match no. 19 pp. 67-87 1986

CHEMICAL COMPLEMENT SIMULATION AND QUANTITATIVE STRUCTURE ACTIVITY RELATIONSHIP (II)

H.R. Karfunkel

Institut für Physikalische und Theoretische Chemie,
Universität Tübingen, Auf der Morgenstelle 8,
D - 7400 Tübingen, FRG

(Received: February 1985)

Summary

The method of Chemical-Complement-Simulation (CCS) deals with the construction of a chemical analogue to a structurally unknown receptor using the binding data of known effectors. This receptor analogue, represented by a 2D-manifold **M**, is used to predict the binding energies of new effectors. The continuous receptor analogue is derived from the overlapping effectors using a structure termed a superenvelope. The steps leading to the simulated receptor are formulated as mathematical algorithms. The minimization of a cost function is recommended for the solution of the matching problem leading to the superenvelope. The simulated receptor is obtained by the solution of a set of geometric programming problems where the constraints reflect the necessity of taking equilibrium conditions into account. A clear distinction between binding and activity data is made. When activity data are supplied in addition to binding data, the new concept, pressure pattern, is introduced for the prediction of biological activity. A generalization (based on the theory of elasticity) of the CCS method to flexible receptors is proposed.

1. Introduction

The method of chemical complement simulation deals with the following standard problem: Given is a set (training set) of sructurally well-defined molecules, called effectors, which bind to the same binding site of a structurally unknown receptor. The binding energies (or the values which correlate with them) are known as well. **A chemical model of the binding site which leads to the prediction of the binding energies of new molecules** (effector candidates) **is desired.** Should the biological activity data be supplied in addition to the binding data

(which is needed for the construction of the receptor model) then the receptor model should also predict the biological activity of new effector candidates. An atomistic representation of the receptor requires more parameters than the information at hand. Due to this reason an atomistic receptor model must be rejected. In the method described below the binding site of the unknown receptor is represented by a continuum. This continuum consists of a 2D - manifold which runs parallel to the front atoms of the binding site.

It was shown in the previous article[1] that the intermolecular interaction between two molecules (effector and receptor) can be described by using the continuum representation for one of them (receptor). In order to reproduce the results of molecular mechanics (MM), three requirements must be fulfilled:

a) A 2D-manifold $\mathbf{M}(u,v)$ (u and v are surface parameters) which replaces the atomistic receptor must be supplied.

b) A reformulation of the interaction equations of MM must be undertaken: The local interaction is described by the scalar product $\rho * \sigma$ in which the components of σ are potential functions **dependent only on the effector**, and ρ is a vector function dependent **only on the receptor**. The global interaction is calculated by the surface integral of the scalar product over the manifold. Here it should be noted that in recent molecular dynamical simulation of water - protein interactions the aforementioned separation ($A_{ij} = \alpha_i * \alpha_j$ and $B_{ij} = \beta_i * \beta_j$) has been performed[2].

c) An algorithm for the parametrization of the manifold using the "experimental" binding data is required. The parametrization of $\mathbf{M}(u,v)$ means: defining functions $\rho_1(u,v)$, $\rho_2(u,v)$ .... on $\mathbf{M}(u,v)$. If we restrict ourselves to electrostatic interactions as well as steric attraction and repulsion of the Lennard-Jones type than the parametrization of the receptor consists of the three functions:

$$\rho(u,v)^T = (A(u,v), B(u,v), e(u,v))$$

The approximation of these functions by a linear combination of suitable basis functions is performed as usual.

The interaction energy of any effector molecule in a given

spatial orientation to the receptor is, in the example above

$$\Delta E = \int_{M} [A(M) \sum_{j} \frac{A_j}{(r_{jM}-r_j)^m} - B(M) \sum_{j} \frac{B_j}{(r_{jM}-r_j)^n} + e(M) \sum_{j} \frac{e_j}{r_{jM}}] dM \qquad (1)$$

r_{jM} is the distance between the j^{th} atom of the effector and a point M on the manifold **M**. r_j, A_j and B_j are constants depending on the j^{th} atom. It should be emphasized that **(1)** is valid for nonequilibrium orientations of the effector as well. The minimum of ΔE with respect to the orientation is the equilibrium state where ΔE corresponds to the measurable binding energy.

The following section deals with point a) i.e. the construction of **M**(u,v).

2. The Supermolecule

Unfortunately the exact spatial orientation of an effector bound to an unknown receptor in a presupposed position is unknown. If the relative orientation of the bound effector molecules to each other is known, then the concept of "excluded volume" according to Marshal et al.[3] or the **supermolecule** according to Balaban et al.[4] can be applied. The supermolecule is the set of atoms resulting from the union of all the effector's atoms in the training set. The construction of the supermolecule is relatively simple when the pharmacophore pattern[5] of the receptor is known (even if parts of the effector molecules are flexible). The construction of the supermolecule is, however, significantly more difficult when the pharmacophore pattern is unknown. The fact that, in certain enzymatical reactions, specified chemical bonds are cleaved, can be used to construct the supermolecule by trying to match this bonds. This problem, also known as the "matching problem", will be dealt with in later sections.

