



Transworld Research Network
37/661 (2), Fort P.O.
Trivandrum-695 023
Kerala, India

Topological Indices for Medicinal Chemistry, Biology, Parasitology, Neurological and
Social Networks, 2010: 53-68 ISBN: 978-81-7895-489-9
Editors: Humberto González-Díaz and Cristian Robert Munteanu

4. QSRR construction of networks for chirality inversion reactions

Sonia Arrasate¹, Nuria Sotomayor¹, Esther Lete¹
and Humberto González-Díaz²

¹*Department of Organic Chemistry II, Faculty of Science and Technology, University of the Basque
Country/Euskal Herriko Unibertsitatea, Apto. 644, 48080 Bilbao, Spain*

²*Department of Microbiology and Parasitology, University of Santiago de Compostela, 15782, Spain*

Abstract. There are many enantioselective reactions of organolithium to imines described with very different substrates, organolithium compounds, chiral ligands, solvents, and specific reaction conditions, such as, temperature, addition times, reaction time, and order of addition of reactants. It implies that we may need to use mathematical tools in order to study this huge amount of information. In this work, we constructed from experimental outcomes large Complex Network, which may be used to perform datamining and quantitatively describe changes in reaction variables that determine the enantiomeric excess and configuration of the stereogenic centre formed in product. Unfortunately, there are not models to predict enantioselectivity for these reactions. Computational chemistry prediction of the reactivity based on Quantitative Structure-Reactivity Relationships (QSRR) may be used in this sense. We developed here a Multiple Linear Regression QSRR (MLR-QSRR) prediction model for the variation on enantioselectivity after changing some of the above-mentioned

Correspondence/Reprint request: Dr. Humberto González-Díaz, Faculty of Pharmacy, University of Santiago de Compostela 15782, Spain. E-mail: humberto.gonzalez@usc.es

reaction conditions and/or reactans. Overall model accuracy was high because the model explains 80.37% of variance (R^2) for 17404 and 79.99% for 26106 reaction pairs used in training and cross-validation respectively. Using the experimental values we constructed a Complex Network for Enantioselectivity (ECN_{obs}) for these reactions. In addition, the outputs of the MLR-QSRR model were used as inputs to predict a ECN_{pred} for these reactions. The ECN_{obs} has 228 nodes (reactions) and 23586 edges (pairs of reactions with high propensity to R/S chirality inversion) whereas the ECN_{pred} has 20374 edges. After edge-to-edge comparison we have demonstrated that the ECN_{pred} is significantly similar to the ECN_{obs} one (Accuracy = 76.82%). The present study opens a new interesting field of research in organic chemistry reporting by the first time the construction and QSRR prediction of ECNs for organic reactions.

Introduction

The asymmetric 1,2-addition of organometallic reagents to imines is a powerful tool to form carbon-carbon bonds. In that way it is possible to introduce a new stereogenic centre in organic molecules [1-10]. Thus, it provides ready access to enantiomerically enriched amines with a stereogenic centre at the α -position, an important structural feature in many biologically active compounds. These optically active amines are also important compounds because of their broad range of applications such as chiral auxiliaries, resolving agents and building blocks for the synthesis of natural and unnatural compounds, and their pharmacological properties [11-15]. In this kind of reactions are implicated many variables, substrates, organolithium reagents, chiral ligands, products and variables of reaction condition for instance. Therefore it exist a huge field of possible reactions to investigate.

In this sense, it is of the major interest the search of rational approaches to predict and describe the high complexity of information generated by the changes on enantioselectivity for large databases of these kind of pairs of reactions. Satoh and Funatsu have created a reaction generator in SOPHIA (System for organic reaction prediction by heuristic approach) considers reaction conditions to recognize suitable AAG (Atoms and/or Atomic Groups) for suitable free bonds by utilizing reaction condition groups obtained by classification based on word combinations of reaction condition descriptions in a reaction database. They also has been extended SOPHIA to employ the reaction condition groups for interpretation of reaction conditions entered by the user and to consider the reaction conditions in reaction prediction procedures [16]. In addition, Patel and coworkers have described a new approach, iterated reaction graphs (IRG) that simulates complex chemical reaction systems. To achieve that they modelling a subset of all possible reactions, i.e., those that are specific to Millard chemistry, modeling of the rate kinetics with reaction rate probabilities and blocking of the

reactions into logical groups [17]. In particular, Quantitative Structure-Reactivity Relationships (QSRR) studies, based on molecular descriptors of chemical structure, may play an important role in the prediction of biological activity or specific property of a reaction. For example, Ignatz-Hoover *et al.* describes QSRR for kinetic chain-transfer constants for 90 agents on styrene polymerization at 60 °C in which three- and five-parameter correlations were obtained with R^2 of 0.725 and 0.818, respectively [18]. Varnek *et al.* propose substructural fragments as a simple and safe way to encode molecular structures in a matrix containing the occurrence of fragments of a given type [19]. Satoh and coworkers have investigated a dataset of 131 reactions focusing on the changes of electronic features on the oxygen atoms at the reaction sites by principal component analysis and selforganizing neural networks analyses [20]. On the other hand, Long and Niu have developed quantitative structure-property relationship/quantitative structure-activity relationship (QSPR/QSAR) for rate constants (k) of alkyl naphthalene reactions with chlorine, hydroxyl and nitrate radicals using partial least squares (PLS) regression [21]. Mu *et al.* have investigated the prediction of oxidoreductase-catalyzed reactions based on atomic properties of metabolites [22].

