Pure & Appl. Chem., Vol. 69, No. 1, pp. 55–69, 1997.
Printed in Great Britain.

Model scenarios for the understanding of molecular recognition

Jürgen Brickmann

Institute of Physical Chemistry and Darmstadt Center for Scientific Computing, Technical University Darmstadt, Petersenstr. 20, D-64287 Darmstadt, Germany

Abstract: There are many factors (energetic, entropic, and kinetic) which have to be taken into account within a theoretical concept for molecular recognition. In this paper a model scenario is presented in which these different components are transferred to a representation for which human pattern recognition abilities can be used. It is demonstrated in particular that the van der Waals surface of molecules can be used as a screen for the representation of information pattern (electrostatics, local hydrophobicity, surface topography, molecular flexibility, etc.) which are relevant in molecular recognition processes. The technology of modern graphics workstations allows us to "see" a molecular scenario from the point of view of a molecule and to interact with this virtual world in a natural way. It is shown that this interaction with simulated molecular reality is not restricted to a local computational environment. New network communication techniques can be used to provide chemistry related information. A new approach is based on the virtual reality modelling language (VRML) which extends the world wide web (WWW) interface to visualize three dimensional (3D) scenarios and interact with the basic elements. It is demonstrated that the human recognition abilities can be transferred, at least partly, to a formal algorithmic concept by using fuzzy logic.

I. Introduction

The specific recognition of a molecule by a molecular scenario plays an important role in many chemical processes, it forms the basis for highly specific reactions in biochemistry and catalysis. There is a large variety of different factors (energetic, entropic, and kinetic, etc.) which come into play in a conceptional model approach to describe the recognition in a proper way (1). From a thermodynamic point of view the specificity of a receptor can be measured by a sub-group A of molecules it recognizes (at a given level of affinity defined by a certain ΔG value) among a larger ensemble B of molecules which in principal have to be considered. This type of recognition is often related to the key and lock image which was first introduced by Emil Fischer (2) in 1894 but this association is not always suitable. The equilibrium constant K_i taken in the direction of an association of a molecule to a receptor complex is (for simple processes) equal to the ratio of two kinetic constants, the association constant $k_{i,on}$ and the dissociation constant $k_{i,off}$

$$K_i = k_{i,on} / k_{i,off} \qquad (1.1)$$

The key and lock analogon works only in those cases when the association constant determines the selective recognition. In many biological reactions, however, the dissociation constant may become more important for the specificity than the association constant (1). In these cases the key and lock image breaks down.

In this paper we will consider the recognition problem from a bit different point of view. According to the formalism of information theory the specificity of a certain class out of a set B of molecules with respect to a given receptor can be measured by the expression (1)

$$S_A = \log(dim(B)/dim(A)) \tag{1.2}$$

where the dimensions dim (A) and dim (B) are simply the number of different molecules in the sets A and B, respectively. High specificity results in high values of S_A. It is obvious that this value drastically depends on the way set B is defined. If, for example, an antibody may selectively bind two of 20 different steroids, one has $S = {}_2\log(20/2) = 3.32$, if, however, 512 potential binding partners are considered $S = {}_2\log 256 = 8$ results. As long as set B can be well defined as in the case of the steroids, the numerical value of S is useful for the characterization of the receptor selectivity. In many cases this is, however, not possible. Two simple examples may demonstrate this fact. If one asks "how selective is the sweetness receptor with respect to the known sweetener one may easily determine the dimension of A but there is no simple selection criterion available for the B because the known sweetener (succrose, sucralose, saccarin, acesulfam etc.) belong to quite different classes of molecules. The situation becomes a bit simpler when two classes X and Y of molecules are compared with respect to the same reference set B. In this case one has

$$\Delta S = S_Y - S_X = \log(dim(X)/dim(Y)) \tag{1.3}$$

i.e. the dimension of B does no longer occur. This situation is realized when the size selectivity of two molecular sieves (two zeolites, for example) with respect to organic molecules is compared. There is possibly a method available to count all organic molecules with a minimum diameter which does not exceed given values (for set X and set Y) and so determine the dimensions of X and Y. In general the classification problem is not an easy task.

This paper does not concern the question whether the kinetics of the association or the dissociation are the selectivity determining factors but it deals with the classification problem, i.e. with the question of how to define the set A and the reference set B (see above). This problem is strongly related to the question of molecular similarity, i.e. instead of considering a set of molecules which belong to a certain molecular class (the steroids for example) we are looking at the molecules which belong to a certain class "from a molecule's point of view".

This work focusses on two aspects of the problem:

(i) How can the molecular principles of molecular similarity recognition be transferred into a scenario wherein human pattern recognition abilities can be applied?

(ii) How can strategies of human recognition be used for the development of algorithms which can be applied in molecular recognition processes?

The paper is organized as follows. In section II new instruments of man-machine communication in molecular science are described. The concept of molecular surfaces is introduced. These surfaces are considered as the interface between different molecules or between a molecule and its solvent. The section also deals with some visualization techniques and the mapping of patterns on molecular surfaces. With the new virtual reality modelling language (VRML) the 3D information on the molecular scenario can be distributed easily on the World Wide Web (WWW). Section III deals with the generation of information which can be mapped on molecular surfaces in order build up a convenient scenario for the recognition process. In section IV it is demonstrated, that the methods from fuzzy logic can be very useful in the classification problem mentioned above (see point (ii)) while in the final section V some conclusions are drawn. Some applications are given and are documented in the figures.

II. Man-Machine Communication Technology in Molecular Science

The human abilities for pattern recognition can only be successfully applied in the field of molecular recognition (see point (i) above) when there are procedures and tools available which allow a transformation of the molecular scenario into one which can be manipulated under visual control. This can be realized on the basis of the concept of molecular surfaces with the aid of modern graphical workstations and new computer network technologies. These conceps are briefly described in the follcwing.

2.1 Molecular Surfaces

All intermolecular interactions can be adequately described, at least in principle, by multi-dimensional scalar- and vector-fields representing the energetics of a molecular system as functions of intermolecular distances and orientations as well as intramolecular structur data. The visualization of these fields, however, has to be done on the basis of a 3D-picture or 2D-projections, because the pattern recognition ability of human beings is strongly related to the 2D- and 3D-world. Consequently, the multidimensional field has to be reduced to a 2D- or 3D-representation. In molecular science this can be done in many different ways.

We will not describe all the different possibilities of molecular visualizations in this contribution but restrict the discussion on molecular surfaces and the mapping of molecular properties on these surfaces.

