (including solvents) used, increase the selectivity towards the given product(s) and also enhance the rate of many organic reactions.

In particular, ball-milling, microwave heating and ultrasonic irradiation are useful tools for promoting organocatalytic reactions under solvent-free conditions.<sup>3</sup> In many cases its application allows for a reduction of the amounts of reagents (in particular solvents), shorter reaction times and increased product yield, thereby resulting in a more efficient use of raw materials and energy.<sup>3</sup> All three techniques affect chemical processes requiring enhanced heat transfer and mass transport. Consequently, heterogeneous reactions can particularly benefit from the application of such techniques.

Several mathematical approaches have been proposed to facilitate the prediction of molecular properties. Equations linking quantitative structure–property (QSPR) relationships are particularly relevant and can be applied to large libraries of compounds for virtual computational screening.<sup>4,5</sup> However, these models require good structural descriptors that reliably represent the molecular features responsible for the property under study.

Molecular Topology (MT) is a way of describing molecular structures. It follows a two-dimensional approach, taking into account the internal atomic arrangement. The structure of each molecule is represented by specific subsets of topological indices (TIs).<sup>6</sup> These indices, when well chosen, provide a unique way of characterizing a molecular structure.<sup>7</sup> TIs are able to characterize the most important features of molecular structure: molecular size, binding and branching. The computation of TIs

is very swift and they also have the advantage of behaving as true structural invariants. This means that TIs are independent of the spatial position of the atoms in a particular moment, although extensions of the TIs giving account of three-dimensional structure have been also devised.<sup>8–10</sup>

MT has been demonstrated to be an excellent tool in the prediction of physical,<sup>11</sup> chemical and biological properties of structurally heterogeneous groups of compounds.<sup>12</sup> Today, it is known that, for most purposes, topostructural and topochemical information explains the most of the predicted properties, and that the inclusion of three-dimensional features results in slightly improved predictive models.<sup>13</sup>

This paper is one of the first attempts to use MT in the practical work of sustainable chemistry. Specifically, it is focused on developing and assessing QSPR models in order to predict the yield and reaction times of organic molecules when reacting under solvent-free and catalyst-free conditions.<sup>14,15</sup> Furthermore, some of these parameters are predicted for reactions under different experimental conditions such as ball-milling or stirring.<sup>3</sup>

## 2. Studied reactions

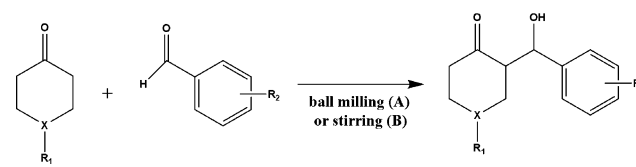
### 2.1 Organocatalytic reactions

In this work we have studied reactions using ball-milling (method A) and conventional magnetic stirring (method B). Ball milling and magnetic stirring are useful tools that allow highly efficient mixing of reagents under solvent-free conditions. In chemical synthesis, ball-milling or magnetic stirring modify the reaction conditions and enhance the reactivity of the reagents, with an enormous increase of the reagent surfaces, which lead to close contact between the starting materials on molecular scale.<sup>3</sup> Table 1 shows the yield obtained by Bolm and co-workers<sup>16</sup> for methods A and B and an aldol reactions under solvent-free conditions.

### 2.2 Synthesis of heterocyclic compounds

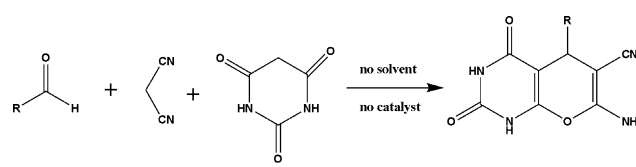
Mashkouri *et al.*<sup>14</sup> have recently presented a route to the synthesis of pyrano[2,3-*d*]pyrimidine-2,4(1*H*,3*H*)-diones with excellent yields (>99%) by simply ball-milling a stoichiometric mixture of an aldehyde, malononitrile, and barbituric acid without any

Molecular Connectivity & Drug Design Research Unit. Department of Physical Chemistry. Faculty of Pharmacy. University of Valencia, Avenida V.A. Estelles s/n, 46100-, Burjassot, (Valencia), Spain.  
E-mail: Jorge.Galvez@uv.es; Fax: (+34)-354-48-92;  
Tel: (+34)-354-48-91