The **superenvelope** which plays a central role in the following discussion, is constructed from the supermolecule. If each atom i of the supermolecule is substituted by a sphere of radius r_i+ R where r_i is an atomic radius and R is a distance parameter, then the surface of this construction can be regarded as a su-

16B,17-BUTANOMORPHINAN-3-OL (AC,B37,188)

MORPHIN C17H19NO3 (AC,B29,1630 COTT.)

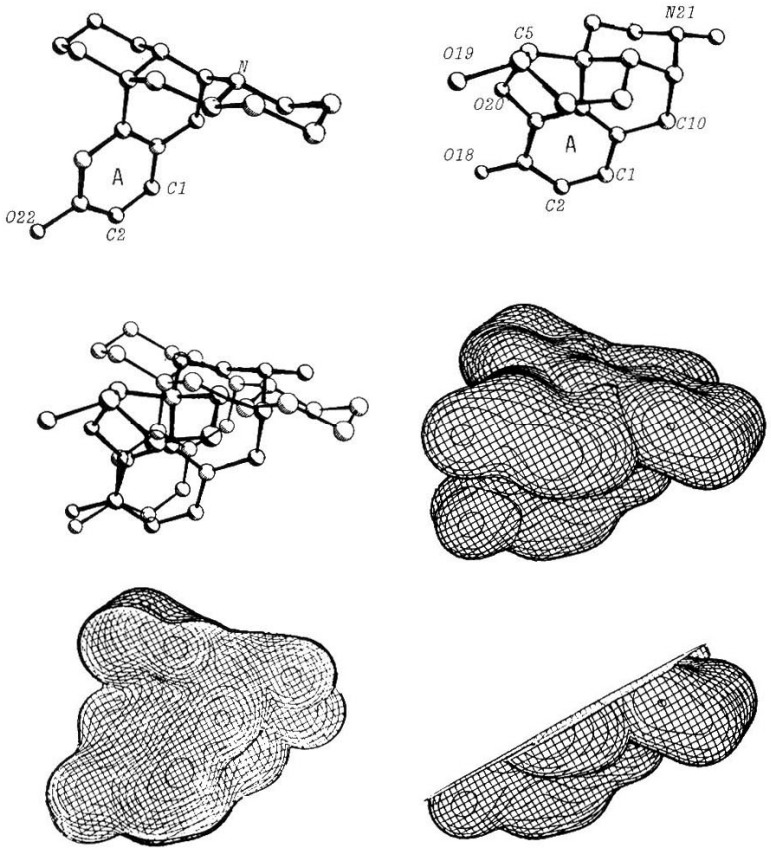

FIGURE 1: Derivation of 2D-manifolds as receptor candidates (an example). Top: The training set (generally 10 to 100 different effector molecules). Middle left-hand: The supermolecule resulting from the matching of effectors. Middle right-hand: The superenvelope. Bottom: Two different manifolds (receptor candidates).

perenvelope $\mathbf{S}_R$ referring to the distance parameter R. Using R as a parameter, the superenvelopes $\mathbf{S}_R$ form a family of nonintersecting closed surfaces. This concept is illustrated for a fixed R in FIGURE 1.

Based on the considerations presented in the previous paper[1] the following statement can be asserted when some rigidity conditions hold: **The superenvelope $\mathbf{S}_R$ runs parallel to the envelope enclosing the front atoms of the receptor at the binding site.** $\mathbf{S}_R$ will coincide (up to minor deviations) with the receptor envelope when a suitable distance parameter is chosen. In other words: A superenvelope can be divided into two parts with the relevant part oriented toward the front atoms of the receptor binding site. **This part of the superenvelope is to be used as a first approximation of the continuous receptor analogue $\mathbf{M}(u,v)$** by parametrizing it with an adequate vector function. Simplified, one would proceed in the following manner: Various manifolds can be "cut" from a given $\mathbf{S}_R$ and then parametrized in a least square sense with a vector function. The manifold delivering the best correlation between experimental and theoretical (predicted) binding energies is the desired approximation for the continuous representation of the receptor.

3. The Optimization Problem

The proposal in the previous section (to use the manifold which yields the best correlation between theory and experiment after parametrization as a continuous representation of the receptor) is plausible. **The results of this proposal are,however,useless.** Numerous numerical experiments have shown that it is impossible to discriminate between the different $\mathbf{M}(u,v)$ candidates. Nearly all $\mathbf{M}(u,v) \subseteq \mathbf{S}_R(u,v)$ which were parametrized with a suitable vector function delivered exellent correlations between theory and experiment. Trying to reproduce the binding energies without taking the equilibrium conditions into account caused the failure. The continuous receptor must be parametrized in such a way that the computed binding energies **are the global minima with respect to any spatial variation of the effectors.** The spatial variations of a rigid effector are the three rotations

and the three translations. In the case of flexible effectors, the internal degrees of freedom must also be considered. For each effector (index k) there are m_k>=6 degrees of freedom designated by $\tau_k^T=(\tau_1,\ldots,\tau_{mk})$. The equilibrium conditions can now be formulated as auxiliary conditions leading to a geometric (quadratic) programming problem.