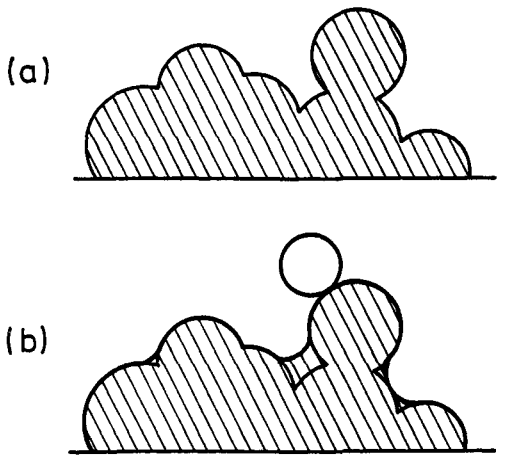

Fig. 1 Hard sphere model of a molecular surface (a) and contact surface (b) the contact surface is generated by rolling a test particle (sphere) over the hard sphere model.

A molecule 'sees' the surface of another molecule as a smooth object. Such a surface can be generated by rolling around another hard sphere model particle on the hard sphere model surface (see fig.1).

This model which was first introduced by Connolly (3,4) forms some reference standard for molecular surface generations in many molecular modelling packages like the MOLCAD program (5) which was developed in the Darmstadt group of the author. The contact surface representation gives the chemist some insight into the molecular shape as it would be seen from a particle of given size. Modern workstation technology allows the real time manipulation (translation, rotation, scaling, stereo projection etc.), i.e. the 3D world can be directly experienced. Surfaces generated with the same test particle (e.g. a water molecule with an effective sphere radius of r=1.4 Å) can be qualitatively and quantitatively compared. Moreover, the contact surface, generated with a water probe, is well suited in order to discuss shape fitting (for example of two proteins).

Formal molecular surfaces have become important tools for the interpretation of molecular properties, interactions and processes (6-9). A detailed review is found in (3).

2.2. Quality Mapping with Texture Mapping Technology

The molecular surface concept is not only useful for a representation of the bulkyness and the shape of molecules. These surfaces can used as screens for the visualisation of arbitrary properties using color coding techniques. Color coding is a popular means of displaying scalar information on a surface. As was demonstrated recently (10), this mapping can be very effectively done by using texture mapping techniques which are available in modern workstations. Texture mapping is a technique that applies an image to an object's surface as if the image were a decal or cellophane shrink-wrap.

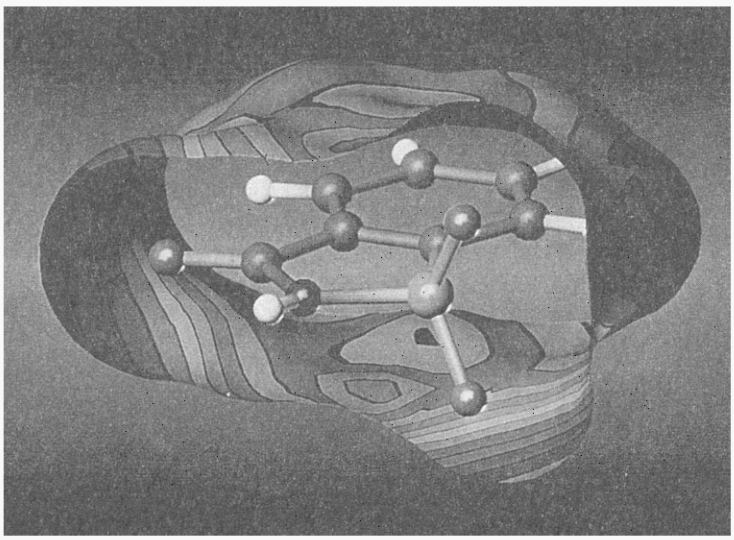

Fig. 2 Electrostatic potential mapped with texture mapping technology (10) onto the contact surface of a molecule.

Texture mapping technology can also be used in order to filter out interactively information from the graphical representation (10). Filtering property information on a molecular surface is able to generate more insight in two different ways: (i) The filter allows the scientist to distinguish between important and irrelevant information, and (ii) the filter puts an otherwise qualitative property into a quantitative context, e.g. the standard deviation from a mean value may provide a hint as to how accurate a represented property actually is. (see fig. 3).

Every three dimensional scalar or vector field which may be generated on the basis of the position of atomic or molecular fragment (see section III) can be visualized by color coding on a given surface.

2.3 Information Transfer on Molecular Scenarios on the World Wide Web (12)

Since the early days of the internet, scientists have used computer networks to exchange their knowledge and their experiences. In the first stage, this was achieved by electronic mail. But this medium permits the transport of information only between a few participants. Later, with the introduction of mailing lists and newsgroups questions and answers could be shared among the scientific community more globally.

With the development of the World Wide Web (WWW) in 1989 the situation has changed dramatically. Within the WWW, it is possible to exchange information in various forms. Data retrival can be made easily. With hyperlinks everything can be referenced from anywhere on the internet. The WWW has turned the entire internet into one large storage for information of any kind. The new information exchange technology has already lead to a rapid growth of electronic publishing media. Not only text and static images can be submitted to an electronic journal, audio sequences and animations can be included as well. The formats used for images and animations are of pixel based nature and the representations are fixed in size and rigid in their behavior. The viewer

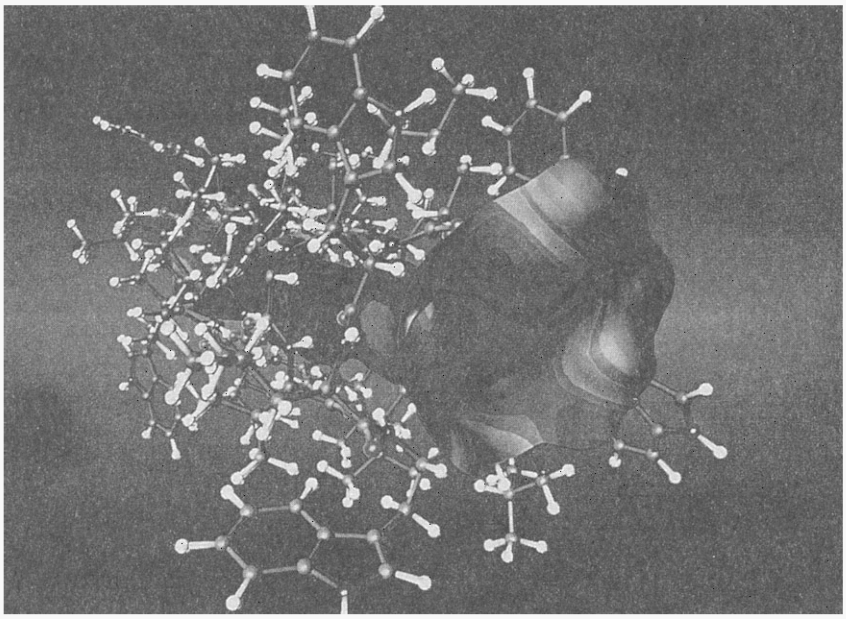

Fig. 3 Ion Channel of the Gramicidin A dimer. The hydophobic part of the molecular surface is clipped out using two dimensional texture mapping technology (11). The electrostatic potential is mapped onto the hydrophilic channel area of the molecule.