As above-mentioned QSRR models may be used to predict effect of changes in reaction variables over enantioselectivity but we also need tools to describe the huge amount of information generated. This sort of problem may be investigated using Complex Networks (CNs) to regroup reactions with inverse results in which the enantiomeric excess and configuration are changed from *R* to *S*. In fact, we can use CNs to study relationships between reactions, proteins, genes, RNAs, organisms, or even non-living objects such as web pages but we can also develop *in-silico* procedures to predict these CNs [23-25].

In principle, we can extend more than 1 600 different molecular descriptors to solve the former problem [26]. Our group has introduced elsewhere a Markov Chain Model (MCM) method named MARkov CHains Invariants for Network Simulation and Design (MARCh-INSIDE). The MARCh-INSIDE approach makes use of MCM to calculate the average values of different molecular physicochemical properties in chemical structures [27]. We propose herein, for the first time, a QSRR model able to predict the difference in enantiomeric excess for *R*-product between two pair of reactions ($\Delta ee(R)\%$), which achieve to similar/dissimilar enantioselectivity after modification of reaction variables. This QSRR may predict the configuration of the new stereogenic centre formed in the synthesis of amines taking into consideration similar reaction pairs in which the enantiomeric excess increases or reduces. The task is difficult but interesting because we pretend to shun using 3D structures of substrates, chiral ligand, and products.

A method independent of these aspects may become notably faster because we do not have to run optimization algorithms to predict the 3D structure; these optimization algorithms are computationally expensive. After developing the MLR-QSRR prediction model we used it to construct an ECN. Last, we compared the ECN predicted with an ECN constructed here based on measured values of product enantiomeric excess. A summary flowchart for all the steps given in the work is presented in **Figure 1** in order to guide the reader.

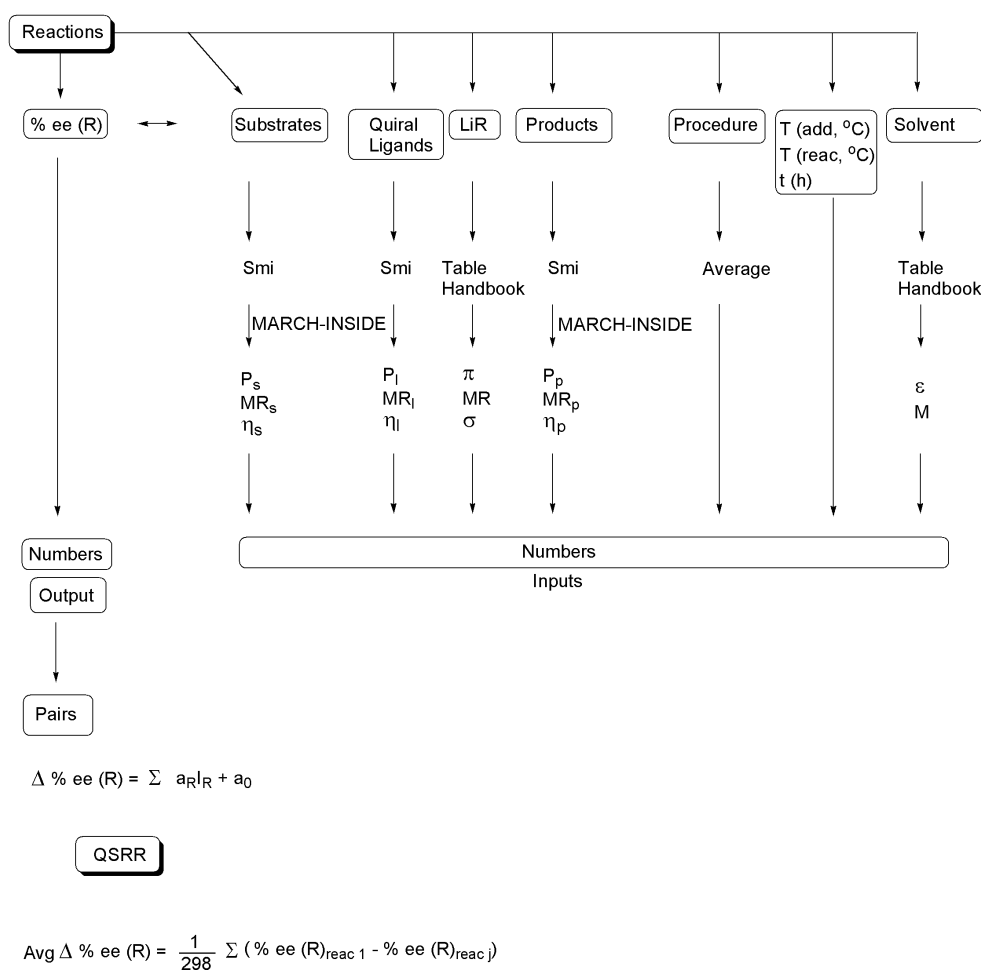


Figure 1. Flow chart for all steps given in the work.

Results and discussion

Training and validation of the MLR-QSRR model.

