



Antimalarial activity of synthetic 1,2,4-trioxanes and cyclic peroxy ketals, a quantum similarity study

X. Gironés, A. Gallegos & R. Carbó-Dorca*

Institute of Computational Chemistry, University of Girona, Campus Montilivi, 17071 Girona, Catalonia, Spain

Received 11 May 2000; accepted 11 January 2002

Key words: artemisinin, malaria, MQSM, peroxy ketals, plasmodium falciparum, quantum similarity, QSAR, 1,2,4-trioxanes

Summary

In this work, the antimalarial activity of two series of 20 and 7 synthetic 1,2,4-trioxanes and a set of 20 cyclic peroxy ketals are tested for correlation search by means of Molecular Quantum Similarity Measures (MQSM). QSAR models, dealing with different biological responses (IC₉₀, IC₅₀ and ED₉₀) of the parasite *Plasmodium Falciparum*, are constructed using MQSM as molecular descriptors and are satisfactorily correlated. The statistical results of the 20 1,2,4-trioxanes are deeply analyzed to elucidate the relevant structural features in the biological activity, revealing the importance of phenyl substitutions.

Introduction

Approximately 300 million people worldwide are affected by malaria and between 1 and 1.5 million people die from it every year [1, 2]. Previously extremely widespread, the malaria is now mainly confined to Africa, Asia and Latin America. The problems of controlling malaria in these countries are aggravated by inadequate health structures and poor socioeconomic conditions, and the situation is becoming even more complex because over the last few years there has appeared an increasing resistance to the drugs used to combat the *Plasmodium* parasite, which causes the disease. Malaria is treated by an assorted set of therapies that act as inhibitors of synthesis and reduction of dihydrofolase in the parasites [3]. However, deteriorating social and economic conditions, along with inadequate sanitation and health care, lead to self-administration of drugs, often with incomplete treatment, which result in an increase in *resistance* of the parasites to commonly used drugs. Fortunately, two natural peroxides, artemisinin and yingzhaosu, which possess potent antimalarial activity, have opened a new door in the chemotherapy of malaria [4–6]. Due

to the complex structure of the natural products, structurally simpler 1,2,4-trioxanes have been synthesized and tested for antimalarial activity [7–9]. In this way, the design of new, both cheap and powerful, compounds able to treat malaria are of major importance. QSAR studies represent one of the best methodologies in computer-based drug design, offering valuable information about biological activity and providing a computationally inexpensive methodology to the design of potential bioactive drugs. Thus this work is devoted to this task.

In the present study, the QSAR models for two sets of 1,2,4-trioxanes and a series of cyclic peroxy ketals, which possess antimalarial activity, are built using Molecular Quantum Similarity Measures (MQSM) [10–16]. This methodology is based upon quantitative comparative measures between molecular density functions and has been applied successfully to QSAR studies within pharmacological [17–23], including also antimalarial compounds [24], and toxicological [25, 26] environments. Quantum similarity measures provide a suitable quantification of the resemblance between two molecular electron distributions, and the similarity matrices, obtained by collecting the MQSM among a molecular set, can then be manipulated to generate QSAR parameters. The mathematical foun-

*To whom correspondence should be addressed. E-mail: director@iqc.udg.es

dation of MQSM, as well as many up-to-date applications into the QSAR field, is compiled in a recent volume by Carbó-Dorca et al. [27].

This paper is organized as follows: first, the basic methodology concerning MQSM and the statistical treatment of the resulting similarity matrices is exposed. Next, the molecular sets composed by 20 and 7 1,2,4-trioxanes, and 20 cyclic peroxy ketals are presented and the QSAR analysis discussed. Finally, the conclusions of this work are given.

Materials and methods

Molecular Quantum Similarity Measures (MQSM) and Carbó indices

The foundation of MQSM lays on the comparison of quantum mechanical density functions. These density functions are, if chosen as first-order, quantum-mechanical observable elements that produce information on the electron distribution of the molecules. Within this framework, two molecular structures are considered to be similar if their electron distributions are. Thus, among other possibilities [12, 15, 16], quantitative measures of the similarity between two molecules can be defined as the direct volume integral between their density functions:

$$Z_{AB} = \int \rho_A(\mathbf{r}) \rho_B(\mathbf{r}) d\mathbf{r}, \quad (1)$$

or as the volume integral weighted by the Coulomb operator:

$$Z_{AB} = \iint \rho_A(\mathbf{r}_1) |\mathbf{r}_1 - \mathbf{r}_2|^{-1} \rho_B(\mathbf{r}_2) d\mathbf{r}_1 d\mathbf{r}_2, \quad (2)$$

where $\rho_A(\mathbf{r})$ and $\rho_B(\mathbf{r})$ are the electron density functions of molecule A and B, respectively. Z_{AB} is the resulting quantum similarity measure. As MQSM depend on the relative position of both compared objects in space, an alignment procedure is required. When studied molecular sets share common structural features, Topo-Geometrical Superposition Algorithm (TGSA) [28] is used as it performs pairwise superpositions according to the molecular backbones.

