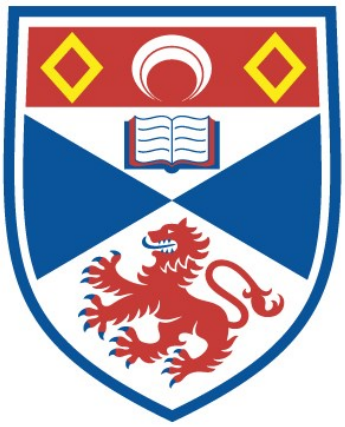
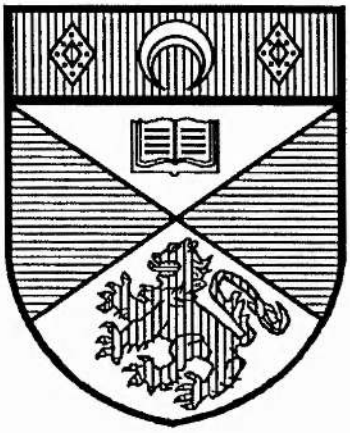
A THEORETICAL STUDY OF TPA-LIKE TUMOUR PROMOTORS AND INOSITOL POLYPHOSPHATES

John Wilkie

A Thesis Submitted for the Degree of PhD
at the
University of St Andrews



1991



A Theoretical Study of TPA-Like Tumour Promoters and Inositol Polyphosphates.

A Thesis presented for the degree of

Doctor of Philosophy

in the Faculty of Science of the

University of St. Andrews

by John Wilkie, BA.

St. Leonard's College

November 1990

Declarations.

I, John Wilkie, hereby certify that this thesis has been composed by myself, that it is a record of my own work, and that it has not been accepted in partial or complete fulfilment of any other degree of professional qualification.

..... Date....25/11/90

I was admitted to the Faculty of Science of the University of St. Andrews under Ordinance General No 12 of 1st October 1987 and as a candidate for the degree of PhD on the 29th September 1988.

..... Date....25/11/90....

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regualtions appropriate to the degree of PhD.

..... Date....27/11/90.....

In submitting this thesis to the University of St.Andrews, I understand that I am giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. I also understand that the title and abstract will be published, and that a copy of the work may be made and supplied to any *bona fide* library or research worker.

..... Date..25/11/90

Acknowledgement.

I would like to thank Dr. Colin Thomson for his supervision during this project and the NFCR and AICR for financial support. I would also like to thank Dr. Graham Richards and Dr. Chris Reynolds who first started my interest in molecular modelling.

To Taid.

Abstract.

In this thesis, the structures and selected properties of TPA-like tumour promoters and of *Myo* -inositol polyphosphates are calculated and compared with their promoting and calcium releasing activities respectively. Three different levels of theory are used in the calculations, namely *ab initio,* semi empirical and molecular mechanics, though the majority of the calculations are performed using the MNDO derived semi-empirical methods implemented in MOPAC.

A detailed description of the phorbol ester / DAG binding site is obtained from the structure / activity relationships derived for the tumour promoters and molecular dynamics simulations of the phorbol ester, TPA, in a lipid bilayer is carried out to investigate the position of the binding groups in relation to the surface of the bilayer. In order to carry out this simulation, a method (HYDRO) has been developed to produce close packed heterogeneous bilayers in which the headgroups of the components lie in random orientations. Thus the simulations are more realistic than calculations in which the lipids are placed in a regular array and interstitial spaces due to the difference in surface area of the bilayer components are kept to a minimum.

The effects of the number and position of phosphate groups on the ring conformations of *myo* -inositol phosphates and the connection between this and calcium releasing activity are studied in chapter 5. As the molecules are particularly flexible with a large number of potential local minima, phosphate groups have been added sequentially to keep the required number of starting points as low as possible and rotation of phosphate hydroxyls has been ignored. The heats of formation, calculated with the different semi-empirical parametrisations, differ considerably, so the final calculations chapter compares energies and selected properties calculated for model organics phosphates using different methods and theory. The results using the new parametrisation, PM3, are compared with those of the earlier AM1 for both the phorbol systems and the inositol phosphates to test its suitability.

Chapter 1

Introduction.

1.1 Cancer

The term Cancer is used to describe the disease caused by uncontrolled growth and division of particular cell types with the mass of cells being referred to as a tumour. In most cases, the affected cells would not normally undergo division, but during the development of cancer, their normal behaviour is altered as the cells enter a transformed state, reflecting the loss of their original function. The actual extent of the cellular changes vary enormously from one tumour to another. In the case of a benign tumour, it may be difficult to distinguish the transformed cells from neighbouring unaffected cells, but in some cases, particularly malignant tumours, the cells are clearly distinguishable from the surrounding tissue. Cells may have significant chromosome damage, especially cross-overs, show poor differentiation and possess highly changed morphology, and they may be dividing rapidly. The extent of a benign tumour is often well defined, it may even be encapsulated, but malignant tumours frequently invade neighbouring tissue or shed cells into the blood stream and lymphatic system, leading to the growth of secondary tumours (metastases). In addition to this morphological classification, tumours can also be classified according to the embryonic origin of the affected cells, in that carcinomas develop from epithelial cells while sarcomas develop from connective tissue.

There are clearly many different forms of Cancer, with radically different morphologies and epidemiologies, with many causes, including viruses, chemicals and ionising radiation. The common factor linking these different carcinogenic agents is their ability to alter or damage cellular DNA or bring about changes in gene expression.

1.2 Chemical Carcinogenesis

Chemicals have been implicated in the development of cancer for many years, in 1775 a report linked the incidence of scrotal cancer amongst chimney sweeps to exposure to soot [1]. Since that time, animal experiments have shown that many other compounds are carcinogenic, including polycyclic aromatic hydrocarbons (initially extracted from coal tar), aromatic amines, many alkylating agents and inorganic substances such as chromates and arsenic [2]. Many of these compounds are electrophiles, or are capable of being metabolised into electrophiles, and can thus attack nucleophilic sites in DNA and proteins leading to genetic mutations or disrupting control of gene expression. The alkylating agents also cause damage and incorrect transcription or replication of the DNA. Some carcinogens are neither electrophiles nor alkylating agents, but are still able to interfere with DNA transcription . Planar molecules are able to lie between adjacent base pairs in the DNA molecule (intercalation), distorting its structure, often partially unwinding it, and so disrupting transcription and replication. Electrostatic interactions between the intercalating molecule and the phosphate groups of the DNA have a major effect on the affinity for such binding, and in some cases, the intercalation has some sequence specificity from interactions with the base-pairs. Chemicals which are sufficient in themselves, whether requiring a single exposure or repeated doses, to transform cells and so produce cancers are referred to as complete (or solitary) carcinogens.

1.2.1 Tumour Promotion and Cocarcinogenesis

It has been suggested that solitary carcinogenesis, as outlined above, is very much the exception rather than the rule [3]. Compounds that are not carcinogenic in themselves have been shown to enhance the ability of solitary carcinogens to transform cells and produce tumours [4]. Such compounds are termed Cocarcinogens or Tumour Promoters according to the timing of administration. Cocarcinogens are administered at

the same time as the carcinogen, while promoters are administered afterwards. There is considerable confusion in the literature over the use of the term cocarcinogen, as some authors use the term to cover all compounds that enhance carcinogenesis, regardless of mechanism [5].

Tumour promotion occurs over a long period of time, many years in man: weeks in mice, and requires repeated exposure to low doses of the promoter. In order to be susceptible to promotion, affected cells must have already been exposed to low, subcarcinogenic, doses of a conventional carcinogen, a process refered to as initiation. It is believed that this initiating dose brings about cellular changes that allow the promoters to have their effect, though has yet no such changes have been identified. Just a single initiating dose is required and the process is irreversible but the effects of the promoters can be reversed merely by ceasing the application of the promoter at any stage up to the development of the tumour. Thus identification and elimination of potential tumour promoters should lead to a reduction in the incidence of cancer. Experimentally, the Initiation / Promotion procedure has been well documented and shown to lead to a high incidence of tumours in mice and many other systems [4, 6, 7].

Though the Initiation / Promotion mechanism is generally accepted, there is some question over whether solitary carcinogens act via this mechanism, possessing both initiating and promoting activities or whether they act via a separate, single-stage, mechanism. Early work [8, 9] produced evidence in favour of both mechanisms as solitary carcinogens could act as both initiators and promoters in separate experiments [8], but retinoic acid was able to inhibit the promoting activity of 12-O-Tetradecanoylphorbol-13-Acetate (TPA), a conventional tumour promoter, but not that of 7,12-Dimethylbenz[a]anthracene (DMBA), a solitary carcinogen [9]. Subsequently the view that solitary carcinogens possess both initiating and promoting activities has become accepted [7], but with the complication that some solitary carcinogens have been shown to induce an endogenous promoter such as the prostaglandins [10] rather than possess the activity themselves.

Hecker and co-workers have developed a standard assay for tumour promoting activity using mouse skin [6]. A small dose of the carcinogen DMBA is applied to a

small shaved area of the skin of a mouse, and then followed by repeated doses of the promoter over a period of weeks. The number of tumours, per survivor, that have developed after 12, 24 or 36 weeks are used to express the activity of the tumour promoter. The results are also compared against control experiments, involving only solvent (acetone) as the promoter or applying the promoter without prior application of an initiator. From this, the promoting activities of different compounds can be compared, as can the effects of varying the dose of promoter. This method has been used by Hecker to assess the effects of modifications to the phorbol moiety of TPA on promoting activity [11], data that forms the basis for the work in chapter 4.

Lifetime treatment of mouse skin with promoting doses of TPA leads to the formation of both benign and malignant tumours, even in the absence of experimentally applied initiator [12]. Whether this low rate of tumour formation is due to a weak initiating activity of TPA, or from the presence of environmental initiators, is not known. It is possible that the results of the assay are distorted by features unique to mouse skin, but similar two-stage carcinogenesis has been reported in both rats and hamsters [13, 14], while TPA has been shown to act as a tumour promoter in tissues other than skin [15].

The concept of two-stage carcinogenesis was further complicated by the discovery that neither mezerein, which is structurally very similar to TPA, or 12-O-Retinoylphorbol-13-Acetate (RPA, see chapter 4 for structural details) are able to promote tumours after initiation with DMBA, but can replace TPA as the promoter part of the way through the promotion stage [16]. As a result, tumour promotion can be regarded as at least two distinct processes, though it is not clear when stage I finishes and stage II begins nor has any biological event been associated with this division. Most known tumour promoters, including TPA, have both stage I and II promoting activities, but mezerein and RPA act only as stage II promoters.

Irritant activity has long been associated with promoting activity as the most active and well known tumour promoters also possess considerable irritant activity [6]. For this reason, ID_{50} values have been used in some cases as a measure of promoting activity [6]. (The ID_{50} value is the dose required to produce reddening of the ears,

within 24 hours of application, of 50% of the animals, usually mice, exposed to the irritant.) Though the promoting activity of the teleocidins was discovered from association with their known irritant behaviour [17], it is questionable whether this is a valid method of assessing promoting activity. For it to be so, the irritant reaction must either be a consequence or precursor of tumour promotion, however work by Hecker *et al* has shown that Kraussianin (Gnidia factor K_{10}) which has a structure similar to that of TPA and ingenol, has a greater irritant activity than TPA but possesses minimal promoting activity [18].

1.2.2 Action of TPA-like Tumour Promoters

All three classes of TPA-like tumour promoters have been shown to bind to Protein Kinase C, discussed in the following section, with high affinity both *in vitro* and *in vivo* [19, 20]. Competitive binding experiments have shown that the tumour promoters bind to the same site as the endogenous activator, DAG, but with a much higher affinity [21, 22] (up to 10^4 greater), such that nM concentrations of tumour promoters are sufficient to completely activate the enzyme. Though inappropriate activation of PKC has been linked to the mechanism of TPA-like tumour promotion [23], subsequent stages in the pathway are not understood. For example, it is not known which of the many substrates of PKC is involved in promotion.

Though DAG activates PKC in exactly the same manner as the TPA-like tumour promoters it does not possess promoting activity itself. Within the cell membrane, DAG is metabolised far more rapidly than is TPA, so its effects are short lasting by comparison. There is evidence that this prolonged activation of PKC by TPA leads to a reduction in PKC levels due to enzymatic hydrolysis of the enzyme, and it is possible that the promoting activity stems from this down-regulation as it fits well with the required dose nature of promotion. Also physiological activation of the enzyme by DAG is always accompanied by calcium release as a result of phosphatidyl-inositol bisphosphate ($PtdInsP_2$) hydrolysis, which does not occur when TPA is the activator.

The two pathways normally have a regulatory effect on each other (see section 1.4) and the loss of this control may contribute to tumour promotion.

1.3 Protein Kinase C.

Protein Kinase C (PKC) was first characterised in 1977, by Nishizuka as a soluble kinase PKM [24]. Subsequently the precursor enzyme, PKC, was shown to be the physiologically active form. It consists of between 592 and 737 residues with a molecular mass between 68 and 83 kdal depending on the subspecies, the most common of which has 672 residues and a mass of 77 kdal. The enzyme possesses two distinct domains, one catalytic, the other regulatory [25] and is both soluble and membrane bound, though it can only be activated when bound to the surface of a lipid bilayer. The active form is vulnerable to proteolytic cleavage by the protease, *Calpain,* which cleaves the enzyme between the two domains. The regulatory domain remains bound to the lipid bilayer, but the catalytic domain becomes soluble and possesses kinase activity, without the need for activation, originally referred to as Protein Kinase M [24]. Activation of the enzyme results in its translocation from the cytosol to the membrane surface.

PKC acquired its name from the fact that it could be activated, *in vitro*, by calcium ions in a mM range [26, 27]. This level is too high to be of any importance physiologically, except in a very few situations: rather the enzyme is activated by binding diacylglycerol (DAG) when bound to the cell membrane. DAG is produced by the hydrolysis of phosphatidyl-inositol (4,5) bisphosphate ($PtdInsP_2$) by $PtdInsP_2$-specific Phospholipase C (also called phosphoinositidase), and in the presence of phosphatidyl-serine (PS), serves to increase the affinity of the enzyme for Ca^{2+}, so that the cytosolic concentration of 1μM becomes fully activating [28]. Hydrolysis occurs in response to receptor binding, the signal being transmitted via a G-protein (called G_p to distinguish it from those involved in the cyclic-AMP pathway). PKC is distinct from

the receptor/G-protein complex so the rate of diffusion and metabolism of DAG determines that the stimulatory effect is short lasting.

The nature of the phospholipid environment of PKC is critical to its activity in that there is an absolute requirement for PS, from which Ganong *et al* have suggested a specific PS-Ca^{2+}-DAG complex [29], but phosphatidyl-ethanolamine (PE) has an inhibitory effect.

PKC has been shown, *in vitro,* to phosphorylate a large number of proteins, and has been assigned several *in vivo* actions, including modulation of ion conductance, control of receptors, involvement in secretion and exocytosis, smooth muscle contraction, gene expression, cell proliferation and tumour promotion [30, 31, 32].

In the short term, PKC activation serves to decrease the Ca^{2+} concentration, both by activating Ca^{2+} transport systems, and providing negative feedback by inhibiting the receptor-mediated hydrolysis of $PtdInsP_2$. A rather more long-term effect of PKC comes from its phosphorylation of the EGF receptor, reducing its affinity of the growth factor, reducing ligand-stimulated tyrosine phosphorylation and so restricting cell growth. PKC has also been shown to have a stimulatory action on gene expression, inducing the interleukin-2 receptor, and has been shown to activate the cellular oncogenes c-myc and c-fos [30, 31].

Activation of PKC leads to a conformational change in the enzyme making it vulnerable to attack by a protease [24, 26]. This separates the catalytic domain from the regulatory / membrane binding domain, allowing it to dissolve in the cytosol. This soluble kinase activity was originally referred to as Protein Kinase M (PKM), until it was shown to be a domain of PKC. Solubilising the catalytic domain of PKC destroys its normal activity as it is removed from its usual (membrane bound) substrates, and is rapidly metabolised by the cytosolic proteases. Coupled with the short half-life of DAG, this serves to heavily control the normal activity of PKC.

Since it was first identified, several subspecies of PKC have been discovered, both from binding information and from the discovery of different functional mRNA's [33, 34]. The different forms of the enzyme possess a high degree of sequence

homology and share common structural motifs, yet differ in their mode of activation, sensitivity to Ca^{2+} and catalytic activity. Fig 1.1, below, shows the distribution of the regions of homology and variable regions in relation to the domain structure and the calcium binding site. In common with all membrane bound proteins, the enzyme is difficult to crystalise and no X-ray structure has been obtained, so little is known about its detailed structure. The calcium binding site can be located because of its distinctive sequence, but there is no such marker for either the catalytic or regulatory sites.

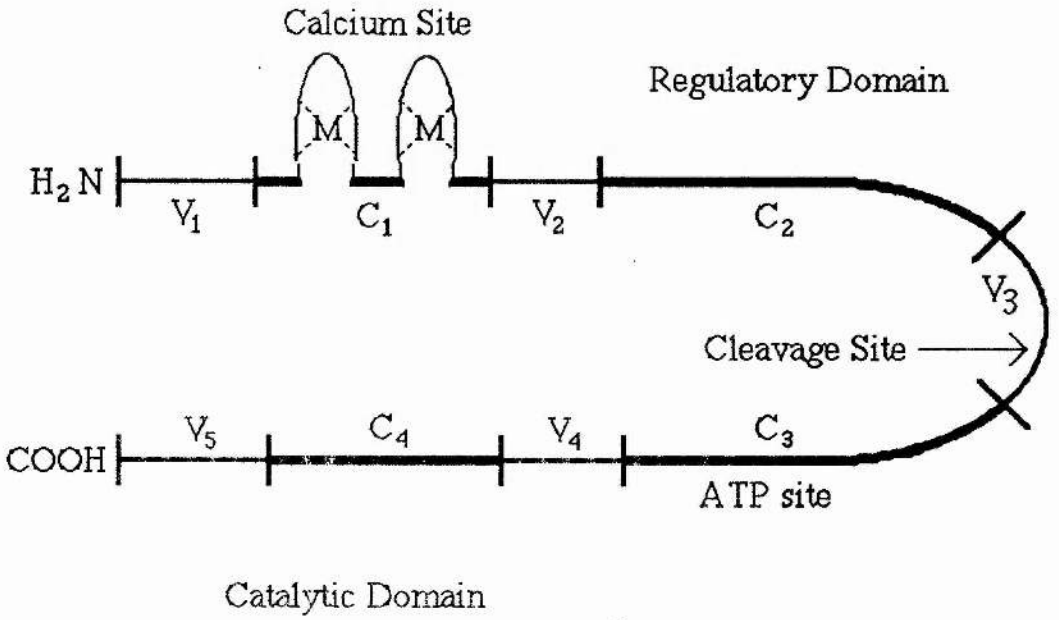


Fig 1.1 Domain Structure of PKC.

The regions labelled C_N show a large degree of sequence conservation between different subtypes, while those labelled V_N show considerable variation.

Northern blot analysis and radiolabelling with tritiated phorbol (12,13) dibutyrate have shown some tissue-specific expression of PKC subspecies and uneven intracellular distribution [33, 35, 36]. Subspecies α, the type first discovered, is found in all tissue types, but the γ subspecies is found only in the brain and spinal cord.

This heterogeneity is almost certainly responsible for the differential binding of phorbol esters shown by Blumberg *et al* [37], with the α subspecies providing the main binding though it is not clear which, or how many of the other subspecies act as the low affinity receptors. It would also be expected that the different subspecies would have different physiological functions, controlled either by their distribution, or by their activities, though as yet there is no biological evidence for this. Further investigation into both the binding of phorbol esters by different subspecies and their role in tumour promotion would be useful as it may be possible to block (or at least reduce) the promoting action if it is restricted to a few subspecies or cell-types. In the absence of this specific information, it is probable that the experimental data used in chapter 4 refers to the α subspecies.

1.4 Inositol Phosphates and Calcium Release

Receptor-mediated hydrolysis of $PtdInsP_2$ also produces *myo* -inositol (1,4,5) trisphosphate (IP_3), which acts as a secondary messenger leading to the release of stored calcium into the cytosol [38, 39], thus physiological activation of PKC is always accompanied by a release of intracellular calcium. These two pathways interact with each other to a considerable extent, and it may be the loss of this interaction that is responsible for the changes to cellular behaviour that result in tumour promotion.

An inositol 1,4,5 trisphosphate concentration of around 1μM is sufficient for half-maximal calcium release, though this varies slightly in different tissues [40]. The IP3 receptor has recently been identified as a 250 kdal glycoprotein with up to 2749 residues (from cDNA analysis) [41, 42]. Scatchard analysis suggests that the receptor has just a single binding site [43], and studies on reconstituted vesicles suggest that it is intimately connected to the calcium channel [42]. In addition to IP_3, the channel also requires potassium ions as the counter-ion and ATP (GTP or ADP can be substituted, but are less effective), though not, it appears, as a phosphate donor [40]. Hydrolysis occurs on the internal surface of the cell membrane and as calcium is released from the

endoplasmic reticulum, the process is limited by diffusion and metabolism of the messenger. Inositol 1,4,5 trisphosphate is rapidly metabolised, bringing about a swift cessation of the signal, though one of the metabolic products, inositol 1,3,4,5 tetrakisphosphate, IP_4, has been implicated in stimulating the entry of extra-cellular calcium, so prolonging calcium-dependant responses [44].

1.4.1 Interaction of Diacylglycerol and Inositol Phosphate Pathways.

In addition to having the same endogenous starting point, the two pathways interact with each other at several different stages. PKC activation enhances the phosphorylation of phosphatidyl-inositol, (PI), to $PtdInsP_2$, but this is contrasted against its inhibitory effect on the receptor / G-protein complex and its ability to stimulate the removal of calcium from the cell. These contrasting mechanisms may well work to cause both a rapid cessation of response and a swift recovery enabling the cell to respond to a further receptor signal. The inositol phosphate pathway does not have a documented *in vivo* effect on PKC, the changes in calcium levels brought about are not sufficient to activate PKC. Even in cells where PKC is activated solely by raised calcium levels (eg. the adrenal medulla), the calcium enters by voltage-sensitive ion channels [40].

1.5 Model Bilayers.

Both hydrolysis of $PstInsP_2$ and activation of PKC take place at the surface of biological membranes, where the nature of the environment is critical to the effects and activities of the molecules concerned. Simple pieces of information, such as partition coefficients, that may have an effect on activity, can be determined experimentally, but it is also important to consider effects such as orientation and accessibility at the membrane surface, and the nature of specific lipid-amphiphile complexes (such as that

suggested by Ganong *et al.* [29]). The complete cell membrane is far too complex a system on which to perform calculations, but an understanding of some of the effects of the environment at the lipid-water interface can be obtained by studying small model systems. At its simplest level, the interface can be modelled by a change in dielectric constant. While such a method would give information on the orientation of molecules within the bilayer for minimal calculation time, it ignores the contribution of specific intermolecular interactions to the conformation and orientation of molecules and so it rather limited in application. The use of Molecular Mechanics / Dynamics simulations including both the bilayer and solvent explicitly overcomes this deficiency but at a substantial cost in CPU time. With the advent of affordable mini-supercomputers such a calculation is now feasible.

1.6 Aims of This Research.

The computational work in this thesis has been split into four sections, all except the last explicitly concerned with a separate aspect of the $PtdInsP_2$ signal pathway, namely:

Chapter 4/	TPA-like Tumour Promoters.
Chapter 5/	Inositol Polyphosphates and Calcium Mobilisation.
Chapter 6/	Amphipathic Molecules in Model Bilayers.
Chapter 7/	Comparison of Charges and Energies obtained using Different Theoretical Methods.

The theory and packages used in the calculations are described chapters 2 and 3, while the latter also carries details of the programming work involved in producing the starting conformations of the model bilayers.

1.6.1 TPA-like Tumour Promoters.

Semi-empirical molecular orbital (MO) theory calculations were carried out on both active and inactive phorbol ester derivatives, giving structural information and, from the point charge approximation (see chapter 2), the electrostatic potentials (ESP). A model for the interactions involved in the binding of the tumour promoter to the regulatory site of the enzyme is developed from a comparison of the ESP's, taking structural differences into account. The structurally dissimilar ingenanes and teleocidins, that both possess comparable promoting activity, are used to confirm the main interactions involved in the site and to provide further detail of the description. The nature of the experimental data means that the description of the site is largely qualitative, but it should be possible, from a knowledge of the principle interactions involved, to identify other potential TPA-like tumour promoters, though it is not possible to provide a useful measure of their activity.

1.6.2 Inositol Polyphosphates.

The ability of *myo* -inositol polyphosphates to stimulate the release of intracellular calcium is entirely dependent on the number and position of the phosphate groups. As the molecules are all extremely flexible, MO structure calculations have been carried out on many conformations of the different physiological phosphates to determine the relative energies of the ring and phosphate conformations favoured by the different phosphates. From this, it may be possible to draw conclusions about the requirements of the binding site on the IP_3 receptor, though as the range of molecules and affinities considered is much less than was the case with the phorbol derivatives, it is not expected that the study will produce a particularly detailed description of the site. In addition, molecular dynamics simulations of *myo* -inositol 1,4,5 trisphosphate have been carried out, both in isolation and in a water bath. This should provide further

information into the relative stabilities of the different conformations of the molecule and will be compared with the semi-empirical results.

1.6.3 Amphipathic Molecules in Model Bilayers.

The TPA-like tumour promoters studied in chapter 4, act at the membrane-cytosol interface and are lipid soluble. Thus it is possible that the conformations found from gas-phase calculations on isolated molecules bear little resemblance to those important for their activity. To investigate this problem further, molecular dynamics simulations have been performed on solvated lipid bilayers containing two molecules of the tumour promoter (one in each monolayer). From such a calculation, the orientation and accessibility of the critical groups, suggested from the gas-phase calculations, can be examined as can the time-dependence of any specific complexes formed. Systems using both model and full size lipids have been examined.

In order to solvate the bilayers efficiently it is necessary to start with a close packed array of lipid molecules. A method has been developed which packs the molecules into a random two-dimensional array, so as to allow simulation of the fluid nature of the bilayer. This packing is achieved by minimising the solvent accessible hydrophobic molecular surface area, and enables efficient packing of a heterogeneous system.

1.6.4 Comparisons of Results from Different Theoretical Methods.

It was found from the calculations described in chapter 5 that the different MNDO-type parametrisations gave considerable differences in the heats of formation and phosphorus charges of the inositol phosphates, especially for the higher phosphates. As this type of simple organic phosphate is outside the range of molecules used for the initial parametrisations of the semi-empirical methods, results from both *ab initio* and semi-empirical calculations on *myo* -inositol 1' monophosphate and a three carbon fragment are compared. For both molecules, the restrictions on disk usage and processor time is such that only single point *ab initio* calculations are feasible, limiting the amount of data that can be compared. A further restriction is the lack of experimental data for this type of molecule, but it proved possible to draw some conclusions concerning the effectiveness of the different methods.

Chapter 2

Theoretical Methods.

2.0 Introduction

The calculations described in this thesis, with the exception of the molecular dynamics work discussed in chapters 5 and 6, are concerned with stationary states of molecules, rather than a time-dependent description, so the Time-Independent Schrödinger equation is appropriate.

$$\mathcal{H}\Psi_T = E_T \Psi_T \qquad (2.1)$$

where $\mathcal{H}$ is the Hamiltonian operator,

E is the energy of the system and

Ψ_T is the wavefunction.

As this thesis is concerned mainly with calculations of quite large molecules, semi-empirical calculations have been used for the most part, but some *ab initio* work has been done, essentially the comparative study presented in chapter 7. Both methods make use of the same basic theory, discussed below, and neglect relativistic effects. The semi-empirical MNDO [45] approximation is also discussed and a brief mention is made of some derived methods.

The mean value of the energy, E, is given by the expectation value of the Hamiltonian operator, $\mathcal{H}$.

$$< E_T > = \frac{\int \Psi_T^* \mathcal{H} \Psi_T \, d\tau}{\int \Psi_T^* \Psi_T \, d\tau} \qquad (2.2)$$

or, using the Dirac notation and assuming Ψ is normalised:

$$E_T = < \Psi_T | \mathcal{H} | \Psi_T >$$

The total Hamiltonian operator, $\mathcal{H}$, includes terms for the potential and kinetic energies of both the nuclei and electrons, in the form of operators. Thus, in atomic units[1], summing over all nuclei and electrons;

$$\mathcal{H} = \sum_{a>b} \frac{Z_a Z_b}{r_{ab}} - \sum_{a,i} \frac{Z_a}{r_{ai}} + \sum_{i>j} \frac{1}{r_{ij}} - \sum_{i} \frac{1}{2}\nabla_i^2 - \sum_{a} \frac{1}{2m_a}\nabla_a^2 \qquad (2.3)$$

where a and b refer to nuclei and i, j refer to electrons and the sums are over all electrons and nuclei.

[1] Atomic units used in eqn. 2.3 are:

Charge in units of electronic charge, e.

Mass in units of the mass of an electron, m.

Distance in Bohr Radii (0.529 Å).

Energy in Hartree.

2.1 The Born-Oppenheimer Approximation.

The kinetic energy of the nuclei is very much less than that of the electrons, on account of the large difference in mass, thus any change in nuclear position will be accompanied by a rearrangement of electron density to give the system with the lowest energy. The Born-Oppenheimer approximation [46, 49, 50] takes advantage of this by treating the electrons and nuclei separately. In quantum mechanical calculations, the kinetic energy of the nuclei is sufficiently small that it can be neglected and the nuclear positions are known. Thus the nuclear repulsion term in the Hamiltonian becomes a simple sum, so the energy can be obtained by adding this sum to the expectation value of the Electronic Hamiltonian, $\mathcal{H}_e$, for the electronic wavefunction, Ψ.

$$E = \sum_{a>b} \frac{Z_a Z_b}{r_{ab}} + \frac{\int \Psi^* \mathcal{H}_e \Psi \, d\tau}{\int \Psi^* \Psi \, d\tau} \qquad (2.4)$$

$$\mathcal{H}_e = \sum_{i>j} \frac{1}{r_{ij}} - \sum_{a,i} \frac{Z_a}{r_{ai}} - \sum_{i} \frac{1}{2} \nabla_i^2 \qquad (2.5)$$

The Born-Oppenheimer approximation is also used in Molecular Mechanics methods, where the electron density is assumed to be distributed to give the lowest energy, and so defined by the nuclear positions. The energy is then treated as a function of nuclear position alone, Thus in molecular mechanics the potential energy hypersurface is a Born-Oppenheimer surface.

The electronic Hamiltonian, $\mathcal{H}_e$, can be split rigourously into one and two-electron terms;

$$\mathcal{H}_e = H_1 + H_2 \qquad (2.6)$$

so that

$$E = \sum_{a>b} \frac{Z_a Z_b}{r_{ab}} + \langle \Psi \mid H_1 \mid \Psi \rangle + \langle \Psi \mid H_2 \mid \Psi \rangle \qquad (2.7)$$

The one electron term includes the electron-nuclear attraction and the kinetic energy, while the two-electron term contains the electron-electron repulsions.

$$H_1 = -\sum_{i} \left(\sum_{a} \frac{Z_a}{r_{ai}} + \frac{1}{2} \nabla_i^2 \right) .$$

$$H_2 = \sum_{i>j} \frac{1}{r_{ij}} \qquad (2.8)$$

2.2 Antisymmetry and the Slater Determinant.

The electronic wavefunction, Ψ, is a function of the coordinates, q_n, (including spin) of all the electrons;

$$\Psi\ (q_1, q_2, q_3, \ldots\ldots\ldots\ldots, q_N)$$

and is antisymmetric with regards to exchange of any two sets of coordinates for particles that possess spin, i.e.

$$\Psi(q_1, q_2, \ldots., q_N) = -\Psi(q_2, q_1, \ldots., q_N) \qquad (2.9)$$

The total electronic wavefunction, Ψ, can be conveniently constructed from a linear combination of the individual electronic wavefunctions (spin orbitals χ_i). The antisymmetry can be expressed mathematically if the electronic wavefunction is expanded in the form of a determinant.

$$\Psi(q_1, q_2, \ldots, q_N) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \chi_1(q_1) & \chi_2(q_1) & \ldots. & \chi_N(q_1) \\ \chi_1(q_2) & \chi_2(q_2) & \ldots. & \chi_N(q_2) \\ . & . & \ldots. & . \\ \chi_1(q_N) & \chi_2(q_N) & \ldots & \chi_N(q_N) \end{vmatrix}$$

(2.10)

where N is the total number of electrons and $(N!)^{-1/2}$ is a normalisation constant. If the orbitals are orthonormal, the overlap integrals $\langle\chi_i | \chi_j\rangle$ are zero, so greatly simplifying the equations. In addition, the orbitals used in the calculations are real so there is no need to use the complex conjugate as;

$$\chi^* = \chi$$

If any two rows or columns are exchanged, the sign but not the magnitude of the determinant is changed. If the orbitals are written sequentially as in eqn. 2.10, the determinant can also be described unambiguously by the diagonal elements:

$$|\chi_1(q_1)\ \chi_2(q_2)\ \cdots\ \chi_N(q_N)| \qquad (2.11)$$

The determinant is referred to as the Slater Determinant and describes a single electronic configuration. If the energy of only a single configuration, eg. the ground state, is required then single determinant methods are sufficient, as is the case of the work described here. However, such a method neglects correlation energy which can be included by using methods such as Configuration Interaction (CI) [46-48].

2.3 The Hartree-Fock Approach.

The Hartree-Fock method treats the electrons as independent particles moving in an average field generated by the nuclei and all the other electrons, thus a complicated many-electron problem is reduced to a series of one-electron problems. The derived Hartree-Fock equation is non-linear so must be solved iteratively in the Self-Consistent Field procedure [46, 49, 50].

For systems in which all the electrons are in pairs, the singly occupied spin orbitals, χ_i, can be replaced by doubly occupied spatial orbitals, ψ_i, adding a spin term to distinguish between the two electrons, thus:

$$\chi_{2i-1} = \psi_i\alpha = \psi_i$$

$$\chi_{2i} = \psi_i\beta = \overline{\psi}_i \qquad (2.12)$$

When substituted into eqn. 2.11, this gives

$$\Psi = \left| \psi_1\overline{\psi}_1, \psi_2\overline{\psi}_2, \ldots, \psi_{N/2}\overline{\psi}_{N/2} \right| \qquad (2.13)$$

In the Restricted Hartree-Fock approach, the electrons are constrained so that they remain in pairs, thus reducing the number of orbitals that must be considered in the final equations. This restriction is valid only if the system studied does not involve unpaired electrons at any stage.

Substituting eqn. 2.13 into the expression for the electronic energy, eqn. 2.7, with the assumption that the orbitals are real, gives the energy in terms of the spatial (molecular orbitals).

The energy due to the one-electron term is given by;

$$E_1 = 2\sum_{i}^{N/2} \int \psi_i H_1 \psi_i \, d\tau \qquad (2.14)$$

or;

$$E_1 = 2\sum_{i}^{N/2} I_i$$

where

$$I_i = \int \psi_i H_1 \psi_i \, d\tau$$

while the two-electron contribution contains two terms; the first, summed over all possible two-electron combinations, corresponds to the classical two-electron electrostatic (coulombic) repulsion; and the second, summed only over two-electron

combinations where both have the same spin, is a consequence of the antisymmetry. This latter term is known as the exchange repulsion and represents the repulsion preventing two electrons with the same spin from having the same spatial coordinates and it has no classical counterpart.

$$E_2 = \sum_{i,j}^{N/2} \left[2\int \psi_i\psi_j \, H_2 \, \psi_i\psi_j \, d\tau - \int \psi_i\psi_j \, H_2 \, \psi_j\psi_i \, d\tau \right]$$

where the orbitals are written in the order $\psi(1)\psi(2)$ throughout. The above equation can also be written as:

$$E_2 = \sum_{i,j}^{N/2} [2J_{ij} - K_{ij}] \qquad (2.15)$$

giving the electronic energy as:

$$E = 2\sum_{i} I_i + \sum_{i,j} [2J_{ij} - K_{ij}] \qquad (2.16)$$

for the Restricted Hartree-Fock formalism.

I_i is the one-electron operator, describing the kinetic energy of an electron and its electron-nuclear attraction. J_{ij} and K_{ij} are two-electron operators and describe the coulombic electron-electron repulsion and the exchange repulsion between electrons of like spin respectively.

The two-electron operators can be expanded as follows:

$$J_{ij} = \langle \psi_j \mid J_i \mid \psi_j \rangle \qquad (2.17)$$

$$J_i\psi_j = \left| \int \psi_i \, r_{12}^{-1} \, \psi_i \, d\tau \right| \psi_j \qquad (2.18)$$

$$K_{ij} = \langle \psi_i \mid K_i \mid \psi_j \rangle \qquad (2.19)$$

$$K_i\psi_j = \left| \int \psi_i \, r_{12}^{-1} \, \psi_j \, d\tau \right| \psi_i \qquad (2.20)$$

According to the variational principle [5i], the expectation value, E, of the Hamiltonian is an upper limit to the exact ground state energy, thus the lower the value obtained for E, the closer it will be to the correct energy. Thus the molecular orbitals are varied to give a minimum energy from eqn. 2.16 with the constraint the wavefunction remains normalised. By defining the Fock operator as

$$F_i = H_1 + \sum_{i}^{N/2} (2J_i - K_i) \qquad (2.21)$$

the eigenvalue equation becomes (in matrix form):

$$F_i\psi_i = \varepsilon_i\psi_i \qquad (2.22)$$

2.4 Linear Combination of Atomic Orbitals.

The spatial orbitals, $\psi(i)$, extend across the entire system and require an extremely complex description in order to be determined accurately, however, it is possible to choose a basis, $\{\phi\}$, such that any function, ψ_i, can be built up from a linear combination of elements of the basis, provided that the basis is complete, according to the formula:

$$\psi_i = \sum_j c_{ij} \phi_j(i) \qquad (2.23)$$

If the basis, $\{\phi\}$, is complete, the eqn. 2.23 is an exact expansion, and any complete basis may be used. However if the function is a molecular orbital, ψ_i, it is convenient to use the atomic orbitals, ϕ_j, as the basis set, hence the term, Linear Combination of Atomic Orbitals (LCAO).

The eigenvalue equation, eqn. 2.22, can be expanded in terms of the basis functions.

$$F_i \sum_j c_{ij} \phi_j = \varepsilon_i \sum_j c_{ij} \phi_j \qquad (2.24)$$

A complete expansion will give the best possible energy for a single-determinant method, a value known as the Hartree-Fock limit. In practice, though, the basis set may be far from complete, resulting in a higher energy than the Hartree-Fock energy (the Hartree-Fock limit). This error can be minimised by a sensible choice of basis and for geometries close to the minimum energy, the potential surface obtained from the basis tends to be parallel to that from a basis near the Hartree-Fock limit. Thus geometries and relative energies obtained from a good, but incomplete, basis are reliable even though the energy is some way above the correct value.

2.5 The Hall-Roothaan Equations.

As the exponents of the atomic orbitals, ϕ_i, are generally fixed, the problem of calculating the energies becomes one of determining the coefficients, c_{ij}. Multiplying eqn. 2.24 by ϕ_k and integrating gives :

$$\sum_j c_{ij} \langle \phi_k | F_i | \phi_j \rangle = \varepsilon_i \sum_j c_{ij} \langle \phi_k | \phi_j \rangle \qquad (2.25)$$

so defining the Overlap Matrix, S_{kj} :

$$S_{kj} = \langle \phi_k \mid \phi_j \rangle \qquad (2.26)$$

and the Fock Matrix, F_{kj} :

$$F_{kj} = \langle \phi_k \mid F_i \mid \phi_j \rangle \qquad (2.27)$$

substitution these into eqn. 2.24 gives;

$$\sum_j F_{kj} c_{ij} = \varepsilon_i \sum_j S_{kj} c_{ij} \qquad (2.28)$$

or in matrix form:

$$Fc = \varepsilon S c$$

where ε is a diagonal matrix of the orbital energies, and c a square matrix of the coefficients, c_{ij}. Equation 2.28 (for the Restricted Hartree-Fock procedure) is known as the Hall-Roothaan Equation [52,53].

2.6 The Density Matrix

The Fock Operator terms include a one-electron Hamiltonian, describing the electron-core attraction and the electronic kinetic energy, and the Hartree-Fock potential which is also a one-electron term representing the average electronic field experienced by a single electron. The latter term is thus determined by the electron density, ρ, and it is therefore convenient to express the Fock Matrix elements in terms of the electron density.

The spatial wave-function, $\psi_a(r)$, is related to the charge density, ρ, due to a single electron at a point, r, by:

$$\rho = \int |\psi_a(r)|^2 \, dr \qquad (2.29)$$

Thus for a closed shell system of N electrons, the total charge density, $\rho(r)$, at position, r, is:

$$\rho(r) = 2\sum_{a}^{N/2} |\psi_a(r)|^2 \qquad (2.30)$$

Integrating gives N, the number of electrons.

By expanding in terms of basis functions, ϕ, the Density Matrix (or Charge-density Bond-order Matrix), P_{ij}, is defined.

$$\rho(r) = \sum_{ij} P_{ij}\phi_i\phi_j \qquad (2.31)$$

$$P_{ij} = 2\sum_{a}^{N/2} c_{ia}c_{ja} \qquad (2.32)$$

The Fock Matrix can thus be written in terms of the density matrix, P_{ij} :

$$F_{ij} = H_{ij} + \sum_{k,l} P_{kl}\left[(ijlkl) - \tfrac{1}{2}(illkj)\right] \qquad (2.33)$$

with;

$$(ijlkl) = \int \phi_i \phi_j \, r_{12}^{-1} \, \phi_k \phi_l \; d\tau$$

and,

$$H_{ij} = \int \phi_i \, H_1 \, \phi_j \; d\tau$$

2.7 Self-consistent Field Approach

For a given geometry, H_{ij} is fixed and so need only be calculated once. However, the two electron part is dependent on the density matrix, itself dependent on the orbital coefficients, making the Hall-Roothaan equations (eqn. 2.28) non-linear with respect to those coefficients i.e.

$$F(c)c = \varepsilon S c \qquad (2.34)$$

Such an equation requires an iterative approach, with repeated solving of the Hall-Roothaan equation (eqn. 2.28)) and evaluation of the Density and Fock Matrices (eqn. 2.32 and eqn. 2.33).

In detail, this procedure is as follows:

1/ The nuclear coordinates, R, number of electrons, N and basis set, $\{\phi\}$, are specified.

2/ All required integrals, S_{ij}, H_{ij}, and (ijlkl) are calculated.

3/ An initial guess is made at the Density Matrix, P_{ij}.

4/ The Fock Matrix is calculated from the integrals H_{ij} and (ij|kl), and from the Density Matrix, using eqn.2.33.

5/ The expansion coefficients, c_{ij}, and orbital energies, ε_i, are then calculated from the Fock Matrix.according to eqn. 2.28.

6/ A new Density Matrix , P_{ij},is then calculated from the coefficients, using eqn.2.32.

7/ The resulting Density Matrix is compared with the previous one, and if it is within specified convergence criteria, the procedure is halted and the energy of the system calculated. At this stage any desired one-electron properties may also be determined from the wavefunction and the appropriate operator (see section 2.11)If the criteria are not reached, steps 4 through to 7 are repeated until it has.

This procedure is known as the Self-Consistent Field (SCF) approach.

Unfortunately, the procedure is not quite so simple as the basis functions are not usually orthogonal to each other, so solving the Hall-Roothaan equation (eqn. 2.28) is not an eigenvalue problem, rather the basis must be orthogonalised. Thus the Overlap Matrix, S, needs to be diagonalised (to matrix, s) and orthogonalised by a transformation matrix, X, eg;

$$X = Us^{-1/2} \qquad (2.35)$$

where U is the unitary matrix.

This transformation matrix must also act on the Fock Matrix and coefficient matrix to leave the Hall-Roothaan equations as eigenvalue problems, thus:

$$F = XF' \qquad \text{and} \qquad c = Xc'$$

and so

$$F'c' = c'\varepsilon \tag{2.36}$$

The initial matrices, F and c, are used in the Density Matrix, so they must be retransformed every cycle of the SCF.

2.8 Atomic Orbitals and Basis Sets.

As stated earlier, a complete basis will give an exact expansion of the molecular orbitals, but is not obtainable as such a basis is infinite. It is necessary therefore to choose a basis which gives results that are an acceptable compromise between accuracy and use of available resources.