**Table 1** Organocatalytic reactions: Effects of ball-milling, microwave and ultrasound irradiation<sup>a</sup>


Comp.	R <sub>1</sub>	R <sub>2</sub>	X	Method A		Method B		MATS4m
				Yield <sub>Exp.</sub> (%)	Yield <sub>Calc.</sub> (%)	Yield <sub>Exp.</sub> (%)	Yield <sub>Calc.</sub> (%)	
1	H	2-MeO	C	65	59	64	62	0.179
2	H	4-Cl	C	87	85	85	86	-0.04
3	H	2-NO <sub>2</sub>	C	97	89	89	93	-0.221
4	H	3-NO <sub>2</sub>	C	94	95	89	89	-0.204
5	H	4-NO <sub>2</sub>	C	99	100	95	85	-0.196
6	<i>t</i> -Bu	4-Cl	C	75	75	85	83	-0.01
7	<i>t</i> -Bu	2-NO <sub>2</sub>	C	66	72	76	79	-0.088
8	<i>t</i> -Bu	3-NO <sub>2</sub>	C	80	78	82	75	-0.077
9	<i>t</i> -Bu	4-NO <sub>2</sub>	C	85	82	58	71	-0.066
10	—	4-Cl	S	72	83	77	83	-0.01
11	—	2-NO <sub>2</sub>	S	59	66	73	62	0.06
12	—	3-NO <sub>2</sub>	S	75	71	59	57	0.084
13	—	4-NO <sub>2</sub>	S	79	77	47	54	0.087

<sup>a</sup> Selected regression equations: Yield A (%) = 224.4 – 82.03MATS4m – 75.94J; *N* = 13, *R* = 0.9043, *R*<sup>2</sup> = 0.8178, *Q*<sup>2</sup> = 0.6784, SEE = 5.82, *F* = 22.94, *p* = 0.0002, where MATS4m = (Moran autocorrelation at lag = 4) / atomic mass weighting, and *J* = Balaban distance index. Yield B (%) = 81.6 – 109.0MATS4m – 1.26 *T*<sub>N-O</sub>; *N* = 13, *R* = 0.8805 *R*<sup>2</sup> = 0.7753, *Q*<sup>2</sup> = 0.6541, SEE = 7.43, *F* = 17.59, *p* = 0.0005, MATS4m = (Moran autocorrelation at lag = 4) / atomic mass weighting, and *T*<sub>N-O</sub> = sum of topological N–O distances.

**Table 2** Synthesis of pyrano[2,3-*d*]pyrimidine-2,4(1*H*,3*H*)-diones under solvent-free and catalyst-free conditions<sup>a</sup>


Comp.	R	Time <sub>Exp.</sub> (min)	Yield <sub>Exp.</sub> (%)	Time <sub>Calc.</sub> (min)	Residual (min)	Time <sub>CV-calc</sub> <sup>b</sup> (min)	Residual <sub>CV</sub> (min)	<i>E</i> <sub>Eig05R</sub>
1	C <sub>6</sub> H <sub>5</sub>	70	>99	83	-13	93	-23	2.90
2	2-ClC <sub>6</sub> H <sub>4</sub>	90	>99	79	11	73	17	2.91
3	4-ClC <sub>6</sub> H <sub>4</sub>	55	>99	52	3	51	4	2.97
4	2-NO <sub>2</sub> C <sub>6</sub> H <sub>4</sub>	60	>99	55	5	54	6	2.96
5	3-NO <sub>2</sub> C <sub>6</sub> H <sub>4</sub>	15	>99	23	-8	27	-12	3.04
6	4-NO <sub>2</sub> C <sub>6</sub> H <sub>4</sub>	25	>99	19	6	15	10	3.05
7	4-CH <sub>3</sub> OC <sub>6</sub> H <sub>4</sub>	30	>99	34	-4	36	-6	3.01

<sup>a</sup> Reaction time (min) = 1357.2 – 439.3 *E*<sub>Eig05R</sub>; *N* = 7, *R* = 0.9459, *R*<sup>2</sup> = 0.8948, *Q*<sup>2</sup> = 0.7477, SEE = 9.59, *F* = 42.51, *p* < 0.0013, and *E*<sub>Eig05R</sub> = Eigenvalue 05 from edge adjacent matrix weighted by resonance integrals. <sup>b</sup> CV = Cross-validation.

catalyst or solvent. Table 2 shows the reaction scheme and the reaction times obtained in each synthesis.

