

RESEARCH ARTICLE

Open Access

# Quantitative structure-activation barrier relationship modeling for Diels-Alder ligations utilizing quantum chemical structural descriptors

Sisir Nandi<sup>1</sup>, Alessandro Monesi<sup>1</sup>, Viktor Drgan<sup>1</sup>, Franci Merzel<sup>2</sup> and Marjana Novič<sup>1\*</sup>

## Abstract

**Background:** In the present study, we show the correlation of quantum chemical structural descriptors with the activation barriers of the Diels-Alder ligations. A set of 72 non-catalysed Diels-Alder reactions were subjected to quantitative structure-activation barrier relationship (QSABR) under the framework of theoretical quantum chemical descriptors calculated solely from the structures of diene and dienophile reactants. Experimental activation barrier data were obtained from literature. Descriptors were computed using Hartree-Fock theory using 6-31G(d) basis set as implemented in Gaussian 09 software.

**Results:** Variable selection and model development were carried out by stepwise multiple linear regression methodology. Predictive performance of the quantitative structure-activation barrier relationship (QSABR) model was assessed by training and test set concept and by calculating leave-one-out cross-validated  $Q^2$  and predictive  $R^2$  values. The QSABR model can explain and predict 86.5% and 80% of the variances, respectively, in the activation energy barrier training data. Alternatively, a neural network model based on back propagation of errors was developed to assess the nonlinearity of the sought correlations between theoretical descriptors and experimental reaction barriers.

**Conclusions:** A reasonable predictability for the activation barrier of the test set reactions was obtained, which enabled an exploration and interpretation of the significant variables responsible for Diels-Alder interaction between dienes and dienophiles. Thus, studies in the direction of QSABR modelling that provide efficient and fast prediction of activation barriers of the Diels-Alder reactions turn out to be a meaningful alternative to transition state theory based computation.

**Keywords:** Quantitative structure-activation barrier relationships, Quantum chemical structural descriptors, HOMO LUMO Energy, Diels-Alder ligations, Multivariate linear regression and artificial neural network modelling

## Background

Activation energy is the minimal energy required to start a chemical reaction. It corresponds to the potential barrier separating the minima of potential energy of reactants and products. The free energy of activation is reversible work required to bring the system from the minimum of reactive well to the top of the barrier. For basically all the condensed phase reactions the transition state is valid and in order to calculate the rate constant it is enough to calculate the activation free energy. Still locating the transition state is not easy and therefore one has to proceed with

computationally less demanding methods. The free energy difference between initial geometry-optimized reactants and the transition state at maximal potential energy barrier is correlated with the rate constant or reactivity of the reactant molecules. Structure-reactivity relationship is a variant of structure-property-activity relationship studies [1,2]. The concept of quantitative structure-property-activity relationship (QSAR/ QSPR) was originated from the idea of Crum-Brown and Fraser as early as 1870 when they proposed that biological response was a function of chemical structure. Although studies in structure-property-activity relationships go back to antiquity since the times of Crum-Brown and Fraser [3], it is only in the recent times that one witnesses a vigorous study of it as an interdisciplinary area of molecular design and modelling by developing quantitative relationship between molecular

\* Correspondence: marjana.novic@ki.si

<sup>1</sup>Laboratory of Chemometrics, National Institute of Chemistry, Hajdrihova 19, Ljubljana 1000, Slovenia

Full list of author information is available at the end of the article

activity or property (such as partition coefficient ( $\log P$ ), boiling point, melting point, acid and base constant, chromatographic retention index, toxicity, or reactivity) and theoretical structural properties such as constitutional, electrostatic, geometrical, topological, or quantum chemical molecular characteristics [4-8]. A significant dimension to such studies has been extended largely due to the approach from the view point of developing quantitative relationship between reactivity and theoretical structural quantum chemical properties by soft computations which increase the probability of success and reduce the time and cost involvement in the chemical design, discovery and modelling of potential reactions and candidates [9-11]. But there are only a few quantitative structure-activation barrier relationship studies using quantum chemical structural parameters concerning the modelling of promising chemical reactions [12,13].

One of the important fields of bio orthogonal chemistry for designing of efficient reactions is well known Diels-Alder ligation because of its higher rate and selectivity in water [14-21]. The Diels-Alder reaction is an organic cycloaddition between a conjugated diene and a substituted alkene, commonly termed the dienophile, to form a substituted cyclohexene system. Otto Paul Hermann Diels and Kurt Alder first documented the novel reaction in 1928 for which they were awarded the Nobel Prize in Chemistry in 1950 for their work on the eponymous reaction. The Diels-Alder reaction is generally considered as one of the more useful reactions in organic chemistry since it requires very little energy to create a cyclohexene ring, which is useful in many other organic reactions [22-26] to gain insights into structure, dynamics and function of bio molecules. Diels-Alder ligation of electron-rich hexadiene and electron-deficient maleimide has been found to be facilitated by electron withdrawing groups (e.g., C=O and C $\equiv$ N) on the dienophile and electron-donating groups (e.g., -R and -OR) on the diene which would be effective for the bio-conjugation of peptides [27], small molecules [28,29], and oligonucleotides [30,31]. It has also been used for the bio-conjugation of carbohydrate to proteins [32,33].

There have been a number of successful explanations regarding the Diels-Alder reactions from the view point of molecular orbital theory. Woodward, Hoffmann, and Fukui used molecular orbital theory to explain that overlapping between  $p$  orbitals of the substituents on the dienophile with  $p$  orbitals of the substituents on the diene is favourable, helping to bring the two molecules together [34-36]. Tang et al. [13] carried out a systematic theoretical study based on M06-2X/6-31 + G(d)//B3LYP/6-31G(d) level on the design of new dienophiles in order to extend the scope of Diels-Alder ligation. A drawback of the Diels-Alder ligation is that the widely used maleimide moiety as a typical Michael acceptor can readily

undergo Michael addition with nucleophiles in living systems. Thus, Tang et al. calibrated a theoretical method to calculate the activation barriers of Diels-Alder reactions by benchmarking the calculations against the available experimental data for 72 non-catalysed Diels-Alder ligations. They have also calculated Diels-Alder barriers of  $\sigma$ -electron withdrawing group substituted alkenes, cyclic alkenes with consideration of electronic and ring strain effect and barriers of Diels-Alder and thiol addition reactions of designed alkenes which are efficient reactions and nucleophile-tolerant in living system. The method is time consuming and due to its complexity sometimes it fails to optimize the reactant complex at a transition state level.

Due to the above reasons, an attempt has been made in the present investigation to find an alternative and cheaper theoretical method to evaluate activation barriers of the Diels-Alder reactions based on quantitative structure-activation barrier relationship (QSABR) modelling utilizing theoretical quantum chemical descriptors calculated solely from the chemical structure of the ligation reactant molecules. The energies of the highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbitals (LUMO) are quantum chemical quantities that can govern the chemical reactions. They are calculated from the structures of reactant molecules utilizing quantum-chemical methods, which can explain reactivity correlated with the activation barriers of a complete molecule as well as of molecular fragments and substituents [37,38]. Computed descriptor based QSABR model produces comparable results as those calculated by Tang et al. at more complicated transition state theory based calculation using M06-2X/6-31 + G(d)//B3LYP/6-31G(d). The QSABR model was validated by introducing training and test set concept and was then applied for the prediction of Diels-Alder barriers of alkenes substituted with  $\sigma$ -electron withdrawing groups, cyclic alkenes and cyclopropene derivatives. The present protocol based on computed quantum chemical descriptors based on HOMO and LUMO energies of reactants can successfully predict activation barriers of  $\sigma$ -electron-withdrawing-group-substituted cyclopropenes, cyclic alkenes and barriers of Diels-Alder reactions studied by Tang et al. [13] at a more computationally demanding and not always successful transition state level. The proposed modelling methodology can be a useful tool to obtain the structure-activation barrier relationships of bio molecules and thus propose new ligations in click chemistry. The computational approach developed is a potential theoretical bench mark for the design of efficient and selective Diels-Alder ligation reactions.

## Results and discussion

### Computation of quantum chemical descriptors

We have calculated 24 quantum chemical properties using HOMO and LUMO energetics of the dienes and

dienophiles involved in 72 reactions considered in this study. The activation barriers calculated using the Eyring–Polanyi equation [2] with experimental reaction rates obtained from literature [19,39–43] have been taken into account to formulate quantitative structure-activation barrier relationships modelling. Quantum chemical structural descriptors were calculated from structures of standalone reactants, different dienes and dienophiles listed in Additional file 1: Table S1. Some of the reactions were chosen for the test set; see labelling of ID numbers of reactions, entries of the first column, and the explanation in the footnote to Additional file 1: Table S1. Chemical structures of dienes and dienophiles along with activation barrier data for all reactions are given in Additional file 1: Table S1.