We used Forward-stepwise to investigate which variables more strongly influence the change on enantioselectivity and construct the MLR-QSRR equation model. The more important variables were the differences between the initial and final reaction for: product partition coefficient (ΔP_p), chiral

ligands hardness (ΔH_l), solvent dipolar moment (ΔD_s), reaction time (Δt_r), reaction temperature (ΔT_r), addition temperature (ΔT_a), average enantiomeric excess for reactions using same procedure (ΔA_e), substrate molar refractivity (ΔM_i), and steric constant (ΔS_o) and hardness of organolithium compounds (ΔP_o) respectively. Using these variables the best model found was:

$$\Delta ee(R)\% = -6.60 + 5.80 \cdot \Delta P_p - 4.63 \cdot \Delta H_l - 23.08 \cdot \Delta D_s + 44.18 \cdot \Delta t_r - 1.23 \cdot \Delta T_r - 0.18 \Delta T_a + 0.24 \Delta A_e + 1.90 \cdot \Delta S_o - 8.22 \cdot \Delta P_o - 0.24 \cdot \Delta M_i$$

$$n = 17404 \quad R^2 = 0.803 \quad R^2_{adjusted} = 0.803 \quad F = 7120.7 \quad p < 0.0000 \quad (1)$$

where, n is the number of cases (reaction pairs) used to train the model, R^2 and $R^2_{adjusted}$ are the train and adjusted square regression coefficients, F is Fisher ratio, and p the level of error. All these reactions were previously reported in the literature [28-40]. This model, with ten variables, predicts correctly 80.3% of variance of the data set with a standard error of 29.35%. Notably, the values of R^2 and $R^2_{adjusted}$ are equal, which indicates that the model is not over-fitted due to incorporating an elevated number of parameters. In the **Figure 2** we plot the observed $\Delta ee(R)$ % values *vs.* the values predicted with the model. In order to validate the model we used it to predict 26106 reactions pairs never used to train the model (validation series). In this series the results were: R^2 79.98%, F 1043E2 and $p < 0.00001$. The model explains correctly 80.0% of variance of the data set with a standard error of 29.79% in the validation series. These results indicate that we developed an accurate model according to previous reports on the use of MLR in QSRR [41-43].

Using this model we can construct QSRR-based charts to depict visually the influence of the change in reaction variables over the enantioselectivity of the reaction [44]. This kind of analysis, known as desirability analysis (DA), allows us to predict which levels of the reactions variables ensure a desired enantioselectivity. [45] It could be used to optimize the reaction changing only one property by organic synthesis modification of substrate or chiral ligand or modifying a reaction condition. In **Figure 3**, we illustrate some of these charts. Note that these charts may refer to only one receptor region or two different regions at the same time. In this sense, if we represent M_i *vs.* P_o we can observe that between -6.6 and 0.7 values of M_i the enantiomeric excess remains in spite of changes on substrates Figure 3 (a). If we represent $T(a)$ *vs.* $t(r)$ we can observed that between -3.44 and 49.9 values of reaction time ($t(r)$), the enantiomeric excess remains although reactive addition temperature change Figure 3 (b).

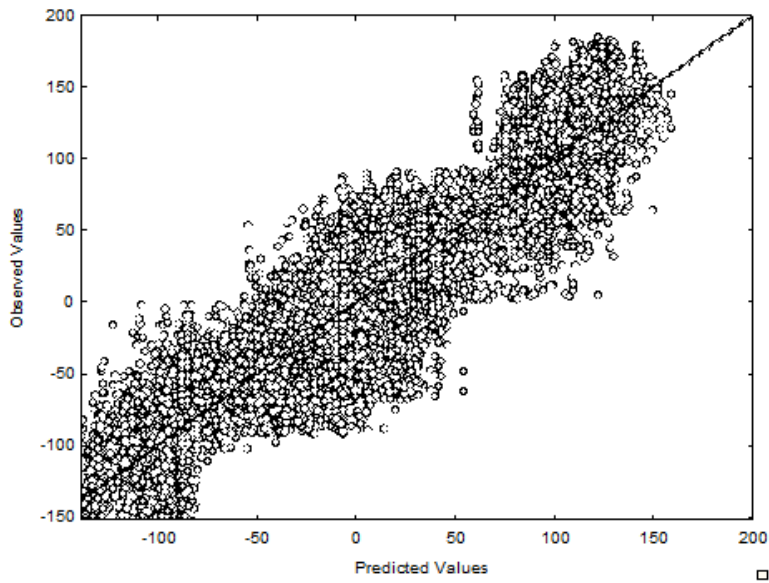


Figure 2. Observed vs predicted values.