TGSA consists of a molecular alignment procedure based on comparisons of atom types and interatomic distances. TGSA finds the largest common molecular substructure to a series of molecules, and orients the molecular superposition towards the matching of such substructure. TGSA program can be obtained upon request [28].

Once superposed, the overall set of MQSM for a series of molecules are computed and collected in *similarity matrix* form, which can be used to extract information and build QSAR descriptors. When the similarity matrix has been computed, a possible scaling of the MQSM can be done by means of the transformation:

$$C_{AB} = Z_{AB}/[Z_{AA}Z_{BB}]^{1/2}, \quad (3)$$

giving the so-called Carbó-index [10]. Carbó indices are comprised within the interval [0, 1]. The closer to one the index is, the more similar to each other the compounds are.

To avoid expensive theoretical calculations, the promolecular Atomic Shell Approximation (ASA) [29–31] has been used here to construct molecular electron density. Within ASA the molecular density is expressed as a sum of discrete atomic density contributions, which are taken as 1S Gaussian functions and fitted to atomic *ab initio* ones. Fitting ASA parameters can be freely downloaded from our web site [32]. Once the overall atomic densities are built, each molecular density function can be constructed by adding as appropriate these elementary building blocks. Since it has been proved that the MQSM from molecular densities built in this way differ by up to a 2% from the *ab initio* ones [30], their use is clearly justified.

Once the similarity matrix has been constructed and, if applicable, Carbó indices calculated, no further calculations are needed before the construction of QSAR models, which obviously requires some computation. This is the main advantage of this method due that only molecular density functions are required, which are straightforwardly obtained using ASA, and the calculation of cross-similarities, which can be easily accomplished. There is no need to obtain or compute further parameters, like log P, Hammett sigma or molecular fields. Given a molecular set, MQSM, which are readily obtained with present computational resources, will constitute the basis for the generation of molecular descriptors. In this way, molecular quantum similarity parameters are unbiased and universal molecular descriptors.

Treatment of quantum similarity matrices

Usual chemometric tools are applied here to deal with similarity matrices. Principal Components Analysis (PCA) [33] is an appropriate technique to reduce the dimensionality of the similarity data. The method considers the objects as points in a multidimensional

space and then finds coordinates such that the inter-point distances match as well as possible the original similarities. Taking the Carbó-index matrix C as a starting point, a coordinate matrix X can be derived in a new multidimensional space. The coordinates in this space are called the Principal Components (PCs) of the system and constitute the descriptors used in this work. Each PC is associated with an eigenvalue, indicating the explained variance of the axis. All those PCs accounting for less than 1% variance are a priori neglected, avoiding unnecessary background noise correlations. Furthermore, the optimal variables to describe the desired property are not necessarily those that account for maximal variance, so a simple selection technique is adopted. The Most Predictive Variable Method (MPVM) [34] quantifies the importance of the descriptors simply by projecting each PC on the external data.

Quantitative models are built by means of multilinear regression using the selected PCs as independent X -variables. The goodness-of-fit is assessed with the conventional r^2 and standard deviation σ_N coefficients. The model obtained is corroborated by leave-one-out cross validation [35]. The cross-validation technique consists of removing one object from the set and then recalculating the predictive model with the remaining objects. This new model is then used to predict the property value for the extracted element. This process is repeated for all molecules of the set, and then a coefficient of prediction q^2 can be defined from the squared cross-validation residuals (PRESS). This coefficient of prediction measures the robustness of the proposed model, and a $q^2 > 0.5$ is commonly accepted as satisfactory.

A note of caution should be given at this point concerning the cross-validation using PCA. When removing the rows which describe each object to be predicted, little part of the information regarding the removed object remains into the objects left to construct the model; as well as the excluded descriptors keep also part of information of the rest of the set. However, as the evaluated property is not present when generating the PCs, this fact is not relevant, as it would be in other statistical procedures, such as Partial least-squares [36]. So, when predicting new molecules, the PCA matrix should be recomputed and the model, with the evaluated property, recalculated to account for the perturbations induced by this new molecule. If the number of molecules to predict is relatively small compared to the training set, the new models generated are not likely to be substantially modified.

Finally, a randomization test [37] is adopted to detect possible chance correlations. In this statistical technique, the dependent Y -variables are randomly permuted in their positions, and new predictive models are built with the altered vectors. If a real structure-activity exists, the only satisfactory results should be obtained for the correctly ordered Y -variables. Otherwise, even if the obtained model seems to correctly describe the system, it cannot be considered significant, because it is built up with an excess of parameters, able to correlate any data set.

Results and discussion

The major goal in QSAR analysis is to establish a simple mathematical relation between the function, namely biological activity, and a few set of molecular descriptors related to the molecular structures involved. Two molecular sets composed by 20 and 7 1,2,4-trioxanes and one group of 20 cyclic peroxy ketals, presented in Tables 1, 2 and 3 respectively, are tested for correlation search between their antimalarial activity and descriptors derived from MQSM.

Molecular modeling

All molecular structures have been constructed using the PC Spartan software package [38] and optimized with the built-in Sybyl Molecular Mechanics force-field. This choice was made according to previous studies [39], where a comparison between Molecular Mechanics and *ab initio* procedures was carried out, proving the superiority of Molecular Mechanics in these molecular sets, and thus saving computational time. A similar study was carried out in a previous work on antimalarial activity [24], where Molecular Mechanics also provided the optimal conformation.

