Monatschefte für Chemie 125, 1259–1266 (1994)

Monatshefte für Chemie
Chemical Monthly

Structure-Property Correlations of a Series of Antimalarial Chloroquine Derivatives

D. T. Manallack* and **D. J. Livingstone**

SmithKline Beecham Pharmaceuticals, The Frythe, Welwyn, Herts, AL6 9AR, United Kingdom

Summary. As part of the development of our in-house molecular modelling package, COSMIC, the structure-property relationships of a series of 13 antimalarial chloroquine derivatives have been investigated using physicochemical properties calculated from molecular models. This has involved the use of various statistical techniques. The compounds were analysed using three different modelling approximations concerning assumptions about the invariant amino alkyl side chain. The first two sets of data were obtained from models with and without the side shain using a formal charge of 0.0. Analysis of these data using principal components resulted in plots of principal component scores in which activity categories were separated. In the third data set, the side chain was modelled with a charge of +1 and an examination of the molecular properties showed poorer clustering of activity categories. We have also investigated relationships between biological activity and physicochemical properties using multiple linear regression. Although significant equations were developed, the correlation coefficients were low and it was not felt that they would be useful for quantitative prediction. The identification of important variables, however, may give some clues to the mechanism of action of these compounds.

Keywords. Chloroquine; Antimalarial drugs; Quantitative Structure-Activity Relationships; Principal Components Analysis; Multivariate Analysis; Multiple Linear Regression.

Beziehungen zwischen Strukturen und Eigenschaften einer Reihe von Chloroquinderivaten mit Antimalariawirkung

Zusammenfassung. Die physikalisch-chemischen Eigenschaften von 13 Chloroquinderivaten mit Antimalariawirkung wurden mittels Molekülmodellrechnungen bestimmt und unter Anwendung verschiedener statistischer Analysenmethoden ausgewertet. Die Untersuchungen sind Teil der firmeneigenen Entwicklung eines Software-Paketes für Molecular Modelling mit Namen "COSMIC". Die konstante Aminoalkylseitenkette der Verbindungen wurde in den Modellierungen mit drei verschiedenen Näherungsannahmen berücksichtigt. Für die beiden ersten Datensätze wurde mit ungeladener Seitenkette und ohne Seitenkette gerechnet. Hauptkomponentenzerlegung der Daten ergab in beiden Fällen klare Trennungen hinsichtlich der Aktivitätskategorien. Bei der Berechnung des dritten Datensatzes wurde die Seitenkette mit einer Ladung von +1 modelliert. Die Clusterung der molekularen Daten in Aktivitätskategorien war in diesem Fall wesentlich schwächer ausgeprägt als in den beiden anderen Fällen. Außerdem wurde die Beziehung zwischen der biologischen Aktivität und den physikalisch-chemischen Eigenschaften durch multiple lineare Regressionsanalyse untersucht. Ungeachtet der Möglichkeit einer Herleitung von signifikanten Gleichungen waren die Korrelationskoeffizienten so gering, daß man die Beziehungen als ungeeignet für eine quantitative Vorhersage

ansehen muß. Die Identifikation der wichtigen Variablen kann aber dessenungeachtet nützliche Hinweise auf die Wirkmechanismen dieser Substanzklasse geben.

Introduction

The application of multivariate structure-property correlation techniques in drug design is gaining increased use. To this end, we have an in-house molecular modelling package, COSMIC [1, 2], to which we have added a module, GENPROP [3], designed to simplify and, to a great extent, automate the process of creating data files from sets of modelled drug structures. This approach has proved to be of considerable value in applications to our research programmes [4] and we are thus continually investigating ways in which we can improve the operation of this module. Development of GENPROP comes about both by the examination of literature data sets and by application to in-house problems. One such data set which has been investigated previously by *Rode* and coworkers [5] contains a series of antimalarial chloroquine derivatives. In their analysis *Rode et al.* employed quantum mechanical methods to calculate a set of descriptors concerning the electronic structure of each compound. Multiple linear regression equations were then derived to relate the electronic structure to antimalarial activity. This study, however, only employed imformation generated from molecular orbital calculations. We report here our investigations of this set of compounds using the GENPROP approach which calculates various physicochemical, steric, and electronic properties.