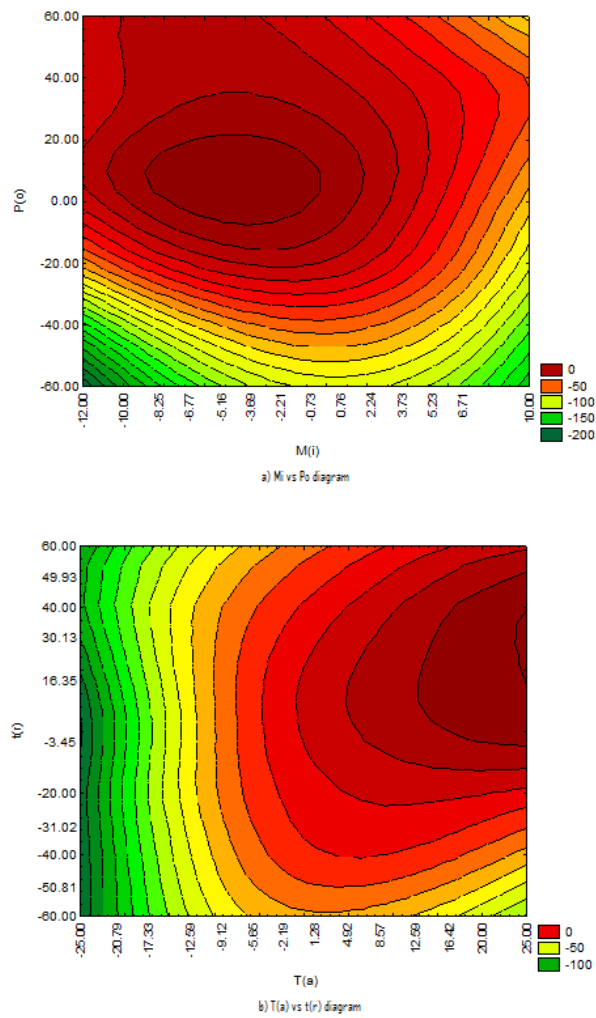


Figure 3. Example of chart used for the desirability analysis of the QSRR model.

Complex networks study. Molecular CNs are used to study large data bases and/or complex systems [46-48]. For instance, proteins, nucleic acids, and small molecules (metabolites) form a dense network of molecular interactions and metabolic reactions in a cell [49]. In order to recall the capacity of the MLR-QSRR to predict new CNs we selected the same data employed for training and validating the QSRR model. With these goals in mind, we constructed first a new ECN_{obs} using the observed values considering the experimental data. Next, we predicted the ECN_{pred} with the QSRR model and last we compared both ECNs. In our CNs we explored the threshold values in a range from -98.31 to -175, obtaining an average values of output node degree from 89.3 to 103.4 respectively (see **Table 2**). Finally, a cut off = -175 was selected to obtain average node degree equal to 103.4; which guarantee that the number of disconnected reactions is 0. Next, we used the MLR-QSRR equation to predict the enantiomeric excess and configuration of some amines. The same as before, we explored the threshold values in a range from -98.31 to -175, obtaining average values of output node degree from 77 to 89.4 respectively, a cut-off = -175, which leads to an average output node degree of 89.4, was selected being 0 the number of disconnected reactions. Additionally, with this threshold, the number of edges is 23586 for the observed network and 20374 for the predicted network. In **Figure 4** we illustrate the complex relationships between ECN drawing coincident edges for both ECN_{obs} and ECN_{pred} . In order to compare the ECN_{obs} and ECN_{pred} , we used the sensitivity, specificity and accuracy a Chi-Square test; the obtained value for the $p < 0.00001$ error level was Chi-square = 293.364.

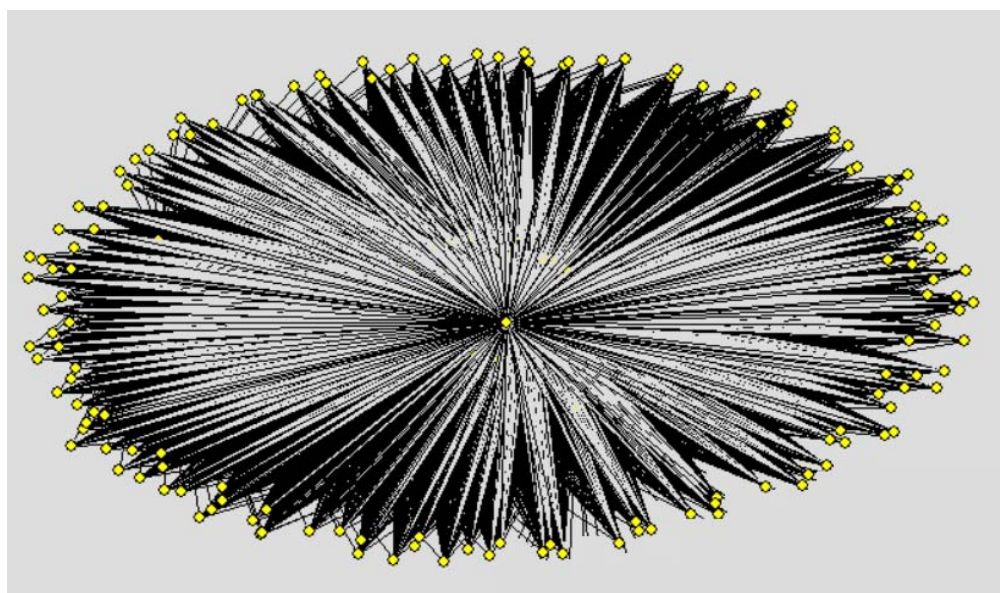


Figure 4. Graphical view of the observed vs. predicted ECNs.

Methods

Computational chemistry methods

The MARCH-INSIDE approach is based on the calculation of the different physicochemical molecular properties (λ_m) for substrates, organolithium reagents, chiral ligands and products (λ_s , λ_o , λ_l , λ_p) respectively. These λ_m are calculated as an average of atomic properties (λ_j). For instance, it is possible to derive average estimations of refractivities (MR_s , MR_o , MR_l , MR_p), partition coefficients (P_s , P_o , P_l , P_p), and hardness (η_s , η_o , η_l , η_p) that we are going to use in this work, as seen in the equation below [50]:

$$\lambda_m = \frac{1}{6} \sum_{k=0}^5 {}^k\lambda = \frac{1}{6} \sum_{k=0}^5 \sum_j p_k(\lambda_j) \cdot \lambda_j \quad (2)$$

