Pure & Appl. Chem., Vol. 60, No. 2, pp. 265–270, 1988.
Printed in Great Britain.

Computer automated structure evaluation (CASE) of flavonoids as larval growth inhibitors

Gilles Klopman and Ruben E. Venegas

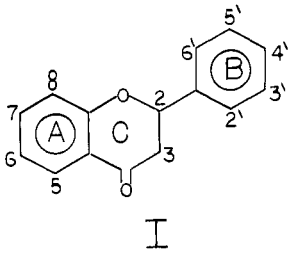
Department of Chemistry, Case Western Reserve University, Cleveland, OH 44106, USA

Abstract : The Computer Automated Structure Evaluation program, CASE, has been used to analyze the inhibition of the larval growth in the H. Zea system by flavonoid derivatives. A quantitative structure-activity relationship was established between molecular fragments generated by the CASE methodology and the activity of the flavonoids. A good correlation with experimental values is obtained.

INTRODUCTION

Although flavonoids (I) have been known to exist in plants as long as alkaloids, by comparison, their pharmacological effects and potential medicinal uses have attracted little attention. Alkaloids often are toxic and their profound physiological properties were known and explored by man even before they were isolated and their structure elucidated. By contrast, the flavonoids are not poisonous and, as dietary ingredients, may sometimes be beneficial. Perhaps another reason for the neglect of the flavonoids was the controversial proposal, in 1936, that citrus bioflavonoids (ref. 1) was an essential vitamin, Vitamin-P. The result of the unfavorable publicity that followed this unfounded claim probably discouraged further medical experiments at the time.

The interest in flavonoids has increased in the last decade, especially as a pharmacologic agent having significant activity in a variety of systems. Some plant flavonoids affect insect behavior, development and growth. Others are responsible for resistance of plants to insect attack. Flavones, glycosides and aglycones in the cotton plant are larval growth inhibitors for Heliothis Zea, Heliothis virescens Fabricius, and Pectinophora gossypiella Saunders (refs. 2,3). Flavonoid polymers are insect resistance factors of oak tree, and corn plant (refs. 4,5).



I

Elliger et al. (ref. 6) analyzed a group of 42 flavonoids for antigrowth activity toward H. Zea, and evaluated them with respect to structural features affecting activity. They found that in general, an adjacent (ortho) substitution of phenolic hydroxyl groups was necessary for inhibitory activity. However the fact that the absence of ortho-hydroxylation does not imply lack of activity could not be explained. Moreover, vicinal hydroxylation alone is not sufficient to produce activity. The analysis presented by Elliger et al. was purely empirical and qualitative and was based on a simple examination of the data base. It does not provide much insight into possible quantitative relationships that may exist between structure and activity.

We have applied successfully our Computer Automated Structure Evaluation (CASE) program to a large number of biological systems (refs. 7-14), and felt that it could provide some insight into the data base presented by Elliger et al. (ref. 6) and help identify the features that are responsible for the antigrowth activity of flavonoid derivatives in the H. Zea system.

METHODOLOGY

The CASE methodology has been described in detail elsewhere (refs. 7-14) and, therefore will only be discussed briefly here. The input to the computer program consists of the structural formulae of the compounds under analysis and, for each of them, an index of their biological activity. The structures are entered in the program using the KLN code as described previously (ref. 15). Each molecule is fragmented into subunits containing 3 to 10

heavy atoms. Each fragment is labeled positive if it belongs to an active molecule and negative if it belongs to an inactive one. When all the compounds have been entered, a statistical analysis of the occurrence of the fragments is performed and significant subunits are selected automatically as potential descriptors of the activity under investigation.

The program also has the ability to perform a quantitative analysis of the data. After removing all the embedded and redundant fragments, a stepwise linear least squares regression analysis can be performed in which the biological index is the dependent variable and the molecular fragments as well as the Log P (partition coefficient in octanol/water system) are the independent variables. The main goal of this procedure is to identify those fragments (together with Log P) that will be the best descriptors in a QSAR equation of the form:

$$\text{Activity} = a + b_i(n_i F_i) + c \text{ Log P} + d (\text{Log P})^2 \quad (1)$$

where a, b, c, and d are regression coefficients, n is the number of times that fragment F_i appears in the molecule. Log P is the partition coefficient (octanol/water) of the molecule, calculated internally by the program, using the Charge Density Method as described previously (refs. 13,16).