The calculated descriptors are briefly described in Table 1. Hardness ( $\eta$ ), softness (*SOF*), electronegativity ( $\chi$ ), electrophilicity ( $\omega$ ), and dipole moment, which are calculated by Gaussian09 [44], and orbital interaction energy difference ( $\Delta E$ ), are the important electronic structure features used to describe stability, reactivity, chemical

potential and other related properties of molecules, see Eqs. (1, 2, 3, 4 and 5). Hardness has been used to understand chemical reactivity and stability of molecules. Electronegativity was introduced by Pauling as a power of an atom in a molecule to attract electron to itself. Softness is a property of molecule that measures the extent of chemical reactivity and is obtained as reciprocal value of hardness. Electrophilicity was proposed as a measure of energy lowering due to maximal electron flow between donor and acceptor. The most obvious and most often used quantity to describe the polarity is the dipole moment of the molecule. The total dipole moment, however, reflects only the global polarity of a molecule.

Inspired by the work of Maynard et al. [45], Parr and co-workers [46] have provided a definition of electrophilicity ( $\omega$ ) as

$$\omega = \frac{\mu^2}{2\eta} = \frac{\chi^2}{2\eta} \quad (1)$$

**Table 1 List of calculated quantum chemical properties used in this study**

No.	Notation	Definition
1	HOMO <sub>d</sub>	Highest occupied molecular orbital energy of the diene
2	LUMO <sub>a</sub>	Lowest unoccupied molecular orbital energy of the dienophile
3	HOMO <sub>a</sub>	Highest occupied molecular orbital energy of the dienophile
4	LUMO <sub>d</sub>	Lowest unoccupied molecular orbital energy of the diene
5	(HOMO <sub>d</sub> - LUMO <sub>a</sub> )	Difference between HOMO <sub>d</sub> and LUMO <sub>a</sub>
6	(HOMO <sub>a</sub> - LUMO <sub>d</sub> )	Difference between HOMO <sub>a</sub> and LUMO <sub>d</sub>
7	$\Delta E$	Orbital interaction energy difference between the corresponding orbitals of the reactants as calculated by the formula proposed by Sustmann.
8	$\eta_d$	Hardness of the diene
9	$\eta_a$	Hardness of the dienophile
10	( $\eta_d - \eta_a$ )	Difference between Hardness of the diene and dienophile
11	$\chi_d$	Electronegativity of the diene
12	$\chi_a$	Electronegativity of the dienophile
13	( $\chi_d - \chi_a$ )	Difference between electronegativity of the diene and dienophile
14	$\omega_d$	Electrophilicity of the diene
15	$\omega_a$	Electrophilicity of the dienophile
16	( $\omega_d - \omega_a$ )	Difference between electrophilicity of the diene and dienophile
17	$\Delta$ HOMO	Difference between HOMO levels of the diene and dienophile
18	$\Delta$ LUMO	Difference between LUMO levels of the diene and dienophile
19	SOF <sub>d</sub>	Softness of the diene
20	SOF <sub>a</sub>	Softness of the dienophile
21	$\Delta$ SOF	Difference between softness of the diene and dienophile
22	DM <sub>d</sub>	Dipole moment of the diene
23	DM <sub>a</sub>	Dipole moment of the dienophile
24	$\Delta$ DM	Difference between dipole moments of the diene and dienophile

<sup>a</sup> dienophile; <sup>d</sup>diene.

where  $\mu$  is the chemical potential, a Lagrange multiplier associated with the normalization of density, defined as

$$\mu = \left( \frac{\partial E}{\partial N} \right)_{V(r)} = -\frac{I+A}{2} = \frac{E_{LUMO} + E_{HOMO}}{2} \quad (2)$$

$\eta$  is the absolute hardness given by

$$\mu = \left( \frac{\partial^2 E}{\partial N^2} \right)_{V(r)} = \frac{I-A}{2} = \frac{E_{LUMO} - E_{HOMO}}{2} \quad (3)$$

and softness (*SOF*) is given by ( $1/\eta$ ). It should be noted that softness, hardness and polarizability are related quantities as molecules with high hardness/low softness have low polarizability and vice versa. Moreover, Vela and Gazquez [47] demonstrated a linear relationship between polarizability and the global softness of a system.

Parr et. al. [46] define the electronegativity ( $\chi$ ) as the negative of chemical potential ( $-\mu$ )

$$\chi = -\mu = \left( \frac{\partial E}{\partial N} \right)_{V(r)} = \frac{I+A}{2} = -\frac{E_{LUMO} + E_{HOMO}}{2} \quad (4)$$

In the above definitions for an  $N$ -electron system with total energy  $E$  and external potential  $V(r)$ ,  $I$  and  $A$  are the ionization potential and the electron affinity, respectively, whereas  $E_{HOMO}$  and  $E_{LUMO}$  are the energies of the highest-occupied and lowest-unoccupied molecular orbitals, respectively. According to the Koopman's theorem [48],  $I$  is simply the eigenvalue of HOMO with a negative sign and  $A$  is the eigenvalue of LUMO with a negative sign.

The interaction energy difference ( $\Delta E$ ) between the corresponding orbital of the dienes and dienophiles was calculated using the following formula proposed by Sustmann [49]:

$$\Delta E = \frac{(HOMO_d - LUMO_a) + (HOMO_a - LUMO_d)}{(HOMO_d - LUMO_a) * (HOMO_a - LUMO_d)} \quad (5)$$

where  $d$  is referring to the dienes and  $a$  indicates dienophiles. In this simplified form of the interaction energy, we assume equal contributions of the atomic orbital coefficients at the centres where the new bonds are formed together with the resonance integral as a measure for the strength of the interaction of the reactants.

Quantum chemical descriptors have long been producing a crucial source of information for modelling chemical reactivity [50-56] and thus have great importance in the development of quantitative structure activation barrier relationships dealing with the chemical, physical, biochemical, and pharmacological properties of chemical compounds. A total number of 24 quantum chemical

descriptors are calculated solely from the structures of dienes and dienophiles, which are subjected to model quantitative structure-reactivity and quantitative structure activation barrier relationships utilizing Stepwise-multiple linear regression methods.

#### Linear QSABR model details

As it is shown in Additional file 1: Table S1, we have 72 Diels Alder reactions along with experimental activation barriers considered from the literatures. Stepwise-MLR method has been applied to develop quantitative structure activation barrier relationship modeling which focus influences of quantum chemical descriptors (described in Table 1) towards activation energy barriers that can be predicted easily with the application of such models avoiding complexity of the calculation at transition state level. QSABR model developed for all Diels-Alder reactions is given as

$$\begin{aligned} (\Delta G) = & -6.4 \pm 8.5 + (12.67 \pm 3.4)\Delta SOF \\ & + (1846.00 \pm 748)\omega_d - (148.00 \pm 45)\Delta HOMO \\ & + (291.00 \pm 73)LUMO_d + (145.00 \pm 54)\omega_a \\ & - (485.00 \pm 252)\chi_d \end{aligned} \quad (6)$$

$N = 72$ ,  $R^2 = 0.831$ ,  $F = 53$ ,  $Q_{Loo}^2 = 0.784$ ,  $PRESS = 437.271$ ,  $SE = 2.30$

In the above equation,  $N$ ,  $R^2$ ,  $Q_{Loo}^2$ ,  $PRESS$  and  $SE$  represent number of observations, squared regression coefficient, leave-one-out (LOO) cross-validated  $R^2$ , predictive sum of square deviations and standard error of the regression, respectively. These are commonly used measures of acceptance of QSAR models [57-59].  $R^2$  and cross validated (LOO)  $Q^2$  of a model can be obtained from:

$$R^2 = 1 - \frac{\sum_{i=1}^N [\Delta G_{exp} - \Delta G_{calc}]^2}{\sum_{i=1}^N [\Delta G_{exp} - \bar{\Delta G}]^2} \quad (7)$$

$R^2$  is a measure of explained variance.

$$Q_{Loo}^2 = 1 - \frac{\sum_{i=1}^N [\Delta G_{exp} - \Delta G_{pred}]^2}{\sum_{i=1}^N [\Delta G_{exp} - \bar{\Delta G}]^2} \quad (8)$$

where,  $\Delta G_{exp}$ ,  $\Delta G_{calc}$ ,  $\Delta G_{pred}$  indicate experimental, calculated and predicted activation barriers respectively and  $\bar{\Delta G}$  indicates mean of experimental activation barrier values.