The following designations will be used:

Let $\mathbf{M}(u,v) \subseteq \mathbf{S}(u,v)$

$k=1,\ldots,K$	the effector index (training set)
$E_{\mathbf{exp}}^T=(E_{expl},\ldots,E_{expK})$	the vector of experimental binding energies
$E^T=(E_1,\ldots,E_K)$	the vector of theoretical binding energies
σ_k^T	the vector potential function of k^{th} effector
I	no. of different potential functions (dimension of σ or ρ)
$i=1,\ldots,I$	index for potential functions
L_i	no. of basis functions for the i^{th} parameter function ρ_i
l_i or $l=1,\ldots,L_i$	running index
$E_k=\int \sigma_k(u,v)*\rho(u,v)dM$	the energy expression

From the Ansatz:

$$\rho_i=\sum c_{il}\phi_{il},\ i=1,\ldots,I \qquad (2)$$

where $\{\phi\}$ are known basis functions and $\{c\}$ are the unknown coefficients, the energy expression E_k is:

$$E_k=\sum_{i}^{I}\sum_{l}^{L_i}\int_M \sigma_{ki}(u,v)*\phi_{il}(u,v)dM*c_{il} \qquad (3)$$

Ordering the unknown c_{il} (by row) in a vector $\mathbf{X}$ of the dimension $N=\sum L_i$ and collecting the basis functions ϕ_{il} into a vector function φ (of dimension N as well), the following holds for the predicted interaction energies:

$$\mathbf{E} = \mathbf{F_{kn}} * \mathbf{X} \qquad (4)$$

where $\mathbf{F_{kn}}$ is a K times N matrix, whose elements, F_{kn}, are com-

posed of the following integrals:

$$F_{kn} = \int \sigma_{ki_n} \phi_n dM \qquad (5)$$

The fact that the index i_n refers to several indices n must be taken into account. This can be illustrated as follows: The K times N matrix is the integral of the following matrix:

$$\begin{matrix} \sigma_{11}\phi_{1} \cdots \sigma_{11}\phi_{l_1} & \sigma_{12}\phi_{l_1+1} \cdots \sigma_{12}\phi_{l_1+l_2} & \cdots \sigma_{1I}\phi_{N} \\ \vdots & & \vdots \\ \sigma_{k1}\phi_{1} \cdots \sigma_{k1}\phi_{l_1} & \sigma_{k2}\phi_{l_1+1} \cdots \sigma_{k2}\phi_{l_1+l_2} & \cdots \sigma_{kI}\phi_{N} \\ \vdots & & \vdots \\ \sigma_{K1}\phi_{1} \cdots \sigma_{K1}\phi_{l_1} & \sigma_{K2}\phi_{l_1+1} \cdots \sigma_{K2}\phi_{l_1+l_2} & \cdots \sigma_{KI}\phi_{N} \end{matrix} \qquad (6)$$

Since there are no unknowns in the above matrix the integration over a manifold **M** can be performed.
The N parameters in **X** replace the I unknown parameter functions ρ_i and characterize the simulated continuous receptor.
Let τ_{km}, $m=1,\ldots,m_k$, $m_k>=6$, be the degrees of freedom of the k^{th} effector. For the sake of clarity only rigid effectors will be considered i.e. $m_k=6$ for all k. The variations of the energy of the k^{th} effector with respect to the three translations and rotations are:

$$\frac{\partial E_k}{\partial \tau_m} = \sum_{n}^{N} \frac{\partial F_{kn}}{\partial \tau_m} * X_n \qquad , m=1,\ldots,6 \qquad (7)$$

where

$$\frac{\partial F_{kn}}{\partial \tau_m} = \int_M \frac{\partial \sigma_{ki_n}}{\partial \tau_m} * \phi_n dM \qquad (8)$$

A spatial variation, $\partial\tau_m$, causes only a change in the value of the effector potential on **M**. The quantities on the right-hand side of **(8)** are known or computable, so that $\partial F_{kn}/\partial\tau_m$ could be computed by numerical integration. The expression in **(8)** can be written (for the whole training set) as a 6K times N matrix **G**. This matrix **G** is the integral of the following elements:

$$\begin{matrix} \vdots & \vdots & \vdots \\ \frac{\partial \sigma_{kl}}{\partial \tau_1} * \phi_1 & \cdots\cdots & \frac{\partial \sigma_{kl}}{\partial \tau_1} * \phi_{l_1} & \cdots\cdots & \frac{\partial \sigma_{kI}}{\partial \tau_1} * \phi_N \\ \vdots & & \vdots & & \vdots \\ \frac{\partial \sigma_{kl}}{\partial \tau_6} * \phi_1 & \cdots\cdots & \frac{\partial \sigma_{kl}}{\partial \tau_6} * \phi_{l_1} & \cdots\cdots & \frac{\partial \sigma_{kI}}{\partial \tau_6} * \phi_N \\ \vdots & & \vdots & & \vdots \end{matrix} \qquad (9)$$