Slater functions, with an orbital exponent of $-\zeta r$ (r is the distance radially from the nucleus), were developed from hydrogen atomic orbitals and closely model atomic orbitals for other systems. However, they are not generally used as there are $K^4/8$ two electron integrals, (ij|kl) (K is the number of basis functions), which are extremely time consuming given the form of the Slater-type orbitals. Instead, gaussian-type functions are generally used (with an orbital exponent of $-\alpha r^2$), to make the integral calculation simpler as the product of two gaussians is a third gaussian centred between them. It must be noted that gaussian functions diverge quite seriously from the actual form of the orbitals, in particular the use of an r^2 exponent leads to a concentration of electron density close to the nucleus , and at the nucleus itself, the gradient of the electron density is zero, whereas, for s orbitals, the electron density should form a cusp. A better fit to the Slater-type function can be obtained by using more than one gaussian to replace each Slater-type orbital, though the problem of the gradient at the nucleus remains

Basis sets that include a single term for each orbital are termed minimal (or single zeta (ζ), as they have only one zeta exponent per orbital) basis sets, an example

being the STO-nG basis sets of Pople [54] which use a single Slater-type orbital to express each atomic orbital, χ_i, and build up the Slater orbital using n gaussians. The most commonly used basis set of this type, the STO-3G basis set, is frequently described as sub-minimal as it gives a worse energy than a minimal Slater basis set. At the other end of the scale are the double and triple zeta basis sets, which use 2 and 3 functions to represent each orbital, and example being the double zeta (9s,5p) Dunning / Huzinaga basis set [55], termed d95 in the GAUSSIAN programs [56]. The split valence basis sets of Pople, eg. 6-31G [57], have also been included in the GAUSSIAN programs, and are commonly used. They form a minimal basis for the core electrons, but are of double zeta quality for the valence electrons.

There is no limit to the number of functions that can be used to represent each orbital, though adding more zeta functions without adding functions of higher angular momentum can distort the results, for example, using a large number of just s and p functions for ammonia produces a planar structure. If a larger basis is required, it is preferable to add functions of higher angular momentum, known as polarisation functions, i.e. d and higher functions are added to first-row atoms and p functions of hydrogen, thus a Pople STO-3G basis with polarisation functions on first-row atoms is termed STO-3G* [58] (a second * is added if polarisation functions on H are added).

The number of gaussian functions can rise rapidly, offsetting much of the speed-up of calculation, so it has become common to use contracted gaussian basis sets rather than all of the gaussian primitives. The process of contraction regards each basis function as a fixed linear combination of gaussian primitives. The contraction coefficients for the function must be chosen so that the form of the function remains consistent with that of the original primitives. This means that a contracted function made up only of functions of s symmetry must itself be of s symmetry. An example of such a contraction is found with the d95 basis of Huzinaga, which in its original form consisted of 9s and 5p gaussian functions optimised for the first row atoms, which has been subsequently contracted by Dunning [55]. This particular method differs from those of Pople in that the contracted coefficients are energy optimised for

atomic calculations whereas Pople uses a least squares fit to Slater-type orbitals. In addition to contraction the Pople basis sets (6-31G, 3-21G etc.) save on computational costs by constraining the p function exponents to be the same as those of the s functions [54].

2.9 Limitations of the Hartree-Fock / SCF Method

As mentioned earlier, in order to reach the Hartree-Fock limit of the solution, a complete basis set is needed and though it is possible to come close to this, the computational effort required limits such calculations to systems containing only a small number of atoms. In addition, the methods used here are based on a single determinant and so do not take into account configurations other than the ground state. By failing to consider the excited states, the energy obtained is greater than the experimental value. Central to the Hartree-Fock method is the assumption that the electrons can be regarded as independent particles moving in a field generated by the other electrons. As this field is treated as an average field, there is no consideration of electron correlation though inclusion of electronic excitation in methods such as CI [46-48] corrects this deficiency.

The RHF approach, where molecular (spatial) orbitals are doubly filled with electrons of opposite spin, only applies to systems in which all the electrons are paired throughout the calculation. Thus it is not applicable to systems in which bonds are broken or orbitals are single occupied, such as radicals and excited states.

2.10 Semi Empirical Methods.

The *ab initio* methods, in which no further approximations are made, are extremely demanding on computer resources. The number of two-electron integrals required increases with the number of basis functions, N, at a rate of N^4. The recent releases in the GAUSSIAN series of programs make use of the Direct SCF approach,

in which the two-electron integrals are calculated only when they are needed and thus do not need to be stored on disk and through efficient calculation the number is proportional only to $N^{2.7}$, but even with this improvement, the processor time required for large systems (>20 heavy atoms) is prohibitive.

One solution has been to replace calculated two-electron integrals with experimentally determined parametric equations, thus the methods have been termed "Semi-Empirical". The parametric equations have been derived from several different sources, some including Pople's INDO [59] and CNDO [60] have been developed to reproduce *ab initio* results, while others such as AAMOM (also known as ZINDO) of Zerner [61] are based on spectroscopic data. The methods used in the work presented here have been developed from the NDDO methods of Dewar and Pople [60], which were parametrised to give accurate heats of formation and molecular geometries for small molecules.

2.10.1 MNDO and other NDDO methods

The semi-empirical calculations in the thesis have used the MNDO (modified Neglect of Diatomic Overlap) [45] method of Dewar, developed from Pople's NDDO (Neglect of Diatomic Differential Overlap) [60] method and made use of ideas and concepts used in INDO (Intermediate Neglect of Differential Overlap) [59].

The valence and core electrons are treated rather differently. The non-valence electrons are treated as part of the nuclear core in that the nuclear charge is reduced by the number of non-valence electrons. In addition, exchange repulsion integrals between the valence and core electrons are neglected and core-core exchange integrals are constant for a given core separation, thus need only to be calculated once and not as part of the SCF. This is collectively known as the "Core Approximation".

The valence electrons are described by minimal basis atomic orbitals, retaining all the valence-valence two-electron integrals except for those involving differential overlap on two different centres (so called Diatomic Differential Overlap). This leads to

a limited orthogonalisation of the orbitals, and the Hall-Roothaan equations (eqn. 2.28) are now modified to be of the form :

$$\sum_{\upsilon}(F_{\mu\upsilon} - \varepsilon_i\delta_{\mu\upsilon})\, c_{\upsilon i} = 0 \qquad (2.37)$$

where δ is the Kronecker Delta and equals one if $\mu=\nu$, zero if μ is not equal to ν. The Fock matrix elements are given by:

$$F_{\mu\mu} = U_{\mu\mu} + \sum_{b} V_{\mu\mu,b} + \sum_{\upsilon}^{a} P_{\upsilon\upsilon}\left[(\mu\mu,\upsilon\upsilon) - \tfrac{1}{2}(\mu\upsilon,\mu\upsilon)\right] + \sum_{b}\sum_{\lambda\sigma}^{b} P_{\lambda\sigma}(\mu\mu,\lambda\sigma) \qquad (2.38)$$

$$F_{\mu\upsilon} = \sum_{b} V_{\mu\upsilon,b} + \tfrac{1}{2}P_{\mu\upsilon}\left[3(\mu\upsilon,\mu\upsilon) - (\mu\mu,\upsilon\upsilon)\right] + \sum_{b}\sum_{\lambda\sigma}^{b} P_{\lambda\sigma}(\mu\upsilon,\lambda\sigma) \qquad (2.39)$$

$$F_{\mu\lambda} = \beta_{\mu\lambda} - \tfrac{1}{2}\sum_{\upsilon}^{a}\sum_{\sigma}^{b} P_{\upsilon\sigma}(\mu\upsilon,\lambda\sigma) \qquad (2.40)$$

with

μ and υ are orbitals centred on atom a.

λ and σ are centred on atom b.

The Fock matrix elements are then expressed in terms of the following parameters:

$U_{\mu\mu}$ One centre, one electron integrals.

$g_{\mu\upsilon}$ One centre, two electron coulomb integrals - $(\mu\mu,\upsilon\upsilon)$

$h_{\mu\upsilon}$ One centre, two electron exchange integrals - $(\mu\upsilon,\mu\upsilon)$

ζ Slater exponents for the orbitals

$\beta_{\mu\lambda}$ One electron resonance integrals, related to $S_{\mu\lambda}$.

α Exponent in an empirical electrostatic term.

The two centre, two electron integrals $(\mu\upsilon, \lambda\sigma)$, are expanded in terms of multipole-multipole interactions, while the two centre, one electron attractions between an electron centred on atom, a, for atom, b, $V_{\mu\upsilon,b}$, are treated as two centre, two electron integrals $(\mu\upsilon, ss)$.

While the Semi-Empirical methods constitute a further approximation of quantum mechanics from *ab initio* theory, and represent the valence electrons by only a minimal basis, they can give good quality results for molecules related to the system used in parametrisation. MNDO derived methods frequently give better results than minimal and sub-minimal basis *ab initio* calculations (eg those using the STO-3G basis set) as errors arising from some of the *ab initio* approximations are partially corrected by the parametrisation. Thus the effects of electron correlation are included within the parametrisation, but require the use of CI or similar methods in *ab initio* calculations.

In addition to MNDO the following Hamiltonians and parameter sets have been used:

AM1 (Austin Model 1) [62], which introduces Gaussian functions for the core-core interactions to correct the excessive long-range repulsion in the MNDO core-core function.

PM3 (Parametric Method 3) [63, 64], which was developed by Stewart and is a reparametrisation of MNDO, with an extra nine parameters, which give better results for the long range core-core repulsions by a similar method to that used for AM1.

2.10.2 Limitations of the MNDO Derived Methods.

The MNDO derived semi-empirical methods are amongst the most recent of the semi-empirical parametrisations and are generally held to be amongst the most accurate in determining molecular geometries and heats of formation. However they do have some significant drawbacks which are described below.

MNDO tends to underestimate the stability of sterically crowded molecules and those containing hypervalent atoms, while overestimating the stability of the four-membered ring. In addition, the hydrogen-bond is poorly represented.

Both the AM1 and PM3 parametrisations have improved the accuracy of the heats of formation (PM3 to a greater degree), and reproduce the experimental value for the energy of the hydrogen bond. The improvements are particularly good for hypervalent phosphorus. Against this, AM1 has a systematic error for alkyl groups as the heat of formation of the CH_2 fragment is around 2 kcal $mole^{-1}$ too low. By comparison, PM3 is a very recent development and though a large range of molecules were used in the original testing and parametrisation, its suitability for systems outside the initial range has been tested for only a minimal range of molecules (e.g. [65]).

2.11 Molecular Properties.

Once the wavefunction has been obtained, it is possible, in principle, to calculate any observable property, O, of a molecule by using the correct operator, $\mathcal{O}$, as the solution to an eigenvalue problem:

$$\mathcal{O}\Psi = O\Psi$$

where the property is given by the expectation value of the operator, $\mathcal{O}$. One such property is the energy of the molecule, obtained from the Hamiltonian operator as already shown.

In the case of one-electron properties, such as the electrostatic potential, the expectation value of the one-electron operator can be calculated from the Density Matrix and the easily calculated one-electron integrals.

$$< O > = \sum_{j,k} P_{jk} \int \chi_j \, O \, \chi_k \qquad (2.41)$$

which are calculated as part of the SCF calculation. For this reason many one-electron properties are calculated by default in the *ab-initio* and semi-empirical packages.

2.11.1 Geometry Optimisation.

The energy of a particular molecular conformation is a two-electron property of the wavefunction, the operator being the Hamiltonian Operator, $\mathcal{H}$, and though the minimum energy conformation is related, it is not an observable as there is no operator which will give that conformation directly from the wavefunction. However, the minimum energy conformation can be obtained from a consideration of the forces acting on an isolated molecule. If the forces acting on the molecule are reduced to zero, that molecule will be in a minimum-energy conformation. The method of Pulay [66], allows the forces exerted on the molecule to be determined analytically, and are given by the negative of the first derivative of the energy expression with respect to the nuclear coordinates. The geometry is altered, using one of several minimisation techniques, to reduce these derivatives, leading to a stationary point on the potential energy surface. In order to determine the nature of this stationary point, the second derivatives are also required; for a minimum, they must all be positive, but for a transition state, one of the second derivatives will be negative, which can be utilised in locating such structures.

The gradient methods fall into two basis classes, Steepest Descent Method and Conjugate Gradients Method, both of which use a first-order Taylor expansion of the

energy term:

$$E(x+\Delta x) = E(x) + g(\Delta x) \qquad (2.42)$$

Steepest descent methods take advantage of the fact that the greatest change in E (ΔE) is obtained when Δx is in the direction of -g (the gradient), so Δx is obtained thus:

$$x_k = x_{k-1} + \lambda_k g_k \qquad (2.43)$$

where λ_k is the step size, which is set initially and then reduced if at any stage, the energy increases. This method produces a rapid reduction in the value of a large gradient, but converges very poorly. The Conjugate Gradient method converges much more rapidly as it is able to locate the minimum energy along a given displacement vector, thus for an n-dimensional surface, it will converge in a maximum of n steps.

The BERNY [67] algorithm of Schlegel has been used as the default in the GAUSSIAN [56] package. First derivatives are calculated directly and though use is also made of the second derivative (Hessian) matrix, an initial estimate is required. The Hessian can then be updated at each step in which case the methods behaves as a conjugate gradient method, but if the magnitude of the gradients are large it may be preferable not to update the Hessian until the gradients have dropped below a suitable threshold, in which case the method acts as a steepest descent minimisation.

2.11.2 Electrostatic Potential Maps.

The electrostatic potential at a point is a physical observable, measurable by electron scattering experiments,and is defined as the force exerted on a unit positive charge at that point [68]. As such it is a one electron property and so can be calculated from the wave-function.

The electron density is given in eqn. 2.29, which when combined with the

nuclear charge gives the net charge on atom, a, as :

$$q_a = Z_a - \int \rho(r)\ dr \qquad (2.44)$$

so, by definition, the electrostatic potential (ESP) at point, r, is:

$$V(r) = \sum_a \frac{Z_a}{|R_a - r|} - \int \frac{\rho(r')\ dr'}{|r' - r|} \qquad (2.45)$$

The magnitude of the ESP at a given point clearly depends on its position in relation to the molecule in question, thus it is useful to calculate the ESP over a plane containing the molecule (particularly if the molecule is planar) or over the Connolly surface [69, 70] of the molecule. When the ESP is plotted onto the surface, features of the ESP can be associated with specific regions of the molecule, while plotting the ESP onto a plane gives an indication of the direction of forces acting on an approaching molecule.

The calculation of the ESP from the wavefunction is a very time-consuming process as the wavefunction must be calculated at each point for which the ESP is to be plotted. If this ESP is plotted onto a Connolly surface of the molecule, a point density between 1 and 5 points $Å^{-2}$ may be required, though rather less if the ESP is plotted over a grid. Thus for a large system, it is necessary to find an approximate means of calculating the ESP. To this end, studies have been carried out calculating the ESP with the atoms represented as point charges, with little apparent effect on the ESP outside the Van der Waals radius [71]. Thus this method has been used to save on computer time. In this case;

$$V(r) = \sum_a \frac{Z_a'}{|R_a - r|} \qquad (2.46)$$

where Z_a' takes into account the electron density around atom, a.

A pictorial representation of the Molecular Electrostatic Potential (MEP) can yield considerable qualitative information about the reactivity of that molecule. As the MEP represents the force on a unit-positive point-charge, regions of positive ESP are likely to be subject to electrophilic attack, while the reverse is true for nucleophilic attack. The MEP can also give insights into the binding of small molecules to their receptors in that regions of appreciable ESP will bind to complementary regions within the receptor binding site. As ligand-receptor binding is generally a non-bonding process, such electrostatic interactions are the main energetic contribution to binding. Thus if the structure of the binding site on the receptor is known, it should be possible to design ligands of high affinity, conversely, a possible structure of a binding site can be built up from comparisons of ESP from ligands of known affinity. Whether it is obtained from the wavefunction of from point charges, the ESP does show varying degrees of basis set dependency, so any attempts to draw qualitative conclusions from the MEP must be done with caution.

2.11.3 Similarity Indices.

The idea of a quantitative measure of similarity was developed by Carbo [72] and expressed in terms of electron density overlap between two molecules A and B, such that:

$$R_{AB} = \frac{\int \rho_A \rho_B \, dv}{(\int \rho_A^2 \, dv)^{1/2} \, (\int \rho_B^2 \, dv)^{1/2}} \qquad (2.47)$$

where ρ_A and ρ_B are the electron densities of atoms A and B. This gives a range of similarity from 0 to 1 (dissimilar to identical), though only compares the shape of the charge distribution not its magnitude. Thus if $\rho_A = n\rho_B$ the Carbo similarity

index remains unity. The magnitude of the electron density is taken into account in the alternative description of Hodgkin [73]:

$$H_{AB} = \frac{2\int \rho_A \rho_B \, dv}{\int \rho_A^2 \, dv + \int \rho_B^2 \, dv} \qquad (2.48)$$

This index also ranges from zero (dissimilar) to unity (identical), but for $\rho_A = n\rho_B$, $H_{AB} = 2n / (1+n^2)$.

Calculation of similarity indices is not restricted to electron density, but can also be performed using the electrostatic potential or electrostatic field of the molecules. The only difference in calculating similarities from these measures rather than electron density is that they are signed so the similarity index runs from -1 to +1, a negative index indicating that the magnitude of the field or potential is similar, but that the sign is reversed. In practice, both the field and potential are calculated using the point charge approximation described in section 2.11.2, in which case the field is given by:

$$F(r) = \sum_j \frac{q_j (r - r_j)}{|r - r_j|^3} \qquad (2.49)$$

The quantities can then be calculated over a grid though it is necessary to exclude the volume occupied by the molecules themselves as the approximation is not valid inside the Van der Waals radius. It is then a simple matter to sum the similarities over the grid point to provide a global value for the similarity index.

2.11.4 Atomic Charges.

Atomic charge is not a physical observable even though its effects can be seen in the electrostatic potential, and the susceptibility of parts of a molecule to nucleophilic or electrophilic attack. The atomic charges can be calculated to reproduce the molecular electrostatic potential at a specified surface, usually a Connolly surface set at a multiple of the Van der Waals radius, or can be calculated via Population Analysis. Calculating atomic charges from the electrostatic potential, ESP, should give the most accurate charges, but is computationally expensive. The method requires that the ESP be calculated at many points and the charges are obtained by a least squares fit of the point charge ESP to the wavefunction ESP. The charges so obtained are largely independent of the positioning of the surface points, provided they are outside the Van der Waals radius and that there are enough of them. The semi-empirical wavefunction used in the MNDO-like methods is not suitable for directly calculating the ESP, but by the technique of deorthogonalising the orbitals, ESP's close to those of *ab initio* calculations can be obtained [74].

Clearly this method is of no use if the reason for using point charges is to reduce the time required for calculation of the MEP. In these circumstances, Mulliken Population Analysis [75] gives a very rapid means of calculation atomic charges. There is no unique definition of how many electrons are associated with each atom in a molecule, but the total number of electrons can be obtained by integrating the total charge density, (eqn. 2.30). i.e.

$$N = 2\sum_{a}^{N/2} \int |\psi_a(r)|^2 \qquad (2.50)$$

which separates the total number of electrons into groups of two per molecular orbital. By substituting in the LCAO basis, the total electrons can be expressed in terms of the Density and Overlap Matrices:

$$N = \sum_{\mu,\nu} P_{\mu\nu} S_{\mu\lambda} \qquad (2.51)$$

or,

$$N = \sum_{\mu} (PS)_{\mu\mu}$$

$(PS)_{\mu\mu}$ can be interpreted as the number of electrons associated with the atomic orbital, ϕ_μ, and so its charge is added to that of the atom on which it is centred. The charge on a particular atom is given by:

$$q_a = Z_a - \sum_{\mu \in a} (PS)_{\mu\mu} \qquad (2.52)$$

The electron density within the overlap region is split equally between the two atoms. The charges obtained by this method are highly dependent on the basis set, in particular, the more functions added to any one atom, the more negative will be its charge regardless of the underlying chemistry of the situation. It is therefore important that the basis set used to obtain atomic charges is balanced between the different atoms of the system to avoid distortion of the charges.

2.11.5 Dipole Moments.

The dipole moment, μ, of a molecule is a one-electron property and is a measure of the turning force produced on the molecule by a uniform electric field and can be determined experimentally from the Stark effect [76]. For the point charge approximation, the dipole moment can be given by:

$$\mu = \sum_{m} q_m x_m \qquad (2.53)$$

where q_m is the net charge of atom m and x_m is the displacement of the atom from an origin. For a neutral molecule, the dipole moment is independent of the location of the origin.. It is more usual to obtain the electronic dipole moment from the wavefunction, summing the nuclear term separately:

$$\mu = e\sum Z_a x_a - e\int r_i |\Psi|^2 \qquad (2.54)$$

The dipole moment can thus be obtained from both *ab initio* and semi-empirical calculations though as experimental dipole moments were included in the MNDO-type parametrisations [45, 62-64] they may be expected to give the more accurate results.

Biologically, dipole moments are significant as they will contribute to the binding energy of a ligand in its polar binding site, while the direction of the dipole will determine, to some extent, the orientation of the ligand as it approaches the binding site, thus both the magnitude and the direction of the dipole moment would be expected to feature in structure / activity relationships.

2.12 Molecular Mechanics

The approximations introduced for the semi-empirical methods can give dramatic savings in terms of computer resources, but for calculations on very large systems (eg >1000 atoms as is the case for a protein or solvated system), it is necessary to change to a completely different method. Molecular mechanics is a fully empirical method, as it uses parameters to reproduce experimental measurements in small molecules without recourse to any quantum mechanical theory. There is no explicit consideration of the electrons and the wavefunction is not calculated [77].

2.12.1 The Molecular Mechanics Force-Field.

Virtually all molecular mechanics force-fields in use today are of the same form, including terms for bond stretching and bending, torsional terms and several non-bonding terms. Some force-fields do not consider all of these, for example ECEPP [78] does not possess bending or stretching terms, rather these variables are fixed at values obtained from crystal studies. In addition cross terms may be included to represent the effects that changing one internal coordinate has on another, these terms are typically bend-stretch, bend-bend and stretch-stretch cross-terms.

It is not possible to model the potential curve for bond stretching or bending over a full range with a simple function, but close to the equilibrium values, the potential is very close to a simple parabolic curve of the form :

$$V = \frac{k}{2}(r_0 - r)^2 \qquad (2.55)$$

where k, is a fixed constant and r_0 the equilibrium value of the parameter. This constraint is acceptable for geometries close to a local energy minimum, but leads to inaccurate energies for highly strained or transition-state geometries. A better approximation for distorted geometries is to include a cubic and / or quartic term as is used in MM2 / MM3 [79, 80]. These extra terms can improve the accuracy of the function at distances (angles) slightly larger than the equilibrium values over that from a quadratic term alone, but produce further complications at excessive values. By contrast, a quadratic term will always give the correct minimum energy structure (within the limitations of the parametrisation), but is not able to reproduce energies of conformations that deviate much from this geometry.

Dihedral rotations can be represented by a cosine series :

$$V = a_1\cos(\phi+\alpha_1) + a_2\cos(2\phi+\alpha_2) + \ldots \qquad (2.56)$$

a_i, is the amplitude of the i'th term, and α_i, is the phase of rotation.

The series is infinite, but amplitudes, a_i, rapidly become very small, so that at most 3 terms are required to adequately represent the rotation potential. Some force fields use a slightly modified cosine series for the rotation potential, but the form of the curve is the same.

The non-bonding forces are represented by a Van der Waals (6-12) term, an electrostatic term, and, where applicable, a hydrogen-bonding (10-12) term.

$$V_{VdW} = \varepsilon \left(a(r_0/r)^{12} - b(r_0/r)^{6} \right) \qquad (2.57)$$

$$V_{El} = \frac{q_1 q_2}{4\pi\varepsilon_r r} \qquad (2.58)$$

$$V_{HB} = \varepsilon \left(c(r_0/r)^{12} - d(r_0/r)^{10} \right) \qquad (2.59)$$

ε is the value of the potential at the equilibrium distance r_0, q refers to the atomic charge and ε_r is the dielectric constant of the medium.

The concept of parametrisation requires that atoms of the same element behave similarly in a wide range of differing local environments, eg. Hydrogen bonded to oxygen is treated similarly whether the oxygen is part of a sugar molecule or a peptide. Normally atoms are allowed a limited number of distinct types. In the case of AMBER [81, 82], 5 or 6 per element is typical, though carbon has nearer twenty and phosphorus just one. Adding more parameters through extra atom types makes the calculation more complicated and increases the computational cost while improving the accuracy and range, so the number of different atom types used is always a compromise between these two facets. Even with the number used within the AMBER force-field, there are a large number of parameters, but the forces are very easy to calculate from the parameters so the calculations remain rapid. Computational efficiency can be further improved by the use of the United-Atom force field, which includes the hydrogen atoms bonded to carbon as part of the carbon atom. Thus

carbons now have increased mass and Van der Waals radius to reflect the number of bonded hydrogen atoms. This approximation is not entirely accurate in all respects as it gives the combined atoms a spherical Connolly surface, and can allow racemisation of chiral carbons with a single hydrogen, though for calculations when specific interactions of the carbon and hydrogen atoms are not important, this approximation can dramatically reduce the size of the problem. One such situation is the study of model bilayers in chapter 6.

2.12.2 Determining Parameters.

The parameters can be obtained directly from experimental observation - equilibrium values of bond lengths and angles from X-ray crystallography, force constants from I.R. spectra. However in many cases it is not practicable to get the parameters in this way and it is much easier to obtain them from *ab initio* or semi-empirical calculation on small molecules. If semi-empirical calculations are used, (or any method for torsion angles), this is done by fitting the changes in energy from small deformations of the coordinate under consideration, to a parabolic curve of the correct form. The effects of the non-bonding forces must be taken into account, as these will vary as bonds and angles etc. are changed. The reason for this is that MOPAC force calculations do not assign the force constants solely to a vibration of a single internal coordinate, but rather calculate vibrations over the whole molecule and give their associated I.R. frequencies. Force calculations using the GAUSSIAN series of programs do give the 2nd derivatives for the internal coordinates, so the force constants for bond stretching and bending can be obtained directly. To get accurate force constants however, a basis set of at least the level of 6-31G* is required, and in most cases gives comparable force constants to those obtained from small perturbations of semi-empirical structures using a quadratic fit.

2.12.3 Molecular Dynamics.

Energy minimisation can be useful for determining molecular structure, but it does not give any information concerning the behaviour of the molecule over a period of time though this information can be readily obtained from Molecular Dynamics calculations, which use the molecular mechanics force fields. The method works by assigning a position and velocity to each particle and calculating the force using the molecular mechanics force-field, which in turn is used to update the position and velocity.

For an infinitely small time step, Δt,

$$x_i(t + \Delta t) = x_i(t) + v_i(t)\Delta t \qquad (2.60)$$

$$v_i(t + \Delta t) = v_i(t) + \frac{f_i}{m_i}(t)\Delta t \qquad (2.61)$$

These equations assume that the force, f_i, is constant throughout the time step, an approximation that becomes less valid as Δt is increased. To overcome this several different methods have been used to update the position and velocity with increasing complexity. One of the more simpler, and reasonable accurate is the Leap-Frog algorithm of Verlet [83], which updates position and velocity alternately.

$$x_i(t + \Delta t) = x_i(t) + v_i(t + \tfrac{\Delta t}{2})\Delta t \qquad (2.62)$$

$$V_i(t + \tfrac{\Delta t}{2}) = v_i(t - \tfrac{\Delta t}{2}) + \frac{f_i(t)}{m_i}\Delta t \qquad (2.63)$$

The time step must be significantly smaller than the fastest vibration in the system under consideration. As the vibrational frequency is related to the force constant, which is highest for bond-stretching, it is common to fix bond lengths in molecular dynamics to allow a larger time-step. This approximation is valid as bond-

stretching is a rapid, low-amplitude motion and does not appear to affect the results obtained from the calculations.

In addition, the velocities and positions can be scaled to maintain a given pressure or temperature, or alternatively the system can be maintained at constant volume and/or energy. Which system is used is determined by the nature of the calculation.

The initial positions for the calculation can be either from a random (or perhaps Monte-Carlo [84]) set-up, or from an energy minimisation calculation. The initial velocities are assigned randomly to give a Boltzmann distribution of energies, and the temperature scaled over a few cycles to give either the required temperature or total energy. This random assignment of starting velocities does not usually give a correct distribution of energy between the different modes, so a period of equilibration must be allowed before any meaningful results can be obtained. The length of this equilibration period depends entirely on the system in question.

2.13 Multiple Minima.

Any system with a large number of degrees of freedom will have numerous local energy minima, and all of the known gradient minimisation techniques will end up in the nearest local minima, which is not necessarily the global minimum. In most molecular calculations, we are interested in the global minimum (or at least one very close to it in terms of geometry and energy), so this is a major problem. Various methods have been tried or are being developed to overcome this or to reduce the chances of error, including the Distance Geometry techniques of Crippen [85]. The two most common techniques are to perform a conformational search or to use Simulated Annealing (a Molecular Dynamics based technique) [86].

2.13.1 Conformational Search.

A conformational search requires that all dihedrals be varied in a stepwise fashion so that all conformational space is sampled, and so is extremely time consuming. The amount of time required can be used by using a larger step size eg. most dihedrals show a 120° or 180° periodicity so the step size is set accordingly. In addition, many such starting points may well reduce to the same minima, so if calculations are to be performed on several very similar molecules, the results for one can be used to set up the starting points for the others.

2.13.2 Use of Molecular Dynamics to Isolate Minima.

In addition to the simulated annealing technique [86], which simulates a process similar to liquids forming crystals on slow cooling, molecular dynamics can be used to investigate the relative stabilities of different minima [84] and the barrier heights between them. If the temperature is sufficient for the molecule to sample a sizeable part of conformational space, then the proportion of the time spent approximating to a given conformation reflects the stability of that conformation. In addition, the greater a barrier between two conformations, the less likely it is that the molecule will cross it, so the average length of time spent in one conformation will increase. This information can be used to direct the conformational search to a particular region of conformational space, as it can only distinguish between quite distinct minima and not between minima with very similar conformations. This is because during a dynamics simulation, the molecule is never in its minimum energy conformation, but always possesses excess potential energy in addition to kinetic energy.

Chapter 3

Computational Aspects of the Project.

3.0 Introduction.

The purpose of this chapter is to discuss some of the reasons behind the choices of packages used in the course of this work, and to briefly describe the nature of the in-house programs used. Also, a section detailing the programming done as part of the research has been included. The program HYDRO, which packs heterogeneous amphipathic molecules into a compact monolayer prior to molecular mechanics / dynamics calculations is described in the final section. (See table 3.2 for a summary of program functions.)

3.1 Computer Resources.

The calculations described in this thesis were performed on several machines,described below.

3.1.1 Machines Operated by the Cancer Research Group (8 users).

A μVAX II / GPX workstation with 13 Mbytes of memory, three hard disks, a TK-50 tape cartridge drive and a 1/2 inch tape drive. The hard disks had a total unformatted capacity of 370 Mbytes, one of which (90 Mbytes) was used as a system disk, another (140 Mbytes) contained the GAUSSIAN executable files and was used as scratch space for the integral files, while the third disk (140 Mbytes) was a general user disk. In normal operation, the free diskspace was typically 80 Mbytes (40 Mbytes scratch space, 10 Mbytes user space and 30 Mbytes free on the system disk).

An FPS-500 minisupercomputer with initially 128 Mbytes (16 Mwords) of memory, later reduced to 64 Mbytes (8 Mwords), 4 hard disks and a high speed 1/2 inch tape drive. Under normal operation the machine had approximately 540 Mbytes free disk space of which around half was used as scratch space for the GAUSSIAN

integral files. The machine ran at approximately 18 MIPS (scalar) and 10 Mflop (vector) under a normal load.

In addition an SCS-40 minisupercomputer was operated by the research group. It initially possessed 4 Mwords (64 Mbytes) of memory, subsequently increased to 16 Mwords (256 Mbytes) and 4 hard disks giving a storage capacity of 1.2 Gbytes. It did not possess a tape drive, but could make use of the drive attached to the μVAX. This machine was used for just under a year and then replaced by the FPS-500 when the manufacturers folded in 1989.

3.1.2 Machines operated by the University of St. Andrews.

The University of St. Andrews operated twin VAX 11/785 machines each with 10 Mbytes of memory, three hard disks of which only one was available to general users, and a 1/2 inch tape drive. Typically the disks had up to 40 Mbytes free space, though during term time this was more usually less than 10 Mbytes. The machines ran approximately 1.5 times the speed of the μVAX and $1/40^{th}$ of the speed of the FPS. As these machines had a large number of users and minimal free disk space, it was rare to have sufficient resources available to run GAUSSIAN *ab initio* calculations, so these were generally performed on either the μVAX or the FPS.

3.1.3 External Machines.

Before the arrival of the SCS-40, which used the CRAY instruction set, the research group had a 50 hour allocation of "pump-priming" CRAY X-MP time at the Rutherford Lab. This was used largely to test code prior to transferring to the SCS-40 and, in view of the considerable computing resources available at St. Andrews, this use was not continued after the arrival of the SCS-40 and its subsequent replacement by the FPS-500.

In addition members of the research group had access to two of the VAX machines at Daresbury in order to use the Protein and Small-Molecule Crystal Structure databases.

3.1.4 Terminals and Networking.

All of the machines were accessible over a pad network from Digital VT200 and VT300 series terminals. Graphics were displayed either on the workstation VR290 terminal , a Tektronix T4109 terminal or a Digital VT340 terminal. The FPS-500 was initially connected to the μVax by a DECNET connection but this was replaced by a faster ethernet connection in 1990.

3.2 Packages Used.

In the course of this thesis, no attempt has been made to compare different packages that operate at the same level of theory, rather the work has been concerned with using the results of calculations in the study of activities of molecules and so a single package at each level has been used throughout the work. The reasons for this lie in the fact that there is either little to choose between rival packages, as is often the case with molecular mechanics, or that one particular package is vastly more suitable to the problem in hand than its rivals (as is the case with MOPAC). The programs can be split into two categories, those widely available, which are usually written by large groups over many years, and are, of necessity, fairly general in application, and those written by various members of Dr. Thomson's research group here in St.Andrews. Programs in this latter category have been written to serve the particular needs of the group and, though written by one person, benefit from considerable input from other members of the group.

3.2.1 AMBER.

There are many molecular mechanics packages available from both commercial sources and via the QCPE [87] (Quantum Chemistry Program Exchange, Indiana University), though all use variations of the force-field described in section 2.13.1. and generally give similar results. The differences lie in the values of the parameters, in the number of terms used in each function and in cost (AMBER [81, 82, 88], ECEPP [78, 89, 90] and MM2 [79] are available for little more that the cost of a tape, while CHARMM [91] is rather more expensive). AMBER, a modular program including both energy minimisation and molecular dynamics, has been the program of choice. It uses only a simple quadratic function for bond stretching and bending rather than the additional cubic terms of MM2 and terms up to power six of MM3 [80]. As a results, it does not model distorted geometries quite so well, but in these cases, the geometry can be constrained to that obtained from semi-empirical or *ab initio* theory. Use of the simple quadratic term prevents errors due to poor initial geometries as each function has only the single minimum.

Though AMBER was initially parametrised for nucleic acids and peptides, it is very flexible and it is a simple matter to include extra parameters into the parameter data file, PARMDAT, that are consistent with those already developed. The modular design of the program means that new residues can easily be introduced using the PREP module and many such residues combined by LINK. Starting geometries can be defined in the EDIT module from files in the Brookhaven Protein Databank (PDB) format [92, 93] (see table 3.1). The size of the system studied depends solely on the available memory and the array sizes in each module can be easily changed to take account of the requirements of the calculation and the facilities available, thus it is possible to perform calculations on systems in excess of 5000 atoms on the FPS-500. The program is readily available from Prof. Kollman and Dr Chandra-Singh, which coupled with its flexibility outlined above, was the main reason for choosing this program. Since starting this work, a highly vectorised version (v 3.1) of the code has been released, greatly improving the speed of calculations on the FPS-500.

3.2.2 MOPAC.

In the field of semi-empirical methods, the MOPAC program [94] of Dr. J.J.P.Stewart is by far the most commonly used and the MNDO-type (MNDO [45], AM1 [62], PM3 [63, 64]) semi-empirical parametrisations, which are fully implemented within the MOPAC program, are generally held to be the most accurate in reproducing molecular properties including geometries, heats of formation, dipole moments and vibration frequencies. A routine has been added to produce output in a modified form of a Cambridge Crystallographic Database [95] format (CSSR) file containing atomic coordinates and charges (a "CHG" file), which can then be read into several graphics packages. In addition, the more usual output file contains information concerning interatomic distances, the bond-order matrix and dipole moments. Though MOPAC can perform CI and UHF calculations, this thesis is concerned only with closed shell molecules, so only RHF calculations have been performed.

3.2.3 The GAUSSIAN series.

The GAUSSIAN [56] series of programs (both 86 and 88 releases) have been used for the limited amount of *ab initio* work done in this thesis. This series of programs were chosen over others like GAMESS [96] and CADPAC [97] because they are simple to use, well documented and tested, and frequently updated to improve their computational efficiency. As most of the molecules studied are too large for anything more than a minimal basis with a small number of Gaussian functions for each orbital, *ab initio* calculations have only been used to determine force constants for molecular mechanics calculations, some timing comparisons using the SCS-40 minisupercomputer, and for the brief comparative study of basis set dependency of several properties of inositol phosphates (chapter 7).

The study of basis set dependency for inositol phosphates was carried out using GAUSSIAN80 (UCSF) as implemented within the QUEST program [98]. At the time

the study was carried out, GAUSSIAN88 was not installed on the FPS-500, but a partly vectorised VAX version of QUEST (v 3.0) had been converted to run on the FPS. The *ab initio* part of QUEST (GAUSSIAN80 - UCSF) is essentially the same as the other members of the GAUSSIAN series except that the input has been changed to be consistent with the other AMBER modules. It is also possible to perform combined molecular mechanical / quantum mechanical (MM/QM) calculations using QUEST.

3.2.4 ASP.

Similarity indices are calculated using the program ASP [99], written by Catherine Burt of Oxford Molecular. The indices can be calculated for both the electrostatic potential and electric field according to the methods of Hodgkin [73] and Carbo [72] using a grid of points outside the Van der Waals radius of the molecules. The similarity index can be optimised using a simplex algorithm [100], but as this is a particularly time consuming operation, the program has been converted to run on the FPS-500 by Michael Charlton. During optimisation, the program will alter non-ring dihedrals as well as moving the molecules relative to each other, though the rotation can be prevented by the user, or an energy weighting placed on the simplex to prevent it .producing a high-energy conformation.

3.2.5 CHEM-X.

CHEM-X [101] is a molecular graphics display package from Chemical Design Ltd., Oxford. Though it is a powerful program, it is limited to a 2-dimensional display by the terminals available, so large systems, such as the lipid bilayers, can become very confusing when displayed though this problem does not arise with small molecules (eg. phorbols and inositol phosphates). It has also been used for geometry manipulations of these small molecules. The FLY option allows the molecule to be moved or rotated and

provides the means of measuring bond lengths, angles and dihedrals. Also present within the FLY option is a weighted least squares fitting procedure which can be used to overlay the similar groups of the otherwise dissimilar molecules. The fitting is performed by selecting pairs of atoms, at least three are required, and assigning a weighting to each pair. If an exact fit of all the pairs cannot be obtained, the priority for fitting each pair is decided by the weighting. The other option that was used extensively was the MODIFY / CURSOR option which allowed the user to alter the structure of a molecule using a "mouse" attached to the terminal. This allowed for the breaking and making of bonds as well as changes in the internal coordinates and the adding and removal of groups to the molecule.

A major drawback of the program lies in its wide range of functions, many of which are not used in this thesis, which makes it rather slow. Its performance is further impaired as it is not optimised for the system at St. Andrews. For this reason, an in-house program MHDRAW has been used to display most of the systems, as this has been tailored specifically for the computers in St.Andrews and the needs of the research group.

3.2.6 In-House Programs.

3.2.6.1 3D2.

3D2 [70], written by Derek Higgins and based on work by Colin Edge [102], is a program that calculates and plots the MEP of a molecule from point charges. Input to the program can be in several forms, the most usual being a CHG format file, (modified CSSR format) containing MOPAC point charges. Alternatively a PDB format file can be read, in which case, only the surface can be calculated as a PDB file does not contain charges, or the calculated MEP itself can be read in (the program can write the MEP to a specified file, or it could be calculated directly from the wavefunction). The MEP is usually plotted onto the molecular surface at the Van der

Waals radius or multiple thereof and displayed as a 2-dimensional colour-coded picture with shading to give it a 3-dimensional quality. The surface is calculated using the Connolly surface algorithm (MS) [69, 70, 103], with a probe radius of 0.0 Å. This algorithm gives the user three surface options for each atom, the surface can be calculated normally, the atom can be used to block the probe but have no surface of its own, or it can be completely ignored.

The convention used in this project and others at St. Andrews is to display areas of negative electrostatic potential in red, positive in blue, with intermediate regions in green, brown and yellow. The MEP can also be displayed in the form of a graph, with the atoms in order along the X-axis, and the ESP on the Y-axis. If the program has not been provided with atomic charges, it can only plot the surface, but colouring can be used to distinguish between different residues or molecules, for example to distinguish between a ligand and its binding site.

3.2.6.2 MHDRAW.

MHDRAW [104] was written to take advantage of the graphics facilities of the μVAX II workstation terminal, and combats the slow speed of CHEM-X by possessing only the features required by the research group. It was written by Michael Charlton, and can display the molecules in ball and stick form, or as shaded spheres (at the bonding radius rather than the Van der Waals radius) and can include all the atoms or just specified elements. The shading gives a 3-dimensional appearance to the pictures and is especially useful in displaying the large bilayer systems. The program is capable of reading structures in a variety of formats including PDB, CSSR,.CHG, Z-Matrix, and MOPAC archive and can write to files in all of these formats except for the MOPAC archive, thus it can be used to generate Z-matrices for running MOPAC or GAUSSIAN calculations, though the program MOPGEN (see table 3.2) is faster, especially if a large number is required.

3.3 Programming Work.

The programming aspect of this work can be divided into three sections, converting and debugging programs for the FPS-500 and SCS-40, writing short programs (mainly file-handling routines for passing data from one program to another), and the program HYDRO, which will be dealt with in a section of its own.

3.3.1 Installation of Packages on the FPS-500 and SCS-40.

V 3.1 of AMBER has been installed on the FPS-500. It arrived as an untested 32-bit vectorised code, so the work involved testing the program against known results from the version already on our μVAX. After the initial debugging, the program was vectorised using the VAST preprocessor and optimised using the compiler optimisation routines, again checking output against that from the μVAX. This revealed that optimisation of some routines, AMINV in NEWTON, MATINV in GIBBS and NNBOND in all modules, introduced errors into the program so they are vectorised, but not passed to the compiler optimisation routine. The bugs discovered included incorrect unit numbers in ANAL and EDIT, incorrect binary record length (512 bytes, which was changed to 1024) and a factor of 4 had been applied to all the database entries for no apparent reason. In addition routines have been included to correctly enable the restart procedure in NEWTON and altered the array sizes to reduce the memory allocation to the minimum required for the systems under consideration.

During normal operation, the EDIT module reorientates the molecules when solvating using the BOX option, to give the smallest amount of solvent for the set cutoffs. While this procedure normally improves the efficiency of the calculation, it causes problems for the bilayer calculations (see chapter 6). Thus an option was added to the input data that would allow or prevent this reorientation depending on the nature of the system (the user specifies which option to take).

It was not so simple to deal with QUEST, as it is only available in 32-bit VAX

specific (v 3.0) and 64-bit vectorised CRAY code, again, both untested. As the SCS-40 uses the CRAY instruction set, the program compiled on the SCS-40 without too much trouble, but it was so heavily bugged that few of the functions worked correctly and more than half of the test jobs failed for a variety of reasons. A further series of test jobs were devised to check the remaining features, and most of the bugs were removed, though the program was still not able to store the 2-electron integrals in a super-matrix, but had to use one of the other options, and the post-SCF routines remained unimplemented. In all, the final working version of the program is in excess of 77,000 lines of FORTRAN.