In this set of reactions, the computed quantum chemical structural descriptors represent a significant impact on the activation energy barriers. Eq. (6) can explain 83% and predict 78% of variance of the activation barriers of the studied reactions. LUMO level of the diene,

difference between softness and difference between the HOMO levels of the dienes and dienophiles, electrophilicity of the dienes and dienophiles and electronegativity of the dienes can produce significant impact on the activation barriers.

Thus it is evident from this study that the application of quantum chemical structural properties can generate quantitative structure activation barrier relationship model with good prediction accuracy of the activation barriers. The model should be statistically validated prior to prediction of new reactions using such QSABR model.

### QSABR model validation and prediction of activation barriers

Further evaluation of model's stability and prediction ability has been performed by dividing the total data set (72 reactions) into training and test sets at a random basis, taking care of equal distribution of the objects over the whole reactivity space. Therefore both sets contain reactions from the available data space. Compounds with asterisk in Additional file 1: Table S1 were selected as test set comprising of 24% of the total reaction data. The quality of the training model is calculated by  $R^2$  and cross-validated  $Q_{Loo}^2$  values for the training set and an external validation was performed by calculating predictive  $R^2$  ( $R_{pred}^2$ ) for the test set reactions. Predictive  $R^2$  ( $R_{pred}^2$ ) for the test set is calculated as

$$R_{pred}^2 = 1 - \frac{\sum_{i=1}^N [\Delta G_{exp(test)} - \Delta G_{pred(test)}]^2}{\sum_{i=1}^N [\Delta G_{exp(test)} - \bar{\Delta G}_{exp(train)}]^2} \quad (9)$$

where,  $\Delta G_{exp(test)}$  and  $\Delta G_{pred(test)}$  indicate experimental and predicted activation barriers of the test set and  $\bar{\Delta G}_{exp(train)}$  indicates mean of experimental activation barriers of the training set. For a predictive model, the value of  $R_{pred}^2$  should be more than 0.5 [60].

The quantitative structure activation barrier relationship model formulated by using stepwise-MLR method for the training reaction set is obtained as

$$\begin{aligned} \Delta G = & -121.3 \pm 43.7 - (195.00 \pm 49)\Delta LUMO \\ & + (1286 \pm 333)\omega_a + (1127.00 \pm 326)(\omega_d - \omega_a) \\ & + (493.00 \pm 147)LUMO_d - (6.10 \pm 2.5)\Delta E \\ & + (7.10 \pm 3.9)SOF_d + (0.21 \pm 0.14)DM_a \end{aligned} \quad (10)$$

$N = 55$ ,  $R^2 = 0.865$ ,  $F = 43$ ,  $Q_{Loo}^2 = 0.800$ ,  $PRESS = 300.461$ ,  $SE = 2.06$ ,  $R_{pred}^2 = 0.880$ .

This QSABR equation can explain and internally predict 86.5% and 80% of variances of the activation barriers of the studied training reactions and it can produce 88% of the external model predictability for the test reactions. Due to rather high error in coefficients shown

in parentheses in Eq. (10), the applicability domain of the MLR model was assessed during the external validation procedure and reported below, in the chapter "Prediction of activation barriers for new reactions". One can observe that many of the descriptors selected when using the whole data set are different from the descriptors in the model built using the training set only. The differences may occur because the descriptors used are correlated and the stepwise method can use one descriptor or a set of descriptors to replace another one while producing a model of similar quality. Due to a large search space several local minima of similar quality can be found, which all provide reasonably good predictions. Also the chemical properties represented in a selected set of descriptors and thus correlating most with the reaction barriers can be interpreted similarly.

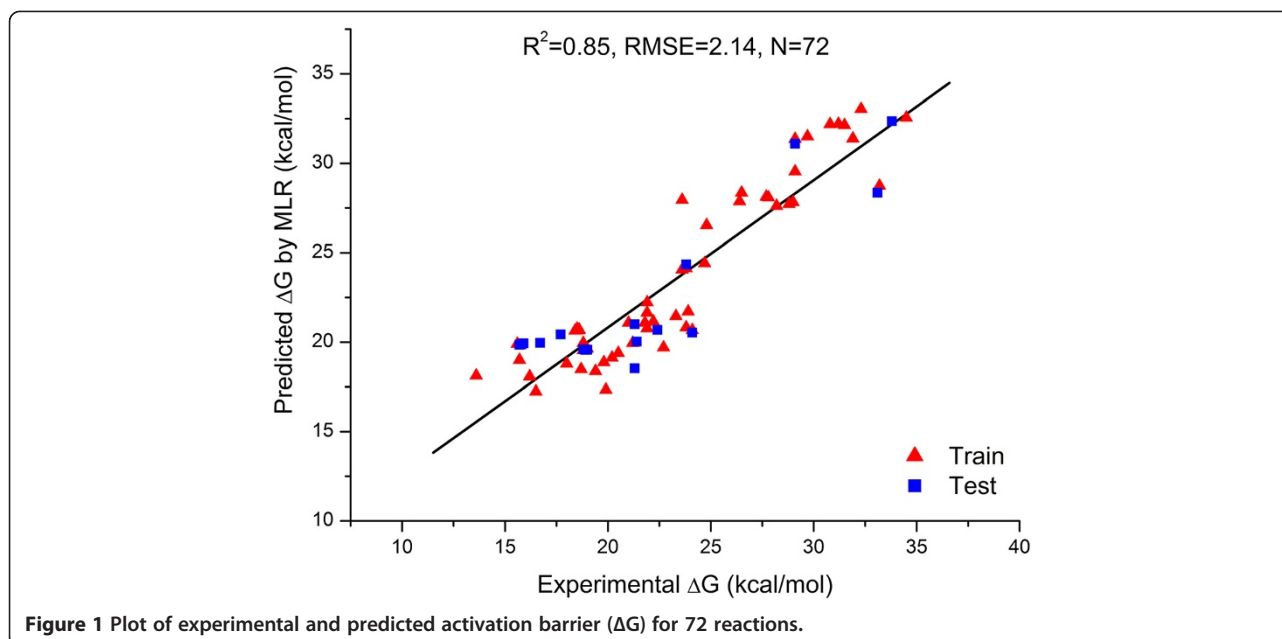
Electrophilicity and dipole moment of the dienophile, electrophilicity differences between the reactants, LUMO level and softness of the dienes have positive impact on the activation barriers. It means that decreasing the values of these variables may produce lower activation barriers for the Diels-Alder reactions and may give higher rate of reactions. Negative coefficients of  $\Delta LUMO$  and  $\Delta E$  indicate that small increase in the difference between LUMO energies and orbital interaction energies of dienes and dienophiles may decrease activation barriers. This distinction is of fundamental importance from the physical chemist's point of view and should be helpful for designing of promising Diels-Alder reactions prior to experimental synthesis.

Eq. (10) has been used to predict activation barriers of the training and test reactions. Predicted activation barriers for all 72 observations are given in Additional file 1: Table S1. The plot of experimental versus predicted activation barriers for training and test reactions is represented in Figure 1. Root mean square error of prediction is 2.14 kcal/mol and linear correlation coefficient between experimental versus predicted activation barrier is  $R^2 = 0.85$  which is almost satisfying according to the linear correlation coefficient ( $R^2 = 0.93$ ) of the same reactions calculated by Tang et al. [13].

From the Additional file 1: Table S1 it is evident that the predicted activation barriers for all the reactions are in good agreement with their corresponding experimental results. A detailed comparison of experimental and predicted activation barriers is evident from Figure 2, which includes also the Artificial Neural Network model results (ANN).