Activity data

For the set composed by 20 antimalarial compounds, the analyzed properties (Y -variables) consist of the concentration (in ng/ml) of the drug able to inhibit 90% of synthesis and reduction of hydrofolate (IC₉₀) in the parasite. *In vitro* experiments have been performed studying two specimens: the *P. falciparum* W2, from Indochina, and D6 clone, from Sierra Leone. Due to the wide range of values present in reference [39], a logarithmic scaling has been adopted to uniform the activity data. For the smaller set formed by 7 compounds, the biological activity is the same

Table 1. Structures and observed in vitro (log IC₉₀) activities of the set of 20 1,2,4-trioxanes.

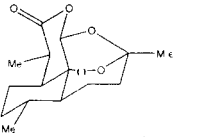
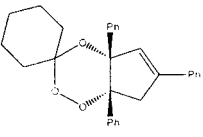
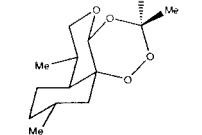
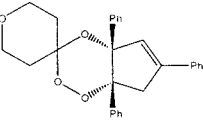
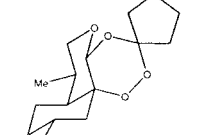
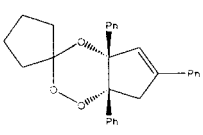
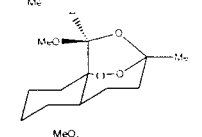
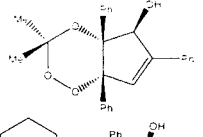
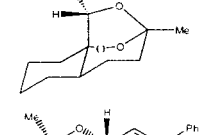
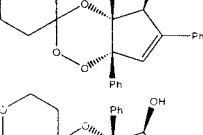
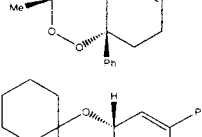
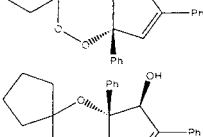
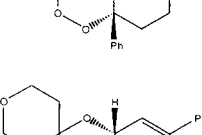
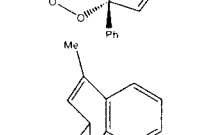
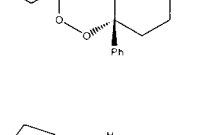
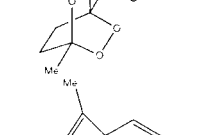
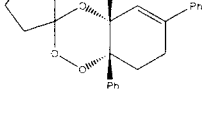
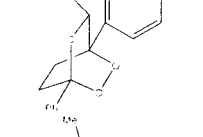
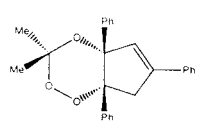
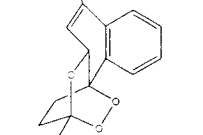
| <i>n</i> | Structure | <i>P. falciparum</i> Indo-China W2 log IC ₉₀ | <i>P. falciparum</i> Sierra Leone D6 log IC ₉₀ | <i>n</i> | Structure | <i>P. falciparum</i> Indo-China W2 log IC ₉₀ | <i>P. falciparum</i> Sierra Leone D6 log IC ₉₀ |
|----------|---|--|--|----------|--|--|--|
| 1 |  | 0.0414 | 0.3617 | 11 |  | 0.6990 | 1.0864 |
| 2 |  | 1.4116 | 1.7767 | 12 |  | -0.3979 | 1.0969 |
| 3 |  | 0.5185 | 1.4829 | 13 |  | 0.0414 | 0.3617 |
| 4 |  | 0.5798 | 1.0607 | 14 |  | 0.9731 | 1.0086 |
| 5 |  | 1.2279 | 2.9420 | 15 |  | 0.3010 | 0.6990 |
| 6 |  | 3.0734 | 3.1004 | 16 |  | 1.0864 | 1.3054 |
| 7 |  | 2.8082 | 3.4125 | 17 |  | -0.2218 | 1.0253 |
| 8 |  | 2.7482 | 3.1483 | 18 |  | 3.0550 | 2.8209 |
| 9 |  | 2.3483 | 2.3856 | 19 |  | 2.7543 | 2.6884 |
| 10 |  | 0.9243 | 1.2856 | 20 |  | 1.4771 | 1.5441 |

Table 2. Structures and observed in vivo (ED₉₀) activities of the set of 7 1,2,4-trioxanes