Methods

Thirteen chloroquine derivatives were chosen for analysis from the study by *Rode et al.* [5] whose structures had various substitutions around the 4-aminoquinoline ring. Antimalarial activity of each compound has been perviously repoted from studies using human volunteers [6]. Compounds were classified as either potent (activity > 50), moderate (10 < activity < 50) or weak (activity < 10) (Table 1).

Each compound was constructed and minimized using the COSMIC molecular modelling package. Electronic properties were calculated using the semi-empirical molecular orbital program CNDO [7].

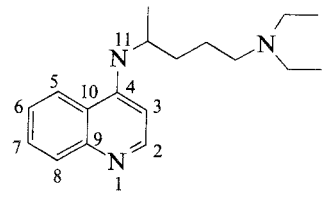
A wide variety of molecular descriptors, comprising whole molecule, atom based and substituent properties, was calculated and collated using the program GENPROP [3]. The atom specific properties were calculated for selected core atoms (indicated by numbers in Table 1) and included partial atomic charges, electrophilic and nucleophilic frontier electron densities, self atom polarizabilities, and nucleophilic and electrophilic superdelocalizabilities. Whole molecule properties included the partition coefficient, molar refractivity, moments of inertia, principal ellipsoid axes, dipole moments and vectors, and the energy of the HOMO and LUMO. The properties calculated for each substituent included data describing the dimensions of the substituent and the sum of their partial charges.

Data handling and analysis was performed using RS1 (BBN Software products UK Ltd., Staines, Middlesex) and the multivariate statistics package, ARTHUR (Infometrix, Inc., Seattle, WA). Calculations were performed on a VAX cluster.

Results and Discussion

The structure of each chloroquine derivative was modelled initially using the 4-aminoquinoline backbone structure (*i.e.* without the aminoalkyl side chain). The rationale for this approach was based on the models described by *Rode* and coworkers [5] who employed the electronic properties of this portion of the

Table 1. Structure and activity of chloroquine derivatives



Compound	Substitution	Activity[a]	Category
1	7-Cl	100	potent
2	7-F	50	potent
3	7-CF_3	50	potent
4	7-OCH_3	14	moderate
5	7-CH_3	7	weak
6	7-H	7	weak
7	6-Cl	100	potent
8	8-Cl	3	weak
9	5-Cl	3	weak
10	7-Cl, 6-CH_3	25	moderate
11	7-Cl, 3-CH_3	15	moderate
12	7-Cl, 2-CH_3	10	moderate
13	7-Cl, 8-NH_2	2	weak

[a] The activity of chloroquine was arbitrarily given the value 100 and each compound was rated relative to this parent compound [6]

molecule. A direct comparison of the performance of the GENPROP module with these models involved just the common 4-aminoquinoline structure. If it is assumed that the omission of the invariant aminoalkyl side chain does not affect the electronic nature of the quinoline ring system, then an additional advantage is gained as the molecular orbital calculations are simplified.

A total of 94 molecular properties were generated using GENPROP for the 13 compounds listed in Table 1. From experience, these wide data sets usually contain parameters with high correlations to other properties in the set. This is known as redundancy, and contraction of the data set may be achieved using statistical procedures which identify these high correlations. Parameters are then removed leaving representative properties of the original data set. In this case we used the program CORCHOP [8] with a correlation coefficient of 0.75. Following this procedure, 45 properties remained which were used for multivariate analysis.

The plots produced from principal components analysis (PCA) of these parameters resulted in some clustering of activity categories which appeared useful (data not shown). A problem that is often encountered using large data sets is that useful patterns may be obscured by way of "noise". In other words, properties unrelated to activity tend to mask those parameters containing useful information. Two methods which have been used to overcome this problem were investigated in this study. The first of these employs a 'supervised' learning technique which chooses those descriptors related to activity (the ARTHUR routine SELECT). SELECT is

