

## Molecular modelling in design of crop protection chemicals

Barbara Odell

*Shell Research Limited, Sittingbourne Research Centre, Sittingbourne, Kent ME9 8AG, U.K.*

**Key words:** Molecular design; Biorational approach; Acetolactate synthase; Crop protection; Molecular modelling; QSAR

---

### SUMMARY

Specific examples from pesticide research are given which illustrate the types of analysis employed to design optimal inhibitors for a given receptor, based on the assumption that a congeneric series of compounds behave in a related mode in the biosystem. The examples illustrate the complementary role played by computational chemistry, X-ray crystallography and computer graphics and also raise questions as to the current limitations of existing molecular mechanics and quantum mechanics techniques.

---

### INTRODUCTION

#### *Scenario for agrochemical design*

In the design of crop protection chemicals the objective is to control various organisms selectively. There are many difficult criteria to fulfill including those of high potency, high selectivity, desirable physical properties (e.g. vapour pressure, solubility, log P, etc.), soil persistence, photostability and low cost. To date, agrochemicals have been discovered largely by empirical synthesis and screening.

In 1950, 2 000 compounds were screened to produce one commercial compound; today it is more like 20 000. Discovery takes 5–8 years and costs tens of millions of dollars. These economic and statistical pressures have increased the emphasis on the *biorational approach* based on a molecular understanding of key biological processes involved in the control of weeds, fungi and insects. Computer-aided molecular modelling (CAMP) techniques, if properly supported by a deeper understanding of the biology and biochemistry at the molecular level, can aid this chemical design process. The approach has generally lagged behind its application in the pharmaceutical industry because of a lack of information on the biochemistry of plant, fungal and insect receptors and enzymes. Investment is needed for molecular biological techniques and rapid examination of the biochemical processes to determine DNA base sequences, amino acid sequences and finally isolating pure enzymes for X-ray determination of molecular structure. Another important fundamental requirement of pesticide design is the necessity to find compounds with broad range

activity. The compounds must inhibit receptor sites in a range of organisms, rather than just one, but still be safe towards non-target species. This multidimensional objective is sometimes very difficult to achieve.

#### *The biorational approach*

Several modes of action have been identified among insect, weed and fungal biosystems, although few are being investigated with the aim of determining X-ray structures of receptor sites. Some companies, are studying specific enzymes in detail (e.g. Dupont [1]; enzyme, acetolactate synthase). Various university groups are studying the structure of the postsynaptic nicotinic cholinergic receptor site in insects [2] and amino-acid sequences of various cytochrome P<sub>450</sub>s involved in 14 $\alpha$ -demethylation of lanosterol in fungi [3]. The 3-D coordinates of the photosynthetic reaction centre from *Rhodopseudomonas viridis* have been published by Miki and Michel at the Max Planck Institute [4]. However, this short list shows the paucity of targets for which detailed information can be expected in the near future.

Although there are many facets to pesticide design involving considerations of absorption, distribution, etc., molecules finally interact specifically with a macromolecular receptor (protein, nucleic acid, membrane lipid, etc.). It is vital, therefore, to obtain good *in vitro* data on intrinsic activity. It is preferable if these data can also be correlated with *in vivo* data. Any divergence from expected behaviour may relate to transport (e.g. penetration) and metabolic factors which could obscure intrinsic potency.

#### *The role of molecular modelling*

What can molecular modelling techniques offer the medicinal or pesticide chemists in terms of concrete results?

The techniques have:

- (1) Descriptive capabilities from which electronic and spatial parameters can be deduced for use in structure-activity relationships (SAR) studies;
- (2) Potential for investigating properties impossible to examine experimentally (e.g. gas phase conformers, transition states);
- (3) Predictive capability and can be used to organise priorities for future synthesis.

Two major analytical tools are used to underpin the rational design approach: (i) quantitative structure-activity relationships (QSAR) and (ii) computer-aided molecular modelling (CAMM) using computer graphics. QSAR uses statistical techniques which attempt to correlate a biological response with calculated or measured physicochemical data (e.g. log P, pK<sub>a</sub>, Hammett substituent constants, molar refractivity, etc.). This approach is useful for yielding predictions and trends but takes little account of stereochemical or spatial information. CAMM is used in tandem with QSAR, providing more mechanistic insights into conformational and electronic requirements for biological activity. Once a hypothesis of the enzyme-substrate or inhibitor interaction is proposed, it is obligatory to test it with more experimental data; as new compounds are synthesised and tested the model becomes more predictive with time.

It is imperative at the early stages of the experimental design process to identify the effects of substituent changes within a congeneric series of compounds. It is thus important to ask the right

sort of questions regarding: (1) *Conformation* (e.g. relative energy differences between conformers, Boltzmann distribution or equilibrium state of the system, importance of flexibility, etc.); (2) *Electronic properties* (requirements for a particular charge distribution molecular electrostatic potential (MEP), dipole moment, etc.), reactivity towards nucleophilic or electrophilic attack, types of non-covalent intermolecular interactions available for binding; (3) *Thermodynamic properties* (dominant enthalpic and/or entropic contributions for binding to a receptor).

#### *Computational choices*

Two main computational choices are available to modellers: (1) quantum mechanics and (2) molecular mechanics. Quantum mechanics involves describing the behaviour of electrons in terms of wave functions  $\Psi$ . The more accurately we can calculate  $\Psi$ , the more nearly we can predict structural and electronic properties prior to synthesis.

The computational chemist must assess how accurately he needs to calculate  $\Psi$ . Ab initio methods using double zeta basis sets with polarisation functions may produce more accurate results, but because computational speed is inversely proportional to  $N^4$  where N is the size of the basis set, it is often impractical to perform this type of calculation on anything but the smallest fragment (<20 atoms) for both economic and time reasons. Defining a fragment to represent part of the molecule that is responsible for the property of interest is risky because it assumes a measure of independence from neighbouring groups and atoms which may or may not be valid. More recently, semiempirical quantum mechanical techniques, e.g. AM1 [5], have taken the lead amongst the fastest and most accurate semiempirical methods yielding results which are in some cases more accurate for energy and geometry than those obtained with ab initio minimum basis sets (e.g. STO-3G). The programs offer a range of output in terms of options (e.g. dipole moments, reaction coordinates, force constants, entropy, etc.), but like all semiempirical techniques are sometimes unable to depict realistic electronic and structural properties of anions of, sulphur-, oxygen- and nitrogen-conjugated systems, of which there are a plethora in current crop-protection agents.

Molecular mechanics is purely empirical and can offer no information regarding electronic properties (e.g. reactivity, excited states, etc.). Its main advantage is its simplicity for describing energy of strain of a molecule in terms of bond stretching, angle bending, twisting around bonds, non-bonded van der Waals interactions, electrostatic and hydrogen bond interactions between groups. It remains the fastest method for determining relative energy differences between conformers. Although it is the most accurate technique for calculating hydrocarbon heats of formation, it relies totally on parameterisation. Most forcefields are deficient in parameters for S, N, O extended conjugated systems, which have not been pursued as much as amino-acid and nucleotide forcefields. Various forcefields are available such as Kollman's AMBER [6] and Allinger's MMP2 [7] but none of the parameters is portable from one forcefield to another.

Having decided on a property of interest, the next step is to decide which computational method to use and when to apply it. The results should be unambiguous and computed structures should always be in accordance with experimental data. Often several techniques are used to confirm various properties such as conformation, dipole moment or hydrolysis rates. Calculation outputs often comprise long lists of numbers which can be simplified in the form of colour displays depicting electronic distributions, MEPs (i.e. molecular electrostatic potential),

superdelocalisabilities or hydrophobicities. Contours can be used to clarify atomic or molecular orbital coefficients or energy probability maps depicting potential energy surfaces. Various conformer changes can be animated to portray dynamic simulations of molecular movement.

#### *Limitations in pesticide design*

Major limitations of current theoretical methods being applied in the area of pesticide design include:

- (1) Computational difficulties in describing S-, O-, N- and P-conjugated and non-conjugated systems, charged groups and hypervalent molecular systems using either quantum or molecular mechanics techniques.
- (2) Calculation of factors influenced by solvation/desolvation effects – computational methods are still based on gas phase situations.
- (3) Lack of data on protein crystallography and biochemistry of target and transport processes.

The process of designing new biologically active molecules in the absence of an X-ray structure of a receptor is a daunting prospect. The modeller's task is analogous to a locksmith [8] designing keys without even seeing the lock, knowing that some keys fit and others do not. Any information that can be gleaned from X-ray data of even a homologous enzyme may be useful in formulating hypotheses for experimental testing. New molecules are conceived by comparisons of structure and biological activity between selected compounds which are known to be active, or by considering the properties of the natural substrate itself. Many false conclusions can be drawn. One area in which we are attempting this approach involves cytochrome P<sub>450</sub> inhibitors. Ideas on the properties of the active site can be gained from the recently resolved X-ray structure of P<sub>450CAM</sub> [9, 10]. The dissociation constants of a number of inhibitors and the substrate, camphor itself, are shown [11, 12] in Fig. 1. If this were the only data available to guide the chemist on to

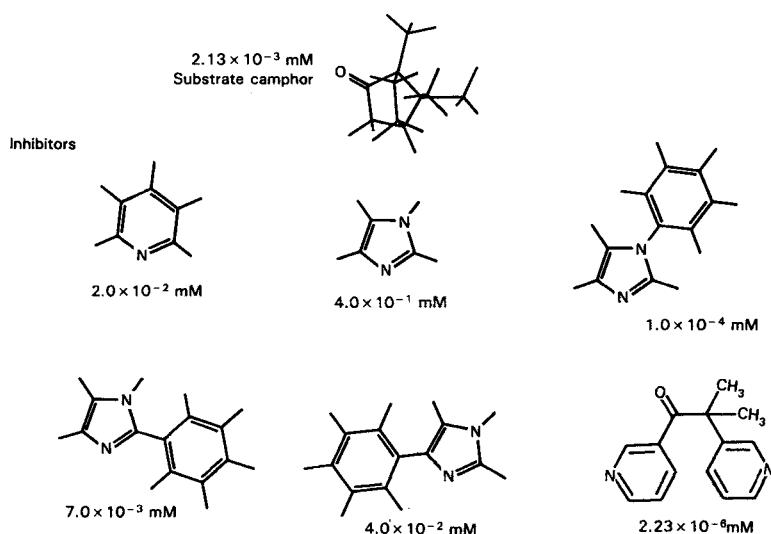


Fig. 1. P<sub>450CAM</sub> substrate and inhibitor dissociation constants. Taken from Griffin and Peterson [11] and Lipscomb [12].

the next stage of synthesis, what sort of structural and electronic requirements might be inferred by inspecting both substrate and inhibitors?