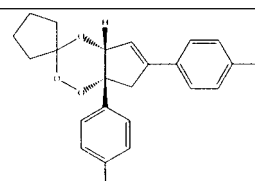
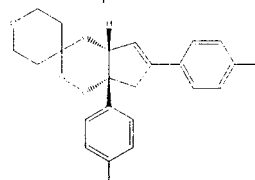
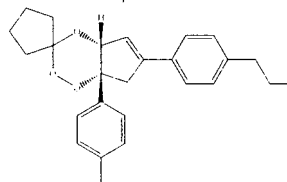
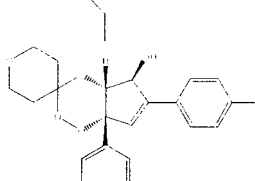
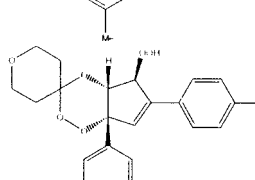
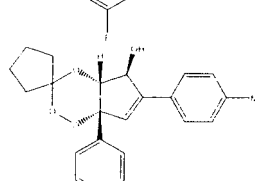
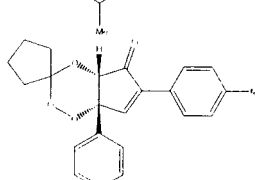
| <i>n</i> | Structure | <i>P. beghei</i> ED ₉₀ |
|----------|---|--------------------------------------|
| 1 |  | 6.8 |
| 2 |  | 19.5 |
| 3 |  | 22.5 |
| 4 |  | 17.0 |
| 5 |  | 13.2 |
| 6 |  | 15.0 |
| 7 |  | 16.0 |

Table 3. Structures and observed in vitro (log IC₅₀) activities of the set of 20 cyclic peroxy ketals

| <i>n</i> | Structure | log IC ₅₀ | <i>n</i> | Structure | log IC ₅₀ |
|----------|-----------|----------------------|----------|-----------|----------------------|
| 1 | | 3.041 | 11 | | 1.929 |
| 2 | | 2.279 | 12 | | 1.892 |
| 3 | | 2.447 | 13 | | 1.491 |
| 4 | | 2.342 | 14 | | 2.255 |
| 5 | | 2.204 | 15 | | 2.204 |
| 6 | | 2.255 | 16 | | 1.748 |
| 7 | | 2.322 | 17 | | 1.663 |
| 8 | | 2.079 | 18 | | 2.000 |
| 9 | | 1.785 | 19 | | 2.301 |
| 10 | | 1.763 | 20 | | 2.146 |

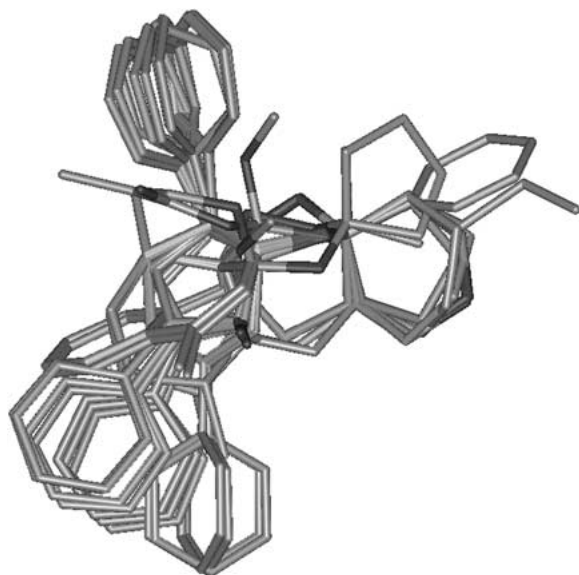


Figure 1. Graphical example of the TGSA alignment solution for artemisinin (molecule 1 from table 1). All molecules have been pairwise superposed to it. (Hydrogen atoms have been suppressed for clarity)

(but in mg/kg). However, it has been measured in the *P. berghei* *in vivo* (ED₉₀). In this case, as the range of values is narrower, no scaling has been performed. The biological property studied in the 20 cyclic peroxy ketals set [40] also consists of inhibition of the metabolism to hydrofolate, but 50% (IC₅₀) in nM concentration units. Similarly to the first set, activity values have been taken in logarithmic scale. All data is presented together with the molecular structures in Tables 1*, 2 and 3.

Molecular alignment

As commented before, MQSM depend on the relative position of the compared molecules in space. In this work, TGSA [28] has been used to perform the pairwise molecular alignments. This method overlays the molecules according the maximal common substructure shared by the analyzed molecules. An illustrative example of the application of TGSA in the 20 antimalarial molecules from the first set is presented in Figure 1.

As seen in Figure 1, the alignment solution provided by TGSA tends to superpose the trioxane groups, and, hence, the rest of the molecular structure is unequivocally overlaid, as it can be seen by

*IC₉₀ value for compound number 1 in Table 1 was not available and IC₅₀ is used instead.

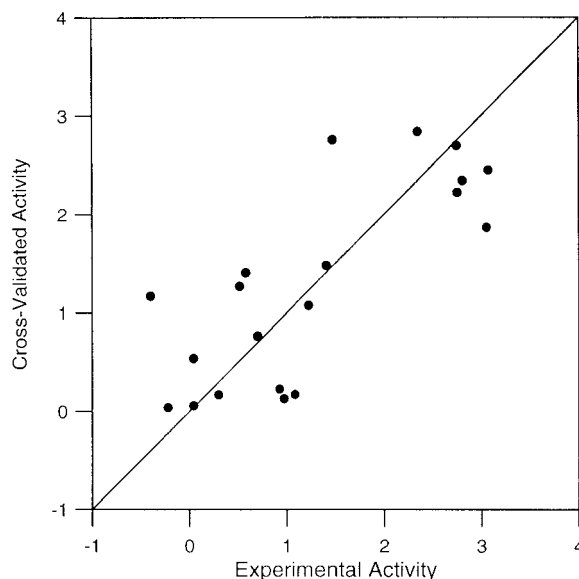


Figure 2. Cross-validated versus experimental antimalarial activity values over *P. falciparum* W2 (1,2,4-Trioxanes).

the phenyl groups, which are located in two clearly differentiated regions.