The equilibrium conditions, i.e. the invariance of the binding energy with respect to spatial variations of the effectors can be expressed as:

$$\mathbf{G} * \mathbf{X} = \mathbf{0} \qquad (10)$$

The optimization problem can be summerized as follows:
For a given manifold, $\mathbf{M} \subseteq \mathbf{S}_R(u,v)$, the optimal parametrization is that which minimizes the quadratic function:

$$SSQ(\mathbf{M},X) = (E_{exp}-F*X)^T*(E_{exp}-F*X) \qquad (11)$$

taking the 6K homogenuous linear auxiliary conditions **(10)** into account. Out of all possible manifolds $\mathbf{M} \subseteq \mathbf{S}_R$, that which delivers the best correlation between experimental and theoretical binding energies is taken as the continuous receptor model. Numerical experiments have shown that the auxilliary conditions **(10)** are too restrictive. Minor errors in the construction of the superenvelope prohibit good correlations. Therefore, the conditions **(10)** are substituted by "milder" auxiliary conditions:

$$\begin{aligned} &\mathbf{G} * \mathbf{X} <= \mathbf{Eps} \quad \text{and} \\ &-\mathbf{G} * \mathbf{X} <= \mathbf{Eps} \end{aligned} \qquad (12)$$

where **Eps** is an empirical predetermined parameter vector which expresses the degree of restrictiveness of the equilibrium conditions. When the components of **Eps** are too small, a poor correlation between experiment and theory will result for all the receptor canditates $\mathbf{M} \subseteq \mathbf{S}_R$. If on the other hand the components of **Eps** are too large, a good correlation for all the receptor

candidates will be the result. In both cases a reliable discrimination between the different receptor candidates would be impossible.

4. Numerical Experiments

In the previous section the mathematical concepts and formalism of the CCS method were established. In this section some practical and numerical aspects will be presented. The first question posed by the practician may be: Can the CCS method work at all? Upon initial examinations a very simple example was required. Simplicity means: A sufficiently large training set of two dimensional rigid molecules (like the bases of nucleic acids) binding to the same (rigid) binding site. Since we have not found an example with unique and reliable binding data, the following method was used: We "invented" simple examples fulfilling all the requirements of the method using MM. Using the interaction energies resulting from the MM calculations of these hypothetical molecules as experimental binding energies and then "forgetting" the atomistic receptor (hypothetical too) in use,an attempt was made to apply the CCS metod to that data. Doubts in conjunction with the "matching problem" were also eliminated since the relative orientations of the effectors are known. The arguments favouring this line of reasoning are: If MM is an adequate tool for the description of intermolecular interactions in the frame of drug design, and if the CCS method is capable of reproducing the results of MM for hypothetical systems, then there is no reason why the CCS method should fail in real situations.

The superenvelope of a training set consisting of two dimensional molecules is a one dimensional manifold $\mathbf{S}_R(t)$ (see FIGURE 2). At this point a general question must be posed: How should a comlex surface like a superenvelope be presented? Generally a superenvelope is composed of hundreds of atoms, making an analytical representation $\mathbf{S}_R(u,v)$ with surface parameters u and v impossible. A representation of a surface suited for numerical purposes is a set of representative surface points. However, in order to substitute surface integrals by surface sums, this set

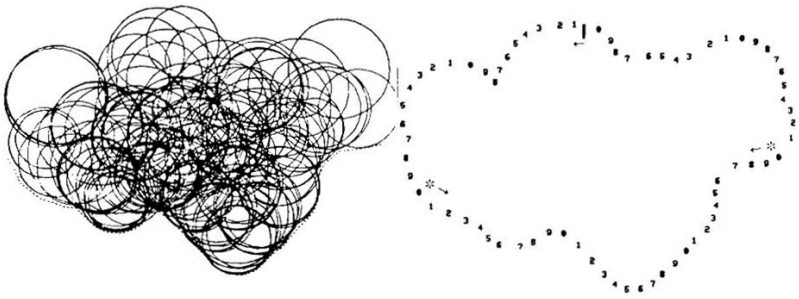

FIGURE 2: Numerical experiments with two-dimensional systems. The training set consists of 8 effectors. Left-hand: The supermolecule. The radii of the circles are the sum of atomic radii and a distance parameter R. The dashed line represents the original receptor (this information is not included in the data set). Right-hand: The superenvelope represented by a set of uniformly distributed points. The points are numbered, but only the last digit is shown. Any receptor candidate can be characterized by the starting point and the length (see FIGURE 3). The receptor candidate delivering the best correlation (Eps=.2) is marked by *.

of surface points should be uniformly distributed. Even covering a simple sphere with a large number of uniformly distributed points is not trivial[6]. A very fast and efficient algorithm for the sampling of a superenvelope with nearly uniformly distributed points, in which atoms not involved with the surface are automatically eliminated, is to be published elsewhere[7].

We are not intending to represent the details of the numerical experiments with these hypothetical systems, (for details see ref. 8) but rather only relevant aspects. Any connected one-dimensional receptor candidate **M**(t) derived from the superenvelope can be defined by two numbers, the starting point on the superenvelope and the length (counterclockwise). In the example represented in FIGURE 2 there are about 4000 possible receptor candidates. For simplicity reasons it was assumed that the steric parameters A and B of the receptor were constants. The distribution of charge e(t) on **M**(t) was expressed with the aid of 6 sine and 6 cosine functions so that any given **M**(t) was char-