### 2.3 Suzuki reaction

Nielsen and coworkers<sup>15</sup> have investigated the coupling Suzuki reaction of diverse haloarenes with phenylboronic acid under ball-milling and solvent-free conditions. Table 4 shows the reaction scheme and the yields obtained in each coupling reaction.

## 3. Molecular descriptors

A set of well-known topological descriptors was used in this work: subgraph Randić–Kier–Hall-like indices up to the fourth order (<sup>*m*</sup>χ<sub>*t*</sub>, <sup>*m*</sup>χ<sub>*t*</sub><sup>*v*</sup>),<sup>17</sup> topological charge indices; TCI, up to the

fifth order, (*J*<sub>*m*</sub>, *G*<sub>*m*</sub>, *J*<sub>*m*</sub><sup>*v*</sup>, *G*<sub>*m*</sub><sup>*v*</sup>),<sup>18</sup> quotients and differences between valence and non-valence connectivity (<sup>*m*</sup>*C*<sub>*t*</sub> = <sup>*m*</sup>χ<sub>*t*</sub>/<sup>*m*</sup>χ<sub>*t*</sub><sup>*v*</sup> and <sup>*m*</sup>*D*<sub>*t*</sub> = <sup>*m*</sup>χ<sub>*t*</sub> – <sup>*m*</sup>χ<sub>*t*</sub><sup>*v*</sup>), Balaban distance index (*J*),<sup>19</sup> Moran autocorrelations.<sup>20</sup> All descriptors used in this work were obtained with the aid of the in-house Desmol11 program (available by e-mail request) and the software DRAGON.<sup>21</sup>

## 4. QSAR algorithms

### 4.1 Multilinear regression analysis

The regression equations were obtained by correlating the experimental yields or times values with the aforementioned TIs, by multilinear regression analysis, MLRA, through the BMDP

**Table 3** Classification functions and classification matrix obtained by linear discriminant analysis with yield (%) and the test group for the Suzuki reaction: Aryl halogenides with phenylboronic acid

Classification functions <sup>a,b</sup>			
Variable	DF <sub>H</sub>	DF <sub>M</sub>	DF <sub>L</sub>
X3P	214.8	181.0	161.1
G2	-38.7	-31.5	-27.8
J3	1306.8	1091.1	953.4
D1	-164.2	-138.2	-121.9
PR1	-49.9	-41.6	-37.0
Constant	-198.7	-145.6	-115.6

Classification matrix

Group	Correct (%)	Number of cases classified in group		
		H	M	L
H	87.5	7	1	0
M	75	1	6	1
L	92.3	0	1	12
Total	86.2	8	8	13

<sup>a</sup>  $N = 29$ ,  $\lambda$  (Wilks' lambda) = 0.163,  $F$ -stat = 6.50. <sup>b</sup> DF = Discriminant Function.

software.<sup>22</sup> The Furnival–Wilson algorithm<sup>23</sup> was followed to find subsets of descriptors, and the minimum value of the Mallows'  $C_p$  parameter was the variable selecting criterion.<sup>24</sup> The program searched subsets with 1, 2, 3... independent variables and selected the equation exhibiting the smallest Mallows'  $C_p$  parameter.

Just to validate the selected prediction functions, a cross-validation was carried out. For the cross-validation test, the leave-one-out algorithm was used, in which one case is eliminated from the data set and then the regression analysis with the  $N - 1$  remaining cases and the original descriptors (the ones selected in the first regression) is performed again. The corresponding property value for the case taken out is then predicted. The procedure is repeated as many times as there are cases in the data.<sup>25</sup> The value of prediction coefficient,  $Q^2$ , will indicate the quality of the prediction function selected.

#### 4.2 Linear discriminant analysis

The objective of the linear discriminant analysis (LDA), which is considered as a heuristic algorithm able to distinguish between two or more categories or objects, is to find a linear function to correctly classify each object.<sup>26</sup> In this work, the LDA for three groups was performed with the yield percent. Thresholds for each category were: High (yields >40%), Medium (yields 10–40%) and Low (yields <10%).