Pyridine and imidazole bind less strongly than the substrate itself. Of the five inhibitors, 1-phenylimidazole binds most effectively. By examining more closely the steric and electronic features of the substrate, one might postulate that the phenyl ring of 1-*N*-phenylimidazole should superimpose onto the lipophilic methyls groups of camphor. However, there is no proton acceptor group present in the inhibitor to match that of the carbonyl oxygen atom of camphor. The activity of P<sub>450</sub> inhibitors, however, is governed by both hydrophobic and steric properties as well as the strength of the bond formed between the nitrogen lone pair on a heterocycle and the prosthetic heme iron.

Would chemists then have been able to design metyrapone, the most active inhibitor of P<sub>450CAM</sub> with a k<sub>D</sub> of  $2.23 \times 10^{-6}$  mM? In this compound, optimisation of both steric, metal binding and hydrogen bonding factors conspire to give powerful inhibitory activity. The X-ray structure of metyrapone (Cambridge database) shows a gauche conformational relationship of two pyridine rings, either of which could bind to Fe-heme. Structure activity relationships [13] have shown, however, that inhibitory activity requires the pyridine A ring (containing the CMe<sub>2</sub> group), but only the steric bulk of the B ring (containing the ketone function). Substitution for ring B by phenyl does not reduce the activity. Conformational analysis using molecular mechanics suggests the *trans* conformer is less favoured over the gauche X-ray form by ca. 6.6 kcal/mol. Goodford's program GRID [14] employing a water and methyl probe for locating favourable binding sites for both substrate and inhibitor (Fig. 2) allows one to map out the pharmacophoric pattern of the active site. For example, both substrate and inhibitor contain the proton acceptor carbonyl oxygen functions which are thought to bind to the proton donor site (e.g. tyrosine OH group). The two molecules contain superimposable methyl groups.

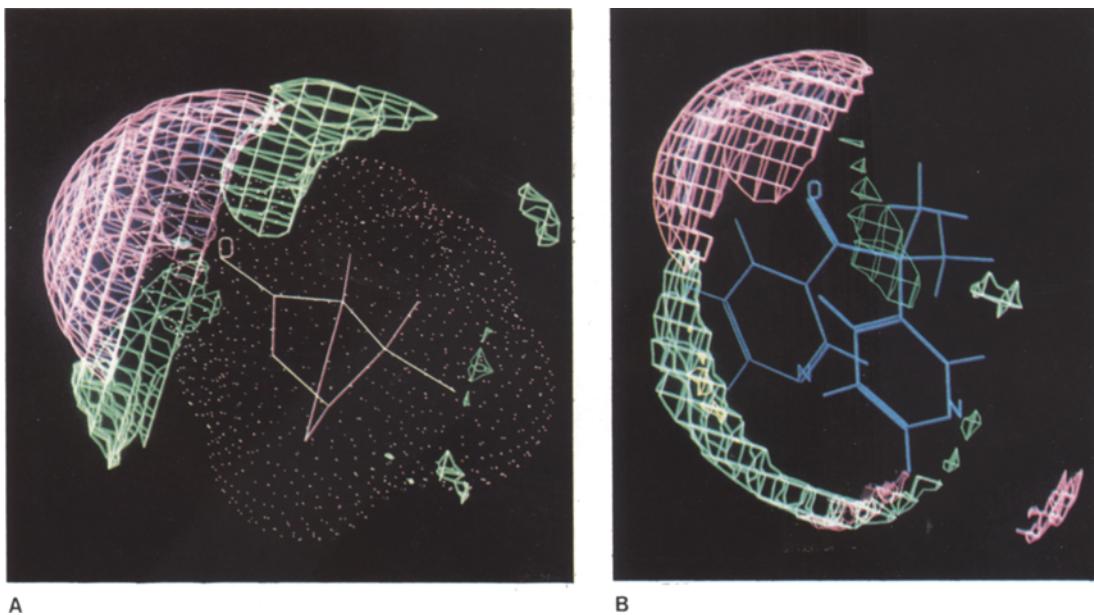


Fig. 2. Favoured contour regions of H<sub>2</sub>O (purple) and Me (green) probes with (A) camphor and (B) metyrapone.

The X-ray structure of P<sub>450CAM</sub>, allows one to clarify assumptions made in the absence of a fungal receptor site. One can examine the amino acids and protein secondary structure comprising the active site in an attempt to assign the types of non-bonded interactions holding the substrate in the active site cleft. One can rationalise factors influencing substrate/inhibitor specificity (comparing these with experimental results) which can ultimately lead to the construction of hypothetical P<sub>450</sub> sites of interest.

For example, focusing on the substrate binding site within a 8 Å radius from the centre of mass of the camphor (Fig. 3) it is apparent that the residues contacting the substrate are essentially lipophilic (Phe-87, Leu-244, Val-247, Val-295, Ileu-395, Thr-252) with only one residue (Tyr-96-OH) forming a hydrogen bond with the carbonyl oxygen of camphor. This hydrogen bond acts as an anchoring point about which the camphor can rotate, provided that unfavourable steric interactions with the remaining hydrophobic residues are not encountered. Overall aromatic and aliphatic side chains provide an excellent 'hand-in-glove' fit for the substrate.

A simple way of examining the types of non-covalent interactions between substrate and active site residues involves using molecular mechanics calculations such as AMBER [15] (employing united atom forcefield and the subroutine ANALYSE) which allows each relative residue contribution to be assessed separately. Obviously there are a number of limitations in the model. Firstly, the relative energies calculated by this method correspond to gas phase interactions (dielectric constant = 1) which are not, to say the least, the most relevant quantities from a biochemical point of view. However, in order to simulate the solvent environment around the active site, one would have to carry out extensive Monte Carlo or molecular dynamics simulations in which both solute and solvent are allowed to relax.

The results in Table 1 show that the active site residues defining the substrate pocket are involved in weak non-covalent dispersion interactions between camphor methyl and methylene groups (e.g. Thr-252, Val-295, Leu-244, Phe-87, Ile-395 and Val-244). There is also a significant hydrogen-bonding interaction between the Tyr-96-OH and the C=O group on camphor. In total these interactions contribute -7.5 kcal/mol to the overall enthalpy of binding in a gas phase environment. Tyr-96 makes the largest single contribution to the interaction energy (-3.54

TABLE 1  
MOLECULAR MECHANICS CALCULATED ENERGY CONTRIBUTIONS FOR THE BINDING OF CAMPHOR IN P<sub>450CAM</sub> ACTIVE SITE  
Total intermolecular non-bonded energy = - 7.5 kcal/mol

Group contribution	VdW	Electrostatics	Hyd-bond	Total energy contribution (kcal/mol)
(1)Phe-87	-1.32	0.29	-	-1.03
(2)Tyr-96	-1.00	-2.51	-0.03	-3.54
(3)Leu-244	-1.58	-0.01	-	-1.59
(4)Val-247	-0.58	0.44	-	-0.14
(5)Thr-252	-0.82	0.44	-	-0.38
(6)Val-295	-0.26	0.21	-	-0.05
(7)Ile-395	-0.68	-0.10	-	-0.78

kcal/mol) via its proton donating ability (the hydrogen-bonding capacity is reflected more by the large electrostatic component). Each of the remaining hydrophobic residues contributes a small, but favourable van der Waals or dispersion interaction with camphor, the sum total of which contributes ca. -4.0 kcal/mol to the total energy of binding. These energies are not absolute values and can only be used in a relative sense to compare group contributions.

Nevertheless, in support of these results, we used Goodford's program GRID [14], to obtain the results in Table 2, to locate potential ligand binding sites for water and methyl probes. The program assumes the total energy of interaction consists of additive individual intermolecular interactions in much the same way as molecular mechanics calculates non-bonded contributions (e.g. electrostatics, Lennard-Jones and hydrogen bonding components). Using the active site devoid of camphor favourable binding sites are located in the camphor substrate binding pocket which contribute again some -7.6 kcal/mol of binding energy.

These semiempirical calculations may have yielded fortuitous results in the light of experimental thermodynamic parameters as derived by Griffin and Peterson [11] from spectrophotometric measurement. Their results indicate that the association is largely entropically driven rather than enthalpically driven. Their results are consistent with the idea that desolvation of both substrate and active site provide the driving force for binding.

Indeed the X-ray structure of the camphor-free P<sub>450</sub> site shows that the substrate pocket is occupied by an array of water molecules which provide a polar environment for the Tyr-96 group. As a result, replacement of the Tyr-96 solvent interactions with the Tyr-96-camphor carbonyl oxygen hydrogen bond is likely to result in a reduced energy difference than predicted by our forcefield calculations. Entropic factors are also likely to govern substrate binding as a result of the release of active site water molecules. However it is not certain whether the same is true for inhibitors, since the binding strengths in forming Fe-heme-N bonds are also believed to be a dominant factor.

Space filling models allow one to rationalise the activity of the phenylimidazoles. For example, 1-phenylimidazole in Fig. 4 can be docked into an empty pocket in which perhaps water molecules are situated in the camphor-free site. The phenyl group can occupy a region coincident with the methyl and methylene groups of camphor. In Fig. 5, however, 2-phenylimidazole encounters steric difficulties, since the phenyl ring would have to penetrate the van der Waals surface of the porphyrin ring to achieve the same binding distance of Fe-N (ca. 2.0 Å) as in 1-phenylimidazole. Consequently, the imidazole ligand would not be able to approach the Fe-heme so closely as in the 1-phenylimidazole case and in fact 2-phenylimidazole may bind in a different mode at the active site.

Thus the X-ray structure and biochemistry of P<sub>450CAM</sub> allows one to deduce certain fundamental criteria for designing new P<sub>450</sub> inhibitor activity which may also be applicable to the fungal P<sub>450</sub> systems.

## APPROACHES TO PESTICIDE DESIGN

The following examples illustrate some of the approaches molecular modelling can offer pesticide design: (1) Methods of describing 'active site' conformations; (2) Investigation of electronic properties; (3) Analysis of factors regulating reactivity; and (4) Methods for depicting intermolecular interactions.