Quantitative structure-activity relationships

MQSM matrices have been constructed as follows: the sets composed by 1,2,4-Trioxanes have been calculated using the volume integral stated in Equation 1, whereas the Coulomb formulation presented in Equation 2 has been used in the cyclic peroxy ketals. The reason for a different MQSM calculation in these sets, beyond the statistical results, often resides in a different chemical behavior of the involved chemical species. Whereas Overlap operator accounts for simple steric effects, Coulomb describes more electrostatic interactions.

In order to quantify the influence on antimalarial activity of the diverse molecules according to their structure, the quantum similarity matrix deals with multivariate analysis techniques as detailed previously. Simple mathematical manipulations yield to reliable models that relate the molecular descriptors to the antimalarial activity.

A quantitative relationship between the structural descriptors and the property can be established by means of a multilinear regression, as pointed out in the theoretical section. Valuable models are achieved using few descriptors, as presented in Tables 4–7. Optimal values of $r^2 = 0.757$ and $q^2 = 0.589$ for the *P. falciparum* W2, and $r^2 = 0.789$ and $q^2 = 0.662$

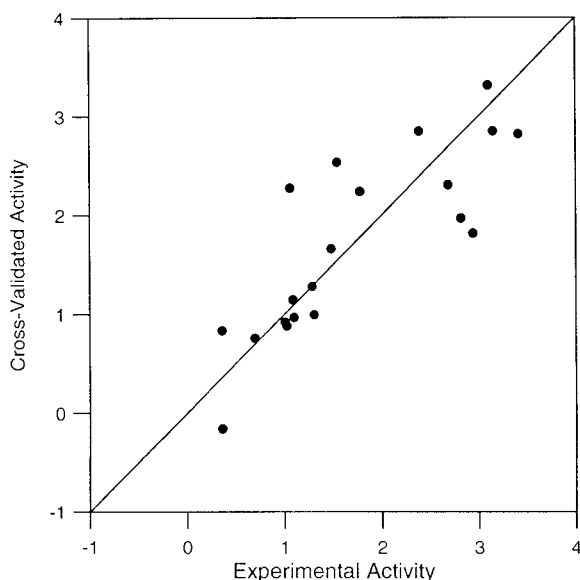


Figure 3. Cross-validated versus experimental antimalarial activity values over *P. falciparum* D6 (1,2,4-Trioxanes).

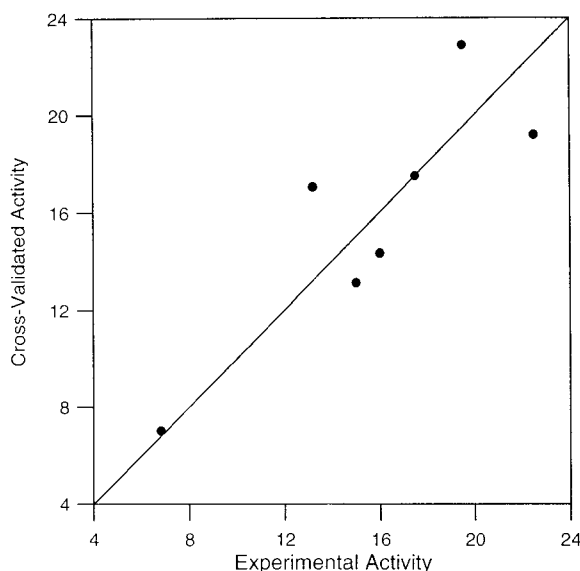


Figure 4. Cross-validated versus experimental antimalarial activity values over *P. berghei* (1,2,4-Trioxanes).

for the *P. falciparum* D6 are obtained when using 4 descriptors for the first molecular set. For the second series, $r^2 = 0.929$ and $q^2 = 0.708$ using three PCs. Finally, in the cyclic peroxy ketals set, $r^2 = 0.778$ and $q^2 = 0.691$ values have been achieved using 4 PCs. As evidenced in the following set of equations, the most descriptive PCs are not necessarily those accounting for maximal variance. The equations of the optimal