The newest feature of the CASE program is its capability to "autodesign". This procedure, as the rest of the data manipulation, is completely automated. The user only has to input a trial structure (this can be the most active compound of the data base for example), and the computer will search through all the descriptors that have been identified to be related to the activity and try to "build" the most effective compound of the data base. A specific fragment (descriptor) is added/replaced into the trial molecule if it leads to improved activity. This process continues until no more enhancement in the activity can be made, and the optimal structure has been reached.

RESULTS AND DISCUSSION

The 42 compounds used in this study along with their experimental activities are presented in Table 1. The results are expressed as the Log of the effective concentration of additive necessary to reduce larval growth to 50% of control values (Log ED_{50}). The concentrations are entered on a molar basis. When Log ED_{50} is less or equal to 1.0 the compound is considered as active. Thus 20 compounds (1 to 20 in Table 1) are listed as actives, and 22 compounds (21 to 42 in Table 1) are listed as inactives. All the compounds are flavones except Orobol (compound number 5) and Pomiferin (compound number 42) which are isoflavones. Four randomly selected compounds were withheld from the calculations in order to test the predictive power of the regression equation generated by the program.

TABLE 1. Flavonoids evaluated by CASE method.

	Compound	Log ED_{50}[a]		Compound	Log ED_{50}[a]
1	Fisetin	1.000	22	7-Hydroxyflavone	> 1.0
2	5,7,2',3'-Tetrahydroxyflavone	0.875	23	Chrysin	> 1.0
3	Luteolin	0.732	24	Primetin	> 1.0
4	Eriodictyol	0.792	25	3'4'-Dihydroxyflavone	> 1.0
5	Orobol[b]	0.806	26	Acacetin	> 1.0
6	Iso-orientin	0.477	27	Apigenin	> 1.0
7	Maysin	0.415	28	Isovitexin	> 1.0
8	Maysin-3'-methyl ether	0.699	29	Naringenin	> 1.0
9	Scutellarein	0.602	30	Naringin	> 1.0
10	Isoscutellarein	0.792	31	Hesperetin	> 1.0
11	Hypolaetin 3'4'-dimethylether	0.845	32	Neohesperidin	> 1.0
12	Tricetin	0.748	33	Resokaempferol	> 1.0
13	Robinetin	0.602	34	Isoscutellarin-4'methyl ether	> 1.0
14	Quercetin	0.544	35	Chrysoeriol	> 1.0
15	Dihydroquercetin	0.544	36	Luteolin 3'4'dimethyl ether	> 1.0
16	Quercitrin	0.653	37	Norartocarpetin	> 1.0
17	Astilbin	0.756	38	5,7,2',5'-Tetrahydroxyflavone	> 1.0
18	Rutin	0.602	39	5,7.3',5'-Tetrahydroxyflavone	> 1.0
19	Catechin	0.716	40	Kaempferol	> 1.0
20	Myricetin	0.491	41	Morin	> 1.0
21	Primuletin	> 1.0	42	Pomiferin[b]	> 1.0

[a] The activity is expressed as the Log ED_{50}, the concentration of additive necessary to reduce larval growth to 50% of control values. If Log ED_{50} < 1.0 active, otherwise inactive. For details concerning the experimental assay see ref. 6.

[b] Isoflavone.

The first step in our analysis is to translate the molecular structures into their KLN code (ref. 15) and label them with an activity index. This index was generated for the active molecules by the following "ad hoc" equation :

$$\text{CASE activity} = 98 - 68\ (\text{Log } ED_{50}) \qquad (2)$$

The inactive molecules are assgined a value of 10 and the actives, value of 30 or higher. The higher the value, the higher activity.

After the training set was entered in the computer, the program performs a rather complex statistical analysis. The CASE program generates the molecular fragments believed to be responsible for the observed activities and the Log P values for each molecule. In this database, 6 uncorrelated molecular fragments were selected as potential descriptors of the inhibition observed in the H. Zea system. Based on these, a stepwise regression analysis was performed. Four descriptors were selected to be particularly significant to the actual potency of the flavonoids. The partition coefficient (Log P) in octanol/water was not found by the CASE program to be an important factor in the description of the observed activity, indicating that Log P does not play an important role in this system. Elliger et al. (ref. 6) suggested that biological effectiveness is not entirely a function of hydrophilic/lipophilic character as expressed by Hansch (ref. 17), whilst such a correlation was proposed to explain the behavior of simpler phenols (ref. 18). Furthermore, Haslem et al. (ref. 19) found in studies with polyphenols interacting with proteins that the enthalpic and entropic terms are negative and they suggested that this, in turn, indicates that hydrophobic interactions cannot be predominant.