is not able to change the point of view with respect to the object shown. Moreover, pixel based formats do not allow the transmission of information in a very compact manner. In order to get a three-dimensional (3D) impression of an object, at least two images are needed and the point of view is still preselected by the producer of the scenario. In molecular science it is absolutely neccessary to have three dimensions in order to transfer complex structural information. Traditionally, the structural information is transfered via standard file formats (e.g., Brookhaven Protein Data Bank, PDB) containing atom types and 3D coordinates and the visualization and manipulation is done locally with the aid of modeling software packages (like SYBYL/MOLCAD (13) or others). There was no way up to now to transfer the 3D scenario directly. This gap can be filled using the Virtual Reality Modeling Language (VRML)(12).

VRML is based on a subset of the Open Inventor File Format. This subset was extended with networking capabilities, such as WWW hyperlinks. With this feature, VRML is an equivalent to the hypertext markup language (HTML)(14). Like HTML files describe the layout of 2D-text pages to be displayed by WWW browers, VRML files describe the layout of 3D scenarios.
Some of the capabilities of VRML are demonstrated at the WWW page (15) of the institution of the author with examples from protein research (see fig. 4), visualization of molecular structures from 3D data bases, and visualization of orbitals.

III. Properties Mapped on Molecular Surfaces

The activity of a drug is very often related to the molecule receptor complementarity. This complementarity is in many cases defined in a quite vague manner. This may be demonstrated by a well-known example. It could be shown by Becket (16) that the (-) form (or l-form) of adrenaline fits much better to a hypothetical receptor than the (+) isomer (or d-form). The author based his interpretation of the drug receptor interaction on the existence of three different interaction types: a flat (hydrophobic) binding site, a hydrogen bond acceptor site, and a cation-anion interaction (see fig. 5).

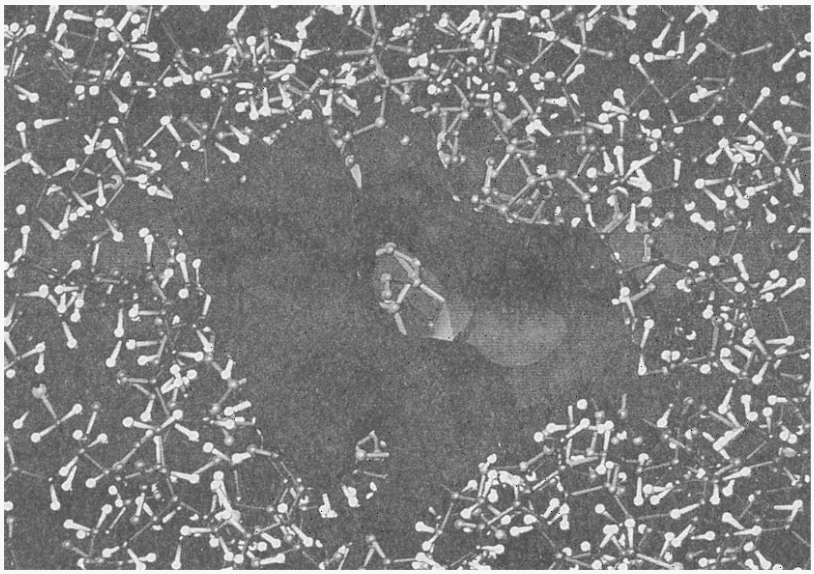

Fig. 4 Scene from the WWW home page of the authors institution (15). It shows part of the Cytochrome P450 enzyme. These proteins play a central role in carcinogenesis and toxicology. The carcinogenic nitrosamines were activated by hydroxylation at the porphyrine binding site of the enzyme. To enter this active site located at the center distant about 10 Å from the surface, substrates have to pass a channel. The surface of this substrate channel is shown. The scene can interactively be inspected and partly manipulated via the World Wide Web.

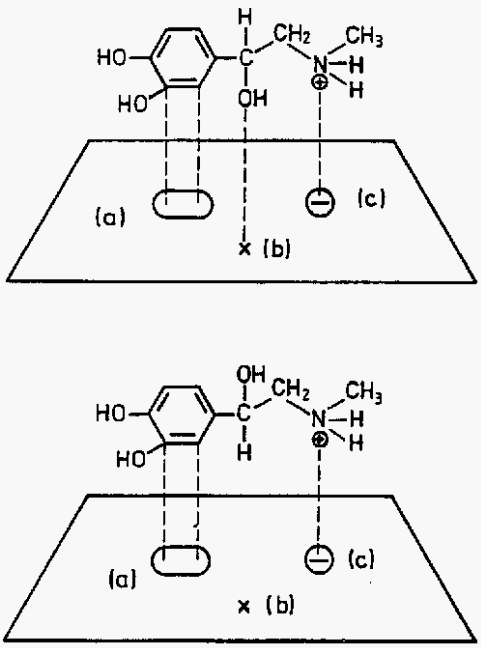

Fig. 5 Orientation of (-)- and (+)-adrenaline at the receptor site. (a) flat bending site, (b) proton acceptor site, (c) anionic group

The different interaction models in this picture are not primarily related to specific structure elements of the adrenaline molecule but to the reaction field between the drug and the receptor. Molecules with completely different structure may generate very similar reaction fields. This is demonstrated with different sweeteners in fig. 6 where the local hydrophobicity (see below) is mapped on the molecular surface of different sweeteners.

What are the properties which can be mapped onto the molecular surface in order to generate patterns which can be used for the similarity analysis? In the following some traditional possibilites as well as new concepts are described.

