match no. 15 pp. 227-249 1984

CHEMICAL COMPLEMENT SIMULATION AND QUANTITATIVE STRUCTURE ACTIVITY RELATIONSHIP (I)

H.R. Karfunkel

Institut für Physikalische und Theoretische Chemie,
Universität Tübingen, Auf der Morgenstelle 8,
7400 Tübingen, West Germany

Summary
Subjects of this contribution are interacting chemical systems of Host-Guest type, in which the Host system is structurally unknown. The method of Chemical-Complement-Simulation (CCS) is represented. The purpose of CCS is the substitution of the unknown Host by a mathematical analogon, whereby sterical complementarity as well as complementarity of intermolecular forces are taken into account. Starting from a series of Guest molecules and corresponding known interaction energies, the Host system is simulated by a two-dimensional manifold M on which a vector function $\underline{\rho}$ is defined, which chemically characterizes the Host. The local contributions to the interaction energy are projected on this manifold and are given by the scalar product of $\underline{\rho}$ with an appropriate vector potential function depending only on the nature and spatial arrangement of the Guest molecule.

1. Introduction

The starting point of any drug design is a set of structurally well-known substances called effectors. They have known activities with respect to a given receptor. The structure of the receptor, which generally is a macromolecular formation (enzymes, membrane proteins, nucleic acids etc.), is unknown.

Only one out of several thousands of substances, which have been tested for a desired biological effect, achieves any practical value today[1]. The rationalisation of such a selec-

tion process requires reliable quantitative structure-activity relationships (QSAR).

The methods for the determination of QSAR might be categorized as non-topographical and topographical methods. In the first case only non-topographical data of the effectors, (i.e. such data which ignore the explicit three-dimensional geometrical circumstances) are correlated to the activities of the effectors using univariate or multivariate[2] regression. The most wide-spread procedures belonging to this category of non-topographical methods are: Hansch-analysis[3,4], Free-Wilson-de-novo-analysis[5] and Pattern-recognition methods[6,7,8]. Simplicity and the fast access to data are the main advantages of these methods, but the ignorance of the three-dimensional nature of the interaction between an effector and the receptor severly limits their applicability. Even the topologically equivalent optical antipode of a chiral effector mostly has a different activity in comparison to the original effector[9].

The more sophisticated topographical methods were known before the computer era began (Croxatto & Huidobro 1956[10]). In the late sixties Amoore et al.[11,12] constructed the PAPA-Machine with a considerable amount of hardware in order to correlate molecular shape with odour. Due to the fast progress in computer performances an increasing importance is being attached to the topographical methods. The trend is the usage of interactive graphical systems[13].

An abstraction of the chemical complementarity, i.e. a complementarity in the sterical sense as well as in the sense of intermolecular forces, is given by the concept of "pharmacophoric-pattern"[14]. This "pharmacophoric-pattern" can be considered as the essential topography that must be part of the topography of an effector in order to be recognized by the receptor. This concept leads to useful and promising algorithms for the derivation of QSAR. However, most of the published topographical methods are only of qualitative nature. Particular attention should be payed in this context to the "MTD-receptor-site-mapping" of Balaban et al.[15] in which an

appropriate superposition of a set of effector molecules delivers, with the aid of integer programming, an approximate picture of the sterically complement receptor groove. Consequently, relevant and irrelevant parts of the effector involved in the steric fit might be distinguished. The quality of the optimized groove is given by the regression coefficient. The MTD method as well as its Monte-Carlo version (MCD) are suffering from the drawback that it only considers the steric fit in a topographical manner. The complementarity of the remaining intermolecular forces, especially the electrostatic interactions, are not taken into account. The counterpart of the MTD-method is the representation of the "pharmacophoric-pattern" by an electrostatic field, as Maayani et al.[16] had done in the case of anticholinergic pharmaca.

2. Purpose and Method

The starting point of the method of chemical complement simulation are N chemical systems of Host-Guest-type[17,18] with the following characteristics:

1) The host subsystem R is the same in all N systems. Its structure is unknown.
2) The structures and properties of the guest molecules EF_1, ..., EF_N are known. The set of molecules EF_1,..., EF_N will be called "Training-Set."
3) The guest molecules bind non-covalently to the host subsystem, producing well-defined one to one complexes EF_i-R.
4) The spatial regions on the host subsystem which the guest molecules bind themselves to, have overlapping parts.
5) The non-covalent binding energy consists of additive local contributions.
6) The N binding energies or quantities which correlate with binding energies are known.

The purpose of the complement simulation is to construct a mathematical analogon to the unknown host subsystem by means of the training set data. As an anticipation this analogon

consists of a two-dimensional manifold on which an inhomogeneous charge distribution is located. In addition to that, this manifold attracts the guest molecules as well as it repells them (with different 1/r powers).

It is self-evident that the items mentioned above correspond to the conventional models of receptor-effector interactions[19] (at least in the early stages of these interactions). The only specific information which refers to the receptor in question is the measured (biological) activity. The activities, however, elicited by an effector, do not necessarily correlate with the binding energies. This might be the weakest point concerning the postulated analogy between host-guest systems with their underlying binding-energies and the receptor-effector systems with their underlying dosis-activity curves. The method of the chemical complement simulation is restricted to binding-energies in the sense of the intermolecular interactions.