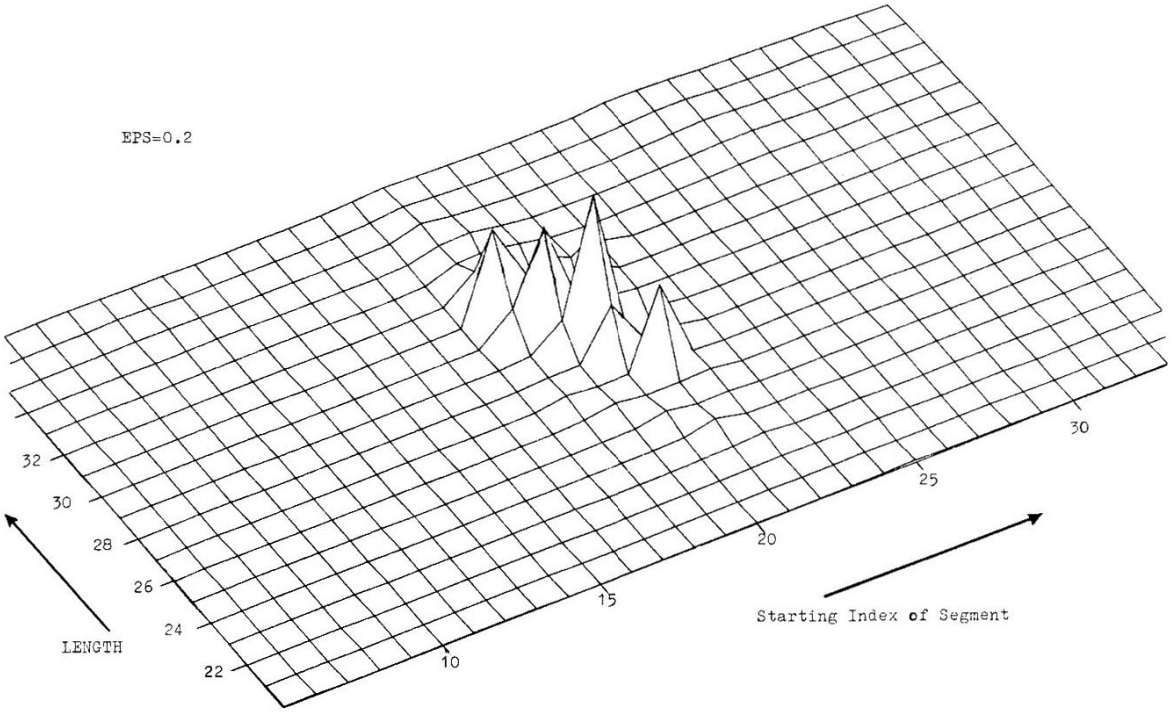

FIGURE 3: The quality of the correlation between thoery and "experiment" for about 500 receptor candidates. Each candidate is represented by a point in the plane. The height above the plane is 1/SSQ (see eq. (11)).

Effector	E-exp	With EPS=0.2		With EPS=0.3	
		E-Theor.	E-Th.Eq	E-Theor.	E-Th.Eq
1	-1.8342	-1.8536	-1.8602	-1.8328	-1.8549
2	-2.3636	-2.4289	-2.4352	-2.4121	-2.4193
3	-2.0369	-1.9046	-1.9093	-1.9820	-1.9907
4	-2.4943	-2.4230	-2.4291	-2.5212	-2.5334
5	-1.7058	-1.8409	-1.8558	-1.7576	-1.7839
6	-3.5730	-3.3703	-3.3759	-3.5479	-3.5611
7	-2.6204	-2.7971	-2.7973	-2.5932	-2.5937
8	-1.3991	-1.3204	-1.3231	-1.3810	-1.3870
Correl. Coef.		.9809	.9807	.9984	.9980
Start. Index		20		21	
Termin. Index		49		48	
9	-1.7343		-1.8147		-1.8016
10	-1.9369		-2.0634		-2.0377

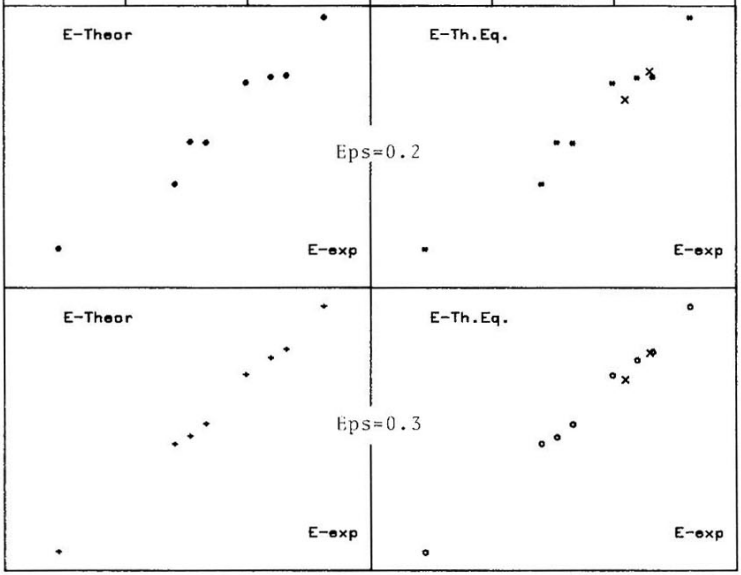

FIGURE 4: The results of the numerical experiments. The verification set consists of effectors 9 and 10. E-Th.Eq is the predicted binding energy after relaxation to equilibrium.