3.1 Electrostatic maps

Molecular recognition is dominantly controlled by free energy changes

$$\Delta G = \Delta H - T\Delta S, \qquad \Delta A = \Delta U - T\Delta S \tag{3.1}$$

where U,H,A,G, and S are the conventional notations for inner energy, enthalpy, free energy, Gibbs free energy, and entropy respectively.

The energetics can adequately be described on the basis of electrostatic interaction. Mostly the electrostatic potential is mapped (17). Hydrogen bond are basically controlled by electrostatic interactions as well, but sometimes it is reasonable to map proton donor- and accepter functionality of the atoms behind the surface, independently (13).

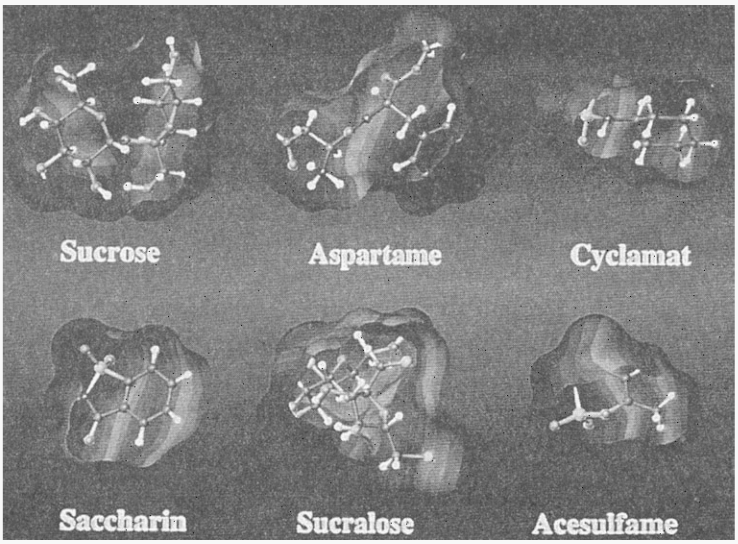

Fig. 6 Local hydrophobicity on the surface of different sweeteners. The left sides of the molecules are hydrophilic, and the right sides are hydrophobic

3.2. Local hydrophobicity (18)

Local hydrophobicity plays an important role in molecular recognition processes. It is generally accepted that hydrophobic interaction between two molecules is related to both, energetic and entropic contribution, but up to now, there is still no *simple* physical model available for hydrophobicity and hydrophobic interaction. However, there are several attempts to define relative hydrophobicity values on the basis of empirical findings.

Quite recently an empirical method for the localisation, the quantification, and the analysis of relative hydropobicity of a molecule or a molecular fragment has been reported from the group of the author (18). Here only a short review is given. The approach is based on two concepts:
(i) that the overall hydrophobicity of a molecule (measured for example by the logarithm of the partition coefficient in an octanol/water system $\log(P)=-RT\Delta G_{transfer}$ with the tranfer free energy $\Delta G_{transfer}$ for one mole substance from one solvent to the other) can be obtained as a superposition of fragment contributions, and
(ii) that this free energy can be represented as a surface integral over the solvent accessible surface of the molecule on the basis of a local free energy surface density (FESD) ρ. This surface density function is represented in terms of a three dimensional scalar field which is composed as a sum of atomic increment functions describing lipophilicity in the molecular environment (18). The empirical model parameters are obtained by a least square procedure using experimental log P- values as reference data. It is found that the procedure does not only work for the prediction of unknown partition coefficient but also for the localisation and quantification of the contribution of arbitrary fragments to this quantity. In addition, the formalism can be used also for an estimate of the hydrophobicity index, a hypothetical log P- value which depends on the actual molecular conformation.
The FESD approach as has been described above can be well used in order to predict unknown partition coefficients of molecules with given structure. This has been demonstrated recently (18). However, the FESD approach is not restricted to the calculation of hydrophobicity index values. The FESD data can be directly used in order to map local hydrophibicity onto the molecular surface and so give the chemist a direct insight on hydrophobic and hydrophylic parts of the molecule. This has been recently demonstrated by Lichtenthaler and cooworkers (19,20) who studied the structure activity (sweetness) relationship of a variety of carbohydrates and other molecules. The authors followed the classical approach in which the sweet taste of organic compounds presumes the existence of a common AH-B-X glycophore (a proton donor, a proton acceptor B and a hydophobic group X arranged in a triangle) in all sweet substances, elicating the sweet response via the interaction with a complimentary tripartite AH-B-X site in the taste bud receptor but they resumed that this "sweetness triangle" concept only holds when the hydrophobic X-part is considered as an entire, obviously quite flexible region rather than a specific corner of the "sweetnes triangle": in sucrose and sucralose encompasssing the outside area of the fructofuranose moiety, in fructose the 1- and $6CH_2$ groups in either linked or separated form. This new concept (19,20) has been developed on the basis of visual inspections of the color coded molecular electostatic potential maps and the FESD maps and it has been tested with a variety of sweet compounds (see Fig.6). Most remarkably, the authors (20) found that the FESD profiles generated for the solid state conformation of a variety of non-carbohydrate high-potency sweeteners, such as the sulfonamides cyclamate, saccarin, and acesulfame, as well as structurally distinctly different dipeptides, e.g. aspartame, exhibite a hydrophobicity distribution strikingly similar to those observed for the sugars (see fig 6).

3.3. Topographical analysis of molecular surfaces (21)

Two molecules of complex structure may only form a stable complex when those parts which are important for the binding can come into close contact. From the point of view of the molecular surfaces this means that both surfaces have to be complimentary to some extent in the binding area. This surface complementarity can be identified in simple cases just by inspection of the computergraphical images, but this technique is not very usefull for systematic searches. For the latter a formal classification is necessary. Several methods for the characterization of surfaces in topological terms have been proposed (21-28). Mezey and coworkers (24) established a method for a topological analysis of contour surfaces and van-der-Waals surfaces represented by fused spheres. One approach of these authors is based on the calculation of a curvature parameter, which is used for the classification of certain domains of the contour surface in terms of different curvature properties.
In a recent work from the group of the author a numerical procedure for the calculation of the *local and global canonical curvatures* (21) at each point on a surface was presented which leads to *domains* of different quality. The domains can be characterized by *curvature profiles*, which provide information about their topology. A comparison of the profiles of different molecules is helpful for