The scarce information about the receptor does not allow its reconstruction as a chemical whole. The receptor analogon can only reflect some special aspects of the receptor-effector interactions, namely those which refer to the binding energies. In spite of this indispensable restriction, however, it must be stated that the construction of a mathematical receptor analogon not only offers a correct reproduction of interrelations. The purpose is not to imitate the behaviour of the receptor restricted to the training set, but to understand how this interrelation comes into being. The desired function of the simulated receptor is to offer predictions about the activity of new effector-candidates apart from the training set. The fundamentals of the interrelation should constitute the beginning of any analysis in order to avoid possible absurdities in QSAR[20-22]. They are the various intermolecular interactions which make up the binding energy.

A note concerning the methodological procedure of this paper: natural receptors as well as the evaluation of actual experimental data are not taken into account. The emphasis is put

on the establishment of an algorithmical frame for new concepts. Drug design is a practical discipline in which only those concepts are valuable that can be formulated as detailed algorithms. Only these algorithms make sense when they are capable of managing the numerical problems in using the computers of today and the near future.

3. Reformulation of the intermolecular interactions

The molecular-mechanic version of the intermolecular interactions

The simplest quantum-mechanical treatment of intermolecular interactions refers to a system consisting of two well-defined molecules. The subject of that investigation is the total energy of the system (which usually depends on six degrees of freedom) compared to the energy of the isolated molecules[23,24]. Such a treatment has deficiencies, for the influence of the medium remains neglected apart from its dielectric features. Furthermore, the statistical aspects must be included in order to get statements about thermodynamic quantities[25,26]. Nevertheless, for the practitioner the results of these analyses are of great importance, especially the simplified approximation formulas of the interaction between larger molecules (see the extensive bibliography in Claverie[27]). In the simplest and most useful formulas the intermolecular potential is composed of a sum of atom-atom interactions. In Molecular Mechanical treatment a given system will move towards a minimum of ΔE_{total} starting from given initial conditions. These minima are equilibrium constellations of the intermolecular complex. (In this place it should repeatedly be emphasized that the intermolecular interaction energies ΔE_i of a set of effectors in contact with the receptor are in reference to the gas phase. The interaction between the effectors and the receptor take place in solution and so a correlation between the above ΔE and the thermodynamic quantities ΔG is only approximate and valid only in certain conditions [28].)

For the interaction between two not too small molecules the

following statements are important for our purpose:

a) The intermolecular interaction energy is composed of local spatial contributions. What concerns Molecular Mechanics, this is a tautology, but it can be shown by rigorous quantum-mechanical treatment that this kind of additivity is valid even in the case of short-range forces[27]. It is the very accumulation of these weak local interactions which cause the specificity[18].

b) Only the atoms neighbouring the contact region, called front atoms, deliver essential contributions to the intermolecular interaction energy.

c) Penetration is prohibited, i.e. the front atoms cannot penetrate the contact region neither locally nor globally. This prohibition expresses only the absolute dominance of the repulsion at short distances and so the possibility of covalent binding is eliminated.

An attempt to substitute the unknown receptor by simulated front atoms bearing appropriate charges is one possibility toward a receptor analogon. Unfortunately, this version of the chemical complement simulation which is based on an atomistic representation of the receptor site is not realistic. The scarce information, namely the experimentally measured activities (in this case the binding energies), are insufficient to produce such a detailed picture of the receptor site.

The concept of the chemical complement simulation, however, is not disposed of: The way out of the many-particle dilemma leads to a continuous representation of the receptor site in place of the atomistic one.

4. Formulation of the continuum representation

In the desired continuum representation of the receptor's binding site this binding site is represented by a two-dimensional manifold M.

For the intermolecular interaction energy between an atomistic effector molecule and the "continuum-represented" receptor the

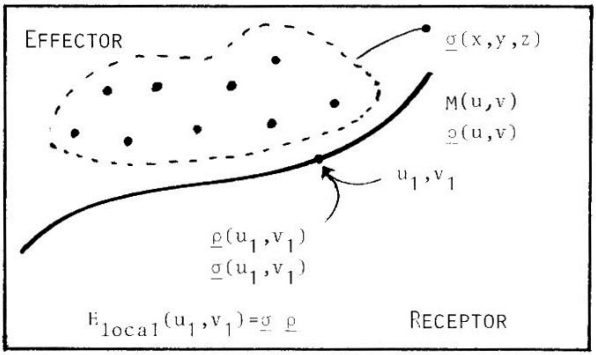

FIGURE 1: Scheme of the interaction between an atomistic effector and the continuum analogon of the receptor. M(u,v) is a two dimensional manifold represented parametrically by (u,v). $\underline{\rho}(M)$ is a vector function on M which characterizes the continuum receptor. $\underline{\sigma}(x,y,z)$ is a vector potential function of the effector and ΔE is defined as the integral of local contributions, E_{local}, over M. E_{local} is given by the scalar product $\underline{\sigma}\cdot\underline{\rho}$, e.g. E_{local} at the point u_1,v_1 on M is given by $\underline{\sigma}(u_1,v_1)\cdot\underline{\rho}(u_1,v_1)$.

following Ansatz will be stated (see the scheme in FIGURE 1):

$$\Delta E = \int_M \underline{\sigma}(x,y,z)\cdot\underline{\rho}(M)\,dM \qquad (1)$$

Comments to formula (1):

M - The two-dimensional manifold which represents geometrically the binding site of the receptor.