### Nonlinear QSABR model details

The error back-propagation neural network model was developed with 55 training data (55 reactions, see Additional file 1: Table S1). The ANN which was chosen as the best model had input layer with seven neurons, five hidden neurons and one neuron in the output layer. Optimal

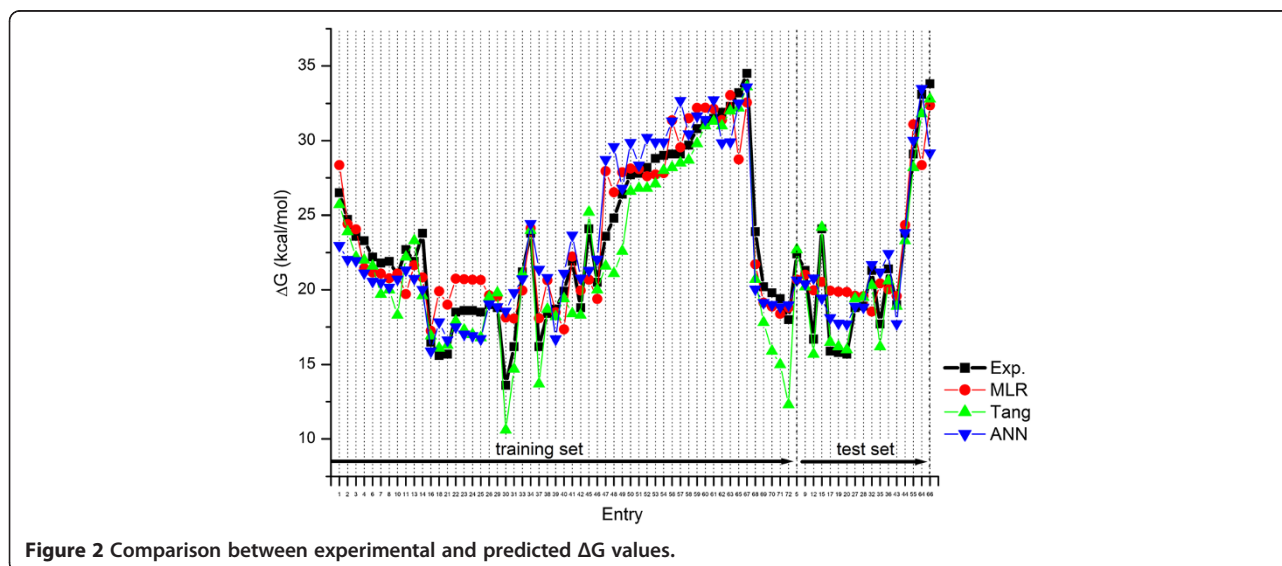


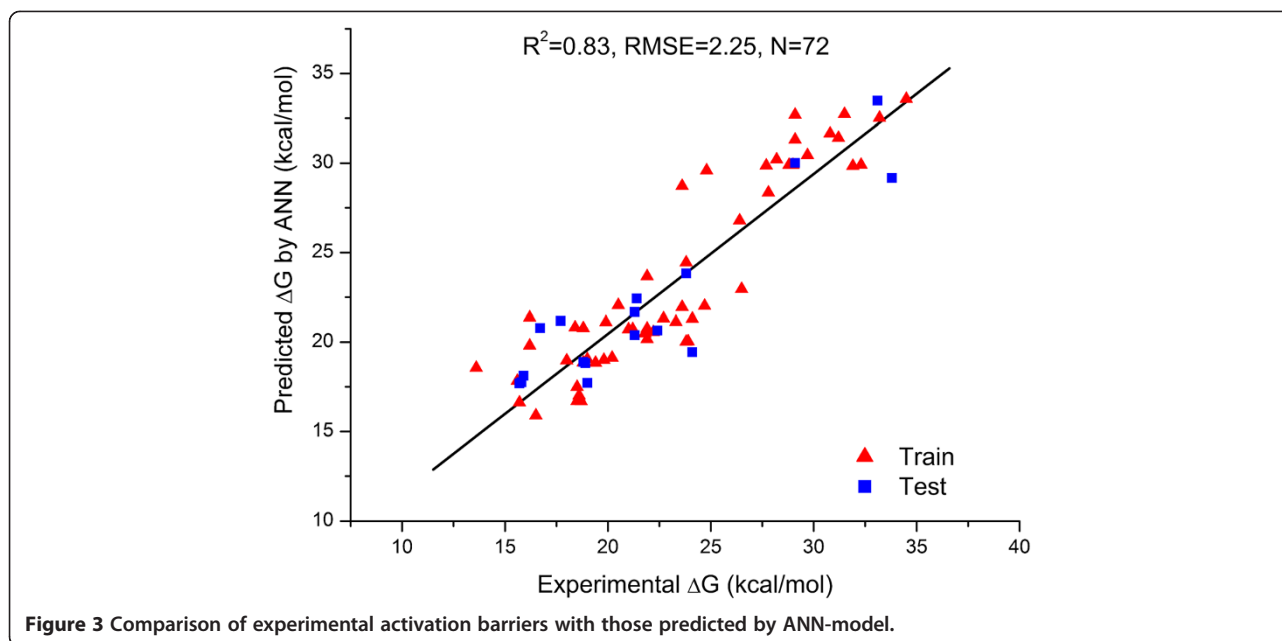
parameters were chosen according to the lowest RMS errors obtained from Leave-one-out on training data, see Additional file 2: Figure S1, taking into account that the residuals should be randomly distributed around zero. The model was tested by 17 test objects as explained in MLR modeling approach. The predictions for all models and calculations from Tang [13] are compiled in Figure 2.

BP-ANN with 5 hidden neurons was trained during 4800 epochs with learning rate 0.6 and momentum 0.2. Observed RMS error was 2.22 kcal/mol and 2.33 kcal/mol for the training set objects and testing set objects, respectively. Leave-one-out for the training set objects resulted in

RMS error equal to 2.74 kcal/mol. Predictive  $R^2_{\text{pred}}$  for test set was 0.83 and  $Q^2_{\text{Loo}}$  was 0.76. The regression plot of experimental versus predicted  $\Delta G$  values by the BP-ANN model is shown in Figure 3.

There is a small difference between RMS errors for ANN and MLR model which is slightly in favour of the developed MLR model. This result implies that the non-linearity in the relationship between the descriptors ( $LUMO_d$ ,  $\Delta E$ ,  $\omega_a$ ,  $(\omega_d - \omega_a)$ ,  $\Delta LUMO$ ,  $SOF_d$ , and  $DM_a$ ) and the property (activation barrier) for the studied Diels Alder reactions is not essential and does not represent a problem for QSABR models. When comparing RMS error





of predictions (training + test set results) for Tang's, MLR and ANN model, we observe the values 1.76, 2.14, and 2.25 respectively. We can observe in Figure 2, that in general  $\Delta G$  values predicted by all three models match well with the experimental ones. While Tang's results show very good fit for the set of reactions we have used as test set objects, their predictions for the reactions from our training set seem for most of the cases a little below experimental values, especially in the case of reactions 30, 37, 70, 71, and 72 (see Figure 2 and Additional file 1: Table S1 for details). In a closer examination of the prediction errors, there are approximately 84% of values lower than experimental ones in Tang's calculations, and around 45% in MLR and ANN models. The average residuals of all predictions are  $-1.28$ ,  $0.29$  and  $0.14$  kcal/mol in Tang's calculation, MLR and ANN, respectively. Although the results are acceptable, they are suggesting a moderate systematic error in Tang's calculations. Since we are interested in the diene-dienophile pairs which give low activation barrier, the values with the predictions with lowest  $\Delta G$  values are the most interesting. And as we shall see in the following section, all three models identify the same new reactions as the most promising ones.

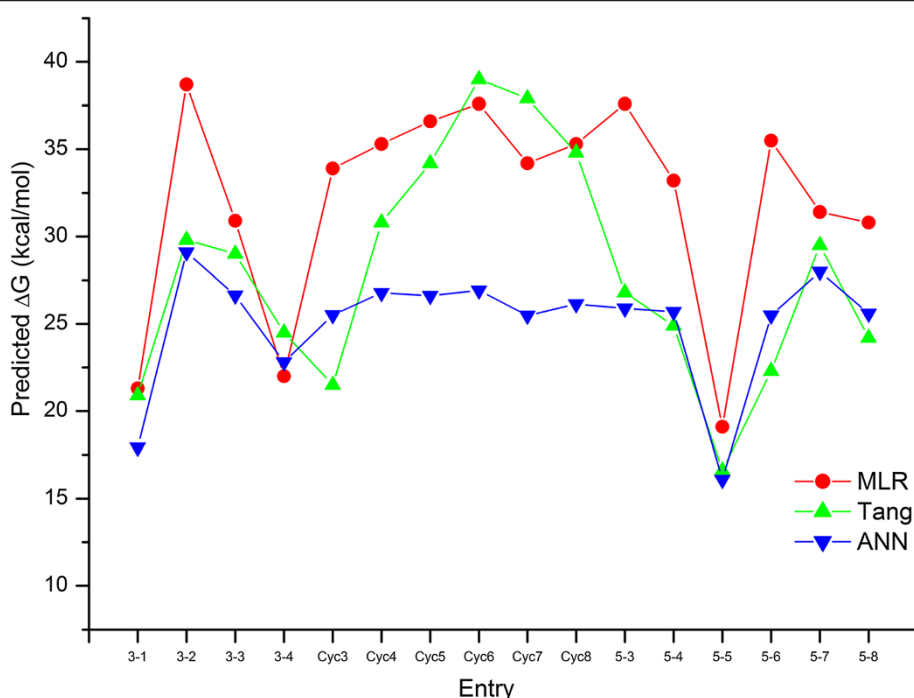
#### Prediction of activation barriers for new reactions