It is possible to consider isolated atoms ($k = 0$) in the estimation of the molecular properties ${}^0\eta$, ${}^0\chi$, 0MR , ${}^0\alpha$, 0P . In this case the probabilities ${}^0p(\lambda_j)$ are determined without considering the formation of chemical bonds (simple additive scheme). However, it is possible to consider the gradual effects of the neighbouring atoms at different distances in the molecular backbone. In order to reach this goal the method uses a MM, which determines the absolute probabilities ${}^k p(\lambda_j)$ with which the atoms placed at different distances k affect the contribution of the atom j to the molecular property in question.

$${}^k\lambda = [{}^0p(\lambda_1) {}^0p(\lambda_2) \dots {}^0p(\lambda_n)] \cdot \begin{bmatrix} {}^1p_{1,1} & {}^1p_{1,2} & {}^1p_{1,3} & \dots & {}^1p_{1,n} \\ {}^1p_{2,1} & {}^1p_{2,2} & {}^1p_{2,3} & \dots & {}^1p_{2,n} \\ \dots & \dots & \dots & \dots & \dots \\ {}^1p_{n,1} & \dots & \dots & \dots & {}^1p_{n,n} \end{bmatrix}^k \cdot \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \dots \\ \lambda_n \end{bmatrix} = \sum_{j=1}^n {}^k p(\lambda_j) \cdot \lambda_j \quad (3)$$

Where, from left to right, the first term is ${}^k\lambda$, which is the average molecular property considering the effects of all the atoms placed at distance k over every atomic property λ_j . The vector on the left-hand side of the equation contains the probabilities ${}^0p(\lambda_j)$ for every atom in the molecule, without considering chemical bonds. The matrix in the centre of the equation is the so-called stochastic matrix. The values of this matrix (${}^1p_{ij}$) are the probabilities with which every atom affects the parameters of the atom

bonded to it. Both kinds of probabilities ${}^0p(\lambda_j)$ and ${}^1p_{ij}$ are easily calculated from atomic parameters (λ_j) and the chemical bonding information:

$${}^0p_{ij} = \frac{\lambda_j}{\sum_{k=1}^n \lambda_k} \quad (4)$$

$${}^1p_{ij} = \frac{\delta_{ij} \cdot \lambda_j}{\sum_{k=1}^n \delta_{ik} \cdot \lambda_k} \quad (5)$$

The only difference is that in the probabilities ${}^0p(\lambda_j)$ we consider isolated atoms by carrying out the sum in the denominator over all n atoms in the molecule. On the other hand, for ${}^1p_{ij}$ chemical bonding is taken into consideration by means of the factor δ_{ij} . This factor has the value 1 if atoms i and j are chemically bonded and it is 0 otherwise. All calculations were performed using the program MARCH-INSIDE version 3.0 [51]; which can be obtained for free academic use, upon request, from the corresponding author of the present work.

Statistical analysis

Given the λ_m the molecular parameters above-mentioned and λ_{orv} other reaction variables such as (T(a), T(r), t(r)) we can calculate the differences $\Delta\lambda = \lambda(r2) - \lambda(r1)$ for any reaction pairs. Using these $\Delta\lambda_m$ and $\Delta\lambda_{orv}$ values as input we performed a MLR analysis to fit the QSRR equation with the form:

$$\Delta ee(R)\%_{pred} = \sum_{s,l,o,p} b_m \cdot \Delta\lambda_m + \sum_{orv} b_{orv} \cdot \Delta\lambda_{orv} + b_0 \quad (6)$$

The parameter $\Delta\%ee (R)_{pred}$ (the prediction of the difference in enantiomeric excess for R -product between two pair of reactions) is the output of the model. In equation (6), b represents the coefficients of the variables in the model determine with MLR module of the software package STATISTICA 6.0 [52]. We used Forward Stepwise algorithm for variable selection. The statistical significance of the MLR model was determined calculating the $R^2 = 0.803$, $R^2_{adjusted} = 0.803$, $F = 7120.7$ and p -level (p) < 0.00001 of error with. The validation of the model was corroborated with external prediction series. The quality of the validation was determined by $R^2 = 0.799$, $R^2_{adjusted} = 0.799$,

F = 1043E2 and p-level (p) < 0.00001 for validation. The data set was conformed by a set of reported organolithium addition to imines in presence of chiral ligands reactions.

Complex network (CN) analysis

In order to achieve the enantiomeric excess and configuration of the product with a network approach where one node represents a reaction and the edges show reactions pairs with high propensity to *R/S* chirality inversion, we carried out the following steps:

1. First, we calculated the observed and QSRR-predicted average-scores that numerically characterize the propensity of one reaction to yield *R/S* chirality inversion. These scores were labelled as Obs. Avg.Δee(*R*)% and Pred.Avg.Δee(*R*)%:

$$Obs.Avg.\Delta ee(R)\%_v = \frac{1}{228} \sum_{w=1}^{w=228} \Delta ee(R)\%_{obs.}(v, w) = \frac{1}{228} \sum_{w=1}^{w=228} (ee(R)\%_{obs.}(v) - ee(R)\%_{obs.}(w)) \quad (7a)$$

$$Pred.Avg.\Delta ee(R)\%_v = \frac{1}{228} \sum_{w=1}^{w=228} \Delta ee(R)\%_{pred.}(v, w) = \frac{1}{228} \sum_{w=1}^{w=228} (ee(R)\%_{pred.}(v) - ee(R)\%_{pred.}(w)) \quad (7b)$$

Where Obs.Avg.Δee(*R*)%_v is the difference between observed *R* enantiomeric excess for reaction *v* minus observed *R* enantiomeric excess for reaction □ and Pred.Avg.Δee(*R*)% is the difference between predicted *R* enantiomeric excess for reaction *v* minus observed *R* enantiomeric excess for reaction □.