$\underline{\sigma}(x,y,z)$ A vector consisting of L components $\sigma_1,\ldots,\sigma_L$. These components are potential functions and they are only dependent on the effector molecule. E.g., the 1^{th} component of $\underline{\sigma}$ can represent the electrostatic potential of the effector molecule, i.e. $\sigma_1=\Sigma e_i/|\underline{r}_i-\underline{r}|$ where the summation is over the atoms of the effector; e_i, $\underline{r}_i$ are the charge and position of the i^{th} effector atom. $\underline{\sigma}$ is specified anywhere in the three-dimensional space, that is, also on the manifold M.

$\underline{\rho}(M)$ - A vector consisting of the components $\rho_1,\ldots,\rho_L$. These components are parameter functions defined on the

manifold M. They are only dependent on the receptor, and so they characterize it, e.g., the l^{th} component of $\underline{\rho}$ can represent a charge distribution e(M) on the manifold M.

$\int_M \ldots dM$ - The surface integral over the manifold.

Notes to the formulation given by Eq. (1):

a) The designation of the vectors by σ and ρ reminds us of the well-known Hammett equation[29]. The fundamental idea behind (1) is similar to that of Hammett, i.e., to decompose interaction terms or differences of such terms into independent parts by separation. Hammett dealt with a series of structural elements (substituents) which were involved in a given reaction proceeding according to the same mechanism. The quantity σ was related to the substituents, the quantity ρ characterized the common reaction mechanism and the considered effect was represented by $\sigma\rho$.

b) The formulation in (1) preserves the additive decomposition of the intermolecular interaction energy into different types (electrostatic interaction, dispersion etc.).

c) The local additivity is retained by the surface integral. The essential point is that the interaction which takes place in the three-dimensional space is projected onto a two-dimensional manifold. In the next section it will be checked by numerical experiments whether such a projection is admissible.

d) Naturally the dimension of the vectors $\underline{\sigma}$ and $\underline{\rho}$ depends on the number of the relevant forces involved in the interaction. In the following computations only three types of forces will be taken into account: Electrostatic interactions, short-range repulsion forces and medium-range attraction forces. Other forces, e.g., hydrophobic interactions, H-bonding etc., should be treated analogously to the above mentioned forces by using potential functions in some 1/r power.

e) The question, whether the different types of interaction energies can be separated so that they are locally representable as a product of effector potential and receptor

parameter, is of importance. This question not only concerns the continuum formulation but also the atomistic description of the system. Obviously this separation is valid for the electrostatic interaction. For the steric attraction and repulsion this question is not at all easy to answer. There are no concepts as to the steric attraction potential of a molecule in form of concrete mathematical expressions. In section 6 this problem will be dealt with numerically.

5. Numerical experiments referring to the projection problem

To simplify matters, in this section rigid two-dimensional fictive molecules are considered. The initial situation for the numerical experiments can be summerized as follows:

a) The receptor region is given by N_r atoms fixed in space and carrying charges e_i, $i=1,\ldots,N_r$ (the atoms on the right hand of the plot in FIGURE 2).

b) A rigid effector molecule, consisting of N_e atoms carrying various charges, is oriented in different ways to the receptor (the position of this effector at the energy minimum can be seen on the left hand of FIGURE 2). The total of K different orientations are taken in regard.

c) All attraction and repulsion parameters are set equal respectively (all $A_{ij}=A$ and all $B_{ij}=B$ in the Lennard-Jones-Potential).

d) Considering Moleclular Mechanics as the standard, the K interaction energies E_k of the "training set" are calculated as follows:

$$E_k = A\sum_{i}^{Nr}\sum_{j}^{Ne}\left(\frac{1}{r_{ij}}\right)^{12}_k - B\sum_{i}^{Nr}\sum_{j}^{Ne}\left(\frac{1}{r_{ij}}\right)^{6}_k + \sum_{i}^{Nr}\sum_{j}^{Ne}\frac{e_i e_j}{(r_{ij})_k} \qquad (2)$$

$k=1,\ldots,K$ is the index of the respective orientation. These E_k represent the "experimental" values.

e) In order to test the reliability of succeeding predictions, the interaction energies of K′ additional orientations of the