The linear and nonlinear models developed were further applied to predict Diels-Alder activation barriers of  $\sigma$ -electron withdrawing group substituted alkenes, cyclic alkenes with consideration of ring strain effect and barriers of Diels-Alder reactions which are promising and nucleophile-tolerant in the living system. These were compared with the activation barriers calculated

by Tang et al. as shown in Additional file 3: Table S2 and Figure 4. Structural optimization and descriptor computations were performed for cyclic alkenes (Cyc3 to Cyc8) considering ring strain effect. Prediction of activation barriers of these cyclic alkenes in MLR model follow the trends as obtained by Tang et al. Increasing the angle, barriers increase in the following sequences: cyclopropene (angle =  $64.6^\circ$ ;  $+33.9$  kcal/mol) < cyclobutene (angle =  $94.4^\circ$ ;  $+35.3$  kcal/mol) < cyclopentene (angle =  $112.1^\circ$ ;  $+36.6$  kcal/mol) < Cyclohexene (angle =  $123.5^\circ$ ;  $+37.6$  kcal/mol). Cyclohexene produces highest barrier prediction which goes to decreasing for larger homologs such as cycloheptene ( $+34.2$  kcal/mol) and cyclooctene ( $+35.3$  kcal/mol).

The above comparative results between  $\Delta G$  calculated (by Tang et al. [13]) and predicted (by our MLR and ANN models) indicated that the current theoretical protocol based on the modelling of quantitative structure activation barriers relationship utilizing computed quantum chemical descriptors. MLR model could satisfactorily predict the activation energy barriers of Diels-Alder ligation reactions if compared to Tang's calculations, except for the entries 3-2, cyc3, 5-3, 5-4 and 5-6 for which deviations from Tang's calculations are higher. Conversely, in the ANN model the four mentioned predictions match well with Tang's calculations, while the deviations are higher for the reactions with cyclic alkenes as dienophiles (Cyc4 - Cyc8, see Figure 4).

Prediction of activation barrier using theoretical quantum chemical descriptors based protocol is quite easy process and less time consuming. So such model can be applied to predict activation barrier of Diels-Alder ligation



**Figure 4** Predicted activation barriers for new reactions listed in Additional file 3: Table S2

reactions before experimental investigation to gain insight for designing of promising Diels-Alder reactions. There is a concern regarding the influence of the solvent in the newly proposed approach. We have observed that the calculated quantum chemical descriptors did not differ significantly for the same reaction performed in different solvents, as for example the reactions 17 and 43 (see Additional file 1: Table S1), although the experimental reaction barriers were significantly different. Consequently, our QSABR model cannot differentiate well these two reactions. On the other hand, we are proposing this fast and efficient new methodology for identification of new click reactions for bio systems, so they are run in water.

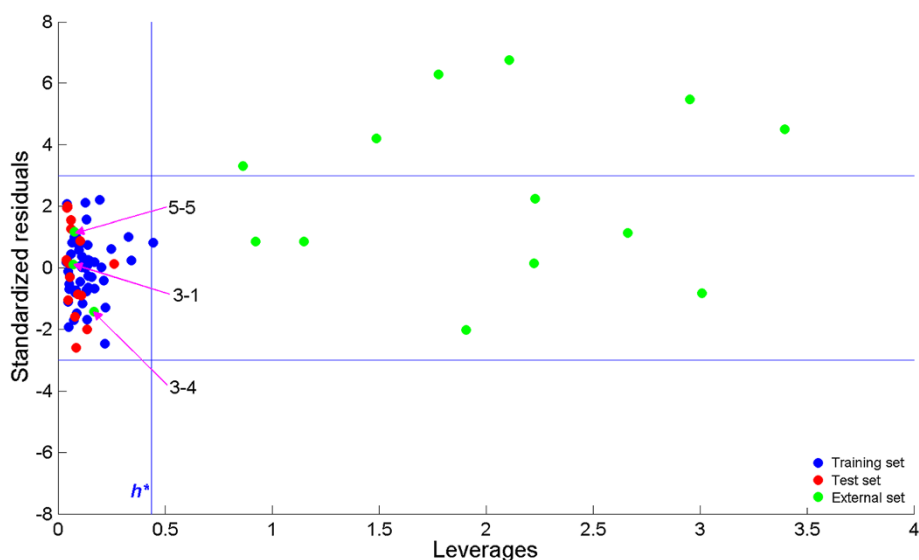
After the MLR model was built, its applicability domain was assessed for the reactions under the study. The applicability domain was investigated using leverage approach method as described in [61-63]. The results are shown in Figure 5 where blue, red and green dots represent training set, test set and prediction set reactions, respectively. Horizontal blue lines represent cut-off value for standardized residuals and the vertical blue line shows the warning value for leverage values obtained from hat matrix. It can be observed that practically all reactions in the training and test set may be considered as inside the applicability domain, while most of the reactions in the prediction set do not belong to the applicability domain of the MLR model. Only for 3 reactions in the prediction set it can be stated that the MLR model is suitable and reliable predictions have been made. The three

reactions are 3-1, 3-4 and 5-5 which have also been found to be reactions with low activation barriers as predicted by the MLR model. Since no experimental values were available for the new reactions, the predictions for the new reactions were compared with Tang's calculations. Williams plot in Figure 5 shows the importance of applicability domain assessment. From the figure it is visible that approximately half of the applicability domain outliers can be considered also as prediction outliers.

### Experimental

*In silico* experiments have been performed for 72 non-catalyzed Diels-Alder reactions of different dienes and dienophiles with experimentally determined reaction rates obtained from literature [19,39-43]. The activation barriers of 72 reactions calculated by Eyring-Polanyi equation [2] were used as target properties in the QSABR modelling, while the quantum chemical structural descriptors were calculated from structures of standalone reactants, different dienes and dienophiles listed in Additional file 1: Table S1. The structures of all the reactant molecules have been prepared in the computer readable form by Chem3D Ultra [64]. The geometries of reactants were fully optimized at the HF/6-31G(d) level to obtain most stable conformation. Quantum chemical descriptors were calculated utilizing Gaussian 09 software [44]. For all quantum chemical calculations, the solvents specified in the study by Tang et al. [13] have been considered. Solvents are introduced by their dielectric constants ( $\epsilon$ )





**Figure 5** Assessment of the applicability domain for the MLR model.

under the Polarizable Continuum Model (PCM) [65]. The solvents used for corresponding reactions were: Benzene ( $\epsilon = 2.2706$ ); Acetonitrile ( $\epsilon = 35.688$ ); Water ( $\epsilon = 78.3553$ ); 1,4-Dioxane ( $\epsilon = 2.2099$ ).

## Conclusions

Quantitative structure-activation barrier relationship studies have been performed utilizing computed quantum chemical descriptors solely calculated from a number of different dienes and dienophiles constituting 72 non-catalyzed Diels-Alder reactions. Stepwise algorithm has been applied for variable selection and the QSABR models have been developed by MLR method. QSABR model reveals that quantum chemical descriptors produce significant impact on the activation barriers for these reaction data. Quantum chemical descriptors such as LUMO level of the diene, difference between softness and difference between the HOMO levels of the dienes and dienophiles, difference between LUMO energies of the dienes and dienophiles, electrophilicity of the dienophiles and dienes and its differences, electronegativity of the dienes, dipole moment of the dienophile, softness of the dienes and orbital interaction energy difference ( $\Delta E$ ) of dienes and dienophiles (as depicted in Eqs. (6) and (10)) have crucial influences on the activation barrier of the Diels-Alder ligations. Reliability of the training QSABR model was confirmed by statistical validation and thus, the developed QSABR model has been used to predict the Diels-Alder activation barriers of a new set of reactions between 1,3-butadiene and  $\sigma$ -electron withdrawing group substituted alkenes and cyclic alkenes designed and calculated by Tang et al. [13].