The discriminant capability was tested by the percentage of correct classifications into each group. LDA was carried out using the BMDP 7M package.<sup>22</sup> The selection of the descriptors was based on the  $F$ -Snedecor parameter, and the classification criteria was the shortest Mahalanobis distance (distance of each case to the mean of all cases used in the regression equation). 7M chooses the variables used in computing the linear classification functions in a stepwise manner. At each step the variable that

adds the most to the separation of the groups is added to (or the variable that adds the least is removed from) the discriminant function. The quality of the discriminant function is evaluated by the Wilks' lambda parameter,  $\lambda$ , which is a multivariate analysis of the variance statistic that tests the equality of group means for the variable(s) in the discriminant function.

## 5. Results and discussion

The first goal of this work was to find the best QSAR models for predicting the yields of processes under ball-milling (method A) or stirring (method B) under organocatalytic reactions. The best linear regression equations obtained, including its statistical parameters, were:

$$\text{Yield A (\%)} = 224.4 - 82.0\text{MATS4m} - 75.9J \quad (1)$$

with

$$N = 13, r = 0.9043, r^2 = 0.8178, Q^2 = 0.6784, \text{SEE} = 5.82, \\ F = 22.9, p = 0.0002$$

and

$$\text{Yield B (\%)} = 81.6 - 109.0\text{MATS4m} - 1.26T_{N-O} \quad (2)$$

with

$$N = 13, r = 0.8805, r^2 = 0.7753, Q^2 = 0.6541, \text{SEE} = 7.43, \\ F = 17.6, p = 0.0005.$$

The values of  $r^2$ , the prediction coefficients in the cross-validation type leave-one-out algorithm,  $Q^2$ , as well as the low SEE (less than 8% of the average values of the property) account for the quality of the models selected.

Table 1 and Fig. 1 show the yields predicted for each reaction studied.

In both equations the Moran index, MATS4m, appears. This descriptor is defined as:

$$\text{MATS } p_k l = \frac{\sum_j \delta_{ij} (p_{ki} - p_k)(p_{kj} - p_k)}{\sum_i (p_{ki} - p_k)} \quad (3)$$

where  $p_{ki}$  and  $p_{kj}$  are the values of property  $k$  of atoms  $i$  and  $j$ , respectively;  $p_k$  is the average value of property  $k$ ;  $l$  is the number of non-zero elements in the sum and  $\delta(l, d_{ij})$  is a Dirac delta function defined as:

$$\delta(l, d_{ij}) = 1 \text{ if } d_{ij} = l \text{ or } \delta(l, d_{ij}) = 0 \text{ if } d_{ij} \neq l$$

where  $d_{ij}$  is the topological distance or spatial lag between atoms  $i$  and  $j$ .

Spatial autocorrelation is a measure of the degree of interdependence between properties, and the nature and strength of that interdependence. It is classified either as positive or negative, depending if all similar values appear gathered or not. In chemical compounds, Moran's spatial autocorrelation evaluates whether the value of an atomic property at a given atom in the molecule is or is not independent of the values of such property at the neighbouring atoms. If that dependence occurs, then the property is considered to show spatial autocorrelation.

Moreau and Broto<sup>27</sup> were the first to apply the autocorrelation indices to the topology of molecular structures. The

**Table 4** Results of classification obtained by linear discriminant analysis with yield (%) and the training group in the Suzuki reaction: Aryl halogenides with phenylboronic acid