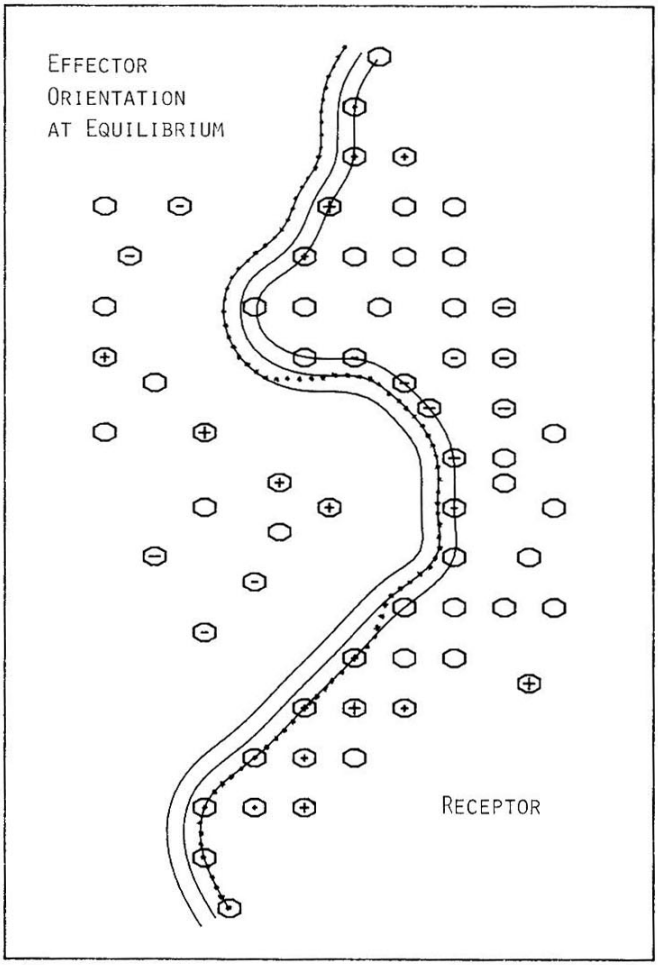

FIGURE 2: Plot to the numerical experiments. The size of the point charges is designated by the size of the symbols + and -. The four manifolds (used in the numerical experiments) can be seen as spline functions. Three of them run either through the front atoms of the receptor or equidistantly to them. The fourth manifold (dotted line) coincides piecewise with the first three manifolds. Fifty atoms were used for the representation of the atomistic receptor site.

effector are calculated according to (2). These K' values, however, won't be taken into account in the succeeding regression.

All data referring to the atomistic receptor (atom positions and charges) as well as the parameter values (A and B) now are "forgotten". The following problem is stated: We are given an effector molecule with N_e atoms carrying charges e_j, $j=1,\ldots,N_e$. The effector molecule is located in K specified orientations in space. For each of these orientations the "experimental" or "measured" value of the intermolecular interaction energy E_k, $k=1,\ldots,K$ is known. Furthermore, some manifold M (in this case a one-dimensional manifold) is given. Which values are to be assigned to the parameters a and b and which form is to be ascribed to the function e(M) so that the K "theoretical" interactions values:

$$E_{k,\text{theor.}} = \int_M \sum \left(\frac{1}{r_{jM}}\right)^{12}_k \cdot a \cdot dM + \int_M \sum \left(\frac{1}{r_{jM}}\right)^{6}_k \cdot b \cdot dM + \int_M \sum \left(\frac{e_j}{r_{jM}}\right)_k \cdot e(M)\, dM \qquad (3)$$

correlate as well as possible with the E_k values given "experimentally"?

It shall be stressed that the manifold M is given. It is a question of the projection problem i.e., can an interaction be projected onto a manifold of lower dimension? The vectors $\underline{\sigma}$ and $\underline{\rho}$ are in this simple case:

$$\underline{\sigma}^T = \left(\sum_j \left(\frac{1}{r_{jM}}\right)^{12}_k \;,\; \sum_j \left(\frac{1}{r_{jM}}\right)^{6}_k \;,\; \sum_j \left(\frac{e_j}{r_{jM}}\right)_k \right)$$

and $$\underline{\rho}^T = (a \;,\; b \;,\; e(M)) \qquad (4)$$

where $(r_{jM})_k$ is the distance between the j^{th} effector atom in the k^{th} orientation and a specified point on M.
Four different manifolds have been tested. M_1, M_2 and M_3 run equidistantly to the front atoms of the "forgotten" receptor (solid lines in the plot), FIGURE 2. The 4^{th} manifold does not

run equidistantly to the front atoms (FIGURE 2, dotted line). All manifolds are represented mathematically by parametric Spline-Functions[30].

Solution of (3): The size of the training set was K=20. The unknown function e(M) is approximated by a linear combination of given basis functions ϕ_1, ϕ_2,... (Chebyshev polynomials, trigonometric functions etc.):

$$e(M) = \sum c_1\phi_1 \qquad (5)$$

By inserting (5) into (3), the resulting "theoretical" interaction energy is a linear function of the receptor parameters and so this is a linear least square problem that can be solved by standard methods[31].

$$E_{k,theor.} = a\int\sigma_1 dt + b\int\sigma_2 dt + c_1\int\sigma_3\phi_1(t)dt + \ldots. \qquad (6)$$

Results: The results of the computations are represented in FIGURE 3 and 4. The number of functions used to approximate the charge distribution was 12. Precise numerical values of the results are not given, since the emphasis is only laid on the qualitative behaviour.
The result for M_1 can be seen in FIGURE 3. The results for M_2 and M_3, that is, for the manifolds running equidistantly to the front atoms, do not differ qualitatively from the result for M_1. In FIGURE 3, at the upper left, the correlation between the theoretical intermolecular energy (computed according to (6)) and the given "experimental" energies is shown. As one can see, the correlation is rather good. At the upper right, the result for the correlation of pure electrostatic interactions is shown. The correlation is excellent. In contradistinction to these manifolds, for the manifold M_4 which is not running equidistantly to the front atoms, no acceptable correlation between the total interaction energies is possible (see FIGURE 4 at upper left). The correlation of the pure electrostatic interactions which is not influenced by the non-equidistancy,