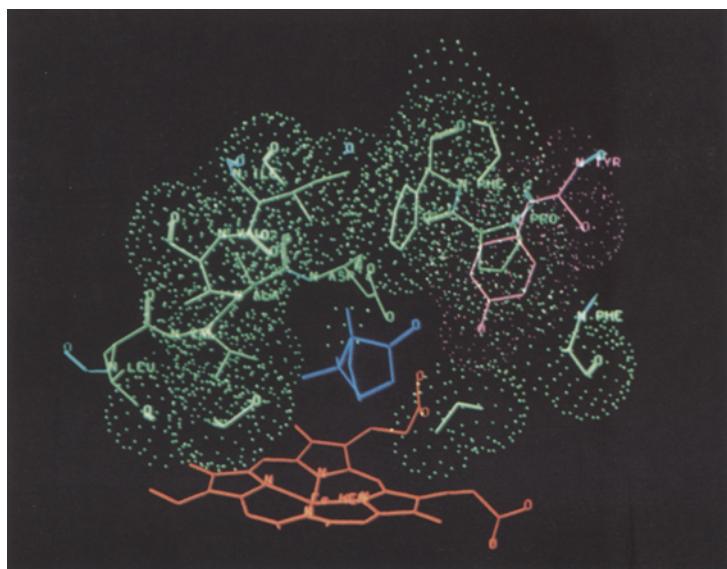


Fig. 3. Active site region of  $P_{450}\text{CAM}$  within 8 Å of centre of mass of camphor.

TABLE 2

GRID/GRIN ENERGIES FOR FAVOURABLE BINDING SITES IN ACTIVE SITE OF  $P_{450}\text{CAM}$  (ABSENCE OF CAMPHOR)

Probe	Energy <sup>a</sup>
$\text{H}_2\text{O}$	Hyd-bond + VdW = -4.88 kcal/mol
Methyl	VdW = -2.68 kcal/mol
Total ( $\text{H}_2\text{O} + \text{methyl}$ )	Hyd-bond + VdW = -7.56 kcal/mol*

<sup>a</sup>Energy values only relative not absolute.

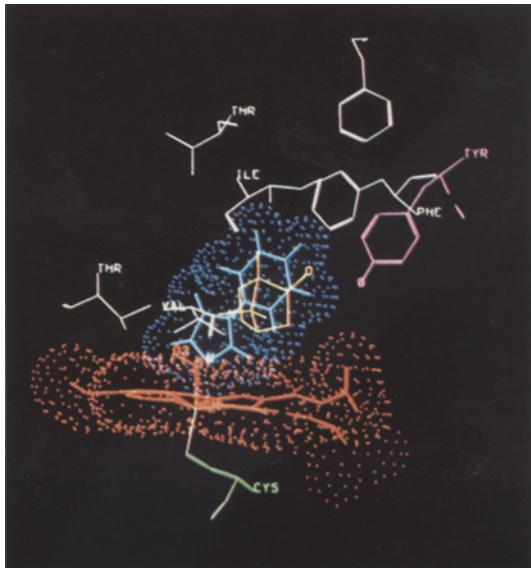


Fig. 4. Docking of 1-phenylimidazole into  $P_{450}\text{CAM}$  active site.

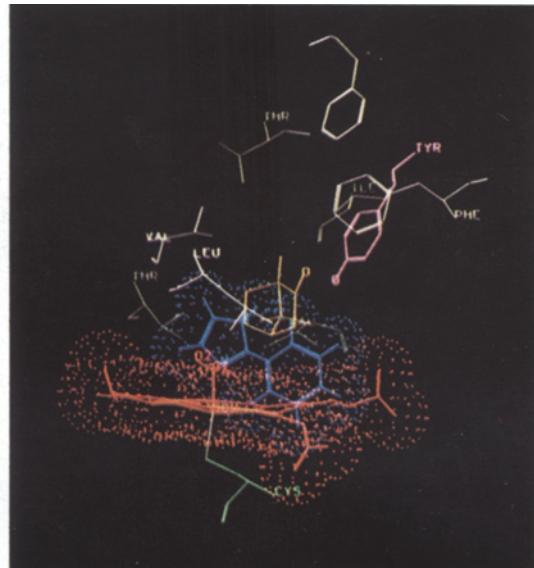


Fig. 5. Model showing docking of 2-phenylimidazole into active site of  $P_{450}\text{CAM}$ .

### *Analysis of 'active site' conformations*

Many azole EBI fungicides which inhibit the P<sub>450</sub> enzymes involved in 14 $\alpha$ -demethylation of lanosterol embody some of the common structural features of metyrapone itself as shown in Fig. 6. EBI fungicides active against plant fungal diseases represent a much wider range of structural types compared to the P<sub>450CAM</sub> inhibitors mentioned earlier. There are no representative parent molecules which incorporate a unique set of salient chemical features from the group. A nitrogen heterocycle is believed to axially ligate axially to the Fe-heme, leading to inhibition of ergosterol biosynthesis.

Additional features of these EBI fungicides include a collection of hydrophobic and hydrophilic groups. Some substituents (e.g. halogenated phenyl rings) are thought necessary to prevent metabolism, whilst others presumably occupy particular regions of the enzyme binding site and/or participate in transport mechanism allowing them to reach that site. The compounds presumably all adopt a shape complementary to the site. Using the 'active analogue' approach based on molecular mechanics, attempts can be made to locate conformers which are reasonably accessible (ca. 5–10 kcal from the global minimum). Results are compared for different molecules noting conformers which are stable for actives versus those for inactives. The conformations that could be adopted by the active site can then be deduced.

Thus efficient computational techniques are required to search conformational space for compounds containing between 2 and 8 rotatable bonds. Two different techniques, both utilising molecular mechanics (AMBER) are illustrated in this text using a well studied system, N-acetyl-alanyl-N-methylamide, the Ala dipeptide. The various conformers are listed in Table 3 along with selected computational techniques. The first search technique is based on the popular

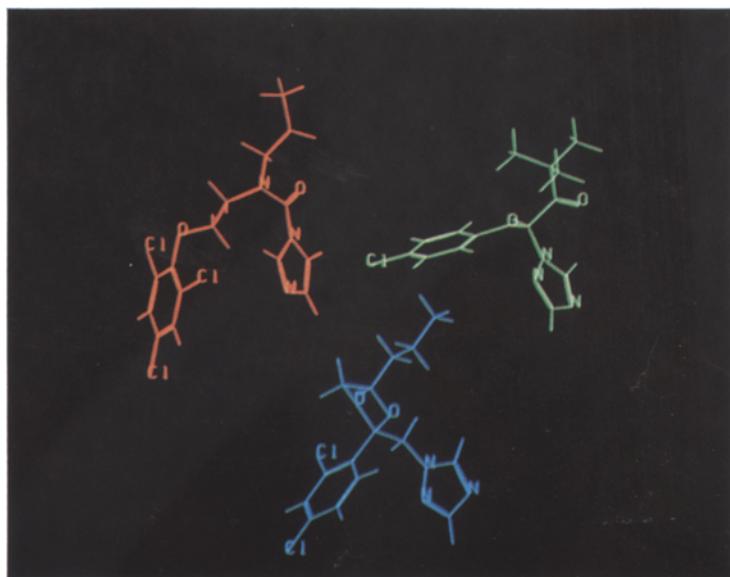


Fig. 6. Representative azole fungicides (EBIs).

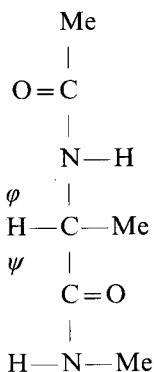
TABLE 3  
CALCULATED INTRAMOLECULAR ENERGIES OF CONFORMERS OF AcAlaNHMe IN kcal/mol

Method	$C_7$	$C_5$	$\alpha_R$	$P_{II}$
Amber <sup>a</sup>	0	2.7	4.5	5.7
MNDO/AMI <sup>b</sup>	0	1.98	3.16	3.07
CHARMM <sup>c</sup>	0	4.5	4.7	4.8
PCILO <sup>d</sup>	0	2.0	3.4	4.0
ECEPP <sup>e</sup>	0	6.0	8.0	8.0
CNDO <sup>f</sup>	0	5.0	5.0	5.0
EHT <sup>g</sup>	0	-2	0	0
4-31g <sup>h</sup>	0	-2.1	7.5	-

<sup>a</sup>P. Kollman and A. Dearing, <sup>b</sup>M. Dewar and J. Stewart, <sup>c</sup>M. Karplus et al., <sup>d</sup>B. Pullman et al., <sup>e</sup>H.A. Scheraga et al., <sup>f</sup>F.R. Momany et al., <sup>g</sup>R. Hoffman and I. Imamura, <sup>h</sup>B. Robson et al.

systematic grid search procedure [15], the second on a stochastic Monte Carlo Metropolis sampling algorithm [16]. Both are able to locate all five minima  $C_{7eq}$ ,  $C_{ax}$ ,  $C_5$ ,  $\alpha_R$  and  $P_{II}$ . As shown the  $C_7$  conformers (eq and ax) are predicted to be the most stable in the gas phase, by both computational techniques [15, 17].

The first method of systematic grid search involves a technique of adiabatic mapping where  $\varphi\psi$  dihedrals are constrained to fixed increments say  $10^\circ$  and the remaining degrees of freedom are then relaxed. In the Ala dipeptide we are essentially only interested in mobility around



since we assume that the *trans* NH to C=O amide conformer is most favoured.

Conformational analysis involves calculating the intramolecular energy as each dihedral is incremented from  $0^\circ$  to  $360^\circ$ . The resulting energies can be contoured as shown in Fig. 7, each colour representing a certain energy value above the global minimum. Values of each of the  $\varphi\psi$  dihedrals are plotted along axes and regions of stability appear as enclosed areas on contour maps illustrating the following:

- (1) Relative energy differences between conformers.
- (2) Conformational mobility available to a molecule in a given region of space.
- (3) Energy barriers separating different regions of stability.

Systematic grid search with energy minimisation at each step requires a large number of separate conformational energy calculations. Two dihedrals scanned at  $10^\circ$  intervals require 1296 separate calculations (1–2 h on a VAX 11/780), three require 46 656 (1 week on a VAX 11/780). Thus for structures containing more than three rotatable bonds, systematic grid search techniques become impractical. An alternative technique based on Monte Carlo methods is proposed for medium-sized molecules. Monte Carlo has been used very successfully in reproducing the thermodynamic and structural properties of solute–liquid and macromolecular systems [18] (e.g. proteins, polynucleotides, crystal lattice geometries, water–solute interaction) but has not been so thoroughly exploited for small, flexible molecules.