		Pd catalyst				
		R <sub>1</sub> -BY <sub>2</sub> + R <sub>2</sub> -X		→ R <sub>1</sub> -R <sub>2</sub>		
		Base				
Comp.	Substrate	Yield (%)	Probability			Class
			H	M	L	
<i>H training group (yields &gt;40%)</i>						
1	2-Bromonaphthalene	96	1.000	0.000	0.000	H
2	3-Bromoanisole	93	0.997	0.003	0.000	H
3	5-Bromo-1,3-benzodioxole	91	0.695	0.304	0.002	H
4	3-Bromoaniline	89	0.134	0.816	0.050	M
5	3-Chlorobromo-benzene	84	0.998	0.002	0.000	H
6	4-Bromo-benzotrifluoride	82	0.999	0.001	0.000	H
7	4-Bromo- <i>N,N</i> -dimethylaniline	50	0.998	0.002	0.000	H
8	3-Bromoquinoline	46	0.998	0.002	0.000	H
<i>M training group (yields &lt;40% and &gt;10%)</i>						
9	4-Bromoanisole	37	0.749	0.248	0.003	H
10	2-Bromo-6-nitrotoluene	27	0.001	0.927	0.071	M
11	3-Bromobenzaldehyde	24	0.034	0.842	0.123	M
12	4-Bromoisoquinoline	20	0.041	0.901	0.058	M
13	3-Bromobenzonitrile	12	0.048	0.905	0.047	M
14	4'-Bromoacetophenone	11	0.000	0.119	0.881	L
15	3'-Bromoacetophenone	10	0.001	0.696	0.304	M
16	3-Bromo-4-nitroanisole	10	0.001	0.769	0.231	M
<i>L training group (yields &lt;10%)</i>						
17	2-Bromobenzonitrile	5	0.000	0.036	0.964	L
18	2-Bromobenzaldehyde	4	0.000	0.052	0.948	L
19	2-Bromo-3-nitrotoluene	4	0.000	0.785	0.215	M
20	4-Bromobenzaldehyde	2	0.000	0.257	0.743	L
21	5-Bromopyrimidine	2	0.000	0.004	0.996	L
22	5-Bromoindole	1	0.000	0.309	0.691	L
23	3-Bromopyridine	1	0.000	0.043	0.957	L
24	4-Bromobenzonitrile	0	0.000	0.486	0.514	L
25	2-Bromonitrobenzene	0	0.000	0.090	0.910	L
26	4-Iodonitrobenzene	0	0.000	0.290	0.710	L
27	4'-Bromoacetanilide	0	0.000	0.329	0.671	L
28	4-Bromophenol	0	0.000	0.216	0.784	L
29	4-Bromopyridine	0	0.000	0.043	0.957	L

autocorrelation vectors represent the level or degree of similarity between/among molecules. In our case, the most representative index in reaction yields prediction was the MATS4m, *i.e.* the Moran autocorrelation index at lag (topological distance) = 4.

A pretty reasonable interpretation can be made on the basis of elemental organic chemistry. Thus, the presence of a donor group (such as MeO) on the 2-position of the aldehyde's benzene ring clearly stabilizes it, so hindering the tautomeric transfer of H in the product. In contrast, the presence on the benzene ring of electron-accepting or -withdrawing groups, such as NO<sub>2</sub> or Cl, promote the reaction. Furthermore, the potency and position of the electron-withdrawing effect is also critical. Thus, the nitro group is a better reaction promoter than Cl, which is not surprising considering that NO<sub>2</sub> is an electron acceptor by both inductive and resonance effects, whereas Cl is only inductive. Regarding the position, it is very clear that the further the electron acceptor from the aldehyde (carbonyl in the product) the better. Compare for example the larger values for compounds with Cl or NO<sub>2</sub> on the 4-position with the smaller values when they are on the 2- and 3-positions. Moreover, as was to be expected, these features are common to both A and B methods.