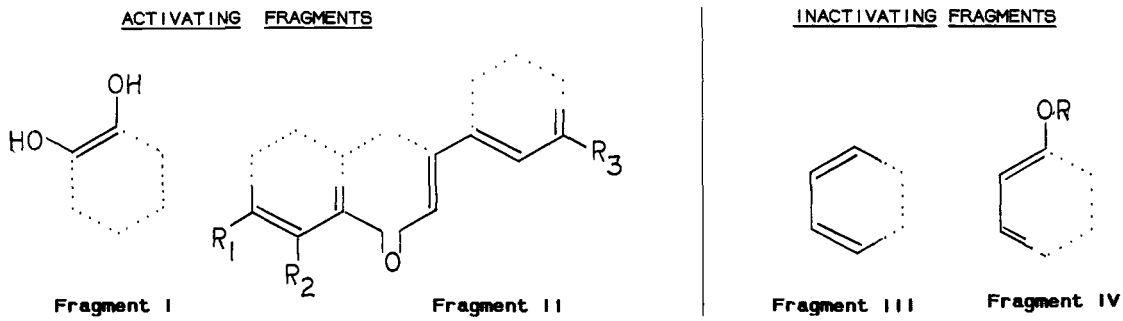


Fig. 1. The activating and inactivating fragments to be believed as important descriptors in antigrowth activity of flavonoids. R_i are any substituents other than H.

The four molecular descriptors selected by the program are shown in Fig. 1. Fragment I and Fragment II are activating (Biophores) whereas Fragment III and Fragment IV are deactivating (Biophobes). Based on these, the following Quantitative Structure-Activity Relationship (QSAR) equation was generated for the activity of the flavonoids as inhibitors of larval growth:

$$\text{Calcd} = 12.79 + 10.21n_I(F_I) + 14.56n_{II}(F_{II}) - 31.72n_{III}(F_{III}) - 11.73n_{IV}(F_{IV}) \qquad (3)$$

where n_i are the number of times fragment F_i appears in the molecule.

With this equation we can explain more than 90% of the variation of the database. The F-test for regression is satisfied at the 0.05 confidence limit; $F(4,33,0.05) = 81.55$, with a correlation coefficient r^2 of 0.91 and a standard deviation of residuals of 10.25. The correlation is significant since the F-test is substantially better than required by our criteria to weed out chance correlations (ref. 20).

It can be seen from Table 2 that the calculated values using eq. 3 seem to be fairly good. All the active compounds are accounted for correctly with only small deviations between actual and calculated activity. Among the inactives, the only compound that was substantially incorrect is Pomiferin (compound number 42). This compound is calculated to be very active, but shows no experimental activity. Pomiferin, as pointed out above, is one of the two isoflavones in a data base consisting mostly of flavones.

TABLE 2. Experimental and Calculated[a] activities of the compounds submitted to CASE analysis[a].

Compound	Actual	Calcd.[b]	Compound	Actual	Calcd.[b]
1	++	+++	23	-	-
2	++	++	24	-	-
4	+++	+++	25	-	-
5	+++	+++	26	-	-
6	++++	++++	27	-	-
7	++++	++++	28	-	-
8	++++	+++	29	-	-
9	++++	+++	30	-	-
10	+++	++	31	-	-
11	+++	++++	32	-	-
12	+++	++++	33	-	-
13	++++	++++	34	-	-
14	++++	+++	35	-	-
15	++++	+++	36	-	-
17	+++	+++	37	-	-
18	++++	+++	39	-	-
19	+++	+++	40	-	-
20	++++	++++	41	-	-
22	-	-	42	-	+++

[a] ++++ : extremely active; +++ : very active; ++ : active; + : marginally active; - : inactive.
[b] Using Equation 3.