2. Then, we used these scores as inputs in a Microsoft-Excell sheet to calculate the elements of the Boolean or Adjacency matrix (A) associated to the ECNs as follows:

$$A \equiv \begin{cases} \text{if } sign(ee(R)\%_{obs.}(v)) = sign(ee(R)\%_{obs.}(w)) & \text{then } a_{vw} = 0 \\ \text{else} \\ \text{if } [Obs.Avg.\Delta ee(R)\%_v - Obs.Avg.\Delta ee(R)\%_v] \leq cut-off & \text{then } a_{vw} = 0 \\ \text{else} & a_{vw} = 1 \end{cases} \quad (8)$$

3. Next, we compared the observed and predicted ECNs pair-by-pair. For this comparison we measured the total number of coincident predictions

(Accuracy), the total number of reactions connected (Sensitivity), and not connected (Specificity) as well as a Chi-square (χ^2) test. For (χ^2) test, we used a contingency table where a, b, c and d are the observed frequencies in our networks.⁷² (See **Table 1**). These frequencies were calculated as follows: f = if (and (obs B2! = 1, pred B2! = 1), 1, if (and (obs B2! = 1, pred B2! = 0), 2, if (and (obs B2! = 0, pred B2! = 1), 3, 4)). Then:

- “a” is the number of pairs of reactions connected in observed and in predicted networks (observed and predicted are 1).
- “b” is the number of pairs of reactions connected in observed but not connected in predicted networks (observed is 1 and predicted is 0).
- “c” is the number of pairs of reactions not connected in observed but connected in predicted networks (observed is 0 and predicted is 1).
- “d” is the number of pairs of reactions not connected neither in observed nor in predicted networks (the elements in the observed and predicted matrices are equal to 0).

4. Chi-square test lets us determine if the variables are associated or not. If they are not associated we could conclude that they are independent. The first step of the chi-square test for independence is to establish hypotheses. The null hypothesis is that the two variables are independent (observed and predicted activity of the ECNs is not associated). The alternative hypothesis to be tested is that the two variables are dependent. χ^2 was calculated as follows [53]:

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^k \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \quad (9)$$

5. Where O_{ij} is the observed frequency, E_{ij} is the expected or theoretical frequency. E_{ij} is calculated as follows

$$E_{11} = \frac{(a + b) \times (a + c)}{n} \quad (10a)$$

$$E_{21} = \frac{(c + d) \times (a + c)}{n} \quad (10b)$$

$$E_{12} = \frac{(a + b) \times (b + d)}{n} \quad (10c)$$

$$E_{22} = \frac{(c + d) \times (b + d)}{n} \quad (10d)$$

6. Then we compared the value calculated in the formula above to a standard set of tables. The value returned from the table is $p < 0.005$. Thus, we can reject the null hypothesis and conclude that there is an association between the variables.

7. The Boolean matrix was saved as a .txt format file. After we had renamed the .txt file as a .mat file we read it with the software CentiBin [54, 55]. Using CentiBin we can not only represent the network but also highlight all nodes connected to a specific node but calculate connectivity parameters including node degree.

8. CentiBin software was used to generate random networks by five different algorithms including: Barabasi-Albert random network, Kleinberg Small World Network (SWN), 2D Lattice network, Erdos-Renyi network and Epsstein power law network (PLN) [55]. These random networks were compared with the observed and predicted networks.

9. Last, all node degrees were used as input in STATISTICA in order to study the distribution of the network and compare it with other ideal network distributions including normal, lognormal, exponential, gamma, and Chi-square [52].

Supplementary material

In the online supplementary material we depict all the parameters necessary to evaluate a reaction with the QSRR as well as the dataset used including corresponding structures, in the form of SMILE codes, for all compounds involved in this study.

Conclusions

Using the MARCH-INSIDE approach is possible to seek a MLR-QSRR classifier to predict the probability of chirality inversion of reactions; which occur by addition of organolithium reagents to imines in presence of chiral ligands. The model can be used as a tool for preliminary screening of reactions without relaying upon geometrical optimization of substrate, organolithium, chiral ligand, or product structure. This MLR-QSRR was also demonstrated to be an efficient tool for computational construction of Enantioselectivity Complex Networks that accurately reproduces the network based on experimental findings. This kind of Complex Networks could

become a valuable approach to explore the complexity of the enantioselectivity in these reactions.

Acknowledgments

Arrasate S. acknowledges sponsorships for a tenure-track research position at the University of Santiago de Compostela from the “Ikertzaileak Hobetzeko eta Mugitzeko/Perfeccionamiento y Movilidad del Personal Investigador” Program of the “Hezkuntza, Unibertsitate eta Ikerketa Saila/Departamento de Educación, Universidades e Investigación, Eusko Jaurlaritzza/Gobierno Vasco”. Financial support from Gobierno Vasco (GIC07/92-IT-227-07) also is gratefully acknowledged. González-Díaz H. acknowledges sponsorships for a tenure-track research position at the University of Santiago de Compostela from the *Isidro Parga Pondal* Program of the “Dirección Xeral de Investigación e Desenvolvemento, Xunta de Galicia”.