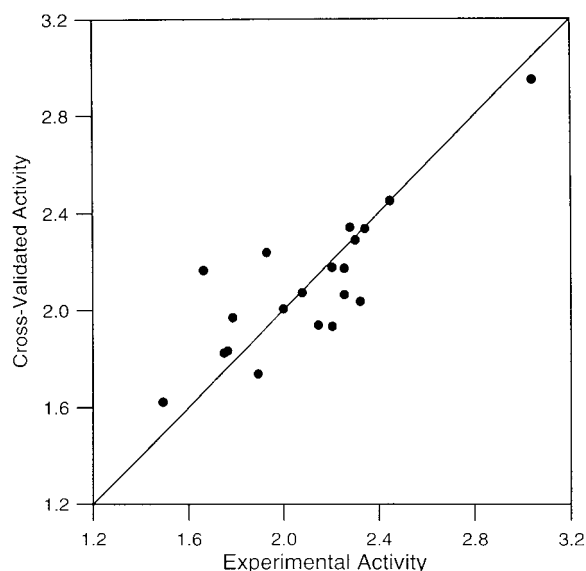


Figure 5. Cross-validated versus experimental antimalarial activity values over *P. falciparum* (cyclic peroxy ketals).

models are:

$$\log IC_{90}^{W2} = 2.912x_1 - 7.166x_2 - 2.083x_3 + 4.948x_7 + 1.272, \quad (4)$$

$$\log IC_{90}^{D6} = 1.806x_1 - 6.102x_2 + 5.966x_7 - 3.420x_{11} + 1.730, \quad (5)$$

$$ED_{90} = -16.53x_4 - 11.67x_5 - 37.52x_6 + 15.79 \quad (6)$$

$$\log IC_{50} = 1.739x_1 - 10.188x_2 - 4.042x_3 - 11.379x_6 + 2.107. \quad (7)$$

Figures 2–5 present the plots of the cross-validated activities of the two activities of the first set and the activities of the second and third sets, respectively. As seen, most of the predicted values are in good agreement to the experimental ones with low error.

Finally, the randomization test was performed to assess that the relationships found were not due to an overparameterization of the models or background noise correlations. Thus, in each case, a hundred new models were built in the same optimal conditions (MPVM and 4/3 PCs were applied), but the Y -variables were randomized permutations of the original activities. The results are compiled in the Table 8, where the average values, as well as the maximum values achieved, of r^2 and q^2 are quoted.

As seen, although many random models achieve high r^2 values, the vast majority of them lead to very

Table 4. Statistical parameters for *P. falciparum* Indo-China W2 clone (1,2,4-Trioxanes)

| No. PCs | r^2 | q^2 | σ_N | PCs used |
|---------|-------|-------|------------|-------------------|
| 1 | 0.524 | 0.446 | 0.771 | 2 |
| 2 | 0.631 | 0.489 | 0.679 | 1, 2 |
| 3 | 0.722 | 0.578 | 0.589 | 1, 2, 7 |
| 4 | 0.757 | 0.589 | 0.550 | 1, 2, 3, 7 |
| 5 | 0.797 | 0.592 | 0.504 | 1, 2, 3, 5, 7 |
| 6 | 0.828 | 0.646 | 0.462 | 1, 2, 3, 5, 7, 11 |

Table 5. Statistical parameters for *P. falciparum* Sierra Leone D6 clone (1,2,4-Trioxanes)

| No. PCs | r^2 | q^2 | σ_N | PCs used |
|---------|-------|-------|------------|-------------------|
| 1 | 0.519 | 0.423 | 0.663 | 2 |
| 2 | 0.575 | 0.427 | 0.623 | 1, 2 |
| 3 | 0.757 | 0.629 | 0.472 | 1, 2, 7 |
| 4 | 0.789 | 0.662 | 0.439 | 1, 2, 7, 11 |
| 5 | 0.795 | 0.621 | 0.433 | 1, 2, 4, 7, 11 |
| 6 | 0.801 | 0.540 | 0.426 | 1, 2, 4, 5, 7, 11 |

Table 6. Statistical parameters for *P. berghei* (1,2,4-Trioxanes).

| No. PCs | r^2 | q^2 | σ_N | PCs used |
|---------|-------|-------|------------|------------|
| 1 | 0.643 | 0.270 | 2.76 | 6 |
| 2 | 0.852 | 0.428 | 1.78 | 4, 6 |
| 3 | 0.929 | 0.708 | 1.23 | 4, 5, 6 |
| 4 | 0.976 | 0.789 | 0.721 | 3, 4, 5, 6 |

Table 7. Statistical parameters for *P. falciparum* (cyclic peroxy ketals)

| No. PCs | r^2 | q^2 | σ_N | PCs used |
|---------|-------|-------|------------|---------------|
| 1 | 0.546 | 0.414 | 0.225 | 2 |
| 2 | 0.592 | 0.418 | 0.214 | 1, 2 |
| 3 | 0.738 | 0.607 | 0.171 | 1, 2, 6 |
| 4 | 0.778 | 0.691 | 0.158 | 1, 2, 3, 6 |
| 5 | 0.795 | 0.656 | 0.151 | 1, 2, 3, 4, 6 |

poor predictive capacity; hence the mean value of q^2 is negative in all cases. However, in a few cases greater q^2 values are obtained, but all of them below the limit of statistical significance ($q^2 > 0.5$), thus conclud-

Table 8. Results of the random test for 100 permutations of the activity vector

| Studied set | Property | Mean r^2 | Max r^2 | Mean q^2 | Max q^2 |
|-------------|---------------------|------------|-----------|------------|-----------|
| 1 | $\log IC_{90}^{W2}$ | 0.308 | 0.628 | -0.267 | 0.358 |
| | $\log IC_{90}^{D6}$ | 0.313 | 0.789 | -0.256 | 0.478 |
| 2 | ED_{90} | 0.477 | 0.890 | -4.746 | 0.397 |
| 3 | $\log IC_{90}$ | 0.215 | 0.541 | -0.435 | 0.090 |

ing that the obtained models are free from fortuitous correlations.

