CROATICA CHEMICA ACTA CCACAA 56 (2) 175—183 (1983)

YU ISSN 0011-1643
CCA-1366 UDC 547:541
Original Scientific Paper

Structure — Activity Relationships in the Antiinflammatory Steroids: A Pattern Recognition Approach*

*Nicholas Bodor,** Alan J. Harget[1] and Edward Phillips*

College of Pharmacy Department of Medicinal Chemistry Box J-4, J. H. Miller Health Center University of Florida Gainesville, Florida 32610

Received August 2, 1982

A pattern recognition technique has been used to determine structure-activity relationships for antiinflammatory steroids. Experimental results using the human vasoconstrictor test of McKenzie and Stoughton and the rat granuloma cotton pellet method of Meier were correlated with the various substructural descriptors. Steroids were classified into two categories according to potency and a pattern recognition method was applied to determine their relative ranking. The resulting structure-activity relationships obtained and the relative contributions of the various structural variables for both bioassays are discussed. A synergistic effect was predicted to be in operation between certain pairs of substituents.

INTRODUCTION

The introduction and development of glucocorticosteroids has been the major therapeutic advance in dermatology in the past fifty years. Early work in corticosteroid research was directed to the synthesis of compounds with high antiinflammatory potency and to the reduction of side effects such as sodium retention. These attempts have been partially successful, since some corticosteroids have been synthesized that are locally active, but that have reduced systemic activity.

Topical antiinflammatory activity has been enhanced by various modifications of the steroid nucleus, the most important being the removal or masking of the hydroxyl groups and fluorination at the 6 and 9 positions. However, it is now believed[2] that the initial importance attached to fluorination has been overestimated since a number of non-fluorinated steroids exhibit high potency, for example, hydrocortisone 17-butyrate and budesonide and some fluorinated steroids display relatively low activity, such as betamethasone and dexamethasone.

In an attempt to rationalize the large volume of data which exists on the antiinflammatory steroids, we have applied a pattern recognition technique in order to determine structure-activity relationships. Such relationships attempt to rationalize the connection between the molecular structure of a

* Presented at *The IUPAC International Symposium on Theoretical Organic Chemistry,* held in Dubrovnik, Croatia, August 30 — September 3, 1982.

** To whom correspondence should be addressed

chemical compound and its measured biological activity. If such relationships could be determined for the steroids, then they would be of considerable practical and theoretical importance because of the significant role that steroids play in medicinal chemistry. For example, the relationships would allow the chemist to predict the biological activity of untested, or unsynthesized, steroids and hence adopt a more rational approach to drug design.

Structure activity studies employed in medicinal chemistry in the past have used the empirical method of Hansch,[3] the mathematical model of the Free-Wilson approach[4] and pattern recognition studies.[5] Despite the criticism leveled[6] at certain pattern recognition studies for their choice of data and its representation, the pattern recognition method, when properly applied, does offer the chemist useful means of determining the relationships existing in a large amount of data.

Steroids, because of their biological importance and relative molecular rigidity, represent a fruitful area for structure-activity studies. Minor molecular modifications to the steroid nucleus cause important changes in biological activity which should result in well-defined structure-activity relationships. Previous studies involving the antiinflammatory steroids have applied Hansch-type studies[7] and the use of de novo constants.[8] In the attempt to obtain more representative structure-activity relationships than those derived by previous investigators, we have considered a much greater number of steroids.

It is well known that certain corticosteroids which are predicted to be potent by animal studies are frequently found to be much less potent by human studies. The reasons for this are not known at the present time, although it might have something to do with the different metabolic transformations of the steroids in different species. Because of this discrepancy, we thought it would be instructive to compare the structure-activity relationships derived both from human and animal studies. Of the assays used, we have selected the McKencie-Stoughton vasoconstrictor test,[9] which is a human study, and the granuloma cotton pellet method of Meier,[10] which is an animal study.

In 1962, McKencie and Stoughton proposed a method of evaluating the antiinflammatory potency of a steroid which in its original and modified forms has come to be widely accepted. In its original form, an alcoholic solution of the steroid is applied to different sites on the human forearm. After evaporation of the alcohol, the area is occluded for 16 hours and then the skin evaluated for relative vasoconstriction (blanching). Because of the variable vasoconstriction response, the method requires a large number of human subjects. Despite the seemingly imprecise nature of the assay, the method gives a good indication of the clinical activity of a drug.[11] The assay measures the combined effect of a number of features, for example, the ability of the steroid to penetrate the skin barrier, its intrinsic activity at and subsequent clearance from the reaction site, local metabolism, binding, etc.