the elucidation of docking procedures. The topological features of a surface can be quantified by the two *canonical* curvatures at each surface point. The canonical curvatures are defined as the eigenvalues of the *local Hessian matrix* (i. e. the matrix of second derivatives).
The global curvatures may be interpreted as average curvatures of the corresponding surface region. They disregard the detailed shape caused by atomic roughness. In other words, the surface can be smoothed, and the grade of smoothing depends on the choice of a selection distance fixing the area to be covered by the paraboloid. This enhancement of the concept of local canonical curvatures allows the classification and characterization of large surface regions, thus opening the possibility to subdivide a surface into *domains* specified by their area and curvatures.
The interactive comparision of two given molecular surfaces on the basis of curvature can be done interactively by using two dimensional texture maps, colour coding the two canonical curvatures calculated for different selection distances along the x- and y-coordinate of the texture map. However, the information from the curvature profile can be further reduced by introducing a surface topography index (STI) as has been recently demonstrated by Heiden (29). The surface topography index s, may be defined on the basis of two global curvatures c_1 and c_2 as follows

$$\begin{aligned} s &= (c_1-c_2)/c_1 && \text{if } c_1 > 0 \text{ and } c_2 > 0 \\ s &= 1+(1-(c_1+c_2)/c_1) && \text{if } c_1 > 0 \text{ and } c_2 \le 0 \text{ and } |c_1| > |c_2| \\ s &= 2+(c_1+c_2)/c_2 && \text{if } c_1 > 0 \text{ and } c_2 \le 0 \text{ and } |c_1| \le |c_2| \\ s &= 3+(1-(c_2-c_1)/c_2) && \text{if } c_1 \le 0 \text{ and } c_2 < 0 \\ s &= 0 && \text{if } c_1 = c_2 = 0 \end{aligned} \qquad (3.2)$$

The STI values vary within the interval $0 \le s \le 4$. Calculated from the relation of both global curvatures (each of which can be either concave (+), flat (0) or convex (-), where $c_1 \ge c_2$, the STI gives an expression of regional shape for every surface point, continuously varying between five basic shape descriptors: bag (+/+), cleft (+/0), saddle (+/-), ridge (0/-), nob (-/-), and as a special case, plateau (0/0). However, an information about the absolute curvature is lost during the process of STI calculation. In a graphical representation this information can be added again to the colour-coded STI display on the molecular surface using two dimensional texture mapping technology, encoding the first dimension with the STI value as a colour and the second with the maximum of c_1 and c_2 as the colour saturation at each surface point (29).
Based on the calculation of regional canonical curvatures, the surface topography index, STI, gives a quite accurate description of local shape, relating surface regions to a set of five basic shape classes. As this definition of the STI - though continuous - already implies a discrete classification, this method is well-suited for completely automatic shape analysis algorithms. This may be accomplished either by keen contour cuts or - for a better characterization of local shape structures - by more sophisticated algorithms (using, for example, fuzzy logic strategies (30)(see next section).
A major advantage of this shape descriptor definition is the freedom of choice of the grade of globality - unfortunately connected with the major disadvantage of rather large computational effort, which rises fast with increasing globality.

3.4 Surface Flexibility (30)

Despite the great usefulness of the concept of surface topography descriptors, almost all approaches suffer from a severe limitation: the flexibility of molecular surfaces is not taken into account. There is no doubt, however, that a rigid surface model does only give a rough impression of the scenario faced, e.g., by a ligand molecule approaching the surface of a protein. Particularly for the selectivity and specificity of enzymatic reactions, the flexibility of the compounds is extremely important. Speaking in terms of the lock-and-key principle (2), neither the lock nor the key are rigid, but may accomodate in such a manner that optimal interaction is ensured.
In a recent work of Zachmann et al (30), two methodically new approaches (termed method I and method II) for the quantification and visualization of surface flexibility have been presented. The basic data for both approaches are supplied by molecular dynamics (MD) simulations and the methods have been applied to the two proteins (PTI and ubiquitin). The calculation and visualization of the local flexibility of molecular surfaces are based on the *solvent accessible surface* (SAS) introduced by Connolly (3,4).

Method I is bases on a statistical analysis of the surface fluctuations during MD-runs taking periodic "snapshots" of the protein. Although applied to proteins the technique is not restricted to molecular systems. Any flexible surface may be analyzed if its position in space is well defined as a function of time. Method II (which is conceptually quite simple and needs only a few seconds computer time) relies on the atomic RMS fluctuations which can easily be calculated from the results of the MD simulations.
This type of representation can be very helpful for the interactive study of protein docking. Even if two surfaces do not fit a local disagreement can be weighted from the knowledge of "local softness" (30).

IV. Fuzzi Logic strategies for molecular recognition

In the last two sections it has been demonstrated, how the molecular interactions which form the basis for all molecular recognition processes, can be transfered to a scenario which can be inspected interactively with the human senses of pattern recognition. It has been shown that, in many cases, microscopic information has to be averaged and thermodynamic concepts have to be extrapolated to a molecular scale, in order to generate pictures which can be handled properly. All these efforts are reasonable as long as the scenario is definitely treated with the senses of human beings. A typical example is the sweetness triangle concept (19,20). The strategy fails when there is a large variety of molecules under consideration. In this cas the interactive treatment is no longer reasonable and one has to deal with the question how the principles, responsible for a certain recognition (from a human point of view) can be transfered to an algorithm which opens up the possibility of transfering the vaguely defined patterns to a computerized strategy. It is known that, the dominant factor for the inhibition of the enzyme trypsin is the shape selectivity of a receptor site which forms a deep bag. A potential inhibitor has either to fit into that bag (like the benzamidine molecule) or it has to have a nop (like the natural trypsin inhibitor PTI) which has this ability. In order to define the class of molecules B which can be considered as possible inhibitors, one has to screen molecular shapes in a systematic manner. This can certainly be done using deterministic algorithms (21-28,31,32) but firstly these algorithms are quite computer time consuming and, secondly, the simple comparison of rigid surfaces may not be adequate to the problem.
In the following we will focus the discussion of the recognition of molecular surfaces, but the principles can be applied to a variety of different molecular properties (see section III) as has been recently demonstrated by Heiden et al. (33)
To sumarize the situation there are obviously two problems in the field of automatic molecular shape recognition:

(i) How can the relevant characteristical properties of a molecular surface be classified such that shape complimentarity can be formulated in a way, which is similar to that controlling the human recognition (like: a big nop fits to a big bag)?
(ii) How can the vagueness inherently incorporated in the definition of the surface of flexible molecules be included in molecular matching strategies?