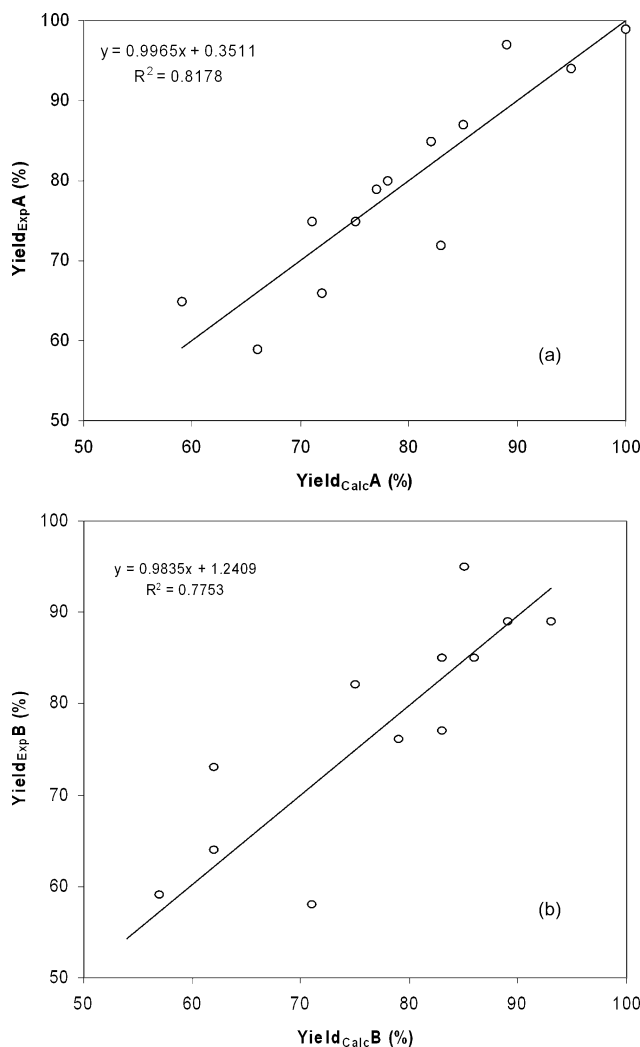
Interestingly, the MATS4m index seems to encode information related to the chemical behaviour. Thus, compound **1**, which contains a methoxy group, shows the largest positive value of MATS4m (0.179), whereas compound **5** (with NO<sub>2</sub> at the 4-position) shows one of the highest negative values (-0.196). The MAT indices range between -1, indicating perfect dispersion in the spatial similarity of the atoms, up to +1, indicating perfect correlation, though in general, the MAT values are close to 0 (random correlation). Also, as deduced from eqn (1) and eqn (2), the larger the MAT values the lower the yield. The substitution of carbon with sulfur on position X clearly balance the atomic weights in the molecule, thereby making the MAT values slightly positive (compounds **11**, **12** and **13**).

Regarding the reaction time for the pyrano-pyrimidone synthesis (Table 2), the best linear regression equation obtained was:

$$\text{Reaction time (min)} = 1357.2 - 439.3 E_{\text{Eig05R}} \quad (4)$$

with

$$N = 7, r = 0.9459, r^2 = 0.8948, Q^2 = 0.7477, \text{SEE} = 9.59, \\ F = 42.51, p < 0.0013$$

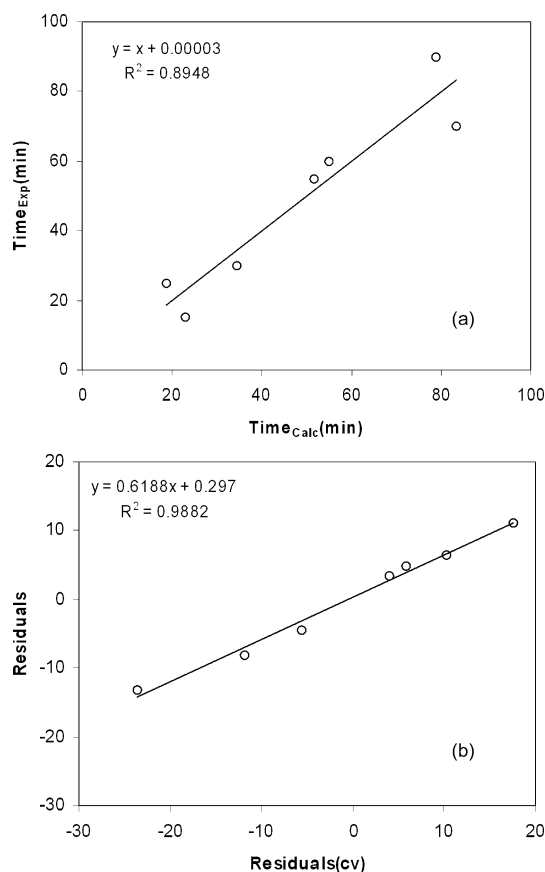


**Fig. 1** Prediction of yield (%) for organocatalytic reactions: effects of (A) ball-milling and (B) stirring.

Table 2 and Fig. 2 show the results of the prediction of reaction times for each reaction studied. The concordance between  $\text{Time}_{\text{exp.}}$  and  $\text{Time}_{\text{calc.}}$  indicate the good quality of the selected model.

