

Alignment-free Descriptors for Quantitative Structure-Rate Constants Relationships of Diels-Alder Reactions

András P. Borosy,^{*1} Balázs Balogh,² and Péter Mátyus²

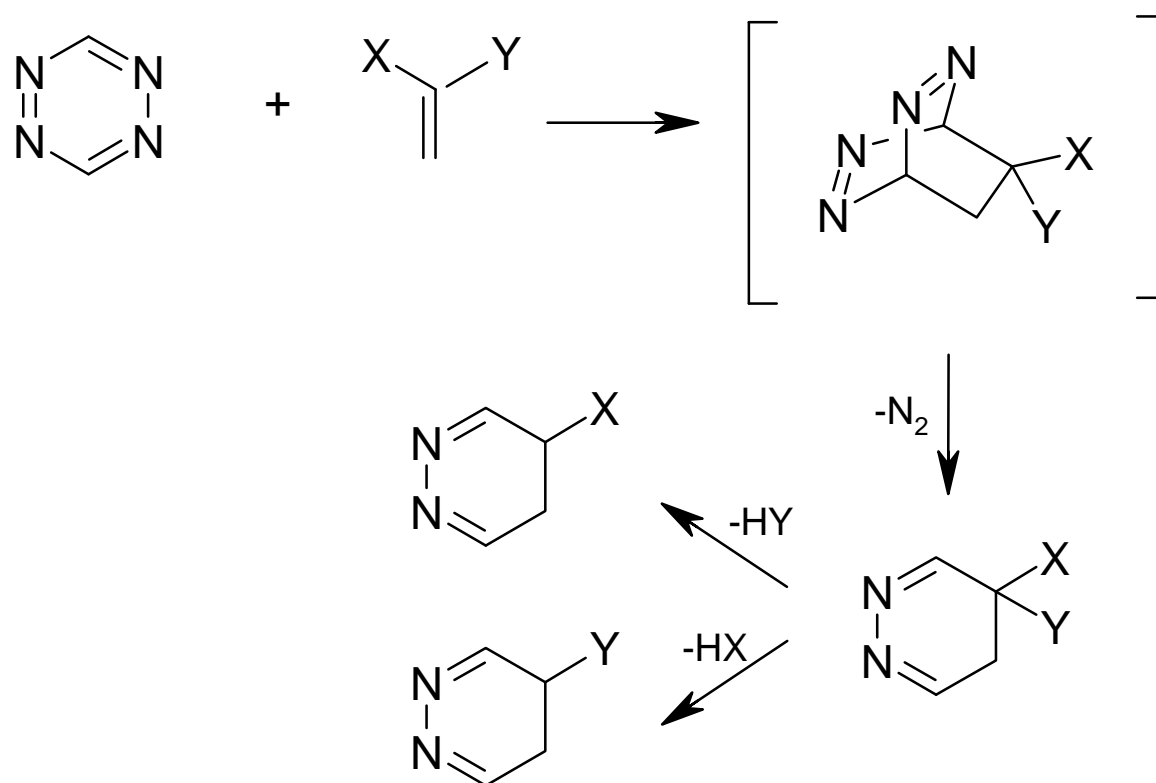
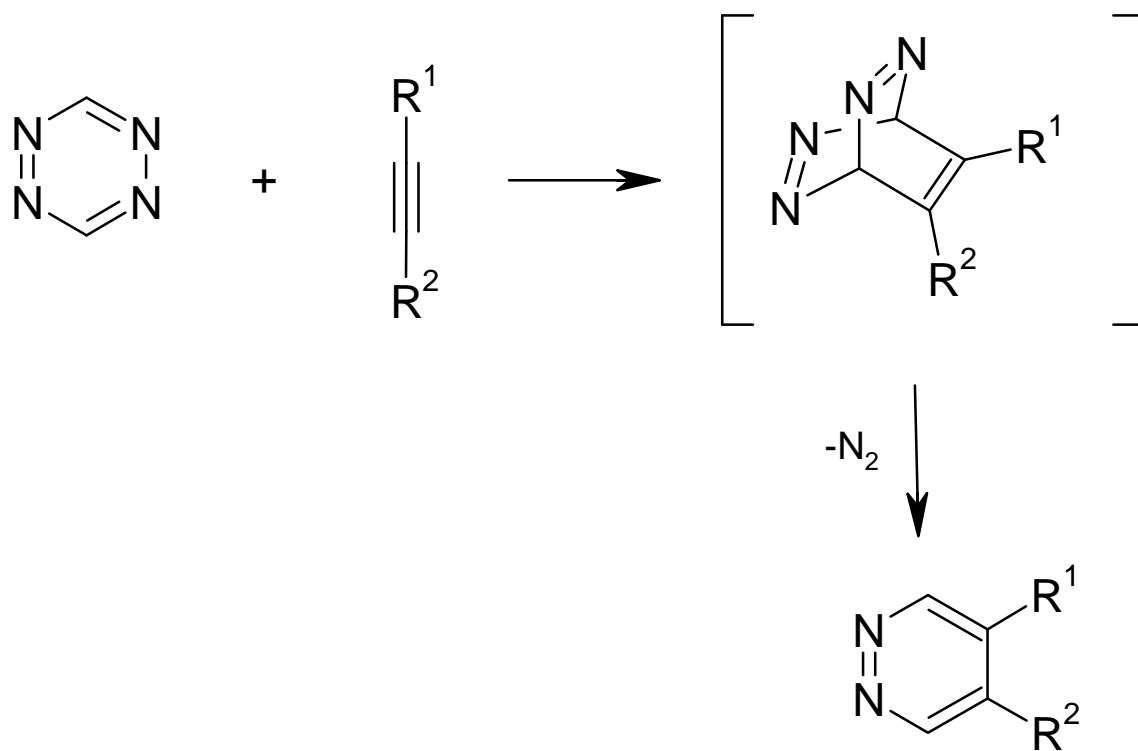
¹*IVAX Drug Research Ltd., 1325 Budapest, POB 82, Hungary*

²*Institute of Organic Chemistry, Semmelweis University of Medicine, 1092 Budapest, Hőgyes u. 7, Hungary*

^{*}To whom correspondence should be addressed. e-mail: borosy@eik.bme.hu, He will join the *Basilea Pharmaceutica ltd* (POB 3255, 4070 Basel, Switzerland, www.basileapharma.com) who financed his participation.

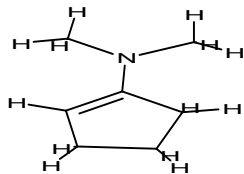
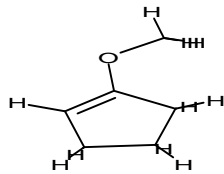
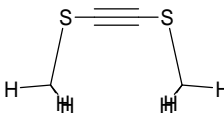
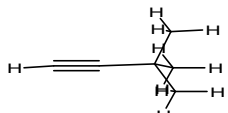
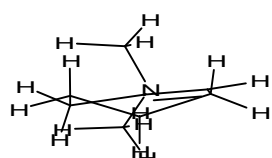
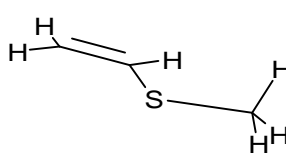
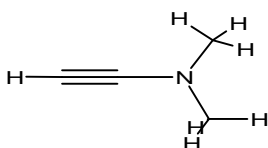
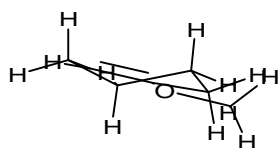
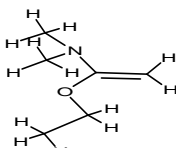
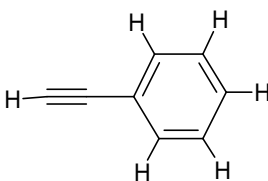
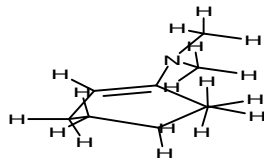
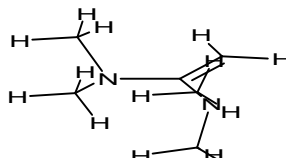
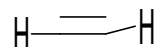
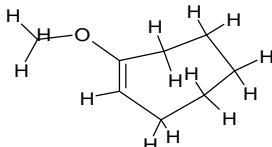
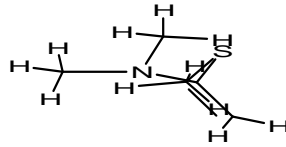
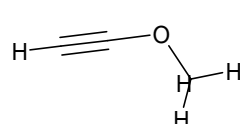
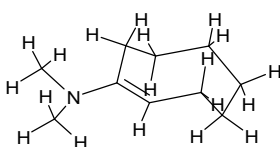
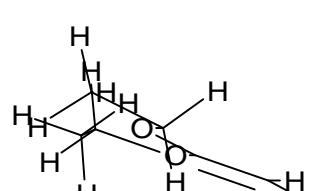
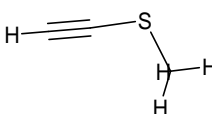
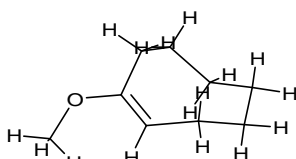
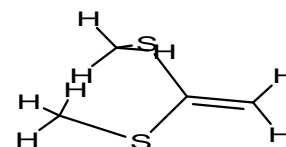
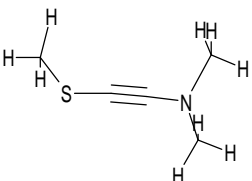
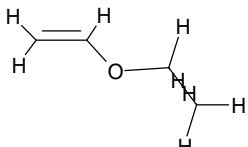
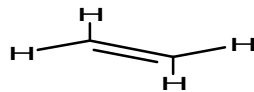
INTRODUCTION