It is evident from comparative study of calculated and predicted activation barriers (reported in Additional file 3:

Table S2) that our theoretical protocol using computed molecular descriptors of reactants alone can provide a good quality predictive model for the Diels-Alder ligation considered in the present investigation. The aim of this work was to show how the molecular orbital theory provides the descriptors of molecular structure that are sufficient to find reasonable correlation with the reaction barriers. Alternatively, the “atoms in molecules” analysis [64] would provide useful information (descriptors) on electronic structure of the molecules. However, regarding its inherent theoretical framework it decouples from molecular orbital concept of bonding and can thus not supplement our present approach. Nevertheless, the critical points data [64,65] might serve as an alternative source of the stand-alone descriptors which could potentially be successful in predicting chemical reactivity.

With the calculation method described in our work, which considers reactant molecules separately, no problems with convergence were observed. It is also less time consuming than the computationally demanding approach based on the transition state theory; yet, the reliable predictions of the activation barriers can be obtained as demonstrated on a large set of Diels-Alder reactions. Since the proposed model for prediction of activation barriers is based on relatively simple calculations of HOMO and LUMO energies, it can become widely used by the experimentalists in organic chemistry labs as a cheap theoretical method to design new promising Diels-Alder reactions.

## Methods

### Computational section

Molecular orbital theory was applied to calculate quantum-chemical descriptors (all based on the HOMO, LUMO

energies and dipole moments). The calculations were performed on high efficient workstation having an Intel(R) Xeon(R) CPU E5620 (12 GB RAM) with the Windows 64-bit operating system. Chem3D Ultra [66] was used to input chemical structures; Gaussian 09 [44] was applied for geometry optimization and quantum chemical calculations; dielectric constants of corresponding solvents were considered in the Polarizable Continuum Model (PCM) [67].

#### Data processing and model development by Stepwise-MLR

Multivariate linear regression (MLR) analysis, one of the oldest data reduction methodologies, continues to be widely used in Quantitative Structure Activity/Property/Reactivity (QSAR, QSPR, QSRR) studies [10,68,69]. Selection of variables having significant influence on the reactivity is a key step in QSPR modelling of these reactants to eliminate the problems like chance correlations and multicollinearity. However, because of the co-linearity problem in MLR analysis, we removed the collinear descriptors. Utilizing every available descriptor that may produce a predictive model with a good correlation coefficient, makes the models difficult to interpret and do not stand up to external validation. An integral aspect of model development is to build the model with a small but appropriate set of descriptors with a view to interpret the relationships. This process forms the basis of a technique known as feature selection or variable selection. Among several search algorithms, stepwise forward-backward based feature selection coupled with MLR is the most popular method for building quantitative models and can explain the situation more effectively [60].

Statistical analyses have been carried out by using Minitab software [70]. One has to choose the values of the F statistic for the partial F tests that will determine if a variable is to enter or be removed from the model;  $F = 0.25$  has been chosen as the threshold for inclusion and exclusion of variables.

Initially, the stepwise method searches for an independent variable which correlates the most with the dependent variable. This variable is used as the first variable in the model. Next, from the remaining set of unused independent variables a new variable is selected so that the most of the response's variance is explained while keeping high values of F-test and correlation coefficient of the model. New variables are being added to the model until no significant improvement of the model can be made. In each consecutive step of adding a variable, the current model is also tested if the quality of the model can be improved by removing any previously selected variable, which is then discarded. Leave-one-out cross-validation was used to evaluate the stability and prediction ability of the generated models in each step.

#### Nonlinear modelling by neural networks

Artificial neural network based on back propagation of errors algorithm was applied as a nonlinear modelling method. The same training data were used as for MLR modelling. Back-propagation artificial neural network model (BP-ANN model) is usually presented as a "black box" which consists of input layer, hidden layer and output layer of neurons. Here only a brief description is given to clarify this "black box", while a more detailed description can be found in literature [71]. The input layer receives input signals that consist of independent variables presenting the inherent properties of the data set objects. The input signals are usually read from a text file; in this study the same descriptors were used as for MLR modelling. The input signals are transferred from the inactive input layer to the active hidden and output layers, in which the signals are transformed through the activation function. The neurons in the hidden and output layers are iteratively modified according to the input signals which are loaded repeatedly to the network until it has been sufficiently trained, which means that the output signals for each input match with the target values (desired properties of each object). The trained network provides us with predictions of one or more predefined properties that are located in the output layer.

Each active neuron transforms input signals using activation function which is usually sigmoidal function. For the first active layer of neurons (hidden layer), the neuron's output signal is calculated as given in Eq. (11), where  $Net$  is defined by Eq. (12) as a sum of descriptors values  $x_i$ , multiplied by the corresponding weights  $w_i$ ,  $n$  being the number of inputs (descriptors). Bias, usually 1, is added to the sum thus  $n + 1$  terms.

$$out = \frac{1}{1 + e^{-Net}} \quad (11)$$

$$Net = \sum_{i=1}^{n+1} w_i x_i \quad (12)$$

For the neurons in the subsequent layers, the inputs ( $x_i$  values) represent the outputs ( $out$  values) of the neurons from the preceding layer, while  $w_i$  values refer to the neurons of the selected layer. As mentioned above, the network needs to be trained with known input-target pairs of data. That means that appropriate values for all weights in the network should be found. Therefore a sufficiently large set of objects with known descriptor and target values is needed to tune the weights so that the outputs in the layer are reasonably close to the target values. Using the training set objects, BP-ANN is trained by back propagation of errors. The

correction of the weights is usually made immediately after each input (single object) and can be summarized using Eq.(13).

$$\Delta w_{jk}^l = \phi \delta_j^l \text{out}_k^{l-1} + \gamma \Delta w_{jk}^{l(\text{previous})} \quad (13)$$

Eq. (13) demonstrates how a correction of a weight is calculated for the neuron  $j$  in the current layer  $l$  for the input received from the neuron  $k$  in the preceding layer  $l-1$ . The correction is made using two terms. The first term shows that correction considers the error produced in the current neuron ( $\delta_j^l$ ) and the output from the neuron in the layer above ( $\text{out}_k^{l-1}$ ) which is multiplied by the learning rate ( $\phi$ ). The second term is used to reduce oscillations during the learning process and to overcome local minima. It takes the most recent correction of the weight ( $\Delta w_{jk}^{l(\text{previous})}$ ) multiplied by a momentum ( $\gamma$ ). When each input object has been used once to correct all weights in ANN, the learning procedure has passed one *epoch*. The learning rate and momentum are the two most important parameters, besides the network architecture, the number of hidden layers and the number of neurons in the hidden layers, which must be chosen before the training. Usually the parameters  $\phi$  and  $\gamma$  are decreasing with time during the learning procedure [71].

For the purpose of this study, BP-ANN had one hidden layer with one to five neurons, subjected further to a standard procedure of optimization of network parameters.

### Applicability domain assessment

As proposed in the report of the European Centre for the Validation of Alternative Methods [62], the applicability domain of a QSAR model is the response and chemical structure space in which the model makes predictions with a given reliability. Or in other words, as written by Eriksson et al. [63], applicability domain of QSAR model is the range within which it tolerates a new molecule. Thus, if new molecules, or as in our case Diels-Alder reactions, represented by chemical descriptors fall outside applicability domain of the model, the predictions of the model become unreliable.

The applicability domain of the MLR model was assessed using leverage approach method, where the leverage of object  $i$  can be calculated as  $h_i = x_i^T (X^T X)^{-1} x_i$  where  $x_i$  is a vector describing object  $i$  and  $X$  is a matrix composed of the objects used in the training. The warning leverage,  $h^*$ , is defined as  $3(n_d + 1)/n_o$ , where  $n_d$  and  $n_o$  are number of descriptors and number of objects used in the training set [62]. If the object's leverage exceeds the warning level value, the prediction for that objects is considered as unreliable, i.e. the object is outside the model's applicability domain.

### Additional files

**Additional file 1: Table S1.** Structures of diene and dienophiles along with activation barrier ( $\Delta G$ ) data.

**Additional file 2: Figure S1.** RMS errors and correlation coefficients of different models and distribution of residuals for the selected model. According to low RMS error obtained with Leave-one-out validation method (a), good correlation coefficient (b) and acceptable distribution of residuals in normal probability plot (c) model 30 was chosen as optimal.