We [19] have recently adapted the stochastic Monte Carlo techniques based on the Metropolis sampling [16] algorithm as an automatic conformational search procedure, which if employed non-adiabatically in a variable temperature mode, can be used to identify various local and global minima and to calculate desired macroscopic properties by weighted averaging [20]. Rotatable bonds of interest are specified and during the simulation these are allowed to assume any value at random between 0 en  $360^\circ$ . A trial conformer is generated from the initial conformer by altering one dihedral at a time in a random fashion. The energy of this trial conformer is then considered for acceptance or rejection as the next step in the simulation by use of the Metropolis algorithm criteria. The conformational energy change is assessed. If the trial conformation is *lower* in energy it is accepted as the next step in the simulation. If, however, it is *higher* in energy, then the trial may or may not be accepted. This decision is made by computing the Boltzmann factor for the energy change:

$$\exp(-\Delta E/RT)$$

and comparing its value with a random number between 0 and 1. If this factor is greater than the random number, then the move is accepted, otherwise it is rejected. The procedure known as Metropolis sampling [17] is designed to achieve sampling of conformations in accordance with the Boltzmann Distribution where the probability of the conformer being included in the simulation is given by  $\exp(-\Delta E/RT)$ . The initial conformers generated are used to move the system away from the arbitrary starting situation. Subsequent steps are generated likewise using the Boltzmann factor criteria until the desired number of steps (say 1000) have been taken or until a given number of trials (say 100 000) have been accumulated, whichever is the sooner. The accepted conformers are then used to calculate the desired properties. Significant features of this method include:

- (1) Only the soft part of the potential function is used (i.e. torsional angles), the hard parts remain invariant (i.e. bond lengths and bond angles). This speeds up the computations considerably without detracting from the significant trends.
- (2) The random nature of deriving new conformers from previous ones enables tunnelling to occur through energy barriers (in one dimension). Molecules are therefore treated more as quantum objects than particulate systems (unlike molecular dynamics).
- (3) The occupation of the potential energy surface depends both on the number of torsions nominated to change in the simulation and on the temperature. To survey conformational space it is therefore necessary to inspect the acceptance/rejection ratio as the temperature is varied. The higher the temperature, the less the occupancy of local and global minima. When many trials produce no additional conformers one can conclude that one has found all the conformers accessible by applying a particular temperature in the simulation. One can thus rapidly obtain a set of available conformations.

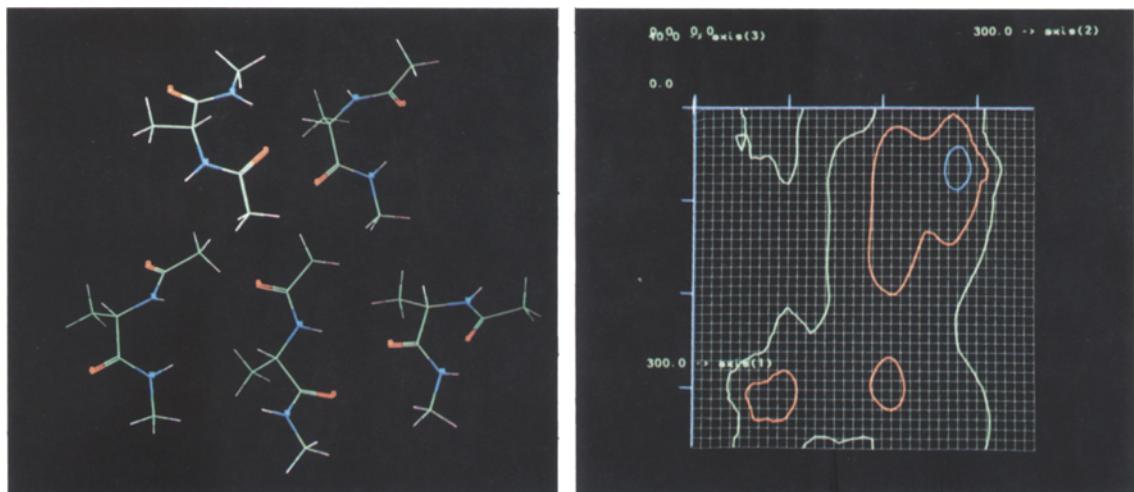


Fig. 7. Minimum energy conformers of Ala dipeptide and GRID search. Contours showing location of minima.

The results of various temperature simulations can be displayed either as a dynamic simulation or as colour-coded histograms of hyperspace depicting either frequency of occupation or calculated partition function surfaces. As an example for the Ala dipeptide, the simulation commenced at 95 K starting with the global minimum energy conformer C<sub>7eq</sub>, which is shown by the frequency histogram (Fig. 8) as the only observable minimum present at this temperature. At progressively higher temperatures, more local minima are observed, until at 950 K all five global and local minima are just about evident including P<sub>II</sub> and  $\alpha_R$  conformers which appear as lumps on the hypersurface. At 2000 K all the various minima which correspond to those observed by grid search become obvious as hills and valleys which are clearly discernable and at 3000 K, the whole of hyperspace is occupied to the maximum. Thus, in this example, the temperature of 2000 K would indicate all the transition probabilities that are discernable from grid search. The advantages of this Monte Carlo technique over grid search is mainly due to speed, the fact that only the softest modes (torsions) are considered, and that one can tunnel through energy barriers to other local minima by virtue of the Metropolis sampling technique.

Both techniques have been actively used in the search for an 'active site' conformation amongst EBI fungicides. The techniques are currently being utilised for predicting the activities of novel (hypothetical) compounds prior to synthesis.

#### *Electronic 'fingerprinting'*

Acetylcholine is a major excitatory neurotransmitter of insect nervous systems. The neurotransmitter binds to a receptor site associated with an ion channel, which causes the channel to open, thus initiating the nerve impulse in the postsynaptic cell. The receptor is a large multimeric protein consisting of 5 subunits of four different monomer types, each containing ca. 500 residues. A group of nitromethylene heterocycles (NMHs) are effective insecticides, and act as agonists

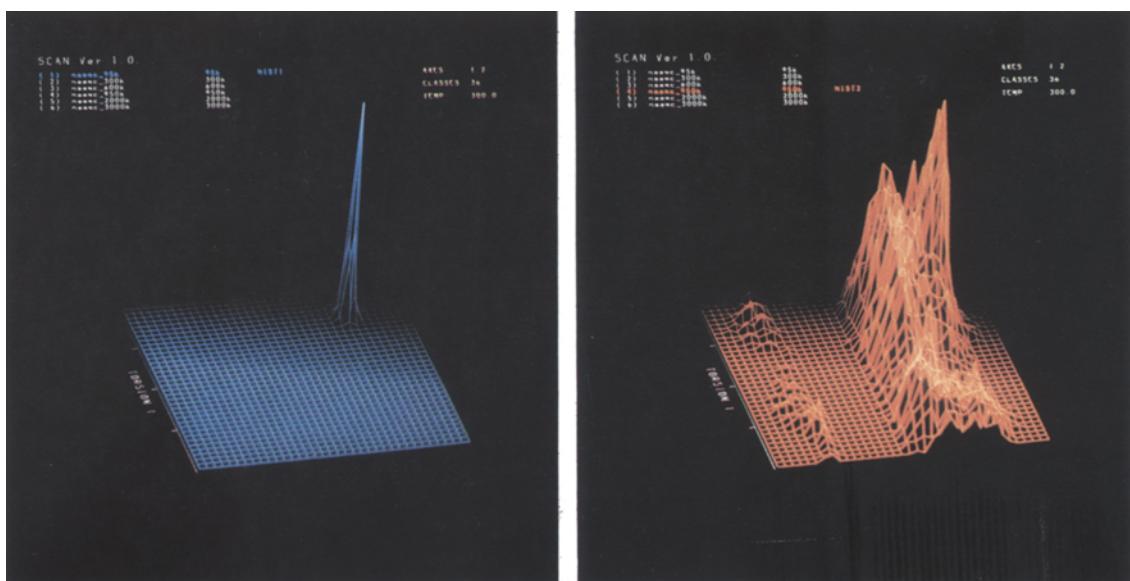
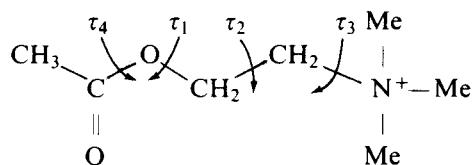


Fig. 8. Frequency plot using Monte Carlo search techniques of Ala dipeptide (blue 95 K, red 950 K).

at the receptor. These compounds along with acetylcholine have been the subject of a pharmacophoric mapping study.

Acetylcholine is a very flexible molecule and there has always been considerable debate [21–23] as to the ‘active site’ conformation of the neurotransmitter when bound in the receptor site. Various minimum energy conformers are to be found in the Cambridge Structural Database [24] as shown in Table 4. Conformational analysis based on molecular mechanics revealed that all X-ray conformers



are accessible for receptor binding (energy difference <3 kcal/mol). The fact that specific conformers are favoured in the solid, as shown by X-ray crystallography, may reflect preferential stabilisation by anion–cation interactions (i.e. hard vs. soft anions).

Structure-activity relationships on acetylcholine analogues have revealed certain requirements for receptor binding [21].

- (1) Agonists require a polarisable cationic function of limited size, sufficient to contain a quaternary alkyl ammonium group.
  - (2) Removal of the ether oxygen in acetylcholine hardly affects receptor binding, whereas the carbonyl oxygen is essential.
  - (3) The distance separating the cationic head  $\text{N}^+\text{Me}_3$  and the electron donor (e.g.  $\text{C}=\text{O}$ ) site is ca 5.2 Å when the acetylcholine receptor is in the activated state.

TABLE 4  
MINIMUM ENERGY CONFORMATIONS OF ACETYLCHOLINE (AcCh) SALTS TAKEN FROM X-RAY DATA [24] AND QUANTUM MECHANICAL/MOLECULAR MECHANICAL (CONFORMATIONAL ANALYSIS) CALCULATIONS

Complex	$\tau_1$	$\tau_2$	( $\tau_1, \tau_2$ )	Distance $\times \text{\AA}$
AcCh <sup>+</sup> Br <sup>-</sup>	78.9	78.4	(g <sup>+</sup> , g <sup>+</sup> )	3.72 <sup>a</sup>
AcCh <sup>+</sup> Cl <sup>-</sup>	193	84	(t, g <sup>+</sup> )	4.79 <sup>a</sup>
AcCh <sup>+</sup> ClO <sub>4</sub> <sup>-</sup>	180	80	(t, g <sup>+</sup> )	
AcCh <sup>+</sup> resorcylate <sup>-</sup>	163	77	(t, g <sup>+</sup> )	5.01 <sup>a</sup>
	201	270	(t, g <sup>-</sup> )	5.05 <sup>a</sup>
AcCh <sup>+</sup> tartrate <sup>-</sup>	164	78	(t, g <sup>+</sup> )	5.05 <sup>a</sup>
AcCh <sup>+</sup> tartrate <sup>-</sup>	124	351	(g <sup>+</sup> →t, g <sup>+</sup> )	4.53 <sup>a</sup>
	103	79	(g <sup>+</sup> →t, g <sup>+</sup> )	4.80 <sup>a</sup>
AcCh <sup>+</sup> I <sup>-</sup>	89	90	(g <sup>+</sup> , g <sup>+</sup> )	
3-Acetoxyquinuclidine	65	120	(g <sup>+</sup> , g <sup>+</sup> →t)	4.42 <sup>a</sup>
AcCh	180	180	(t,t)	5.19 <sup>a,b</sup>
AcCh	40	120	(g <sup>+</sup> , g <sup>+</sup> →t)	4.86 <sup>b</sup>
AcCh	172	44	(t, g <sup>+</sup> )	4.87 <sup>b</sup>

<sup>a</sup>X-ray data [24].