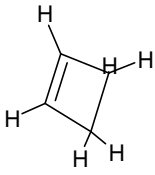
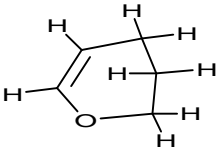
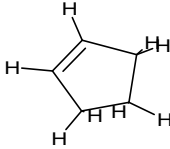
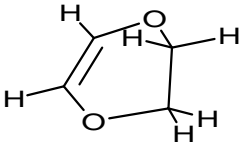
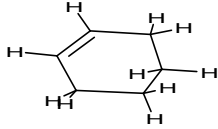
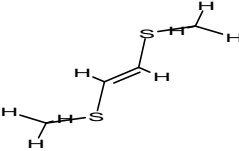
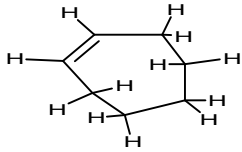
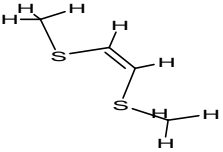
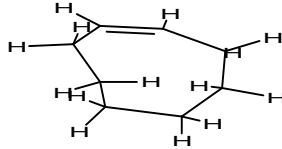
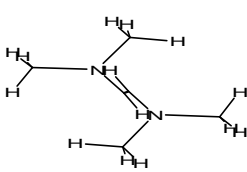
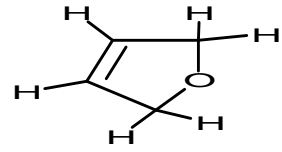
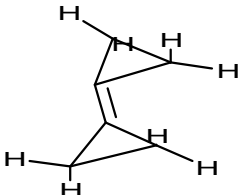
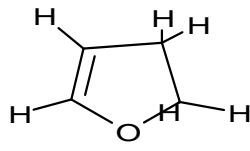
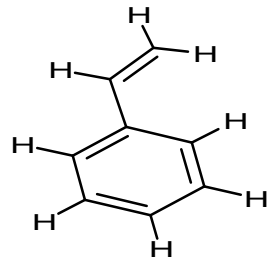
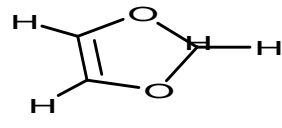
Full details on the reactivity of 1,2,4,5-tetrazine as 4π component in inverse-type *Diels-Alder reactions*, including kinetic data, are reported [1]. Donor-substituted alkynes, alkenes, donor-substituted and unsubstituted cycloalkenes, ketene acetals and amins, as well as several cyclic enol ethers were used as dienophiles in these investigations. A number of 4-mono- and 4,5-disubstituted pyridazines can easily be obtained by this method. As shown for a large number of open-chain and cyclic dienophiles, 1,2,4,5-tetrazine can be used as an electron-poor diene in inverse-type Diels-Alder reactions to yield a great variety of pyridazine derivatives not easily accessible by other synthetic methods:



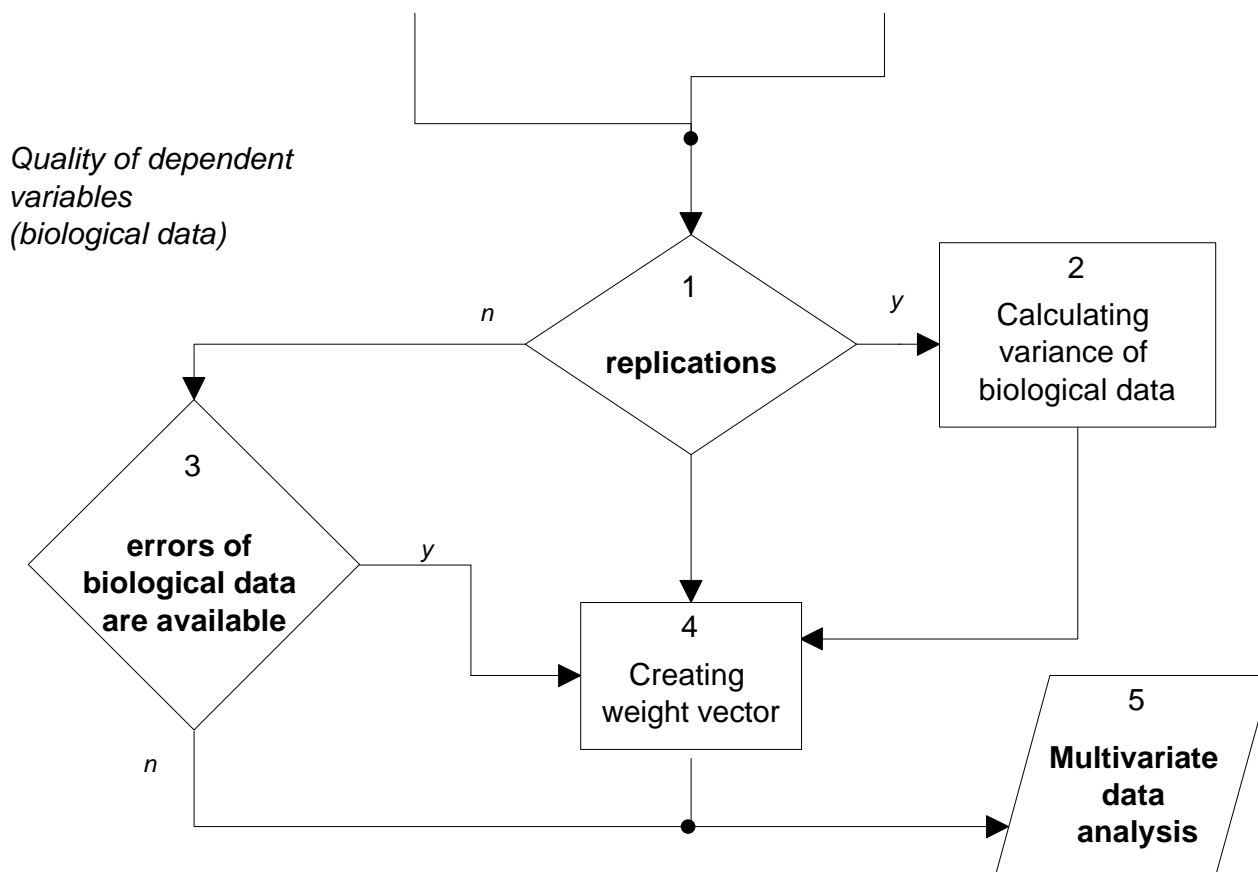
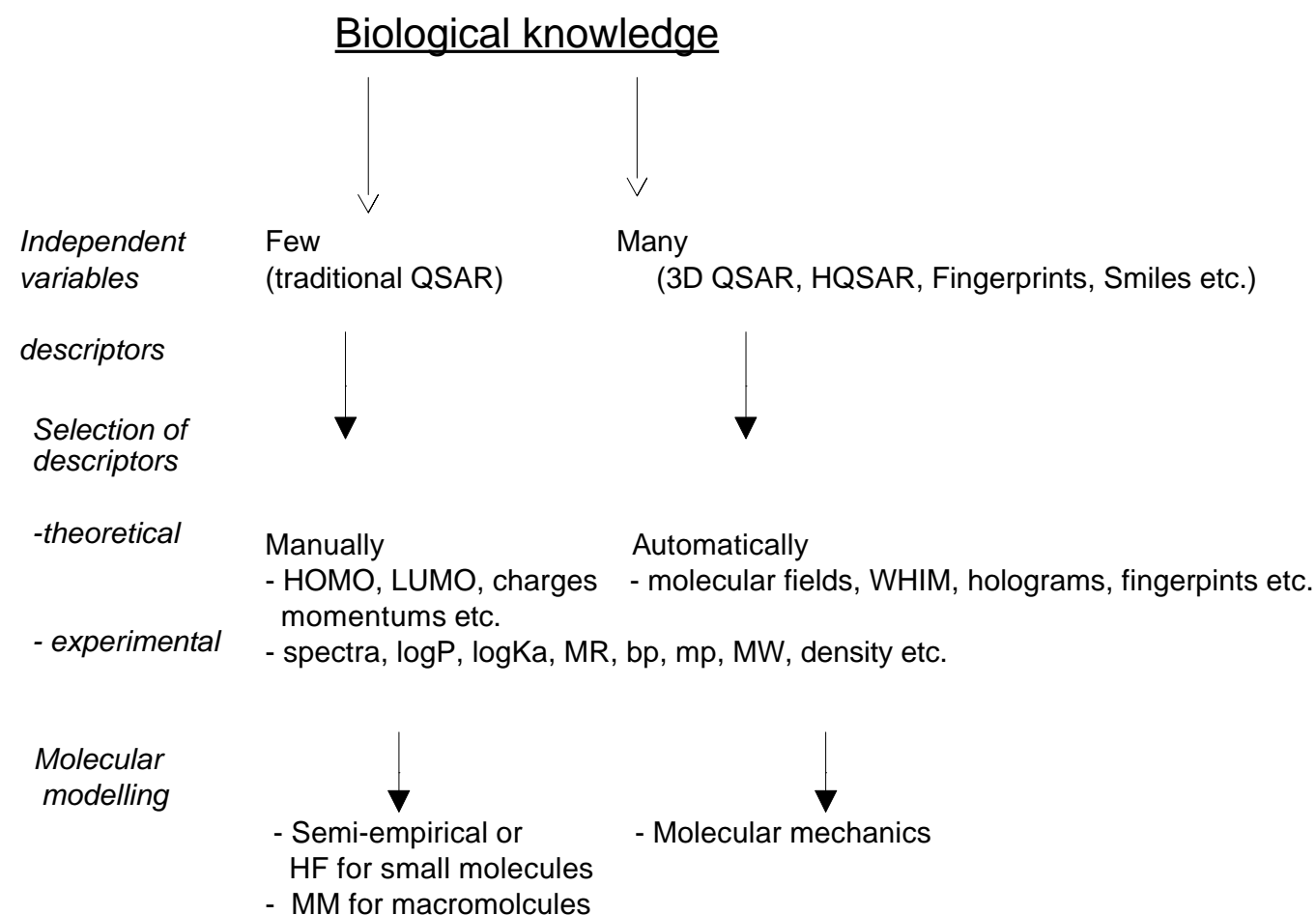
The rate constants obtained by extensive kinetic measurements lead to quantitative rules for the influence of steric and electronic substituent effects in the dienophile. The results obtained using 1,2,4,5-tetrazine as diene can, in principle, be applied to other tetrazines. **To plan organic synthesis reliably, the understanding and prediction of the rate of these processes are of worthwhile goals**