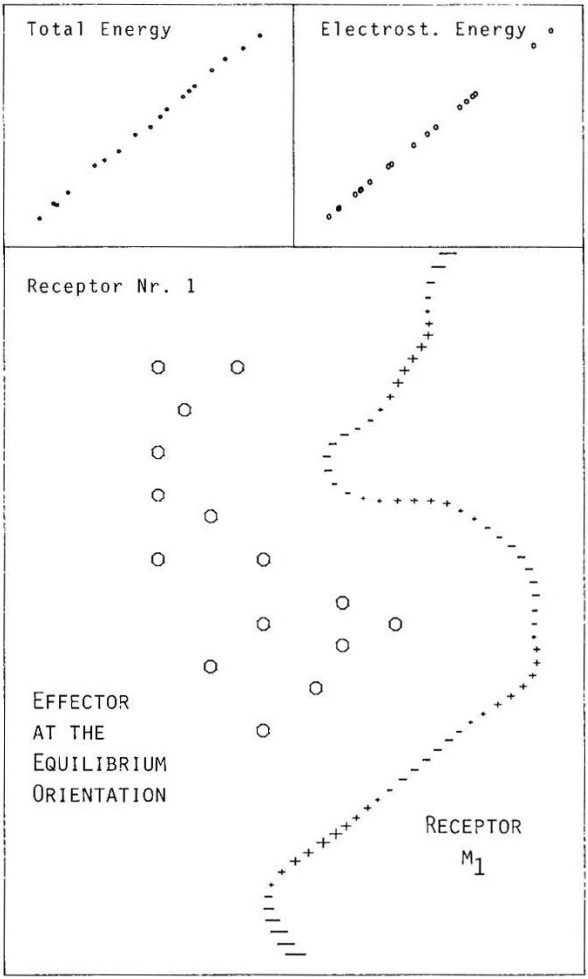

FIGURE 3: Result with the manifold M1 running through the front atoms. The computed charge distribution is shown symbolically. The correlation between the theoretical and experimental total energies (upper left) and electrostatic interaction energies (upper right, 3rd term of (3) and (2)) is represented.

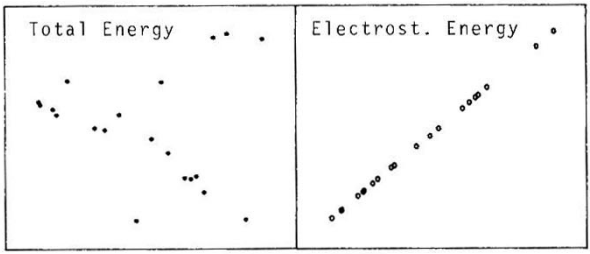

FIGURE 4: Result with the fourth manifold M4 (non-equidistant). For further explanations, see FIGURE 3.

is still excellent (FIGURE 4 at the upper right).
In order to test the reliability of the continuum receptor, the interaction energies of the additional K´ orientations, not having been taken into regard in the regression above, were computed according to (6) and compared with the "experimental" values. In all K´ orientations the predictions were inside the confidence interval of the regression.

In summarizing, the numerical experiments showed in the case of simple systems that the continuum representation of the receptor by a manifold of lower dimension is capable of reproducing the results of Molecular Mechanics. The condition however is that the considered manifold runs equidistantly to the front atoms of the receptor. If the front atoms are differing which is usually the case, the differing sizes of the front atoms must be taken into account. "Equidistantly to the front atoms" then means equidistance to a molecular envelope which takes into regard the atom sizes. As an example of a molecular envelope here the van der Waals envelope shall be mentioned.

6. Reformulation of the Lennard-Jones potential

In this section a reformulation of the Lennard-Jones potential (7) is undertaken:

$$E_{ij} = A_{ij}/r_{ij}^{12} - B_{ij}/r_{ij}^{6} \qquad (7)$$

where A_{ij} and B_{ij} are constants depending on the nature of the atoms i and j (non-bonded).
The aim of the transformation is to approximate the steric interaction energy E_{ij} in (7) by the following formula:

$$E_{ij} = \frac{A1_i \cdot \tilde{A1}_j + A2_i \cdot \tilde{A2}_j}{(r_{ij} - r_i - r_j)^m} - \frac{B1_i \cdot \tilde{B1}_j + B2_i \cdot \tilde{B2}_j}{(r_{ij} - r_i - r_j)^n} , \qquad (8)$$

where $A1_i$, $A2_i$, $B1_i$, $B2_i$ and r_i are parameters depending only on the nature of the index atom (in this case i). $\tilde{A1}_j$, $\tilde{A2}_j$, $\tilde{B1}_j$ $\tilde{B2}_j$ is a different parameter set which also depends only on the nature of the index atom (in this case j). The first term in (8) corresponds to the repulsion, the second term to the attraction; m and n are power parameters which shall be valid for all types of atoms. The justification for the choice of this rather complicated formula (8) will be given step by step below.

Since there is no rigorous mathematical equivalence between (8) and (7) one should confine oneself to concrete atoms. The following six types of atoms have been parametrized for (8): H, C, N, O, S and AH (aromatic hydrogen atom). We do not claim having chosen the optimal parametrization. The purpose is to demonstrate that there are parametrizations for (8) which approximate the Lennard-Jones curves sufficiently for practical purposes. A set of Lennard-Jones potential parameters used in Molecular Mechanics has been taken from Stuper et al.[32].

The numerator of the terms in (8) is the result of an attempt to separate the A_{ij} and B_{ij} parameters of (7) in a product of two independent terms. In the case of a simple separation in the form of:

$$A_{ij} = A_i \cdot \tilde{A}_j \quad , \quad i,j \varepsilon \{H,C,N,O,S,AH\} \qquad (9)$$

only four parameters (A_i, B_i, $\tilde{A}_i$, $\tilde{B}_i$) would be related to each

atom so that one has only 24 parameters instead of 42 Lennard-Jones parameters in the case of six atom types. But curve fitting attempts show that the reproduction of the original Lennard-Jones curves are not very satisfactory. However, for the needs of the drug design the simple separation in (9) might be useful. The approach:

$$A_{ij} = A1_i \cdot \tilde{A1}_j + A2_i \cdot \tilde{A2}_j \qquad (10)$$

offers a greater flexibility with the disadvantage of an increased number of parameters. Eight parameters are now related to each atom. These 48 parameters (in the case of six atoms) however, are not independent of each other. A symmetry, as $A_{ij} = A_{ji}$, is not demanded. With the intermediate formula:

$$E_{ij} = \frac{A1_i \cdot \tilde{A1}_j + A2_i \cdot \tilde{A2}_j}{r_{ij}^{12}} - \frac{B1_i \cdot \tilde{B1}_j + B2_i \cdot \tilde{B2}_j}{r_{ij}^{6}} \qquad (11)$$

the original Lennard-Jones curves can be approximated fairly well by using appropriate parameters. The quality of the approximation is comparable to that in FIGURE 6.
The advantage of equation (11) is that now the steric interaction between two molecules can be formulated by potential functions. Yet there are two potentials for the repulsion as well as for the attraction.