acterized by 15 parameters. The results for a training set consisting of 8 effectors and a verification set consisting of 2 molecules are represented in FIGURE 3 and 4. In FIGURE 3 the results of the optimization (parametrization) can be seen for about 500 of the most relevant receptor candidates. The points in the xy-plane represent the different receptor candidates and the height above this plane corresponds to the correlation between experimental and theoretical binding energies. The choice of the parameter **Eps** which expresses the restrictiveness of the equilibrium conditions was not very critical. Out of 4000 possible candidates only a very limited number of **M**(t) delivered a significant correlation between experiment and theory,and belonged to the original binding region. The results are summarized in FIGURE 4. It should be rememered that the effectors of the training set are not in exact equilibrium with the simulated receptor (due to **Eps**>0). When the effectors were allowed to relax into their new equilibrium with respect to the continuous receptor model, no significant changes in the computed binding energy (E-Th.eq in FIGURE 4) nor in the spatial orientation were caused. By allowing the receptor model to interact with the 2 molecules of the verification set, the reliability of the predictions were demonstrated as well.

5. The Matching Problem

The success of the CCS method depends to a great extent on a reliable superenvelope construction. The construction of the superenvelope is a point where a large amount of additional information about the considered receptor could be integrated in a natural way. A very useful concept mentioned above, is that of a pharmacophoric pattern. Complications arise when dealing with flexible molecules or molecules which may adopt several stable conformations. Forcing the set of K effector molecules to match at presupposed centers will cause an unrealistically high strain in some of these molecules. Allowing them on the other hand to adopt their exact (but flat) minimum energy conformation will result in a poor match. The matching problem can be successfully managed by minimizing a suitable cost function

which takes both effects into account. The cost function depends on the 6(K-1) translations and rotations of the effectors (except the first one) and the internal degrees of freedom. If q is a measure for the lack of matching then:

$$\text{Costs} = w_1 q + w_2 \sum \Delta St_i \qquad \textbf{(13)}$$

where w_1 is the relative weight of the matching, w_2 the relative weight of the strain and ΔSt_i the additional strain energy of the i^{th} effector molecule. ΔSt_i depends only on the internal degrees of freedom of the i^{th} effector whereas q depends on the translations and rotations as well.

If some of the effectors have flexible parts which do not influence q (i.e. large substituents not involved in the pharmacophoric pattern), then cost function **(13)** is not sufficiently general for a useful superenvelope construction. In such a case we recommend the usage of the following cost function:

$$\text{Costs} = w_1 q + w_2 \sum \Delta St_i + w_3 N_s \qquad \textbf{(14)}$$

where w_3 is a weight factor and N_s the surface area of the superenvelope. N_s depends on the internal and external dgrees of freedom. **A measure of N_s is the number of uniformly distributed points needed to cover the superenvelope when the mean nearest neighbour distance is held constant.** The algorithm used for the superenvelope representation[7] is ideally suited for this purpose (A FORTRAN 77 source can be supplied by the auther on request).

The need for the cost function **(14)** were aroused when structure activity relationship related to the opiate receptors[9,10] were investigated. The effectors of the various opiate receptors are on the one hand the rigid morphine like molecules (e.g. FIGURE 1) and on the other hand the endorphines. The pentapeptides Met and Leu enkephaline are members of the endorphine family. The enkephalines are very flexible and can adopt a very large number of conformational minima[11,12]. The pharmacophoric pattern is supposed to be known[13]. In view of the flexibility of the enkephalines, it is not surprising at all that attempts to match morphine derivatives with enkephalines in their local energy minima have failed[12]. Even the usage of the cost func-

tion **(13)** delivered unconclusive results due to the fact that the postulated pharmacophoric pattern refers only to a small portion of these molecules. Preliminary tests with cost functions including the surface area were quite promising, although the amount of computation is rather extensive[14].

6. Predicting Biochemical Activities: The Pressure Pattern

The binding energy (ignoring entropic effects or neglecting differences in entropic effects) is a physical observable, derivable from intermolecular interactions. The fact that intermolecular interactions are governed by laws which are valid for all chemical species allowed us to develop a general method for the construction of a receptor model out of given binding data. When considering biochemical activities or biological responses the situation is completely different. There are no general rules valid for all biochemical species which relate activities to structure. Every receptor must be treated as a special case and the investigator has to elucidate the special rules governing the activity expressed by the receptor. When an investigator relates biological responses to the binding energies, this expresses in many cases his helplessness rather than the desired structure activity relation.

Let us suppose that a receptor model reliably predicting binding enrgies was established. Assume further that biological activity data are supplied for some training set. How should the continuous receptor model be combined with this activity data in order to make activity prediction possible?

In the following we will present a very speculative concept termed **pressure pattern**. When a given effector is in eqilibrium with the receptor model we can calculate at any point on the 2D manifold the force exerted by the effector on this point. The fact that the overall resultant force between receptor and effector is zero does not exlude the possibility of large local forces. The local force or pressure at a point $\mathbf{r}_0$ (equilibrium constellation) on the manifold is the 3D-vector:

$$\mathbf{P}(\mathbf{r}_0) = \mathbf{Grad}\{\sigma(\mathbf{r}_0)*\rho(\mathbf{r}_0)\} \qquad (15)$$

The pressure pattern of a given effector is the 3D-vector function P(**M**) on the manifold representing the receptor. Each effector has its own characteristic pressure pattern and our speculative assertion is that the **pressure pattern of an effector correlates with the biological response elicited by its binding to the receptor.**

The total interaction energy between non-covalently bound effectors (excluding small ions) and a receptor is the sum over small and local contributions. The pressure pattern expresses the redistribution of the interaction forces at equilibrium e.g. when two parts of an effector are pulled over a central obstruction).