**Additional file 3: Table S2.** Predicted activation barriers for new reactions.

### Abbreviations

ANN: Artificial neural network; BP-ANN: Back-propagation artificial neural network; HOMO: Highest occupied molecular orbital; LUMO: Lowest unoccupied molecular orbital; MLR: Multiple linear regression; PRESS: Predicted residual error sum of squares;  $Q_{\text{LoO}}$ : Leave-one-out cross-validated regression coefficient; QSABR: Quantitative structure-activation barrier relationship; QSAR: Quantitative structure-activity relationship; QSPR: Quantitative structure-property relationship; QSRR: Quantitative structure-reactivity relationship; R: Regression coefficient; SE: Standard error.

### Competing interests

The authors declare that they have no competing interests.

### Authors' contributions

SN calculated quantum chemical structural parameters for all reactants in 72 Diels-Alder reactions and developed the initial MLR model. AM selected the dataset from literature and advised the chemical optimization steps in QC calculations. VD developed ANN models, assessed applicability domain of MLR model. FM assisted in optimization of QM calculations and interpretation of the results. MN initiated the research and analysed the results. All authors contributed to the preparation of the manuscript. All authors read and approved the final manuscript.

### Acknowledgements

Authors are thankful to Slovenian Ministry of Education, Science, Culture, and Sport for financial support (Grant P1-0017) and to the European Union for the BioChemLig research grant (FP7-PEOPLE-ITN-2008-238434).

### Author details

<sup>1</sup>Laboratory of Chemometrics, National Institute of Chemistry, Hajdrihova 19, Ljubljana 1000, Slovenia. <sup>2</sup>Laboratory of Biomolecular Structure, National Institute of Chemistry, Hajdrihova 19, Ljubljana 1000, Slovenia.

Received: 24 July 2013 Accepted: 23 October 2013

Published: 30 October 2013

### References

1. Todeschini R, Consonni V: *Molecular Descriptors for Chemoinformatics, Revised and Enlarged Edition*. 2nd edition. Weinheim: Wiley-VCH; 2009.
2. Evans MG, Polanyi M: Some applications of the transition state method to the calculation of reaction velocities, especially in solution. *Trans Faraday Soc* 1935, **31**:875–894.
3. Crum-Brown A, Fraser TR: On the connection between chemical constitution and physiological action. Part 1. On the physiological action of the ammonium bases, derived from Strychia, Brucia, Thebaia, Codeia, Morphia and Nicotia. *Trans R Soc Edinburgh* 1868, **25**:151–203.
4. Randić M: On molecular identification numbers. *J Chem Inf Comput Sci* 1984, **24**:164–175.
5. Randić M: On characterization of molecular branching. *J Am Chem Soc* 1975, **79**:6609–6615.
6. Basak SC: *Mathematical descriptors in the prediction of property, bioactivity, and toxicity of chemicals from their structure*. *Curr Comput Aided Drug Des: A Chemical-Cum-Biochemical Approach*; 2013. in press (PMID: 24138422).
7. Pompe M, Novič M: Prediction of gas-chromatographic retention indices using topological descriptors. *J Chem Inf Comput Sci* 1999, **39**:59–67.
8. Mlinšek G, Novič M, Hodošček M, Šolmajer T: Prediction of enzyme binding: human thrombin inhibition study by quantum chemical and