<sup>b</sup>Quantum mechanical/molecular mechanical calculations.

During this study attempts were made to address the following questions:

(1) Was there a pharmacophore which could accommodate the binding of both acetylcholine and the NMHs since there was no obvious atom-for-atom alignment between the two types of compounds (see Figs. 9 and 10)?

(2) NMHs containing six-membered rings have high insecticidal activity. Why is this activity drastically reduced in the five-membered ring analogues? This could not be explained on the basis of lipophilicity, conformation or steric effects.

Electronic descriptors were sought in the form of quantum mechanically derived molecular electrostatic potentials (MEPs). The MEP is a global potential, which represents the work needed to bring a unit positive charge from infinity to that point, providing the point charge does not perturb the electronic distribution of the molecule. MEPs have physical meaning and can be visualised as a cloud or surface surrounding the molecule (of varying density of interaction energy), suitably colour-coded (e.g. blue: nucleophilic, red: electrophilic, and green: neutral). These surfaces [26] can be compared/overlaid to investigate common patterns rather than atom-for-atom alignments. These representations, however, are static and involve purely electrostatic considerations, not polarisation, exchange repulsion, charge-transfer, or dispersion.

The MEP of acetylcholine (in the extended *trans, trans* (t,t) conformation) as shown in Fig. 9 suggests that a clear charge separation exists in the molecule, separating positive NMe<sub>3</sub><sup>+</sup> from the negative potential comprising the OCOCH<sub>3</sub> group. This charge separation is also apparent in the other conformers although the distance between positive and negative potentials is different (t,t ca. 5.2 Å; t,g<sup>+</sup> ca. 4.4 Å). A similar charge separation is evident in both E- and Z-isomers of the six-membered NMH which contain positive potential surrounding one half of the heterocycle from S to N-H. Both isomers would be accessible to the receptor by virtue of the extended

conjugation comprising the nitromethylene moiety (energy difference between E- and Z-isomers being ca. 5–8 kcal/mol). The E-isomer is stabilised by the intramolecular hydrogen bond formed between NH and NO<sub>2</sub>, whilst the Z-isomer gains some stabilisation from a coulombic through space interaction between the positive potential on S and the negative potential surrounding the NO<sub>2</sub> group. NMR studies [25] indicate the presence of the two isomeric forms, which are interchangeable at room temperature. The distances between centres of positive and negative potentials, respectively, in E- and Z-isomers correspond to the long and short distances of various conformers of acetylcholine (i.e. t,t 5.2 Å and t,g<sup>+</sup> 4.4 Å, respectively). Although specific atoms cannot be aligned, the MEPs are superimposable. In Fig. 10 the MEPs of the less active five-membered ring NMH show that there is a clear charge separation in the Z-isomer, but this is absent in the E-isomer, since fragments of negative potential reside over the N-H region of the heterocycle. Thus the heterocyclic ring angles, particularly those involving the C-S-C angle (six-membered ring 103° and five-membered ring 91.9°), are instrumental in determining the gross electronic properties for this group of compounds.

Certain electronic properties appear to be important for determining nicotinic cholinergic activity. The requirements include two types of conformers, one containing positive and negative potentials separated by a long distance (ca. 5.2 Å) and the other containing these moieties separated by a shorter distance (ca. 4.6 Å).

These results support the hypothesis that ready interchange between the two conformations is required to bring about agonist receptor binding. Like acetylcholine, the NMHs can bind in one conformation and remain bound like acetylcholine, while the protein flips to create the ion channel, by virtue of its ability to undergo a spatial change without changing its ‘electronic spots’.

#### *Factors regulating redox potential*

It is an aim of any modelling study to be able to rationalise cause versus effect. One such example is shown in the herbicidal properties of a group of PSI inhibitors, the heteropentalenes shown in Fig. 11. They have a similar action to paraquat (MV<sup>2+</sup>) and are believed to intercept the electron transport processes in plant photosynthetic Photosystem I proteins by competing with the natural substrate, ferrodoxin, for electrons emanating from P<sub>700</sub> primary electron donor protein. Once reduced, the anion radical reacts rapidly with dioxygen to regenerate MV<sup>2+</sup> and superoxide O<sub>2</sub><sup>·-</sup> which in the presence of hydrogen peroxide generates phytoxic OH<sup>·</sup> radicals lethal to plants. Heteropentalenes exhibit a similar symptomatology to paraquat and have redox potentials in the same range. Phytotoxic requirements for the heteropentalene series include: (1) a redox potential in the region of –450 mV (anything more negative would prevent the compound accepting electrons, energetics would be unfavourable), whilst compounds with reduction potentials less negative than ca. 200 mV would be thermodynamically inefficient at transferring the electron to oxygen (O<sub>2</sub>/O<sub>2</sub><sup>·-</sup> couple at –155 mV against NHE); (2) appropriate pK<sub>a</sub> of radical species; and (3) fast kinetics for electron transfer to oxygen to give the superoxide anion radical.

Reduction was proposed and subsequently proved by ESR [27] to involve the highly electron-deficient S=C linkage embedded in the centre of the heteropentalene ring (as observed by ESR quintet coupling of the anion radical with four methylene protons). The redox potential could be fine-tuned by altering the 4-substituent in the six-membered ring. X-ray data and molecular

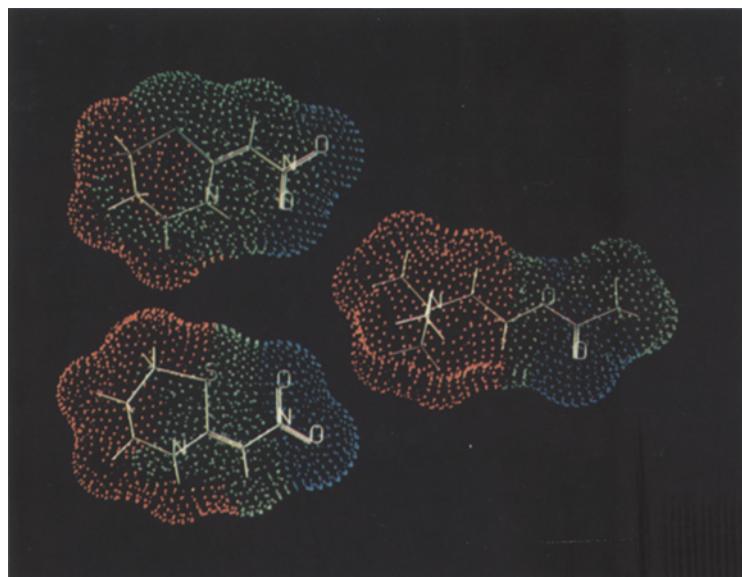


Fig. 9. Molecular electrostatic potentials of the NMH insecticide and acetylcholine (computed on a Connolly surface).

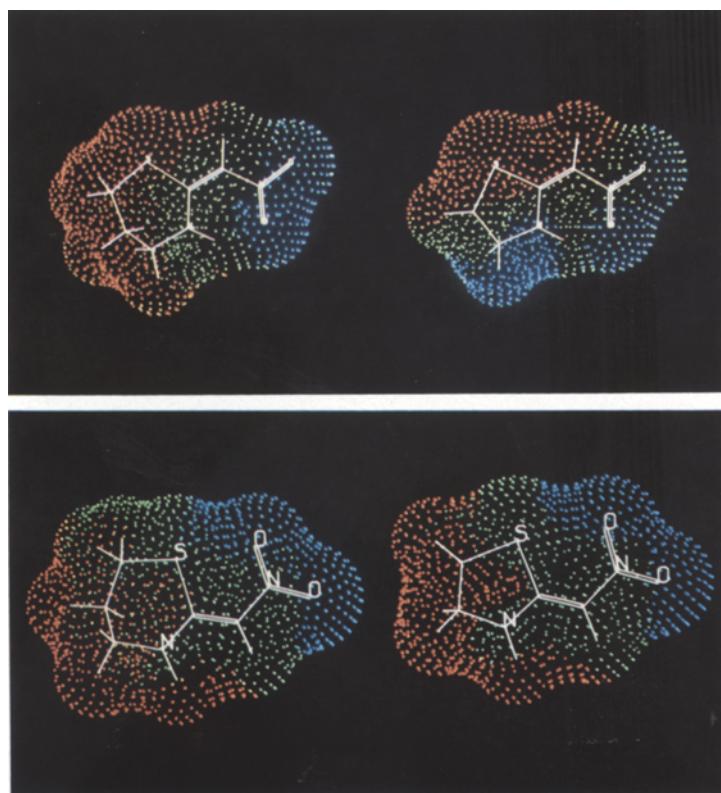


Fig. 10. MEPs of E- and Z-isomers of 6- and 5-membered rings of NMH insecticides.

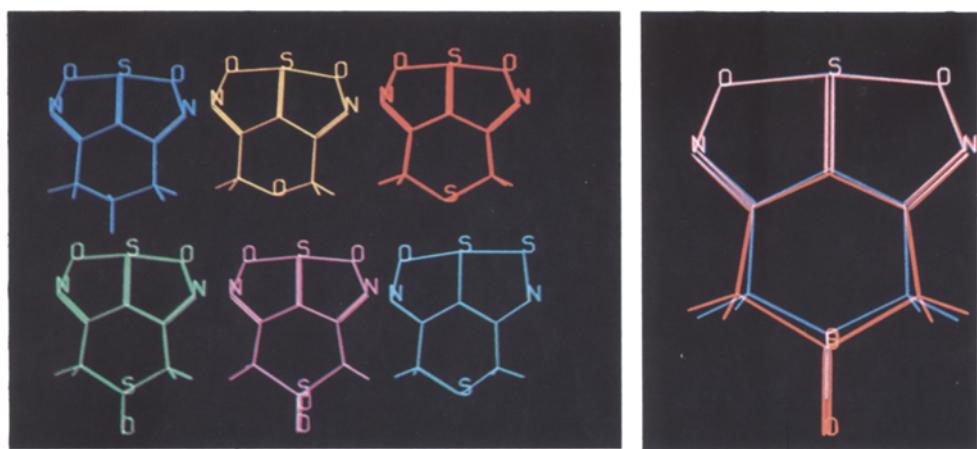


Fig. 11. Heteropentalene herbicides.

