Asian Journal of Chemistry Vol. 16, No. 1 (2004), 67–73

3D-QSAR Analysis of Some Cinnamic Acid Derivatives as Antimalarial Agents

A.K. GUPTA*, L.K. SONI, P. HANUMANTHARAO, S.V. SAMBASIVARAO, M. AROCKIA BABU and S.G. KASKHEDIKAR
Department of Pharmacy, Shri Govindram Seksaria Institute of Technology and Science 23, Park Road, Indore-452 003, India
E-mail: arunkg_73@rediffmail.com

A set of sixteen cinnamic acid derivatives with antiprotozoan activity against *Plasmodium falciparum* was subjected to the three dimensional quantitative structure activity relationship studies using various combinations of thermodynamic, electronic and spatial descriptors. Among several 3D quantitative structure activity relationship models, one model gave good statistical significance ($r > 0.83$, F-test = 9.47, $S < 0.33$, chance correlation < 0.01 and cross validated $r^2 > 0.43$) involving the descriptors, *viz.*, van der Waals' 1,4 energy, energy of lowest unoccupied molecular orbital and principal moment of inertia of x-component. The predictive power of this model was discerned by successfully testing the four compounds constituting the test set with predictive $r^2 > 0.61$. Thus, this validated model brings important structural insight to aid the design of novel antimalarial agents.

Key Words: 3D-QSAR, Antimalarial, Cinnamic acid derivatives.

INTRODUCTION

Malaria was and still is recognized as a disease of worldwide incidence. It is a highly infectious disease caused by several species of the genus *Plasmodium*, a protozoan parasite that is transmitted to humans by *Anopheles* mosquitoes. *Plasmodium falciparum* is the most virulent human malaria parasite and is responsible for higher death rates, particularly in tropical and subtropical regions like Africa, Asia and South America[1]. Despite extensive efforts to eradicate the insect vector using insecticides and the development of several types of synthetic anti-malarial agents, the incidence of malaria is still increasing in large parts of the world due to the development of resistance in parasites and mosquitoes to available drugs and insecticides[2, 3].

Thus, there is great need for novel antimalarial agents and insecticides, ideally with modes of action and chemical structures different from the currently used compounds, and hence, to study the quantitative structure activity relationships[4, 5] (QSAR) of some cinnamic acid derivatives as anti-malarial compounds so that associated molecular properties could be identified and exploited to optimize anti-malarial activity.

EXPERIMENTAL

The antimalarial data of cinnamic acid derivatives (Tables 1 and 2 and Fig. 1) were taken from reported work of Wiesner *et al.*[6,7] (excluding compounds with biological activities numerically not well defined). The derivatives are divided into two sets, *viz.*, training set of 16 compounds and test set of 4 compounds using random selection method. All the biological activity data (IC_{50} in μM) have been converted to negative logarithmic mole dose (pIC_{50}) for QSAR analysis. For molecular modeling and calculation of various descriptors, we have used different modules provided in software[8]. All structures of the compounds (**1** to **16** in Table-1 and **T-1** to **T-4** in Table-2) were built using molecular stetching facility provided in the modeling environment of the software. The energy minimization was carried out by fixing maximum iteration limit 1000 and RMS gradient value less than 0.001 kcal/mol/Å using Austin model (AMI) Hamiltonian approximation method[9]. The geometry optimization of the lowest energy structure was carried out using MOPAC. All the descriptor calculations were performed on this geometrical optimized structure. The following thirty-two descriptors[10–13] were calculated for QSAR analysis: bend energy (E_b), heat of formation (H_f), principal moment of inertia—x-axis (PMIX), principal moment of inertia—y-axis (PMIY), principal moment of inertia—z-axis (PMIZ), boiling point (BP), Henry's law constant (H), critical pressure (P_c), HOMO energy (HOMO), LUMO energy (LUMO), critical temperature (T_c), ideal gas thermal capacity (C_p), repulsion energy (NRE), critical volume (V_c), standard Gibb's free energy (G), Connolly accessible area (SAS), log P, stretch energy (E_s), Connolly molecular area (MS), melting point (mp), stretch bend energy (E_{sb}), Connolly solvent excluded volume (CSEV), molar refractivity (MR), torsion energy (E_t), dipole (DPL), molecular weight (mw), total energy (E), dipole-dipole energy (DDE), non-1,4-van der waals' energy (E_v), van der waals'-1,4-energy (VDWE14), electronic energy (ElcE), ovality exact mass (Mass), partition coefficient (PC).