In order to deal with these questions one has to deal with vaguely defined objects on the one side and vaguely defined strategies to compare these objects, on the other. It has been demonstrated recently (33,34) that at least parts of the answers can be given using the technology of fuzzi logic. In the following two subsections it is shown that this scheme can, indeed, be adequately used within an algorithmic treatment of molecular recognition.

4.1 Fuzzi logic and linguistic variables

The concept of fuzzy logic was introduced almost 30 years ago by Zadeh [35]. Lying dormant for many years, it has been rediscovered in the mid 80's for regulation in micro electronics, automatic process regulation or in operation research. By now, fuzzy set theory has many applications in a large variety of different fields. We refer to the literature [36-37] for detailed representations. Here we only present those concepts which are directly relevant for the molecular reconition problem.
Fuzzy set theory may be seen as a generalization of classical set theory, each element of a fuzzy set A being defined by a function value x in definition space X together with its degree of membership to A. The latter is defined by a membership function $\mu_A(x)$, whose values lie normally within a range $0 \leq \mu_A(x) \leq 1$ between zero and complete membership.

$$A = \{(x,\mu_A(x)) \mid x \in X\} \tag{4.1}$$

In classical (crisp) sets $\mu_A(x)$ can only be 0 or 1, while fuzzy logic allows almost any type of function for membership definitions.
One of the most important tools in applications of fuzzy set theory is the concept of *linguistic variables* (LV) [33,34]. These are groups of fuzzy sets with (partially) overlapping membership functions over a common (crisp) basic variable x. In order to represent several classes within a LV the membership functions should cover all the relevant definition space of the basic variable x with membership function values $0 < \mu_A(x) < 1$. (Values of 0 or 1 are assigned to the rest of the definition space in all membership functions). The overlap of these functions defines the fuzziness. Generally, a linguistic variable L, classified by n fuzzy sets A_i, can be defined as

$$L = \{(x,\mu_{A_1}(x)),.....,(x,\mu_{A_n}(x)) \mid x \in X\} \tag{4.2}$$

Usually, the information a decision should be based upon, is given by crisp function values (for molecular surface segmentation, this means certain scalar qualities assigned to every node point on a triangulated surface). Also the decision itself shall again lead to a crisp value (in this case the binary decision between continuation or limitation of a surface domain). However, in order to apply fuzzy logic tools to a problem, it has to be defined by linguistic variables. Thus decision making requires three steps:

(1) fuzzification of crisp basic variables into linguistic variables;
(2) fuzzy decision from different LV using fuzzy operators;
(3) defuzzification back to a crisp value.

The details of these steps are discussed with the specific application patterns as far as necessary. For further details see ref. [37].

4.2 Shape analysis of molecular surfaces using linguistic variables

The shape analysis presented here is based on the surface topography index (STI) presented above. Following Heiden et al. (33) a six class linguistic varable

$$\begin{aligned} L = \{ & (bag, \mu_B(s)); (cleft, \mu_C(s)); (saddle, \mu_S(s)); \\ & (ridge, \mu_R(s)); (nob, \mu_N(s)); (plateau, \mu_P(\max(c_1,c_2))) \} \end{aligned} \tag{4.3}$$

is introduced. The membership functions are schematically shown in Fig. 7.

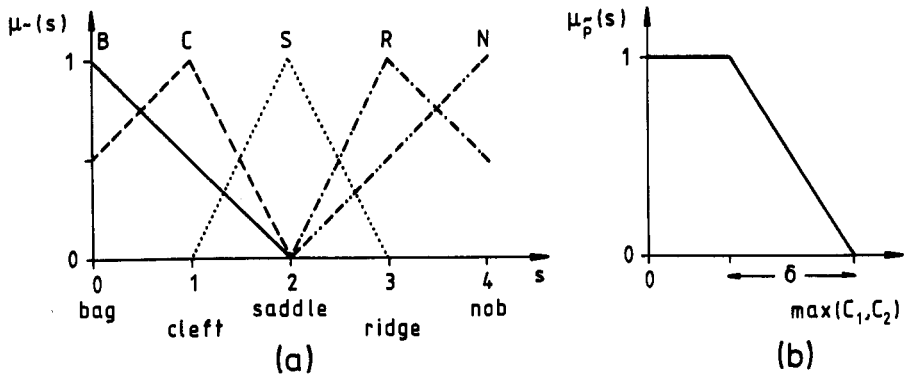

Fig.7 Membership functions for linguistic variables describing molecular shape

An automatic segmentation of molecular surfaces into distinct domains can be performed using dissimilarity measures D introduced by Heiden et al. (33) for linguistic variables. In the practical

calculations we started from a representation wherein the molecular surfaces are given as a triangle mesh in 3D-space with location-dependent qualities assigned to each surface point (which is a node between adjacent triangles) are divided into separate homologous domains. Neighbouring domains

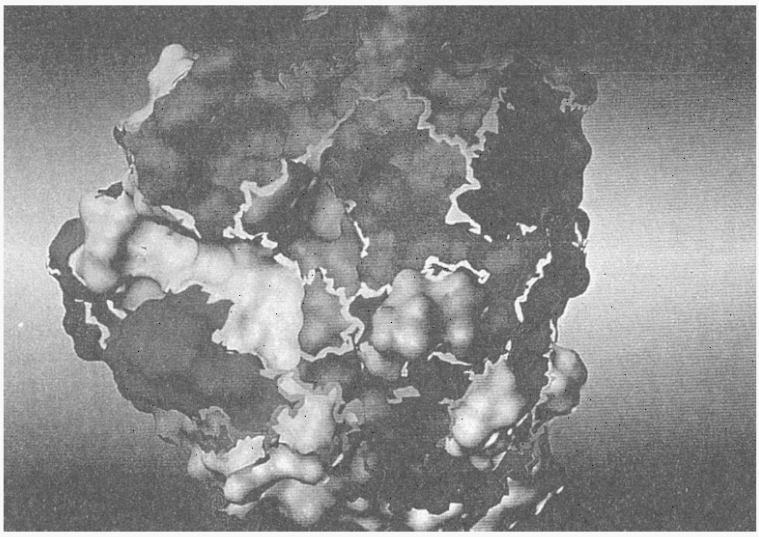

Fig. 8 Segmentation of the surface of the trypsin molecule into domains of different qualities (shown in different intensities of grey).