An extensive literature survey was undertaken to determine the vasoconstrictor activity of corticosteroids as measured by the McKenzie-Stoughton method or one of its derived methods. After the elimination of those structures, which contained unique substituents, 122 steroids remained. The elimination process was necessary if meaningful structure-activity relationships were to

be obtained. The remaining steroids were classified into two categories, potent and non-potent, the classification being determined relative to the potency of hydrocortisone 17-butyrate which was selected arbitrarily as the standard corticosteroid. Of the 122 steroids considered, 74 were classified as potent and 48 as non-potent.

In the cotton pellet granuloma test of Meier, the granuloma tissue is produced by implanting a cotton pellet of known weight into a subcutaneous pocket in rats. The degree of inflammation can be measured by the accumulation of granuloma tissue around the pellet and the antiinflammatory activity of the drug determined by the inhibition of granuloma tissue. Drugs can be administered systemically by oral or subcutaneous administration, or locally by direct absorption into the pellet. Several modifications have been proposed[12] to the cotton pellet method in order to improve the precision of the assay. The granuloma pouch technique,[13] which is similar to the cotton pellet method, was initially considered as a likely candidate for the animal study, but it was rejected because it appears that the accuracy of this assay is questionable and its effectiveness for granulomatous reaction has been criticized.[14]

After elimination of structures which contain unique substituents, the remaining 78 steroids were classified into two categories of approximately equal size by assigning steroids containing a potency greater than 15 to the potent category (31 steroids) and the remainder to the non-potent category (47 steroids).

THE PATTERN RECOGNITION ANALYSIS USED

During the past ten years, a number of papers have appeared describing chemical applications of pattern recognition. The usefulness of the technique has been tested in such diverse areas as the analysis of mass spectra,[15] assistance in material production problems,[16] and the determination of pharmacological activity.[17] Pattern recognition is useful for dealing with data of high dimensionality where the deduction of relationships is difficult. Although the technique is empirical, it is capable, when properly used, of providing the experimentalist with some insight into the relationships contained within the experimental data. The sole assumption made by the technique is that a relationship exists between the observed experimental data and the defined categories, although even this assumption will be investigated by the technique.

In the present study, we have applied a pattern recognition technique, the linear learning machine method, in an attempt to develop classification rules capable of distinguishing between potent and non-potent steroids. Essentially, a training set was composed of a number of steroids of known activity and classified according to potency into one of two categories. The linear learning machine method was then applied in an attempt to create a linear decision surface that would be capable of separating the potent steroid of the training set from the non-potent ones.

The linear learning machine method has been widely discussed in the literature,[18,19] so we shall present only the essential elements of the method here.

The experimental data to be classified is represented by a pattern vector of the form:

$$X = X_1, X_2, X_3 \ldots X_n$$

where each element, X_i, of the pattern vector represents a physically measurable quantity. The dimensionality, n, of the vector indicates the number of features, or observations, necessary to describe the pattern. If each compound in the training set is represented by a point in n-dimensional space, it may be expected that compounds of similar biological activity would lie in one limited region of the space and be separated from compounds of different biological potency from the remainder.

Linear separable data can be separated by a linear discriminant function of the form:

$$S = \sum_{i=1}^{n+1} w_i X_i$$

where X_i is an element of the pattern vector and w_i is the element of the weight vector associated with X_i. A $(n + 1)$ component is added where $(X_n + 1 \equiv 1)$ so that the category is determined by the sign of the dot product, S, that is,

$S > 0$ implies category 1

$S < 0$ implies category 2.

in the area of the pattern classifier, the effect of which is to maximize the separation of the two categories. In this case classification using the dot product proceeds according to the relationships

$S > 0$ implies category 1

$S < 0$ implies category 2.

The main problem is to determine the weight vector, that is the set of weights $(w_1, w_2 \ldots w_{n+1})$, such that each member of the training set is assigned to the correct category according to the relationships given above. The weight vector is determined by an error-correction feedback algorithm. One popular algorithm and the one used in this study, modifies the weight vector such that the linear decision surface is reflected about the misclassified point, that is, the dot product S has the same magnitude, but the opposite, and therefore correct, sign. This algorithm guarantees convergence if the data is linearly separable, although the rate of convergence cannot be predicted. The sole criterion for convergence is the correct classification of all members of the training set. If convergence is not obtained, the training process terminates after a predetermined number of iterations.