Table 1. Dienophiles

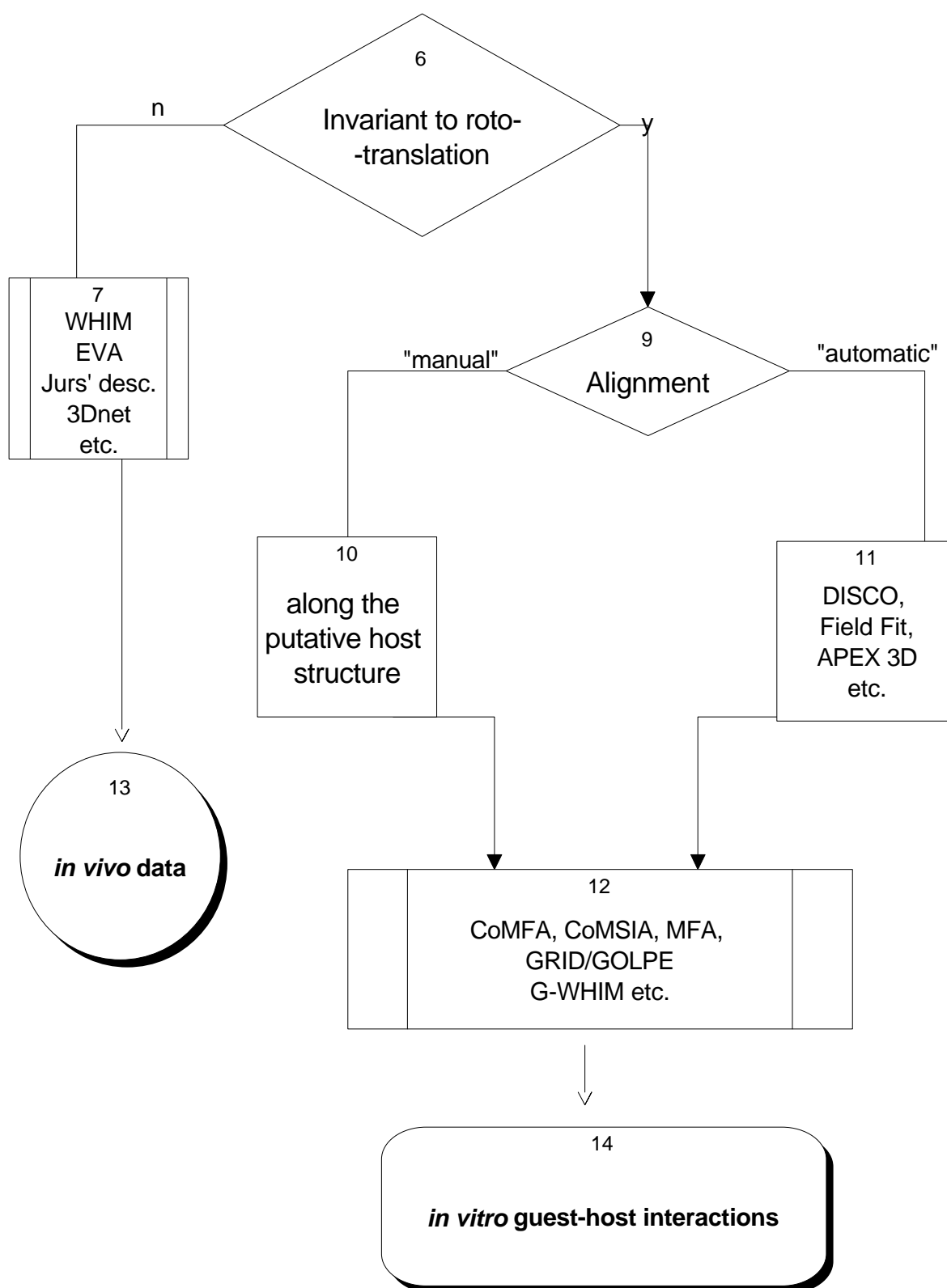
ID	<i>p</i> k	STRUCTURE	ID	<i>p</i> k	STRUCTURE	ID	<i>p</i> k	STRUCTURE
1	-6		10	-0.54		9	1.74	
2	0.77		11	-4.33		18	-1.16	
3	-4.7		12	1.77		19	-6.13	
4	0.77		13	-4.83		20	-5.83	
5	-0.46		14	-1.08		21	-4.79	
6	0.16		15	-6.33		22	-2.08	
7	-0.17		16	-1.1		23	0.61	
8	-3.34		17	-1.91		24	-3.33	

25	-3.55		33	-0.24	
26	-2.44		34	1.01	
27	-0.44		35	1.78	
28	-2.56		36	1.66	
29	-2.76		37	-3.6	
30	-1.38		38	-2.67	
31	-2.64		39	-1.85	
32	-1.86				

Multivariate modelling in QSAR-QSPR



3D QSAR-QSPR



DESCRIPTORS, VARIABLES

Dependent variable:

- Negative logarithm of rate constant (pk)

Independent variables:

- Sybyl's EVA descriptors [2]
- Descriptors of 3Dnet [3]

MODEL BUILDING METHODS

- Partial Least Squares (PLS) regression for EVA and
- Artificial Neural Network (ANN) regression for 3Dnet descriptors (variable selection was performed by ANN, too).
- For comparison the percent relative prediction error (Rel. Err.) were used as a measure of predictive ability[13]:

$$Rel.Err.\% = 100 * \frac{\hat{y}_i - y_i}{y_i} \quad (2)$$

where n equals to the number of objects in the training or in the prediction sets, y_i and \hat{y}_i refer to the measured and estimated values of the property for the object i , respectively. This measure were used as a least - worse alternative being the real degrees of freedom for PLS and ANN are not known [14].