After the departure of the SCS-40 in Autumn 1989 and its replacement by the FPS-500, it was necessary to repeat the installation for the new machine. The FPS uses 32-bit integers, rather than 64-bit like the SCS-40 and CRAY machines, which caused problems as the central array is used to store both integers and 64-bit real numbers, thus the CRAY version failed on the FPS as the indexing of the reals and integers were out of step and some reals were being interpreted as integers and vice versa. There were also some statements that the FPS compiler did not understand (eg. a computed GO TO with a non-integral index), but had been acceptable on the SCS-40, and unlike its predecessor, the FPS did not accept subroutines with the same name as common blocks. All of these were relatively simple to correct. There were several possible approaches to the problem of different real and integer sizes, the first tried, was to reduce the size of the reals to 32 bits using single precision which immediately caused problems as the code had been written with double precision rather than generic function calls. Even with these alterations, the program failed as it packs several items of data into a single (64-bit) variable. To combat this, routines were written that imitated packing into and unpacking from 64-bit integers and the indexing of the integer sections of the main array was altered accordingly, in effect trying to get the machine to perform as a 64-bit integer machine. Though the packing and unpacking functioned correctly, the program still would not work.

The 32-bit VAX code contains several VAX specific routines, mainly for asynchronous I/O, but with the failure of the attempts to convert the 64-bit code, work

was done to convert this to run on the FPS-500. Several of the VAX-specific routines could be replaced, but those involved in the post-SCF procedures could not, so the procedures were not implemented. That this should be the case is not surprising as these procedures have not been implemented in the 64-bit version either. The remaining problems with VAX specific I/O were replaced by conventional I/O, though a problem with addressing of direct access I/O had to be dealt with. The final problem to be cured was caused by the handling of character data. Character variables on both machines have 8 bits, but the FPS stores them as the rightmost 8 bits of a 32-bit word while the VAX uses the leftmost 8 bits. Thus when bitwise operations are performed on character data on the FPS, they miss the stored data completely, or alternatively, if the data is stored as a result of setting specific bits, the FPS always interprets the data as a "null string".

Once the program was working, the major time consuming routines were rewritten to improve their vectorisation, though vectorisation of some routine using the VAST preprocessor introduced bugs into the program (FORHF and HELFEY are examples of such routines). The computation time could also be improved by expanding short routines in-line using the Procedure Integrator (PI) facility on the FPS. In all these changes were able to save over 40% of the computation time for a set of test cases.

3.3.2 File Handling Routines.

Derek Higgins has already added a routine to MOPAC to produce files in a modified form of the Cambridge Crystallographic Database format (CSSR), containing nuclear coordinates, connectivities and charges. For such a file, the suffix ".CHG" is routinely used on the VAX systems (and the file is referred to as a CHG file - see table 3.1)). This type of file can be read into CHEM-X, MHDRAW and 3D2 and as such is rather versatile. As this type of file is the closest we have for a standard geometry input, it became the normal means of communication between the different programs

used in this work. The exception to this is the PDB file that is used to pass information from HYDRO (or BIL) to the EDIT module of AMBER, this is because AMBER is designed primarily for calculations on peptides and proteins for which the PDB format is standard.

The CHG file is produced from the coordinate file from the PARM or BORN modules of AMBER by my program, CONVERT. It requires input from five different files, a Z-Matrix file (produced in EDIT using the keyword GAUSS), the coordinate file (INPCRD from PARM or RESTRT from BORN), a file containing bonding-radii, one containing the charges for the AMBER residues and a control file which gives the total number of atoms, the type of force-filed (all or united atom) and the residue sequence. From this a CHG file is created containing the minimised nuclear coordinates, the connectivity and the charges used in the AMBER calculation. A similar file can be created from the restart files from the GIBBS and NEWTON modules as the coordinate information is stored before the velocities, though in both cases the velocity information is ignored.

The output from the NEWTON module is designed to produce a moving image on a real time graphics display, but as there is no such a device available at St.Andrews, the program DPIC was written to produce the coordinate data in a form suitable for the graphics packages available (CHEM-X and MHDRAW). Essentially this program also produces a CHG file (or several), but has the option to place each time-step into a separate file or into the same one and it can also add the edges of the periodic box if boundary conditions are in operation. By placing several time-steps into the one file, some idea of the relative mobilities of different parts of the molecule can be obtained. As this program may be called on to handle a very large number of atoms (it has been dimensioned for up to 10,000 at 20 points), the connectivities are not calculated directly, but are taken from the CHG file produced from the BORN module, which always precedes a dynamics calculation and contains the same atoms as for each point of the dynamics. On the FPS this program has generated output from 20 points each of 5373 atoms in under a minute.

MOPAC Z-matrices can be generated from CHG files via MHDRAW, but if

several matrices need to be calculated, this can be rather slow. Instead, a short program, MOPGEN, has been written to generate a MOPAC format Z-matrix from the coordinates and connectivity contained within the CHG file. Both files possess the same prefix, and numerous files can be generated with the same keyword line by giving the program the names of a many CHG files as are required. The title line for each file is taken from the relevant CHG file.

HYDRO, the monolayer packing program produces a PDB file, which can be read directly into the EDIT module of AMBER, but usually a bilayer is required, so the routine BIL is used to produce a bilayer using the monolayer and its double reflection (X'=X, Y'=-Y, Z'=-Z) which is then transformed so that its centre of mass lies at the origin. This is required for the correct functioning of the subsequent AMBER molecules, and is normally done as part of the reorientation in EDIT, which has been disabled to give correct solvation of the bilayer.

3.4 HYDRO.

The purpose of HYDRO is to generate pseudo-random, close-packed monolayer arrangements of heterogeneous lipids, arranged around a single central test molecule. It was preferable to find a computational means of doing this as visual positioning using a graphics package does not produce sufficiently close packed monolayers that solvent is not able to penetrate during the early stages of a dynamics calculation. Brasseur *et al* [105] devised a method whereby single lipids are added to the growing monolayer in turn, and energy minimisation performed. This would prove too time consuming for the relatively large systems, and unnecessary if dynamic information is the prime concern.

The main reason for producing the packing program was the need to study phorbol esters in an environment closer to the physiological environment than the gas phase as was the case for the initial calculations (chapter 4). The results of the program would be used as the starting point for AMBER minimisation and dynamics.

3.4.1 Algorithm.

The monolayer is packed by minimising the solvent-accessible hydrophobic surface area, which is calculated using the Connolly surface algorithm [69, 70, 103], and takes advantage of the 2 dimensional nature of a monolayer. The molecules are treated as rigid cylinders, with the axis along the Z-axis and distributed randomly in the X-Y plane. The molecules are constrained to their position on the Z-axis, but are free to move along both X and Y axes. Atoms are designated as hydrophobic or hydrophilic according to their AMBER atom types, and the surface of the hydrophobic atoms calculated using a probe radius of 1.75 Å, the larger radius of a water molecule. The Connolly algorithm gives the molecular surface in terms of point coordinates, the area associated with that point, the associated atoms and a vector perpendicular to the surface.

The molecules are moved by summing the X and Y components of the vectors, but any that point away from the central molecule are ignored (otherwise the system would fly apart rather than contract), and the molecules moved in the direction of the sum of the vectors, by a distance given by:

$$\text{Displacement} = \frac{V\,R}{20\,N} \qquad (3.1)$$

where: V = Sum of Surface Vectors.

R = Distance from Molecular Axis to origin.

N = Number of Surface Vectors.

At the end of each cycle, a Van der Waals distance check is carried out to check for molecules that overlap. If such an overlap is found the displacement of the molecule furthest from the centre is reduced by half, and its position recalculated. The packing proceeds until the total hydrophobic surface is reduced to below a specified proportion of the starting value (set by the user to take account of edge effects due to the size of the system and shape of the molecules), or the displacement vectors of all of

the molecules have been reduced to essentially zero due to Van der Waals contacts. As minimum energy conformations of the individual molecules are used, it is expected that this program will slightly underestimate the surface area of the monolayer, but this should be corrected during a dynamics simulation at constant pressure.

3.4.2 Computational Aspects.

The initial positions of the molecules are obtained by estimating the minimum radius which enables all the molecules (except the first which is placed at the origin) to be placed in a ring around the origin without coming into Van der Waals contact, and placing them there. The radial component of the position is then increased by a random factor of up to the magnitude of the radius again. The molecules are also rotated to a random degree about their own central axis, and the rotation updated by a small amount during each cycle.

Connolly's surface algorithm uses conventional cartesian coordinates, but it is convenient to use a different set for updating molecular and atomic positions. Each molecule is able to move in both the X and Y directions, but not the Z direction, so its central axis need only have these two coordinates, in addition it is given a rotational coordinate, ϕ, which is the current rotation from an initial arbitrary direction. Within each molecule, the atoms are assigned local coordinates relative to the central molecular axis. The concept of the program treats the molecules as rigid cylinders, so the local coordinates are in the form of cylindrical polar coordinates, (r,θ,Z).

It is fairly simple to convert these local cylindrical coordinates to global cartesians by the following transformation:

Atomic coordinates	r,θ,Z
Molecular coordinates	X,Y,ϕ
Transformed coordinates	X',Y',Z'

$$X' = X + r \sin(\phi + \theta) \quad (3.2)$$

$$Y' = Y + r \cos(\phi + \theta) \quad (3.3)$$

$$Z' = Z \quad (3.4)$$

3.4.3 Input.

The program requires a control input file that specifies the test molecule and the lipids to be packed around it. In addition the control file is used to specify the convergence criteria, time limit, maximum number of cycles and initial Z-offset of the test molecule.

Coordinate data can be in two forms, from the PREP input file of AMBER, or from a file containing the cylindrical coordinates and an the surface option for each atom (CPC format). This later file is produced by the program when the molecule is encountered for the first time, and can be used subsequently to save on computation time. In both cases, the files are called by their AMBER residue name with the relevant suffix to describe their format.

Finally, the Van der Waals radii are stored in a separate file, making it easier to change values if necessary. This is the case with changing from an all-atom approach to the united-atom approach, when the radius of carbon changes from 1.6 Å to 2.1 Å.

3.4.4 Output.

The main output of the program is a PDB file containing the coordinates of the packed monolayer in a form suitable for input into the EDIT module of AMBER. If however a bilayer is required, the routine BIL will duplicate the selected monolayer and place the centre of mass of the bilayer at the origin. If the packing fails to converge

within the time or number of cycles specified, a restart file is generated (and updated every cycle) giving the molecular coordinates and there is an option in the control file which allows the initial position to be taken from this file. In some cases, intermolecular distances may be useful (surface areas can be estimated from the distances of nearest neighbours) and so are calculated from the molecular coordinates and output for the packed monolayer.

3.5 Tables.

Table 3.1 A Summary of File Types.

File Type	Description of Format / Function
CSSR (CSR)	Cambridge Crystallographic Database format structure file, giving unit cell dimensions, atomic names, cartesian coordinates (either in A or fraction of the unit cell) and connectivities.
CHG	Modified CSR file which includes charges after the connectivity data. The standard input file for all the graphics packages used.
PDB	Brookhaven Protein Databank format, contains atomic names, coordinates and residue information.
CPC	Input file for HYDRO (also created by the program) containing atomic numbers, cylindrical polar coordinates and a surface calculation index for use by HYDRO.
PRP	Residue information for the PREP module of AMBER. Contains atomic numbers, atom types and coordinates in a Z-Matrix.

Table 3.2 A Summary of Programs Used.

Program Name	Function
MOPAC	Semi-empirical quantum chemistry package, incorporating MINDO/3, MNDO, AM1, PM3 parametrisations
GAUSSIAN88	*Ab initio* quantum chemistry package with a wide range of basis sets (Pople, Dunning / Huzinaga etc.)
AMBER	Molecular mechanics (energy minimisation / dynamics / free energy perturbation) package.
ASP	Similarity Index Package.
CHEM-X	Molecular graphics display and geometry manipulation package.
3D2	In-house MEP and surface calculation and display.
MHDRAW	In-house molecular graphics display.
CONVERT	Produces standard graphics input from BORN output.
DPIC	Produces standard graphics input for NEWTON output.
MOPGEN	Produces MOPAC input from a standard graphics input and user-specified information.
HYDRO	Packing of amphipathic monolayers by minimisation of solvent-accessible surface area.
BIL	Produces bilayer in PDB format suitable for input into EDIT.

Table 3.3 - AMBER Modules.

Module	Function
PREP	Residue preparation module of AMBER.
LINK	Connects residues together for AMBER.
EDIT	Solvation, coordinate setting and preparation of data for minimisation etc. for AMBER.
PARM	Generates parameter set for subsequent AMBER calculations.
BORN	Energy minimisation module of AMBER.
NEWTON	Molecular dynamics module of AMBER.

Chapter 4

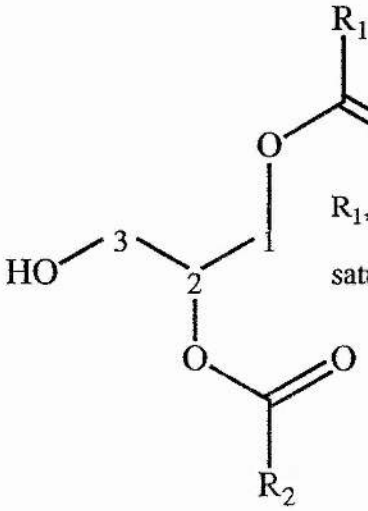
TPA-Like Tumour Promoters

4.0 Introduction.

Phorbol esters are extremely potent tumour promoters, and are the active agent of Croton Oil, extracted from *Croton Flavens,* a shrub widespread in the Caribbean and surrounding areas [6]. They have been found to be the major cause of oesophagal cancer in the Dutch Antilles and are used extensively to promote tumours experimentally [106]. A receptor with an affinity for tetradecanoylphorbol acetate (TPA) in the nM range has been found within the cell membrane and subsequently identified as protein kinase C (PKC). Phorbol esters are able to displace the endogenous activator, diacylglycerol (DAG), from its site on the enzyme suggesting that both bind to the same site, now thought to be part of the membrane-binding / regulatory domain of the enzyme [19, 20]. The affinity of the enzyme for phorbol esters is some 10^4 greater than for DAG [20]. While the kinetics of binding of phorbol esters by PKC have been characterised and activation of PKC correlates well with the extent of tumour promotion [107], little is known about the actual binding site or about how inappropriate levels of PKC activity leads to tumour promotion. By correlating changes in structure and electrostatic potential with differences in activity, it should be possible to identify critical features of the molecules that are responsible for their tumour promoting activity and to increase the understanding of the nature of the binding site on PKC. It is not, however, desirable to develop inhibitors of PKC activation as the enzyme is intimately involved in many normal cellular functions.

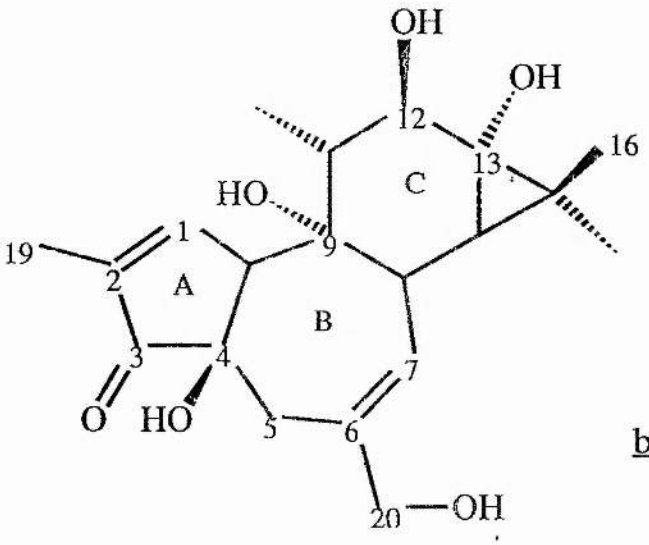
More recently, several other classes of compound have been shown to act as tumour promoters in a manner analogous to phorbol esters and to activate PKC by binding to the DAG site. Notable amongst these are the ingenane and teleocidin families [17, 108] (the basic structure and numbering systems for these 4 classes of molecule are shown in fig 4.1.). The teleocidins; in particular, bear little obvious resemblance to the phorbol esters, yet possess very similar levels of activities, a situation contrary to general enzyme biochemistry where molecules with similar activities possess considerable structural similarity and any differences in structure or ESP result in a large change in the ability to activate or inhibit the enzyme. The lack of

similarity between the different PKC activators suggests that a large proportion of the molecule is not involved in binding to the enzyme, but rather, in view of the cellular distribution of the enzyme and activators, most of the molecule may be involved in a non-specific interaction with the cell membrane which ensures that the relevant parts of the activator are presented to the enzyme in the correct orientation.

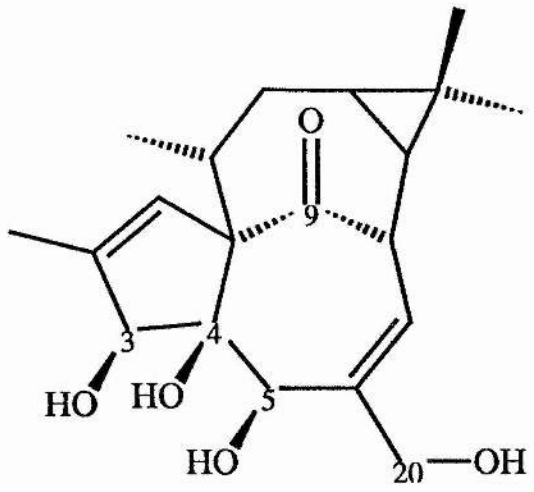


R_1, R_2 are variable length saturated or unsaturated acyl chains

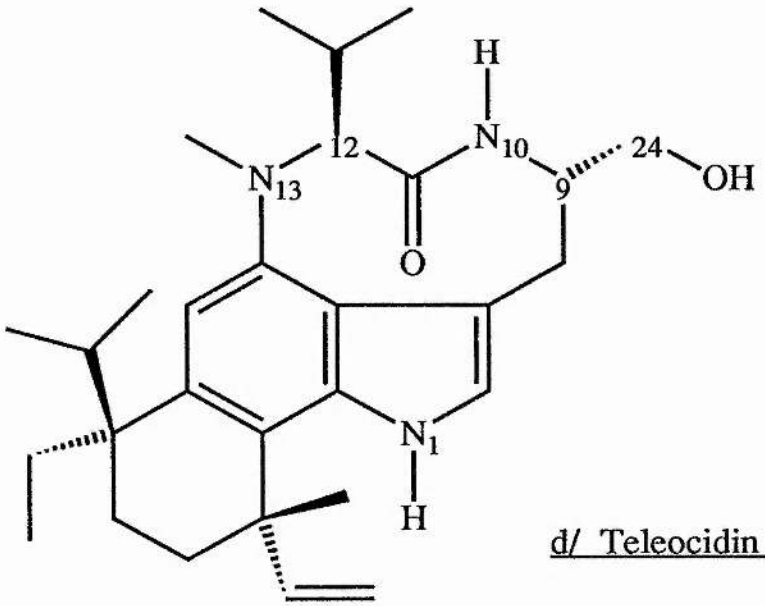
a/ Diacylglycerol



b/ Phorbol



c/ Ingenol



d/ Teleocidin B

Fig 4.1 Structures of DAG and the TPA-like Tumour Promoters.

4.1 Activity Data.

A large amount of activity data has been published for phorbol ester derivatives, however in many cases, it is difficult to compare the results of one study to another. Often the extent of enzyme activation is measured at single concentrations of activator (e.g. [109]) and cannot therefore be compared with data for derivatives taken at a different concentration. A further problem with activity data measured at a single

concentration is that changes in enzyme activity can be brought about by an alteration of either K_m or V_{max}, which cannot be distinguished from just a single point but require a complete concentration / activity plot.

Table 4.1 Tumour Promoting Activities of Phorbol Ester Derivatives.

Derivative	Activity
TPA	++++
1α2α Epoxy TPA	None
1β2β Epoxy TPA	None
1,2α Dihydro TPA	None
1,2β Dihydro TPA	None
3α Hydroxy TPA	++
3β Hydroxy TPA	++
4β Methoxy TPA	None
4β Deoxy TPA	++++
5β Hydroxy TPA	+++
6α,7α Epoxy TPA	++
6β,7β Epoxy TPA	++
16' Hydroxy TPA	++++
20' TPA Aldehyde	+
20' TPA Carboxylic Acid	None
Phorbol	None
Phorbol dibutyrate	++

Alternatively, activities of the phorbol esters have been expressed as ID_{50} values [6] or in terms of their ability to induce ornithine decarboxylase [108]. Neither are adequate measures of tumour promoting activity as they do not appear to be a direct

part of the promotion process. Deficiencies in the correlation of ID_{50} values with promoting activity have been shown with the phorbol esters [18] but ornithine decarboxylase induction does show good correlation with the promoting activities of the known TPA-like tumour promoters. It does not however correlate with promoting activity per se, as palytoxin, an extremely toxic non-TPA-like tumour promoter does not induce the enzyme [110], so it is only valid as a measure of the activity of the TPA-like tumour promoters.

The most consistent and reliable activity data is the direct measure of the promoting activity using the standard assay of Hecker [111], though the data is largely qualitative as the measurement in terms of number and rate of tumour development is of necessity fairly imprecise. This assay has been used for a wide range of phorbol ester derivatives [112], the results of which a summarised in table 4.1 above (the '+' terms used in the activity data are taken from [112]).

By comparison with the phorbol ester activity data, the information concerning the ingenol derivatives and the teleocidins is rather limited, reflecting their more recent discovery as tumour promoters. The ingenol activities (table 4.2) of Fujiki *et al* [108] are expressed in terms of the ability of different ingenol esters to induce ornithine decarboxylase and can be compared with the phorbol results as the Hecker assay [111] has shown similar promoting activities for TPA and 3' ingenol esters [112].

Table 4.2 Promoting Activities of Some Ingenol Derivatives.

Derivative	Activity
20-O-Isobutyl,3-O-Propionyl Ingenol	++++
20-O-Isobutyl Ingenol	None
3,5,20 Tri-O-Acetyl Ingenol	None
Ingenol	None

Table 4.3 shows the effect of changing chirality of the two asymmetric carbons (C9 and C12. See fig 4.1) within the lactam ring of teleocidin. The results, shown, from Fujiki *et al* [113] are for the model compound, Indolactam V, (fig 4.2) which is common to all four members of the teleocidin family, which differ in the substituents of the tetramethyl ring. Though the activity of the SS indolactam V isomer is much lower than that of TPA, the activity of teleocidin B, with the same stereo-isomeric structure for the lactam ring, is comparable to that of TPA so the values given in the table are for the complete teleocidin as extrapolated from the indolactam V results. The olivoretins differ from the teleocidins only in being methylated at position 24 (fig 4.1) yet are completely inactive as both the indolactam V model and as the complete molecule [114].

Table 4.3 Promoting Activities of Teleocidin Conformers.

Derivative	Chirality at C9, C12	Activity
Teleocidin	S, S	++++
Teleocidin	R, S	None
Teleocidin	S, R	None
Teleocidin	R, R	None
Olivoretin	S, S	None

The geometries of these molecules have been calculated and the resulting electrostatic Potentials (ESP's) determined. Though the groups of molecules differ considerably from one another, it is hoped that certain common features will emerge from the calculations and will thus pinpoint the particular areas of the molecules involved in specific interactions with the binding site. From this it should be possible to develop a model for the binding site including the main electrostatic interactions and, from a consideration of the inactive derivatives, some ideas as to the positions of residues on the edge of the pocket.

4.2 Methodology.

As the smallest molecules studied in this section consist of 54 atoms, it was not practicable to use *ab initio* methods for the calculations, so the semi-empirical methods PM3, AM1 and MNDO were used as implemented within the MOPAC program. This work was initially performed using version 3.1 of MOPAC, with the PM3 studies carried out using the later version 5.0. The optimisations were carried out over internal coordinates and continued until the gradients had fallen below 0.01 kcal rad^{-1} or kcal $Å^{-1}$ (for angles and bonds lengths). In some cases it proved impossible to lower the gradients to this value as either the heat of formation was constant (to 4 or 5 decimal places) for many cycles of optimisation, or the optimisation routine was unable to find a point of lower energy than the most recent. In both these circumstances the calculation was stopped and the value for the energy used as it is clearly the (local) minimum value.

The electrostatic potentials were determined using the point charges obtained from the diagonals of the density matrix. In the descriptive sections, the AM1 potentials are used as, for the types of molecule under consideration, they are the most reliable. AM1 is a more recent method than MNDO and gives a consistently better reproduction of properties for small molecules [62, 64, 115] and though PM3 is a the most recent of the three, there has been little published concerning its reliability for molecules outside the range used in the initial parametrisation. Furthermore, testing of the MOPAC5 code and the PM3 parameters on their arrival in St. Andrews, suggests that the charges obtained from the density matrix with PM3 on molecules containing nitrogen is particularly unreliable. For example, PM3 gives a positive charge on the nitrogen of ammonia in contrast to MNDO and AM1, yet manages to give a dipole moment of the correct magnitude [64]. A desire to use the same parameter sets for all the ESP plots and this doubt over the reliability of PM3 for the nitrogen containing molecules, such as the teleocidins, is responsible for the choice of the AM1 potentials.

Crystal structures of phorbol [116], dihydroteleocidin B monobromoacetate [117], ingenol 3,5,20 triacetate [118] and 1,2 *sn* dilauroyl glycerol [119] were

obtained from the Cambridge Crystallographic Database at Daresbury. These structures were then modified to give the required derivatives, and missing hydrogens added, using the MODIFY / CURSOR option of CHEM-X. Initial gradients for the modified molecules were determined using a single SCF calculation within MOPAC, and the structures were refined with molecular mechanics if the value of the gradient norm was over 200 (a quantity without dimensions as the individual gradients are measured either in kcal rad^{-1} or kcal $Å^{-1}$). The molecular mechanics refinement was found to reduce the computer time required by the semi-empirical calculation to a fifth for particularly poor initial geometries. The AMBER force field is such that accurate energies and geometries can only be obtained for unstrained systems with geometries similar to those used in the initial parametrisation. The cyclopropyl (C) ring of phorbol is highly strained with the carbon bond angles distorted well away from the usual sp^3 angle of 109.5° (and therefore the equilibrium bond angle). However, AMBER gives the correct 60° bond angles but with a higher energy than would be expected from the true potential energy curve. The lack of angle-bond cross-terms and the nature of the 3-membered ring mean that the strain on the bond angles cannot be relieved by altering the bond lengths from their equilibrium values. Several parameters required for the preliminary AMBER optimisations were not included in the supplied AMBER parameter file and so were derived from semi-empirical and *ab initio* calculations (see chapter 2).

As TPA consists of 100 atoms, geometry optimisations on the whole molecule using MOPAC are rather time consuming, so it is necessary to devise suitable models for which geometry and electrostatic potential calculations are feasible. An earlier study by Thomson and Cuthbertson [120] used a model consisting of just the A and B rings of phorbol, but this did not retain the crystal geometry for the B ring. To overcome this problem, the whole phorbol molecule was chosen as the model, replacing the ester functions with hydroxyl groups. It was felt that such a model would be justified as the promoting activity has been shown to vary with chain length in a manner that correlates well with lipid solubility. The validity of the phorbol model was tested by comparing the structure and electrostatic potential of the model with that of a short chain ester,

phorbol dibutyrate, which possesses a moderate tumour promoting activity [121].

The ingenol derivatives considered all had short chains so the structure of the whole molecule could be optimised, though, from experimental data, the acyl chains are required to confer lipid solubility rather than being involved in specific interactions with the binding site [6].

The model system chosen for the teleocidin studies consisted of the indole and lactam rings, a molecule known as Indolactam V and shown in Fig 4.2. Though this molecule does not possess the hydrophobic region found in teleocidin B, and thus has a much lower lipid solubility, it is able to act as a tumour promoter but with a reduced activity [113].

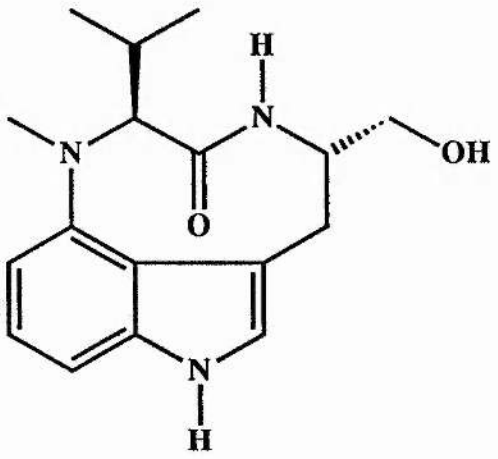


Fig 4.2 Indolactam V.

When the calculations of the structures of the TPA-like tumour promoters were started, MOPAC was limited to molecules of less than 80 atoms (40 non-hydrogen). 1, 2 *sn* dilauroyl glycerol is rather larger than this, so the length of the lipid chains was reduced to 11 carbon atoms each, to fit within this limit. With the arrival of the minisupercomputers, larger MOPAC calculations were feasible, but it was not thought necessary to use a larger model for DAG as the chain carbons do not have a noticeable effect on the ESP of the head group after the first two or three. Chains as long as eleven carbons were used to give some indication of the orientation of the molecule,

and the headgroup in particular, with respect to the lipid bilayer. No other modification was made to the conformation of the DAG prior to the minimisation, though subsequently the structure was altered to correspond to the phorbol ester bound to PKC and reminimised. The use of short acyl chains in the model DAG is valid as physiologically DAG has a heterogeneous mix of chain types and there appears to be no chain specificity for binding.

The large difference between the different families of molecules lead to problems in comparing their ESP's particularly in overlapping the parts of the molecules involved in the same interaction with the site. Much of each molecule does not appear to be involved in binding at all and thus the ESP over these regions would not correlate with activity but would distort the results of optimising the field or potential similarity of the different molecules using ASP. The use of model compounds would also cause problems with optimisation of the similarity indices as each of the families retain a varying proportion of the parent molecules. Fortunately the ESP's of both DAG and phorbol possessed only a few areas where the magnitude of the potential was at all sizeable. Thus it was possible to overlap similar functional groups with equivalent potentials from different molecules using the weighted fit found in the FLY option of CHEM-X. For most overlaps, the same weighting was used for all the groups fitted, but when comparing the phorbol esters with DAG, it was convenient to use a lower weighting for the acyl chains than for the groups involved in more specific interactions with the site.

4.3 Diacylglycerol.

The MOPAC optimised structure and electrostatic potentials on Connolly surfaces plotted at the Van der Waals radius and at twice that, are shown in figure 4.3. As would be expected for calculations on the isolated molecule, the chains are almost completely straight in an all-*trans* form and diverge towards the ends as a result of weak non-bonding repulsions between them. In a bilayer, the conformation of the chains would be expected to differ slightly from this, but the overall orientation should be essentially the same. The minimum energy position of the 3' O-H group varies with differing parameter sets, but the rotation barrier is small (<2 kcal mol^{-1}) and, because of the polarisation of the bond, dependent on the electrostatics of the local environment.

The Electrostatic potential in the vicinity of the hydrocarbon chains is essentially equal to zero, being shown in yellow and green. At the larger radius, the potential on the head group is generally negative, with the carbonyl atoms and 3' O-H group being the only significant charge centres, though polarisation of the individual bonds can be seen at the smaller radius. Care must be taken when interpreting the ESP's at the Van der Waals radius as the point charge model breaks down when considering the ESP in regions of space with significant (i.e. statistically non-zero) electron density. At this radius, the potential around the first carbons in the chains is positive, as a result of polarisation of the carbon-oxygen bonds, likewise the hydrogen of the 3' hydroxyl group also possesses a positive potential, which at larger radii is overshadowed by the negative potential of the oxygen.

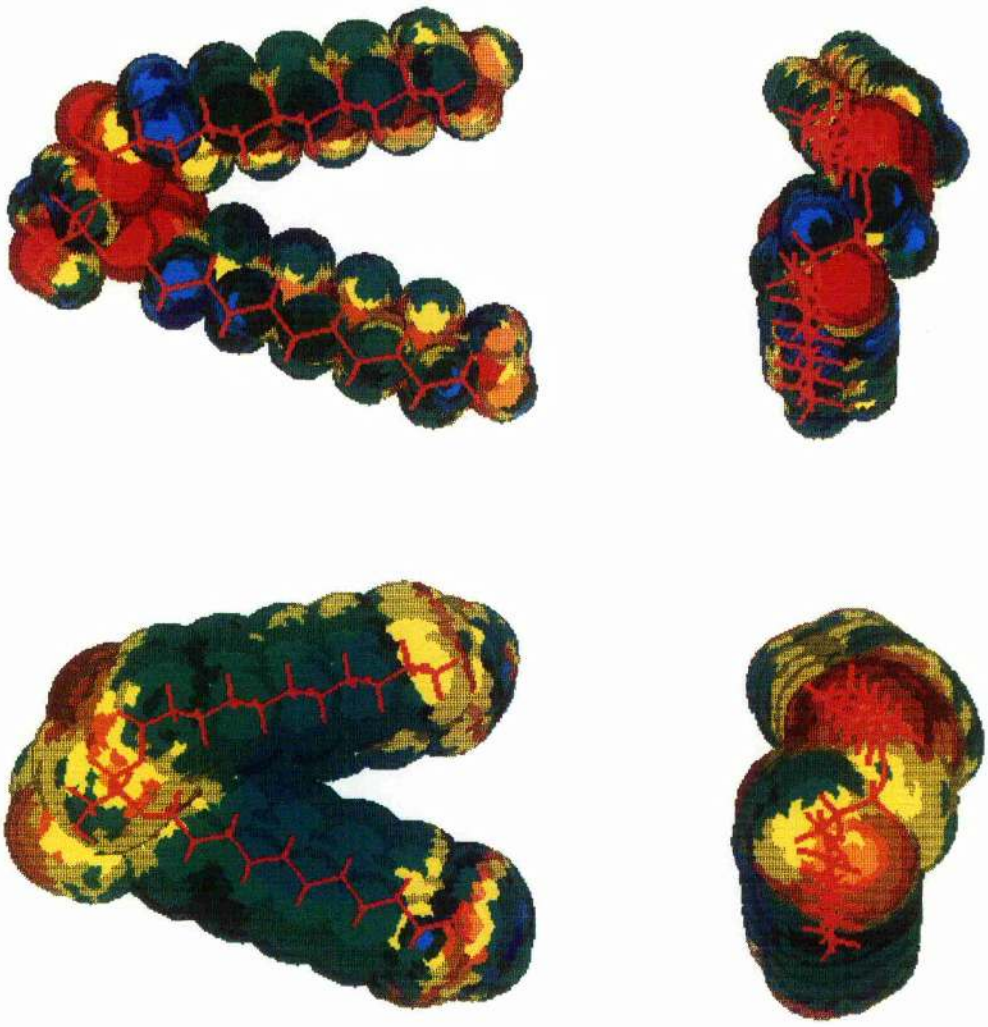


Fig 4.3. EPM of Diacylglycerol.

4.4 Phorbol Esters and their Derivatives.

In this section, the AM1 derived Electrostatic potentials of the phorbol ester derivatives are considered. Firstly the ESP of phorbol is compared with that of a short chain ester (phorbol dibutyrate - PDBu) that possesses tumour promoting activity (though not at the same level as TPA on account of the short chains), to test the validity of using phorbol derivatives as models for the esters. Then the chain orientation of the ester is compared with that of DAG and attempts are made to overlap areas of similar ESP. Finally ESP's of the different phorbol derivatives are compared in the light of known promoting activities of the esters, and used to produce a description of the DAG binding site. The modifications used for the structure / activity relationships derived in the last part of this discussion, are limited, with one exception, to the A and B rings.

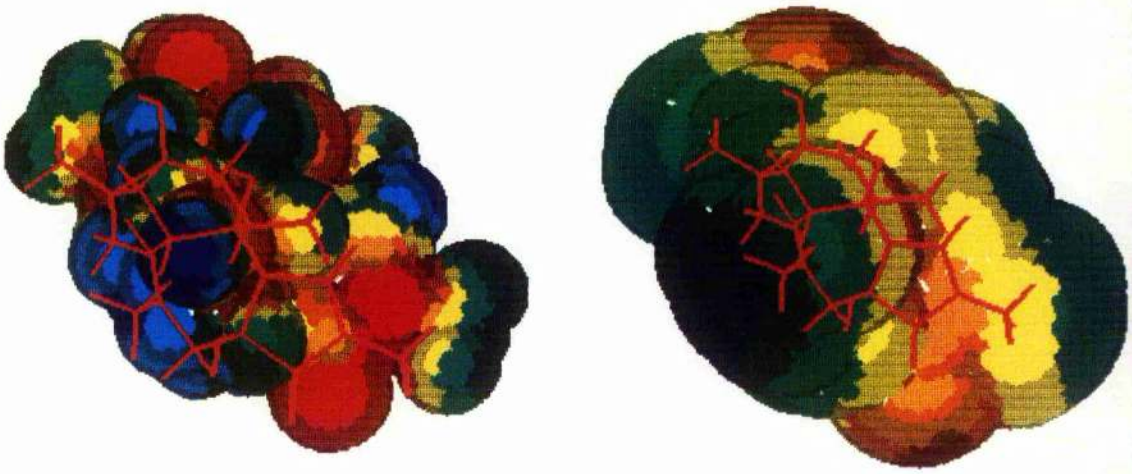
To test the validity of phorbol as a model for the phorbol esters, the similarity index for phorbol and PDBu were calculated over the whole molecules and over a fragment containing just the A and B rings. Indices were calculated according to both Carbo [72] and Hodgkin [73] for the electrostatic potential and the resulting electric field.

Similarity indices for A+B rings of phorbol and PDBu:

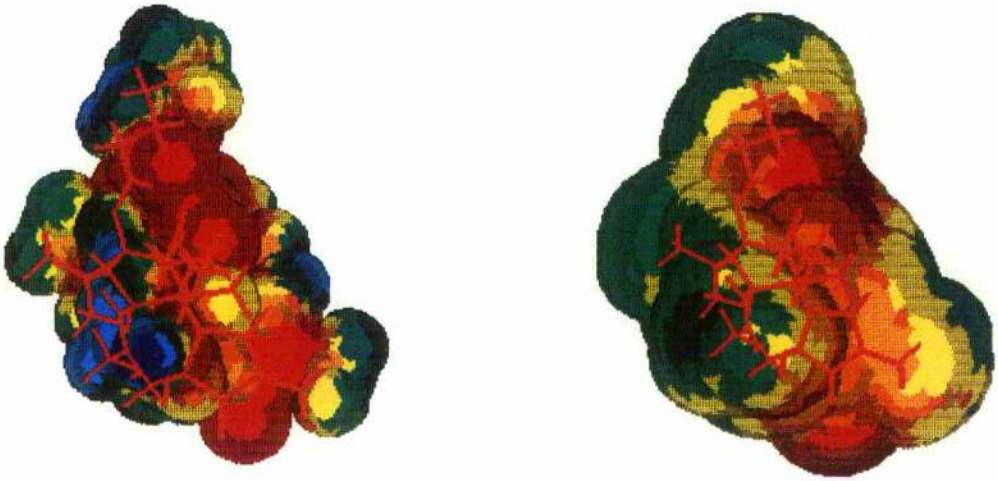
Carbo / Potential	0.909
Hodgkin / Potential	0.903
Carbo / Field	0.954
Hodgkin / Field	0.954

Similarity indices for phorbol and PDBu

Carbo / Potential	0.424
Hodgkin / Potential	0.394
Carbo / Field	0.503
Hodgkin / Field	0.494



a/ Phorbol.



b/ PDBu.

Fig 4.4. EPM's of Phorbol and PDBu.

The structures and ESP's of the two molecules are shown in fig 4.4, and it can be seen that the ring substituents lie in the same orientations. The only difference in the potentials is a slight reduction in the magnitude of the potential for the ester in the vicinity of the 9' hydroxyl group and that produced by the presence of ester groups at 12' and 13'. The similarity indices also show close agreement between the potential and field for the A and B rings, but considerable difference for the complete molecules, due to the large magnitude of the potential at the ester groups. Thus phorbol is a valid model for the ester structure and potential only in the vicinity of the A and B rings, the two molecules diverging considerably at the C and D rings for which phorbol is no longer a valid model. The low calculated similarity of phorbol and PDBu despite the substantial visible similarity of the ESP's over much of the molecule highlights some of the problems involved in the meaningful measurement of the similarity of molecules with corresponding activities but with superficially different structures. When only part of each molecule is involved in specific binding, only those particular parts can be included in the similarity calculations as the remaining functionalities present on the molecules will serve to give a false value for the calculated similarity. Thus calculated similarity indices are only of use when the whole molecule binds to the site, or when enough is known about the nature of the site so as to suggests parts of the molecules actually involved in binding.

4.4.1 Comparison of PDBu and DAG.

There have been several reported comparisons of the structure of DAG with that of phorbol or its ester, claiming to have found good geometric fits of the three DAG functional groups with the phorbol oxygens at 9', 12' and 13' [122], 4', 9' and 20' [123] or 3', 4', 9' and 20' [120, 124] though most are based on the use of space-filling models rather than energy calculations. The approach used here was to produce an initial fit using the conformations calculated in previous sections for DAG and PDBu, comparing both the positions of oxygen atoms and, to a lesser weighting, the

orientations of the acyl chains. This fit was then refined by altering the conformation of DAG to fit that of PDBu as closely as possible and reoptimising with the constraint that the close fit was retained. If the fitting was correct, this energy cost of this second step should be no more than a few kcal mol^{-1}.

Attempts to fit the 9' hydroxyl (PDBu) to the 3' hydroxyl (DAG) and the ester groups of the two molecules gave a mean deviation of 0.529 Å but the positions of the hydrophobic 1', 2' and hydrophilic 20' regions of the phorbol were not consistent with the orientation of the lipid bilayer (as inferred from the DAG conformation), while the 4', 9', 20' fit gave a much higher mean deviation and could be discounted from the activity of the 4' deoxy derivative [112]. The most productive fitting involved the superposition of the 20' hydroxyl (PDBu) with the 3' hydroxyl of DAG which, by applying a low weighting to the fitting of the acyl chains, brought the oxygens at positions 3 and 9 (PDBu) close to the 1' and 2' carbonyl oxygens of DAG respectively (mean deviation 0.398 Å). With the molecules in this orientation, DAG does not occupy the equivalent positions of the 4' group and hydrophobic region of the A-ring (1', 2' and 19') of PDBu.

The conformation of DAG was then altered to minimise the deviation of the hydroxyl and carbonyl oxygens from their equivalents in the PDBu molecule, while the acyl chains were constrained to lie in the same direction as for the fitting, so as to reproduce the orientation of the molecule in a lipid bilayer. This raised the energy of DAG from -306.0 kcal mol^{-1} to -46.619 kcal mol^{-1}, but on optimisation the energy fell to -304.1 kcal mol^{-1} with a mean deviation in the oxygen positions of less than 0.1 Å. Thus the energy cost for DAG to take up a conformation equivalent to the of the phorbol esters is less than 2 kcal mol^{-1}, which is within an acceptable range to allow binding especially as DAG is 10^4 less active than TPA. Thus, from the comparison with DAG, it is proposed that the main interactions between the binding site and the phorbol esters involve a 3-point hydrogen bonding system with the 3' and 9' groups accepting bonds while the 20' hydroxyl can act either as a donor or an acceptor. The electrostatic potentials of phorbol and the corresponding DAG conformation are shown in fig 4.5.

orientations of the acyl chains. This fit was then refined by altering the conformation of DAG to fit that of PDBu as closely as possible and reoptimising with the constraint that the close fit was retained. If the fitting was correct, this energy cost of this second step should be no more than a few kcal mole^{-1}.

Attempts to fit the 9' hydroxyl (PDBu) to the 3' hydroxyl (DAG) and the ester groups of the two molecules gave a mean deviation of 0.529 Å but the positions of the hydrophobic 1', 2' and hydrohpilic 20' regions of the phorbol were not consistent with the orientation of the lipid bilayer (as inferred from the DAG conformation), while the 4', 9', 20' fit gave a much higher mean deviation and could be discounted from the activity of the 4' deoxy derivative [112]. The most productive fitting involved the superposition of the 20' hydroxyl (PDBu) with the 3' hydroxyl of DAG which, by applying a low weighting to the fitting of the acyl chains, brought the oxygens at positions 3 and 9 (PDBu) close to the 1' and 2' carbonyl oxygens of DAG respectively (mean deviation 0.398 Å). With the molecules in this orientation, DAG does not occupy the equivalent positions of the 4' group and hydrophobic region of the A-ring (1', 2' and 19') of PDBu.