differ with regard to a certain surface quality, whose value is characteristic for each domain (within a fuzzy limit). The algorithm is based principally on the growth of a surface domain, starting at a characteristic reference point (for example, the point with the highest STI absolute not yet assigned to another domain). Linguistic variables are assigned in advance to each surface point and are updated continuously for an average of the actual domain. Following the neighbourhood information given by the triangle mesh, the domain ends when the dissimilarity of a surface point to the domain average, or its direct neighbour within the domain, exceeds a given limit. The borders of other domains already defined also put an end to segment growth. Working its way through all triangle node points sequentially, the program achieves complete segmentation of a triangulated surface.
The result of the segmentation is a set of surface patches of given surface area which can be uniqely related to the linguistic variables bag (B), cleft (C), saddle (S), ridge (R), nob (N), and plateau (P) (see eq. 4.3). It has been demonstrated (33) that the linguistic variables can be adequately used in order to discuss topographical and energetical differences of proteins (for example Trypsin/trypsinogen) as well as complementarity (for example in the trypsin/PTI complex). Moreover, the linguistic descriptions can be well applied for first guesses in automatic docking procedures (34). Herein the coordinates of the centers of masses of the surface patches are used in order to characterize the position of the domain in space. The first guess fitting of a set of domains from a molecule A to one of a molecule B with complementary properties is performed by matching the points pairwise using an analytical minimum distance least square algorithm commonly applied in standard molecular modelling procedures for atom/atom fittings (38). The complementarity is therein completely expressed in linguistic terms. A simple example may demonstrate this: A *big nob* (at position x_1) *imbedded* in a *big plateau* (at x_2) is matched to a *big bag* (at y_1) *imbedded* in a *big plateau* (at y_2). One cannot expect that this type of matching leads to prediction of molecular complexes with atomic precision but the effort for a systematic screening of possible arrangements of the two molecules (3 translational and 3 rotational degrees of freedom) can be drastically reduced. In order to optimize the positions of both molecules in space one has to match the molecular surfaces and then proceed according to energy minimization procedures. The latter are not discussed here. In the next subsection it is demonstrated, however, that fuzzy logic may also successfully be applied in surface matching procedures.

4.3 Matching of molecular surfaces with fuzzy logic strategies

Molecular surfaces as can be defined on the basis of the Connolly algorithm (3,4) (see section 2.1) can be well used in the interactive treatment of molecular scenarios. These surfaces are a rough representation of repulsive interaction of molecules i. e. two molecules can be moved towards each other with reasonable energetic effort as long as these surfaces do not substantially interfere. The situation of closest approach where parts of the molecular surfaces are in close contact can be realized in an interactive treatment with reasonable effort. However, there is no simple way to transfer this strategy to an automatic procedure. This is related to the following problems.

(a) There is no a priori way of deciding which part of one surface should be compared with a particular part of another.

(b) The surfaces of molecules cannot be uniquely defined since a given molecule's surface is dependent on the properties of the probe molecule (what ever the surface generation procedure may be).

(c) There is no unique way to quantify the matching of two surfaces even if they are defined with arbitrary accuracy.

(d) A matching procedure should take into account the "softness" of molecular surfaces. If one knows that a certain part of a surface is quite flexible, one should not worry about local disagreement of the surfaces to be matched and take this fact into consideration when designing the matching strategy.

There are a number of techniques which have been proposed to solve the surface matching problem (31,32,39,40) but they all suffer from the fact that an inherent uncertainty is replaced by ad hoc procedures. Even if the molecular surface concept is replaced by a one parameter family of isosurfaces (3) this does not lead to a unique matching technique. In a recent work of the author (34) a matching procedure has been suggested which is based on fuzzy logic in order to take the incertainties formulated above into account, at least in principle. This concept is based on a soft definition of a surface by defining membership functions (see section 4.1) $\mu_s(\mathbf{r})$ and $\mu_v(\mathbf{r})$, measuring to what extent a given space point belongs to the surface and the bulk of a molecule, respectively. Following this concept the matching of two molecules A and B can be calculated from the Carbo indices (3,41)

$$O_{AB} = \int \mu_{SA}(\boldsymbol{r})\mu_{SB}(\boldsymbol{r})dv \; / \left(\int \mu_{SA}^2(\boldsymbol{r})dv \cdot \int \mu_{SB}^2(r)dv \right) \tag{4.4}$$

and

$$V_{AB} = \int \mu_{va}(\boldsymbol{r})\mu_{VB}(\boldsymbol{r})dv \; / \left(\int \mu_{VA}^2(r)dv \int \mu_{VB}^2(r)dv \right) \tag{4.5}$$

An optimal match of the two molecules is reached when O_{AB} is maximal while V_{AB} takes a minimum value. The new technology has been tested in a first application by matching the surfaces of two flexible proteins Tripsin and PTI (42). In this application the membership functions M_S and M_V have been calculated from molecular dynamic simulations similar to that reported earlier (30). It turned out that the structure of the trypin-PTI complex is very close to that which was found in x-ray studies. Studies for the refinement of the method in particular in connection with the domain decomposition are in progress.

V. Conclusions

It was demonstrated that the capacity of modern graphical work-stations (like texture mapping technology) enables the chemist to 'see' molecular scenarios from a molecule's point of view. Model experiments as the docking of a substrate to a receptor or the comparison of molecules of different structure but similar chemical activity (like sweeteners) can be performed on a time scale of human interaction. It was shown that in particular the concept of molecular surfaces as screens for the representation of different properties is very helpful for the discussion of molcular recognition i.e. specific intermolecular interactions. This mapping is extremely effective when texture mapping capacities of modern work-stations are used. It was demonstrated that the introduction of quantities describing local hydrophobicity, surface roughness, surface curvature and surface flexibility lead to a better understanding of the "molecular language". The communication

with the molecular scenarios is not restricted to a local workstation enviroment. This can be done also nonlocally using new network technologies like the World Wide Webb and an object oriented programming language (virtual reality modelling language, VRML). Finally, it was shown the the pattern recgnition abilities of human beings (which are of dominant importance in the field of interactive modelling) can be transfered to a formal algorithmic concept by using fuzzi logic strategies. This latter field is, however, still in an early phase of development.