Since the approach is an empirical one and not constrained by theory, relationships may be determined from the weight vectors which may not have otherwise been considered. Furthermore, convergence in the training phase may allow the potency of unsynthesized, or untested, steroids to be predicted with some confidence. The predictive power of the method is enhanced if a large and representative training set is taken.

The application of the linear learning machine method to the antiinflamatory steroids and the results obtained is discussed in detail below.

RESULTS AND DISCUSSION

The computational description of molecular structure is a major problem in chemical pattern recognition studies. At the present time, we believe there is no satisfactory means of describing chemical compounds of diverse structure for pattern recognition requirements. In this study we have considered compounds of similar structure for which computational description is facilitated.

The relatively rigid structure of the steroid nucleus allows the structure of each steroid to be unambiguously described by employing substructural descriptors which are used to indicate the presence or absence of the associated substructure. In the case of the vasoconstrictor study, it was found that the 122 steroids could be described by a total of 33 descriptors. In order that meaningful structure-activity relationships could be determined, non-essential descriptors were eliminated by the weight-sign change feature selection technique.[20] Using this technique, a descriptor was retained if the sign of its weight vector component was found to be invariant to the initial weight value taken for the learning machine. This procedure reduced the number of descriptors

from 33 to 22. In the granuloma study, 17 descriptors were originally required to describe the 78 steroids for which granuloma information was available. Using the weight-sign change feature selection technique, the 17 descriptors were reduced to 13.

A requirement of pattern recognition studies is that the number of compounds in the training set should exceed, by at least a factor of 3, the number of descriptors if chance separation is to be avoided.[21] A further requirement states that the population of the least populated category should be greater than the number of descriptors.[22] Both of these conditions were fulfilled by the vasoconstrictor and granuloma data sets and hence the resulting weight vectors obtained for both of these studies must be considered meaningful.

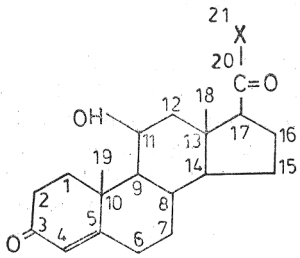
The linear learning machine method was applied to the vasoconstrictor and granuloma training sets to determine those weight vectors which would correctly classify each steroid according to the potencies experimentally determined. Complete convergence was obtained in the training procedure for each study. The resulting weight vectors obtained for both sets of data are shown in Table I. The magnitude of the weight vector indicates the contribution the feature makes to the classification, while the sign of the weight vector indicates whether the contribution enhances potency (given by positive values) or decreases potency (given by negative values). That is, a quantitative ranking of the substituents was obtained. If we first consider the weight vectors obtained for the vasoconstrictor study, Table I, it can be seen that there is general agreement between the results and experiment. For example, the 6-fluoro, 9-fluoro and 16, 17 acetonide grouping are predicted to enhance potency while saturation of the 1,2-double bond is predicted to decrease potency. It must be remembered that the weight vectors are *relative values,* that is to say, for example, that the replacement of an ester group in the 17-position by the hydroxyl group will lead to a decrease in potency, but not necessarily to a non-potent steroid, potency here being measured relative to that of of hydrocortisone 17-butyrate.

For a compound to be an effective topical antiinflammatory agent, it must remain in the epidermis and migrate only slowly into the dermis. This property is enhanced by the conversion of hydroxyl groups in the steroid to more lipophilic derivatives. The results shown in Table I confirm this, since most of the lipophilic groups like the esters are assigned positive weight vector values while the hydroxyl groups are assigned negative values. For the 17-ester groups the results show a similar parabolic dependence with activity as that found by Wieriks.[23] Although our results predict 17-propionate to be the most potent of the 17-ester substituents in contrast to that of 17-butyrate found by Wieriks our results were derived from a much larger set of steroids and are therefore considered to be more representative.