In this case is noteworthy that, in contrast to the reaction yield, the electronic properties of the moieties involved are not relevant. So, compound **6**, containing the nitro motif at the 4-position of the benzene substituent, and compound **7**, which has a methoxy group, exhibit very similar reaction times. What does play a key role is the position of the substituents on the benzene ring (compare compounds **2** and **3**, as well as **4** and **5**). In the second case, the reaction time falls from 60 to 15 min just by moving the chlorine or  $\text{NO}_2$  from the 2-position to the 3-position.

These features are consistent with the regression equation obtained, eqn (4), which contains the index  $E_{\text{Eig}05\text{R}}$ , which is a measure of the eigenvalues of the edge adjacency matrix weighted by the resonant integrals. Given the conjugated nature of the analysed compounds, it makes sense that an index related to the  $\pi$  electron energies (delocalized molecular  $\pi$ -orbital energies) should play a dramatic role. In fact, compounds with



**Fig. 2** Prediction of reaction times in the synthesis of pyrano[2,3-*d*]pyrimidine-2,4(1*H*,3*H*)-diones under solvent-free and catalyst-free conditions. (a) Graphical representation of experimental time against calculated time from eqn (3). (b) Graphical representation of the residuals obtained in the training series against that obtained in the cross-validation.

$E_{\text{Eig}05\text{R}}$  values above 3.00 show reaction times below 30 min. Moreover, since it is the edge adjacency matrix – and not the vertex adjacency matrix – that is involved, the conjugation effects (interactions between a bond and its adjacent bond) are better depicted.

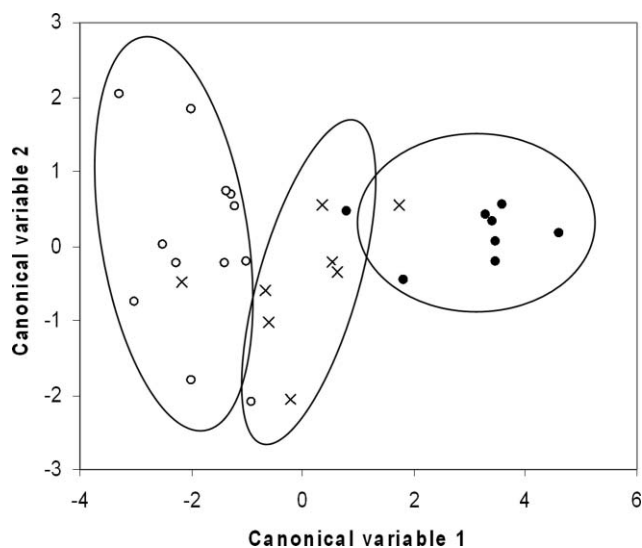
Finally, LDA performed very well in the Suzuki reaction. In order to get the discriminant function, we applied LDA to a training set comprised of 29 cases included in the three subsets mentioned above. As shown in Table 3, a 5-variable equation led to an average correct classification of 86.2% in the prediction of reaction yield.

Table 4 shows the classification obtained for each case in order to get the probability parameter, and Fig. 3 illustrates the spatial region of stability for each group using the canonical variables obtained in the LDA.

Typical topostructural indices, such as  ${}^3\chi_p$  and PR1 (pairs of ramification at lag one), as well as topological charge indices, such as  $G_2$ ,  $J_3$  and  ${}^1D$ , play an important role. This can be interpreted in the sense that both structural and electronic properties determine the reaction yield.

Though the interpretation of these results is difficult, some thought reveals that the presence of electron acceptor-donor couples results in larger values of yield. Compare for example





**Fig. 3** Plot of canonical variable 2 versus canonical variable 1 for LDA with three groups (● = H group (yields >40%); ○ = L group (yields <10%); × = M group (yields <40% and >10%).

the high yield for 3-bromoanisole and 3-bromoaniline with the – much lower – value for 4-bromo-*N,N*-dimethylaniline, in which the tertiary amine hinders the basicity of the amine. In contrast, the presence of two acceptors virtually stops the progress of the reaction (see compounds **17** to **29**).

