A Method for Describing the Molecular Electrostatic Potential in Determining Structure–Activity Relationship

Igor V. Svitan'ko,*[a] Mikhail I. Kumskov,[b] Ivan L. Zyryanov[a] and Ivan A. Suslov[a]

[a] *Higher Chemical College of the Russian Academy of Sciences, 125820 Moscow, Russian Federation. Fax: +7 095 135 8860; e-mail: svitanko@ncc.free.net*

[b] *N. D. Zelinsky Institute of Organic Chemistry, Russian Academy of Sciences, 117913 Moscow, Russian Federation.*

Electrostatic complementarity is more important than geometrical complementarity in quantitative structure–activity relationship (QSAR) calculations. The similarity of molecular electrostatic potentials (MEP) (the occurrence of combinations of points possessing a definite electrostatic potential) describes the structure–odour relationship as a particular case of such calculations at a sufficient probability level.

The synthesis of compounds possessing predetermined properties is of great importance. To perform a directed synthesis, it is desirable to establish the most reliable relationships between the structures of chemical compounds and their properties. Two general approaches to the QSAR problem can be distinguished.

1. To find geometric relations between the preliminarily selected elements of a molecular structure ensuring the "key–lock" molecule–receptor complementarity[1,2] on the basis of molecular mechanics and quantum-chemical calculations of the molecular geometry.

2. To use topological parameters[3] that usually relate well the structure and activity of a compound but afford a limited view of the spatial picture of intermolecular interactions.

Evidently, the "key–lock" complementarity of the molecule and receptor, which is, in the opinion of many researchers, responsible for various types of biological activity of compounds,[4–6] cannot be unambiguously determined by the geometric structure of a molecule. Moreover, it is impossible to represent the correspondence of a molecule and a receptor solely on the basis of the molecular topology. First, it is necessary to take into account the spatial electrostatic complementarity of two or several molecules, or a molecule and a receptor (long-range interaction). For this purpose, it is necessary to calculate the electrostatic field created by the molecule and then supplement the model thus obtained by the geometric complementarity (short-range interaction).

We have for the first time proposed a method for determining the structure–activity relationship. In this method, calculations of the electrostatic field created by a molecule are combined with geometric restrictions imposed on the molecule–receptor (or molecule–molecule) system. In our opinion, the recurrence of these parameters primarily governs the mathematical model of a given property.

It is important in this respect that similar electric fields can be created by molecules differing in structure and in the type of atoms they contain.

The approach proposed in the present work is based on the following principal assumptions:

1. No chemical reaction occurs during the coordination of a molecule to a receptor, *i.e.*, old bonds are not broken and new bonds are not formed either in the molecule itself or between the molecule and the receptor. This implies that the minimum distance separating the molecule and receptor atoms equals the sum of the van der Waals radii of the respective atoms.

2. At these distances, the energy of interaction between the molecule and the receptor is appreciably governed by the electrostatic component. Thus, it is assumed that the overlap integrals of the molecule and receptor orbitals are equal to 0; dispersion interactions are not taken into account, either.

Let us construct a surface with an optimum size where the active centres of the receptor can be located. For this purpose, we shall make the following definitions: (*a*) a probe is a sphere with a radius of 1.4–1.5 A, which corresponds to the average van der Waals radii of atoms most frequently met in organic compounds; (*b*) a van der Waals surface is the surface obtained by covering the molecule by rigid interpenetrating spheres built around the atomic centres and having the radii approximately equal to the van der Waals radii for the respective atoms;[7] (*c*) a covering surface is a locus of points created by the centre of a probe sphere when it is encircled along the van der Waals surface.[7]

The covering surface so defined represents the geometric restrictions imposed on the molecule–receptor coordination, since the centres of the receptor atoms cannot be located inside the surface. After that, we set the molecular electrostatic potential (MEP) on the surface for the electrostatic field created by the molecule. Let us call the surface thus obtained the potential energy surface. Points corresponding to local extremums of the MEP are situated at this surface. We shall call these points "representation points" of a molecule. A molecule is described as a combination of all its representation points and the MEP values in these points.

Let us assume that the occurrence of combinations of points with definite coordinates and potentials on the potential energy surface defines the desired electrostatic complementarity, *i.e.* represents an essential property; in other words, a factor governing a given property of a molecule.

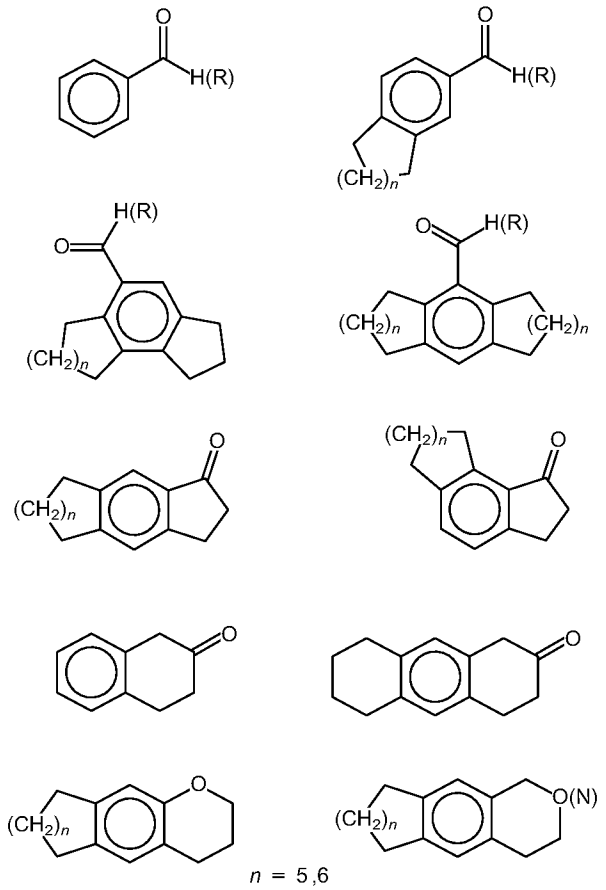

Fig. 1 Compounds used in the training and control sets.