Observations which support our assertion are:

a) Binding enrgies are usually insensitive to small variations in the effector's structure, whereas biological activities are often extremely sensitive to such variations. Numerical experiments with hypothetical model molecules have shown that the pressure pattern is very sensitive to small structural variations in contradistinction to the computed binding energies.

b) Effectors with very high binding energies are often poor agonists but strong antagonists. This fact coincides with Jenck's concept of enzym-substrate strain destabilization and transition state stabilization[15]. The interaction energy should not be "wasted" but rather "reinvested" at the right local region (high pressure regions).

c) Koshland's concept of induced fit is widely accepted in biochemistry. The pressure pattern can be regarded as the driving force for an induced fit in its early stages.

7. Flexible Receptors

Although the amount of computation increases when flexible effector molecules are present in the training set, the flexibility of the effectors is managable in the frame of CCS without introducing a new concept. But what about the flexibility of the receptor's binding site? X-ray studies of the structure of enzymes have shown that nearly all active sites are in clefts,

pockets or boundaries between domains[15] so that there is an opportunity for a tight fit of substrates. Ideally the enzyme or the receptor should be rigid in order to minimize entropy losses.But real receptors are flexible and in many cases the flexibility is greater than necessary for the effector's entrance. The local flexibility of the receptor can be taken into account (at least partially) by the choice of appropriate steric interaction expressions, i.e. in those regions where the steric fit is not critical, the steric interaction function (Morse curve) should be very shallow (in the neighbourhood of the minimum). This simple, but very crude method has its justification only if the local motions of the receptor's constituents are small and not correlated, or if the amount of available binding data does not permit the construction of a more detailed model.
Considering the simulated receptor manifold as an elastic continuum would be a completely new approach. The 2D-manifold representing the receptor is characterized not only by the vector function $\rho(\mathbf{M})$ but also by quantities which express its elastic properties. It is clear that the local strain-stress relations of such a flexible membrane are much more complicated than the strain-stress relations in an isotropic rubber membrane. In the frame of linear elasticity theory the local elastic properties of any material are described by a fourth order tensor dependent on at most 21 elastic coefficients[16]. In many examples of anisotropic matter, a smaller number of elastic coefficients is sufficient (9 in the case of orthotropic continua and 5 in the case of transverse isotropy). A lot of preliminary research would be necessary until the concepts of elasticity theory could be applied to problems concerning drug design or receptor mapping. It is obvious that one would start with very simple molecules and their molecular envelopes. Our preliminary investigation can be outlined as follows:Deform the molecular skeleton by well-defined external forces (using quantum chemistry or molecular mechanics). The deformation of the molecular skeleton is accompanied by a deformation of the molecular envelope from which the strain tensor at any surface point can be derived.
From a series of strain-external forces relations derive the

elastic properties of this "envelope material".
This kind of computer experimentation with the fictive "envelope material" is not trivial at all. Some additional assumption concerning the deformation process are necessary in order to identify the point on the deformed envelope which belongs to a given point on the undeformed envelope.
Investigating medium to large molecular systems by classical newtonian methods (MM, molecular dynamics) was extremely useful in the past years[17], and there is no reason why this classical point of view should'nt be extrapolated beyond point mechanics.

8. Summary and Discussion

Most problems of molecular chemistry can be assigned to one of the following two classes: I) Given is the molecular structure; predict the observables. II) Given are the values of some observables of a chemical system; derive the molecular structure (spectroscopy, analytical chemistry). The problems of drug design when dealing with an unknown receptor are of different nanure. Given is a set of molecular systems (the effector receptor complexes). Each element of this set has a structurally unknown part, the receptor, which is common to all elements. With each element of the set only one (or two when activity is also considered) observable is assosiated: the binding energy. The drug designer is asked to predict the value of exactly this observable (or these two observables) for a system composed of a structurally well-defined part (effector candidate), and the same structurally unknown part by using the data of the given training set. The CCS method tries to solve the last-mentioned problem by the construction of a continuous chemical analogon which substitutes the unknown receptor. It is obvious that no other observable,but the binding energy (and biological activity activity, if activity data are supplied) can be predicted by this chemical analogon.
The receptor analogon ($\mathbf{M}$,ρ) consists of a 2D-manifold $\mathbf{M}$ which characterizes the geometry, and a vector function ρ on $\mathbf{M}$ which characterizes the chemical properties relevant for the interaction energy (i.e. charge distribution, H-bonding capacities,

dispersion forces etc.). A clear distinction is made between binding and activity data. Since the laws governing the binding process are valid for all receptors, only data related to the binding process are used for the construction of $(\mathbf{M},\rho)$. Generally there is **no correlation** between binding data and biological activity. The most common exception to this statement is when antagonists are considered. In that case a clear correlation between binding and activity (inhibition) can be observed. Hence, most of the so-called drug design methods should be correctly termed "antagonist design methods".