The conformation of DAG was then altered to minimise the deviation of the hydroxyl and carbonyl oxygens from their equivalents in the PDBu molecule, while the acyl chains were constrained to lie in the same direction as for the fitting, so as to reproduce the orientation of the molecule in a lipid bilayer. This raised the energy of DAG from -306.0 kcal mole^{-1} to -46.619 kcal mole^{-1}, but on optimisation, the energy fell to -304.1 kcal mole^{-1} with a mean deviation in the oxygen positions of less than 0.1 Å. Thus the energy cost for DAG to take up a conformation equivalent to the of the phorbol esters is less than 2 kcal mole^{-1}, which is within an acceptable range to allow binding especially as DAG is 10^4 less active than TPA. Thus, from the comparison with DAG, it is proposed that the main interactions between the binding site and the phorbol esters involve a 3-point hydrogen bonding system with the 3' and 9' groups accepting bonds while the 20' hydroxyl can act either as a donor or an acceptor. The electrostatic potentials of phorbol and the corresponding DAG conformation are shown in fig 4.5.

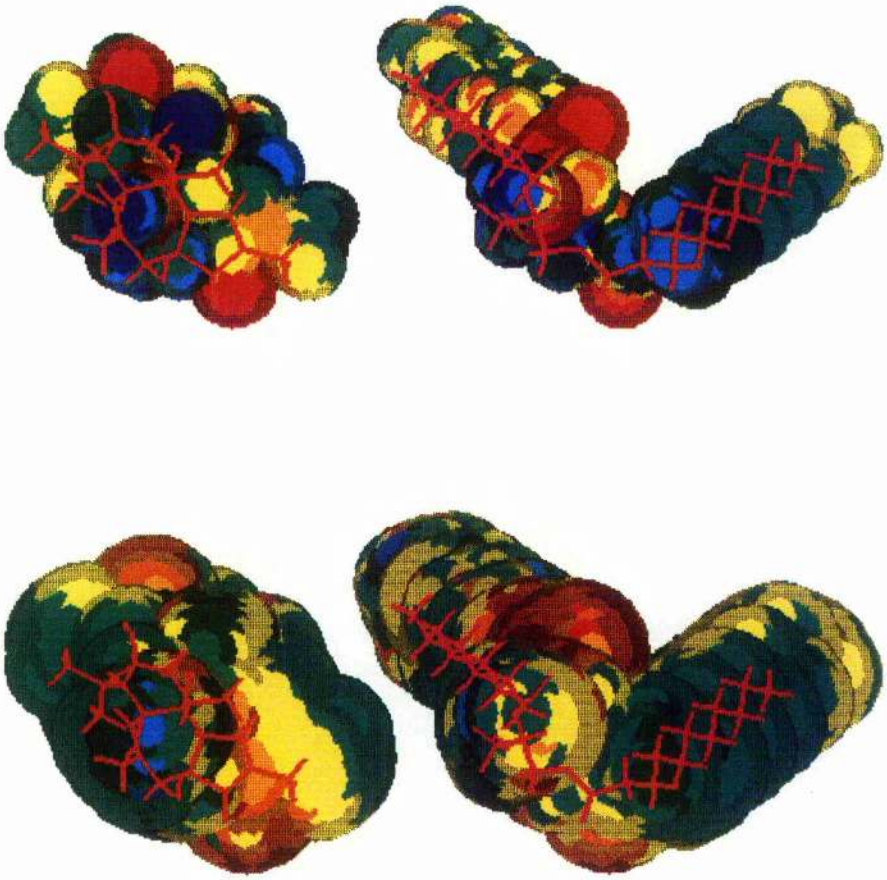


Fig 4.5. EPM's of Phorbol (left) and DAG (right) in comparable orientation after fitting with CHEM-X.

4.4.2 Comparisons of the Structures and ESP's of the derivatives.

The structures of the phorbol derivatives, listed in table 4.1, were calculated using the MNDO, AM1 and PM3 parameter sets using MOPAC. Possible rotations of ring substituents were considered, and the lowest energy obtained for each molecule is shown in table 4.4.

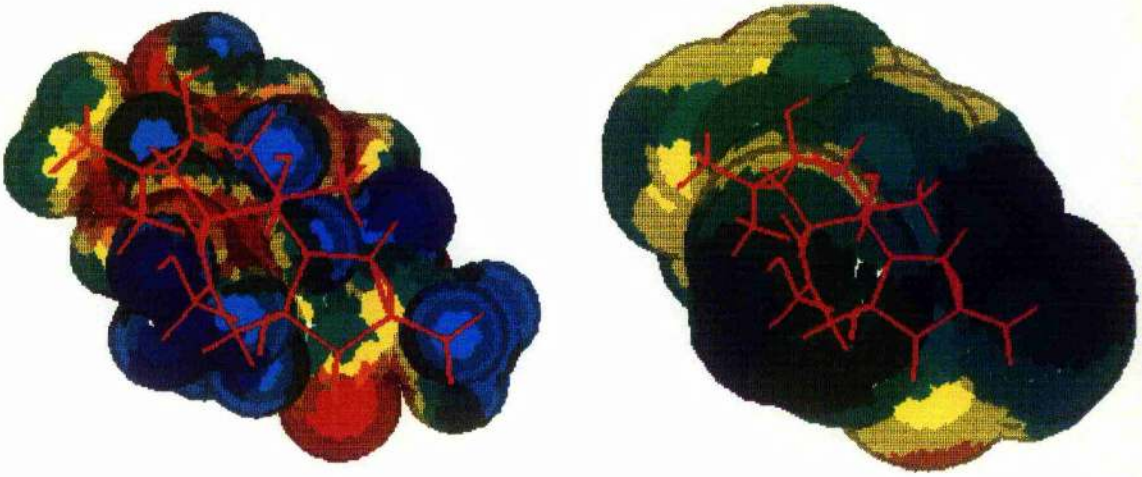
Table 4.4 Heats of Formation (kcal mol^{-1}) of Phorbol Derivatives.

	MNDO	AM1	PM3
Phorbol	-183.7	-214.0	-214.0
1α,2α Epoxy phorbol	-198.7	-225.7	-223.4
1β,2β Epoxy phorbol	-199.8	-227.9	-226.6
1,2α Dihydro phorbol	-203.1	-243.2	-236.2
1,2β Dihydro phorbol	-202.9	-242.2	-236.2
3α Hydroxy phorbol	-194.9	-237.0	-224.5
3β Hydroxy phorbol	-198.5	-240.2	-226.3
4β Methoxy phorbol	-172.4	-207.3	-206.5
4β Deoxy phorbol	-150.5	-177.3	-178.4
5β Hydroxy phorbol	-222.9	-259.3	-252.5
6α,7α Epoxy phorbol	-202.6	-232.6	-230.9
6β,7β Epoxy phorbol	-204.8	-233.4	-232.4
16' Hydroxy phorbol	-223.2	-261.1	-252.3
20' Phorbol aldehyde	-165.1	-197.3	-201.8
20' Phorbol acid	-222.4	-256.2	-256.9

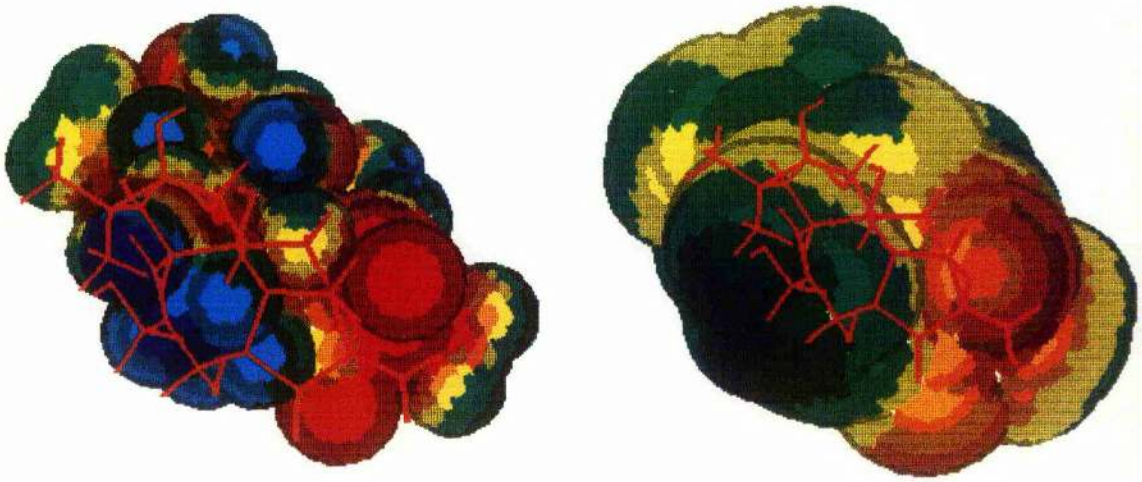
The most surprising results of these heats of formation is the close agreement between the AM1 and PM3 values for many of the derivatives. The reports of the PM3 parametrisation [63, 64] and subsequent discussion by Stewart [115, 125], which claim a 40% increase in the accuracy of PM3 heats of formation over AM1, lead one to expect some difference in the results. That the only difference greater than 2 kcal mol^{-1} is for those derivatives with an extra hydroxyl group and the 1',2' saturated derivative suggests that PM3 is not a significant improvement over AM1, though unfortunately it is not possible to compare the values with experimental results as they have not been determined.

Table 4.5 Similarity Indices for Phorbol Derivatives.

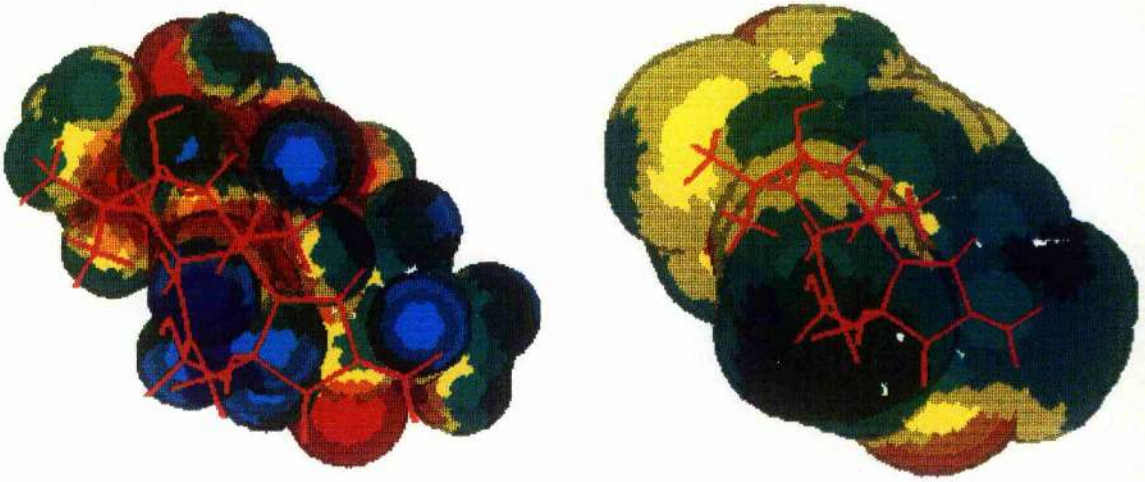
	Activity	Carbo / Potential	Hodgkin / Potential	Carbo / Field	Hodgkin / Field
16' Hydroxy phorbol	++++	0.998	0.997	0.995	0.995
4β Deoxy phorbol		0.913	0.911	0.881	0.880
5β Hydroxy phorbol	+++	0.924	0.924	0.876	0.873
6α,7α Epoxy phorbol		0.793	0.780	0.663	0.660
6β,7β Epoxy phorbol		0.717	0.696	0.617	0.613
3α Hydroxy phorbol	++	0.706	0.704	0.667	0.665
3β Hydroxy phorbol		0.689	0.676	0.703	0.701
20' Phorbol aldehyde	+	0.764	0.761	0.766	0.765
1,2α Dihydro phorbol	None	0.936	0.933	0.927	0.927
4β Methoxy phorbol		0.930	0.928	0.815	0.812
1,2β Dihydro phorbol		0.881	0.881	0.770	0.769
1α,2α Epoxy phorbol		0.777	0.728	0.679	0.676
1β,2β Epoxy phorbol		0.728	0.717	0.791	0.789
20' Phorbol Acid		0.711	0.701	0.688	0.675



a/ 1a Epoxy Phorbol

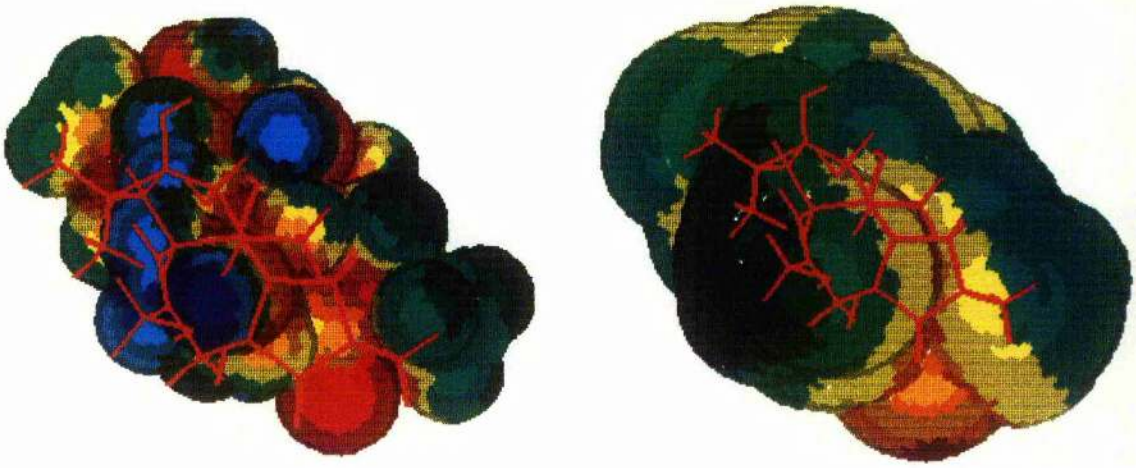


b/ 1b Epoxy Phorbol

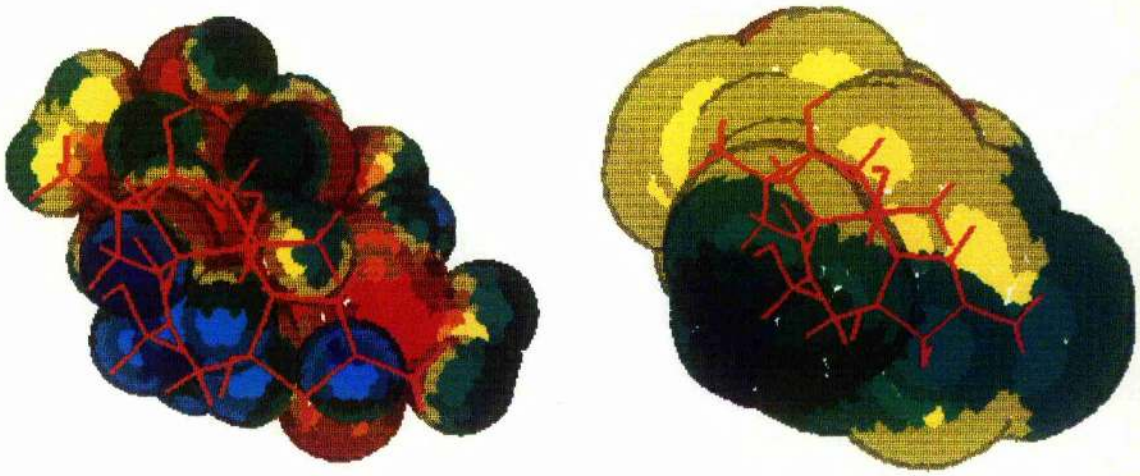


c/ 1a Saturated Phorbol

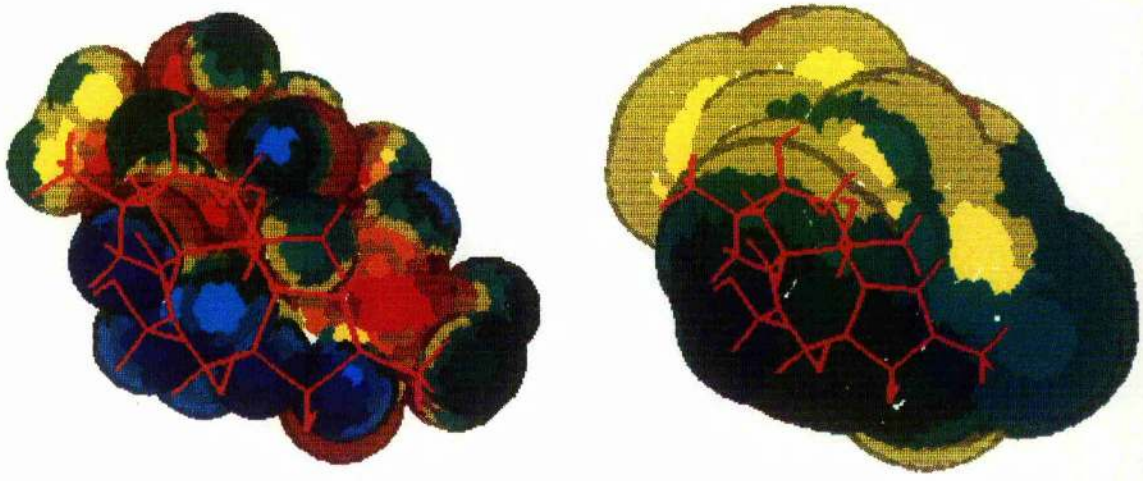
Fig 4.6. EPM's of Phorbol Derivatives.



d/ 1b Saturated Phorbol

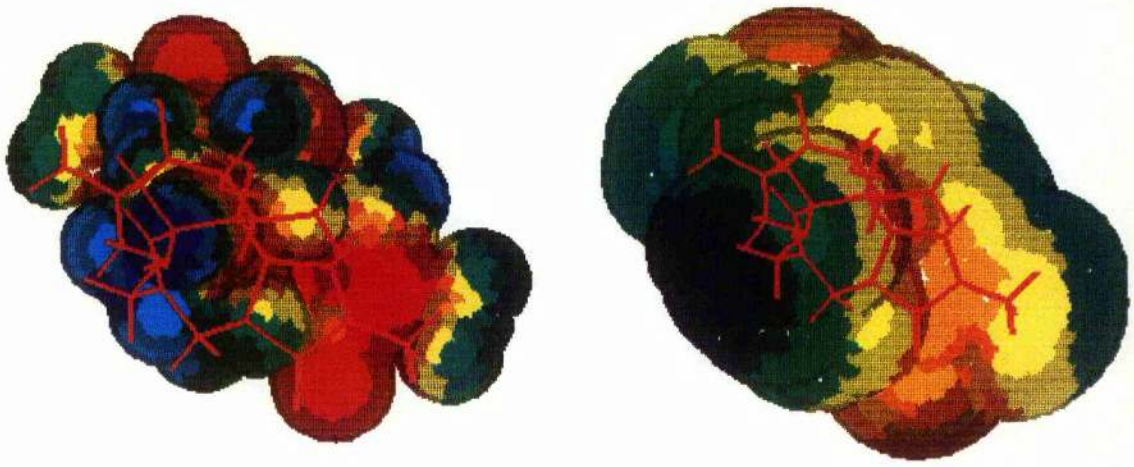


e/ 3a Hydroxy Phorbol

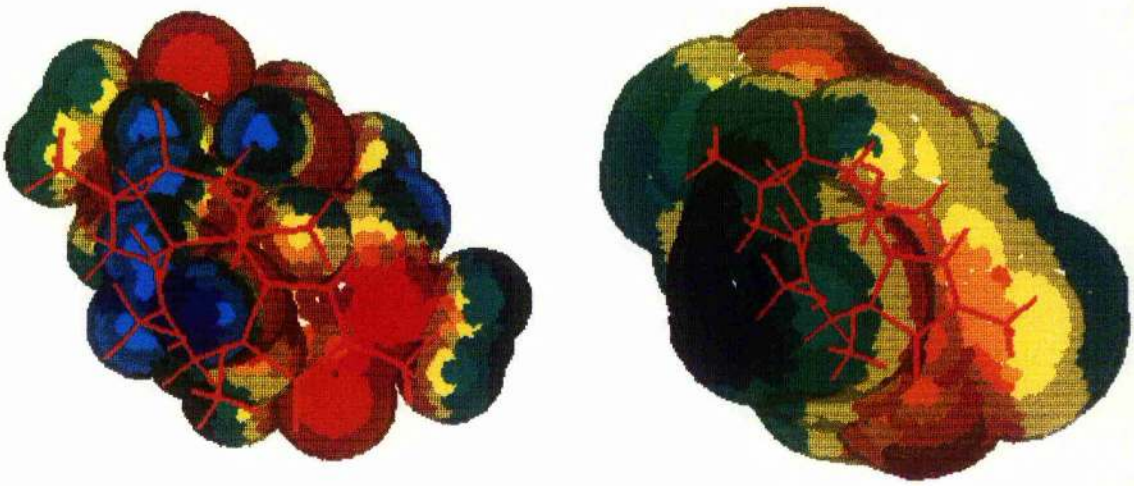


f/ 3b Hydroxy Phorbol

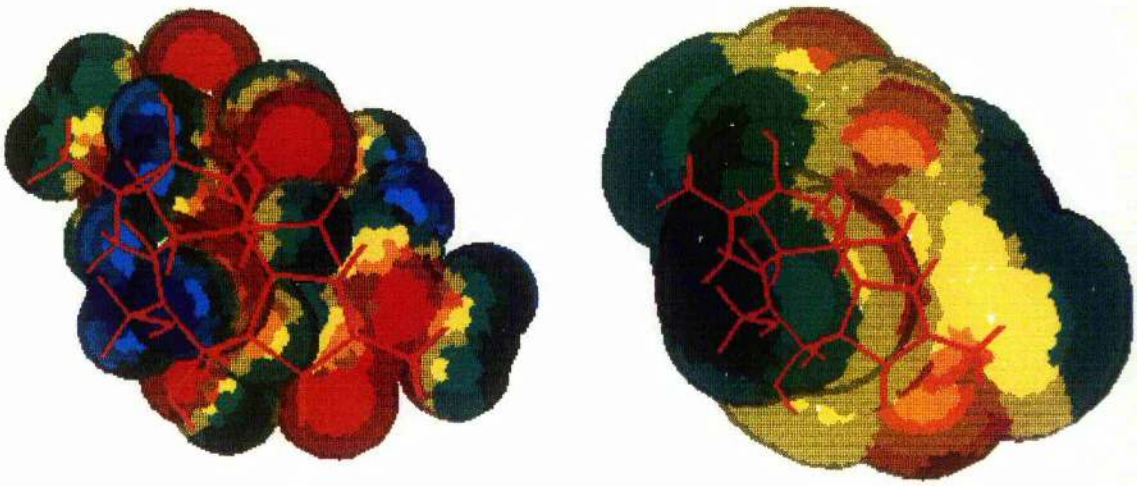
Fig 4.6. (Continued)



g/ 4b Deoxy Phorbol

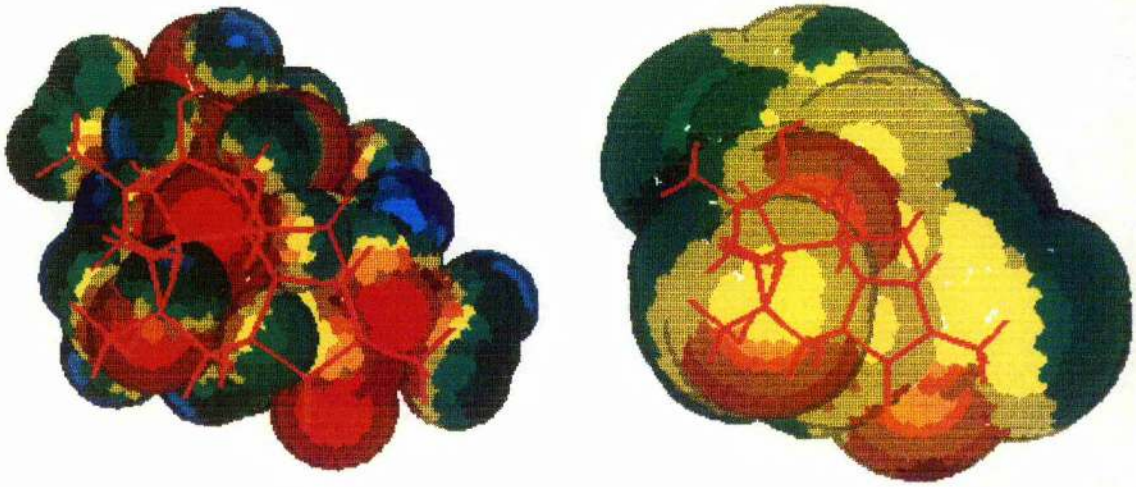


h/ 4bO Methoxy Phorbol

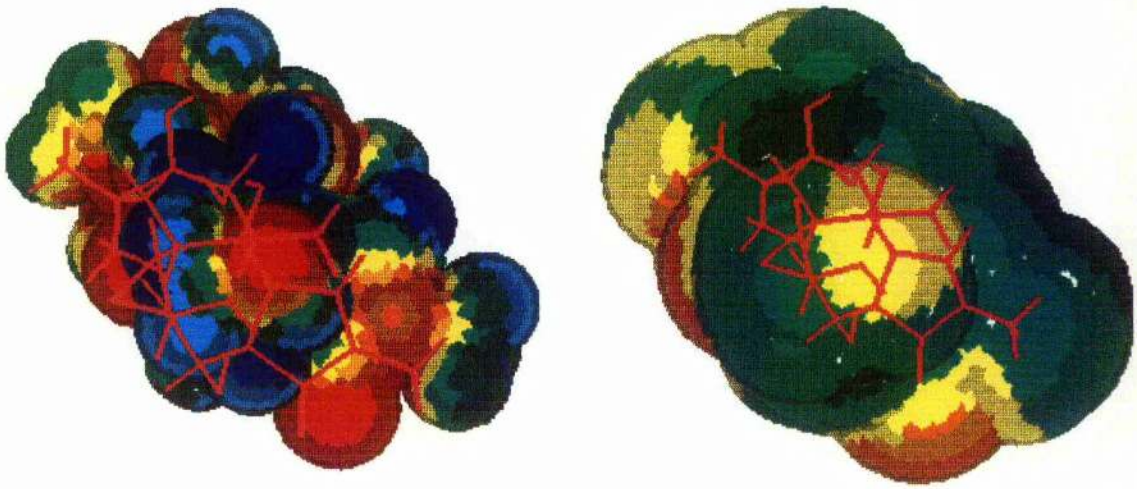


i/ 5b Hydroxy Phorbol

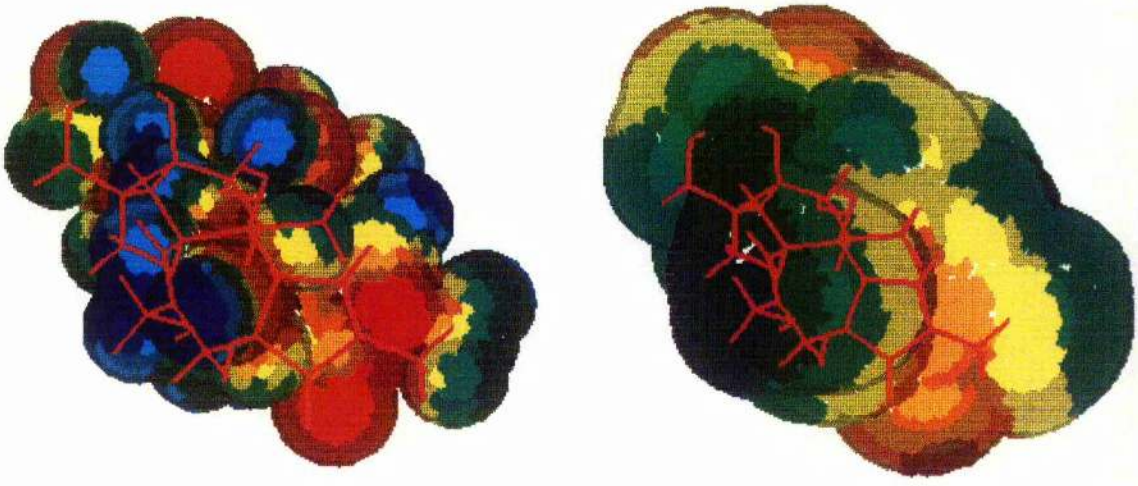
Fig 4.6. (Continued)



j/ 6a,7a Epoxy Phorbol

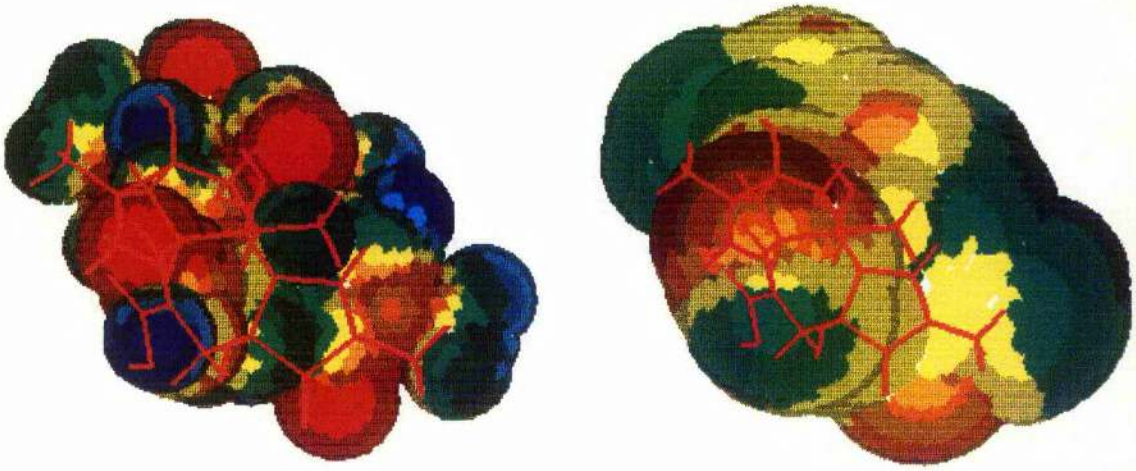


k/ 6b,7b Epoxy Phorbol

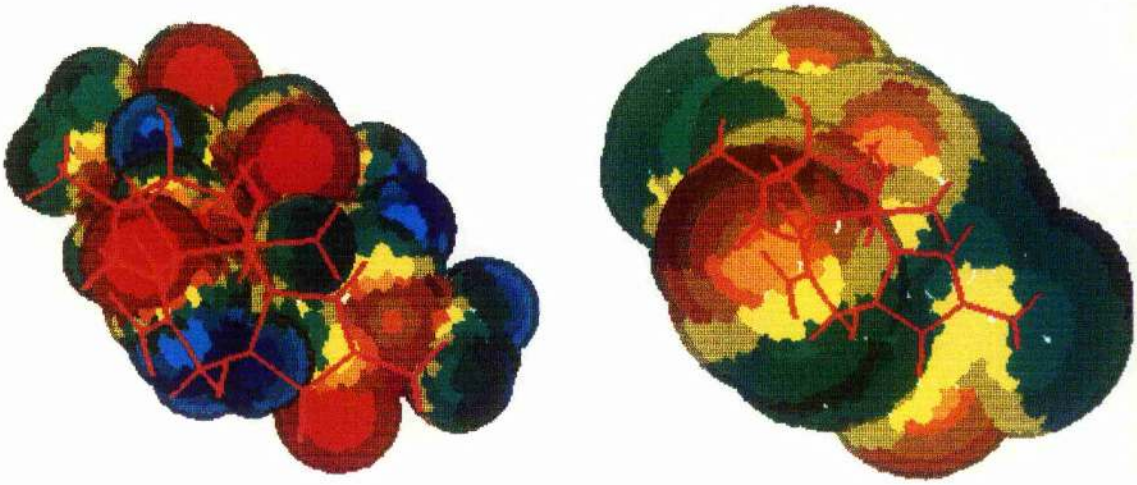


l/ 16′ Hydroxy Phorbol

Fig 4.6. (Continued)



m/ 20′ Phorbol Acid



n/ 20′ Phorbol Aldehyde

Fig 4.6. (Continued)

The AM1 point charges of the derivatives were used to generate electrostatic potentials at both the Van der Waals radius and at twice the radius, shown in fig 4.6. Similarity indices, given in table 4.5, for the A and B rings, compared with that of phorbol, were calculated from these point charges without optimisation by the simplex method [101]. Instead the different molecules were fitted using the procedure in CHEM-X, which gave very small mean deviations for equivalent groups. The similarity indices do not show a good correlation with promoting activity showing ESP is not equally important for binding over the whole two rings. This may be caused by some parts of the rings not being involved in the site, or because an unfavourable steric interaction prevents a molecule with a complementary ESP from entering the binding site. In this particular case it is more useful to consider the detail of the ESP rather than similarity indices over sizeable fragments of the molecule.

The discussion of the ESP's themselves is largely qualitative both because the activity data is also essentially qualitative and because most of the derivative potentials differ from that of phorbol only in the immediate vicinity of the structural change. The potentials shown in fig 4.6 are all drawn to show the 3', 9' and 20' groups with the convention that negative potentials are displayed in red and positive in blue, while the terms 'α' and 'β' are replaced by 'a' and 'b', as the colour printer is unable to produce the Greek characters. At twice the Van der Waals radius, the ESP of phorbol shows very little detail, being dominated by the strongly negative potentials of the 3' carbonyl and the 12', 13' hydroxyls. At the single radius, rather more detail is shown, the polarisation of the 9' and 20' hydroxyls can be seen clearly, while H_1 has a strong positive potential. Apart from these regions, most of the carbon skeleton possesses a neutral potential.

Both of the modifications considered to the A-ring double bond (1',2') lead to complete inactivation of the ester, while structurally both alter the geometry of the A-ring, forcing the 19' methyl group out of the plane of the ring. Epoxidation includes a second strongly electronegative atom into the A-ring and thus produces a substantial effect on its potential. The potential on the same side of the ring becomes more negative, while that on the opposite side, especially H_1, becomes more positive, which

is also accompanied in the β form by an increase in the positive potential on the hydrogen of the 9' hydroxyl. By comparison, saturation of the double bond does not change the ESP of the A-ring to any extent, except to reduce the positive potential at position 1. Thus it appears that the change in A-ring geometry alone is sufficient to abolish activity, rather than being an electrostatic effect.

As the 3' carbonyl was one of the key contacts identified in section 4.4.1, a change in ESP, or modification would be expected to have an effect on activity. Conversion of a hydroxyl group would not abolish its ability to accept a hydrogen bond, but as the positions of the lone pairs are different with the lone pair-oxygen-carbon angle decreasing from 120° to nearer 109° typical of tetragonal geometries, the energy of such an interaction would be altered. The rotation of the hydroxyl group is also restricted by the presence of neighbouring groups so that there is only one minimum energy position for the β form, while those of the α form lie close together and in neither case would the lone pairs be expected to lie in the plane of the A-ring. That both forms have comparable activities is in keeping with the absence of change to other parts of the ESP and suggests that the optimum positions of the lone pairs for hydrogen bonding lies roughly midway between the positions which, from the conformations shown in fig 4.6, is virtually in the plane of the ring.

Some earlier studies [120, 123, 124] suggested that the 4' hydroxyl was one of the key hydrogen bonding contacts with the site, but this view is not supported by the results of the fitting to DAG, section 4.4.2, and by the activity data [112] which shows that the deoxy derivative possesses full activity. The potentials of both the 4' deoxy and methoxy derivatives are virtually identical, and are similar to that of phorbol, thus the difference in activity cannot be due to an electrostatic effect. Rather, the larger methoxy group projects further from the rest of the molecule than does either the hydroxyl group or hydrogen atom, and thus the reason for the inactivity of the methoxy derivative lies in the fact that it is unable to fit within the site. This requires the presence of site groups close to the 4' position, though little can be said about the potential except that it clearly does not interact strongly with either the hydroxyl group or hydrogen atom, and in view of the proximity of the 3' hydrogen bonding contact,

this site group may well be part of the hydrogen bond donor. Rotation of the 4' substituent is severely restricted (barrier height >200 kcal mol^{-1}) by the presence of the A and B rings and results in the oxygen substituents pointing almost directly away from the rest of the molecule, so the effect of the extra size of the methyl group is exaggerated, and it is simultaneously prevented from rotation to relieve the unfavourable steric interaction. The 5' hydroxyl group also projects from the molecule in a position close to that of the 4' substituent, but the derivative retains almost complete activity. Like the 4' derivatives, there is little change in the ESP except at the site of modification itself, but the 5' group is not constrained by the rest of the molecule and is free to rotate. As it is also further from the 3' contact, it would appear to lie further from the site groups and not form an unfavourable steric interaction.

Epoxidation of the B-ring double bond (6',7') as for its A-ring counterpart alters both the geometry of the ring and its ESP. However, unlike the corresponding A-ring modification, it does not abolish activity, merely reducing it by a small extent. As with the A-ring, the epoxide group causes the ring to acquire a negative potential on the same side and a positive potential on the opposite side of the ring to the substitution, with the effect that the α form has a region of negative potential between the 9' and 20' hydrogen bonding contacts, while the equivalent region in the β form has a positive potential. As both forms possess a high activity, this region cannot be involved in binding to the site, except perhaps as part of a non-specific polar interaction. Though the geometry of the ring, including the position of the 20' hydroxyl relative to the other hydrogen bonding contacts, is also altered, the flexibility of the 20' group is such that the energy cost on changing the conformation at 6' and 20' to optimise the hydrogen bond would be quite small, and apparent in the small reduction in activity.

Oxidation of the 20' hydroxyl to either the aldehyde or acid reduces the promoting activity considerably, but without changing the ESP except at position 20. The almost complete inactivity of the aldehyde is consistent with the 20' group being a hydrogen bond donor. This derivative is still able to form the other hydrogen bonds to the site and would not suffer from the unfavourable steric interactions in the manner of the 4b methoxy derivative and derivatives with a saturated 1',2' bond, so should still be

able to bind to the site though with a much lower affinity. Its low level of activity can be compared with that of the 3' hydroxyl derivatives, where the hydrogen bond is weakened rather than being lost completely. The acid derivative should be able to both accept and donate hydrogen bonds unless it is deprotonated, which is unlikely at physiological pH, and though the orientation of the hydroxyl group is the opposite of that of phorbol, the difference between the two acid conformations is only 0.04 kcal mol^{-1}. Thus the lack of promoting activity of the acid is more likely to stem from an unfavourable interaction with the binding site. The potential at the acid carbonyl oxygen is the most negative found for the phorbol derivatives, so will be repelled by the hydrogen bond acceptor within the site if it too possess a sizeable negative potential.

The only modification outside the A and B rings to be considered was that of hydroxylation at position 16. The calculated similarity indices show that the A and B rings are essentially identical to those of phorbol which is consistent with their identical activities. This is not sufficient proof that the interactions involve solely these two rings as this is the assumption underlying the choice of model.