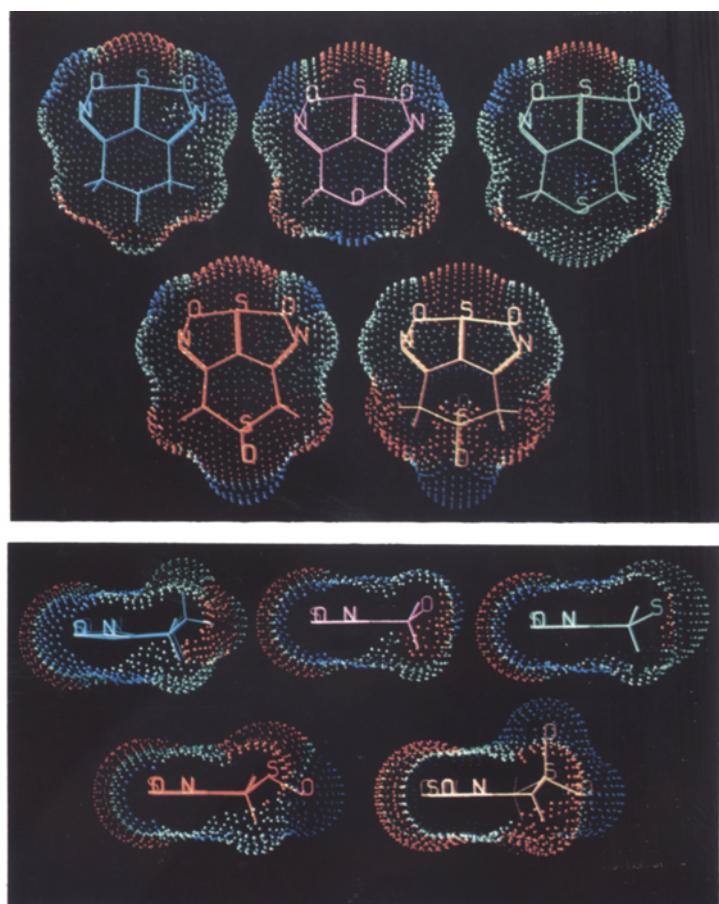


Fig. 12. MEPs of heteropentalene herbicides.

models showed that the 4-substituent is appreciably separated (by ca. 4.6 Å) from the S, indicating that a through space electronic effect between the two functions would be highly unlikely. It was suggested [28] that one reason for the variation in redox potentials amongst heteropentalenes was related to changes in strain which are relieved or increased upon reduction. We attempted to analyse this group of compounds armed with the latest X-ray data and MNDO/AMPAC parameters for S, N and O.

The X-ray geometries were not used directly (because of irregularities in bond lengths and angles) but were first optimised. The optimised geometries could be obtained through use of the symmetry option in the program [29]. Comparisons of resulting optimised structures show that as strain increases so the redox potential becomes more negative. The strain (Fig. 11) is evident in the small differences in distances between C<sub>2</sub> and C<sub>6</sub> (2.54 Å in the methylene compound, where E = 693 mV and increases to 2.57 Å in the SO<sub>2</sub>-substituted compound, where E = 523 mV), and in the internal C<sub>2</sub>-C<sub>1</sub>-C<sub>6</sub> ring angle which varies from 127.2° in the methylene-substituted compound to 131° in the SO<sub>2</sub> compound. In the disubstituted open ring acyclic system, where there is least strain, the redox potential falls to -250 mV. Thus, by altering the 4-substituent, one in turn alters the C<sub>2</sub>-C<sub>6</sub> distance and internal C<sub>2</sub>-C<sub>1</sub>-C<sub>6</sub> ring angle. The 4-substituent behaves as a mechanical crow-bar increasing or decreasing the strain of the heteropentalene ring system.

What are the repercussions of this internal ring strain on the electronic properties of the molecules in Fig. 12 as a whole? A view taken of the MEPs of various molecules shows a large protruding positive potential region associated with the electron-deficient central sulphur atom, which could act as an attractive site for electron transfer. The central core of the heteropentalene ring comprising delocalised O, N and C atoms has a veil of negative potential. As shown, in Fig. 12, the redox potentials become less negative as the overall positive potentials in the molecules increase on going from left to right from CH<sub>2</sub> to O, S, SO and finally SO<sub>2</sub>. The greater the electron deficiency associated with the 4-substituent (i.e. larger valency on S) the greater the redox potential. Another interesting phenomenon is the positive potential surrounding the two lots of methylene protons on C<sub>3</sub> and C<sub>5</sub>. These also appear to become more positive (as shown by the values of residual charge for compounds with X=CH<sub>2</sub>, 0.029; to X=SO<sub>2</sub>, 0.095) as the redox potentials become more positive. Their inherent acidity has been demonstrated with proton

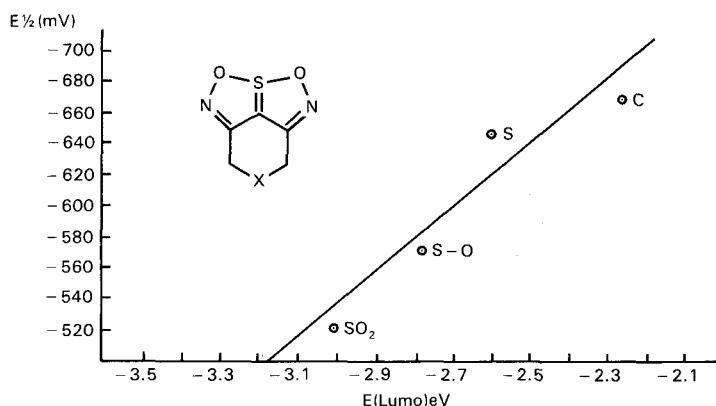


Fig. 13. Heteropentalenes: Redox potential vs. E (LUMO).

deuterium NMR exchange experiments. Furthermore, if one then examines C<sub>2</sub>-C<sub>3</sub> and C<sub>5</sub>-C<sub>6</sub> bond length differences, the compounds exhibiting least strain, highest redox potentials, also exhibit appreciable C<sub>2</sub>-C<sub>3</sub>/C<sub>5</sub>-C<sub>6</sub> bond shortening (ca. 0.02 Å). This effect is reminiscent of a hyperconjugative effect which increases with enhanced electron deficiency and stabilisation of the heteropentalene moiety, through increased extended conjugation in the system.

The MEPs do not reveal anything about reactivity, for example, location of the unpaired electron under kinetic control and the energetics of the reduction process. To address these questions, we must examine both eigenvalues and eigenvectors. The energy level that is of prime concern is the frontier orbital, the LUMO into which the unpaired electron is donated. The lower the energy of this eigenvalue, the more susceptible the compound to accepting the electron. H. Rzepa [27, 29] observed a trend as in Fig. 13 between the energy of the LUMO (which in this case is negative because the compounds are electron-deficient) and the redox potentials. The differences in heats of formation between radical anion and that of the starting heteropentalene (which signifies heat of reaction for the electron transfer process) also follows the trend in redox potential (Fig. 14). The more exothermic the reaction (i.e. more negative the heat of reaction to form anion radical), the less negative is the redox potential (i.e. easier it is to reduce). Thus the heat of reaction differences between 4-methylene (redox potential -693 mV) and 4-SO<sub>2</sub> (-523 mV) spans 0.95 eV or 21.8 kcal/mol.

In order to locate the position(s) into which the electron is most likely to reside upon electron transfer, it is necessary to view the orbital coefficient or eigenvectors within the LUMO energy level. The program ORBIT (T. Dickens and H. Prout, Glaxo) [30] allows a simplistic analysis in terms of the atomic (rather than molecular) orbital coefficients consisting of discrete s, p and sp orbitals. As shown in Fig. 15, a dominant p-orbital coefficient is located on C<sub>1</sub> (d orbitals on S are not included in MNDO/AMPAC orbital coefficients). The spin densities [29] on the anion radical also indicate highest values occur on C<sub>1</sub> (0.575) compared to S (0.17). Comparing atomic orbital coefficients among a series of compounds indicates that the actual location of the eigenvectors in the LUMO are invariant to substitution, with the largest contribution consistently on C<sub>1</sub>. It is only the energy value of the LUMO itself that varies.

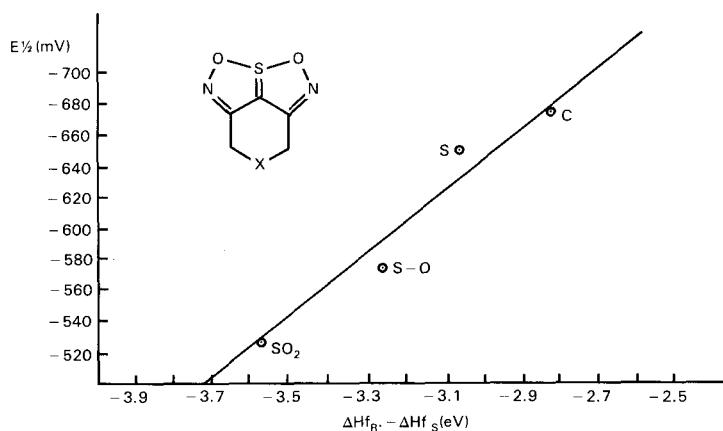


Fig. 14. Heteropentalenes: Redox potential vs.  $\Delta H_f^{\circ} - \Delta H_f^{\circ}$  to form anion radical (MNDO).

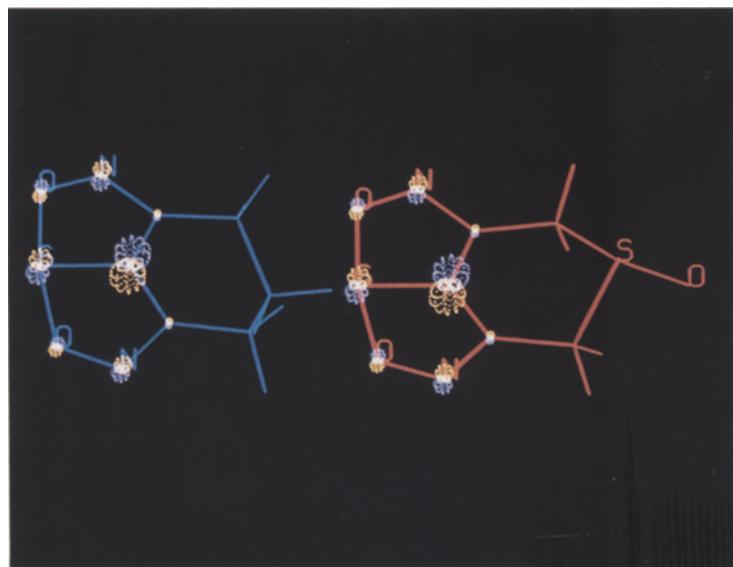


Fig. 15. Atomic orbital coefficients of heteropentalene herbicides using ORBIT (T. Dickens and H. Prout, Glaxo [28]).