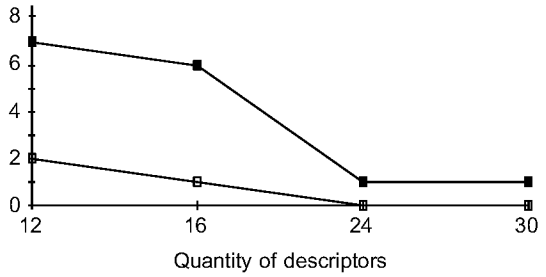

Fig. 2 Errors distribution diagram. ■, number of type-1 errors; □, number of type-2 errors.

This assumption was verified and the procedure tested for musk odour as an example of biological activity.

A set of 110 compounds was chosen, 35 of which possess pronounced musk odour, while the others do not. The compounds included in the set are structures with different substituents presented in Fig. 1. The following groups were used as substituents: H, Alk, NO_2, OAlk, Hal, CHO, C(O)Alk, CN, C(O)Hal.

Geometry optimisation was performed for each molecule included in the set. The optimisation included two stages: initial optimization by the molecular mechanics methods using the MMX force field; final optimisation by quantum-chemical AM1 methods.

Each molecule was surrounded by a covering surface using a probe sphere of 1.5 A radius. The surface was represented by points spaced so that the distance between the closest neighbours was 0.2 A. The MEP value was calculated for each point of the surface. Points were then found where the MEP has minimum and maximum values, both in the point itself and in its closest neighbours. The set of such points for each molecule was used to represent the shape of the MEP.

The range of the MEP values in the representation points was divided into intervals depending on the frequency of occurrence of different MEP values, and the optimum discretisation step of the range (0.2 A) was chosen.

A set of descriptors of the type $(E_1, E_2, R), N$ was constructed for each molecule, where E_1 and E_2 are the numbers of intervals of the MEP range for the first and the second point in a pair; R is the number of interval in the range between the representation points; and N is the occurrence of a given descriptor for the molecule. The descriptors were constructed for each pair of representation points for each molecule.

The combination of descriptors was determined for all of the molecules belonging to the training set. A unique name was assigned to each descriptor. Thus, a $(X_1, \ldots X_N)$ vector, where X_i is the occurrence of the ith descriptor for the molecule, corresponded to each molecule.

A stepwise regression procedure was performed using the Bibigon system[8] in order to select significant descriptors and construct linear models describing a property of the molecule:

$$P_i = a_0 + \sum(a_j X_{ij}), \quad i = 1 \ldots n,$$

where j is the number of the criterion used for constructing the model.

The classification assumed that the jth compound possesses odour if $P_i > 0.5$, otherwise the compound was regarded as odourless. A type-1 error is the wrong attribution of an odorous compound to the odourless class of compounds, while a type-2 error is the wrong attribution of an odourless compound to the odorous class.

As a result, the models given in Fig. 2 were obtained with different number of descriptors differently fitting the property–structure dependence.

Comparative calculations showed that computing the recurrence of solely geometric parameters (account of spatial atomic coordinates and charges) without an account of electrostatic parameters gives a much lower probability of predicting the desired property.

Thus, the results obtained imply that the occurrence of similar points (which have identical mutual arrangement and potentials) on the potential energy surface defines the given parameter of the set of compounds at a sufficient probability level, and the method proposed for forming descriptors adequately describes the type of biological activity studied.

References

1 M. Yu. Gorbachev, A. S. Dimoglo, I. B. Bersuker, P. F. Vlad and M. N. Koltsa, *Teor. Eksper. Khim.*, 1986, 355 (in Russian).
2 V. B. Luzhkov and G. I. Bogdanov, *Usp. Khim.*, 1986, **55**, 3 (*Russ. Chem. Bull.*, 1986, **55**, 1).
3 M. I. Stankevich, I. V. Stankevich and N. S. Zefirov, *Usp. Khim.*, 1988, **57**, 337 (*Russ. Chem. Bull.*, 1988, **57**, 191).
4 Y. C. Marttin, *J. Med. Chem.*, 1992, **35**, 2145.
5 G. M. Grippen, *J. Med. Chem.*, 1979, **22**, 988.
6 G. M. Grippen, *J. Med. Chem.*, 1980, **23**, 599.
7 F. M. Richards, *Ann. Rev. Biophys. Bioeng.*, 1977, **6**, 151.
8 L. A. Ponomareva, M. I. Kumskov, E. A. Smolenskii, L. K. Maslova and N. S. Zefirov, *Tezisy dokladov pervoi Vsesoyuznoi konferentsii po teoreticheskoi organicheskoi khimii (Abstracts of the First All-Union Conference on Theoretical Organic Chemistry)*, VATOKh, Volgograd, 1991, p. 537 (in Russian).

Received: Moscow, 11th April 1994
Cambridge, 10th May 1994; Com. 4/02265C