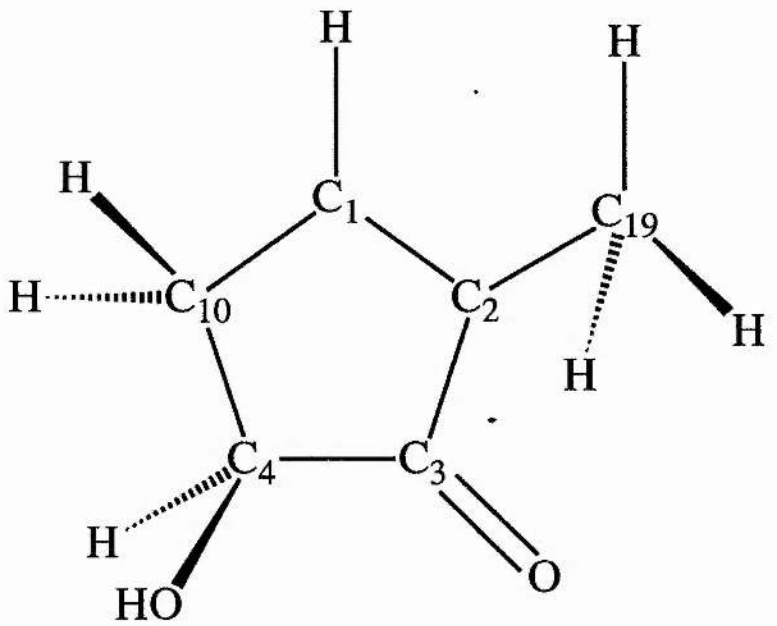
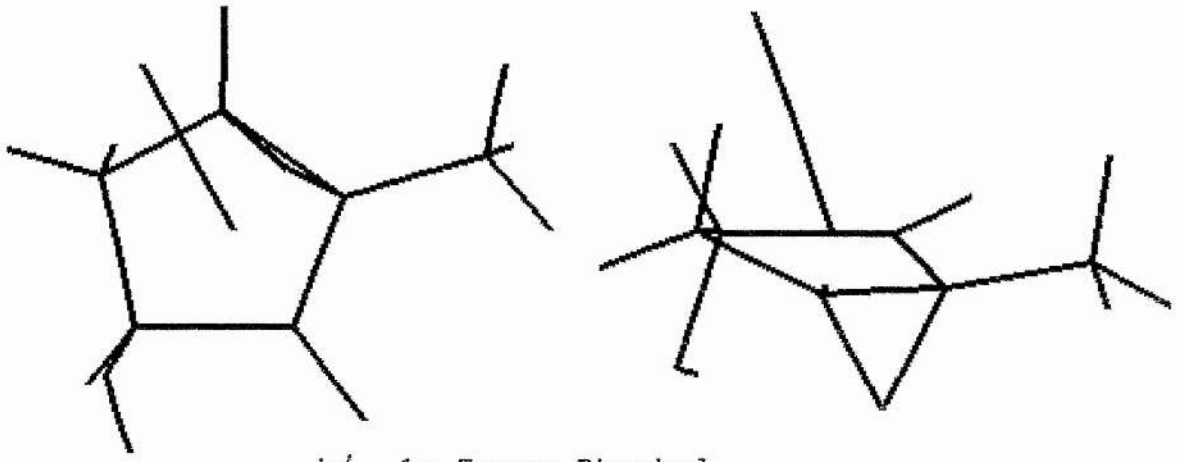
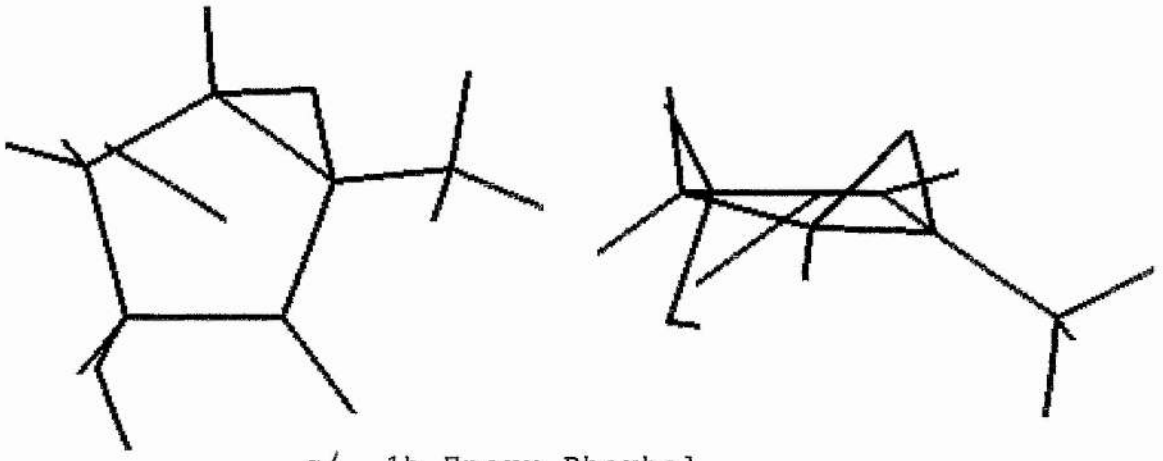


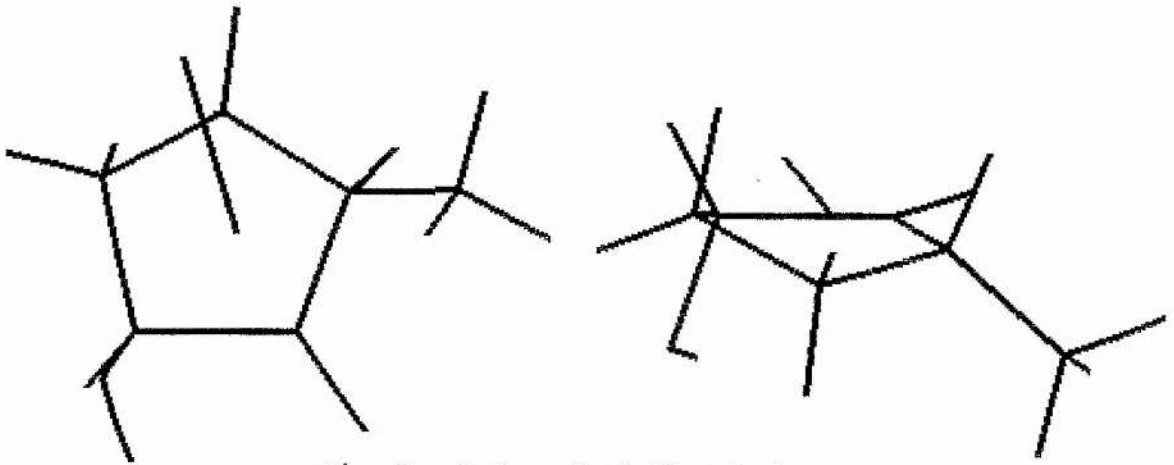
Fig 4.7a Orientation of A-ring in Following Diagrams.



b/ 1a Epoxy Phorbol

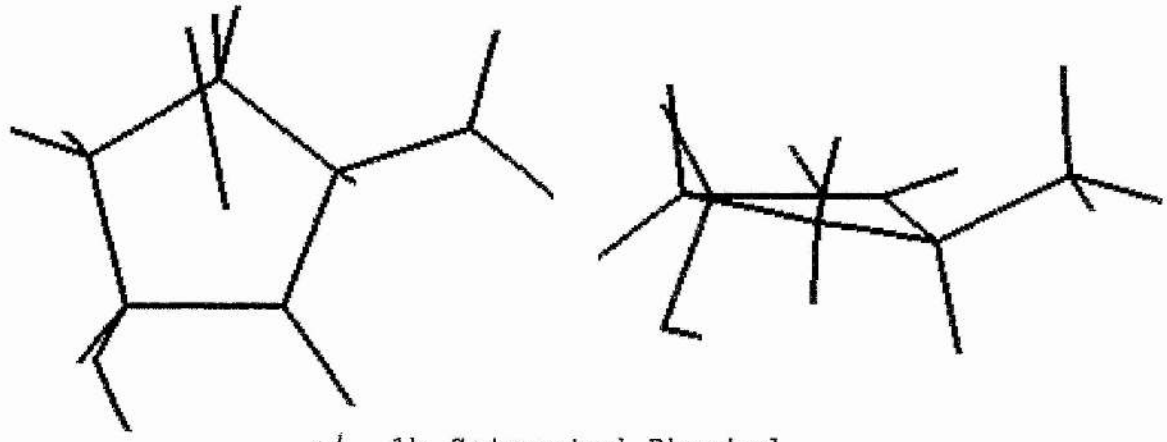


c/ 1b Epoxy Phorbol

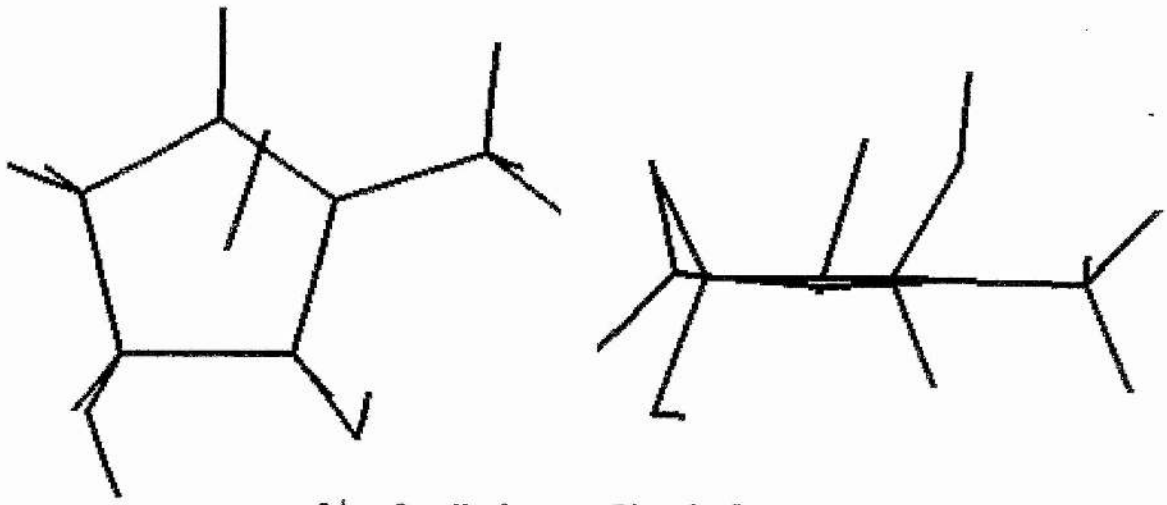


d/ 1a Saturated Phorbol

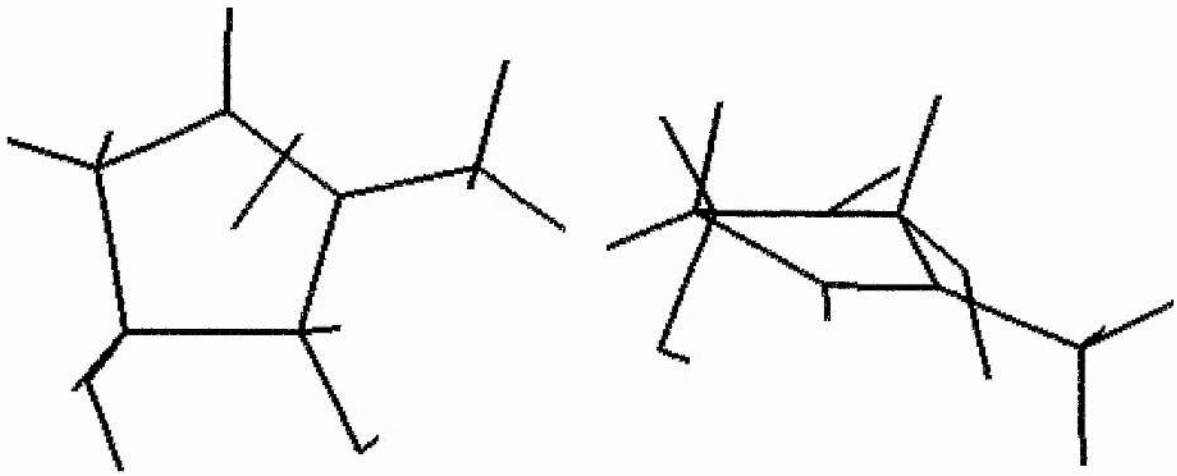
Fig 4.7. Direction of A-Ring Dipole Moments.



e/ 1b Saturated Phorbol

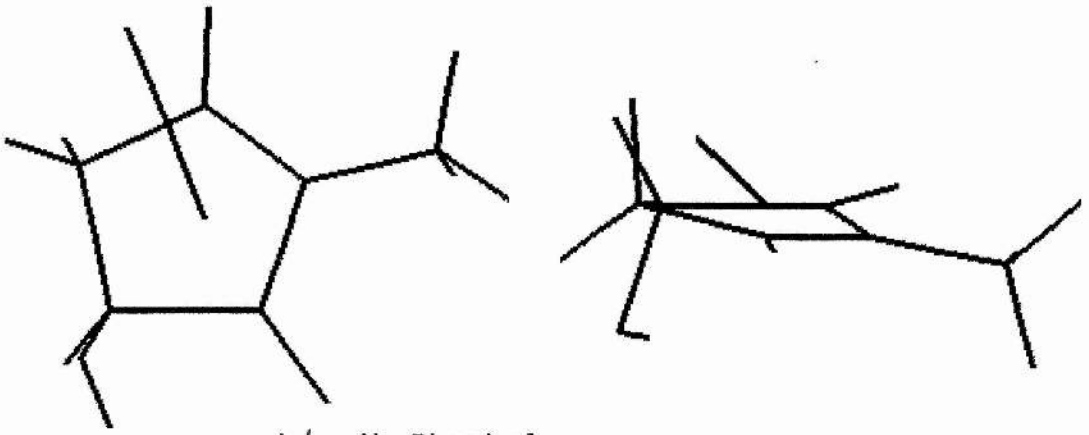


f/ 3a Hydroxy Phorbol

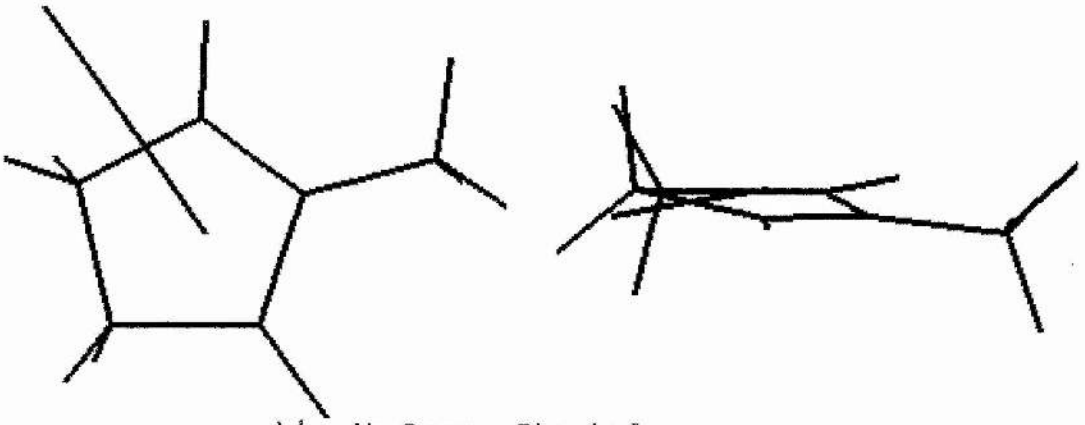


g/ 3b Hydroxy Phorbol

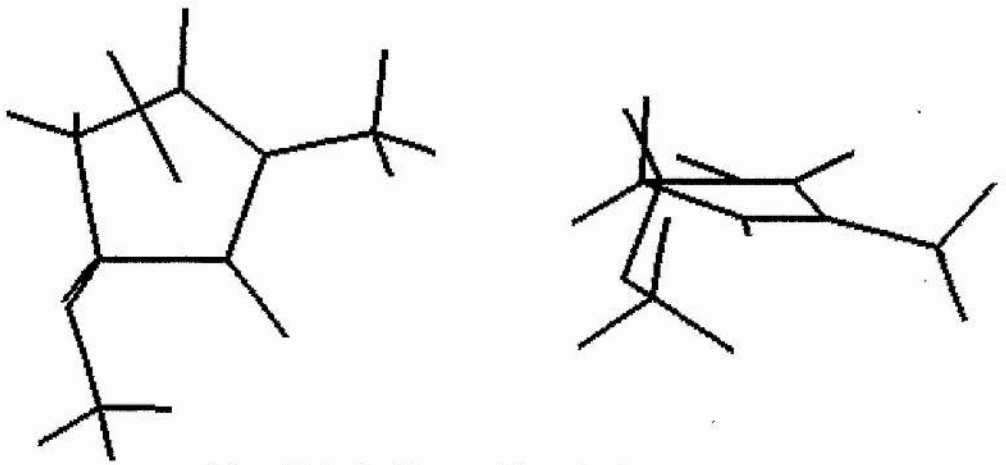
Fig 4.7. (Continued)



h/ 4b Phorbol



i/ 4b Deoxy Phorbol



j/ 4bO Methoxy Phorbol

Fig 4.7. (Continued)

The plotted electrostatic potentials show a dipole moment across the A-ring, the magnitude and direction of which was calculated from an AM1 single point calculation on the A-ring fragments using the geometries found in the parent molecules. The magnitudes are given in table 4.6 and directions in fig 4.7, while the potential derived charges for H_{20} are included in table 4.6.

Table 4.6 Activity vs. Dipole Moments.

	A-Ring	In plane of A-ring	Whole molecule	H_{20} charge
Phorbol	2.275	2.168	3.188	0.322
4' Deoxy	3.016	2.999	1.235	0.324
16' Hydroxy	-[1]	-	2.583	0.322
5β Hydroxy	-	-	1.026	0.314
6α,7α Epoxy	-	-	1.751	0.312
6β,7β Epoxy	-	-	2.162	0.337
3α Hydroxy	1.747	1.173	4.037	0.320
3β Hydroxy	1.162	1.084	4.169	0.320
20' Aldehyde	-	-	3.101	-0.035
1,2α Saturated	1.878	1.795	3.618	0.324
1,2β Saturated	1.914	1.859	3.550	0.326
1α,2α Epoxy	2.883	1.811	4.913	0.311
1β,2β Epoxy	1.751	1.474	2.303	0.327
4β Methoxy	1.859	1.831	1.559	0.318
20' Acid	-	-	2.730	0.109

1 The entries marked with a hyphen do not differ from phorbol in the A ring so the dipole moment was not calculated.

Charges for the oxygen atoms as obtained from Mulliken analysis do not vary significantly between derivatives, but the charges obtained from the wavefunction ESP for a small fragment of the molecules have been compared below. The ESP's for the fragments have been calculated using a minimal basis (STO-3G) as a basis with more primitive Gaussian functions requires more disc space than was available.

Table 4.7 Activity vs. Oxygen Charges

	O_3	O_4	O_9	O_{20}
Phorbol	-0.397	-0.523	-0.574	-0.543
4' Deoxy	-0.402	-	-0.556	-0.545
16' Hydroxy	-0.397	-0.523	-0.574	-0.543
5β Hydroxy	-0.410	-0.486	-0.588	-0.523
6α,7α Epoxy	-0.390	-0.499	-0.571	-0.490
6β,7β Epoxy	-0.396	-0.504	-0.576	-0.536
3α Hydroxy	-0.567	-0.569	-0.570	-0.541
3β Hydroxy	-0.591	-0.588	-0.574	-0.540
20' Aldehyde	-0.394	-0.514	-0.573	-0.316
1,2α Saturated	-0.390	-0.562	-0.576	-0.538
1,2β Saturated	-0.383	-0.552	-0.560	-0.542
1α,2α Epoxy	-0.386	-0.556	-0.605	-0.518
1β,2β Epoxy	-0.391	-0.489	-0.560	-0.546
4β Methoxy	-0.394	-0.293	-0.569	-0.524
20' Acid	-0.409	-0.491	-0.557	-0.628

It is noticeable, from fig 4.7, that the dipole moments for both phorbol and 4' deoxy phorbol lie almost on the plane of the A ring, while for 3β hydroxy phorbol and the inactive 1,2' epoxides the dipole moment has a substantial component perpendicular to the ring. The dipole moment for 3α hydroxy phorbol points between C_1 and C_2 rather than between C_1 and C_{10} as is the case with phorbol. The A-ring dipole moments of the 1,2 saturated derivatives differ from that of phorbol only in being slightly smaller, while the dipole moment of the 4' methoxy derivative is, not surprisingly, almost identical. These differences could account for some of the differences in activity as the direction of the dipole moment will have some influence on the orientation of the A-ring on its approach to the site, while the magnitude will contribute to the binding energy. By comparison, the dipole moments for the whole molecule are not useful as a guide to activity. Though many of the low activity and inactive derivatives possess dipole moments that differ from that of phorbol, the magnitude of the dipole moment is very sensitive to rotation of key groups, for example rotation of the 20' aldehyde changes the magnitude of the dipole moments from 3.101 Debye to 3.935 and a change on a similar scale can be seen for a rotation of the 16' hydroxyl group, without affecting the activity.

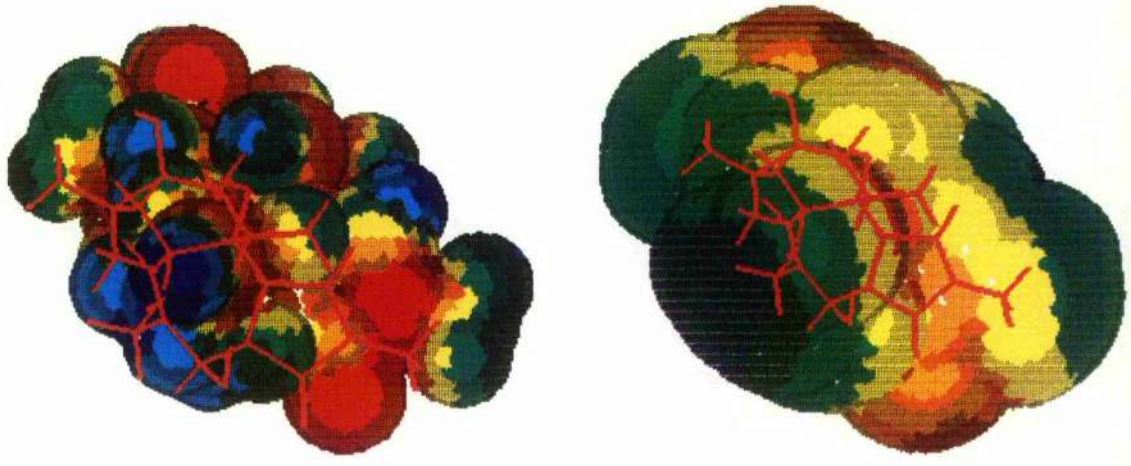
It is difficult also to draw any conclusions from the oxygen charges beyond the observation that deviation from the charges found for phorbol is associated with a reduction in activity. None of the derivatives shows a change in the O_9 charge and its inclusion as one of the three hydrogen bonding contacts is based solely on the comparison with DAG. The 20' hydrogen of the aldehyde carries a positive charge when the charges are calculated using Mulliken analysis on the semi-empirical wavefunction, but direct calculation of the ESP from the wavefunction gives it a small negative charge. It is noticeable that at twice the Van der Waals radius, the potential over the hydrogen atom is completely dominated by the oxygen charge even when the potential is calculated from the Mulliken charges, so this difference should not alter the conclusions drawn from the ESP's.

4.4.3 Parameter Dependency of Results.

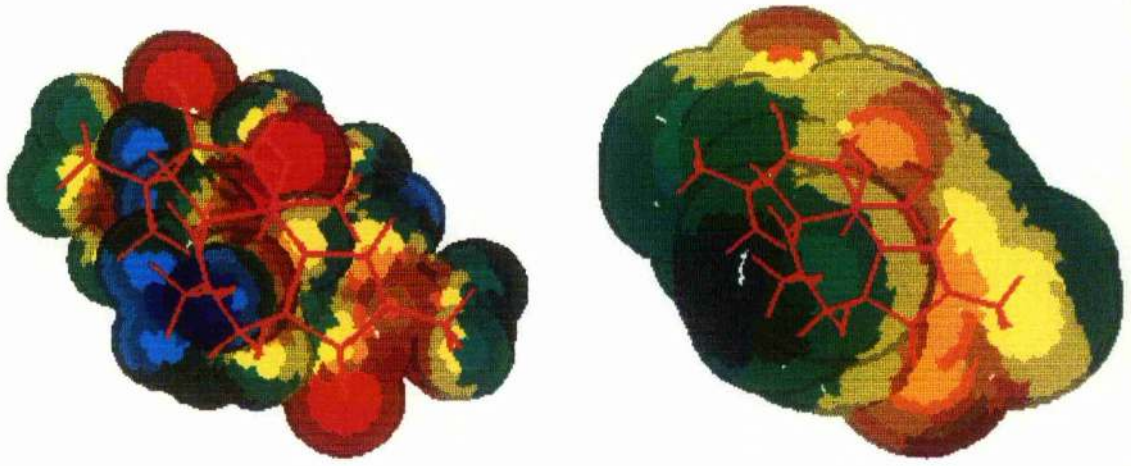
The ESP's, charges and dipole moments discussed above are calculated using the AM1 parametrisation for the reasons mentioned earlier. In this section, some of the results obtained using the three different parameter sets are compared with each other and, where possible, with experimental data.

The phorbol ESP's calculated using each of the three parameter sets are shown in fig 4.8 and are remarkably similar. The potentials are calculated at the minimum energy geometries for each method yet the conformations differ only in the rotation of the 9' hydroxyl (and some small rotations elsewhere), which is able to freely rotate over a limited range before coming into close contact with the neighbouring ring systems. The potentials themselves differ only in that the AM1 potential is of a slightly higher magnitude over much of the molecule, but without significant changes of sign. As all three parameter sets use a well balanced minimal basis set, and have been tested and parametrised using a wide range of carbon, oxygen and hydrogen systems, the Mulliken analysis would be expected to give similar atomic charges for all three methods.

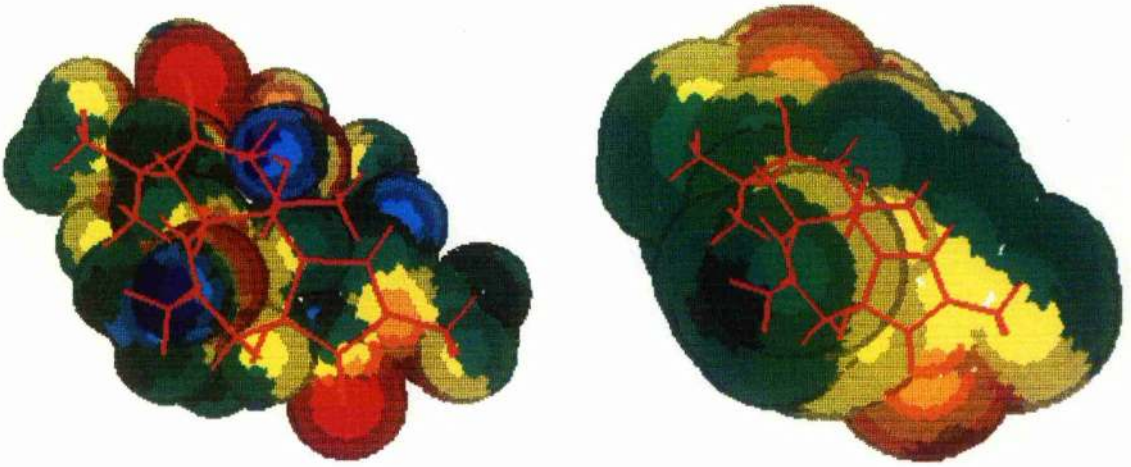
The only experimental data for the phorbols, apart from tumour promoting activities, is the crystal structure [116]. In table 4.8, below, the deviations between the crystal structure and the optimised geometries are compared. AM1 and PM3 give equally good results for the bond lengths, but MNDO is rather less accurate. It is noticeable however that MNDO consistently overestimated bond lengths especially for bonds within the ring systems, as is shown by the signed mean being rather large compared to the unsigned mean. The calculation of the bond angles is rather different as all three methods give similar accuracy with MNDO being most accurate. The signed mean is small in all cases showing that there are no systematic errors for the angles considered, but as the molecules are both rigid and in places strained, there is little scope for large variation in the bond angles. As the hydrogen atoms are not placed correctly in the crystal structure, hydrogen containing bonds and angles were not considered in the analysis.



a/ AM1



b/ PM3



c/ MNDO

Fig 4.8. Phorbol EPM's using AM1, PM3 and MNDO.

Table 4.8 Mean Deviations of Calculated Structures from Experimental Values.

	AM1	PM3	MNDO
Bond Length Errors (in Angstrom)			
Unsigned Mean	0.01886	0.02003	0.03259
RMS	0.02106	0.02294	0.03979
Signed Mean	-0.00603	-0.00217	+0.01741
Bond Angle Errors(in degrees)			
Unsigned Mean	1.255	1.345	0.991
RMS	1.532	1.797	1.365
Signed Mean	-0.447	-0.383	-0.277

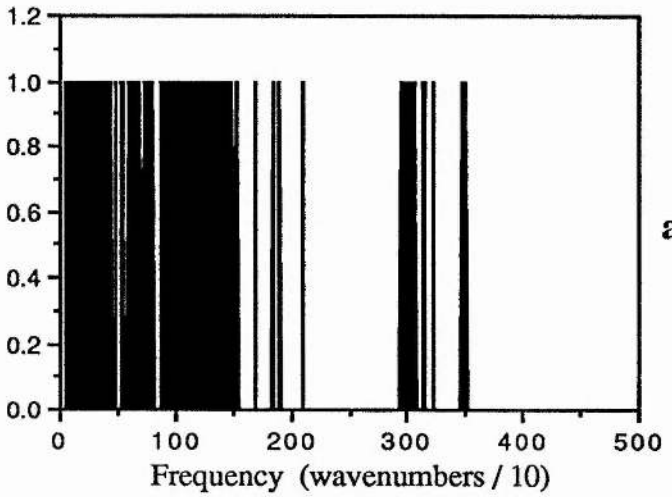
Though the I.R. spectrum of phorbol has not been obtained, it is possible to compare the calculated I.R. spectra, shown in fig 4.9, with typical experimental values. As the frequencies of angle vibrations are subject to the greatest variation, due to changes in the reduced mass arising from substitutions, they are not considered. Frequencies of bond stretching, especially involving hydrogen or terminal atoms, are confined within relatively small ranges, though for long chains and ring systems, carbon-carbon stretching frequencies can vary widely. Experimental values were taken from the "Aldrich Library of I.R. Spectra" [126] and are compared with the calculated values below in table 4.9.

Of the three methods, AM1 gives values that agree most closely with the experimental frequencies while MNDO consistently overestimates the stretching frequencies. The only bond for which the MNDO frequencies coincide with the experimental values is the C-C bond where the range of experimental values is very large. The most surprising results are those of the C=C and C=O bonds for which all three methods give very similar results and all overestimate the frequency considerably. These results suggest that PM3 is not an improvement over AM1 for this sort of

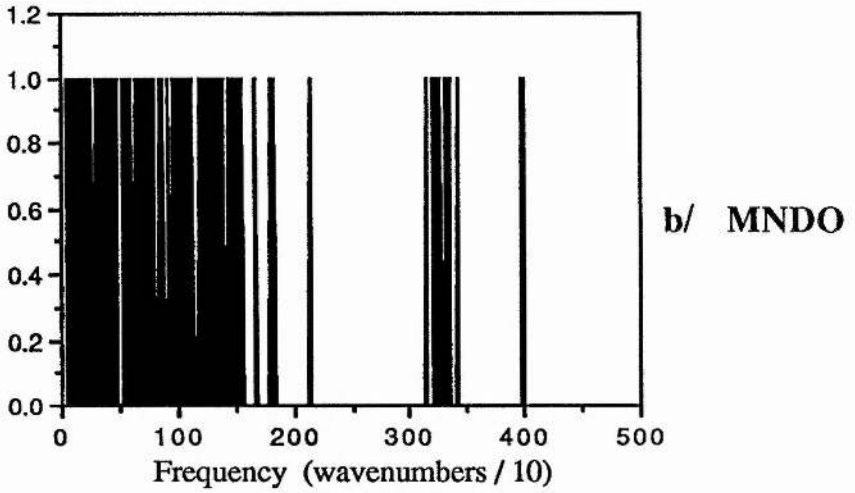
Table 4.9 Phorbol Bond Stretching Frequencies.

Bond	AM1	MNDO	PM3	Experimental
O-H	3466-3504	3984-4007	3842-3908	3335-3635
C-H	2944-3226	3149-3429	2771-3174	2855-2940
C-O	1175-1453	1320-1511	1103-1367	1000-1250
C=O	2106	2150	2009	1670-1750
C-C	817-1697	778-1677	531-1604	910-1250
C=C	1847-1902	1801-1833	1856-1892	1450-1695

Frequencies are in wavenumbers (cm^{-1}).



a/ AM1



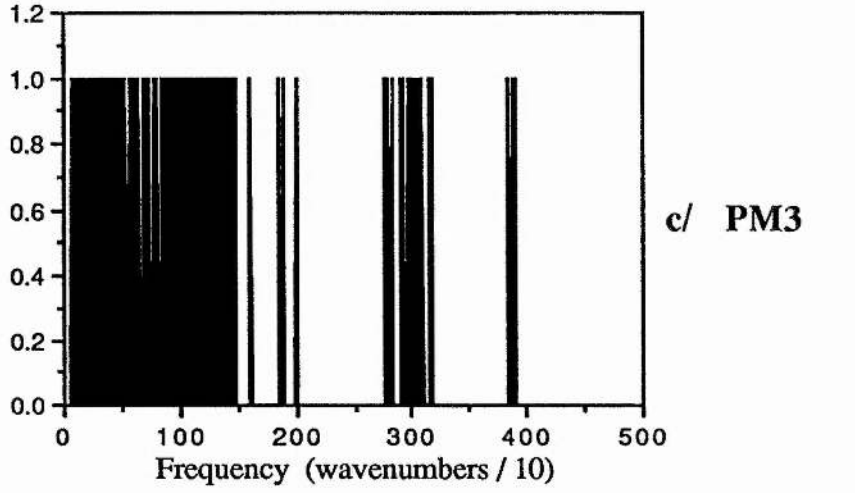


Fig 4.9 Phorbol I.R. Spectra.

Tables 4.10 and 4.11 give summaries of some of the phorbol properties calculated with the different parameter sets. There is however, no experimental data with which they can be compared.

With the exception of the dipole moments, the three methods give largely similar results for the energies and thermodynamic properties. The discrepancy in the dipole moments is rather surprising as this forms part of the experimental data used in the parametrisations [45, 62-64] and so for molecules containing only carbon, hydrogen and oxygen, it would be expected to be both consistent between the methods

and accurate, but it clearly is not.

Table 4.10 Calculated Phorbol Properties.

	AM1	MNDO	PM3
Total Energy (eV)	-4868.3	-4885.5	-4562.9
Ionisation Potential (eV)	9.598	9.952	9.697
HOMO Energy (eV)	-9.598	-9.952	-9.697
LUMO Energy (eV)	-0.076	-0.229	-0.133
Dipole Moment (Debye)	3.349	1.462	1.604
Enthalpy (kcal mol^{-1} at 300 K)	17.441	16.937	17.353
Heat Capacity (kcal K^{-1} mol^{-1} at 300 K)	0.1029	0.1027	0.1063
Entropy (kcal K^{-1} mol^{-1} at 300 K)	175.24	169.54	172.05

By comparison, the HOMO's and LUMO's, significant eigenvalues of which are shown in table 4.11, are consistent in both energy and location between the methods. The eigenvectors for both orbitals are normalised and the contributions from each of the p orbitals have been summed together. In each case only the atomic orbitals with significant eigenvalues have been included so as to show the areas of the molecules most susceptible to electrophilic (HOMO) and nucleophilic (LUMO) attack. The location of the LUMO shows that the molecule accepts electrons most readily on the A-ring, particularly in the region of the double bond. This is particularly interesting as modifications to this region have been shown to abolish activity. In order to investigate this further, the LUMO of the 1,2β saturated derivative was calculated. In this case, the LUMO was centred on both C3, as for phorbol, but also on the B-ring

double bond rather than its A-ring equivalent, thus implying that there is an electron donor in the site that interacts with the A-ring. By comparison, the HOMO is delocalised over a large proportion of the A and B rings covering almost the entire region involved in binding, thus it reveals nothing of the nature of the site.

Table 4.11 Location of the Phorbol HOMO and LUMO.

a/ AM1

HOMO			LUMO		
Atom	Orbital	Eigenvector	Atom	Orbital	Eigenvector
C1	p	0.374	C1	p	0.604
C2	p	0.448	C2	p	0.392
O3	p	0.179	C3	p	0.489
H5	s	0.103	O3	p	0.402
C6	p	0.347	H10	p	0.115
C7	p	0.334			
C8	p	0.118			
H8	s	0.114			
C9	p	0.214			
C10	p	0.207			
C15	p	0.114			
C19	p	0.134			
H19	s	0.159			

b/ MNDO

HOMO			LUMO		
Atom	Orbital	Eigenvector	Atom	Orbital	Eigenvector
C1	p	0.321	C1	p	0.636
C2	p	0.371	C2	p	0.452
O3	p	0.160	C3	p	0.430
H5	s	0.118	O3	p	0.371
C6	p	0.455	H10	p	0.120
C7	p	0.431			
C8	p	0.119			
H8	s	0.112			
C9	p	0.190			
O9	p	0.152			
C10	p	0.157			
H10	s	0.115			
C15	p	0.124			

c/ PM3

HOMO			LUMO		
Atom	Orbital	Eigenvector	Atom	Orbital	Eigenvector
C1	p	0.341	C1	p	0.603
C2	p	0.382	C2	p	0.383
O3	p	0.147	C3	p	0.480
H5	s	0.130	O3	p	0.404
C6	p	0.459			
C7	p	0.451			
H8	s	0.119			
C9	p	0.178			
C10	p	0.177			
H10	s	0.138			
C19	p	0.113			
H19a	s	0.120			
H19b	s	0.111			

4.4.4 A Model Binding Site.

The major contribution to the binding energy is a three-point hydrogen bonding system with phorbol accepting bonds at position 3 and 9 while donating one from position 20 (DAG accepts bonds at the two ester carbonyls and donates from the 3' hydroxyl group). The site group that accepts the hydrogen bond from position 20 is believed to carry a negative charge, so preventing the acid derivative from binding, while little can be said about the site hydrogen bond donors, except that the donor for position 3 (phorbol) lies in the same plane as the bound phorbol A-ring. This same donor is believed to lie close to the phorbol 4' substituent and is responsible for the

unfavourable steric interaction with the 4' methoxy derivative. It is possible that the donor for the 3' bond lies of the other side of the phorbol group and that this steric block is caused by another group within the site. The A-ring appears to have a further contribution to the binding energy, highlighted by the abolition of activity on changing the position of the 19' methyl group, the location of the LUMO and the effects of different orientations of the A-ring dipole, though the nature of this contribution is not clear from the phorbol structure / activity data, except that the LUMO suggests that the site group is electron rich. The optimium fit of DAG to the phorbol structure enables DAG to mimic the hydrogen bonding network almost exactly at a minimal energy cost. There is also no conflict with the steric block opposed to the phorbol 4' groups, but much of the A-ring site is vacant, except that the 1' acyl chain occupies the same region of space as does the hydrophobic C19 of phorbol. Fig 4.10 shows the suggested positions of the site groups and phorbol bound to the site, though for clarity only the A and B rings are shown.

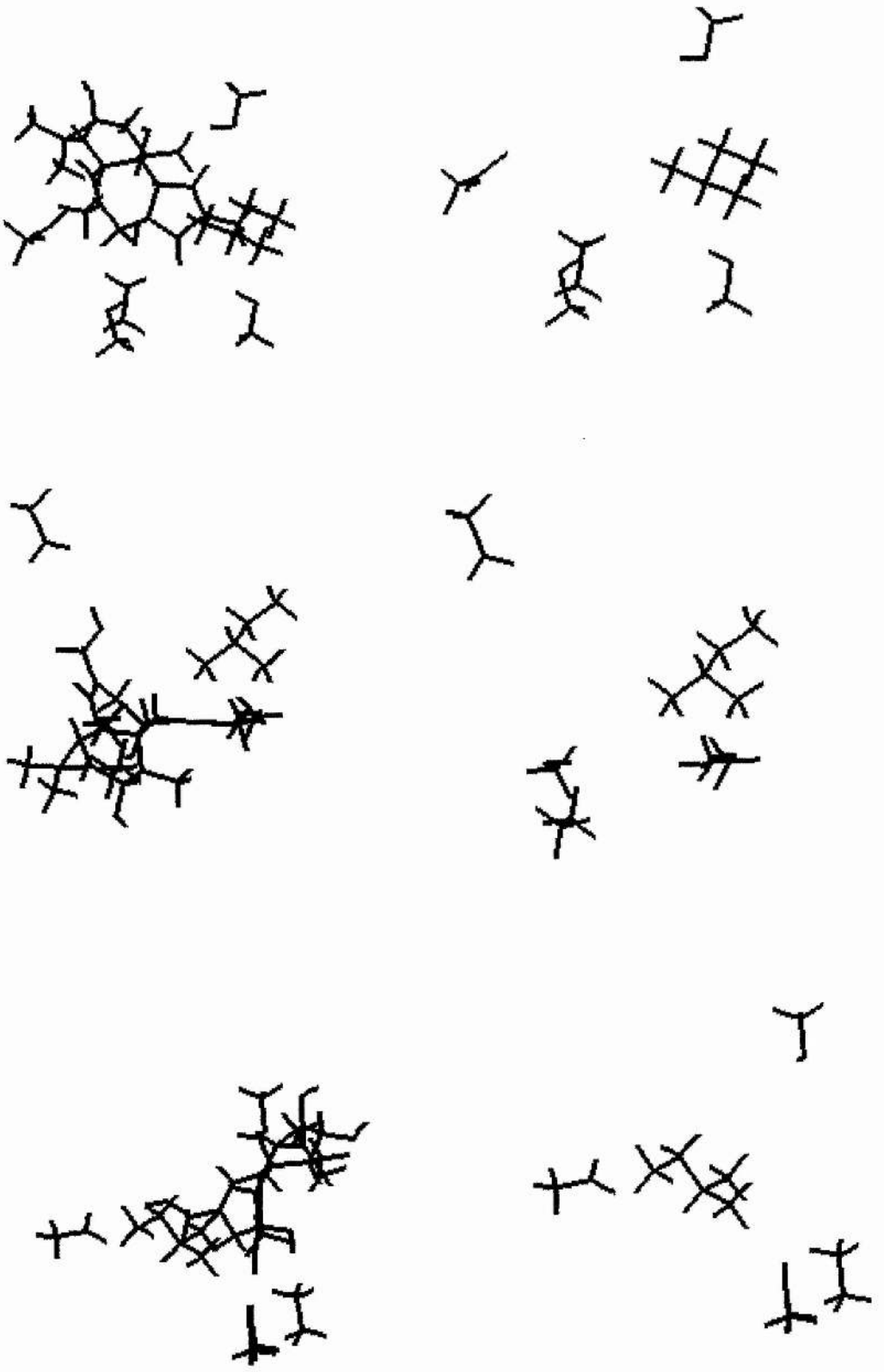
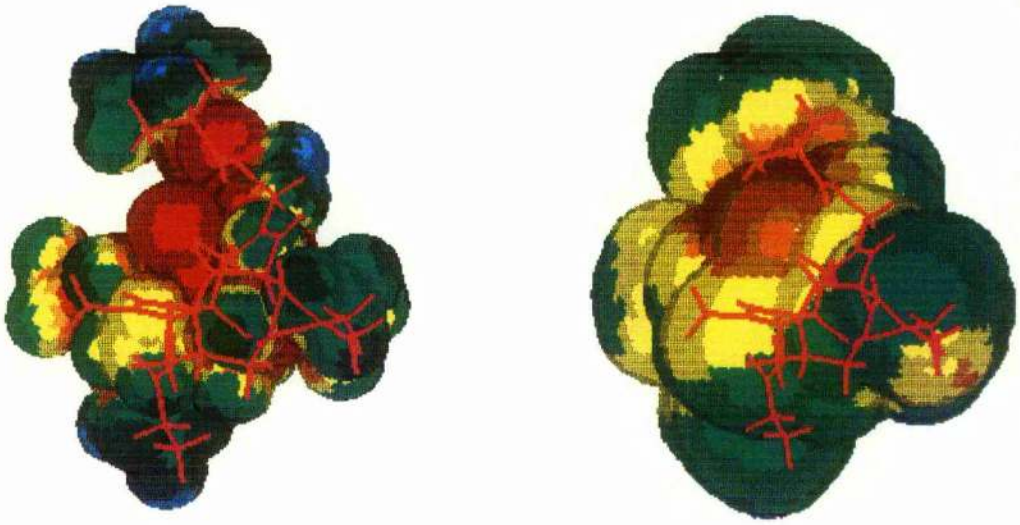
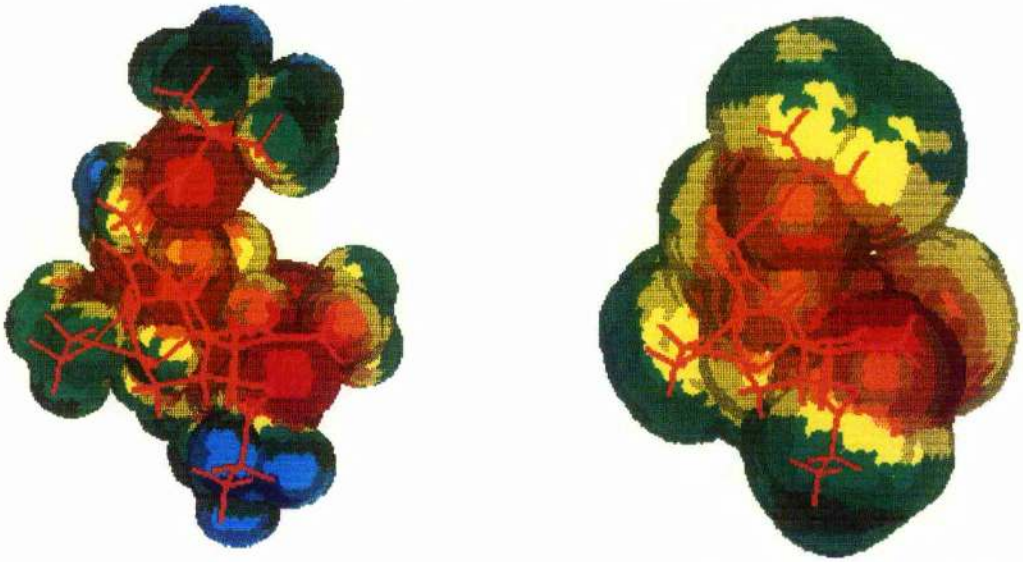


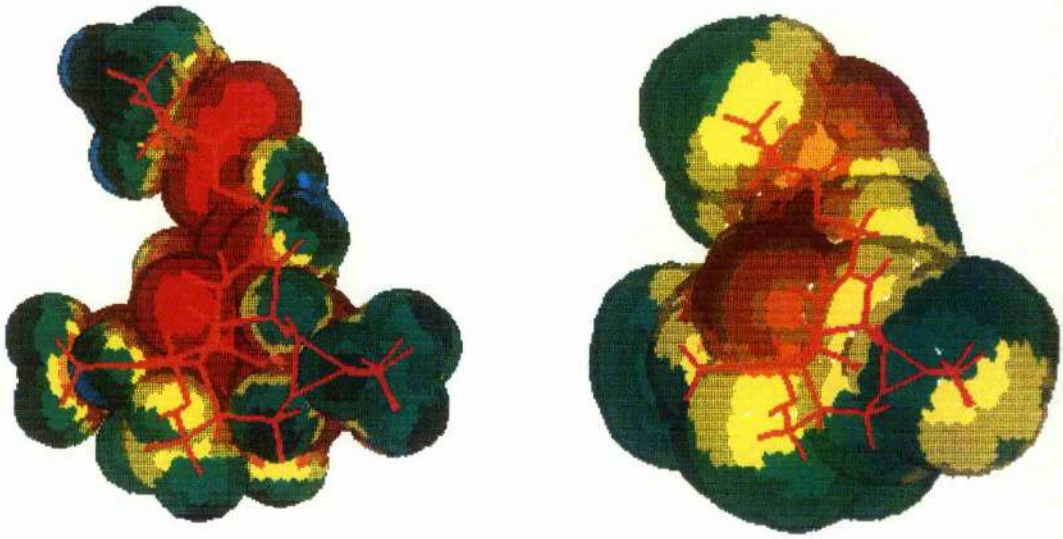
Fig 4.10. Proposed Structure of the Binding Site.



a/ 3-O Propionyl, 20-O Isobutyl Ingenol

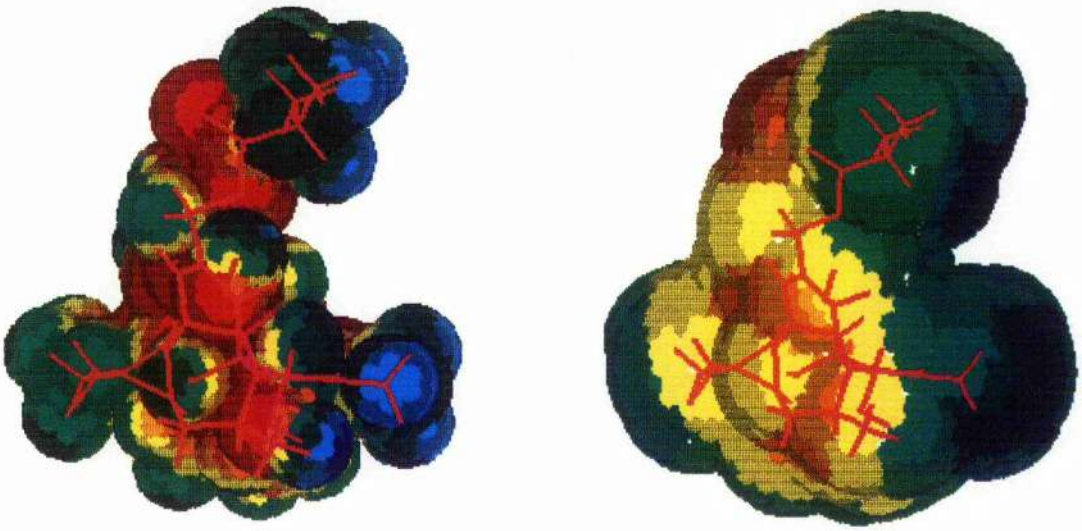


b/ 3-O Propionyl, 20-O Isobutyl Ingenol (Rotated)