To summarise, in spite of the fairly primitive nature of these calculations, various pertinent trends are apparent. The model used for the primary event of reduction, the addition of an electron to C<sub>1</sub>, suggests that a method of fine-tuning the redox potential involves inducing strain in the molecule. A consequence of the strain about the C<sub>2</sub>-C<sub>1</sub>-C<sub>6</sub> angle is to alter the energy levels. Heteropentalenes with least strain exhibit an enhanced type of hyperconjugative effect which in turn stabilises the system. There may be other secondary factors which also influence the redox potential for this series of compounds, such as local inductive effects of substituents which can alter the overall electron deficiency of the heteropentalene, physicochemical properties (e.g. log P, solubility) and pK<sub>a</sub>. Nonetheless, it is satisfying that fairly straightforward theoretical techniques can be applied giving results consistent with experimental observation [27].

#### *Intermolecular interactions, solvation and the binding site problem*

Interactions between molecules whether non-covalent or covalent provide the basis for chemical and biological specificity whether through base pairing in DNA, ligand protein interactions or hydrolysis via tetrahedral intermediates formed by nucleophilic attack of H<sub>2</sub>O or OH<sup>-</sup>. There is an increasing need to be able to analyse the nature of interactions. Ab initio methods can be used whereby calculated energies can be expanded using energy decomposition as proposed by Morokuma [31]. Semiempirical techniques like MNDO/AMPAC now allow calculations to be carried out using dynamic reaction coordinates. One can also analyse changes in conformation and electronic properties at each step along the reaction coordinate. An example is illustrated in a thiazinone derivative, depicted in Fig. 16. Its susceptibility to undergo nucleophilic attack with subsequent hydrolysis at the electrophilic carbon atom C<sub>2</sub> to give methanol and thiazin-2-4-dione, was of relevance in determining biological activity. Solvation effects were considered important

for deciding whether hydrolysis or binding takes place prior to or upon binding at a receptor site. Under kinetic control, the properties of the transition state can be evaluated (i.e. the activation energy to attain the transition state, charge and orbital distribution), where the transition state in this case is a tetrahedral intermediate formed by nucleophilic attack of  $\text{H}_2\text{O}$  or  $\text{OH}^-$ . In

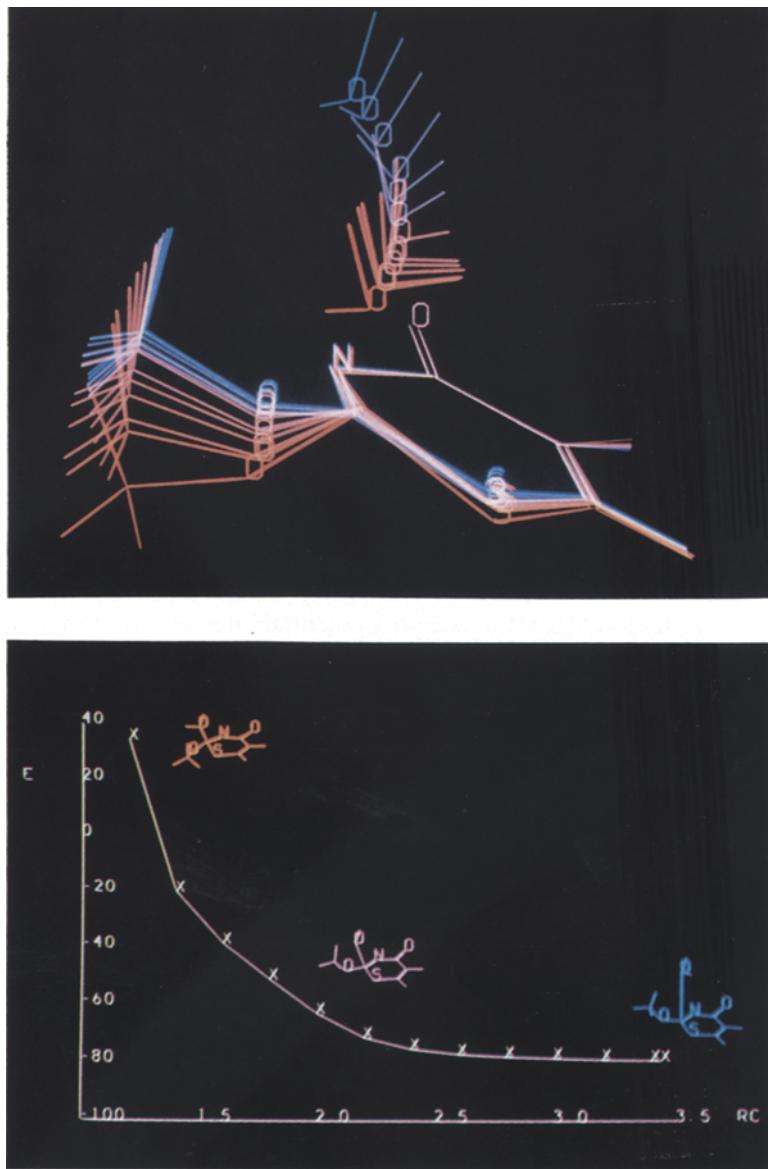


Fig. 16. Following a reaction path for hydrolysis of a thiazinone derivative.

attempting to correlate electronic factors likely to vary with hydrolysis rates; it is important to understand the way in which water molecules can interact with the thiazinone moiety, in addition to the overall thermodynamic factors (i.e.  $\Delta H^\circ$  and  $\Delta S^\circ$ ) involved.

A quantum mechanical simulation or Morokuma energy component analysis of solvation effects on a molecular system as complex as thiazinone would be too large a task. We have used MNDO to model hydrolysis, following the approach of a water molecule travelling along the reaction coordinate perpendicular to the thiazinone ring, towards the reactive C<sub>2</sub> atom. One can observe the formation of the tetrahedral transition state. The energy for this process can be plotted as a function of the reaction coordinate. As shown in Fig. 16, there appears no definite transition state due to neglect of solvation, which would, in fact, stabilise developing zwitterionic forms encouraging the formation of a transition state. We [29] are in the process of modelling this type of complex with several water molecules.

It is hoped that these recent developments in semiempirical techniques especially, as shown by the new parameters and functions in AM1, will advance our understanding of the quantum-mechanical aspects of solvation of such heteroatomic systems, which cannot be easily simplified into smaller fragments for ab initio calculations.

Other approaches for investigating intermolecular non-covalent interactions are based on semiempirical potential energy functions as seen in molecular mechanics energy minimisation techniques and energy partitioning functions as exemplified by Goodford's program. Non-bonded components are split between the van der Waals (Lennard-Jones) term which can be attractive or repulsive, electrostatic (Coulombic ion-pairing) and hydrogen bonding (which is partly electrostatic with directional character). There are also hydrophobic forces which may be entropically driven (i.e. requiring desolvation of both inhibitor/substrate and receptor active site).

Although Goodford's program GRIN/GRID has usually been used to locate potential ligand binding sites in protein structures, we have found it particularly useful when used as a complementary tool alongside MEPs for assessing potential interactions of small molecules with probes representing component amino acids of the receptor site (e.g. H<sub>2</sub>O, CH<sub>3</sub>, OH, NH<sub>2</sub>, etc.). We have used it as a comparative technique to analyse ligand binding sites amongst compounds known to bind to a common receptor. Potential proton donor or acceptor sites may be identified, some of which may be shielded from binding by neighbouring hydrophobic groups (e.g. Me, <sup>iso</sup>Pr, etc.). It can also be informative for predicting trends in lipophilicity with conformational variability, since solvent accessible surfaces of proton donor and acceptor groups can change with conformation. GRID predicts higher energy of interaction of a water probe with conformers containing greater exposed surface areas of hydrophilic groups than those conformers in which these groups are shielded. In certain cases, these interaction energies correlate with the observed differences in measured log P values between isomers.

We illustrate the application of GRID as a fast routine for calculating 'hot spots' in molecules which are likely to interact with proton donor/acceptor probes (e.g. water). The example used is that of N-acetyl-alanyl-N-methylamide, Ala dipeptide and a water probe. Earlier work by Hagler et al. [32] using Monte Carlo simulations of water behaviour around various conformers of the Ala dipeptide, predicted that whilst the C<sub>7</sub> conformer is favoured in the gas phase (by ca. 4.5 kcal/mol), the converged total energy for the  $\alpha_R$  form was -5.6 kcal/mol more stable than the C<sub>7</sub>, C<sub>5</sub> and P<sub>II</sub> conformers in water. Beveridge et al. [18] proposed that the total free energy is actually similar for all conformers, which are all thermally accessible at room temperature. Although the

major factor contributing to the stabilisation of the internal energy of hydration of  $\alpha_R$  and  $P_{II}$  in aqueous solution is due to hydration of the accessible carbonyl oxygens, that are not internally hydrogen bonded, there is a compensating entropy change for hydration which is negative for both  $P_{II}$  and  $\alpha_R$ .

This is presumably due to the increased ordering of water molecules in the region of the hydration spheres of these open forms. Goodford's program located contour regions which correspond to the most favourable (probable) binding positions for water (Fig. 17). These agree with the probability density maps of water molecules obtained by using Hagler's Monte Carlo simulation. The 'hot spots' are recorded giving the most favourable energies of interaction in terms of electrostatics, van der Waals and hydrogen bonding contributions. For the  $C_7$  conformers, the program predicts a binding energy of ca.  $-6.1$  kcal/mol suggesting that a water molecule can share hydrogen bonds with the intramolecular seven-membered hydrogen bond system by using  $C=O$  and  $N-H$ . In addition the accessible  $C=O$  group acts as a proton acceptor to water. The  $C_5$  conformer can also form intermolecular hydrogen bonds to water, where the favourable energy of interaction is  $-7.36$  kcal/mol. This indicates that the local geometry of a five-membered ring is more accessible to water than the seven-membered ring. The two other local minima  $P_{II}$  and  $\alpha_R$  form weaker 'hot spots' in binding to a water probe with favourable binding energies of only  $-5.1$  kcal/mol for both  $P_{II}$  and  $\alpha_R$  conformers, respectively. This is perhaps unexpected in view of Hagler's results which suggest that the intermolecularly hydrogen-bonded systems are favoured in water.