Let us regard the denominator of equation (8). This reformulation was caused by the results of the numerical experiments of the preceding section. There the sensitivity of the continuum representation has been demonstrated in respect to deviations from the equidistancies of the front atoms. Let us consider the contact region of two molecules in equilibrium which sterically fit well. Since the front atoms of the two molecules are of different size the lines (see FIGURE 5 at right) going through the front atoms do not run equidistantly to one another (more precisely: the surfaces going through the front atoms are not equidistant). Putting a sphere of an appropriate radius, according to the atom type, around each atom and regarding

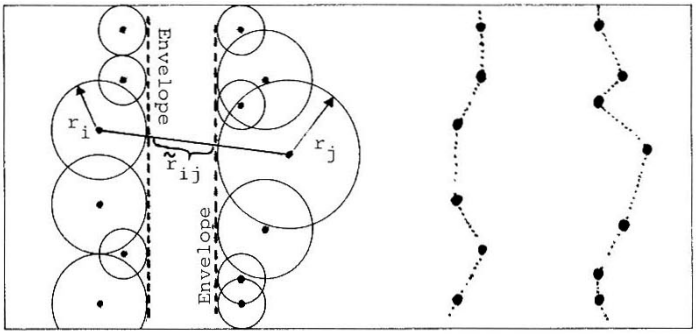

FIGURE 5: Schematic representation of steric complementarity in the case of good steric fit. The equilibrium position of the front atoms are shown on the right (no equidistancy). The equidistancy of the envelopes in the contact region is demonstrated on the left for a suitable choice of atom radii.

the envelopes of these spheres, the equidistancy in the contact region can be stated. This is schematically represented in FIGURE 5 (at left). Now steric fit can be defined in the sense of the equidistancy of the envelopes in the contact region. The equilibrium distance of the envelopes depends on the choice of the used atom radii. By (8) the repulsion as well as the attraction is referred to the "new" distance $\tilde{r}_{ij} = r_{ij} - r_i - r_j$. The next publication deals with systems in which the atoms of one molecule are unknown (r_j unknown), but there is some idea of the envelope. I.e. instead of r_{ij} there is given r_{iM}, the distance between the i^{th} effector atom and a point on the envelope lying in front of the j^{th} atom of the receptor:

$$\tilde{r}_{ij} = r_{iM} - r_i = r_{ij} - r_i - r_j \qquad (12)$$

Obviously these atom radii r_i have nothing to do with the absolute atom size. They shall only cause equidistancy of the front atoms in a steric fit. If a constant value is added to each of these radii, the equidistancy would not be changed, only the distance between the fronts.
The calculation of a set of atom radii is quite simple:

Referring to a constant distance, R_{const}, the following system is fitted by least square:

$$r_i + r_j + R_{const} = d^\circ_{ij}, \qquad i,j=1,\ldots,6 \tag{13}$$

where d°_{ij} is the equilibrium distance (minimum of the Lennard-Jones curve). For $R_{const}=2\mathring{A}$ the results, r_i, of the least square calculation are shown in Table 1.

i	j	d°_{ij}	E°_{ij}	$r_i+r_j+R_{const}$	r_i
H	H	2.373	-.197	2.284	0.1422
H	C	2.687	-.233	2.697	
H	N	2.680	-.175	2.649	
H	O	2.550	-.224	2.548	
H	S	2.950	-.220	2.881	
H	AH	2.050	-.221	2.149	
C	C	3.120	-.283	3.110	0.5549
C	N	3.020	-.331	3.062	
C	O	2.900	-.375	2.961	
C	S	3.330	-.366	3.293	
C	AH	2.687	-.233	2.562	
N	N	3.030	-.218	3.014	0.5068
N	O	2.960	-.285	2.913	
N	S	3.180	-.452	3.245	
N	AH	2.680	-.175	2.513	
O	O	2.820	-.345	2.813	0.4064
O	S	3.080	-.455	3.145	
O	AH	2.580	-.225	2.413	
S	S	3.500	-.458	3.477	0.7385
S	AH	2.950	-.220	2.745	
AH	AH	1.960	-.196	2.013	0.0067

TABLE 1: Result of the calculation of the atom radii, r_i, with $R_{const}=2A$. E°_{ij} and d°_{ij} are energy minimum and equilibrium distance in the Lennard-Jones (LJ) curves (see also [32]).