Acknowledgment
The author likes to thank Wolfang Heiden, Bonn, Horst Vollhard and Carl-Dieter Zachmann, both Darmstadt, for fruitful discussions and technical assistence, as well as Ines Osterloh for carefully reading the manuscript. This work was supported by the Fonds der Chemischen Industrie, Frankfurt

REFERENCES

1. M. Delaage, in *Molecular Recognition Mechanisms*, Delaage M., pp 1-13, Editor, VCH Publishers, New York, (1991)
2. E. Fischer, *Chem. Ber.* **27**,2985 (1894)
3. P.G. Mezey, *Molecular Surfaces*; Rev.Comp.Chem, Lipkowitz,Boyd (Eds.) Verlag Chemie, Weinheim **1990**, 265-294
4. M. Connolly, *Science*, **211**, 709-713 (1983)
5. M. Waldherr-Teschner, Th. Goetze, W. Heiden, M. Knoblauch, H. Vollhardt and J. Brickmann "MOLCAD - Computer Aided Visualization and Manipulation of Models in Molecular Science", in *Adv. in Scientific Visualisation*, F.H. Post, A.J.S. Hin, Eds., Springer, Berlin, pp 58-67 (1992)
6. R.Langridge, T.E.Ferrin, J.D.Kunz, M.L.Connolly, Science, **211**, 661-666 (1981).
7. L.Pang, E.Lucken, J.Weber, G.Bernardelli, J.Comp.Aided.Mol.Des. **5**, 285-291 (1991)
8. W.Heiden, M.Schlenkrich, J.Brickmann, J.Comp.Aided.Mol.Des., **4** 255-269 (1990)
9. W. Heiden, T. Goetze, J. Brickmann, *J. Comp. Chem.* **14**,246(1993)
10. Ref. 201
11. M.Waldherr-Teschner, Chr. Henn, H. Vollhard, S. Reiling, and J.Brickmann, *J.Mol.Graphics* **12**, 98 (1994)
12. H. Vollhardt, Chr. Henn, G. Moeckel, M. Teschner, and J. Brickmann *J.Mol.Graphics*, 1995, in press
13. J. Brickmann, T. Goetze, W. Heiden, G. Moeckel, S. Reiling, H. Vollhardt, and C.-D. Zachmann, Interactive Visualization of Molecular Scenarios with MOLCAD/SYBYL. In: *Data Visualization in Molecular Science*. (J. E. Bowie, Ed.) Addison-Wesley Publishing Company, Reading (1995) pp. 84-97
14. T. Berners-Lee, D. Connolly, Hypertext Markup Language - 2.0. (1995) (http://www.w3.org/hypertext/WWW/MarkUp/html-spec/html-spec_toc.html)
15. http://www.pc.chemie.th-darmstadt.de
16. A.H.Becket, *Fortsch.Arzneimittelforschung*, **1**, 455 (1959)
17. Náray-Szabó, G., J. Mol. Graphics 7, 2 (1989), 76-81.
18. Pixner, P., Heiden, W., Merx, H., Moeckel, G., Möller, A., Brickmann, J., *J. Mol. Inform. Comput. Sci.*, **34**,1309 (1994)
19. F.W.Lichtenthaler, S.Immel, U.Kreis in *Carbohydrates as Organic Raw Materials*, F.W.Lichtenthaler Ed., VCH Publishers, Weinheim/New York 1991, 1-32; Staerke/Starch 43,121 (1991); F.W.Lichtenthaler, S.Immel, D.Martin, V.Mueller, in *Carbohydrates as Organic Raw Materials*, Vol. 2, G.Descotes Ed., VCH Publishers, Weinheim/New York, 1993, in press
20. F.W.Lichtenthaler, S.Immel *Sucrose, Sucralose and Fructose: Correlations Between Hydrophobicity Potential Profiles and AH-B-X Assignments*, in "Sweet Taste Chemoreception", G.G.Birch, M.A.Kanters, M.Mathlouti, Eds., Elsevier Publ., Amsterdam, 1993, in press
21. C.-D. Zachmann, W. Heiden, M. Schlenkrich,and J. Brickmann, *J.Comput.Chem.* **13**, 76 (1992)
22. S. E. Leicester, J. L. Finney and R. B. Bywater, *J. Mol. Graphics*, **6**, 104 (1988).

23. P. Bladon, *J. Mol. Graphics*, **7**, 130 (1989).
24. P. G. Mezey, *J. Comput. Chem.*, **8**, 462 (1987); A. Arteca and P. G. Mezey, *J. Comput. Chem.*, **9**, 554 (1988).
25. R. L. DesJarlais, R. P. Sheridan, G. L. Seibel, J. S. Dixon, I. D. Kuntz and R. Venkataraghavan, *J. Med. Chem.*, **31**, 722 (1988).
26. P. M. Dean, P. Callow and P.-L. Chau, *J. Mol. Graphics*, **6**, 28 (1988).
27. H.Nakamura, K. Komatsu, S. Nakagawa and H. Umeyama, *J. Mol. Graphics*, **3**, 2 (1985).
28. N.Colloc'h and J.-P. Mornon, *J. Mol. Graphics*, **8**, 133 (1990).
29. W.Heiden, Dissertation, TH Darmstadt, 1993
30. C.-D. Zachmann, S. Kast, and J. Brickmann, *J.Mol.Graphics*, **13**,89 (1995)
31. P.L.Chau, P.M.Dean *J.Mol.Graphics*, **5**, 152 (1987)
32. F.Blaney, C.Edge, R.Phippen, C.Burt in *Data Visualization in Molecular Science*. (J. E. Bowie, Ed.) Addison-Wesley Publishing Company, Reading (1995) pp. 99-129
33. W.Heiden, J.Brickmann *J.Mol.Graphics*, **12**, 106 (1994)
34. J.Brickmann, The Use of Linguistic Variables in the Molecular Recognition Problem, in *Fuzzi Logic in Chemistry*, D.H.Rouvray Ed., Academic Press, San Diego, to be published 1996
35. L.A.Zadeh, *Information and Control*, **8** , 338 (1965)
36. H. Schildt, *Artificial Intelligence Using* C. Osborne McGraw-Hill, Berkeley, 1987
37. H./J.Zimmermann, *Fuzzy Set Theory and Its Applications* Kluwer, Boston 1991
38. D.R.Ferro, J.Hermans *Acta Christ.*, **A33**, 345 (1977)
39. P.G.Mezey, *J.Comp.Chem.* 8,462 (1987)
40. G.A.Arteca, T.M. Gund, M.A. Hermmeier, V.B. Jaumal, P.G.Mezey, and J.S.Yadav, *J.Mol. Graph.* 6,45,1988
41. R.Carbo, L.Ledda, A.Arnau, *Intern.J. Quantum Chem.* 17,M85
42. C.-D.Zachmann, and J.Brickmann, to be published