effectively a forward stepping regression procedure, however, a decorrelation step is applied to the remaining properties after each property is selected. When applied to the 45 properties chosen above, 11 parameters remained. Using PCA on the 11 selected properties, an improved clustering of activity categories was observed (data not shown). While this selection procedure improved clustering, additional techniques are available that allow simplification of the components which often improves clustering further. One such method is known as Varimax rotation which involves rotation of the components so as to maximize high correlations of descriptors with the principal components [9]. Interpretation of the variable loadings is simpler following these rotations thus rendering them more useful for explaining QSAR's [10]. Fig. 1 shows a plot of PC's 2 and 3 derived from the 11 selected properties following Varimax rotation and demonstrates that each activity group resides in different regions of the PC plot. Interestingly, the features CSe(5) [Se is the electrophilic superdelocalizability of atom 5 (see core atoms Table 1), a theoretical measure of the susceptibility of the atom to attack by an electrophile] and CMR [calculated molar refractivity, a measure of bulk] were found to be related strongly with activity demonstrating that activity is described by both structural and electronic properties. It should be remembered that these properties represent a family of parameters and should not be considered individually as influencing activity.

In the previous study analysing the compounds listed in Table 1, the variables included to develop a QSAR regression model were the partial charges of the atoms in the 4-aminoquinoline nucleus [5]. In our study we have included various properties to describe bulk, lipophilicity and additional electronic parameters. Inclusion of these properties has shown to be very useful and has resulted in plots which could be used to predict the activities of novel chloroquine derivatives. That is, the molecular properties of a novel untested compound may be calculated and simply added as a test compound to the principal components analysis. The location of the test compound on the PC plot relative to the activity category clusters already identified, should provide clues to its activity.

The second phase of our analysis was conducted to examine the assumption that exclusion of the aminoalkyl side chain from the calculations would not greatly affect the electronic properties of the quinoline ring system. To achieve this, the aminoalkyl side chain was constructed on to each molecule, employing the same configuration. The CNDO calculations were repeated and the GENPROP program was used

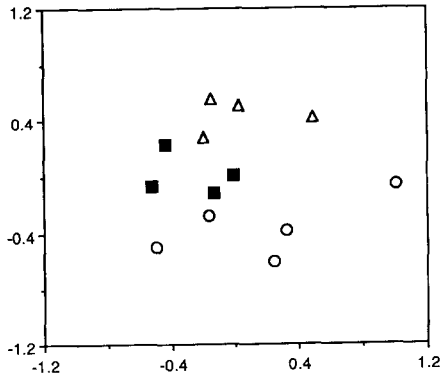


Fig. 1. Plot of the 13 chloroquine derivatives on the second and third principal components from 11 parameters following Varimax rotation; (■) good, (△) moderate, (○) weak

to regenerate the properties listed above. Following CORCHOP and the SELECT procedures, 10 properties remained. Analysis of these properties using PCA resulted in clustering of activity categories (Fig. 2). Once again, features CSe(5) and CMR were found to be strongly related to activity and it appeared therefore that inclusion of the side chain did not affect the clustering observed previously.

There is, however, an additional complication. The aminoalkyl side chain may also exist in the charged state. We therefore remodelled each compound altering the nitrogen in the side chain to be in the charged state (*i.e.* a hydrogen atom was placed on the terminal diethyl nitrogen atom). This would then allow us to examine the molecule in its expected charge state under physiological conditions. The CNDO calculations were repeated using a charge of + 1 for each of the molecules. Following data manipulation to remove redundant parameters and the selection of useful properties, 8 remained for further investigation. PCA was used to display the remaining properties but the plots involving the first three components did not give such pronounced clustering of activity categories as shown in Figs. 1 and 2 (see for example, PC 1 vs. 2, Fig. 3). Among the selected properties shown to be strongly related to activity was CMR, however, the electronic property selected in this analysis of this dataset was the energy of the lowest unoccupied molecular orbital (ELUMO). This suggests that some change in the electronic nature of the quinoline ring may have occurred in the models of the charged species. The CNDO calculations for the charged species were investigated further and compared to the calculations using the uncharged species and compounds modelled without the side chain. The results demonstrated that in models without the side chain and those with an uncharged sidechain the HOMO and LUMO resided on the 4-aminoquinoline ring system. However, when the side chain was charged the LUMO was located on the protonated nitrogen. This change in electronic characteristics altered the molecular orbitals on the 4-aminoquinoline ring system and ultimately their relationship to activity.