References

1. Klein J. The Chemistry. In: Patai S, ed. *The Chemistry of Double-bonded Functional Groups: Supplement A*. Chichester: Wiley 1989.
2. Volkmann RA. In: Schreiber SL, ed. *Comprehensive Organic Synthesis, Additions to C-X π -Bonds, Part 1* Oxford: Pergamon Press 1991.
3. Kleinman EFV, R. A. In: Heathcock CH, ed. *Comprehensive Organic Synthesis, Additions to C-X π -Bonds, Part 2*. Oxford: Pergamon Press 1991.
4. Berrisford DJ. *Angew Chem, Int Ed Engl.* 1995;34:178-80.
5. Risch NA, M. In: Helmchen GH, R. W.; Mulzer, J.; Schaumann, E., ed. *Methods of Organic Chemistry Stereoselective Synthesis [Houben-Weyl]*. Stuttgart: Thieme 1996.
6. North M. *Contemp Org Synth* 1996;3:323-43.
7. Denmark SEN, O. J.-C. *Chem Commun.* 1996:999-1004.
8. Enders DR, U. *Tetrahedron: Asymmetry.* 1997;8:1895-946.
9. Bloch R. *Chem Rev.* 1998;98:1404-38.
10. Denmark SEN, O. J.-C. In: Jacobsen ENP, A.; Yamamoto, H., ed. *Comprehensive Asymmetric Catalysis*. Berlin: Springer-Verlag 1999.
11. Seyden-Penne J. *Chiral Auxiliaries and Ligands in Asymmetric Synthesis*. New York: Wiley 1995.
12. Jacques JC, A.; Wilen, S. H. *Enantiomers, Racemates, and Resolution*. New York: Wiley 1981.
13. Eliel ELW, S. H.; Mander, L. N. *Stereochemistry of Organic Compounds*. New York: Wiley 1994.
14. Moser HR, G.; Santer, H. *Z. Naturforsch.* 1982;37B:451-62.

15. Ariëns EJS, W.; Timmermans, P. B. M. W. M. *Stereochemistry and Biological Activity of Drugs*. Oxford: Blackwell Scientific 1983.
16. Satoh H, ; Funatsu, K. Further Development of a Reaction Generator in the SOPHIA System for Organic Reaction Prediction. Knowledge-Guided Addition of Suitable Atoms and/or Atomic Groups to Product Skeleton. *J Chem Inf Comput Sci*. 1996;36:173-84.
17. Shail Patel JR, Stephen Russell, Jos Tissen, and Werner Klaffke. *J. Chem. Inf. Comput. Sci*. 2001;41:926-33.
18. Ignatz-Hoover F, Petrukhin R, Karelson M, Katritzky AR. QSRR correlation of free-radical polymerization chain-transfer constants for styrene. *J Chem Inf Comput Sci*. 2001 Mar-Apr;41(2):295-9.
19. Varnek AF, D.; Hoonakker, F.; Solov'ev, V. P. Substructural fragments: an universal language to encode reactions, molecular and supramolecular structures. *J Comput Aided Mol Des*. 2005;19:693-703.
20. Hiroko Satoh OS, Tadashi Nakata, Lingran Chen, Johann Gasteiger, Funatsu K. *J. Chem. Inf. Comput. Sci*. . 1998 38:210-9.
21. Long XN, J. Estimation of gas-phase reaction rate constants of alkylnaphthalenes with chlorine, hydroxyl and nitrate radicals. *Chemosphere*. 2007;67:2028-34.
22. Mu FU, P. J.; Unkefer, C. J.; Hlavacek, W. S. Prediction of oxidoreductase-catalyzed reactions based on atomic properties of metabolites. *Bioinformatics*. 2006;22(24):3082-8.
23. Bornholdt S, Schuster HG. *Handbook of Graphs and Complex Networks: From the Genome to the Internet*. Weinheim: WILEY-VCH GmbH & CO. KGa. 2003.
24. Boccaletti S, Latora V, Moreno Y, Chavez M, Hwang DU. Complex networks: Structure and dynamics. *Phys Rep*. 2006;424:175-308.
25. Réka A, Barabasi A-L. Statistical mechanics of complex networks. *Rev Mod Phys*. 2002;74(1):47-97.
26. Todeschini R, Consonni V. *Handbook of Molecular Descriptors: Wiley-VCH* 2002.
27. González-Díaz H, Vilar S, Santana L, Uriarte E. Medicinal Chemistry and Bioinformatics – Current Trends in Drugs Discovery with Networks Topological Indices. *Curr Top Med Chem*. 2007;7(10):1025-39.
28. Arrasate S, Lete S, Sotomayor N. Synthesis of enantiomerically enriched amines by chiral ligand mediated addition of organolithium reagents to imines *Tetrahedron: Asymmetry*. 2001; 12(14):2077-82.
29. Denmark SES, C. M. Effect of Ligand Structure in the Bisoxazoline Mediated Asymmetric Addition of Methylithium to Imines. *J Org Chem*. 2000;65:5875-8.
30. Hasegawa MT, D.; Tomioka, K. Facile Asymmetric Synthesis of α -Amino Acids Employing Chiral Ligand-Mediated Asymmetric Addition Reactions of Phenyllithium with Imines. *Tetrahedron*. 2000;56:10153-8.
31. Taniyama DH, M.; Tomioka, K. A facile asymmetric synthesis of 1-substituted tetrahydroisoquinoline based on a chiral ligand-mediated addition of organolithium to imine. *Tetrahedron: Asymmetry*. 1999;10:221-3.