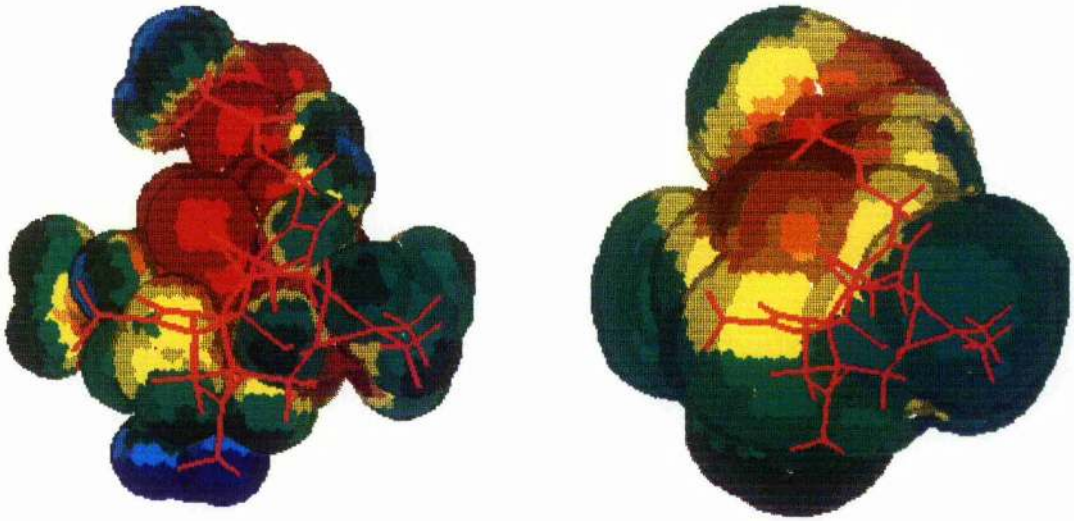


c/ 20-O Isobutyl Ingenol

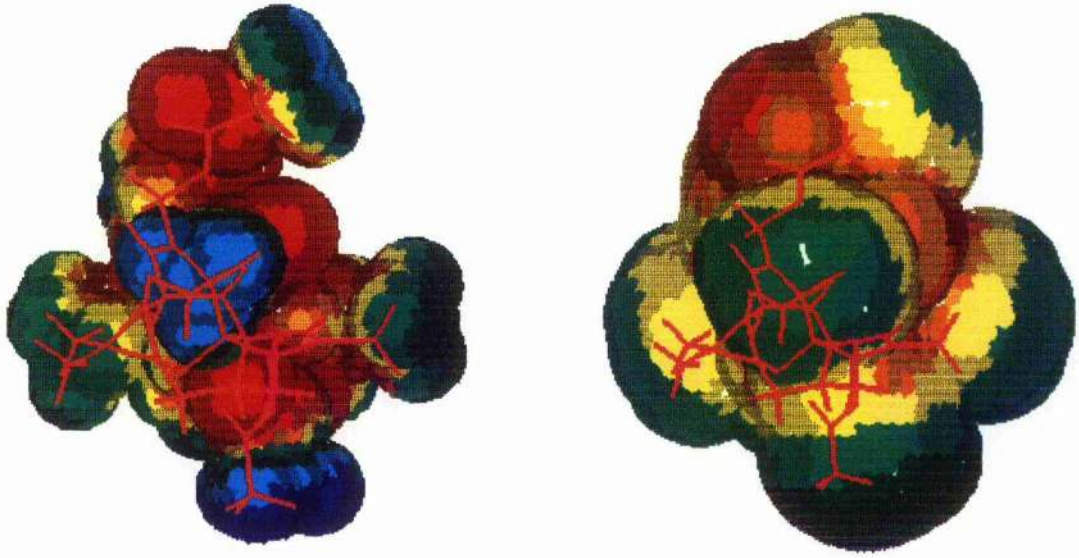
Fig 4.11. EPM's of Ingenol Derivatives.



d/ 20-O Isobutyl Ingenol (Rotated)

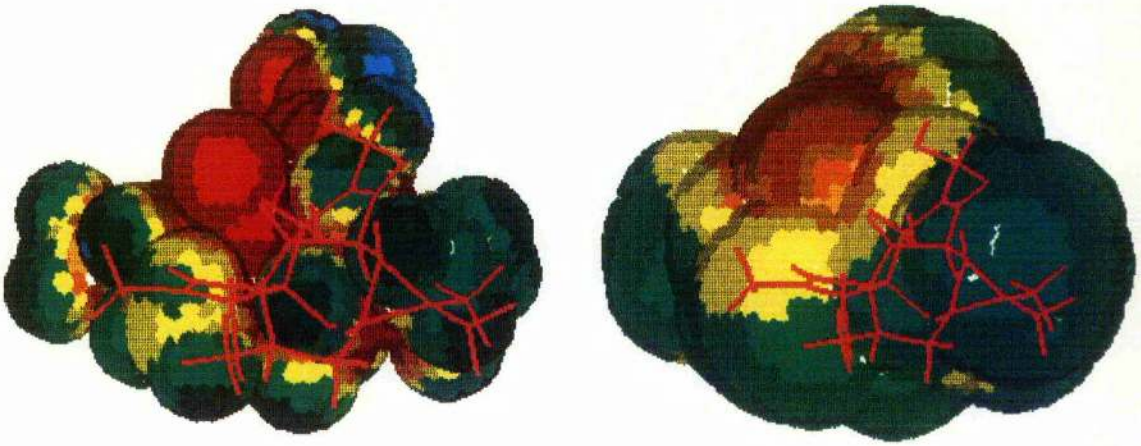


e/ 3′,5′,20′ Ingenol Tri Acetate

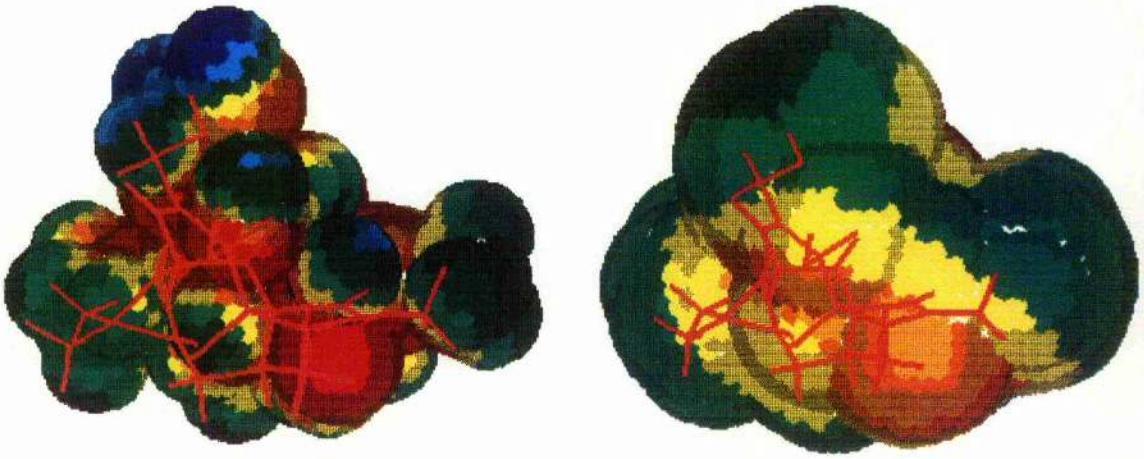


f/ 3′,5′,20′ Ingenol Tri Acetate (Rotated)

Fig 4.11. (Continued)



g/ Ingenol



h/ Ingenol (Rotated)

Fig 4.11. (Continued)

4.5 Ingenols.

As the activity data for the ingenol derivatives is considerably less extensive than for the phorbol esters, they have been used to provide further detail for the phorbol site rather than for a repeat of the analysis.

The active ingenol derivative, 20-O-isobutoyl, 3-O-propionoyl ingenol, has a similar structure to the phorbol esters and a close fit of the key areas of the ESP could only be obtained with the 3', 9' and 20' oxygens overlapping the same positions on the phorbol molecule. This fit is compatible with the view that positions 3 and 9 accept hydrogen bonds, but the ester at position 20 is not able to donate a hydrogen bond. The corresponding phorbol 12', 13', 20' triesters are fully active and there is evidence to suggest that the 20' group is readily hydrolysed [112] leaving the active species as the 12', 13' diester. If the 20' ester of the ingenols undergoes a similar hydrolysis, there is no conflict with the suggested nature of the binding site. Either the ether or carbonyl oxygen at position 3 can be in a position to accept the proposed hydrogen bond from the site., The carbonyl oxygen requires a substantial rotation of the ester-link away from its minimum energy position in order to optimise the fit but the energy barrier is less than 2 kcal mol^{-1}. This ester group is also in a position which, from the phorbol study, would be expected to be close to the membrane surface. However, low-energy minima have been found where the acyl chain takes up a conformation consistent with the expected orientation at the membrane surface, in particular, the chain conformation corresponds closely to the position of the 1' acyl chain of DAG when fitted into the binding site.

The ESP's of all the derivatives considered are shown in fig 4.11, the two orientations centred on positions 9 (upper pictures) and 4 (lower pictures). The upper part of each diagram would be visible, though rotated by 90° around the X-axis, in the orientation used for the phorbols. As is the case with the phorbols, the potential is dominated by the carbonyl oxygen at twice the Van der Waals radius, and by all the oxygen atoms at the smaller radius. Very little difference can be seen in the ESP's of the different derivatives except differences in the extent of esterification, a position

consistent with the idea that the electrostatics of the site - promoter interaction are not important except near the three-point hydrogen bonding contacts and in terms of the orientation of the molecules at the bilayer surface. Thus perhaps the activity of the ingenol esters depends on the position of the esters and the effects they have on the orientation of the molecule in the bilayer. When the ingenols are fitted to the phorbol esters in the manner described above, a 5' ester is in a position that would be expected to be exposed to solution. Thus, if this ester is present, the entropic cost of exposing the acyl chain to solution may be sufficiently large to prevent it binding to the site and so be responsible for the inactivity of this derivative. An alternative view is to suggest that possession of a 5' ester alters the orientation of the ingenol in the bilayer so that the 3', 9' and 20' groups are no longer accessible to PKC thus preventing the molecule from entering the binding site.

From the activity data it can be seen that short-chain ingenol esters are fully active while the corresponding short-chain phorbol esters possess only a reduced activity. Against this it can be seen that the C ring of the ingenols has an ESP that is much smaller in magnitude than it is for the phorbol esters and so the ingenols will sit more readily within the membrane at the correct orientation than the corresponding phorbol esters.

4.6 Teleocidins.

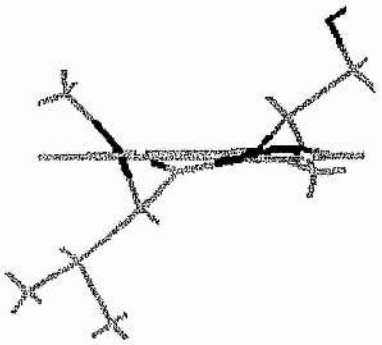
The structures of the teleocidins were derived from the crystal structure of dihydroteleocidin B monobromoacetate [117], with the hydrophobic region being deleted. NMR data suggests that several conformations of the the lactam ring are found in solution [127] and six possible conformations were found from initial model building, namely two "boat" conformations, two "chair" conformations, the twisted form found in the crystal structure and its reverse, all of which are shown in Fig 4.12, displayed with the indole ring projected towards the observer.. All of these conformations contained a *cis* amide-bond. Conformations containing a *trans* amide-

bond were not considered as the rotation barrier for the amide bond is sufficiently large to prevent interconversion of the two forms except at high temperatures [128] and there is no evidence that such structures are found naturally.

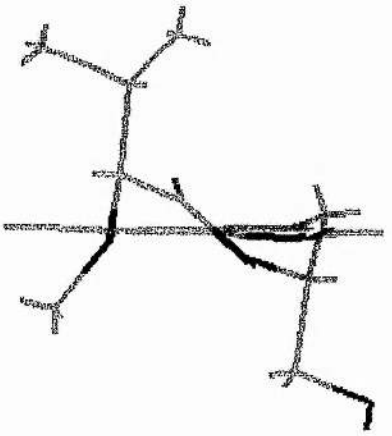
As carbons 9 and 12 in the lactam ring are chiral (see fig 4.1), there are four possible diastereoisomers, though only one, the SS form, is active. The heats of formation of all six possible conformations for each diastereoisomer was calculated to give an indication of which conformations were present in the equilibrium, the results of which are given in the section below.

4.6.1 Heats of Formation of Indolactam V Diastereoisomers.

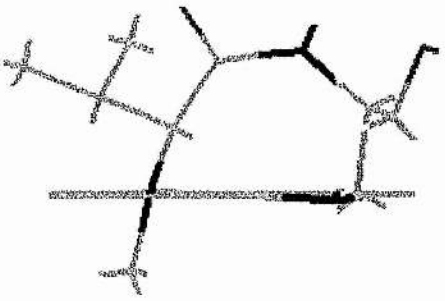
The heats of formation of the six gas-phase conformations of each of the enantiomers are given in table 4.12. Several starting points were used for each point, allowing for rotation of the substituents of the lactam ring, so the energies given should be close to the minimum energy possible for each conformation. It is not practicable to perform a complete conformational search for these molecules as the indole ring alone possesses many close lying local energy minima that differ by no more than 2 kcal mol^{-1}. For this reason, the relative energies are an approximate guide to the equilibrium nature of the conformations. The three semi-empirical methods were used to calculate the energies of the conformations both to compare the reliability of the methods and to reduce the possibility of distortion of the results due to errors in the parametrisation.



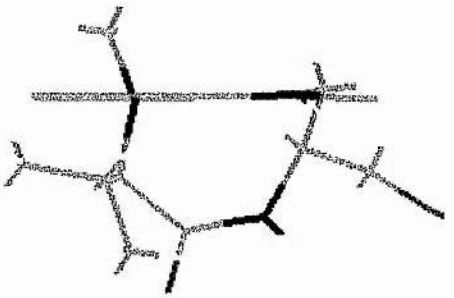
a/ Crystal Conformation



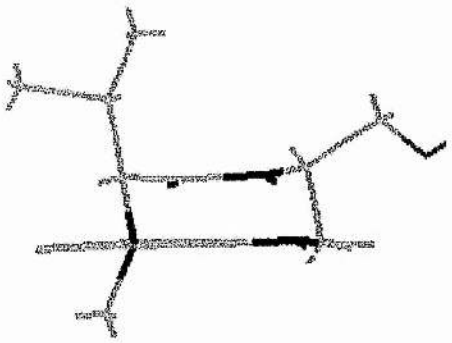
b/ Reverse of Crystal Form



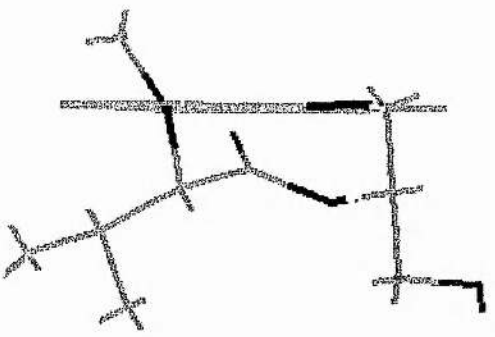
c/ "Boat Up" Conformation



d/ "Boat Down" Conformation



e/ "Chair Up" Conformation



f/ "Chair Down" Conformation

Fig 4.12. Cis-Indolactam V Conformers.

Table 4.12 Conformational Analysis of Indolactam V.

a/ SS Diastereoisomer.

	AM1	PM3	MNDO
Boat up [2]	-21.0	-47.6	-15.1
Boat down	-23.6	-46.6	-12.2
Chair up	-25.7	-48.6	-12.4
Chair down	-27.7	-51.5	-14.2
Crystal	-27.3	-49.6	-13.4
Reverse	-19.4	-42.1	-2.3

b/ RR Diastereoisomer

	AM1	PM3	MNDO
Boat up	-21.5	-48.8	-10.0
Boat down	-21.7	-46.4	-15.1
Chair up	-25.3	-51.3	-15.0
Chair down	-26.2	-50.4	-13.9
Crystal	-15.2	-38.8	-4.5
Reverse	-25.1	-47.9	-12.0

c/ SR Diastereoisomer

	AM1	PM3	MNDO
Boat up	-14.6	-42.4	-10.7
Boat down	-23.7	-49.0	-16.4
Chair up	-29.5	-54.8	-18.7
Chair down	-24.8	-39.7	-7.9
Crystal	-19.9	-43.9	-6.9
Reverse	-21.0	-44.5	-8.3

d/ RS Diastereoisomer

	AM1	PM3	MNDO
Boat up	-24.8	-50.1	-17.2
Boat down	-22.3	-47.3	-9.6
Chair up	-20.7	-46.1	-12.6
Chair down	-29.5	-56.2	-18.9
Crystal	-25.2	-48.0	-10.9
Reverse	-24.4	-45.9	-10.9

[2] For the meaning of the terms "up" and "down" see figure 4.11.

Energies are in kcal mol^{-1}.

Whereas the AM1 and PM3 results were largely similar for the phorbol derivatives, there is a consistent difference in the heats of formation obtained with the two methods for these systems. Some doubt has been cast over the ability of PM3 to accurately model nitrogen containing systems as it does not give convincing charges from Mulliken analysis (see earlier), while it requires the use of a molecular mechanics

correction to the amide torsional potentials to keep it flat. It has been suggested that even with this correction, the results are not entirely reliable [128], and the use of artificial potentials in strained system such as these is undesirable. For these reasons, the AM1 potentials were were used to refine the model of the binding site. There are some discrepancies in the relative energies of different ring conformations using the different methods, so a particular conformation was not discarded from consideration unless all three methods gave it a high energy. These discrepancies are rooted in the strain found in the 9-membered ring and in the distortion of the amide bond, which is not flat in any of the conformations. This distortion is likely to introduce errors into the energies of all three methods, the most vulnerable being PM3 with the molecular mechanics correction (though the quoted energies are without this correction), as none were parametrised for similar systems

The teleocidin and indolactam molecules possess several centres that are able to participate in hydrogen-bonding, namely the hydroxyl group at position 24, the carbonyl oxygen and position 11 and nitrogens at positions 1, 10 and 13. In addition, the entire indole ring is noticeably electron-rich and so may interact favourably with a region of positive ESP, such as a hydrogen bond donor, within the binding site.

This proliferation of potential hydrogen-bonds makes the procedure of fitting to the phorbol esters rather complicated as there are many potential orientations. If however, the positions of the hydrophobic tetramethyl ring and isopropyl groups are considered, the fitting can be somewhat simplified. There is a requirement that the equivalent group to the phorbol 20' hydroxyl is some distance from the hydrophobic regions and directed away from them. Thus it becomes unlikely that the indole nitrogen (N_1) is the hydrogen-bond donor while the hydroxyl group at position 24 is best placed to mimic the 20' phorbol hydroxyl. This group also possesses a near identical degree of flexibility to the phorbol group and so is an obvious choice as the hydrogen-bond donor. The amide nitrogen (N_{10}) is also some distance from the hydrophobic region and in some conformations can be found in similar orientations to the 20' phorbol hydroxyl, however its conformational freedom is severely limited and the distances to potential hydrogen-bond acceptors do not fit too well with those in DAG and the

phorbol esters.

If the 24' hydroxyl group is chosen as the hydrogen-bond acceptor, the 11' carbonyl is well placed to act as one of the hydrogen-bond acceptors. For some conformations, the O_{11}-O_{24} distance fell within the possible range for both the O_2-O_3 and O_1-O_3 distances of DAG, so both orientations were fitted to DAG and phorbol. None of the nitrogens were well placed to fulfil the role of the other hydrogen-bond acceptor, but in many cases, the indole ring, with its hugely negative ESP was well placed to interact favourably with the positive potential found from a hydrogen-bond donor.

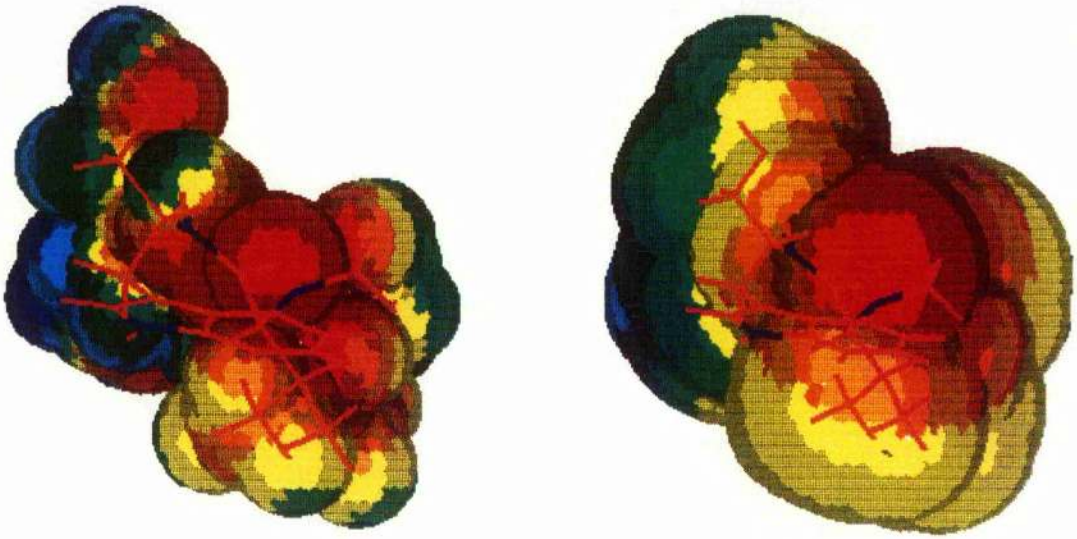
The energies obtained for the active SS form suggested that all of the conformations, except that with the lactam ring in the reverse conformation to the crystal structure (referred to as the "reverse" conformation), would be present in the equilibrium mixture, so these five conformations were all fitted to phorbol and DAG. Both possible fits for the "chair up" conformation and that of O_{11} (indolactam) to O_3 (phorbol) for the "boat down" conformation gave a very poor overlap of the hydrogen-bonding system with a mean deviation greater than 0.72 Å. It was felt that this difference was sufficiently large to severely reduce the contribution of these interactions to the binding energy and so abolish binding. Fitting of O_{11} (indolactam) to O_3 (phorbol) for the "boat up" and "chair down" conformations placed the isopropyl substituent of the lactam ring in a position analogous to that of the 4' methyl group of 4βOMe-phorbol suggesting that the SS form would not be able to bind in these conformations and orientations. The same conformations could be assumed not to bind in the other possible orientation (O_{11} corresponding to O_9) as the hydrophobic regions would be forced into a position which would, from the phorbol study, be expected to lie at the membrane surface or even within the cytosol. The remaining three possibilities, both fits for the crystal conformation and the O_{11} (indolactam) to O_9 (phorbol) fit for the boat down conformation are consistent with both the proposed hydrogen-bonding system and the nature of the site inferred from the phorbol study.

These three fits differ in many details and it is unlikely that all three orientation are possible within the binding site. It is hoped therefore, that consideration of the

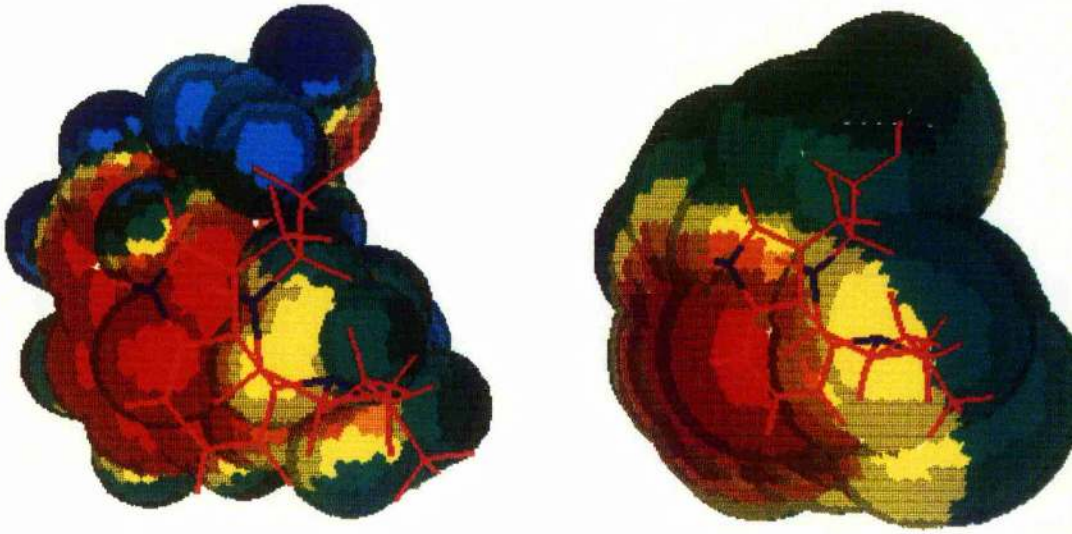
inactive diastereoisomers may shed further light on the nature of binding and give a more detailed picture of the binding site.

As was the case with the active SS form, many conformations of the inactive diastereoisomers would not be expected to bind because the 24' and 11' oxygens fitted poorly with the phorbol and DAG equivalents (both chair conformations, the "boat down" and "reverse" conformations for the RR form and the "boat up" conformation for the RS form), while other fits (RS "chair down":O_{11}-O_9[3], SR "chair up":O_{11}-O_9, SR "boat down":O_{11}-O_3) gave a poor orientation of the hydrophobic or hydrophilic groups relative to the expected position of the membrane. Of the remaining three conformations, the "chair down" conformation of the RS form gave a good fit (O_{11}-O_3) with both phorbol and DAG (mean deviation 0.090 Å), but the isopropyl group occupied the same position as the 4' methyl group of 4βOMe-phorbol. The rotation barrier for this isopropyl group is sufficient (>15 kcal mol^{-1} for ±30° rotation) to prevent rotation to fit within the site so accounting for the inactivity of this enantiomer. When the SR "chair up" conformation gives a good fit when O_{11} corresponds to O_3 of phorbol, but the methyl group of N_{13} is raised above the position of the five-membered ring in a position similar to the methyl group of the 1'2' saturated phorbol so would be expected to prevent binding.

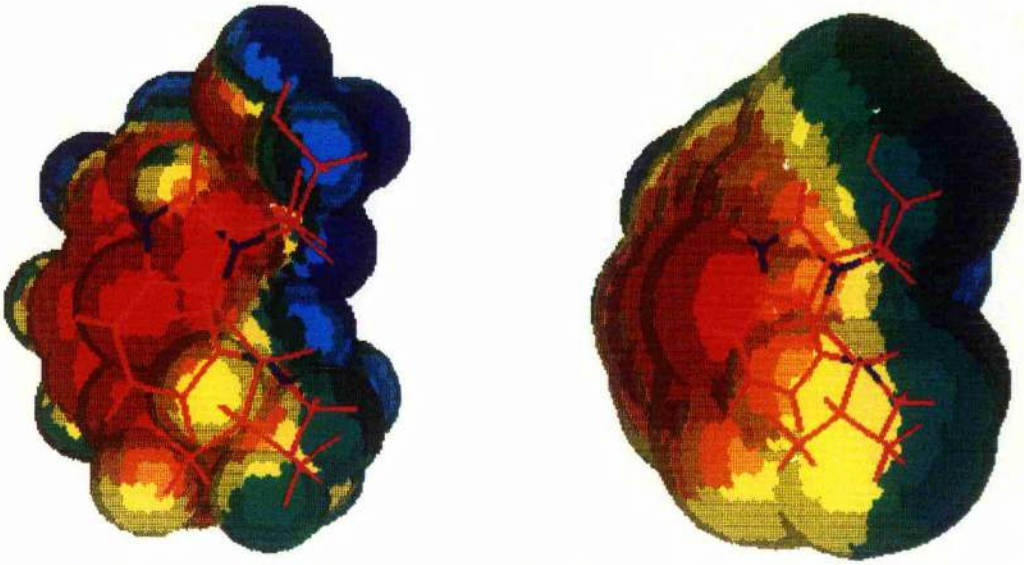
The "boat down" conformation of the inactive SR form is identical to that of the active SS form except for the position of the isopropyl ring. As this group is well away from any part of the binding site, it would follow that neither conformation is able to bind to the site and that the active SS form binds in a conformation similar to that found in the crystal. The "boat down" conformations fit comfortably into the site proposed from the phorbol study without coming into conflict with any of the steric blocks already proposed. The only difference from phorbol in the region of the site is the presence of bulky hydrophobic groups below the plane of the five-membered ring. It has been proposed that this ring occupies a largely hydrophobic pocket so it is possible that this conformation of the indolactams would interfere sterically with the lower bound of this pocket and so be unable to bind. The ESP's of the conformations compatible with the phorbol site are shown in fig 4.13.



a/ "Crystal" SS Form. O11 fitted to O3 (Phorbol)



b/ "Boat Down" SS Form. O11 fitted to O9 (Phorbol)



c/ "Boat Down" SR Form. O11 fitted to O9 (Phorbol)

Fig 4.13. EPM's of Indolactam V.

In the preceding figure, the indolactams are not shown in the same orientation as the phorbols relative to the binding site, rather they are rotated by 90° so as to present the indole ring face-on revealing the areas of negative potential. The picture of the SS diastereoisomer in the crystal conformation is shown with O_{11} pointing towards the observer, thus the molecule is in the equivalent orientation as phorbol centred on O_3. The other two conformations are shown with O_{11} also in the centre, but as these two have been fitted the other way around, they are in an equivalent orientation to phorbol centred on O_9.

[3] RS "chair down":O_{11}-O_9 refers to the "chair down" (see fig. 4.11) conformation of the RS diastereoisomer, with O_{11} fitted to O_9 of phorbol.

4.6.2 Olivoretin.

A similar conformational analysis was carried out on an indolactam model for olivoretin. The parent molecule is inactive and differs structurally from the teleocidins only in that the 24' oxygen is methylated [114]. According to the binding mechanism already proposed, this would be expected to be the cause of their inactivity as the ability to donate a hydrogen bond has been lost, however, it is possible that the modification would change the conformational equilibrium and affect the activity accordingly.

Table 4.13 Conformational Analysis of Olivoretin.

	AM1	PM3	MNDO
Boat up	-18.4	-44.5	-9.7
Boat down	-20.8	-42.5	-8.5
Chair up	-21.9	-44.6	-8.1
Chair down	-21.4	-44.9	-8.7
Crystal	-23.4	-45.8	-7.1
Reverse	-14.8	-35.4	+1.0

The energies of the different Olivoretin conformations, and hence their relative stabilities, closely follows that of the equivalent conformation of the active (SS) teleocidin form in that the reverse crystal geometry is extremely unstable, while the other conformations are of quite similar energy. This suggests that the sole reason for the inactivity of the olivoretins lies in the O_{24} methylation, and is consistent with the view that this group serves the same function as the 20' OH of the phorbol esters.

4.6.3 Consideration of Properties of the Teleocidins.

Thermodynamic and other properties have been calculated for the teleocidins in the same manner as for the phorbols, both to compare the results from the different parametric methods and to compare them with the properties of phorbol. In this latter category, the energies and location of the HOMO and LUMO are particularly important. The minimum energy conformation based on the crystal structure was used for all three parameter sets.

Table 4.14 Thermodynamic Properties of Indolactam V.

	AM1	**MNDO**	**PM3**
Total Energy (eV)	-3754.4	-3764.2	-3443.1
HOMO Energy (eV)	-8.039	-8.297	-8.200
LUMO Energy (eV)	0.094	-0.086	-0.111
Dipole Moment (Debye)	5.842	5.281	5.196
Enthalpy (kcal mol^{-1} at 300 K)	13.563	13.808	13.935
Heat Capacity (kcal K^{-1} mol^{-1} at 300 K)	0.0817	0.0818	0.0852
Entropy (kcal K^{-1} mol^{-1} at 300 K)	0.1462	0.1491	0.1482

The main difference in the calculated properties lies in the total energies, where that of PM3 differs from the other two by around 10%, and corresponds to the known problems with using PM3 for systems containing an amide bond, especially one that is distorted as is the case here. Of the properties that would be expected to affect the binding energy, the HOMO and LUMO energies both compare well with the phorbol equivalents , but the dipole moment is significantly larger, and unlike the phorbol dipole, consistent between the different methods. The change in the dipole moment may not be significant in terms of activity as the phorbol dipole in particular is very small, and varies considerably for rotation of groups not involved in binding. The HOMO and LUMO eigenvectors agree closely between the three methods with both orbitals covering much of the indole ring, the HOMO also extending to N_{13}. The largest LUMO eigenvectors are on the six-membered part of the indole system, which occupies the same area of the site as the phorbol A-ring.

The I.R. Spectrum of the indolactams is potentially the most complicated of all of the molecules considered, with a large number of atoms in quite different bonding confirgurations. As before, the semi-empirical stretching frequencies have been compared with typical experimental values [126] shown in table 4.15.

The frequencies are again in units of cm^{-1}, and for bonds involving carbon, oxygen and hydrogen, the results are similar to those of phorbol. The experimental ranges of the frequencies of nitrogen containing bonds are rather wide and all three methods give results that are close to or lie within these ranges with the exception of the AM1 amine N-H frequency and in general AM1 generally frequencies close to the upper limit of the range. There is no corresponding consistency with the other two methods. The known problems with using PM3 for amide systems [128] is partly reflected in the low value it gives for the amide stretch frequency.

Table 4.15 Indolactam I.R. Spectra.

	AM1	MNDO	PM3	Experimental
C-H	2968-3228	3160-3455	2826-3181	2855-2940
C-C	1054-1401	1026-1391	973-1436	910-1250
Amine C-N	1405-1495	1095-1539	1220-1454	1065-1235
Amide C-N	1621	1499	1437	1585-1695
C-O	1325	1442-1452	1107-1378	1000-1250
C=O	1986	1669-2080	1915	1670-1750
Aromatic C-C	1197-1799	1213-1760	1342-1797	1430-1665
Aromatic C-N	1562-1591	1351-1587	1090-1504	1250-1660
O-H	3499	4008	3910	3335-3635
N-H	3432-3458	3541-3629	3351-3459	3450

4.7 Discussion.

The structure / activity data for the ingenols and teleocidins is consistent with the model for the binding site suggested from the phorbol data, with the modification that the interaction with the 9' groups of ingenol and phorbol or the indole ring of teleocidin is a dipole-dipole or charge-dipole interaction rather than a straight forward hydrogen bond. The teleocidin results also suggest steric interaction below the phorbol A-ring in a position that prevents the bulky tetramethyl ring entering the site, but does not interact with the flexible 1' acyl chain of DAG. The locations of the LUMO for both phorbol and indolactam V suggest that the A-ring is also the target of a nucleophilic interaction with the binding site. The need to accommodate the 1' acyl chain of DAG into this region suggests that the interaction comes from above the plane of the A-ring (in the sense of fig 4.6), thus binding is abolished for modifications that move the location of the LUMO away from this area, or that raise groups above the

plane so as to interfere sterically with the nucleophile. This nucleophilic interaction may also be responsible for the effect on activity of changes in the direction of the dipole moment of the A-ring. The refined model for the binding site is shown in fig 4.14 at the end of this chapter, which follows the same protocol as fig 4.9.

When considering the implications of structure / activity relationships, it is important to consider that the bound conformation is not necessarily the same as the free minimum-energy conformation, so the flexing of rings and rotation of side-chains must be considered. However, the three classes of tumour promoters all contain highly strained ring system that allow only minimal rotation and so the conformations of these rings can be considered fixed, while the case of the *cis* amide lactam ring of teleocidin has already been considered. The hydroxyl groups at positions 4 of phorbol and ingenol and 9 of phorbol are allowed only limited rotation as the ring system folds around them, but the 20' hydroxyl group and its 24' equivalent of teleocidin is allowed almost completely free rotation. The minimum energy conformations of DAG and phorbol have the equivalent hydroxyl groups pointing in quite different directions, but they can be rotated to point the same way at a minimal energy cost. From this it can be seen that the assigned position of the hydrogen-bond acceptor within the binding site is purely arbitrary and could be positioned anywhere following rotations of the C_6-C_{20} and C_{20}-O_{20} bonds. According to this model the function of the large hydrophobic regions of the molecules is merely to partition the promoter into the membrane and to present it to the binding site in a reasonable orientation.

There have been several attempts to correlate calculated structures with activity data and thus to suggest important interactions of the tumour promoters with the binding site. These have lead to differing conclusions.

Jeffrey and Liskamp [124] proposed a four-point hydrogen-bonding system, the additional one being at position 4, though subsequent data has shown that the absence of OH_4 has no effect on binding. This particular study was carried out before the crystal structure of the teleocidin series had been determined and like an earlier work by Weinstein *et al* [109] found a best fit of the teleocidins to phorbol for the RR diastereoisomer. This conflicts with the fitting data shown here, but it is clear from the

energies quoted by Jeffrey and Liskamp that there is some difference in the conformations studied. It is more likely that the earlier work contains distortions to the geometry as the work was carried out empirically using the MM2 force field on a highly strained system quite different from those used in the parametrisation. Also of course the use of a different hydrogen-bonding system will lead to a different fit as they compare O_4 of phorbol with N_{13} of teleocidin, though from the calculation shown here, N_{13} is not accessible for hydrogen-bonding in the SS diastereoisomer.

By comparison, Wender *et al* [123] propose a three-point system involving OH_4 (phorbol) rather than O_3 and have synthesized analogues that bind to the site with low affinity. This activity could alternatively be attributed to a relatively poor fit to the 3', 9', 20' three-point system proposed here. Wender compared the structures of DAG and phorbol and was able to obtain a reasonable fit for the key groups, but there remains some doubt as to the viability of the orientations of the molecules as he neglected the acyl chains of the phorbol esters and his proposed fit would result in the chains being exposed to the solution of tightly folded back over the molecule, neither of which is likely.

The most recent work, by Itai *et al* [129], has made use of the electrostatic potential of the teleocidin series of tumour promoters, though only through single point calculations on geometries calculated with molecular mechanics. They present a similar hydrogen bonding system to that shown here, but include the possibility of a fourth hydrogen-bond donated by NH_{10} in a near-*trans* amide conformation based on the crystal structure of olivoretin B. The description of the rest of the site differs from that shown here. As Itai *et al* obtained an understanding of the site from a consideration of teleocidin, they have proposed a compact hydrophobic pocket covering the hexamethylene ring and then folded the 12' myristate chain of TPA to fit within the pocket. This is at odds with the concept of a membrane anchor developed here and is dynamically far less likely.

In this chapter, the description of the site has been primarily concerned with specific and localised interaction of the tumour promoters with the binding site. Both active and inactive derivatives have been considered whereas the previous studies

mentioned above have been concerned only with the active derivatives. As the phorbol derivatives considered differ mainly in the A and B ring substituents, it is not surprising that the detailed description if the site presented here should be confined to regions surrounding these two rings. However it is noticeable that all of the different hydrogen-bonding systems suggested previously involve only groups within this region, while the only derivative modified elsewhere on the molecule possess full activity so the binding site may well only cover the A and B rings of the phorbol esters. The structures of the different classes of tumour promoters used in this study vary considerably in the regions corresponding to the phorbol C and D rings and in the exact positions and conformations of the hydrophobic groups suggesting that they are responsible for presenting the tumour promoter to the site in the correct orientation rather than acting as part of a specific interaction as suggested by Itai *et al.*

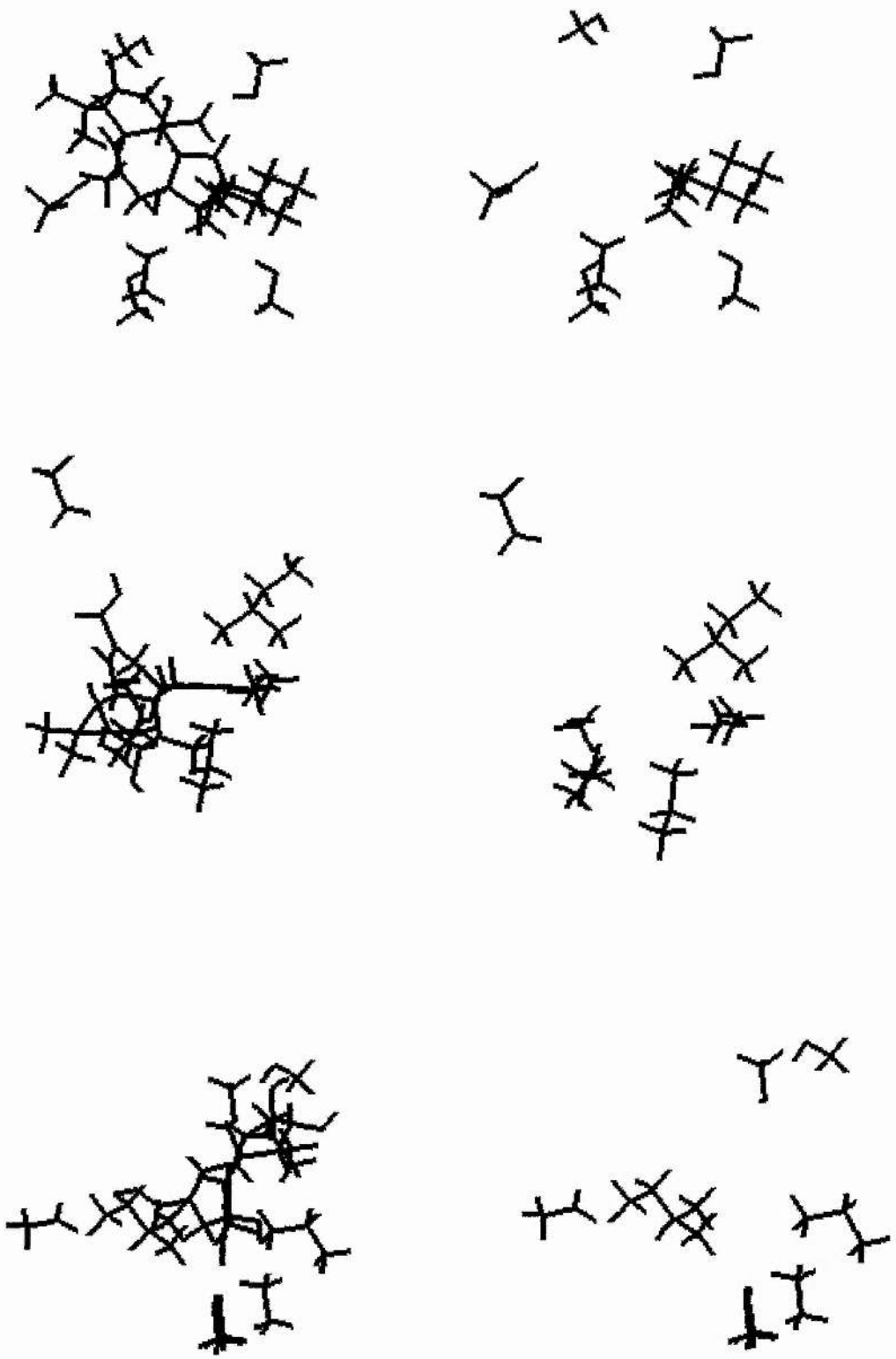


Fig 4.14. Refined Binding Site Structure.

Chapter 5

Inositol Polyphosphates.

5.0 Introduction.

The purpose of this chapter is primarily to compare the conformations of the different *myo* -inositol polyphosphates and relate differences to differences in function. These molecules, unlike the phorbol esters, ingenols and teleocidins dealt with in the previous chapter, are extremely flexible. A six-membered ring with all six atoms being non-equivalent, as is the case with most of the inositol phosphates, possesses eight basic ring conformations, six "boat" and two "chair" conformations. In addition, rotation is possible about both the ester-link and within the phosphate group itself, leading to a very large number of potentially stable conformations. Thus methods have been used to reduce the number of conformations examined to a reasonable number, while still ensuring that statistically significant conformations are not ignored.