Fragment I is the most prevalent. With it, about 60% of the data base can be explained. It is present in 20 molecules, of which 17 are active and 3 inactive. This fragment represents the ortho quinolinic group, claimed to be responsible for the activity of the flavonoids by Elliger et al. (ref. 6). However, as was pointed out in their work, the observation of their presence is not sufficient to produce significant conclusions regarding the mechanism of action of flavonoids. As is shown in eq. 3, the structural features responsible for activity seem indeed to be more complex. The second most important descriptor is Fragment II. With Fragments I and II together, about 75% of the variation of the database can be explained. The structure of Fragment II seems to indicate that coplanarity of the flavonoid ring system may impart significant activity to the flavones in which it exists. As can be seen from structure I, unsaturation at C_2-C_3 bond forces the molecule to be planar and the C ring becomes conjugated with the A and B rings. Saturation of this bond disrupts the planarity and conjugation in the C ring. Elliger et al. (ref. 6) believed that coplanarity is not a requirement for activity. They based their conclusions on two observations. Firstly, Catechin (compound 19) lacking the 2,3-double bond has the same activity as Luteolin (compound 3), which contains the C_2, C_3 unsaturation and secondly, 2,3-dihydro compounds such as 4, 15, 17, and 19 that are non planar show significant activity. However, since compounds 29, 30, 31 and 32 which also have the 2,3-dihydro bond, are inactives, they concluded that the coplanarity of the system is not related to the activity of the flavonoids. The question then, is why do we find Fragment II to be related to activity ? What we find is that even though Luteolin and Maysin (compound 7) are both planar, the difference in their activities is significant (Log ED_{50} are 0.73 and 0.42 for Luteolin and Maysin respectively). The only differences between these compounds is that Luteolin is unsubstituted at position 6, and thus does not contain Biophore II. It can be argued that the substitution/unsubstitution at position 6 is the reason for the activity/inactivity of the molecules. However, as can be seen from Table 2, isovitexin (compound 28) is inactive and iso-orientin (compound 6) is very active, although they both are planar and substituted at position 6 (actually the substituent in both molecules is the same sugar, glucosyl). The key differences between these two compounds is that the first one is unsubstituted at position 3', and as a result does not contain Fragment II. The two most actives molecules (compound 6 and 7) of the data base contain Biophore I and II. Compound 8, Log ED_{50} = 0.70, contains only Biophore II and lacks the ortho hydroxyl group. Thus, in conclusion, we can say that coplanarity alone is not absolutely required for the activity but, when it exists at the same time as substitutions at positions 5, 6 and 3', it allows the molecule to interact with the receptor in a more efficient way. It can thus be postulated that a possible mechanism of action can involve the simultaneous participation of the three sites of the molecule.

The presence of Biophobe III appears to be very important and strongly deactivating. It can be seen from the magnitude of its regression coefficient (eq. 3) that even a molecule containing both activating fragments (Fragments I and II) will be predicted to be inactive if

Fragment III is also present. Thus, it is not surprising to find that compound 25, although possessing an ortho quinolinic group (Frag. I) has no activity at all. This descriptor (Biophobe III) indicates that a string of four unsubstituted aromatic carbon atoms, whether occurring in the A or B ring system will prevent activity. A possible explanation for this observation is that an alternative metabolic path exists that sidetracks the flavonoids before they can interact with the growth receptor of the larvae. It should be emphasized that Biophore II and Biophobe III are mutually excluding since the substitution pattern in Fragment II precludes the existence of a string of an unsubstituted aromatic carbon atom.

Finally, Fragment IV (see Fig. 1) suggests that the replacement of the hydroxyl groups at position 6 or 4' by methoxy groups markedly decreases the activity of the molecule. For instance, it can be seen that Luteolin (having OH groups in 5, 7, 3', and 4' positions) is very active, but becomes inactive when methylated at position 3' and 4' (compound 36 in Table 1). However, such substitution does not necessarily deactivate a molecule. Indeed, compound 11, Hypolaetin 3',4'-dimethyl ether (containing OH groups at 5, 7, 8 positions) remains active, since the powerfull Biophore I is still present. However, the overall activity value is lower than what would have been expected, were Fragment I alone be present.

Equation 3 was used to calculate the inhibitory potency for the compounds originally excluded from the analysis. The results are shown in Table 3. Overall, the predictions are quite good. Only compound 16, Quercitrin, which is found experimentally to be "extremely active" was predicted to be only "very active". These results show that the QSAR equation generated by the CASE program could be used to screen new compounds that can be used as potential larval growth inhibitors with a high confidence level.

TABLE 3. Predictions of inhibitory potency of the test compounds[a].

	Compound	Actual	Calcd.[b]	Prob.[c]
3	Luteolin	+++	+++	82.0%
16	Quercitrin	++++	+++	82.0%
21	Primuletin	-	-	17.0%
38	5,7,2',5'-Tetrahydroxyflavone	-	-	NB[d]

[a] See Table 2 for explanation of the symbols.
[b] Using Equation 3.
[c] Overall probability of larval growth inhibition activity, based on fragments with probability p > 85%.
[d] No basis to support activity; the compound is assumed to be inactive.