The validation of results was done by using a prediction set (7 elements). First Hierarchical Cluster Analysis (HLC) of Sybyl was performed on EVA descriptors, and then the prediction set was chosen by experts including all clusters and covering a wide range of descriptors and including extremes, as well as.

Sybyl's EVA Descriptors

The EVA descriptor [4-10] is derived from fundamental IR and Raman range molecular vibrational frequencies. EVA is sensitive to 3-D structure, but has an advantage over field-based 3-D QSAR methods inasmuch as it is invariant to both translation and rotation of the structures concerned and thus structural superposition is not required pharmacophore. This reduced sensitivity is a consequence of the use of a *Gaussian* smearing function to develop the descriptor (as described below) and as a result EVA might be described as a '2.5-D' descriptor.

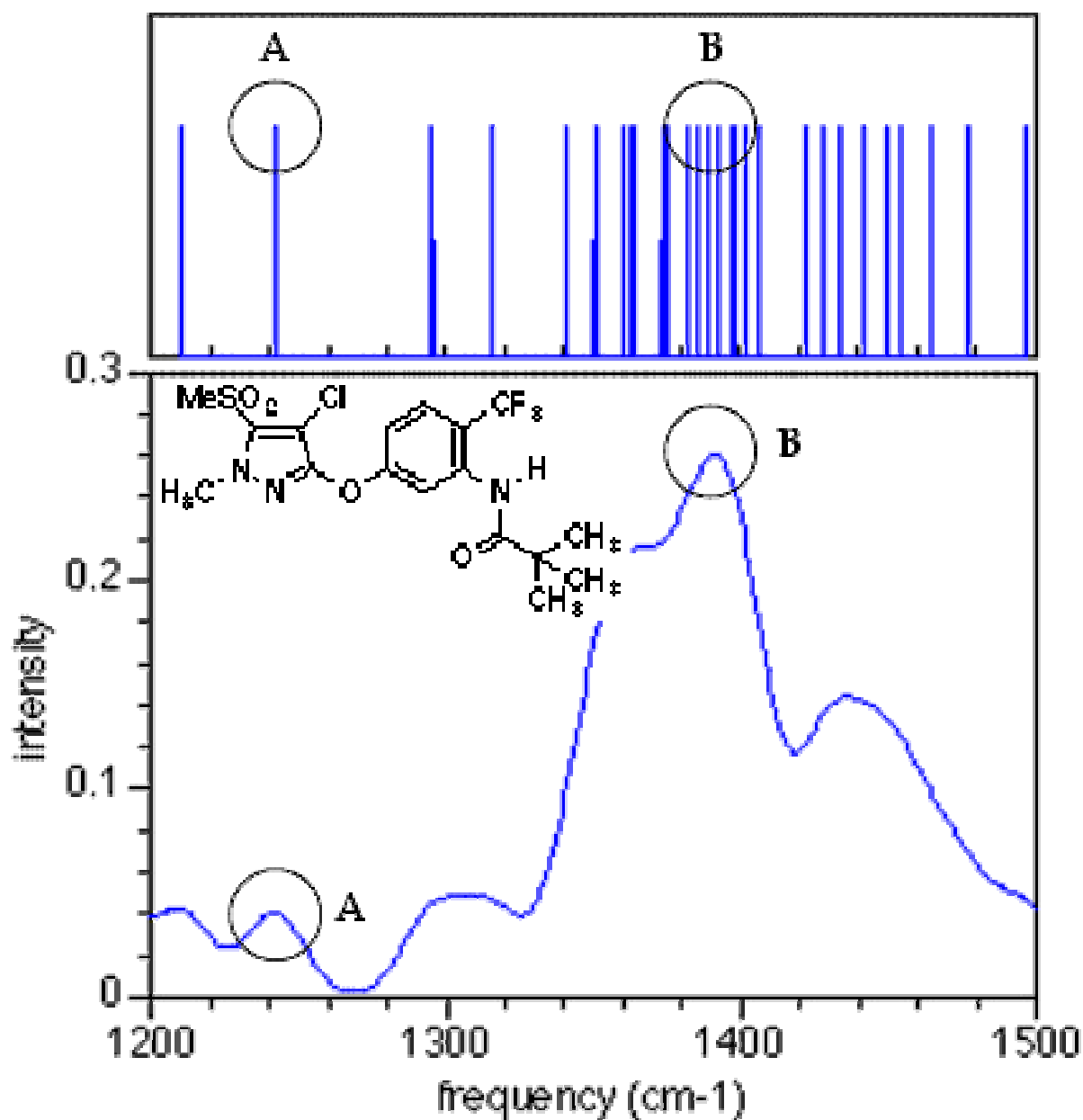
Once determined, from whatever source, the set of vibrational wave numbers (vwn) for a given structure is projected onto a linear bounded frequency scale (BFS) typically covering a range from 1-4 000 cm^{-1} . The use of this range means that all fundamental vibrational normal modes are included in the analysis, should a vwn exceed 4 000 cm^{-1} then either the BFS can be extended or all vwns from all molecules can be scaled according to scale factors. Next, a Gaussian kernel of fixed standard deviation (σ) is placed over each and every frequency value. The BFS is then sampled at fixed increments of $L \text{ cm}^{-1}$ and the value of the resulting EVA descriptor, $EVA(x)$, at each sample point, x , is the sum of the amplitudes of the overlaid kernels at that point:

$$EVA(x) = \sum_{i=1}^{3N-6} \frac{1}{s\sqrt{2p}} \exp[(x - f_i)/2s^2] \quad (1)$$

where f_i is the i th normal mode frequency of the compound concerned. This procedure is repeated for each dataset compound and then combined to provide a matrix with M rows (compounds) and $4000/L$ (columns) descriptor variables. Typically, a descriptor set has been derived using a σ of 10 cm^{-1} and an L of 5 cm^{-1} giving 800 descriptor variables. Thus, for a QSAR dataset of typical size the number of variables is very much larger than M and a method such as **Partial Least Squares (PLS) regression** in conjunction with cross-validation is required to provide a robust regression analysis. As such, the results obtained with EVA QSAR are usually dependent upon the chosen kernel width (σ) since this parameter determines whether or not, and the extent to which, proximal kernels overlap.

Consider, for example, the pyrazole trifluorotoluidide ether [11] shown below. The molecule's 136 normal modes were calculated in MOPAC using AM1; those falling between 1200 and 1500 cm^{-1} are shown as the line spectrum at the top. The bottom

of the figure shows the profile obtained by summing across all 136 Gaussians at each frequency in this range.



Small peaks in the profile arise from relatively isolated normal modes, such as the one at 1242 cm^{-1} (peak A). Large peaks result when several normal modes are separated by less than 2σ from each other, as occurs near 1390 cm^{-1} (peak B).

The input descriptors used in 3Dnet calculations [12]:

Molecular Mass

Dipole Moment

Polarizability

Degree of Freedom

Double Bonds Equivalent

Lipophilicity

Hildebrand Solubility Parameter

Electrostatic Acidity & Basicity

Electrostatic Total Acidity & Basicity

Wiener, Randic, Bodor Indices

Minimum, maximum average of ElectroStatic Potential (ESP)

Minimum, maximum average of Molecular Lipophilicity Potential (MLP)

WHIM-descriptors (Moments of spatial distribution of atomic values)

- Mass
- Position
- Van der Waals
- Electronegativity
- Polarizability
- Electrotopological Index
- Lipophilicity

Autocorrelation Histograms

- Local Charge
- Van der Waals Volume
- Lipophilicity contribution
- Polarizability contribution
- pI function
- Electrotopological Index

Quantum chemical descriptors:

- AM1 heat of formation
- AM1 HOMO & LUMO
- 6-31 G* energy
- 6-31 G* HOMO, HOMO-1 , HOMO-2
- 6-31 G* LUMO, LUMO+1, LUMO+2

Results

Variables selected by 3Dnet:

SORTED SIGNIFICANCE OF STANDALONE_INPUTS [%]

GRAVIT_INDEX	100.00
MIN_OF_MLP	94.16
GLOBULARITY	87.00
AVR_OF_ESP	73.09
K_VDW (WHIM)	64.59
AVR_OF_MLP	52.85
K_LIPO (WHIM)	52.33
K_ETPI (WHIM)	45.13
MIN_OF_ESP	42.21
MAX_OF_MLP	34.96
QTOT_BODOR	33.16
K_EN (WHIM)	30.56
HOMO-1	26.83
K_MASS (WHIM)	25.59
QN_BODOR	25.33
DOUBLE_BOND_EQUIV	19.89
MAX_OF_ESP	16.17
DIPOLE_MOMENT	11.35
LIPOPHILICITY	7.96

X,Y graphics for

X=GRAVIT_INDEX

and

Y=PK(exp.)