The method predicts 21-butyrate and 21-isobutyrate to be the most potent of the substituents studied. We believe their importance has been overemphasized, however, for the following reason. Betamethasone 21-butyrate and betamethasone 21-isobutyrate have potencies measured[24] by the blanching test of 85 and 90, respectively, which is only slightly greater than the potency of the reference compound, hydrocortisone 17-butyrate (potency = 50).[24] The inclusion of the 17-hydroxyl group into the betamethasone nucleus has the effect, as can be seen from Table I, of substantially reducing potency. Thus, for the

TABLE I

Sample Descriptor Values of the Various Substituents in the Steroids Studied



Human Vasoconstrictor Data		Rat Granuloma Data	
21-$CH_2OCOC_3H_7$ (i)	1.11	16,17-acetonide	0.88
21-$CH_2OCOC_3H_7$ (n)	1.10	6-F	0.77
21-$COOCH_2Cl$	0.71	9-F	0.56
6-F	0.68	17-OH	0.49
21-CH_2Cl	0.54	16-βCH_3	0.47
17-$OCOC_2H_5$	0.51	9-Cl	0.30
16,17-acetonide	0.50	6-αCl	0.20
21-CH_2OCOCH_3	0.47	16-αCH_3	0.16
16-αCH_3	0.38	11=O	0.03
9-F	0.37	saturated 1,2 bond	—0.33
17-$OCOC_3H_7$	0.22	16-OH	—0.40
21-CH_2OH	0.17	21-CH_2OH	—0.81
21-$CH_2OCOC_2H_5$	0.03	21-CH_2OCOCH_3	—1.17
17-$OCOCH_3$	—0.06		
16-βCH_3	—0.07		
saturated 1,2 bond	—0.25		
17-$OCOC_4H_9$	—0.38		
21-OCH_2SCH_3	—0.53		
21-OC_2H_5	—0.74		
21-OH	—1.12		
21-OC_3H_7	—1.26		
17-OH	—1.32		

potencies of betamethasone 21-butyrate and betamethasone 21-isobutyrate to be correctly determined, excessively large positive values must be assigned to the 21-butyrate and 21-isobutyrate weight vectors.

The results obtained for the granuloma study show some agreement with the vasoconstrictor study since both predict that the 16, 17-acetonide group and the 6-fluoro and 9-fluoro substituents will enhance potency, while saturation at the 1,2-position will reduce potency. In contrast to the vasoconstrictor study, however, the 17-hydroxyl group is predicted to increase potency while the 21-acetate group is predicted to decrease potency. The different contribution made by the 17-hydroxyl group is exemplified by betamethasone and dexamethasone which have different potencies measured by the two assays.

The linear pattern classifier obtained after training can be tested according to its ability to correctly classify the potency of unknown steroids not contained in the training set. This has been achieved for both studies by the so-called "leave-one-out" procedure.[25] In this procedure, a steroid was removed from the training set and the remaining steroids subject to training in the usual manner. The steroid was then classified and returned to the training set and

a second steroid removed and the training and classification procedure repeated. This process was repeated for all members of the training set, whereupon the predictive ability of the linear pattern classifier could be determined. Although this procedure is computationally expensive, it does give a good indication of the performance of the linear pattern classifier. The predictive ability was found to be 87.7% for the vasoconstrictor study and 88.5% for the granuloma study. Prediction is therefore good since the probability of guessing the correct potency is, of course, 50% for a binary decision maker. This suggests that the derived pattern classifiers could be used to predict the antiinflammatory potency of untested, or unsynthesized, steroids. It is not suggested that the pattern recognition technique should eliminate biological testing, but rather that its role be complementary to that of testing. For example, in overtaxed testing programs it could establish a ranking order in which the steroids were to be tested. A more important feature of the technique, however, is that by its analysis of previously tested steroids it allows the medicinal chemist to adopt a more rational approach to drug design.

Since all the structures studied contain more than one substituent, the linear learning machine program was extended to study the effect of combining certain pairs of substituents for both the vasoconstrictor and granuloma studies. The choice of which pairs of substituents to study was dictated by the chemical interest of the pairing and by the frequency with which that pairing occurred in the data.