The power parameters m and n in (8) have been obtained in an intuitive way by numerical experiments:

$$n=5 \text{ and } m=7.5 \tag{14}$$

Having $r_1,\ldots,r_6$, m, and n, the attempt can be undertaken to determine the 48 parameters for the considered atoms. 36 curves of type (8) must be simultaneously adapted to the corresponding

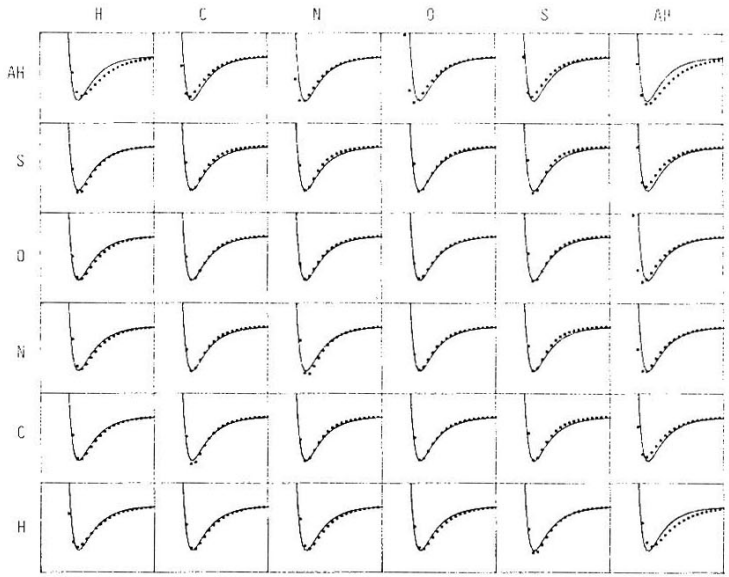

FIGURE 6: Atom-atom interaction curves for H, C, N, O, S and AH. The solid line is the Lennard-Jones curve. The dotted lines correspond to (8a) with the fitted parameters of table 1, 2a and 2b. The curve i-j (according to (8a)) is not necessarily identical with the curve j-i, but the deviations are small. The plot units of the curves are different. The limits of each square are: Xmin=0.5d°_{ij}, Xmax=2d°_{ij}, Emin=1.5E°_{ij}, and Emax=-0.5E°_{ij}; the values of d°_{ij} and E°_{ij} must be taken from Table 1.

LJ curves (because of the non a priori symmetry there are 36 instead of 21 curves). The previously mentioned difficulty about the interdependence of the parameters could not be eliminated by the arbritrary choice

$$\tilde{A1}_H = \tilde{A2}_H = \tilde{B1}_H = \tilde{B2}_H = 1.$$

A non-linear least square fit has been performed using the NAG (National Physical Laboratory Algorithms Library) routine E04BGF. For reasons of scaling the equation to be fitted is:

$$E_{ij}=\frac{100(A1_i\cdot\tilde{A1}_j+A2_i\cdot\tilde{A2}_j)}{(r_{ij}-r_i-r_i)^{7.5}}-\frac{10(B1_i\cdot\tilde{B1}_j+B2_i\cdot\tilde{B2}_j)}{(r_{ij}-r_i-r_i)^{5}} \qquad (8a)$$

Because of the parameter interdependence there are many possible combinations of parameter values producing the global least square minimum. One possible result is given in the Tables 2a and 2b. In FIGURE 6 the 36 curves of (8a) with parameters of Tables 2a and 2b as well as (14) are compared with the corresponding LJ curves (solid lines). The qualitative conformity of these curves can be considered as satisfactory and certainly it can be improved (e.g. by other exponents m and n, by other r_i and finally by fit methods not in the sense of pointwise least square).

i	A1	A2	B1	B2
H	1.84161	-1.06662	0.26044	1.70084
C	2.00731	-1.08519	0.40352	1.97661
N	2.22414	-1.35496	0.40431	1.65870
O	2.33025	-1.39237	0.41976	1.94503
S	2.45380	-1.30935	0.56416	2.19758
AH	1.51968	-0.83812	0.22141	1.63063

TABLE 2a: Result of the non-linear least square fit of (8a). The parameters A1, A2, B1 and B2 refer to the atoms of the effector molecule and are used for the computation of the "steric effector potential".

j	$\tilde{A1}$	$\tilde{A2}$	$\tilde{B1}$	$\tilde{B2}$
H	1.00000	1.00000	1.00000	1.00000
C	1.73945	2.13877	3.06100	0.93069
N	0.64741	0.30293	3.80558	0.63020
O	0.90678	0.68634	3.43938	0.86371
S	2.27091	2.84793	5.57822	0.76961
AH	1.17798	1.39489	0.42621	1.02363

TABLE 2b: Result of the non-linear least square fit of (8a). The parameters A1, A2, B1 and B2 refer to the atoms of the (atomistic) receptor. The parameters of TABLE 2a and this one are to be used only in connection with the atom radii in TABLE 1 and the power parameters of (17).

9. Discussion

In section 2 (Purpose and Method) the aim of the chemical complement simulation was stated. In order to formulate a useful algorithm some preliminary requirements had to be satisfied. These were: 1) The validity of the projection of the intermolecular interaction onto a two-dimensional manifold, and 2) the separability of the interaction terms (i.e. the $\underline{\sigma}\cdot\underline{\rho}$ Ansatz). Using simple interaction models based on Molecular Mechanics as a standard, the applicability of these two approximation principles was demonstrated. As a consequence, the mathematical problem can be stated in a linear manner. In section 6 it has been shown that a very good approximation to the Lennard Jones curves could be obtained by introducing two additional potential functions. We believe, however, that the simple separation (only two steric interactions) as stated in (9) combined with a suitable choice of atomic radii are completely sufficient for the needs of drug design. In this case the chemical complement simulation is equivalent to the following problem:
Find a two-dimensional manifold (considered as fixed in space) and three functions e(M), A(M), B(M) defined on this manifold such that the interaction between a guest (or effector) molecule in <u>any orientation</u> in space and the binding site on the unknown host is given by:

$$\Delta E = \int_M \{(\Sigma \frac{e_j}{r_{jM}})e(M) + (\Sigma \frac{A_j}{(r_{jM}-r_j)^m})A(M) - (\Sigma \frac{B_j}{(r_{jM}-r_j)^n})B(M)\}dM \quad (15)$$