From an examination of the principal components plots shown here it would appear that there is little difference between the two data sets produced from the models with and without the aminoalkyl side chain. Inclusion of a charge of + 1 on the aminoalkyl side chain resulted in poorer separation of activity classes (Fig. 3) However, these displays simply give one view of the data sets and it is possible that the use of other analytical methods may show that a particular modelling approach is more appropriate for a problem of this nature. Having carried out PCA it is

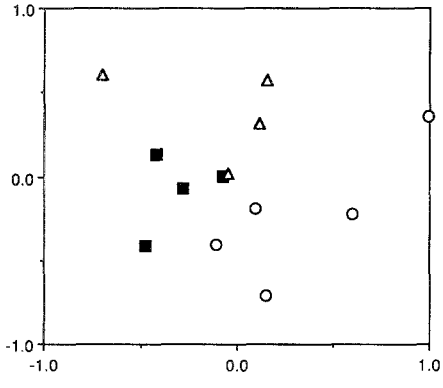


Fig. 2. Plot of the 13 chloroquine derivatives on the second and third principal components from 10 parameters; (■) good, (△) moderate, (○) weak

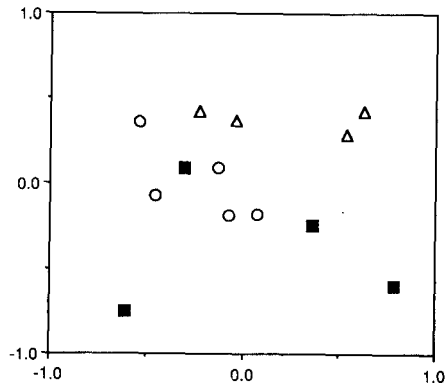


Fig. 3. Plot of the 13 chloroquine derivatives on the first and second principal components from 8 parameters; (■) good, (△) moderate, (○) weak

possible to use the principal components scores as new independent variables in supervised learning [9] procedures such as multiple linear regression (MLR). We carried out MLR using the natural logarithm of activity (as employed by *Rode et al.* [5]) for principal components derived from all three data sets and found the following significant regression equation[1]:

$$\ln(Activity) = -1.74\,PC2 + 2.20\,PC4 + 2.67 \qquad 1$$
$$n = 13 \quad s = 0.90 \quad F = 8.20 \quad R^2 = 0.62$$

This was obtained from the data set using the compounds modelled with the uncharged side chain; none of the other sets yielded significant regression equations. Although Eq. 1 is significant, the correlation coefficient is low and we would not expect that this would be useful in prediction. One reason why the correlation coefficient is so poor may be due to the use of logarithms of what is essentially percentage data (activity has been quoted relative to chloroquine which was arbitrarily given the value 100 [6] and thus the data are scaled between 0 and 100). A commonly adopted procedure with percentage data is to perform a logit transformation [11], but this gave a very similar regression equation.

In their original report, *Rode* and coworkers [5] looked at the relationships between ln(*Activity*) and the partial charges of the ring atoms, calculated dipole moment and the overall charge of the quinoline ring using MLR. They obtained equations containing from 4 terms plus a constant to 9 parameters and a constant, with correlation coefficients ranging from $R = 0.298$ to 0.995. For comparison, we report below the results[1] of our regression analysis involving CNDO derived partial charges for the eleven core atoms (Table 1) and the dipole moment. In each of the three series of modelled compounds a three term equation was obtained, but it was not possible to calculate significant regression equations with a higher number of terms.

No side chain:

$$\ln(Activity) = -31.6(Atom\,8) + 120.3(Atom\,10) + 908.9(Atom\,11) + 229.0 \qquad 2$$
$$n = 13 \quad s = 0.76 \quad F = 9.07 \quad R^2 = 0.75$$

[1] The *t*-statistics for individual coefficients were significant at better than 95% confidence

Side Chain Uncharged:

$$\ln(Activity) = -32.3(Atom\,8) + 130.9(Atom\,10) + 343.2(Atom\,11) + 81.8 \quad 3$$
$$n = 13 \quad s = 0.85 \quad F = 6.72 \quad R^2 = 0.69$$