The results obtained for both the vasoconstrictor and granuloma studies are shown in Table II. The results predict that a *synergistic effect* is in operat-

TABLE II

Sample Descriptor Values Obtained for Pairing Certain Substituents

Vasoconstrictor Data		Granuloma Data	
21-$CH_2OCOC_3H_7$ (i)	1.00	16,17-acetonide	0.80
1-$CH_2OCOC_3H_7$ (n)	0.82	6-F	0.67
21-$COOCH_2Cl$	0.70	9-F	0.50
17-$OCOC_2H_5$	0.48	17-OH	0.47
21-CH_2Cl	0.44	16-βCH_3	0.40
21-CH_2OCOCH_3	0.43	9-Cl	0.39
16-αCH_3	0.41	6F + 9F	0.25[a]
6-F + 9-F	0.37[a]	6-αCl	0.17
16,17-acetonide	0.35	16-αCH_3	0.16
6-F	0.29	11=O	0.01
17-$OCOC_3H_7$	0.23	saturated 1,2 bond + 17-OH	—0.02
21-CH_2OH	0.19	saturated 1,2 bond	—0.29
16-βCH_3	0.15	16-OH	—0.34
6F + 21-CH_2OCOCH_3	0.11[a]	21-CH_2OH	—0.73
9F + 16βCH_3	0.10[a]	21-CH_2OCOCH_3	—1.06
9F	0.02		
21-$CH_2OCOC_2H_5$	—0.04		
17-$OCOCH_3$	—0.11		
saturated 1,2 bond	—0.27		
17-$OCOC_4H_9$	—0.35		
21-OCH_2SCH_3	—0.40		
21-OC_2H_5	—0.66		
17-OH	—0.98		
21-OH	—0.99		
21-OC_3H_7	—1.15		

[a] Synergistic increments

ion for certain pairings of substituents. For example, synergism is predicted between the 6- and 9-fluoro substituents with the synergistic increment predicted to be 0.37, that is to say, the combined effect of a 6-fluoro and 9-fluoro pairing is 0.68, which is 0.37 more than the sum of the individual descriptors. A smaller synergistic effect is predicted to be in operation between the 6-fluoro and 21-acetate substituents and between the 9-fluoro and 16β methyl substituents, the synergistic increment being 0.11 and 0.10 respectively.

In the granuloma study, synergism is also predicted between the 6-fluoro and 9-fluoro substituents, the synergistic increment being 0.25. The results predict an absence of synergism between the saturated 1,2 bond and the 17-hydroxyl group.

It is difficult to rationalize the mechanism for synergism, although it is well known that small changes in the molecular structure of a steroid can have profound effects on biological activity. It may be that slight changes in conformation occur with the synergistic pair, that give the steroid a more favorable conformation for binding with the receptor. Another possible explanation may be the operation of long range through space interaction between the substituents of the synergistic pair. Whatever the mechanism of synergism, both assays seem to be sensitive to the effect. Nevertheless, our results predict that the potency of a steroid can be enhanced by the inclusion of certain pairs of substituents in the steroid nucleus, the most effective of those studied being the 6- and 9-fluoro compounds. It must be remembered, however, that our results ought not to be used to predict the overall clinical effectiveness of a steroid since this study has not considered the occurrence of possible side effects. Synergism may play an important role, not only in enhancing potency, but also in decreasing unwanted side effect of a drug. If this proves to be the case, then the medicinal chemist will be in a position to synthesize a drug which most fully meets the clinical requirements or demands specified.

CONCLUSIONS

A pattern recognition technique has been successfully applied to the elucidation of structure-activity relationships in the antiinflammatory steroids. In addition to determining the contribution each individual substituent makes to potency, a synergistic effect was predicted to be in operation between certain pairs of substituents. These results allow the medicinal chemist to adopt a more rational approach to the design of potent antiinflammatory steroids.

Acknowledgement. — Financial support by Otsuka Pharmaceutical Company is gratefully acknowledged. A. J. Harget is grateful for the financial support provided by the University of Florida.

REFERENCES

1. Visiting Professor from the Computer Center, University of Aston, Birmingham, B4 7PA, England.
2. M. D. Stewart, J. O. Runikis, S. C. Verma, and S. Wallace, *Can. Med. Assoc. Journ.* **108** (1973) 33; G. Bruni and G. Peverelli, *J. Int. Med. Res.* **2** (1974) 240.
3. C. Hansch, *J. Med. Chem.* **19** (1976) 1; C. Hansch and T. Fujita, *J. Amer. Chem. Soc.* **86** (1964) 1616.
4. S. M. Free and J. W. Wilson, *J. Med. Chem.* **7** (1964) 395.
5. P. C. Jurs and T. L. Isenhour, *Chemical Applications of Pattern Recognition,* John Wiley, New York, 1975.