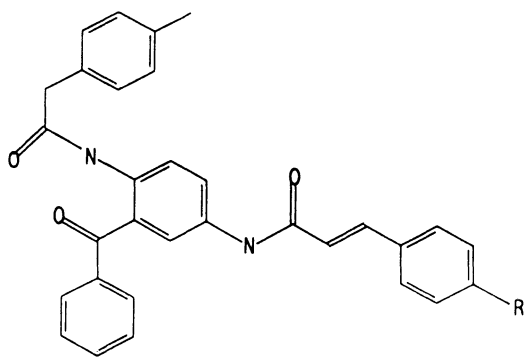

Fig. 1. Cinnamic acid analogs used in the present study

TABLE-1
COMPARISON OF OBSERVED AND LEAVE ONE OUT PREDICTED pIC_{50} VALUE OF COMPOUNDS USED IN TRAINING SET

Compound No.[a]	Substitution	$IC_{50}^{b}(\mu M)$	Observed pIC_{50}	Predicted (LOO)[c] pIC_{50}
1[d]	—	2.70	5.569	5.742
2	—H	5.80	5.237	5.620
3	$-NO_2$	6.50	5.187	4.675
4	$-COOCH_3$	1.00	6.000	5.629
5	$-CF_3$	5.70	5.244	5.330
6	—Br	3.20	5.495	5.519
7	$-NH_2$	5.50	5.260	5.060
8	$CH{=}C(CN)_2$	4.30	4.367	5.187
9	$-CH_3$	1.40	5.854	5.430
10	$-OCH_3$	1.30	5.886	5.998
11	$-C_2H_5$	1.20	5.921	5.529
12	$-CH(CH_3)_2$	1.20	5.921	6.081
13	$-C(CH_3)_3$	3.00	5.523	5.993
14	$-OC_3H_7$	0.20	6.699	6.112
15	$-OC_4H_9$	1.10	5.959	6.440
16	$-OCH_2C_6H_5$	0.50	6.301	6.144

[a]Data for compounds **1–15** and **16** have been taken from references 6 and 7 respectively.
[b]Inhibitory activity against intraerythrocytic form of the *P. falciparum* strain Dd2 using a semi-automated micro dilution assay.
[c]Leave one out (LOO) method
[d]N-[3-Benzoyl-4(2-*p*-tolyl-acetylamino)-phenyl]-3-phenyl propionamide.

TABLE-2
COMPARISON OF OBSERVED AND PREDICTED pIC_{50} VALUE OF COMPOUNDS USED IN TEST SET

Compound No.[a]	Substitution	IC_{50}^{b} (μM)	Observed pIC_{50}	Predicted pIC_{50}
T-1	—CHO	4.00	4.398	5.154
T-2	—Cl	5.50	5.260	5.529
T-3	$-OC_2H_5$	0.85	6.071	5.814
T-4	$-C_6H_5$	0.65	6.187	5.881

[a]Data for compounds **T-1** to **T-3** and **T-4** have been taken from references 6 and 7 respectively.
[b]Inhibitory activity against intraerythrocytic form of the *P. falciparum* strain Dd2 using a semi-automated micro dilution assay.

All the calculated parameters were considered as independent variables and biological activity was taken as dependent variable. Stepwise regression analysis method[14, 15] was used to develop equations; following statistical parameters were considered to select best QSAR model: coefficient of correlation (r), standard deviation (s) and sequential Fischer test (F-test). In addition to this, the best QSAR model was subjected to internal validation *via* leave one out cross validation process, bootstrapping and randomization test[16], further validated by external means using test set.

RESULTS AND DISCUSSION

When training set was subjected to stepwise multiple linear regression analysis, in order to develop QSAR model, various statistical equations with different values of coefficient of correlation (r), standard deviation (s) and F-test were obtained.

$$pIC_{50} = 0.107244*VDW14E - 0.205987*DDE + 0.056215*E_b + 1.689183$$

$$n = 16,\quad r = 0.87,\quad s = 0.30,\quad F = 12.75 \quad \text{and} \quad r^2_{cv} = 0.49 \qquad (1)$$

$$pIC_{50} = 0.146606*VDW14E - 0.139111*DDE + 1.792659*HOMO + 16.858826$$

$$n = 16,\quad r = 0.86,\quad s = 0.31,\quad F = 11.04 \quad \text{and} \quad r^2_{cv} = 0.18 \qquad (2)$$

$$pIC_{50} = 0.275909*VDW14E + 3.011498*HOMO - 0.014593*CSEV + 31.982547$$

$$n = 16,\quad r = 0.86,\quad s = 0.31,\quad F = 11.74 \quad \text{and} \quad r^2_{cv} = 0.51 \qquad (3)$$