The main objective in QSAR studies is to develop potential new drugs that are more effective than those already known. The fragments (descriptors) generated by the CASE program are sometimes difficult to visualize in new molecules and therefore the design of new drugs remain a rather complicated problem. We applied the newly developed "autodesign" feature of the CASE program to find out if it could generate better inhibitors of H. Zea larval growth. The starting molecule used to **build** the best compound was the most active compound of the database, Maysin (compound 7 in Table 1), with an activity value of Log ED_{50} = 0.41. A compound was indeed found by CASE to be extremely active, i.e. projected Log ED_{50} = 0.13, which would be 0.29 log units more effective as inhibitor of larval growth than the most active compound in the database. To the best of our knowledge, this compound has not yet been synthesized nor tested as a larval growth inhibitor for H. Zea system. We intend to synthesize and report on its inhibitory properties in the future.

CONCLUSION

The Computer Automated Structure Evaluation (CASE) method has been applied successfully to generate the molecular fragments believed to be relevant to the H. Zea inhibitory potency of flavonoids. A regression equation was generated based on four descriptors. All the tests seem to indicate that the equation is significant and that its predictive power is accurate.

A concerted mechanism of action is proposed on the basis of the structures of the fragments found by the CASE program to be relevant to activity. In this mechanism, the relative geometry of position 5 and 6 of ring A and position 3' of ring B appear to be

crucial in the activity/inactivity of the flavonoids inhibiting larval growth. Although the existence of adjacent phenolic hydroxyl groups is important for inhibitory activity, some peripheral effects must be considered as well. It was also found that a compound will be inactive, if ring A and/or B have no substitution at all, independently of the presence of biophores in the molecule.

Acknowledgement We thank the Environmental Protection Agency and the Office of Naval Research under the Selected Research Opportunities program for support of this work.

REFERENCES

1. M. Shils and R. Goodhart, The Flavonoids in Biology and Medicine. The National Vitamin Foundation, Incorporated. Nutrition Monograph Series, Number 2, 1956.
2. B.G. Chang, A.C. Waiss, R.G. Binder, C.A. Elliger, Ent. Exp. and Appl. 24, 294-300 (1978).
3. P.A. Hedin, D.H. Collum, W.H. White, W.L. Parrot, H.C. Lane and J.N. Jenkins. The chemical basis for resistances in cotton to Heliothis insects. "Proceedings International Conference on Regulation of Insect Development and Behavior. Wroclaw Technical Univ. Press, Wroclaw, Poland, pag. 1071 (1980).
4. P.O. Feeny, J. Insect Physiol. 14, 805-817 (1968).
5. B.G. Chan, A.C. Waiss and M. Lukefahr, J. Insect Physiol. 24, 113-118 (1978)
6. C.A. Elliger, B.G. Chan and A.C. Waiss, Naturwissenschaften, 67, 358-360 (1980).
7. G. Klopman and H. Rosenkranz, Mut. Res. 126, 227-238 (1984).
8. G. Klopman and R. Contreras, Mol. Pharmacol. 27, 86-93 (1985).
9. G. Klopman, A. Kalos and H. Rosenkranz, Mol. Toxic. 1, 61-81 (1987).
10. G. Klopman and O. Macina, J. theor. Biol. 113, 637-648 (1985).
11. G. Klopman, J. Amer. Chem. Soc. 106, 7315-7321 (1984).
12. G. Klopman, M. Frierson, and H. Rosenkranz, Environ. Mutagen. 7, 625-644 (1985).
13. G. Klopman and R.E. Venegas, Acta Pharm. Jugosl. 36, 189-209 (1986).
14. G. Klopman, R. Contreras, H. Rosenkranz and M. Waters, Mut. Res. 147, 343-356 (1985).
15. G. Klopman and M. McGonigal, J. Chem. Inf. Comput. Sci. 21, 48-52 (1981).
16. G. Klopman, K. Namboodiri, and M. Schochet, J. Comput. Chem. 6, 28-38 (1985).
17. C. Hansch, Accounts Chem. Res. 2, 232-239 (1969).
18. T. Fujita, J. Med. Chem. 9, 797-803 (1966).
19. E. Haslem and T. Lilley, Plant Flavonoids in Biology and Medicine: Biochemical, Pharmacological, and Structure-Activity Relationship, pag. 53-65, 1986, Alan R. Liss, Inc.
20. G. Klopman and A. Kalos, J. Comp. Chem. 6, 492-506 (1985).