In addition molecular dynamics calculations are carried out on the physiologically important inositol (1,4,5) trisphosphate (IP_3), to give further insight into the relative stability of different conformations. These calculations have been carried out on the molecule both in isolation and immersed in a water bath.

The shortage of structure / activity data for activators of the IP_3 receptor (also called p400 [130]) coupled with the considerable flexibility have made it impossible at the moment to describe the binding site in the manner of the previous chapter.

The structure and numbering system used for *myo* -inositol is shown below in fig. 5.1.

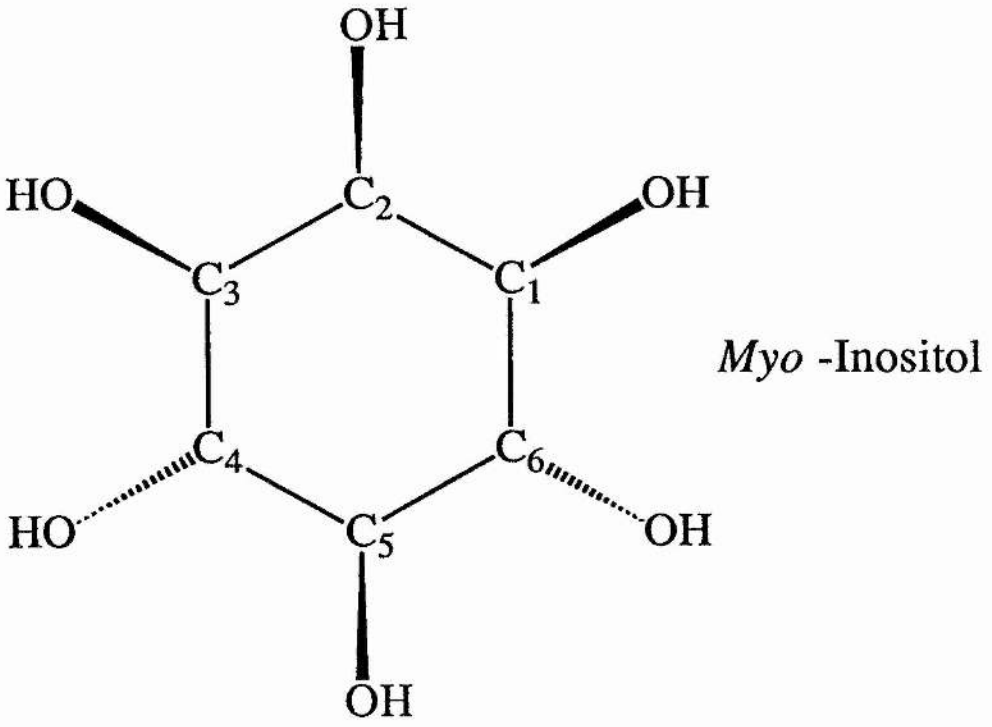


Fig 5.1 *Myo* -Inositol Showing Numbering Scheme.

5.1 Physiological Action of Inositol Polyphosphates.

The inositol phosphates studied in this chapter are all part of the metabolic cycle involving IP_3 [40] and as such would be expected to be present, in varying proportions, subsequent to the formation of IP_3 in response to receptor binding (see fig 5.2).

The key step in the pathway is the agonist-stimulated hydrolysis of $PstInsP_2$ to produce both DAG and IP_3, which occurs on response to the binding of a wide range of messengers to their receptors. As can be seen from fig 5.2, IP_3 is metabolised by two separate routes [40], it can be phosphorylated to the (1,3,4,5) tetrakisphosphate (IP_4) [131] before being successively dephosphorylated at positions 5 [132], 1, 3 and 4 [40], while the other path starts with the loss of the 5' phosphate [133] followed by the other two in any order to give free inositol [134] which can be reincorporated into phosphatidyl inositol when it is rephosphorylated to return $PstInsP_2$. It has been

suggested that hydrolysis of the ester link in $PstInsP_2$ can make use of the free 2' hydroxyl group of the inositol moiety rather than water and so produce cyclic inositol 1:2,4,5 trisphosphate, cIP_3 [40]. This reaction has been demonstrated *in vitro* using $PstInsP_2$ and purified phosphoinositidase [135], but the evidence for its occurrence *in vivo* is rather less certain [136]. It is, however, known that cIP_3 can be dephosphorylated in a similar manner to IP_3, to give cyclic inositol 1:2 phosphate [137], which is a substrate for a D-*myo*- inositol 1:2-cyclic phosphate 2-inositol phosphohydrolase so opening the cyclic ester to leave inositol 1 phosphate [138]. Cyclic IP_3 is hydrolysed more slowly than IP_3, but it is not known whether this serves as a means of prolonging the stimulus for calcium release, as the ability of cIP_3 to activate the calcium channels is not known.

In this chapter, the conformational equilibria of all the free inositol phosphates, in this metabolic cycle, have been studied, except for inositol (3, 4) bisphosphate as this part of the pathway is the least well defined [40].

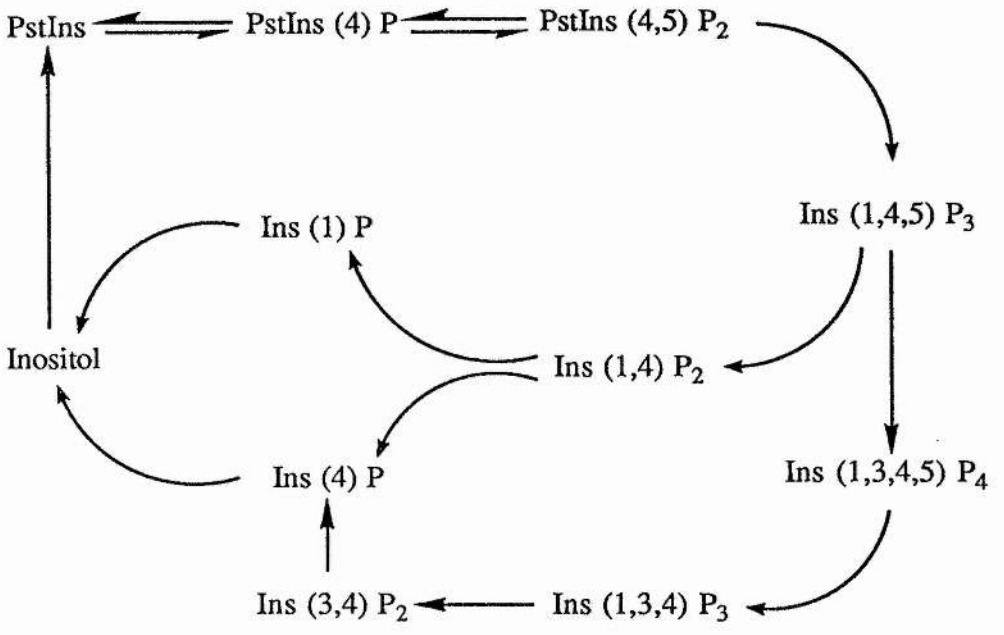


Fig 5.2 *Myo* -Inositol (1,4,5) Trisphosphate Metabolic Pathway.

The cellular receptor for IP3 has been purified, isolated [130] and its sequence determined from cDNA analysis [41]. Due to an error in the initial mass measurements it is usually referred to as p400 though its molecular weight is somewhat less than this. As it was only recently identified as the receptor, little is known about its structure and action at a molecular level. There is also very little known about the physical properties of the inositol phosphates, though crystal structures exist for the 2' monophosphate [139], the hexakisphosphate [140], found in plants, and inositol itself [141-143]. All of these molecules are found in a chair conformation, but there are 5 equatorial hydroxyl groups for the 2' monophosphate and inositol and only one (2') for the hexakisphosphate. The reason behind this change in chair conformations is undoubtedly to relieve close steric contacts and ionic repulsions between the phosphate groups of the hexakisphosphate. Extrapolating from this, it is possible that the differences in activity of the different inositol polyphosphates is dictated by differences in the relative energies of different ring conformations.

5.2 Computational Strategy.

The structural calculations presented in this chapter have all been carried out using the semi-empirical parameter sets contained within MOPAC. Of these three, MNDO is known to give poor core-core repulsions [63] and be unreliable for high valence phosphorus compounds [144], so would be likely to give the poorest quality results for molecules with severe steric crowding such as the higher inositol phosphates. While AM1 generally gives a more accurate representation of the long-range repulsions, it is reported to be less accurate for hypervalent compounds [144]. Phosphorus was not considered in the original AM1 parametrisation and although parameters were subsequently published [145], the results may not be as accurate for compounds outside those used for the parametrisation than would be the case for methods with all the atoms parametrised together. Early AM1 calculations of phosphorus compounds (including Stewart's comparison with PM3 [64]) used the

MNDO parameters for phosphorus while retaining the AM1 parameters for the other atoms. Version 3 of MOPAC gives results similar to those for MNDO parameters alone, but a bug in version 5 leads to highly spurious results, for example heat of formation of -13000 kcal mol^{-1} and a value for GNORM of over 1000 for IP_3.

Two approaches have been used in the calculations of the monophosphates. In the first approach, termed "anionic" approach, the energies were first minimised with the phosphate groups fully ionised, using the AM1 parameter set. Final energies were then obtained for all three parameter sets using these initial geometries as starting points, and adding the hydrogens to the structures. By ignoring the hydrogens in the initial stages, the number of necessary starting points can be reduced as there is no need to consider phosphate rotation as all the oxygens are equivalent. As the main phosphate-phosphate interaction was thought to be a steric and ionic repulsion, this simplification should not lead to a substantial change in the final geometry. The second approach, termed "hydrogen" approach, includes the hydrogens from the beginning. In this case, extra starting points were included to allow for rotation of the phosphate groups and of the phosphate hydroxyls, though in the case of the monophosphates, rotation of the hydroxyls could be ignored as the rotation barriers are very small and other considerations, such as electrostatics, tend to dictate the nature of the energy minimum for these rotations. This method was used to confirm (or deny) the validity of the first approach.

Both approaches have potential drawbacks for the higher phosphates. The MNDO-like semi-empirical methods proved not to be reliable for highly charged species, in particular large positive energies were be obtained and energies for many occupied orbitals became positive. However, the geometries obtained from these calculations are reasonable and appear less prone to error than the energies and as the hydrogens were replaced for the final calculations used in the structure / energy comparisons, there is good reason to have faith in the results. If on the other hand, the hydrogens are included from the beginning of the calculations, the problem of steric crowding is added to and there is a real danger of incorrect placement of the hydrogens, which can only be overcome by considering rotation of the phosphate group and

possibly also requiring consideration of rotation of the phosphate hydroxyls, so adding to the already large number of starting points. The problem this could pose can be seen from the fact that adding only the possibility of phosphate rotation to IP_4 increases the number of necessary starting points by a factor of 3^4. For this reason alone, this first method is preferable for higher phosphates.

5.2.1 Choice of Starting Points.

Initial model building shows eight basic ring conformations for a non-symmetrical six-membered ring such as is the case for most *myo* -inositol phosphates, namely two "chair" and six "boat" conformations. In addition, a phosphate group has potentially three minima for rotation about the "P-ester O" bond corresponding to the three non-equivalent staggered conformations, and a similar three potential minima produced by rotation of the phosphate group itself. It is reasonable to neglect the effect of phosphate hydroxyl rotation as the barrier is so small that the relative energies of different orientations depend almost entirely on the position of neighbouring charged groups. There is a multiplicative relationship between these rotations so that the number of potential minima that would need to be examined rapidly rises to unmanageable proportions. (There are potentially over 52000 minima for IP_4 even when hydroxyl rotation is ignored.) Against this, the higher phosphates in particular suffer from considerable steric crowding so that many of these potential minima disappear from steric considerations alone. There remains, however, the problem of ensuring that the energies of all of the important minima are calculated for each compound, so starting points corresponding to most, if not all, of the minima must be used. It was found that the barriers for rotation of the phosphate groups were generally rather small (<2 kcal mol^{-1}) except for those conformations that showed strong hydrogen bonding between phosphate or hydroxyl groups. In this situation however, the different starting points corresponding to phosphate rotations would usually all give the same final minima or be close to it, so it was reasonable to ignore the possibility of

phosphate rotation when setting the starting points for the calculations. This reduced the number of starting points by a factor of 3N (where N is the number of phosphate groups).

Reduction of the number of starting points in this way left a total of 24 starting points required to ensure the calculation of all of the important minima for each of the mono-phosphates, though there still remained a considerable (and impracticably large) number of starting points for the higher phosphates. The required starting points for calculation on the higher phosphates are built up from the final structures obtained for the compound with one fewer phosphate group. Thus for each molecule, only three new starting points are required (corresponding to rotation about the phosphate C-O bond) for each distinct conformation obtained for the previous molecule. This method is able to reduce the number of necessary calculations only if several starting points yield the same final geometry at the earlier stages, a situation which can easily arise in sterically crowded molecules where many potential minima are discounted by close contacts between groups. This method has been particularly useful in this study, reducing the potential 52000 starting points of IP_4 to a mere 9. It is felt that this method is valid for the inositol phosphates as the interaction between the phosphate groups is primarily repulsive in nature owing to the considerable negative charge on the oxygen atoms. In this situation, it is unlikely that minimum energy conformations for the higher phosphates will have no corresponding minimum energy structure in the lower phosphorylated analogs, thus important minima should not be missed using this method.

5.3 *Myo* -Inositol.

Initial calculations were performed using all three parameter sets on the eight basic ring conformations to examine their relative stability and to look for potential distortions from the simple geometries brought about by the particular positions of the *myo* -inositol hydroxyl groups, especially that at position 2 as it is on the same side of

the ring as its neighbours.

Table 5.1 *Myo* -inositol energies and geometries.

Starting Geometry	AM1		MNDO		PM3	
	Final Geometry	Energy kcal mol^{-1}	Final Geometry	Energy kcal mol^{-1}	Final Geometry	Energy kcal mol^{-1}
I	I	-307.852	I	-282.231	I	-273.672
E	E	-303.607	E	-273.233	E	-272.098
1A	1A	-301.252	2E	-276.159	1A	-267.682
2E	2E	-305.134	2E	-276.173	2E	-270.856
3A	3A	-301.252	2E	-276.050	3A	-267.684
1E	1E	-303.465	1E / 3A	-273.192	1E	-269.049
2A	2A / 3E	-303.613	2A	-272.876	2A	-271.800
3E	3E	-303.470	1A / 3E	-273.189	3E	-269.049

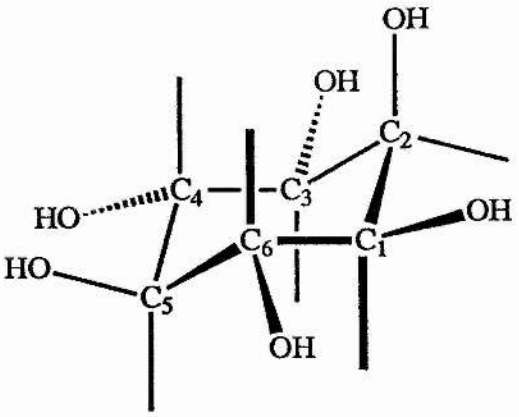
Thus for all three parameter sets, the "5 eq / 1 ax" chair conformation (termed "I", see table 5.2) is the most stable, though there is some difference in the positions and relative energies of the other local minima. The excessive long-range core-core repulsion present in the MNDO parametrisation is shown by the disappearance of minima involving eclipsed hydroxyl groups ("1A", "1E", "3A" and "3E" boat conformations) and the relative stability of the "2E" boat, which apart from the "I" chair has the greatest separation of the hydroxyl groups. Reduction of this core-core repulsion for the AM1 parameter set does not alter the relative stability of the eight ring conformations, but reduces the importance of the hydroxyl-hydroxyl distances on the positions of the local minima, thus minima are found corresponding to the eclipsed conformations though at relatively high energy. The further changes to the evaluation

of the core-core potential within the PM3 parametrisation resulted in a further reduction in the importance of hydroxyl-hydroxyl repulsion in fixing the position of the local minima, such that minima are found corresponding to all of the conformations and the "2E" conformation is no longer the most stable of the boat conformations. The enhanced stability of the "2A" conformation comes from the greater contribution of hydrogen bonds to the final energy using PM3. *Myo* -inositol is a symmetrical molecule with the "1A" and "3A" boat conformations being mirror-images (the same is true of the "1E" and "3E" boat conformations), though the energies obtained differ slightly. This comes from the fact that hydroxyl rotation is not considered when setting the starting points, so the small difference in energy between the mirror-images comes from different rotations of the hydroxyl groups and as the largest difference is 0.1 kcal mol^{-1} these rotations are not significant in terms of the effect on conformation.

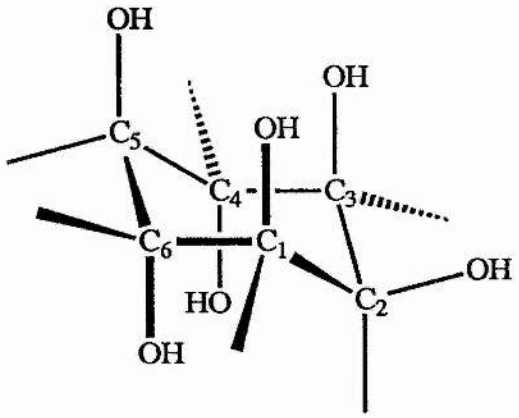
The following abbreviations for the ring conformations have been used in all the tables in this chapter including the one above:

Table 5.2 Abbreviations of Ring Conformations.

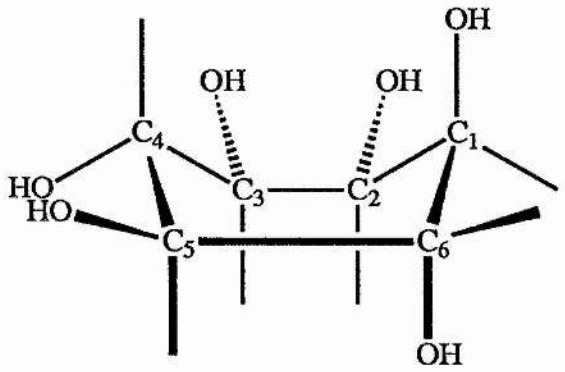
Abbreviation	Meaning
"I"	1 axial / 5 equatorial chair conformation (also called "2'ax chair")
"E'	5 axial / 1 equatorial chair (also called "2'eq chair")
"1A"	Boat conformation with 1' OH (or H_2PO_4) axial
"1E"	Boat conformation with 1' OH (or H_2PO_4) equatorial
"2A"	Boat conformation with 2' OH (or H_2PO_4) axial
"2E"	Boat conformation with 2' OH (or H_2PO_4) equatorial
"3A"	Boat conformation with 3' OH (or H_2PO_4) axial
"3E"	Boat conformation with 3' OH (or H_2PO_4) equatorial



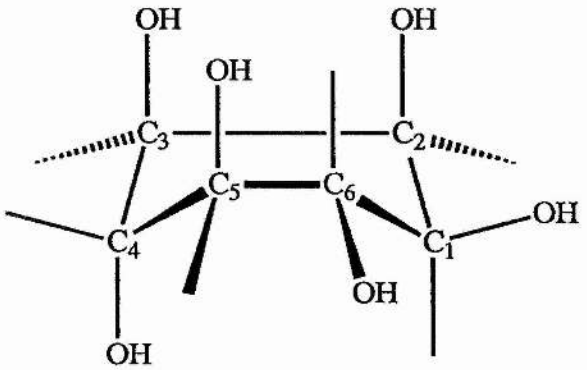
a/ "I" conformation



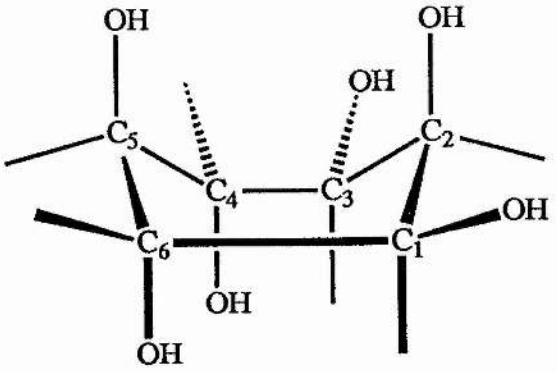
b/ "E" conformation



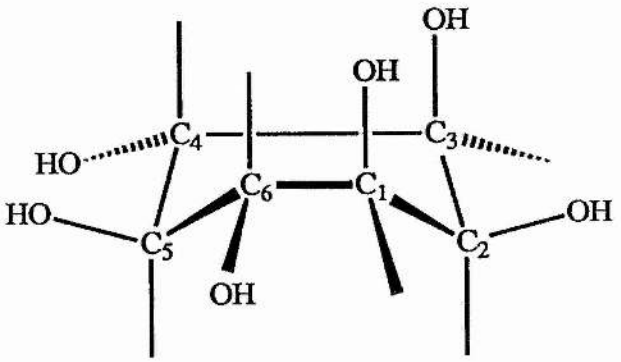
c/ "1A" conformation



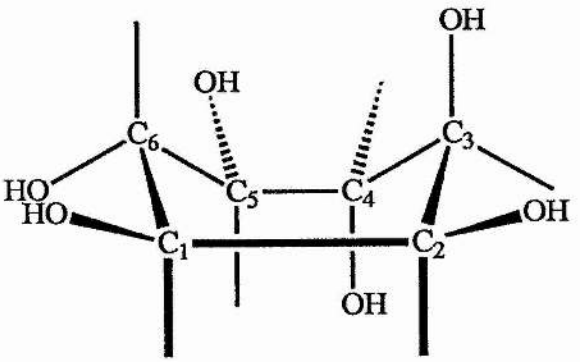
d/ "1E" conformation



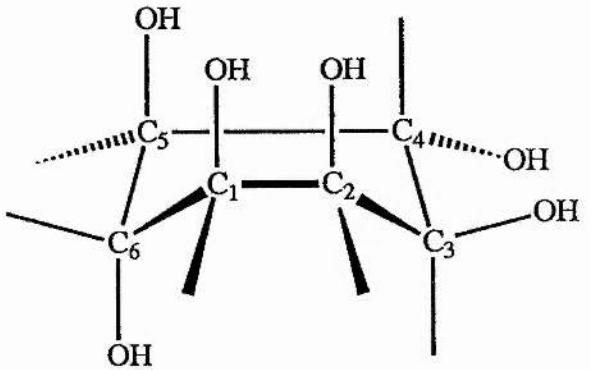
e/ "2A" conformation



f/ "2E" conformation



g/ "3A" conformation



h/ "3E" conformation

Fig 5.3 Myo -Inositol Ring Conformations.

Where the ring conformation is given by two of the above symbols (always two of the "boat" conformations), the actual conformation is someway between the two (produced by twisting the parallel ring bonds of either named conformation in the direction of the other). Such a twisting away from the classical boat conformations usually occurs to relieve close contacts between ring substituents. This twisting of the boat conformations is reflected in the order in which they are listed in the tables as a 60° rotation of the parallel ring bonds of any conformation will transform it to either of the adjacent conformations in the table.

In addition, for the inositol phosphates, the phosphate position is always given as a dihedral angle measured conventionally in the sequence P_N-O_N-C_N-C_{N+1} around the ring.

5.4 Inositol Monophosphates.

In this section, a comparison of the minimum energy conformations of 1', 3', 4' and 5' *myo* -inositol phosphates are presented. Both of the calculation strategies described in section 5.2 are used for the 1' phosphate and comparisons made between the two. The "anionic" method is not used for the other monophosphates as it did not lead to a reduction in the number of starting points and though orbital energies remain negative for the monophosphate, it was felt preferable to use the "hydrogen" method.

There are 24 starting points for each of the monophosphates using either method, but 4 were discarded when using the "anionic" method at the hydrogen addition stage as they produced the same conformations as other starting points. There was no comparable reduction using the "hydrogen" method as only one calculation was performed for each starting point, but the 24 starting points yielded the following numbers of distinct final geometries: 1' - 16, 3' - 19, 4' - 17, 5' - 20, while in addition (for both methods) several starting points gave similar final ring conformations with different phosphate orientations.

Table 5.3 1' Inositol Monophosphate (PM3 using "hydrogen" method).

Ring conformation	Energy of Minimum	P_1-O_1-C_1-C_2
I	-473.539	226°
E	-472.723	221°
1A	-473.303	222°
1A / 2E	-472.620	254°
2E	-474.742	220°
2E / 3A	-	-
3A	-462.983	102°
3A / 1E	-472.708	237°
1E	-	-
1E / 2A	-	-
2A	-472.125	237°
2A / 3E	-	-
3E	-	-
3E / 1A	-458.160	51°

No minima was found for conformations which have a hyphen in place of the energy, even though there were starting points for all the classical boat conformations with phosphate groups at 60°, 180° and 300° to the next ring carbon atom.

Table 5.4 3' Inositol Monophosphate (PM3).

Ring conformation	Energy of Minimum	P_3-O_3-C_3-C_4
I	-477.541	255°
E	-475.744	246°
1A	-	-
1A / 2E	-473.377	227°
2E	-473.723	249°
2E / 3A	-	-
3A	-471.018	243°
3A / 1E	-471.800	225°
1E	-	-
1E / 2A	-464.734	25°
2A	-474.406	257°
2A / 3E	-	-
3E	-465.676	44°
3E / 1A	-473.764	257°

Table 5.5 4' Inositol Monophosphate (PM3).

Ring conformation	Energy of Minimum	P_4-O_4-C_4-C_5
I	-476.572	98°
E	-473.580	124°
1A	-	-
1A / 2E	-473.383	129°
2E	-473.275	94°
2E / 3A	-	-
3A	-466.677	292°
3A / 1E	-472.805	93°
1E	-	-
1E / 2A	-	-
2A	-475.882	114°
2A / 3E	-	-
3E	-467.777	308°
3E / 1A	-473.835	90°

Table 5.6 5' Inositol Monophosphate (PM3).

Ring conformation	Energy of Minimum	P_5-O_5-C_5-C_6
I	-475.238	260°
E	-473.789	251°
1A	-471.215	269°
1A / 2E	-471.982	226°
2E	-471.845	214°
2E / 3A	-	-
3A	-470.983	205°
3A / 1E	-	-
1E	-472.016	242°
1E / 2A	-471.498	243°
2A	-	-
2A / 3E	-	-
3E	-	-
3E / 1A	-468.361	245°

Table 5.7 1' Inositol Monophosphate conformations using the "Anionic" Method.

	AM1		MNDO		PM3	
Ring Conformation	Energy	P_1-C_2	Energy	P_1-C_2	Energy	P_1-C_2
I	-528.344	235°	-425.321	234°	-477.395	241°
E	-521.783	231°	-414.786	244°	-475.274	235°
1A	-	-	-	-	-	-
1A / 2E	-	-	-	-	-	-
2E	-525.236	225°	-418.433	230°	-475.458	233°
2E / 3A	-	-	-416.656	242°	-473.194	240°
3A	-519.336	228°	-	-	-	-
3A / 1E	-	-	-415.734	279°	-	-
1E	-517.437	236°	-	-	-469.580	245°
1E / 2A	-514.457	87°	-415.340	229°	-	-
2A	-	-	-414.374	268°	-472.944	76°
2A / 3E	-523.148	257°	-	-	-	-
3E	-522.087	248°	-	-	-475.151	244°
3E / 1A	-506.482	105°	-412.667	228°	-	-

Comparison of the results for the PM3 conformations using the two methods provide some surprising results, namely that the "anionic" method gives consistently lower energies for comparable ring conformations and that there is a large degree of dissimilarity in which of the ring conformations are represented in the local minima. Mostly this dissimilarity means nothing more than a high energy conformer using one method is not represented in the results of the other method (e.g. the "3A" conformation using the "hydrogen" method) or the equivalent conformation has a small change in the rotation of the ring bonds so that it appears as the adjacent conformation (e.g. the "2E / 3A" conformation using the "anionic" method). A clear exception to this

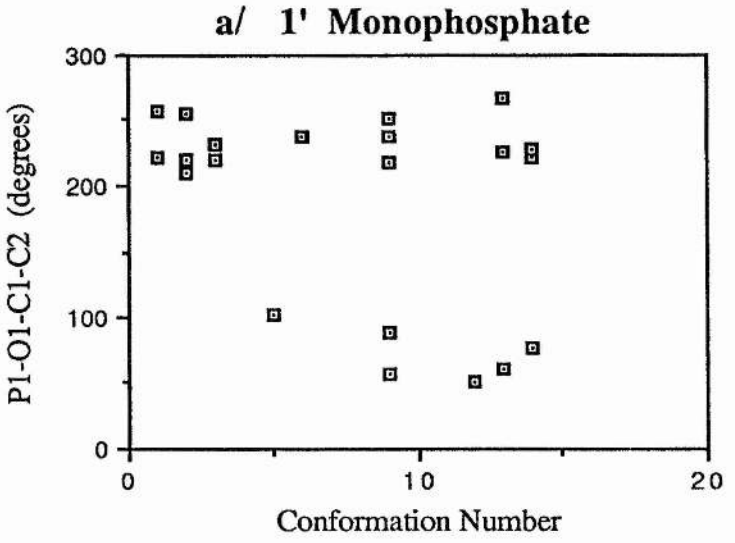
is the "3E" conformation produced using the "anionic" method. It is a low energy conformation, while the only remotely similar conformation produced by the "hydrogen" method is of considerably higher energy though it was produced from a starting point with a 3E ring conformation (and the phosphate dihedral was 60°). It is noticeable from the calculations, that all minima with P_1-O_1-C_1-C_2 dihedrals around 60° have relatively higher energies, while the "3E" starting points for the "hydrogen" method with larger phosphate dihedrals both produced a final ring "1A / 2E" conformation.

As the 1' phosphates are mirror images of the equivalent 3' phosphates, comparison of these two sets of structures would prove valuable (equivalent structures are produced by exchanging "1" for "3" in the names used for the conformations). The first thing to notice is that the energies of the 3' phosphate chair conformations (table 5.4) compare very well with the results using the "anionic" method. The agreement is less good for the boat structures where there is rather more scope for changes in conformation, but it is still noticeably better than the 1' "hydrogen" method results.

All of the mono-phosphates show a large degree of consistency in the orientation of the phosphate groups, being (approximately) 225° for P_1, 250° for P_3, 105° for P_4 and 240° for P_5 (all defined as P_N-O_N-C_N-C_{N+1}), with all angles being measured around the ring so as to increase the carbon index. These angles reflect the fact that P_1, P_3 and P_5 all lies on the same side of the ring with P_4 on the other side, thus the phosphate groups tend to point directly away from the ring (to point exactly away from the ring would require dihedral angles of 240°, or 120° for P_4). The deviations come from the tendency of P_1 and P_3 to be repelled by the hydroxyl at position 2, while P_4 is attracted towards that at position 6.

Comparison of the results produced using the three different parameter sets and the "anionic" method gives a similar picture to that obtained for inositol. The "I" conformation is the most stable, followed by the "2E" conformation. As is the case for the earlier results, phosphate-hydroxyl distance (compared with hydroxyl-hydroxyl distance for inositol) is the main deciding factor in the relative stability of the conformations, but its effect is greatest for MNDO and least for PM3. The relatively

low energy of the "3E" conformations for AM1 and PM3 calculations (and also the "2A / 3E" conformation with AM1) is interesting as this has an eclipsed geometry between P_1 and O_2 such as would be required for the formation of a 1:2 cyclic phosphate (viz. hydrolysis of $PstInsP_2$ and [135]). Its relative stability is greater than for the equivalent inositol conformation. It is important to note that the minimum energy obtained for a given ring conformation is not necessarily produced from the same starting point for each of the parameter sets. Either the ring conformation may move to a different classification as for example is the case with the "3E / 1A" conformation produced by MNDO which comes from the same starting point as produced the minimum energy for the "3E" conformation using AM1 or PM3, or the relative stabilities of minima with similar ring conformation but different phosphate orientation may be changed by the different parameter sets. There is no example of this latter case for the 1' monophosphate, the change in angle for the "2A" conformation between the MNDO and PM3 results is produced by the starting point that gave the minimum energy for "2A" with MNDO parameters giving instead a "3E" ring conformation with PM3.



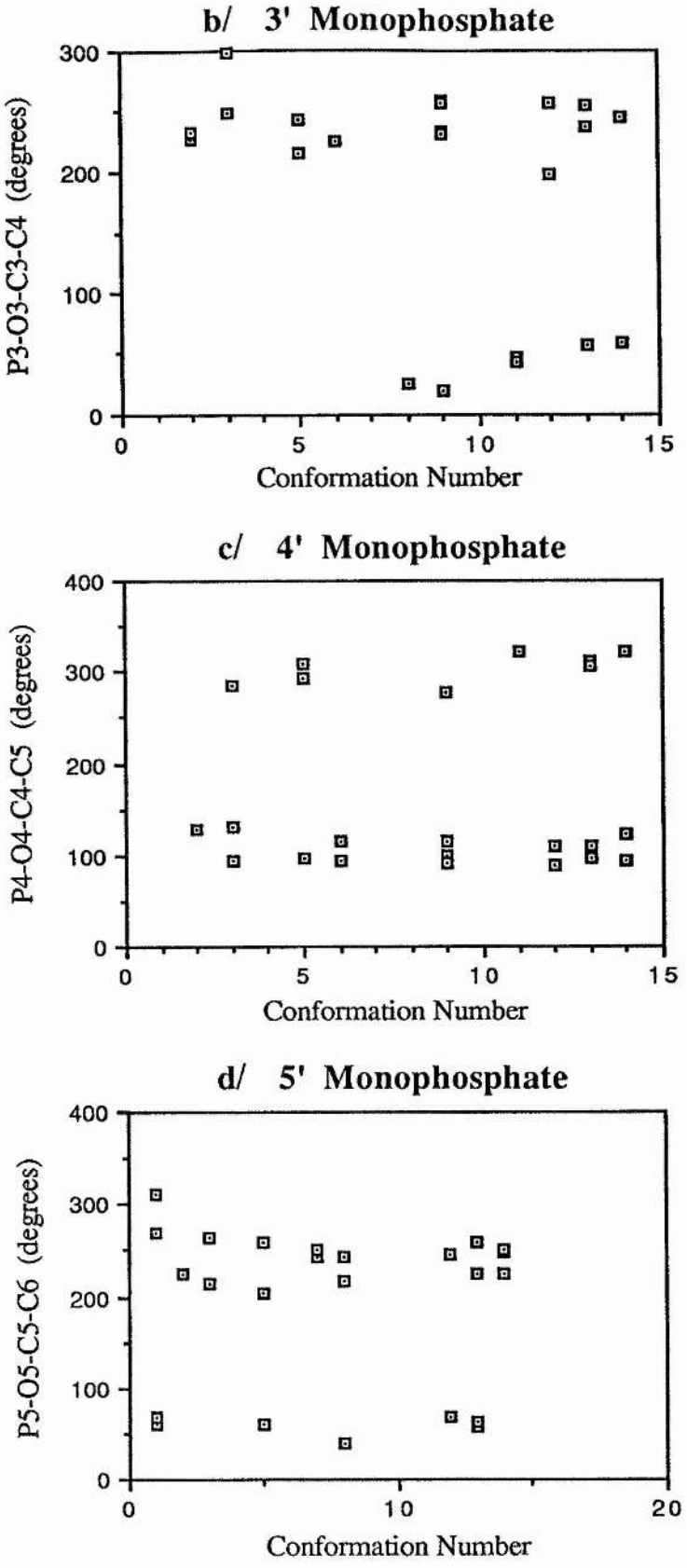


Fig 5.4 Comparison of Phosphate Dihedrals Using the "Hydrogen" Method.

The conformation numbers used in the above and following figure refer to the ring conformations from table 5.2 thus:

Number	Conformation	Number	Conformation
1	"1A"	8	"1E / 2A"
2	"1A / 2E"	9	"2A"
3	"2E"	10	"2A / 3E"
4	"2E / 3A"	11	"3E"
5	"3A"	12	"3E / 1A"
6	"3A / 1E"	13	"I"
7	"1E"	14	"E"

All of the conformations had 3 starting points (dihedrals at 60°, 180° and 300°) so providing a grid across the graphs above. In all cases, there is a favouring of a particular angle which is largely independent of the ring conformation and results in a P-O-C-H dihedral of 0°. The occasional conformations which differ from this generally have a higher energy, though the results for the 4' monophosphate are an exception. This phosphate group is on the opposite side of the ring to all the others and as such suffers from less crowding from the hydroxyl groups and it is not in a position to adopt and eclipsed ring conformation with an adjacent hydroxyl as can happen for the 1' and 3' phosphates.

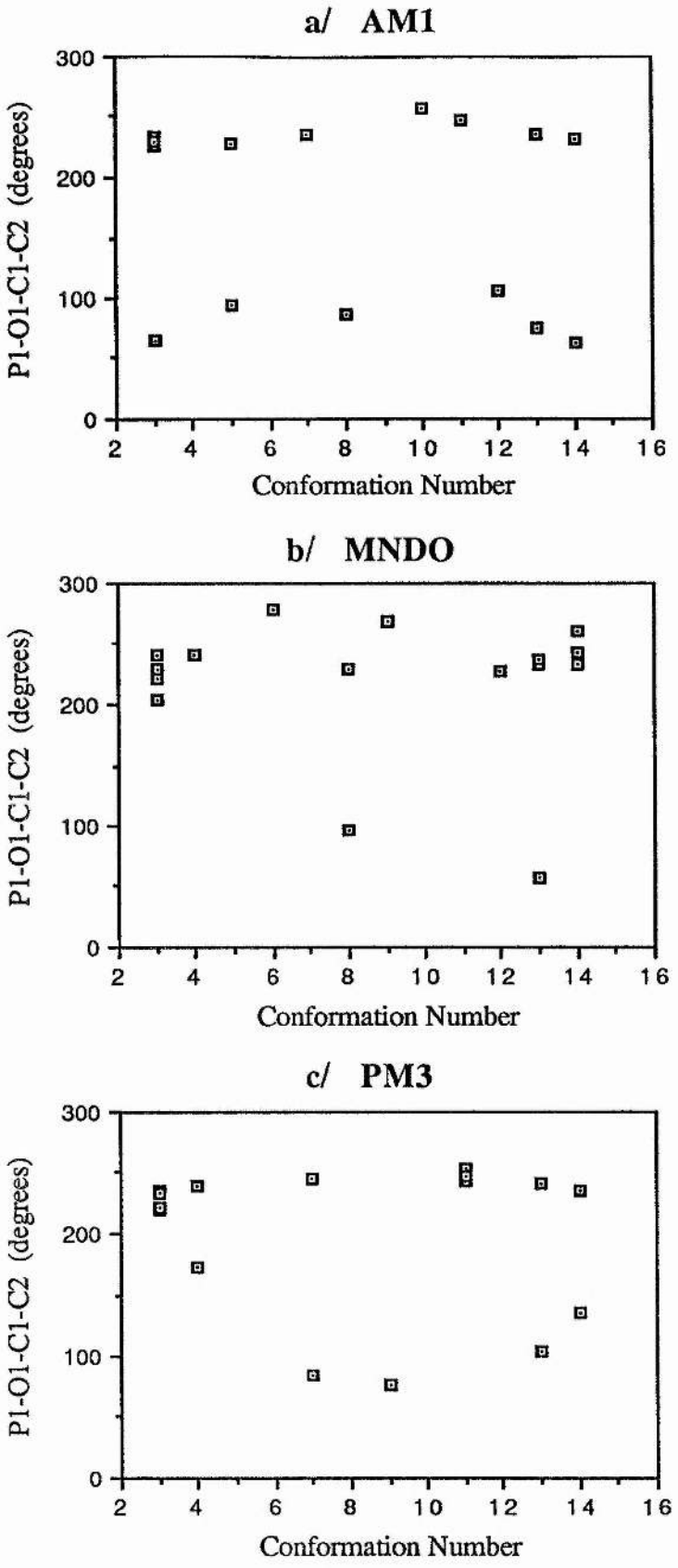


Fig 5.5 Comparison of Phosphate Dihedrals Using the "Anionic" Method.

The distribution of minima with PM3 and MNDO agree well with those from the "Hydrogen" method for the 1' monophosphate, but AM1 gives an equal distribution of minima with dihedrals around 230° and 80° though the lower angled conformations are still of high energy. Even as the monophosphate the 1' phosphate group shows a narrow spread of low energy minima.

5.5 Inositol (1,4) Bisphosphate.

Table 5.8 Inositol (1,4) Bisphosphate Conformations.

	AM1			MNDO			PM3		
	Energy	P_1-C_2	P_4-C_5	Energy	P_1-C_2	P_4-C_5	Energy	P_1-C_2	P_4-C_5
I	-744.867	234	132	-567.204	236	123	-679.261	241	131
E	-742.009	239	119	-560.143	230	101	-680.144	230	123
1A	-	-	-	-	-	-	-	-	-
1A / 2E	-	-	-	-	-	-	-	-	-
2E	-744.710	233	115	-562.727	216	132	-678.256	223	135
2E / 3A	-742.819	226	114	-	-	-	-677.898	236	118
3A	-726.328	16	312	-550.323	75	318	-662.466	86	286
3A / 1E	-	-	-	-557.226	68	97	-	-	-
1E	-	-	-	-551.305	69	126	-669.504	73	119
1E / 2A	-732.365	110	102	-552.866	91	110	-669.696	85	103
2A	-735.025	236	318	-	-	-	-677.132	232	118
2A / 3E	-736.142	241	129	-555.953	259	105	-	-	-
3E	-	-	-	-555.869	262	292	-673.326	229	302
3E / 1A	-	-	-	-	-	-	-	-	-

The study of *myo* -inositol 1 monophosphate summarised in table 5.7 left 20 distinct conformations so generating 60 starting conformations for the study of the 1,4 bisphosphate. Calculations were performed using the "anionic" method which left 24 distinct conformations prior to the addition of the hydrogens, the results of which are shown in table 5.8 above.

As the two phosphate groups are on opposite sides of the molecule they are unlikely to interact to any great extent, so it was expected that the relative energies for the different ring conformations would show a similar pattern to those for inositol and the 1' phosphate. It is surprising therefore to see the "E" chair being more stable than the "I" chair with the PM3 calculations. Though the crystal structures [139-143] suggest that this change should occur as more phosphate groups are added to the inositol ring, the large distance between those at positions 1 and 4 would lead one to expect the change to come with rather more phosphate groups on the ring than with this example. Though neither the AM1 nor the MNDO calculations give the "E" conformation the lower energy, both give it a greater relative stability than was the case for the 1' phosphate alone. The stability of the boat conformations seems to be governed largely by phosphate-phosphate and phosphate-hydroxyl repulsions as the only low energy conformations lie around the "2E" position with the exception of the PM3 "2A" conformation. However in this latter case, both phosphate groups are pointing directly away from each other and the low energy is as a result of a good hydrogen bond between the 2' hydroxyl group and O_5. It is questionable whether this conformation would be an important structure in solution as the 2' hydroxyl group would be able to form hydrogen bonds with the solvent so reducing the stabilising effect of this particular internal hydrogen bond.

In contrast to the earlier results, the "3A" and "3E" conformations which have O_1 and O_2 eclipsed, as in the 1:2 cyclic phosphate, have a significantly higher energy than the minimum (between 7 and 19 kcal mol^{-1} depending on the parameter set).