$$pIC_{50} = 0.123048*VDW14E + 4.170752*HOMO - 0.033016*H + 38.895878$$

$$n = 16,\quad r = 0.85,\quad d = 0.32,\quad F = 10.32 \quad \text{and} \quad r^2_{cv} = 0.14 \qquad (4)$$

$$pIC_{50} = 0.172130*VDW14E - 0.161392*DDE - 0.000177*PMIX + 1.791083$$

$$n = 16,\quad r = 0.84,\quad s = 0.33,\quad F = 9.96 \quad \text{and} \quad r^2_{cv} = 0.29 \qquad (5)$$

$$pIC_{50} = 0.155327*VDW14E - 0.145298*DDE + 0.318179*LUMO + 1.92239$$

$$n = 16,\quad r = 0.84,\quad s = 0.32,\quad F = 9.88 \quad \text{and} \quad r^2_{cv} = 0.28 \qquad (6)$$

$$pIC_{50} = 0.143391*VDW14E + 0.548392*LUMO - 0.000262*PMIX + 4.342180$$

$$n = 16,\quad r = 0.83,\quad s = 0.33 \quad \text{and} \quad F = 9.47 \quad \text{and} \quad r^2_{cv} = 0.43 \qquad (7)$$

All the above equations were screened on the basis of inter-correlation within the descriptors (> 0.3 as mentioned in Table-3), leave one out cross validated squared correlation coefficient (<0.4) and intercept of best fit line. The equations (1) to (6) did not comply. Hence, eqn. (7) was considered as best model with $r > 0.83$, standard deviation < 0.33, chance correlation < 0.01 and better statistical significance > 99% with $F_{(3, 16)} = 9.47$ against the value for 99% significance ($F_{3, 16\,\alpha\,0.01} = 5.29$).

TABLE-3
CORRELATION MATRIX OF PARAMETERS USED IN EQUATIONS

	VDW14E	DDE	Eb	LUMO	HOMO	PMIX	CSEV	H
VDW14E	1.000							
DDE	0.088	1.000						
Eb	0.692	0.111	1.000					
LUMO	0.239	0.608	0.093	1.000				
HOMO	0.307	0.514	0.298	0.909	1.000			
PMIX	0.074	0.503	0.156	0.292	0.381	1.000		
CSEV	0.894	0.214	0.532	0.106	0.245	0.168	1.000	
H	0.156	0.311	0.166	0.457	0.507	0.024	0.268	1.000

The model was tested for outlier by Z-score method; no compound was found to be outlier. To ascertain the predictivity of the model, internal validation using leave one out cross validation process, bootstrapping technique and randomization test were performed. The satisfactory values of internal validation, cross validated squared correlation coefficient (r^2_{cv}) > 0.43, standard deviation of prediction (S_{press}) = 0.45, standard deviation of error of predictions (S_{DEP}) = 0.39, bootstrapping squared correlation coefficient (r^2_{bs}) = 0.72 and chance correlation < 0.01 in the randomized biological activity test revealed that the results were not based on chance correlation. The model's r^2_{cv} > 0.43 supported the predictive ability and significance of the model (Fig. 2 and Table-1). The r^2_{bs} was at par with r^2, supported the robustness of the model, as well as indicated that, no single compound of the series contributed much more to the model. The validity of this model was checked using the test set of 4 compounds that were not included in model development. The models having predictive squared correlation coefficient (r^2_{pred}) > 0.61 were in agreement with the accepted criteria as more than 0.3 (Table-2 and Fig. 3).

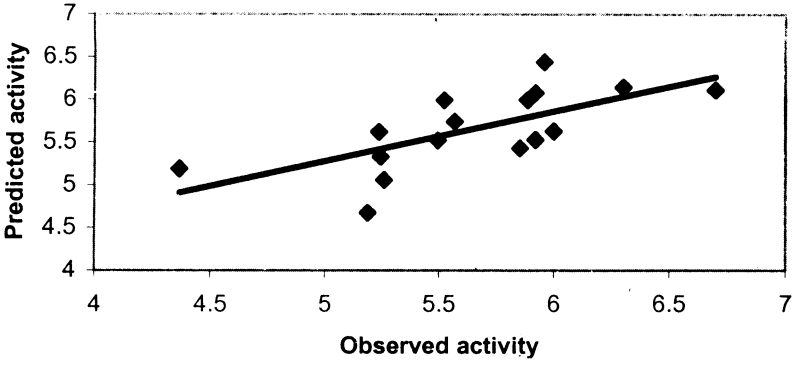

Fig. 2. A plot of observed pIC_{50} *vs.* predicted (leave one out) pIC_{50} for training set