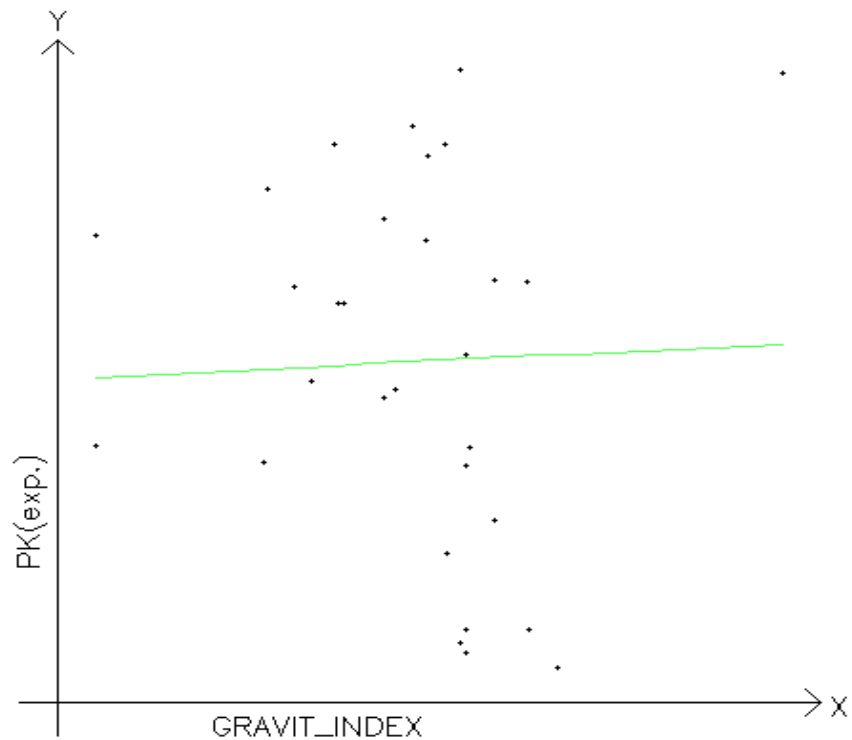
X data range :

11.0097 to 40.1861

Y data range :

-6.3304 to 1.7747

The curve displays the network's response to the change of the X input while keeping the other inputs at their average value.



(Press P to write PCX file or any other key to continue.)

X,Y graphics for

X=GLOBULARITY

and

Y=PK(exp.)

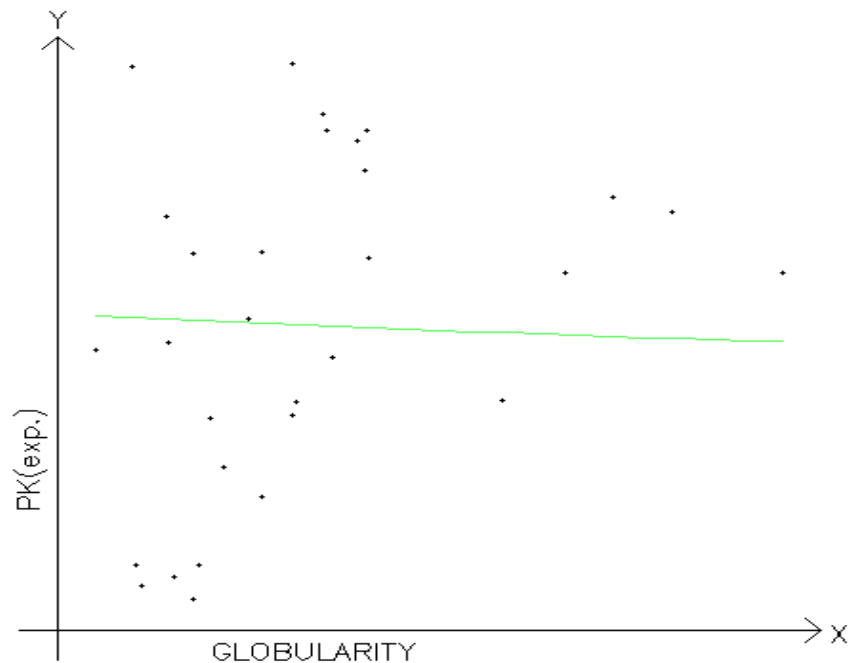
X data range :

0.6190 to 0.9797

Y data range :

-6.3304 to 1.7747

The curve displays the network's response to the change of the X input while keeping the other inputs at their average value.



(Press P to write PCX file or any other key to continue.)

X,Y graphics for

X=AVR_OF_ESP

and

Y=PK(exp.)

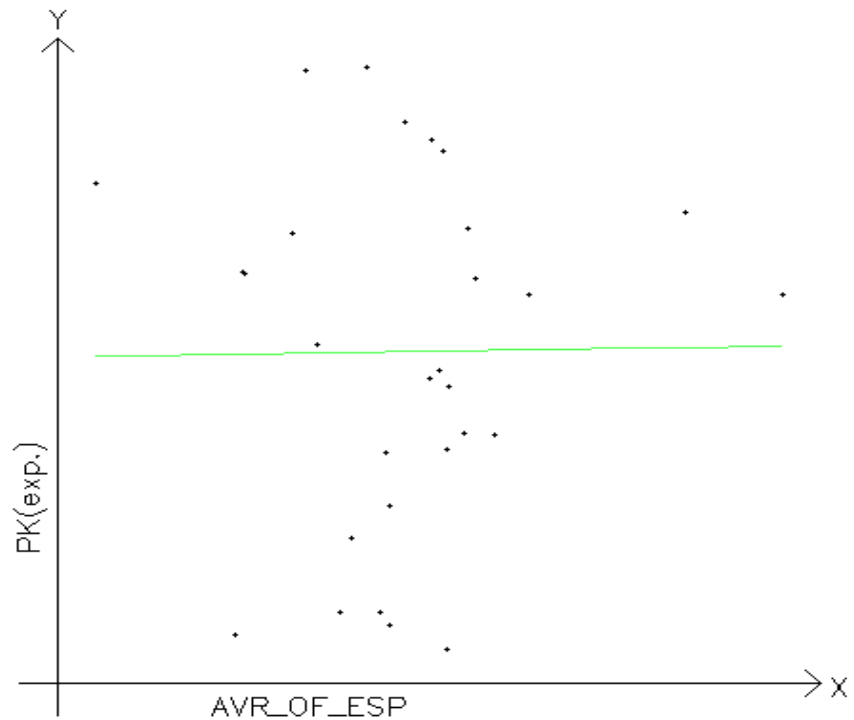
X data range :

-4.3205 to 3.3469

Y data range :

-6.3304 to 1.7747

The curve displays the network's response to the change of the X input while keeping the other inputs at their average value.



(Press P to write PCX file or any other key to continue.)

X,Y graphics for

X=AVR_OF_MLP

and

Y=PK(exp.)