## 6. Conclusions

The results outlined here demonstrate that molecular topology can be successfully employed to predict reaction parameters playing key role in green chemistry. The topological indices involved in the regression equations, particularly the Moran autocorrelation indices, provide a good framework to interpret the results in terms of structure–reactivity relationships. As far as we are aware, the results described herein represent one of the first attempts to apply molecular topology to the framework of green chemistry. The outcome of this work demonstrates that *in silico* approaches based on molecular topology can be a very useful tool for experimentalists. Thereby, we should be able to predict greener reactions prior to experimental tests, so saving time and expense.

## Acknowledgements

We thank the Ministerio de Educación e Innovación, Spain (project SAF2009-13059-C03-02), for support of this work.

## References

- 1 D. C. Dittmer, *Org. Synth. Chem. Ind.*, 1997, **19**, 779–784.
- 2 F. Toda, *Acc. Chem. Res.*, 1995, **28**, 480–486.
- 3 A. Bruckmann, A. Krebs and C. Bolm, *Green Chem.*, 2008, **10**, 1131–1141.
- 4 J. V. de Julian-Ortiz, J. Galvez, C. Munoz-Collado, R. Garcia-Domenech and C. Gimeno-Cardona, *J. Med. Chem.*, 1999, **42**, 3308–3314.
- 5 R. Lahana, *Sci. Am., Ed. Française*, 1997, **241**, 56–58.
- 6 A. T. Balaban, *Pure Appl. Chem.*, 1983, **55**, 199–206.
- 7 L. B. Kier and L. H. Hall, *Molecular connectivity in structure–activity analysis*, John Wiley & Sons, 1986.
- 8 J. V. de Julian-Ortiz, C. de Gregorio Alapont, I. Rios-Santamarina, R. Garcia-Domenech and J. Galvez, *J. Mol. Graphics Model.*, 1998, **16**, 14–18.
- 9 J. A. Castillo-Garit, Y. Marrero-Ponce, F. Torrens, R. Garcia-Domenech and V. Romero-Zaldivar, *J. Comput. Chem.*, 2008, **29**, 2500–2512.
- 10 A. Golbraikh, D. Bonchev and A. Tropsha, *J. Chem. Inf. Comput. Sci.*, 2002, **42**, 769–787.
- 11 Z. Tomovic and I. Gutman, *J. Chem. Inf. Comput. Sci.*, 2001, **41**, 1041–1045.
- 12 R. Garcia-Domenech, J. Galvez, J. V. de Julian-Ortiz and L. Pogliani, *Chem. Rev.*, 2008, **108**, 1127–1169.
- 13 S. C. Basak, D. R. Mills, A. T. Balaban and B. D. Gute, *J. Chem. Inf. Comput. Sci.*, 2001, **41**, 671–678.
- 14 S. Mashkouri and M. R. Naimi-Jamal, *Molecules*, 2009, **14**, 474–479.
- 15 S. F. Nielsen, D. Peters and O. Axelsson, *Synth. Commun.*, 2000, **30**, 3501–3509.
- 16 B. Rodriguez, A. Bruckmann and C. Bolm, *Chem.–Eur. J.*, 2007, **13**, 4710–4722.
- 17 L. B. Kier and L. H. Hall, *J. Pharm. Sci.*, 1983, **72**, 1170–1173.
- 18 J. Galvez, R. Garcia, M. Salabert and R. Soler, *J. Chem. Inf. Comput. Sci.*, 1994, **34**, 520–525.
- 19 A. T. Balaban, *Chem. Phys. Lett.*, 1982, **89**, 399–404.
- 20 P. Moran, *Biometrika*, 1950, **37**, 17–23.
- 21 *DRAGON 5.4*, Talete S. R. L., University of Milano-Bicocca, Milan, Italy, 2006.
- 22 W. J. Dixon, M. B. Brown, L. Engelman and R. I. Jennrich, *7M Package*, University of California Press, San Francisco, 1990, vol. 1.
- 23 G. M. Furnival, *Technometrics*, 1971, **13**, 403–408.
- 24 R. Hocking, *Technometrics*, 1972, **14**, 967–970.
- 25 E. Besalú, *J. Math. Chem.*, 2001, **29**, 191–204.
- 26 S. Wold and L. Eriksson, *Statistical validation of QSAR results in Chemometric methods in molecular design* ed. H. Van de Waterbeemd, VCH, New York, 1995.
- 27 G. Moreau and P. Broto, *Nouv. J. Chim.*, 1980, **4**, 359–360.