QSAR models interpretation

A further study can be performed to the obtained results by analyzing the obtained equations. The first set is ideal to carry out this analysis due that it is composed by assorted different substitutions. This molecular set, as studied, includes two different biological activities, which correlate fairly well ($r^2 = 0.796$), when applying the logarithmic transformation. This fact explains both the resemblance in the statistical results, and the elections and relevance of the optimal PCs, which were chosen in the same order, see Tables 4 and 5, except for the last one.

The subsequent step involves an analysis of the PCs chosen to see how the molecular point clouds are distributed in the lower dimensional space. As example, the first and second PCs are plotted in Figure 6.

As seen in Figure 6, a clear 4-cluster pattern is present, grouping molecules according to clear substructural features:

- Molecules 1–5, which present small aliphatic, but no aromatic, substitutions, present high activity.
- Molecules 6–9, which have an aromatic substitution and a phenyl group in the region where the two main ring fuse, present low activity.
- Molecular 10–17, which present phenyl groups at both sides of the fusion region, present high or very high activity.
- Molecular 18–20, which have a benzene ring fused to the main backbone, present low activity.

Thus, it can be seen the influence of these phenyl groups in the biological activity and how it is reflected in this simple 2-dimensional space. From this observation, it can be thought that any novel structure falling in the first, or better in the third, cluster would possess

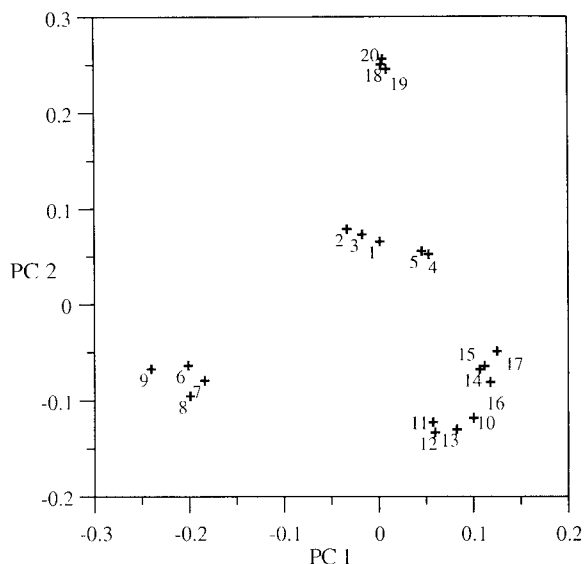


Figure 6. Plot of the first versus de second PC for the molecular set of Table 1.

a high activity against both breeds of *P. Falciparum*. In this way, computational design should be guided to those structures substituted with two phenyl groups at both sides of the ring fusion area and tested with different functional groups.

Conclusions

In the present study, molecular quantum similarity measures have been applied to correlate systematically the antimalarial activity of 1,2,4-trioxanes and cyclic peroxy ketals series. Satisfactory quantitative models have been obtained using little number of descriptors based on Principal Components regressions, achieving also good results in leave-one-out cross validations and random tests.

In addition, a qualitative analysis of the results for the 20 antimalarial set composed by 20 1,2,4-Trioxanes has been carried out, revealing structural information about the data set. The molecules were clustered according to common structural features, which in turn explained the biological activity. When two phenyl substitutions are present in the molecule, it seems that the biological activity tends to increase. Hence, a pattern for computer-aided molecular design is given.

Acknowledgements

This research was partially supported by the European Commission Project number ENV4-CT97-0508 and the CICYT contract SAF2000-0223-C03-01. One of us, X.G., benefited from a predoctoral fellowship from the University of Girona. Financial support from the *Fundacio Maria Francisca de Roviraltà* is also acknowledged. The authors wish to acknowledge the referee's efforts due to their comments have improved the original manuscript in many ways. Lively discussions with Dr Emili Besalú and Mr Lluís Amat are also acknowledged.