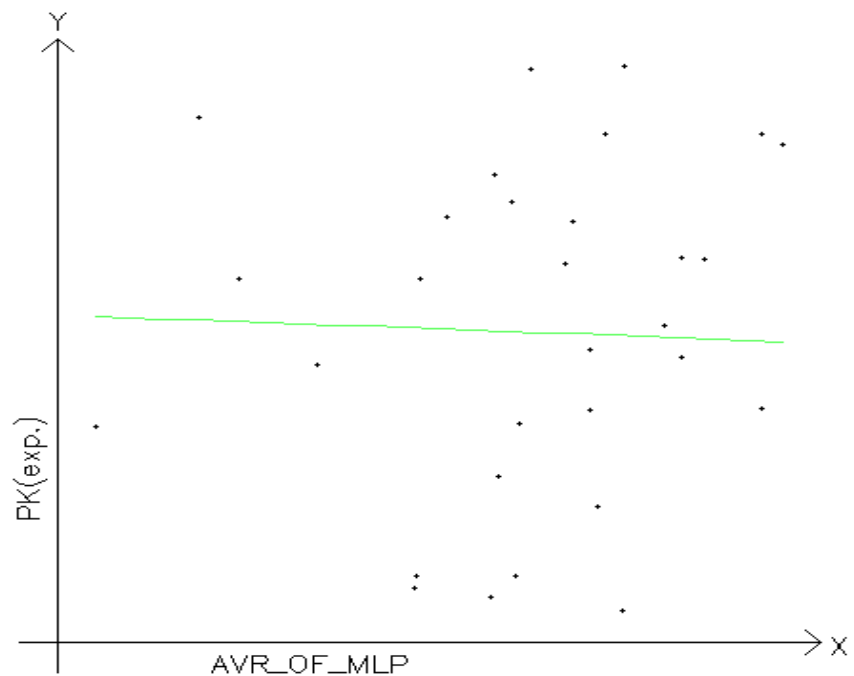
X data range :

-0.0744 to 0.5168

Y data range :

-6.3304 to 1.7747

The curve displays the network's response to the change of the X input while keeping the other inputs at their average value.



(Press P to write PCX file or any other key to continue.)

X,Y graphics for

X=MIN_OF_ESP

and

Y=PK(exp.)

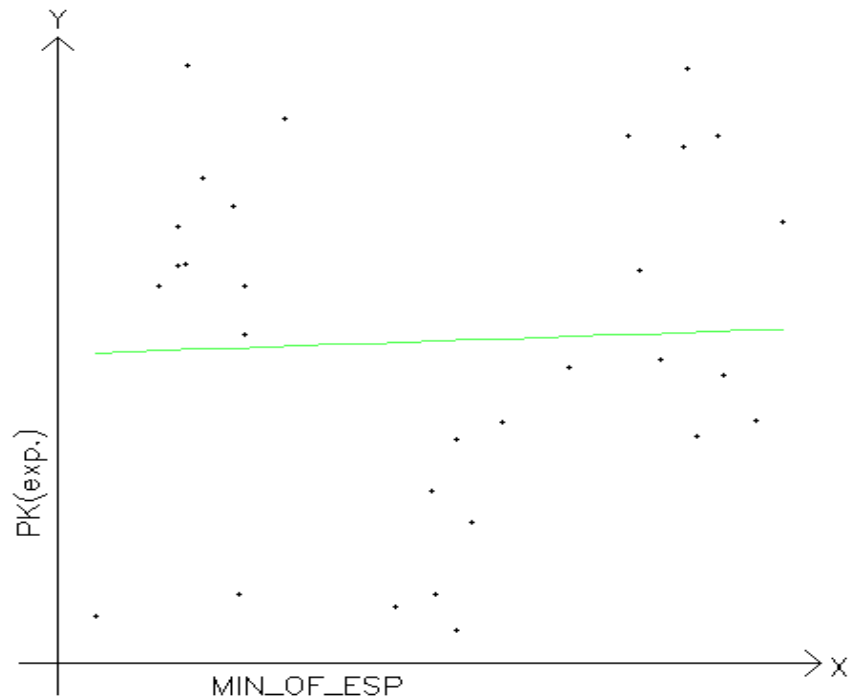
X data range :

-25.9153 to -1.9752

Y data range :

-6.3304 to 1.7747

The curve displays the network's response to the change of the X input while keeping the other inputs at their average value.



(Press P to write PCX file or any other key to continue.)

X,Y graphics for

X=EXPL_HOMO_1

and

Y=PK(exp.)

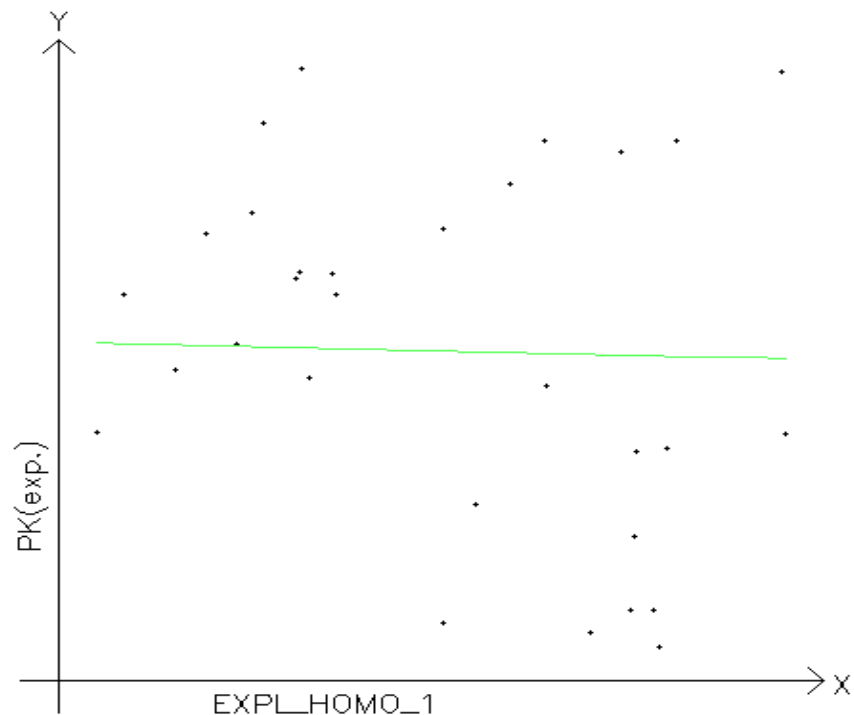
X data range :

-0.4945 to -0.3109

Y data range :

-6.3304 to 1.7747

The curve displays the network's response to the change of the X input while keeping the other inputs at their average value.



(Press P to write PCX file or any other key to continue.)

Regression

Training set

ID	pk_exp	pk_3Dnet	Rel.Err.	pk_EVA	Rel.Err
1	-6.000	-5.433	0.095	-5.901	0.017
2	0.772	0.500	0.352	0.78	0.010
4	0.775	1.392	0.796	0.915	0.181
5	-0.464	-0.398	0.142	-0.483	0.041
6	0.163	0.156	0.042	0.32	0.963
8	-3.337	-3.488	0.045	-3.136	0.060
9	1.745	1.388	0.204	1.737	0.004
10	-0.537	-0.418	0.221	-0.625	0.165
11	-4.335	-5.412	0.248	-4.091	0.056
12	1.775	1.276	0.281	1.806	0.018
14	-1.076	-0.718	0.332	-0.944	0.122
15	-6.330	-5.335	0.157	-6.269	0.010
16	-1.100	-2.256	1.051	-1.105	0.004
18	-1.158	-1.296	0.119	-1.249	0.078
19	-6.134	-5.397	0.120	-6.139	0.001
20	-5.833	-5.551	0.048	-5.872	0.007
21	-4.788	-4.798	0.002	-4.731	0.012
22	-2.079	-1.715	0.175	-2.129	0.024
23	0.607	0.667	0.098	0.509	0.162
24	-3.332	-2.508	0.247	-2.753	0.174
25	-3.545	-3.442	0.029	-3.296	0.070
26	-2.441	-2.476	0.015	-2.407	0.014
28	-2.555	-2.456	0.039	-2.417	0.054
30	-1.382	-1.532	0.109	-1.449	0.048
33	-0.243	0.068	1.281	-0.061	0.749
34	1.013	0.958	0.055	0.917	0.095
36	-3.599	-5.548	0.542	-3.532	0.019
37	-2.669	-2.605	0.024	-2.697	0.010
Rel.Err. %	25			11	

Prediction set

ID	pk_exp	pk_3DNet	Rel.Err.	pk_EVA	Rel.Err
3	-4.701	-5.311	0.130	-3.948	0.160
7	-0.167	-0.303	0.808	-0.191	0.142
17	-1.905	-2.467	0.295	-1.766	0.073
29	-2.760	-2.449	0.113	-2.157	0.218
32	-1.863	-1.894	0.017	-2.059	0.105
35	1.656	1.310	0.209	1.031	0.377
38	-1.848	-1.338	0.276	-0.813	0.560
Rel.Err. %	26.4			23.4	

Optimal models:

3Dnet

19 input, 4 hidden, 1 output nodes

EVA

$L=20\text{ cm}^{-1}$, $\sigma=20\text{ cm}^{-1}$, 11 latent variables

Kinetic Results

[4+2] Cycloadditions of 1,2,4,5-tetrazines are $\text{LUMO}_{\text{diene}}\text{-HOMO}_{\text{phil}}$ -controlled reactions. [1] For a particular diene, the reactivity change of the system parallels the HOMO energy of the dienophiles. Donor substituents raise the HOMO energy of the dienophiles and by decreasing the $\text{LUMO}_{\text{diene}}\text{-HOMO}_{\text{phil}}$ gap, increase the rate constants of the cycloaddition step. In principle, any exchange of hydrogen in the dienophile for a larger substituent has an impeding steric effect. So, as a net result, the substitution of hydrogen by a substituent in the dienophile component, depending on its electron-donating power, can lead to an increase or decrease of the dienophile's reactivity.