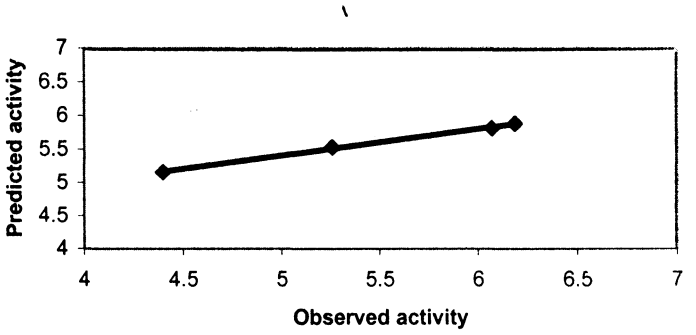

Fig. 3. A plot of observed pIC_{50} *vs.* predicted HOMO pIC_{50} for test set

The above studies reveal that the VDW14E, LUMO descriptors are positively contributing whereas PMIX is negatively contributing to the antimalarial activity of cinnamic acid derivatives, out of which LUMO, that is indicative of π-bonding interaction of species crucial for the reactivity and PMIX, describes mass distribution over the molecule on x-component in spatial arrangement. VDW14E is sum of pairwise van der Waals interaction energy terms for atoms separated by exactly three chemical bonds which explains the depth of the attraction potential energy well and how easy it is to push atoms together.

From the above analysis, it may be inferred that eqn. (7) can be used for theoretical prediction of antimalarial activity of the new molecules.

Conclusion

The study identifies common features responsible for variation in the antimalarial activity for cinnamic acid derivatives. Although substitution with groups which increase LUMO energy and also groups which result in increased 1,4 dihedral van der Waals interaction while substitutions with decreasing steric effect enhance the activity, the optimum combination of these three will yield better inhibitory activity. Thus, 'R' substitutions in phenyl ring may be helpful for designing novel molecules with better antimalarial activity prior to synthesis.

ACKNOWLEDGEMENT

The authors thank the Director, S.G.S.I.T.S., Indore for providing facilities for this work.

REFERENCES

1. D. Wirth, *Nature Med.*, **4**, 1360 (1998).
2. F. Chandre, F. Darriet, M. Darder, A. Cuany, J.M. Doannio, N. Pasteur and P. Guillet, *Med. Vet. Entomol.*, **12**, 359 (1998).
3. R.P. Penilla, A.D. Rodriguez, J. Hemingway, J.L. Torres, J.I. Arredondo-Jimenez and M.H. Rodriguez, *Med. Vet. Entomol.*, **12**, 217 (1998).
4. C. Hansch, Comprehensive Medicinal Chemistry, Vol. 4, Pergamon Press, New York, p. 579 (1990).

5. H.J. Smith, and H. Williams, Introduction to Principle of Drug Design, John Wright & Sons Ltd., Bristol, p. 216 (1983).
6. J. Wiesner, A. Mitsch, P. Wißner, H. Jomma and M. Schlitzer, *B.M.C.L.*, **11**, 423 (2001).
7. J. Wiesner, P. Wißner, H. Dashe, H. Jomma and M.J. Schlitzer, *Bioorg. Med. Chem.*, **9**, 785 (2001).
8. C.S. Chem. Office, Version 6.0, Cambridge Soft Corporation, Software Publishers Association, 1730 M Street, NW, Suite 700, Washington D.C. 20036 (202) 452–1600, USA.
9. L.B. Kier, Molecular Orbital Theory in Drug Research, Acadmic Press, New York, p. 62 (1971).
10. C. Hansch and A. Leo, Substituent Constants for Correlation Analysis in Chemistry and Biology, John Wiley & Sons, New York, p. 114 (1979).
11. W.J. Dann, *Eur. J. Med. Chem.*, **12**, 109 (1977).
12. H. Van de Waterbeemal and B. Testa, *Adv. Drug Res.*, **16**, 85 (1987).
13. J.A. Pople and G.A. Segal, *J. Chem. Phys.*, **44**, 3289 (1966).
14. N.R. Draper and H. Smith, Applied Regression Analysis, John Wiley & Sons, New York, p. 193 (1981).
15. J.E. Freund, Mathematical Statistics, Prentice-Hall of India Pvt. Ltd., New Delhi, p. 494 (1991).
16. H. Kubinyi, Hansch Analysis and Related Approaches, Weinheim, New York, p. 91 (1993).

(*Received*: 3 March 2003; *Accepted*: 20 August 2003) AJC-3145