References

1. T.D.R. News (News from the WHO Division of Control of Tropical Diseases), 46 (1994) 5.
2. The Malaria Foundation International website, at <http://www.malaria.org>.
3. Bruce-Chwatt, L.J. (Ed.), *Chemotherapy of Malaria*, 2nd ed., W.H.O., Geneva, 1981.
4. Klayman, D.L., *Science*, 228 (1985) 1049.
5. Butler, A.R. and Wu, Y.L., *Chem. Soc. Rev.*, 21 (1992) 85.
6. Shen, C.C. and Zhuang, L.G., *Med. Res. Rev.*, 4 (1984) 47.
7. Jefford, C.W., Velarde, J.A., Bernadinelli, G., Bray, D.H., Warhust, D.C. and Milhous, W.K., *Helv. Chim. Acta*, 76 (1993) 2775.
8. Jefford, C.W., Wang, Y. and Bernadinelli, G., *Helv. Chim. Acta*, 71 (1988) 2042.
9. Testa, B., Kyburz, E., Fuhrer, W. and Giger, R. (Eds.) *Perspectives in Medicinal Chemistry*, Vol. 25, VCH Publishers, Amsterdam, 1992, pp. 459–472.
10. Carbó, R., Arnau, J. and Leyda, L., *Int. J. Quantum Chem.*, 17 (1980) 1185.
11. Carbó-Dorca, R. and Besalú, E., *TEOCHEM*, 451 (1998) 11.
12. Carbó-Dorca, R. and Mezey, P.G. (Eds.) *Advances in Molecular Similarity*, Vol. 2, JAI Press, Greenwich, CT, 1998, pp. 1–42.
13. Carbó-Dorca, R. and Mezey, P.G. (Eds.) *Advances in Molecular Similarity*, Vol. 2, JAI Press, Greenwich, CT, 1998, pp. 43–72.
14. Besalú, E., Carbó, R., Mestres, J. and Solà, M., *Top. Curr. Chem.*, 173 (1995) 31.
15. Carbó, R. (Ed.) *Molecular Similarity and Reactivity: From Quantum Chemical to Phenomenological Approaches*, Kluwer, Amsterdam, 1995, pp. 3–30.
16. Carbó-Dorca, R. and Mezey, P.G. (Eds.) *Advances in Molecular Similarity*, Vol. 1, JAI Press, Greenwich, CT, 1996, pp. 1–42.
17. Carbó, R., Besalú, E., Amat, L. and Fradera, X., *J. Math. Chem.*, 18 (1995) 237.
18. Fradera, X., Amat, L., Besalú, E. and Carbó-Dorca, R., *Quant. Struct.-Act. Relat.*, 16 (1997) 25.
19. Lobato, M., Amat, L. and Carbó-Dorca, R., *Quant. Struct.-Act. Relat.*, 16 (1997) 465.
20. Amat, L., Robert, D., Besalú, E. and Carbó-Dorca, R., *J. Chem. Inf. Comput. Sci.*, 38 (1998) 624.

21. Robert, D., Amat, L. and Carbó-Dorca, R., *J. Chem. Inf. Comput. Sci.*, 39 (1999) 333.
22. Robert, D., Gironés, X. and Carbó-Dorca, R., *J. Comput. Aid. Mol. Des.*, 13 (1999) 597.
23. Gironés, X., Amat, L. and Robert, D. and Carbó-Dorca, R. *J. Comput. Aid. Mol. Des.*, 14 (2000) 477.
24. Gironés, X., Gallegos, A. and Carbó-Dorca, R. *J. Chem. Inf. Comput. Sci.*, 40 (2000) 1400.
25. Robert, D. and Carbó-Dorca, R., *SAR QSAR Environ. Res.*, 10 (1999) 401.
26. Gironés, X., Amat, L. and Carbó-Dorca, R., *SAR QSAR Environ. Res.*, 10 (1999) 545.
27. Carbó-Dorca, R., Robert, D., Amat, L., Gironés, X. And Besalú, E. (Ed.), *Molecular Quantum Similarity and Drug Design*, Springer, Lecture Notes in Chemistry 73, Berlin, 2000.
28. Gironés, X., Robert, D. and Carbó-Dorca, R., *J. Comp. Chem.*, 22 (2001) 255. An operative version can be obtained upon request at quantum@iqc.udg.es.
29. Constants, P. and Carbó, R., *J. Chem. Inf. Comput. Sci.*, 35 (1995) 1046.
30. Amat, L. and Carbó-Dorca, R., *J. Comput. Chem.*, 18 (1997) 2023.
31. Amat, L. and Carbó-Dorca, R., *J. Comput. Chem.*, 20 (1999) 911.
32. Fitting ASA parameters form an assorted collection of basis sets can be downloaded from our web site: <http://iqc.udg.es/cat/similarity/ASA/basisset.html>.
33. Cox, T.F. and Cox, M.A.A. (Eds), *Multidimensional Scaling*, Chapman & Hall, London, 1994.
34. Cuadras, C.M. and Arenas, C.A., *Commun. Stat. Theor. Method.*, 19 (1990) 2261.
35. Allen, D.M., *Technometrics*, 16 (1974) 125.
36. See for example this detailed tutorial by Geladi, P. and Kowalsky, B.R. *Anal. Chim. Acta*, 185 (1986) 1.
37. Van der Waterbeemd, H. (Ed.), *Chemometric Methods in Molecular Design*, VCH, New York, 1995, pp. 309–318.
38. PC Spartan Pro, version 1.4, 1997, Wavefunction Inc, Irvine.
39. Grigorov, M., Weber, J., Tronchet, J.M.J., Jefford, C.W., Milhous, W.K. and Maric, D., *J. Chem. Inf. Comput. Sci.*, 37 (1997) 124.
40. Posner, G.H., O'Dowd, H., Ploypradith, P., Cumming J.N., Xie, S. and Shapiro, T.A., *J. Med. Chem.*, 41 (1998) 2164.