Theories and concepts in drug design are of minor value when they are not combined with algorithms. We summarize now the CCS method in a computer adequate formulation:

Step 0: Preliminary step. Choose a training set and a verification set with reliable binding data. Gather additional information about the nature of the intermolecular forces between the effectors and the receptor (i.e. is H-bonding relevant? Is there a tendency to form charge transfer complexes?). If additional information is not available or insufficient, a reasonable hypothesis should be made. Define the interaction energy in terms of $\sigma * \rho$. Fix the parameters R, **Eps**. Choose suitable basis functions for ρ.

Step 1: Construction of the superenvelope. Take additional information, like the parmacophoric pattern etc.,into account and solve the matching problem with the aid of a suitable cost function. Represent $\mathbf{S}_R$ by a set of uniformly distributed surface points.

Step 2: Compute the integrands of the **F** and **G** matrices (see **(6)** and **(9)**) at the relevant points of $\mathbf{S}_R$ (the relevant part is that which serves as a source of receptor candidates. It may consist of all points representing $\mathbf{S}_R$).

Step 3: Choose a receptor candidate $\mathbf{M} \subseteq \mathbf{S}_R$. The receptor candidate (exept the first) can be derived by systematic variations of predecessors.

Step 4: Compute the **F** and **G** matrices by integrating (summing) the stored results of step 2 over **M**.

Step 5: Solve the geometric programming problem (equations **(11)**

referring to the basis functions) of the receptor candidate.

Step 6: Equilibrium relaxation. Minimize the interaction energy between each effector and $(\mathbf{M},\rho)$ with respect to the external and internal degrees of freedom. Result: A set of $\Delta E_{theor.}$ for this receptor candidate.

Step 7: Correlate the ΔE's of step 6 with the experimental binding energies. Is the correlation between theory and experiment the best one (global minimum) with respect to all possible receptor candidates? If not, return to step 3, otherwise go to the next step.

Step 8: The best simulated receptor analogon is at hand. Predict the binding energies of the effectors in the verification set by allowing these molecules to interact with the simulated receptor. Are the predictions satisfactory? If not, return to step 0 and revise the model (data are insufficient, assumptions about the interactions are wrong, parameters are wrong, superenvelope is wrong or the method of CCS is not applicable); otherwise go to next step.

Step 9: If "antagonist design" was desired (i.e. if activity is directly related to binding energies) then the goal has been achieved. If biological data are supplied, compute the pressure patterns of the training set's molecules and try to relate these patterns (especially regions on $\mathbf{M}$ where most active effectors exert the greatest pressure) to the biological activities. If reliable activity predictions are possible (test with the verification set), then the goal in in this second case has been achieved.

The CCS method has been formulated in an algorithmic fashion, but has not yet reached a stage which allows its application to concrete problems of drug design, and a great deal of programming efforts must still be invested. Since the CCS method is a topographic method based on the generally accepted principles of intermolecular interactions,and since this method allows the integration of almost any piece of additional information, we belive that the above mentioned programming efforts will pay off.

References

1) H.R. Karfunkel, Match, 15, 227 (1984)

2) J. Hermans, H.J.C. Brendson, W.F. Van Gunsteren & J.P.M. Postma, Biopolymers, 23, 1513 (1984)

3) G.R. Marshall, C.D. Barry, H.E. Bosshard, R.A. Dammkoehler & D.A. Dunn in: Computer-Assisted Drug Design, ACS Symposium Series 112, E.C. Olson & R.E. Christoffersen (eds.), American Chemical Society: Washington (1979)

4) A.T. Balaban, A. Chiriac, I. Motoc & Z. Simon, Steric Fit in QSAR. Lecture Notes in Chemistry, No. 15, Springer-Verlag: Berlin (1980)

5) P. Gund, Progress in Molecular and Subcellular Biology, F.E. Hahn (managing ed.), Vol. 5, p. 117, Springer-Verlag: Berlin (1977)

6) H. Meschkowski, Ungelöste und unlösbare Probleme der Geometrie, B.I.-Wissenschaftsverlag: Mannheim-Zürich (1975)

7) H.R. Karfunkel, submitted to J. Comp. Chem. (1985)

8) H.R. Karfunkel, Habilitationsschrift, Universität Tübingen (1983)

9) S.R. Childers in: Neurotransmitter Receptors, part 1, S.J. Enna & H.I. Yamamura (eds.), Chapman and Hall: London-New York (1980)

10) F.E. Bloom, Ann. Rev. Pharmacol. Toxicol., 23, 151 (1983)

11) Y. Isogai, G. Némethy & H.A. Scheraga, Proc. Natl. Acad. Sci. USA, 74, 414 (1977)

12) G.H. Loew & S.K. Burt, Proc. Natl. Acad. Sci. USA, 75, 7 (1978)

13) A.S. Horn & J.R. Rodgers, Nature, 80, 795 (1976)

14) H.R. Karfunkel, to be published

15) W.P. Jencks in: Molecular Biology Biochemistry and Biophysics, Vol. 32, Chemical Recognition in Biology, F. Chapeville & A.L. Haenni (eds.), Springer-Verlag: Berlin (1980)

16) A.E. Green & W. Zerna, Theoretical Elasticity, At the Clarendon Press (1968) (second edition)

17) N. Allinger & U. Burkert, Molecular Mechanics, ACS Monograph 177, American Chemical Society: Washington (1982)