All cycloadditions studied kinetically in this contribution cleanly follow a second-order rate law between less than 10 % and mostly more than 90% conversion. The second-order rate constants illustrate the tremendous influence of structural variations in the dienophile component. As was shown [1] these substituent effects on the cycloaddition rate of are typical for other tetrazine cycloadditions.

Discussion of 3Dnet regression:

Important variables depends on

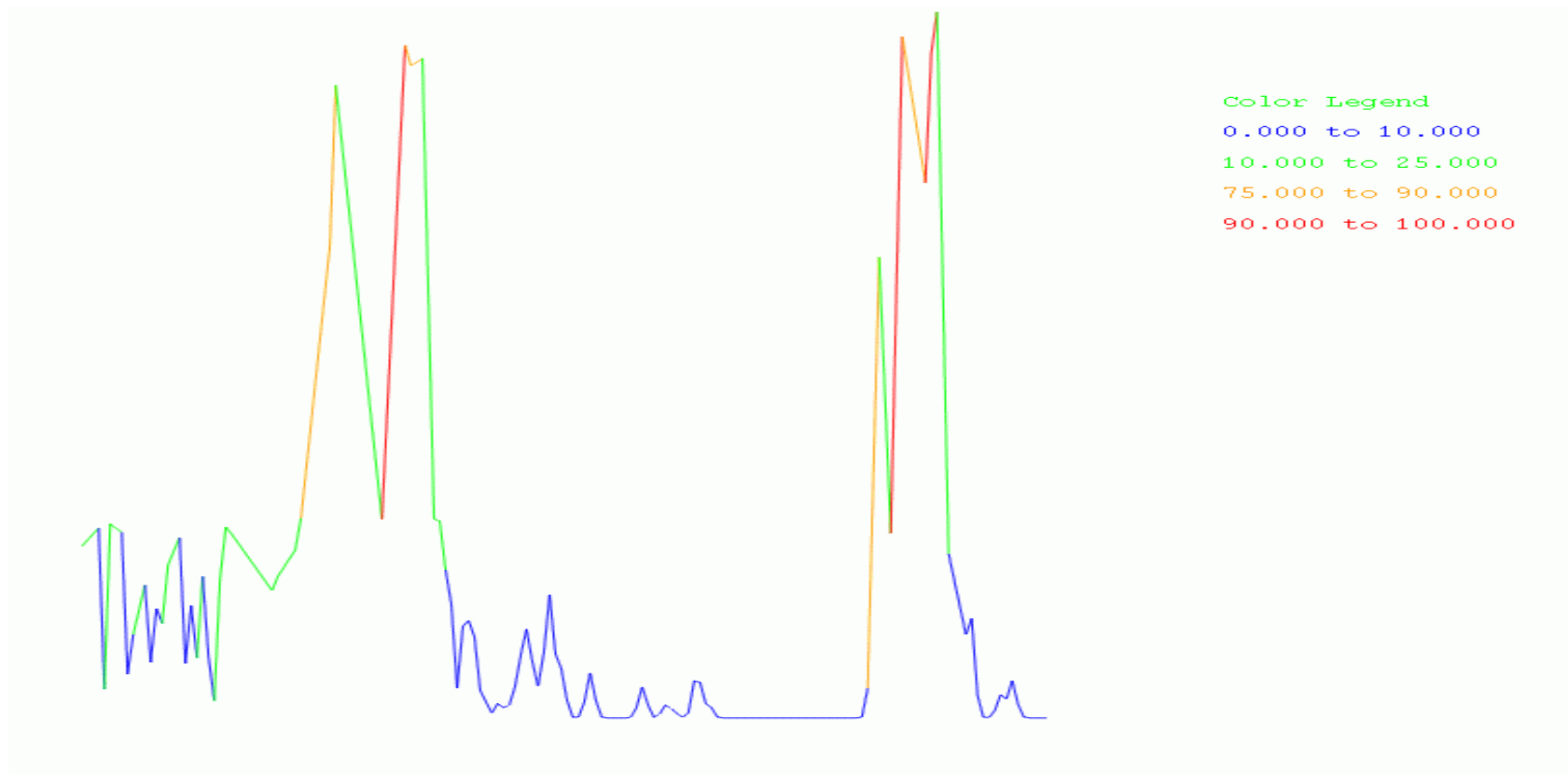
- the non-directional mass and size
- minimal, maximal and average values of molecular electrostatic as well as lipophilic potenciales
- HOMO-1

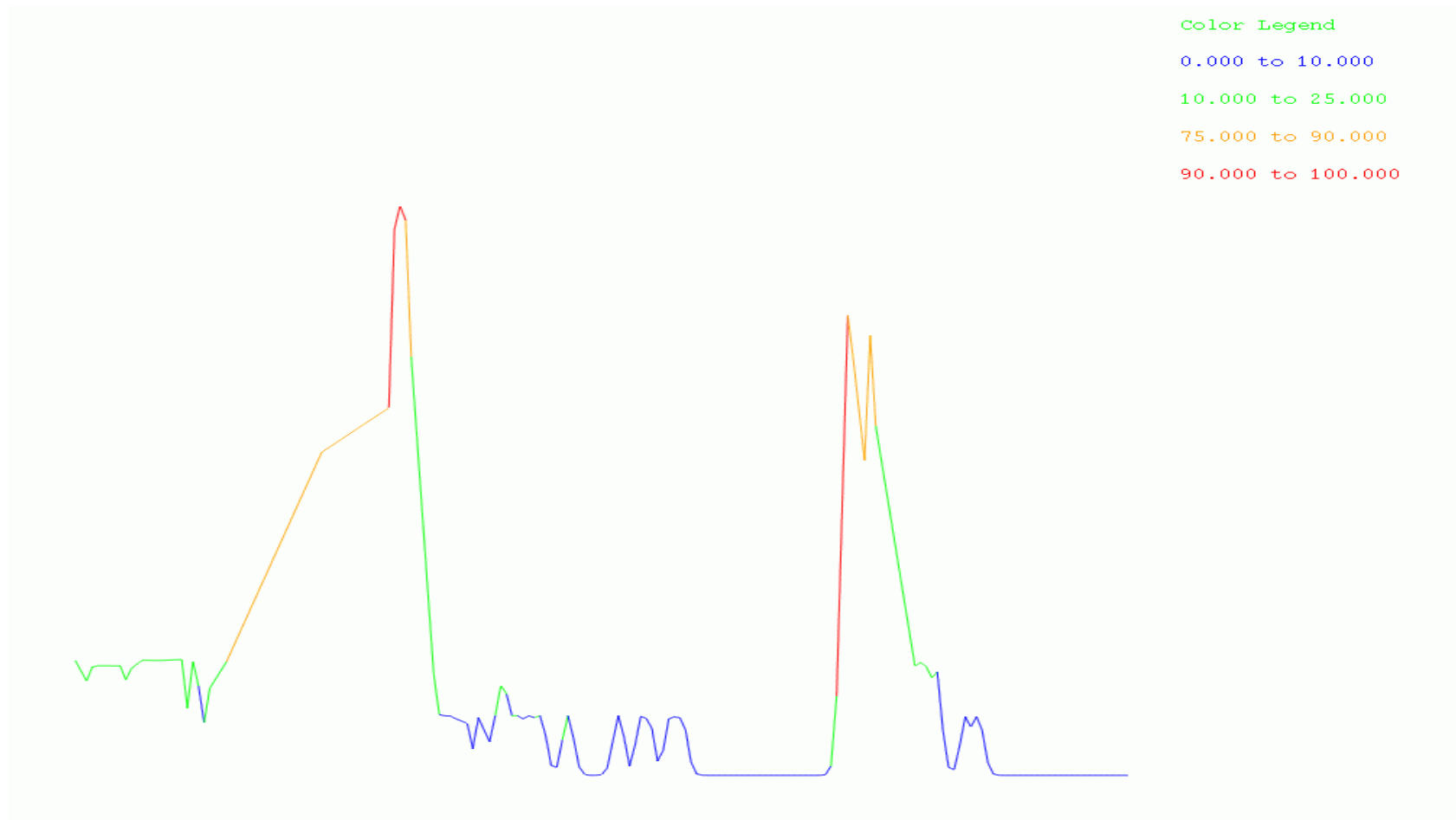
which correlate the experimental observations [1]

Unfortunately experimental errors of the rate constants are not know. Model might be improved by split of the data set.