Side Chain Charged:

$$\ln(Activity) = -16.6(Atom\,8) + 20.2(Atom\,6) + 285.7(Atom\,11) + 63.1 \quad 4$$
$$n = 13 \quad s = 0.91 \quad F = 5.36 \quad R^2 = 0.64$$

Just as for the principal components regression (Eq. 1), the correlation coefficients for these equations are quite low and predictions would be expected to be poor. It was interesting to note that the best MLR models were obtained using the datasets from the uncharged models; these correspond to the most useful plots of principal components scores (Figs. 1, 2). It is also interesting to see that these two equations contain the same terms with very similar coefficients. Thus, as might be expected, the addition of the side chain has had very little effect on the calculated properties. This may be used as an argument for the omission of common features when carrying out modelling calculations, particularly if such calculations involve the use of compute intensive methods such as *ab initio* quantum chemical routines.

The partial charges of atoms 11, the amino nitrogen of the 4-aminoquinoline nucleus, 10, a bridgehead carbon, and 8, a 'benzene' ring carbon, were selected as useful contributing variables in these equations. This may provide some useful clues to the mode of action of these compounds; for example, it may indicate those atoms which take part in binding interactions. A note of caution should be sounded here as these properties are related to others in the set and it may be these properties which are important determinants of the antiparasitic mechanism. Equation 4, derived from the data set in which the side chain was modelled with a formal charge of +1.0, is seen to have the lowest correlation coefficient. This result is consistent with the principal components scores plot (Fig. 3) in which the clustering of activity categories is less well defined than Figs. 1 and 2. Addition of the formal charge to these models has resulted in the selection of a different 'benzene' ring carbon (6 for 10) but retention of the partial charges of atoms 11 and 8. It is difficult to draw a general conclusion from this result as to whether compounds should be modelled with a formal charge when the pK_a of functional groups suggest that they would exist in the charged state at neutral *pH*. An argument may be put forward that the *pH* of the micro-environment of a binding site is significantly different to that occurring in biological tissues. Alternatively, it may also be argued that a binding site is likely to contain charges which will interact with the formal charge and thus counteract the effect of a formal charge on the computed properties. Thus, ignoring formal charges may be the most appropriate way to model biologically active compounds.

Conclusions

These results serve to highlight two issues concerning QSAR studies. Firstly, care should be exercised when considering whether to use the entire structure of a

compound or to perform calculations on a substructure. The danger of using a substructure is that long range effects, although assumed not to influence the results of molecular orbital calculations, may be missed. It appears from the work reported here that omission of the side chain has little or no effect on the relationships found. In addition, the state of formal charge of a molecule should be carefully considered. In the case of the compounds reported here it would appear that the addition of charge to the models results in data sets which were less well able to describe the biological data.

Acknowledgements

The authors would like to thank *M. R. Saunders* for his helpful discussions.

References

[1] Vinter J. G., Davis A., Saunders M. R. (1985) J. Comput.-Aided Mol. Design **1**: 31
[2] Morley D., Abraham R. J., Haworth I. S., Jackson D., Vinter J. G., Saunders M. R. (1991) J. Comput.-Aided Mol. Design **5**: 475
[3] Livingstone D. J., Evans D. A., Saunders M. R. (1992) J. Chem. Soc. Perkin Trans. II 1545
[4] Camilleri P., Livingstone D. J., Murphy J. A., Manallack D. T. (1993) J. Comput.-Aided Mol. Design **7**: 61
[5] Rode B. M., Schwendinger M. G., Kokpol S. U., Hannongbua S. V., Polman S. (1989) Monatsh. Chemie **120**: 913
[6] O'Brien R. L., Hahn F. E. (1965) Antimicrobiol Agents and Chemotherapy **5**: 315
[7] Pople J. A., Beveridge D. L. (1970) Approximate molecular orbital theory. McGraw-Hill, New York
[8] Livingstone D. J., Rahr E. (1988) Quant. Struct.-Act. Relat. **7**: 103
[9] Livingstone D. J. (1991) Meth. Enzymol. **203**: 613
[10] Harmon H. H. (1976) Modern factor analysis. University of Chicago Press, Chicago, Illinois, pp 487
[11] logit(Activity) = log(Activity + 0.5/100.5 + Activity)

Received February 19, 1994. Accepted May 22, 1994