32. Inoue IS, M.; Koga, K; Kanai, M.; Tomioka, K. . Enantioselective Reaction of An Imine with Methyllithium Catalyzed by A Chiral Ligand. *Tetrahedron: Asymmetry*. 1995;6:2527-33.
33. Denmark SEN, N.; Nicaise, O. J.-C. . Asymmetric Addition of Organolithium Reagents to Imines. *J Am Chem Soc*. 1994;116:8797-8.
34. Anderson PGJ, F.; Tanner, D. Enantioselective Addition of Organolithium Reagents to Imines Mediated by C2-Symmetric Bis(aziridine) Ligands. *Tetrahedron*. 1998;54:11549-66.
35. Inoue IS, M.; Koga, K; Tomioka, K. Asymmetric 1,2-Addition of Organolithium to Aldimines Catalyzed by Chiral Ligand. *Tetrahedron*. 1994;50:4429-38.
36. Perron QA, A. Synthesis and application of a new pseudo C2-symmetric tertiary diamine for the enantioselective addition of MeLi to aromatic imines. *Tetrahedron: Asymmetry*. 2007;18:2503-6.
37. Cabello NK, J.-C.; Gille, S.; Alexakis, A.; Bernardinelli, G.; Pinchard, L.; Caille, J.-C. Simple 1,2-Diamine Ligands for Asymmetric Addition of Aryllithium Reagents to Imines. *Eur J Org Chem*. 2005:4835-42.
38. Kizirian J-CC, N.; Pinchard, L.; Caille, J.-C.; Alexakis, A.; . Enantioselective addition of methyllithium to aromatic imines catalyzed by C2 symmetric tertiary diamines. *Tetrahedron*. 2005;61:8939-46.
39. Gille SC, N.; Kizirian, J.-C.; Alexakis, A.;. A new pseudo C2-symmetric tertiary diamine for the enantioselective addition of MeLi to aromatic imines. *Tetrahedron: Asymmetry*. 2006;17:1045-7.
40. Cabello NK, J.-C.; Alexakis, A.; Enantioselective addition of aryllithium reagents to aromatic imines mediated by 1,2-diamine ligands. *Tetrahedron Lett*. 2004;45:4639-42.
41. Cheng Z, Ren J, Li Y, Chang W, Chen Z. Study on the multiple mechanisms underlying the reaction between hydroxyl radical and phenolic compounds by qualitative structure and activity relationship. *Bioorg Med Chem*. 2002 Dec;10(12):4067-73.
42. Pompe MV, M.; Randic, M.; Balaban, A. T. Using variable and fixed topological indices for the prediction of reaction rate constants of volatile unsaturated hydrocarbons with OH radicals. *Molecules (Basel, Switzerland)*. 2004;9:1160-76.
43. Ren YL, H.; Yao, X.; Liu, M. Prediction of ozone tropospheric degradation rate constants by projection pursuit regression. *Anal Chim Acta*. 2007;589:150-8.
44. González-Díaz H, Saiz-Urra L, Molina R, Santana L, Uriarte E. A Model for the Recognition of Protein Kinases Based on the Entropy of 3D van der Waals Interactions. *Journal of proteome research*. 2007 Feb 2;6(2):904-8.
45. Cruz-Monteagudo M, González-Díaz H, Agüero-Chapin G, Santana L, Borges F, Domínguez RE, et al. Computational Chemistry Development of a Unified Free Energy Markov Model for the Distribution of 1300 Chemicals to 38 Different Environmental or Biological Systems. *J Comput Chem*. 2007; 28:1909-22.
46. Bonchev D, Buck GA. From molecular to biological structure and back. *Journal of chemical information and modeling*. 2007 May-Jun;47(3):909-17.
47. Bonchev D. On the complexity of directed biological networks. *SAR QSAR Environ Res*. 2003 Jun;14(3):199-214.

48. Park J, Barabasi AL. Distribution of node characteristics in complex networks. *Proc Natl Acad Sci U S A*. 2007 Nov 13;104(46):17916-20.
49. Spirin V, Mirny LA. Protein complexes and functional modules in molecular networks. *Proc Natl Acad Sci U S A*. 2003 Oct 14;100(21):12123-8.
50. Santana L, Uriarte E, González-Díaz H, Zagotto G, Soto-Otero R, Mendez-Alvarez E. A QSAR model for in silico screening of MAO-A inhibitors. Prediction, synthesis, and biological assay of novel coumarins. *J Med Chem*. 2006 Feb 9;49(3):1149-56.
51. González-Díaz H, Molina-Ruiz R, Hernandez I. **MARCH-INSIDE** v3.0 (**MARK**ov **CH**ains **IN**variants for **SI**mulation & **DE**sign); Windows supported version under request to the main author contact email: gonzalezdiazh@yahoo.es. 3.0 ed 2007.
52. StatSoft.Inc. STATISTICA (data analysis software system), version 6.0, www.statsoft.com. Statsoft, Inc. 6.0 ed 2002.
53. Hill T, Lewicki P. STATISTICS Methods and Applications. A Comprehensive Reference for Science, Industry and Data Mining. Tulsa: StatSoft 2006
54. Koschützki D. CentiBiN Version 1.4.2. 2006:CentiBiN Version 1.4.2, Centralities in Biological Networks © 2004-6 Dirk Koschützki Research Group Network Analysis, IPK Gatersleben, Germany.
55. Junker BH, Koschutzki D, Schreiber F. Exploration of biological network centralities with CentiBiN. *BMC bioinformatics*. 2006;7:219.