6. C. L. Perrin, *Science* **183** (1974) 551; J. T. Clerc, P. Naegeli, and J. Seibl, *Chimia* **27** (1973) 639.
7. M. E. Wolff and C. Hansch, *Experientia* **29** (1973) 1111; R. A. Coburn and A. J. Solo, *J. Med. Chem.* **19** (1976) 748; P. Ahmad and A. Mollers, *J. Steroid Biochem.* **7** (1976) 19.
8. J. Fried and A. Borman, *Vitamins and Hormones* **16** (1958) 303.
9. A. W. McKenzie and R. B. Stoughton, *Arch. Derm.* **86** (1962) 608.
10. R. Meier, W. Schuler, and P. Desarilles, *Experientia* **6** (1950) 469.
11. J. Reid and D. B. Brookes, *Brit. J. Derm.* **80** (1968) 328; K. H. Burdick, *Acta. Derm. Venerol. Suppl.* **52** (67), (1972) 19; B. W. Barry and R. Woodford, *Brit. J. Derm.* **97** (1977) 555.
12. V. Robinson and J. M. Robson, *Brit. J. Pharmacol.* **23** (1964) 420; G. Benzi and G. M. Frigo, *Farmaco, Ed. prat.* **19** (1964) 327; K. F. Swingle and F. E. Shideman, *J. Pharmacol.* **183** (1972) 226.
13. H. Selye, *J. Amer. Med. Assoc.* **152** (1953) 1207; A. Robert and J. E. Nezamis, *Acta. Endocr. (Kbh)* **25** (1957) 105.
14. M. W. Whitehouse, *Progress in Drug Research,* Vol. VIII, Basel-Stuttgart, Birkhäuser, 1965.
15. P. C. Jurs, B. R. Kowalski, T. L. Isenhour, and C. N. Reilley, *Anal. Chem.* **42** (1970) 1387.
16. B. R. Kowalski, *Chem. Technol.* (1974) 300 (May).
17. K. Yuta and P. C. Jurs, *J. Med. Chem.* **24** (1981) 241.
18. N. J. Nilsson, *Learning Machines,* McGraw-Hill, New York, N. Y., 1965.
19. T. L. Isenhour and P. S. Jurs, *Anal. Chem.* **43** (10), (1971) 20A.
20. T. L. Isenhour, B. R. Kowalski, and P. C. Jurs, *CRC Crit. Rev. Anal. Chem.* **4**, (1974) 1.
21. C. F. Bender, H. D. Shepherd, and B. R. Kowalski, *Anal. Chem.* **45** (1973) 617; D. H. Foley, *IEEE Trans. Inf. Theory, it-18,* (1972) 618.
22. E. K. Whalen-Pedersen and P. C. Jurs, *J. Chem. Inf. Comput. Sci.* **19** (1979) 264..
23. J. Wieriks, W. Hespe, K. D. Jaitly, and B. L. van Kan, *Derm.* **152** (1976) 181, (suppl. 1).
24. G. H. Phillips, *Mechanisms of Topical Corticosteroid Activity,* Ed. L. Wilson and R. Marks, Churchill Livingstone, 1976.
25. B. R. Kowalski and C. F. Bender, *J. Amer. Chem. Soc.* **94** (1972) 5632.

SAŽETAK

Odnos strukture i aktivnosti u antiupalnim steroidima: Pristup prepoznavanjem obrasca

N. Bodor, A. J. Harget i E. W. Phillips

Tehnika prepoznavanja obrasca upotrijebljena je za određivanje odnosa strukture i aktivnosti za antiupalne steroide. Eksperimentalni rezultati ljudskog vazokonstriktorskog testa po McKenzieu i Stoughtonu, kao i Meierove metode induciranja granuloma u štakora pamučnom kuglicom, bili su korelirani s raznim substrukturnim deskriptorima. Steroidi su bili klasificirani u dvije kategorije prema intenzitetu djelovanja, a metoda prepoznavanja obrasca bila je upotrebljena za određivanje njihova relativnog poretka. Razmatrani su dobiveni odnosi strukture i aktivnosti i relativni doprinosi raznih strukturnih varijabli u oba biološka testa. Predviđen je sinergistički efekt u djelovanju određenih parova supstituenata.