- artificial intelligence methods based on X-ray structures. *J Chem Inf Comput Sci* 2001, **41**:1286–1294.
9. Wilcox CF, Carpenter BK: Quantitative prediction of structure-reactivity relationships for unimolecular reactions of unsaturated hydrocarbons. development of a semiempirical model. *J Am Chem Soc* 1979, **101**:3897–3905.
  10. Hemmateenejad B, Sanchooli M, Mehdipour A: Quantitative structure-reactivity relationship studies on the catalyzed Michael addition reactions. *J Phys Org Chem* 2009, **22**:613–618.
  11. Craig D, Slavov NK: A quantitative structure-reactivity relationship in decarboxylative Claisen rearrangement reactions of allylic tosylmalonate esters. *Chem Commun* 2008, (45):6054–6. doi: 10.1039/b812306c. Epub 2008 Oct 14.
  12. Michaelides A, Liu Z-P, Zhang CJ, Alavi A, King DA, Hu P: Identification of general linear relationships between activation energies and enthalpy changes for dissociation reactions at surfaces. *J Am Chem Soc* 2003, **125**:3704–3705.
  13. Tang S-Y, Shi J, Guo Q-X: Accurate prediction of rate constants of Diels-Alder reactions and application to design of Diels-Alder ligation. *Org Biomol Chem* 2012, **10**:2673–2682.
  14. Lindström UM: Stereoselective organic reactions in water. *Chem Rev* 2002, **102**:2751–2772.
  15. Otto S, Engberts JBFN: Hydrophobic interactions and chemical reactivity. *Org Biomol Chem* 2003, **1**:2809–2820.
  16. Breslow R, Zhu Z: Quantitative antihydrophobic effects as probes for transition state structures. 2. Diels-Alder reactions. *J Am Chem Soc* 1995, **117**:9923–9924.
  17. Breslow R, Dong SD: Biomimetic reactions catalyzed by cyclodextrins and their derivatives. *Chem Rev* 1998, **98**:1997–2011.
  18. Breslow R, Groves K, Mayer MU: Antihydrophobic cosolvent effects for alkylation reactions in water solution, particularly oxygen versus carbon alkylations of phenoxide ions. *J Am Chem Soc* 2002, **124**:3622–3635.
  19. Meijer A, Otto S, Engberts JBFN: Effects of the Hydrophobicity of the Reactants on Diels-Alder Reactions in Water. *J Org Chem* 1998, **63**:8989–8994.
  20. Otto S, Blokzijl W, Engberts JBFN: Diels-Alder reactions in water. effects of hydrophobicity and hydrogen bonding. *J Org Chem* 1994, **59**:5372–5376.
  21. Sun XL, Yang LC, Chaikof EL: Chemoselective immobilization of biomolecules through aqueous Diels-Alder and PEG chemistry. *Tetrahedron Lett* 2008, **49**:2510–2513.
  22. Diels O, Alder K: Synthesen in der hydroaromatischen Reihe. *Justus Liebigs Ann Chem* 1928, **460**:98–122.
  23. Kloetzel MC: The Diels-Alder reactions with Maleic Anhydride. *Org React* 1948, **4**:1–59.
  24. Holmes HL: The Diels-Alder reaction: ethylenic and acetylenic dienophiles. *Org React* 1948, **4**:60–173.
  25. Kagan HB, Riant O: Catalytic asymmetric Diels Alder reactions. *Chem Rev* 1992, **92**:1007–1019.
  26. Nicolaou KC, Snyder SA, Montagnon T, Vassilikogiannakis G: The Diels-Alder reaction in total synthesis. *Angew Chem Int Ed* 2002, **41**:1668–1698.
  27. Braun K, Wiessler M, Waldeck W, Kliem C, Pipkorn R: The Diels-Alder-reaction with inverse-electron-demand, a very efficient versatile click-reaction concept for proper ligation of variable molecular partners. *Int J Med Sci* 2010, **7**:19–28.
  28. Hill KW, Taunton-Rigby J, Carter JD, Kropp E, Vagle K, Pieken W, McGee DPC, Husar GM, Leuck M, Anziano DJ, Sebesta DP: Diels-Alder bioconjugation of Diene-modified oligonucleotides. *J Org Chem* 2001, **66**:5352–5358.
  29. Husar GM, Anziano DJ, Leuck M, Sebesta DP: Nucleosides: covalent modification and surface immobilization of nucleic acids via the Diels-Alder bioconjugation method. *Nucleotides Nucleic Acids* 2001, **20**:559–566.
  30. Marchan V, Ortega S, Pulido D, Pedrosa E, Grandas A: Diels-Alder cycloadditions in water for the straightforward preparation of peptide-oligonucleotide conjugates. *Nucleic Acids Res* 2006, **34**:e24.
  31. Steven V, Graham D: Oligonucleotide conjugation to a cell-penetrating (TAT) peptide by Diels-Alder cycloaddition. *Org Biomol Chem* 2008, **6**:3781–3787.
  32. Thorson JS, Langenhan JM: Recent Carbohydrate-based chemoselective ligation applications. *Curr Org Synth* 2005, **2**:59–81.
  33. Tiefenbrunn TK, Dawson PE: Chemoselective ligation techniques: modern applications of time-honored chemistry. *Biopolymers* 2010, **94**:95–106.
  34. Hoffmann R, Lipscomb WN: The Boron hydrides; LCAO-MO and resonance studies. *J Chem Phys* 1962, **37**:2872–2883.
  35. Fukui K: Role of frontier orbitals in chemical reactions. *Science* 1982, **218**:747–754.
  36. Woodward RB, Hoffmann R: The conservation of orbital symmetry. *Angew Chem Int Edit* 1969, **8**:781–932.
  37. Gothelf KV, Jorgensen KA: Asymmetric 1,3-Dipolar cycloaddition reactions. *Chem Rev* 1998, **98**:863–909.
  38. Fukui K, Yonezawa T, Shingu H: A Molecular Orbital Theory of Reactivity in Aromatic Hydrocarbons. *J Chem Phys* 1952, **20**:722–725.
  39. Craig D, Fowler RB, Shipman JJ: The rate of reaction of maleic anhydride with 1,3-Dienes as related to diene conformation. *J Am Chem Soc* 1961, **83**:2885–2891.
  40. Rulisek L, Sebek P, Havlas Z, Hrabal R, Capek P, Svatos A: An experimental and theoretical study of stereoselectivity of furan-maleic anhydride and furan-maleimide diels-alder reactions. *J Org Chem* 2005, **70**:6295–6302.
  41. Engberts JBFN, Rispens T: Micellar catalysis of Diels-Alder reactions: substrate positioning in the micelle. *J Org Chem* 2002, **67**:7369–7377.
  42. Sauer J: Diels-Alder reactions. II. The reaction mechanism. *Angew Chem Int Ed* 1967, **6**:16–33.
  43. Sauer J, Wiest H, Mielert A: Diels-Alder reaction. I. Reactivity of dienophiles towards cyclopentadiene and 9,10-Dimethyl anthracene. *Chem Ber* 1964, **97**:3183–3207.
  44. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Scalmani G, Barone V, Mennucci B, Petersson GA, Nakatsuji H, Caricato M, Li X, Hratchian HP, Izmaylov AF, Bloino J, Zheng G, Sonnenberg JL, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Vreven T, Montgomery JA Jr, et al. Gaussian 09, Gaussian, Inc.; 2009.
  45. Maynard AT, Huang M, Rice WG, Covell DG: Reactivity of the HIV-1 nucleocapsid protein p7 zinc finger domains from the perspective of density-functional theory. *Proc Natl Acad Sci USA* 1998, **95**:11578–11583.
  46. Parr RG, Chattaraj PK: Principle of maximum hardness. *J Am Chem Soc* 1991, **113**:1854–1855.
  47. Vela A, Gazquez JL: A relationship between the static dipole polarizability, the global softness and the Fukui function. *J Am Chem Soc* 1990, **112**:1490–1492.
  48. Koopmans T: Über die Zuordnung von Wellenfunktionen und Eigenwerten zu den Einzelnen Elektronen Eines Atoms. *Physica* 1, **1**:104–113.
  49. Sustmann R: Orbital energy control of cycloaddition reactivity. *Pure Appl Chem* 1974, **40**:569–593.
  50. Brown RE, Simas AM: On the applicability of CNDO indices for the prediction of chemical reactivity. *Theor Chim Acta* 1982, **62**:1–16.
  51. Gruber C, Buss V: Quantum-mechanically calculated properties for the development of quantitative structure-activity relationships (QSARs). pKA-values of phenols and aromatic and aliphatic carboxylic acids. *Chemosphere* 1989, **19**:1595–1609.
  52. Bodor N, Gabanyi Z, Wong C-K: A new method for the estimation of partition coefficient. *J Am Chem Soc* 1989, **111**:3783–3786.
  53. Buydens L, Massart D, Geerlings P: Prediction of gas chromatographic retention indexes with topological, physico-chemical, and quantum chemical parameters. *Anal Chem* 1983, **55**:738–744.
  54. Murugan R, Grendze MP, Toomey JE Jr, Katritzky AR, Karelson M, Lobanov VS, Rachwal P: Predicting physical properties from molecular structure. *Chem Tech* 1994, **24**:17–23.
  55. Liu P-Y, Wu Y-J, Pye CC, Thornton PD, Poirier RA, Burnell DJ: Facial Selectivity in the Diels-Alder reactions of 2,2-disubstituted cyclopent-4-ene-1,3-dione derivatives and a computational examination of the facial selectivity of the Diels-Alder reactions of structurally related dienes and dienophiles. *Eur J Org Chem* 2012, **2012**:1186–1194.
  56. Kahn SD, Hehre WJ: Modeling chemical reactivity. 5. Facial selectivity in Diels-Alder cycloadditions. *J Am Chem Soc* 1987, **109**:663–666.
  57. Roy K, PP R: On Some aspects of variable selection for partial least squares regression models. *QSAR Comb Sci* 2008, **27**:302–313.
  58. Roy PP, Roy K: Comparative chemometric modeling of cytochrome 3A4 inhibitory activity of structurally diverse compounds using stepwise MLR, FA-MLR, PLS, GFA PLS and ANN techniques. *G/Eur J Med Chem* 2009, **44**:2913–2922.
  59. Roy K, Chakraborty P, Mitra I, Ojha PK, Kar S: Das RN some case studies on application of “rm2” metrics for judging quality of quantitative structure-activity relationship predictions: emphasis on scaling of response data. *J Comput Chem* 2013, **34**:1071–1082.
  60. Nandi S, Bagchi MC: In silico design of potent EGFR kinase inhibitors using combinatorial libraries. *Mol Simul* 2011, **37**:196–209.

61. Minovski N, Zuperl S, Drgan V, Novič M: **Assessment of applicability domain for multivariate counter-propagation artificial neural network predictive models by minimum Euclidean distance space analysis: A case study.** *Anal Chim Acta* 2013, **759**:28–42.
62. Netzeva TI, Worth AP, Aldenberg T, Benigni R, Cronin MTD, Gramatica P, Jaworska JS, Kahn S, Klopman G, Marchant CA, Myatt G, Nikolova-Jeliazkova N, Patlewicz GY, Perkins R, Roberts DW, Schultz TW, Stanton DT, van de Sandt JM, Tong W, Veith G, Yang C: **Current status of methods for defining the applicability domain of (Quantitative) structure-activity relationships.** *ATLA* 2005, **33**:155–173.
63. Eriksson L, Jaworska J, Worth AP, Cronin MTD, McDowell RM, Gramatica P: **Methods for reliability and uncertainty assessment and for applicability evaluations of classification- and regression-baes QSARs.** *Environ. Health Persp.* 2003, **111**:1361–1375.
64. Bader RFW: **Definition of molecular structure: by choice or by appeal to observation?** *J Phys Chem A* 2010, **114**:7431–7444.
65. Kržan A, Mavri J: **Atomic volume as a descriptor for carbon electronic structure and stability.** *Org Chem* 2011, **76**:1891–1893.
66. Mills N: **ChemDraw Ultra 10.0.** *J. Am. Chem. Soc* 2006, **128**:13649–13650.
67. Tomasi J, Mennucci B, Cammi R: **Quantum mechanical continuum solvation models.** *Chem Rev* 2005, **105**:2999–3093.
68. Katritzky AR, Petrukhin R, Tatham D, Basak S, Benfenati E, Karelson M, Maran U: **Interpretation of quantitative structure–property and–activity relationships.** *J Chem Inf Comput Sci* 2001, **41**:679–685.
69. *Challenges and Advances in Computational Chemistry and Physics.* Springer, Dordrecht: Puzyn T, Leszczynski J, Cronin MTD; 2009.
70. Minitab® Statistical Software: *Minitab.* ; 2010. [www.minitab.com](http://www.minitab.com) (accessed October 29th 2013).
71. Zupan J, Gasteiger J: *Neural Networks for Chemists: an introduction.* Weinheim: VCH; 1993.

doi:10.1186/1752-153X-7-171

**Cite this article as:** Nandi *et al.*: Quantitative structure-activation barrier relationship modeling for Diels-Alder ligations utilizing quantum chemical structural descriptors. *Chemistry Central Journal* 2013 **7**:171.

Publish with **ChemistryCentral** and every scientist can read your work free of charge

*“Open access provides opportunities to our colleagues in other parts of the globe, by allowing anyone to view the content free of charge.”*

W. Jeffery Hurst, The Hershey Company.

- available free of charge to the entire scientific community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:  
<http://www.chemistrycentral.com/manuscript/>



**ChemistryCentral**