Discussion of EVA regression:

Discriminating power: fraction of the variation in intensity at each frequency which contributes to the model.

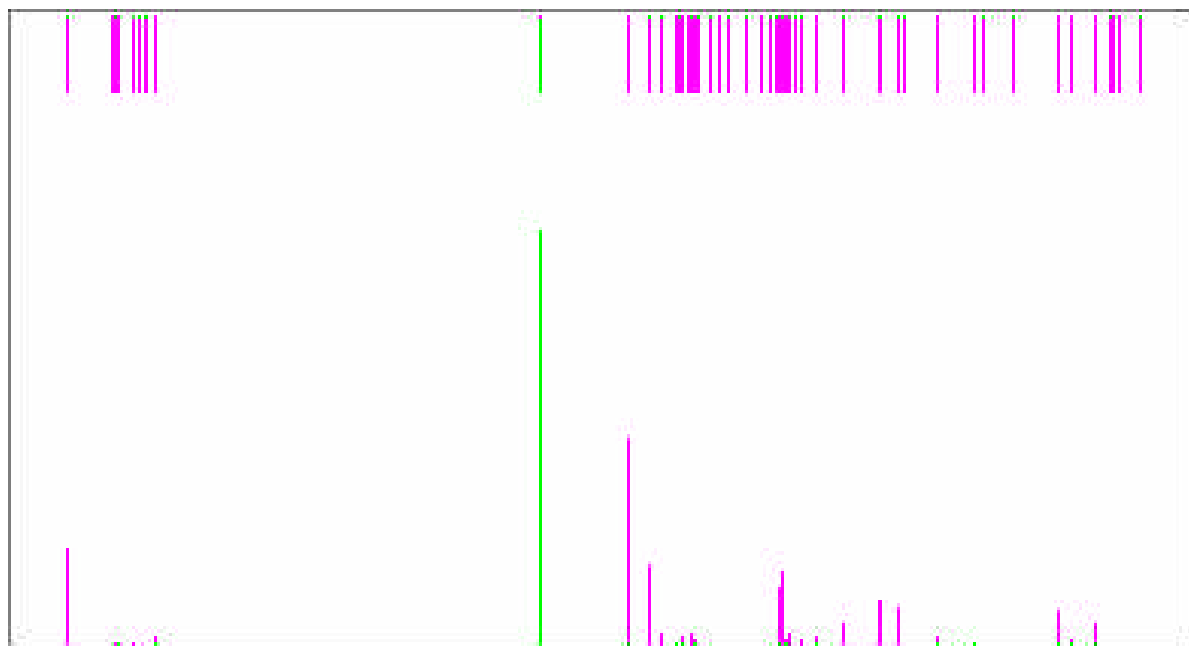




Max field: at each frequency the highest intensity in EVA profiles used to derive the model.

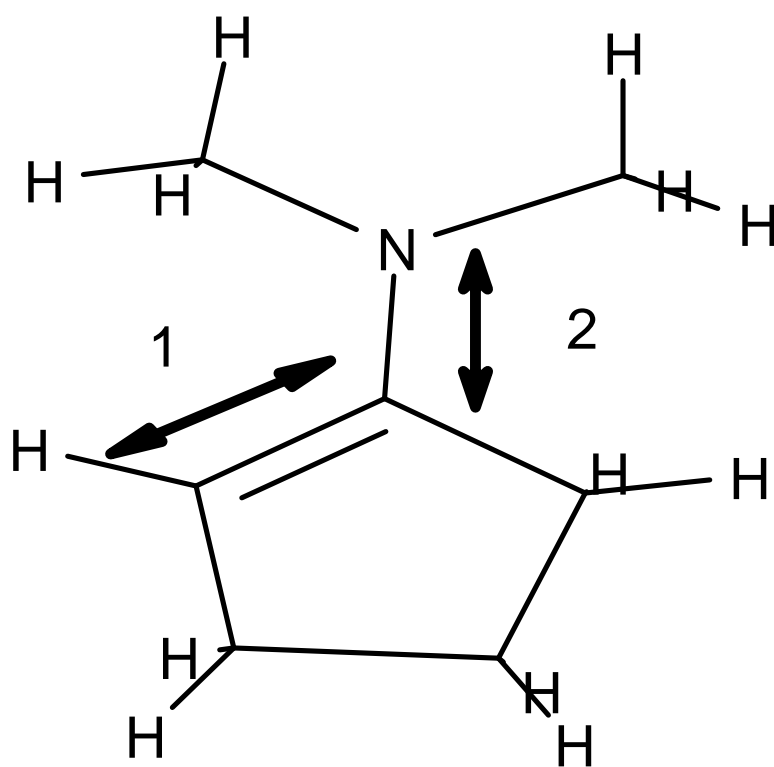
Most discriminating frequencies show that the stretching vibrations of double bonds and the single bonds attaching to this are the most important ones.

Dienophile **1**



Vibrations

1 2



Conclusions

As shown for a large number of open-chain and cyclic dienophiles, 1,2,4,5-tetrazine [1] can be used as an electron-poor diene in inverse-type Diels-Alder reactions to yield a great variety of pyridazine derivatives not easily accessible by other synthetic methods.

The rate constants obtained by extensive kinetic measurements lead to quantitative rules for the influence of steric and electronic substituent effects in the dienophile.

The EVA and 3Dnet descriptors configurations, are comparable in prediction Rate effects can hence be easily predicted for synthetic purposes in the tetrazine field.

Although experimental IR spectra might not be optimal for QSAR [15] but EVA descriptors seem to be as applicable as anything else

References:

- [1]J. Sauer et al. *Eur. J. Org.* 1998, 2885-2896.
- [2]Sybyl, Version 6.4, Tripos Inc. USA.
- [3]3Dnet 1.0, Budapest, Hungary, Vichem Ltd, 2001. ikovesdi@eqnet.hu
- [4]A. M. Ferguson, T.W. Heritage, P. Jonathon, S. E. Pack, L. Phillips, J. Rogan, P.J. Snaith, *J. Comp.-Aided Mol. Des.* 11 (1997), 143-152.
- [5]Jonathan P., McCarthy W.V., Roberts A.M.I., *J. Chemom.* 10 (1996), 189-214.
- [6]Turner D.B., Willett P., Ferguson A.M., Heritage T., *J. Comput.-Aided Mol. Des. II* (1997) 409-422.
- [7]Turner D.B., Willett P., Ferguson A.M., Heritage T., *J. Comput.-Aided Mol. Des.* 13 (1999) 271-296.
- [8]Turner D.B., Willett P., *J. Comput.-Aided Mol. Des.* 14 (2000) 1-21.
- [9] Ginn C.M.R., Turner D.B., Willett P., Ferguson A.M., Heritage T. W.,*J. Chem. Inf. Comput. Sci.* 37 (1997) 23-37.
- [10]Turner D.B., Willett P. *Eur. Med. Chem.* 35 (2000), 367-375.
- [11]D. Clark, *J. Agric. Food Chem.*, 44 (1996), 3643-3652.
- [12] M. Karleson, *Molecular Descriptors in QSAR/QSPAR* , Wiley-Interscience, New York, 2000.
- [13]S. Sekulic, M. B. Seasholtz, Z. Wang, B. K. Kowalski, S. E. Lee, and B. R. Holt, *Anal. Chem.*, 65 (1993), 835A.
- [14]H. van der Voet, *J. Chemometrics*, 13 (1999), 195-208.
- [15]R. Benigni, A. Giuliani, L. Passerini, *J. Chem. Inf. Comput. Sci.*, 41 (2001), 727-730.