The round brackets are potentials of the known effector molecule at a point M on the manifold. r_{jM} is the distance between the j^{th} atom of the effector and a point M on the manifold. r_j, A_j and B_j are known constants depending on the j^{th} atom. It should be emphasized that (15) is valid for nonequilibrium orientations of the effector as well.
An algorithm for the computation of the manifold and the parameter functions using the experimental equilibrium interaction energies (or quantities which are proportional to them) of a given training set will be given in the subsequent paper.

References

1) V. Brodbeck, Nachr. Chem. Tech. Lab. 30, 695 (1982).
2) P.P. Mager & A. Seese, Pharmazie in unserer Zeit, 10, 97 (1981).
3) N.B. Chapman & J. Shorter (Eds.), Advances in Linear Free Energy Relationships, Plenum Press, New York (1972).
4) C. Hansch in: Structure Activity Relationships, C.J. Cavallito (Ed.), Vol. 1, p. 75, Pergamon Press, New York (1973).
5) S.M. Free & J.W. Wilson, J. Med. Chem., 7, 395 (1964).
6) B.R. Kowalski in: Computers in Chemical and Biochemical Research, C.E. Klopfenstein & C.L. Wilkins (Eds.), Vol. 2, p. 1, Academic Press, New York (1974).
7) K-L. H. Ting, R.C.T. Lee, G.W.A. Milne, M. Shapiro & A.M. Guarino, Science, 180, 417 (1973).
8) W.J. Dunn III & S. Wold, Bioorg. chem. 9, 505 (1980).
9) A. Pedro & F. Lehmann in: Receptors and Recognition Series, A.P. Cuatrecasas & M.F. Greaves (Eds.), Vol. 5, p. 2, Chapman & Hall, London (1978).
10) R. Croxatto & F. Huidobro, Arch. Int. Pharmacodyn., 106, 207 (1956).
11) J.E. Amoore, G. Palmieri & E. Wanke, Nature, 216, 1084 (1967).
12) J.E. Amoore, G. Palmieri, E. Wanke & M.S. Blum, Science, 165, 1266 (1969).
13) E.C. Olson & Christoffersen (Eds.), Computer-Assisted Drug Design, ACS Symposium Series 112, American Chemical Society, Washington D.C. (1979).
14) P. Gund, Progress in Molecular and Submolecular Biology, F.E. Hahn (managing Ed.), Vol. 5, p. 117, Springer-Verlag, Berlin (1977).
15) A.T. Balaban, A. Chiriac, I. Motoc & Z. Simon, Steric-Fit in Qualitative Structure-Activity Relations. Lecture Notes in Chemistry, Nr. 15, Springer-Verlag, Berlin (1980).
16) S. Maayani, H. Weinstein, S. Cohen & M. Sokolovsky, Proc. Nat. Acad. Sci. USA, 70, 3103 (1973).

17) F. Vögtle (Ed.), Host-Guest Complex Chemistry II, Springer-Verlag, Berlin (1982).
18) D.J. Cram in: Structure and Dynamics in Chemistry, P. Ahlberg & L.O. Sundelöf (Eds.), Almquist & Wiksell, Stockholm (1978).
19) A.J. Hopfinger, Intermolecular Interactions and Biomolecular Organization, John Wiley & Sons, New York (1977).
20) P. Kent & T. Gaumann, Helv. Chim. Acta, 58, 787 (1975).
21) C.L. Perrin, Science, 183, 551 (1974).
22) J.T. Clerc, P. Naegeli & J. Seibl, Chimia, 27, 639 (1973).
23) H. Margenau & N.R. Kestner, Theory of Intermolecular Forces, Pergamon Press, New York (1971).
24) B. Pullman (Ed.), Intermolecular Interactions: From Diatomics to Biopolymers, John Wiley and Sons, New York (1978).
25) E. Clementi, Computational Aspects for Large Chemical Systems, Lecture Notes in Chemistry, Vol. 19, Springer-Verlag, Berlin (1980).
26) S. Fraga, K.M.S. Saxena & M. Turres, Biomolecular Information Theory, Elsevier, Amsterdam (1978).
27) P. Claverie in: Intermolecular Interactions: From Diatomics to Biopolymers, B. Pullman (Ed.), John Wiley & Sons, New York (1978).
28) R.S. Mulliken & W.B. Person, Molecular Complexes, Wiley-Interscience, New York (1969).
29) L.P. Hammett, Physikalische Organische Chemie, Verlag Chemie, Weinheim (1973).
30) H. Späth, Spline-Algorithmen, Oldenburg Verlag, München, (1973).
31) J.V. Beck & K.J. Arnold, Parameter Estimation in Engineering and Science, John Wiley & Sons, New York (1977).
32) A.J. Stuper, T.M. Dyott & G.S. Zander in: Computer-Assisted Drug Design, ACS Symposium Series 112, E.C. Olson & R.E. Christoffersen (Eds.), American Chemical Society, Washington D.C. (1979).