Nevertheless, one must appreciate that these interaction energies signify the hottest spots and are not total energies or interaction of free energies. There are, in fact, several binding sites within each conformer and the volumes of contours generated and energy values of each point should be taken into account before attempting to assess the total enthalpic contribution to binding of a probe. The program is useful in that it immediately identifies a more favourable binding site for the five-membered intramolecular hydrogen-bonded conformer than the seven-membered conformer. Different probes can be used (e.g.  $H_2O$ ,  $OH^-$ ,  $NH_2$ , etc.) and if overlayed it is possible to construct a 3D picture of receptor site residues likely to be surrounding the molecules of interest. When different classes of molecules are compared in their minimum energy conformations, structure activity patterns can emerge.

## CONCLUSIONS

We have illustrated how various molecular modelling techniques can be used to analyse substrate binding properties and to design an optimum inhibitor structure for a given receptor site. Various important points have emerged during these studies. Most important, we rely on good and dependable numbers for biological activity. Without this, even the most wondrous hypotheses and correlations are meaningless. Secondly, in the absence of X-ray structures of a receptor site of interest, we are working blindfold. X-ray data on receptor-inhibitor complexes, even homologous ones, provide new insights and inspiration. These two aspects go hand in hand, increasing our understanding at the molecular level by advancing theoretical and computational studies to interpret the complex factors involved in determining why certain compounds are active, whilst others are not.

Whilst molecular modelling techniques can provide mechanistic insights and physicochemical

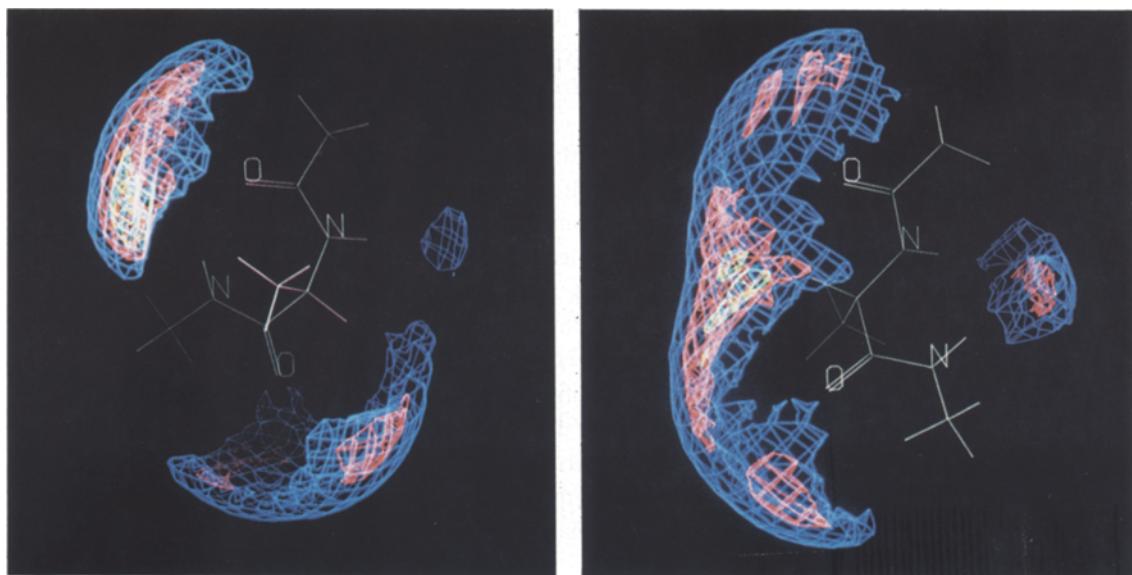


Fig. 17. Contours showing favourable interactions with a  $\text{H}_2\text{O}$  probe and the Ala dipeptide using GRID/GRIN.

explanations for biological activity, results from such studies have often been accused of being introspective. They help to explain results but often do not go beyond, to offer vital links to completely different classes of compounds. These connections are the biggest prize in the discovery process. Such thought processes are often exercised effectively on congeneric series of compounds, but are not usually extrapolated to a wider range of chemical functionalities.

We have illustrated the complementary role of combinations of technique ranging from X-ray, NMR data, QSAR, quantum mechanics, molecular mechanics, etc. in order to establish essential conformational and electronic requirements for biological activity. This new discipline of molecular design will always rely on a multidisciplinary approach; it is always dangerous to rely on only one technique! In the early stages of a project, there is the urgent need to establish the critical parameters which directly influence biological activity. These must be probed quickly and effectively through the use of good experimental design. At this stage, one is seeking mechanistic explanations for activity, which may involve common shape, electronic or reactivity properties. Here, molecular modelling, physicochemical measurement and statistical methods must be employed together until a hypothesis emerges, whose worth must be iteratively tested against wider groups of compounds.

Current techniques are being stretched to deal with compounds containing an abundance of hetero-atoms, involved in extended conjugation. Compounds also vary in their degree of flexibility. Thus, the techniques available must be improved to meet the demands of the user. Ab initio quantum mechanical techniques combined with more sophisticated computers, will gradually be able to solve problems of greater complexity and give reliable results for the types of heteroatomic systems of interest. Parameterisation of forcefields for heteroatomic/conjugated systems has been slow and inconsistent, due to preoccupation with protein and polynucleotide systems, where much effort has been spent on experimenting with the energy minimum problem,

in the form of molecular dynamics and Monte Carlo simulations. Greater investment is needed from companies working on medium-sized heteroatomic systems, and this may transpire from the various consortia that have appeared during recent years. The new areas of molecular dynamics and Monte Carlo techniques will undoubtedly enhance our understanding of solvation effects and intermolecular interactions and hopefully this will coincide with advances in quantum mechanical methods for systems of pesticidal interest.

#### ACKNOWLEDGEMENTS

B. Odell would like to thank Dr. H. Rzepa and Dr. D. Williams from Imperial College for helpful discussion and considerable input into this work, Dr. M. Tute from Pfizer, Sandwich, Dr. P. Goodford of Oxford University, and past and present members of the molecular modelling group at Shell Research (A. Dearing, R. Ryan, J. Vilkauls, P. Camilleri, M. Hilton, A. Bracher and J. Rogan).

#### REFERENCES

- 1 Vorpagel, E.R., In Ragdale, N.N. and Kuhn, P.J. (Eds.) *Pesticides and Minimising the Risks*, ACS Symposium 336, American Chemical Society, Washington, D.C., 1987.
- 2 Lunt, G., Bath University, private communication.
- 3 Kelly, K., Sheffield University, private communication.
- 4 Deisenhofer, J., Epp, O., Miki, K., Huber, R. and Michel, H., *Nature*, 318 (1985) 618.
- 5 Dewar, M.J.S., Zoebisch, E.G., Healy, E.F. and Stewart, J.J.P., *J. Amer. Chem. Soc.*, 107 (1985) 3902–3909 (QCPE Program No. 506).
- 6 Weiner, S.J., Kollman, P.A., Nguyen, D.T. and Case, D.A., *J. Comp. Chem.*, 1 (1986) 230 (and references therein).
- 7 Burkert, U. and Allinger, N.L., *Molecular Mechanics*, ACS, Washington, DC, 1982; MMP2, Molecular Design Ltd; version 6.0, 1985; QCPE Program No. MMP2 (85).
- 8 Kollman, P. and Blaney, J., *Top. Mol. Pharmacol.*, 3 (1986), 285–305.
- 9 Poulos, T.L., Finzel, B.C., Gunsalus, I.C., Wagner, G.C. and Kraut, J., *J. Biol. Chem.*, (1985) 16122.
- 10 Poulos, T.L., Finzel, B.C. and Howard, A.J., *J. Mol. Biol.*, 195 (1987) 687–700.
- 11 Griffin, B.W. and Peterson, J.A., *Biochemistry*, 11 (1972) 4740.
- 12 Lipscomb, J.D., *Biochemistry*, 19 (1980) 3950.
- 13 Rossi, M., *J. Med. Chem.*, 26 (1983) 1247.
- 14 Goodford, P., *J. Med. Chem.*, 28, (1985) 849.
- 15 Weiner, S.J., Singh, U.C., O'Donnell, T.J. and Kollman, P.A., *J. Am. Chem. Soc.*, 106 (1984) 6243–6245.
- 16 Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N. and Feller, A., *J. Chem. Phys.*, 21 (1953) 1087.
- 17 Beveridge, D.L., Ravishankar, G., Mezei, M. and Gedulin, B., *Biomolecular Stereodynamics III*, Proceedings of the Fourth Convers, in Discipline of Biomolecular Stereodynamics, State University of New York, New York, NY, 1985 (publ. 1986), Vol. 3, pp. 237–252.
- 18 Hagler, A.T. and Moult, J., *Nature*, 272 (1978) 222.
- 19 Ryan, R.P. and Odell, B., publication in preparation.
- 20 Saunders, M., *J. Am. Chem. Soc.*, 109 (1987), 3150–3152.
- 21 Klimkowski, J.V., Schafter, L. and Scarsdal, J.N., *J. Mol. Struct. (Theochem.)*, 109 (1984) 311.
- 22 Genson, D.W. and Christoffersen, R.F., *J. Am. Chem. Soc.*, 95 (1973) 362.
- 23 Port, G.N. and Pullman, A., *J. Am. Chem. Soc.*, 95 (1973) 4059.
- 24 Cambridge Structural Database, Cambridge Crystallographic Data Centre and Allan, F.A., *Acta Crystallogr., Ser. B*, 35 (1979) 2331–2339.
- 25 Rajappa, S., Nagarajan, K., Venkatesan, K., Kamath, N., Padmanabhan, V.M., von Philipsborn, W., Chin Chen, B. and Müller, R., *Helv. Chim. Acta*, 67 (1984) 1669.

- 26 Weinstein, A., Meayani, S., Srebrenik, S., Cohen, S. and Sokolovsky, M., Mol. Pharmacol., 11 (1975) 671.
- 27 Camilleri, P. and Rzepa, H., publication in preparation.
- 28 Dearing, A., personal communication.
- 29 Rzepa, H., personal communication.
- 30 Dickens, T. and Prout, H., personal communications.
- 31 Politzer, P. and Truhlar, D.G., Chemical Applications of Atomic and Molecular Electrostatic Potentials, Plenum Press, New York, NY, 1980.
- 32 Hagler, A.T., Osguthorpe, D.J. and Robson, B., Science, 208 (1980) 599.