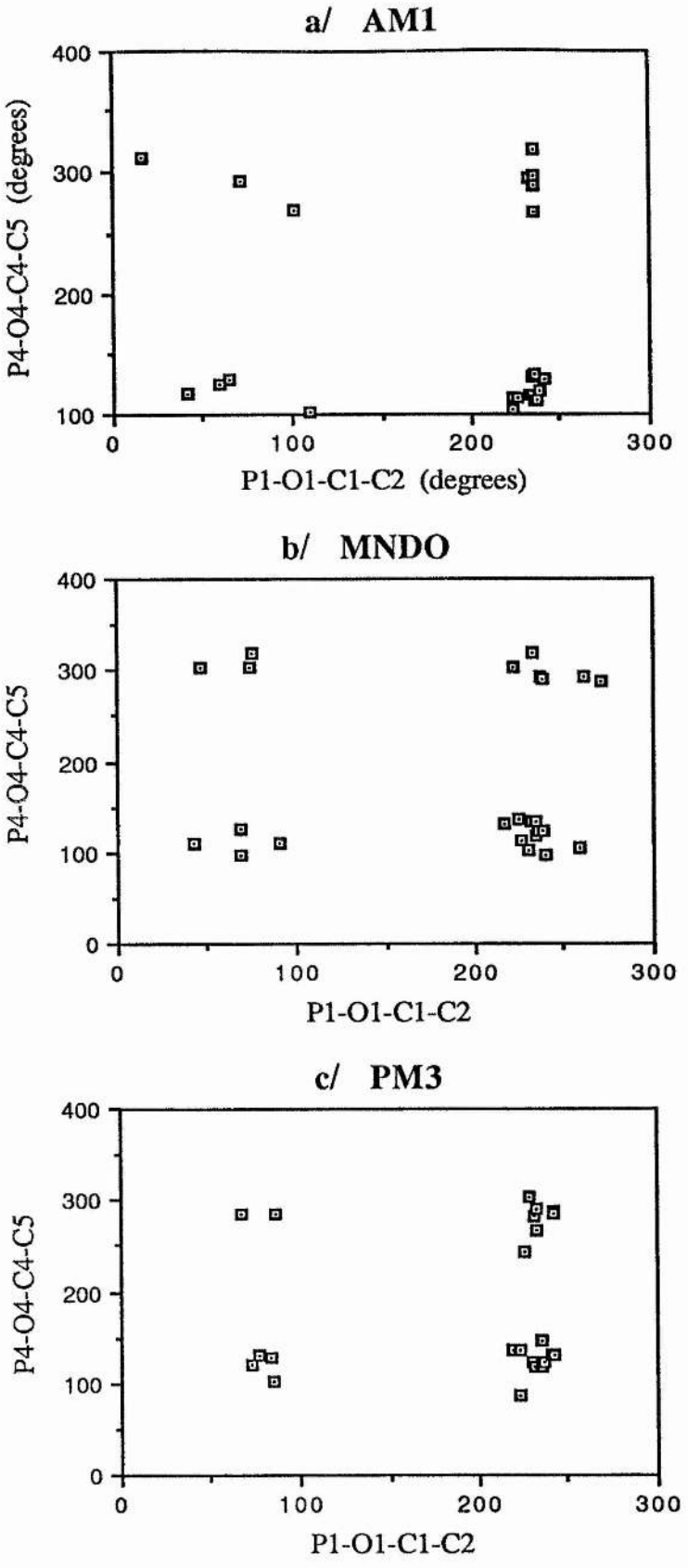


Fig 5.6 Distribution of Minima for *Myo* -Inositol (1,4) Bisphosphate.

As would be expected from the distributions for the monophosphates, there are essentially 4 clusters of minima for all three parameter sets, with most of the low energy conformations found in the bottom right cluster. The graphs are very similar to the those that would be produced by overlaying the 1' and 4' monophosphate results suggesting that there is little interaction between the two phosphate groups, which is not surprising in view of the distance between them.

5.6 *Myo* -Inositol Trisphosphates.

Both the (1,3,4) and (1,4,5) trisphosphates could be built up from the (1,4) bisphosphate of section 5.5. This earlier stage had produced a total of 24 distinct minima resulting in 72 starting points for each of the trisphosphates and as before, the "anionic" approach to the calculations was used. The phosphate positions in the two trisphosphates are essentially mirror-images of each other along the 1'-4' axis, the only difference between the molecules being the relative position with respect to the ring of the 2' and 6' hydroxyl groups (they lie on opposite sides of the ring). That this difference is sufficient to give the two molecules different equilibrium conformations is clear from the fact the the (1,4,5) derivative is active whereas the (1,3,4) derivative is not. It is this asymmetry in the hydroxyl groups that is responsible for there being only 15 distinct minima after the first stage of the calculations for the (1,4,5) trisphosphate against the 17 for the (1,3,4) trisphosphate.

The minimum energies corresponding to the different ring conformations of the two isomers are given below in table 5.9.

Table 5.9a Energies of *Myo* -Inositol (1,4,5) Trisphosphate.

	AM1				MNDO			
	Energy kcal mol^{-1}	P_1-C_2	P_4-C_5	P_5-C_6	Energy kcal mol^{-1}	P_1-C_2	P_4-C_5	P_5-C_6
I	-967.392	239	114	237	-711.495	241	113	227
E	-963.513	240	110	231	-701.904	223	125	231
1A	-	-	-	-	-	-	-	-
1A / 2E	-	-	-	-	-701.311	64	114	243
2E	-961.058	222	118	248	-703.503	232	115	248
2E / 3A	-965.735	228	110	248	-706.639	275	125	253
3A	-960.857	27	103	244	-	-	-	-
3A / 1E	-948.319	359	124	30	-699.673	83	108	241
1E	-	-	-	-	-700.844	235	138	237
1E / 2A	-957.349	238	130	241	-702.403	239	111	253
2A	-	-	-	-	-	-	-	-
2A / 3E	-952.886	260	132	244	-698.433	275	125	253
3E	-	-	-	-	-	-	-	-
3E / 1A	-	-	-	-	-	-	-	-

	PM3			
	Energy (kcal mol^{-1})	P_1-C_2	P_4-C_5	P_5-C_6
I	-883.113	245	114	240
E	-884.105	219	124	256
1A	-	-	-	-
1A / 2E	-	-	-	-
2E	-881.390	211	124	274
2E / 3A	-883.725	233	115	238
3A	-879.629	228	124	240
3A / 1E	-873.604	249	130	49
1E	-876.260	109	121	249
1E / 2A	-882.615	250	111	232
2A	-878.557	254	151	253
2A / 3E	-880.565	254	111	251
3E	-	-	-	-
3E / 1A	-	-	-	-

Table 5.9b Energies of *Myo* -Inositol (1,3,4) Trisphosphate.

	AM1				MNDO			
	Energy kcal mol^{-1}	P_1-C_2	P_3-C_4	P_4-C_5	Energy kcal mol^{-1}	P_1-C_2	P_3-C_4	P_4-C_5
I	-964.335	252	238	123	-711.004	255	209	116
E	-956.819	237	237	119	-699.638	228	258	124
1A	-	-	-	-	-	-	-	-
1A / 2E	-	-	-	-	-	-	-	-
2E	-963.575	219	242	127	-	-	-	-
2E / 3A	-959.771	230	251	118	-	-	-	-
3A	-958.377	238	235	117	-705.158	260	265	91
3A / 1E	-954.022	264	249	127	-703.179	238	240	118
1E	-953.366	86	224	121	-701.637	195	213	121
1E / 2A	-957.400	233	242	116	-703.527	204	220	137
2A	-968.012	232	257	130	-	-	-	-
2A / 3E	-959.283	236	239	111	-701.624	265	223	123
3E	-	-	-	-	-	-	-	-
3E / 1A	-	-	-	-	-	-	-	-

	PM3			
	Energy (kcal mol^{-1})	P_1-C_2	P_3-C_4	P_4-C_5
I	-883.803	273	215	124
E	-880.680	224	238	114
1A	-	-	-	-
1A / 2E	-	-	-	-
2E	-	-	-	-
2E / 3A	-878.278	237	238	120
3A	-877.422	265	265	124
3A / 1E	-879.483	267	250	128
1E	-	-	-	-
1E / 2A	-	-	-	-
2A	-883.907	238	244	106
2A / 3E	-	-	-	-
3E	-	-	-	-
3E / 1A	-	-	-	-

One of the most notable differences between the two isomers is the relative stability of the two chair conformations. With AM1 and PM3, the "I" conformation has its relative stability over the "E" form favoured by around 4 kcal mol^{-1} in going from the 1,4,5 isomer to the 1,3,4 isomer and in the case of the PM3, this is sufficient to bring the energy of the "I" form to below that of the "E" form. With MNDO, the difference is only about 1.5 kcal mol^{-1}. As would be expected with an increasing number of phosphate groups, the variation in the P_N-O_N-C_N-C_{N+1} dihedral angles becomes increasingly more restricted. This is especially true for positions 3 and 4, while minima with dihedrals for 1' and 5' phosphates that differ sharply from the usual values generally have noticeably higher energy. There are some exceptions to this, such as the "3A" conformation of the (1,4,5) isomer found using the AM1 parameters,

when the structure is stabilised by inter-phosphate hydrogen bonding.

The asymmetry of the two isomers is shown in the relative energies of the different boat conformations. The "2A" conformations for the (1,3,4) isomer is the most stable conformation with both the PM3 and AM1 parameters (no minimum was found using MNDO and neighbouring minima at "1E / 2A" and "2A / 3E" are not of especially low energy). This conformation allows for strong hydrogen bonding between the 3' and 4' phosphate groups and a weaker, longer bond between 1' and 3'. This latter bond will not be important at all in solution as, presumably, it will be readily solvated. That MNDO does not give a low energy minima in this region is therefore not surprising as it is known not to reproduce hydrogen bonding and the exaggerated core-core repulsion will tend to push the relatively close 3' and 4' phosphates apart towards a "1E" conformation. The converse of this can be seen from the scarcity of minima in regions close to the "2A" region with PM3 suggesting that the energy contribution of the 3'-4' inter-phosphate hydrogen bond is sufficiently large to overcome any other factor in deciding the location of minima in that area. Of the 17 conformations used in the second stage of the PM3 calculation, 5 gave final structures with an "I" ring conformation and 7 with a "2A" ring conformation. Stewart, in the initial PM3 publication, claims that PM3 more accurately reproduces hydrogen bonds [63, 64], a claim which appears to be supported by the results shown here. While the "2A" conformation has a greater relative stability with AM1, the hydrogen bond is opposed by the greater core-core repulsion, thus other, albeit higher energy, minima are found in the vicinity (and only two of the second stage calculations gave a final "2A" conformation). From the point of view of the phosphate groups, the equivalent (1,4,5) conformation to the "2A" conformation of (1,3,4) is "3A". However, this conformation is eclipsed across position 1 and 2 and so for all three parameter sets, is rather less stable than the neighbouring "2E / 3A" conformation, which relieves much of the close contact between positions 1 and 2, but allows strong hydrogen bonding between the 4' and 5' phosphate groups. A crucial difference between the two isomers is that this conformation is still less stable than the chair form.

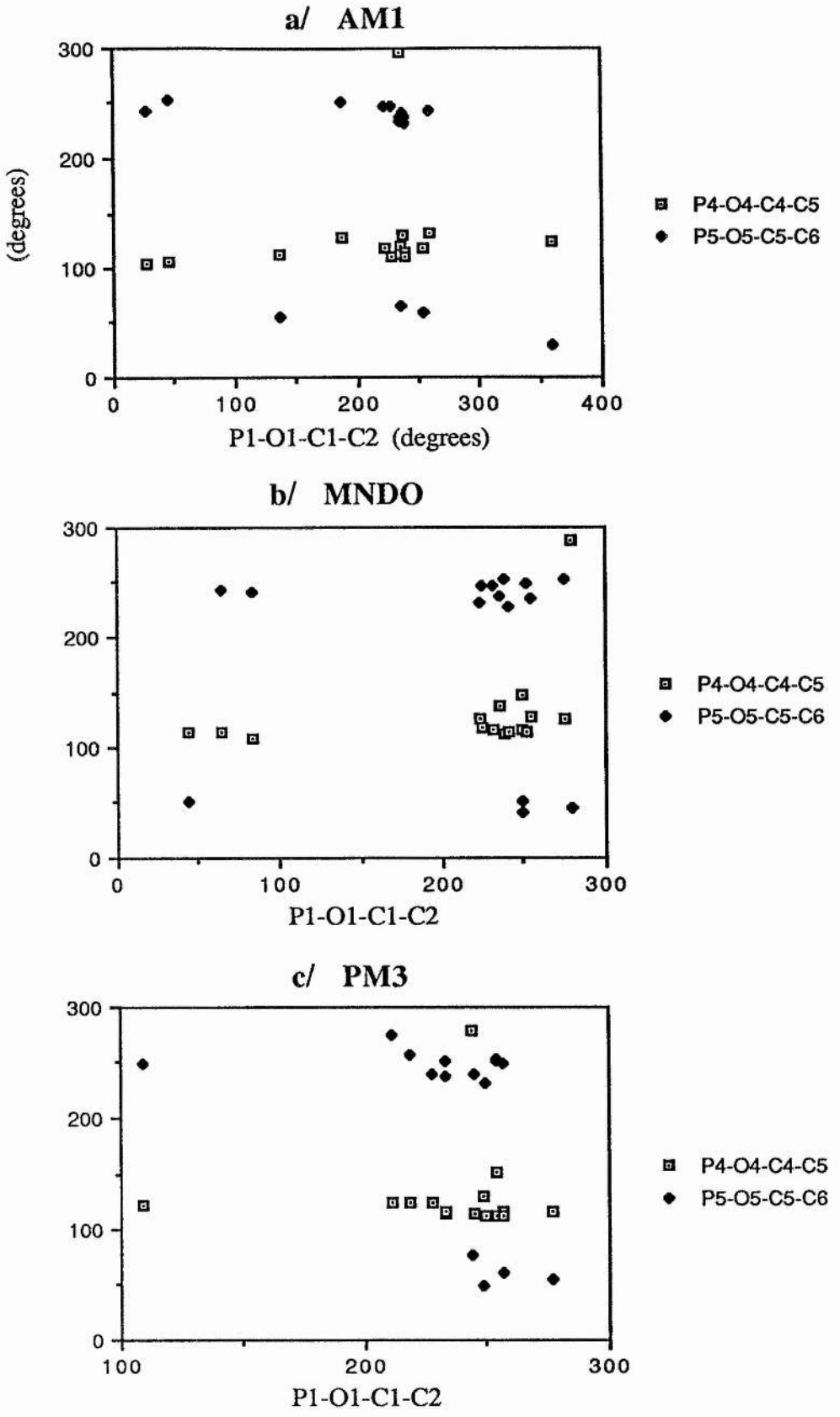


Fig 5.7 Distribution of Minima for *Myo* -Inositol (1,4,5) Trisphosphate.

In both this (fig. 5.7) and the following (fig. 5.8) sets of graphs, the 1' phosphate dihedrals are given on the X-axis and the remaining dihedrals on the Y-axis and the points marked by different symbols corresponding to the angle used for the Y-axis. The different range of the X-axis for different parameter sets is an artefact of the plotting program which automatically sets the range depending on the largest and smallest values to be plotted.

The most interesting feature, that is common to all three parameter sets, is that the 4 clusters of minima for the 1' vs 4' plot (from the bisphosphate results) are reduced to just 2 (with the exception of a single point). A dihedral in the region of 300° brings the 4' phosphate close to the 5' phosphate so that repulsion between them tends to abolish the minimum. The one such minima that survives has an "I" ring conformation but is of extremely high energy (>10 kcal mol^{-1} higher than the lowest energy form with the same ring conformation) while the 5' dihedral is close to 60° so the group points almost directly away from the 4' phosphate. Against this, the 1' vs. 5' dihedral plot shows the same 4 clusters as seen earlier with the lowest energy conformations in the top right. As would be expected, this region also contains the greater number of minima while the higher energy of minima in the bottom left is reflected in their comparative scarcity. With one exception all of the minima with small values of the 5' dihedral have an "I" ring conformation (the other is in a "3A / 1E" conformation) for all three parameters sets. This conformation allows the phosphate groups to form a strong hydrogen bond with the 2' hydroxyl group (the other such point forms a strong hydrogen bond with the 3' hydroxyl group but is opposed by the relatively close approach of the 4' phosphate.

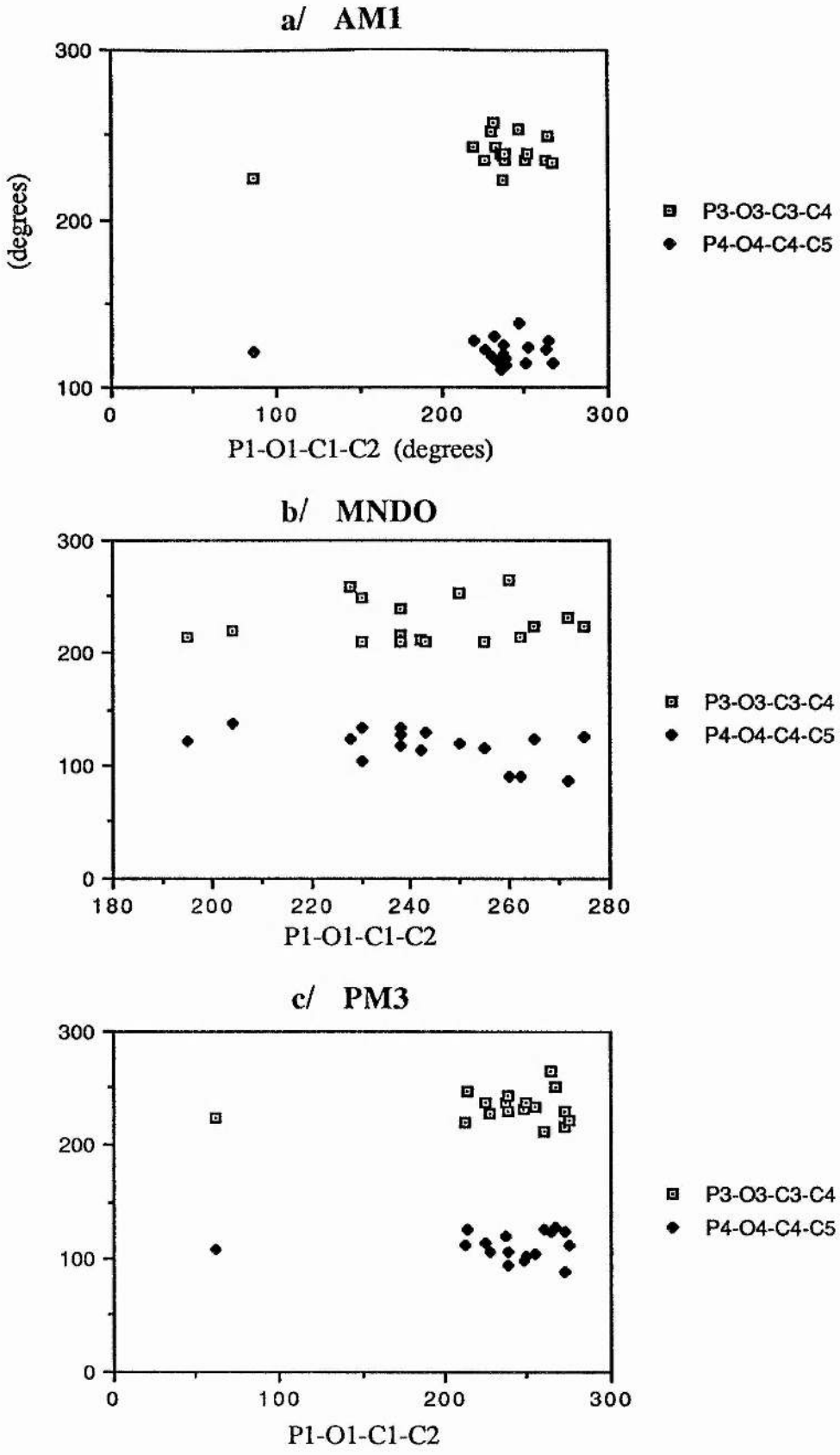


Fig 5.8 Distribution of Minima for *Myo* -Inositol (1,3,4) Trisphosphate.

The minima for the (1,3,4) phosphate are much more restricted than for the (1,4,5) isomer. Neither the 3' nor the 4' phosphate groups show any variation in the dihedrals in the minima, while only a single conformation diverges from the norm for the 1' dihedral (PM3 and AM1 calculations only). Unlike the 5' phosphate of (1,4,5) trisphosphate, the 3' phosphate is heavily constrained by the 4' phosphate and the 2' hydroxyl and so in turn places a restriction on the position of the 4' phosphate. The MNDO results for the 1' dihedral are rather different from those of the other two parameter sets. The small angle minimum does not exist, the same starting point as produced it for the other two sets gives a dihedral of 238°, but the cluster of minima centred on 240° is rather more spread out and shows no correlation with ring conformation. This may well lie with the lack of a hydrogen-bonding potential with MNDO calculations leading to the position of the minima being dictated by a large number of core-core repulsion interactions rather than a single (or small number) of stronger hydrogen-bonding interactions, thus leading to a greater spread in the position of minima within a given cluster.

5.7 *Myo* -Inositol (1,3,4,5) Tetrakisphosphate.

The starting points for the tetrakisphosphate can be built from either of the two trisphosphates considered in the above section. The (1,4,5) trisphosphate calculations were completed before those on the (1,3,4) isomer, so they were used to generate 45 starting points for the tetrakisphosphate (the 1,3,4 isomer would give 51 starting points). The number of starting points could have been reduced still more by taking only starting points compatible with both the (1,4,5) and (1,3,4) isomers, but it it was preferable to begin the tetrakisphosphate calculations (on the VAX 11/785's) before the (1,3,4) trisphosphate calculations were completed on the FPS-500.

The 45 starting points yielded only 9 different conformations at the end of the first stage as the problem of steric crowding severely restricted the allowed rotations of the phosphate groups. In view of this small number of final distinct conformations, the

regions without a local minima have not been included in the table summarising the relative energies (table 5.10).

Table 5.10a AM1 Energy Minima for *Myo* -Inositol (1,3,4,5) Tetrakisphosphate.

	Energy (kcal mol^{-1})	P_1-C_2	P_3-C_4	P_4-C_5	P_5-C_6
I	-1183.779	241	252	125	255
2E	-1180.505	233	229	132	234
3A	-1180.187	242	249	104	242
3A / 1E	-1172.073	108	250	123	230
1E	-1175.895	192	262	109	185
1E / 2A	-1171.224	244	235	123	231
2A / 3E	-1182.187	240	250	131	261

Table 5.10b MNDO Energy Minima for *Myo* -Inositol (1,3,4,5) Tetrakisphosphate.

	Energy (kcal mol^{-1})	P_1-C_2	P_3-C_4	P_4-C_5	P_5-C_6
I	-852.252	238	232	110	253
E	-844.641	221	252	131	246
2E / 3A	-846.857	273	240	128	240
3A	-847.823	241	275	127	215
3A / 1E	-844.378	196	222	123	237
1E	-843.572	230	228	103	222
1E / 2A	-842.446	266	227	120	234

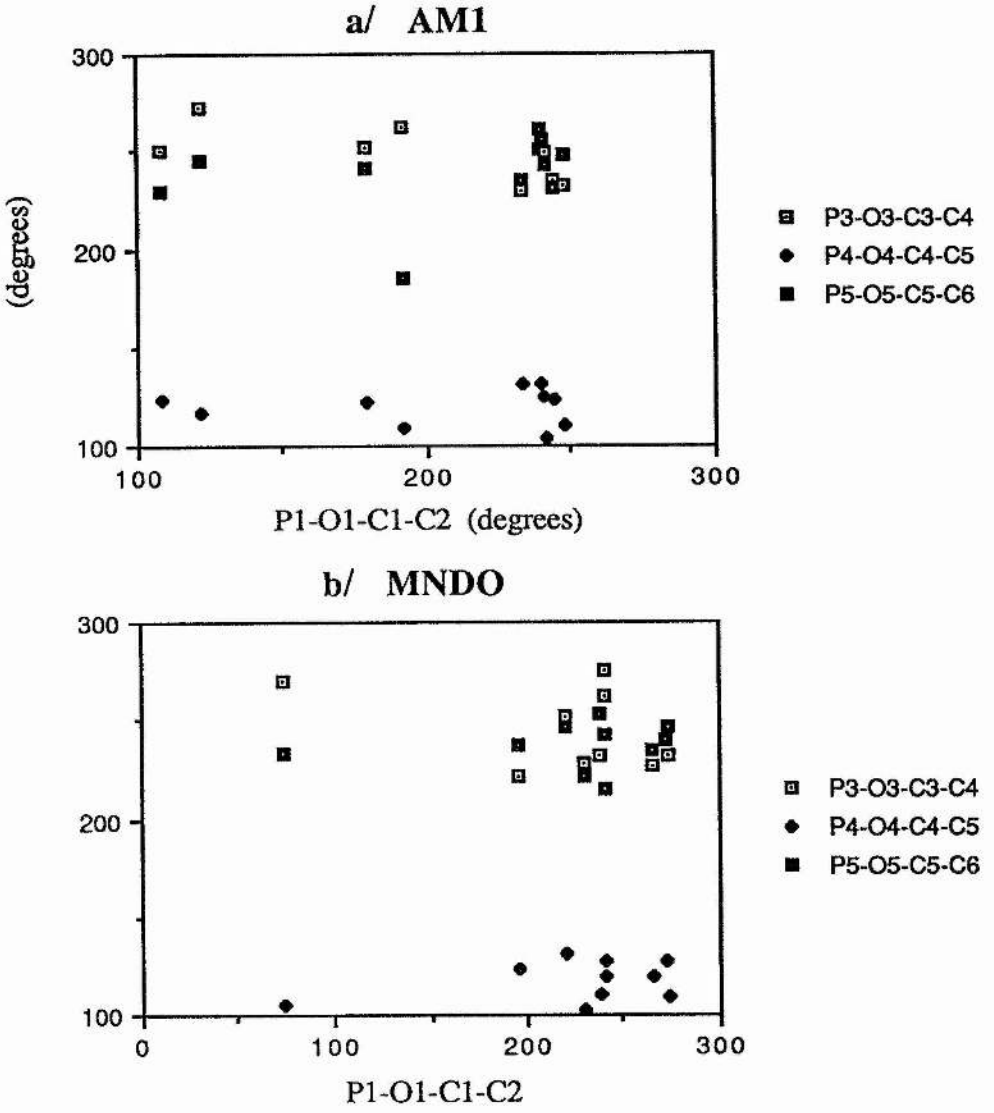
Table 5.10c PM3 Energy Minima for *Myo* -Inositol (1,3,4,5) Tetrakisphosphate.

	Energy (kcal mol^{-1})	P_1-C_2	P_3-C_4	P_4-C_5	P_5-C_6
I	-1085.238	269	220	115	260
E	-1084.061	222	266	124	253
3A	-1084.656	247	239	127	232
3A / 1E	-1085.148	270	268	128	240
1E	-1084.133	254	255	112	211
1E / 2A	-1083.449	262	213	114	250
2A	-1087.034	253	203	112	223

Only 10 distinct minima were represented at the end of stage 1 of the calculations, mostly with different ring conformations and so most are represented in the stage 2 results shown above. Calculation of minima in the "E" conformation caused considerable problems. From the earlier results and from consideration of the crystal structures [139-143], this conformation would be expected to be relatively stable, but none of the final geometries possessed an "E" ring conformation. All the starting points with this conformation had a "3A" ring conformation at the end of the first stage, so new starting points were generated with an "E" ring conformation and with the phosphate dihedrals set to fit with previous results (phosphate hydrogens were included). These new starting points gave "E" ring conformations with PM3 and MNDO but a "2A / 3E" conformation with AM1 and only the latter was of lower energy than the other boat conformations.

The tetrakisphosphate minima have a greater similarity to the (1,3,4) trisphosphate minima than to those of the (1,4,5) trisphosphate they were derived from. This is particularly interesting as both the (1,3,4,5) tetrakisphosphate and (1,3,4) trisphosphate lack the intracellular calcium releasing activity of the (1,4,5) trisphosphate and produced as a by-product of its metabolic inactivation. It also lends credence to the results of the "anionic" method as it shows that the final geometries are not constrained

by the choice of build-up path. The implications of the relative energies of the minima in terms of activity are discussed in section 5.10.



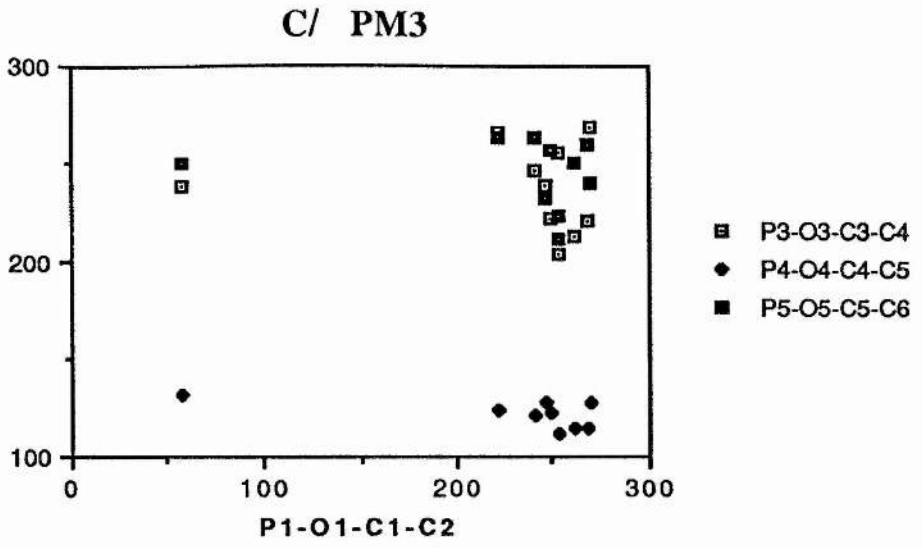


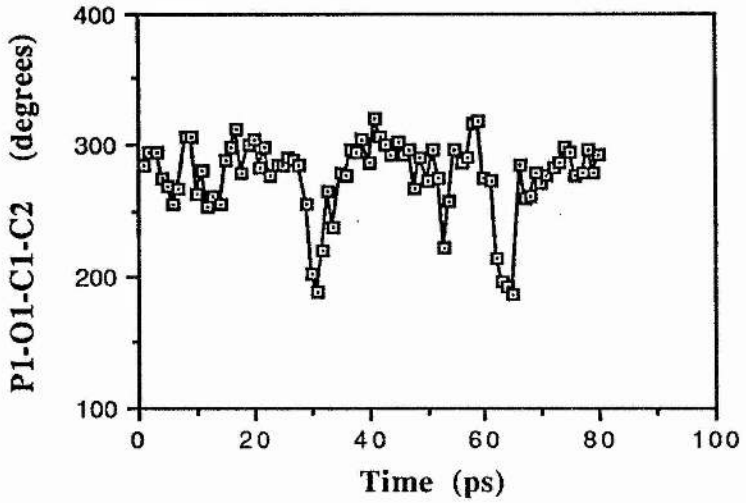
Fig 5.9 Distribution of Minima for *Myo* -Inositol (1,3,4,5) Tetrakisphosphate.

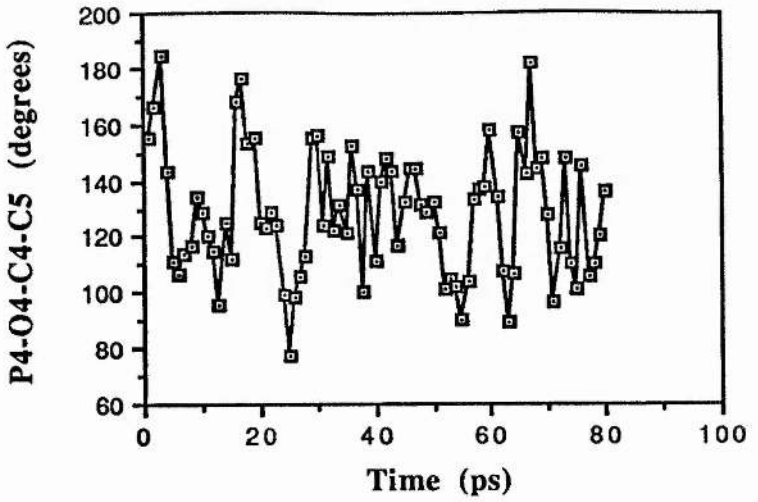
Again all the dihedrals are superimposed on the one graph for each parameter set in the same manner as for the trisphosphates and as would be expected show considerable restriction in the range of 3', 4' and 5' dihedrals represented in the minima. The exact angles vary only slightly with the three parameter sets reflecting that steric considerations are paramount in deciding the position of the minima. Most noticeable is the lack of minima with 5' dihedrals less than 180°. The increased steric crowding mitigates against conformations that differ from the alternating up and down phosphate groups (taken with the ring horizontal) and the 4' phosphate is less able to move to accommodate the 5' phosphate with a dihedral in the range 0°-180°.

Rather more variation is found in the range of 1' dihedrals are here there is less steric crowding in this part of the molecule, so the balance between the different long-range forces in the three parametrisations are able to have a greater effect on the position of the minima.

5.8 Molecular Dynamics.

Molecular dynamics calculation were performed on *myo* -inositol (1,4,5) trisphosphate both in a water bath and in the gas phase. Both sets of calculations were performed with a fixed number of particles at constant volume and a temperature of 300 K. A value of 16 was chosen for the dielectric constant to reflect the polarisation of the solvent by the phosphate groups. For the gas phase calculation, a distance dependant dielectric constant was used to mimic the electrostatic shielding effect of the water molecules. Thus the results from the gas phase calculation would be expected to be similar to those from the water-bath calculations except for solvent-solute hydrogen bonding. In both sets of calculations a time step of 0.002 ps was used and the bond lengths were constrained. Equilibration lasted for 40 ps with data points every 2 ps followed by data collection over 80 ps sampling every 1 ps. The ring and phosphate dihedral angles were monitored over this period and the results are presented in the following graphs.





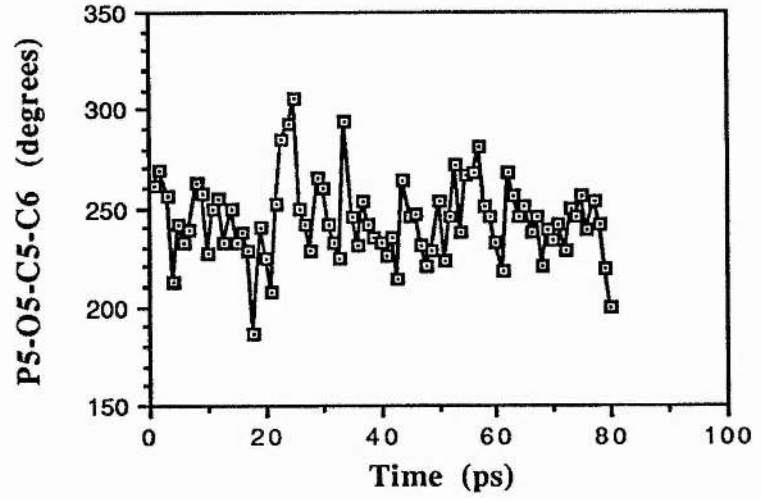


Fig 5.10 Phosphate Dihedrals vs Time for Dynamics in a Water Bath.

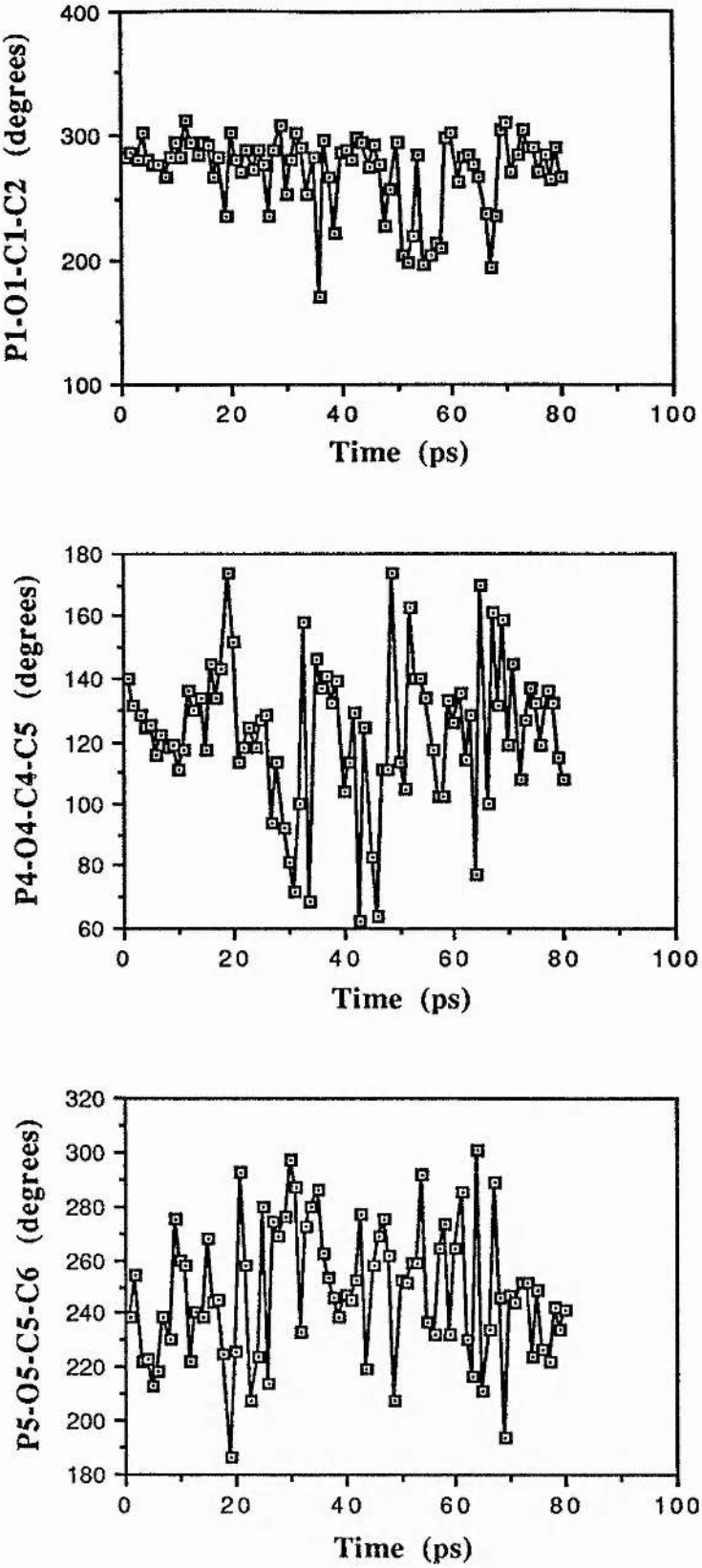


Fig 5.11 Phosphate Dihedrals vs. Time for Gas-phase Dynamics.

The ring conformation remained in the "I" chair form throughout both calculations showing that the barrier heights for conversion between the two chair forms or for the chair-boat transition are rather high. To confirm this, the gas-phase calculation was repeated at 800 K when the ring conformation still remained in the "I" form and the only change from the lower temperature calculation was increased mobility of the phosphate groups (results not shown).

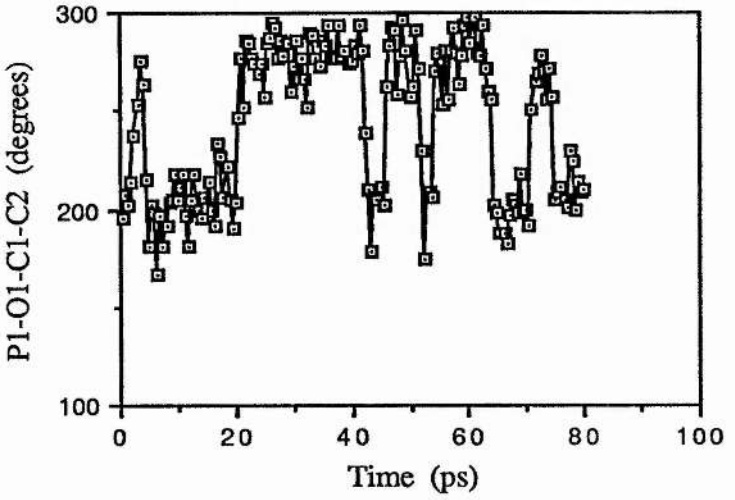
At 300 K in the water bath, both the 1' and 5' phosphates show little mobility being confined to dihedrals of around 280° and 240°. Conformations with dihedrals substantially different to these values (sufficient to suggest that they lie in different minima) are short lived suggesting that the corresponding minima are comparatively shallow. The consistency of the dihedrals compare well with the results from the semi-empirical calculations which show these two dihedrals both confined to values around 240° (AM1) or slightly higher in the case of MNDO and PM3 calculations for the 1' phosphate. The most common conformation in the gas-phase calculation also has a dihedral for the 1' phosphate of about 280° but shows more frequent deviations from this, suggesting a lower barrier height than when in solution, though a much longer calculation with considerably more transitions would be needed to give a reliable estimate of the barrier height. The lack of any long-lived dihedral for the 5' phosphate in the gas-phase shows that there are no strong interactions between it and other groups of the molecule. By comparison, the conformation in solution is extremely stable suggesting the presence of a bridging water molecule between the 5' phosphate and another group, potentially the 1' phosphate as it lies on the same side of the molecule. In both calculations, the 4' phosphate is considerably more mobile than the other two (which are both on the other side of the molecule). The longest-lived conformation has a dihedral of 130° which compares well with the minima found in the semi-empirical studies and the rapid fluctuations in the dihedral can be attributed to a shallow potential at the minimum rather than several minima all of the other conformations are short-lived. There does not appear to be any correlations between deviations in two or more of the dihedrals except at around 30 ps in solution 1' and 5' dihedrals show substantial divergence from the most common conformation. There is however no corresponding

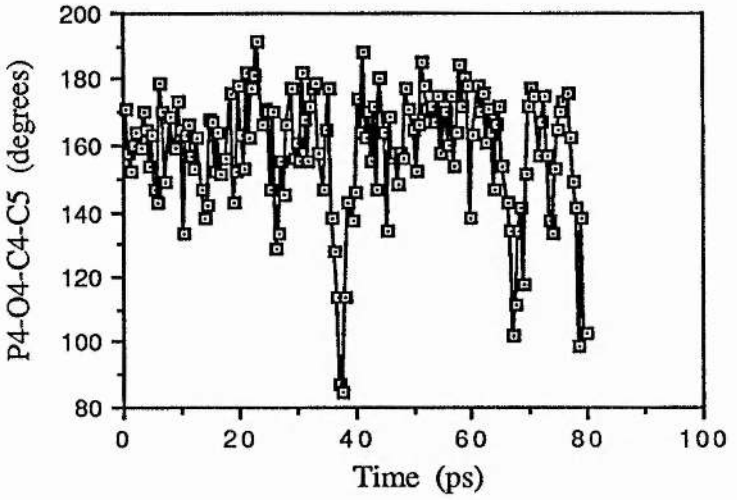
change in the 5' dihedral at 60 ps suggesting that the 30 ps event is coincidental.

The results for a similar calculation starting with a boat conformation ("2E") in a water bath is shown in fig.5.12. Several differences from the results using a chair conformation are noticeable, the first being that a particular boat conformation is rather less stable than the chair conformation. After equilibration, the initial "2E / 3A" ring conformation had rotated to a "1A / 2E" conformation which also turned out to be the more stable form during the data collection. The ring conformations found differ from the results of the semi-empirical study reflecting the reduced electrostatic repulsion between the phosphate groups due to the change in dielectric constant. Differences are also found in the rotation of the phosphate groups with the chair and boat conformations. In the latter case both the 1' and to a lesser extent the 5' phosphates show bistable conformations, though for both conformations, the 4' phosphate does not. The equilibrium dihedral of the 4' phosphate is increased slightly to 160° as a result of the greater separation from the 3' hydroxyl, while the two bistable dihedrals for the 1' phosphate are the same as for the chair conformation. However the small angle dihedral (200°) is considerably very short lived in the "I" conformation, its extra stability with the "2E" coming from 3'-1' hydrogen bonding (though the results suggest that it is still less stable than the large dihedral conformation especially for the "1A" ring conformation). The most stable 5' dihedral for the chair conformation was around 240°, but with the boat conformation, this becomes the less stable of the two bistable dihedrals, the more stable being at 280°-290°, again different from the position in the semi-empirical studies. This is surprising as there is essentially no change in the immediate environment of the 5' phosphate in changing from an "I" ring conformation to a "2E" conformation, though changes at the 1', 2' and 3' positions may be communicated to the 5' phosphate via the 4' and 6' groups.

It has been mentioned that the higher value of the dielectric used in the dynamics calculations would have an effect on the phosphate-phosphate repulsions compared with the semi-empirical studies, so dynamics calculation was run with a dielectric constant of 2. in this case the electrostatic repulsion between the phosphate groups was sufficient to hold the ring conformation rigidly in the "2E" form and to fix the 1', 4'

and 5' dihedrals at 190°, 125°, and 240° respectively. The latter two values correspond almost exactly to those found from the semi-empirical studies while that for the 1' phosphate is slightly lower, a parameter dependent effect, it is clearly the equivalent small angle minima rather than the large angle minima favoured in the dynamics calculation with a higher dielectric. This shows clearly the problems associated with the choice of dielectric, and the dependence of conformation of these highly polar molecules on their environment.





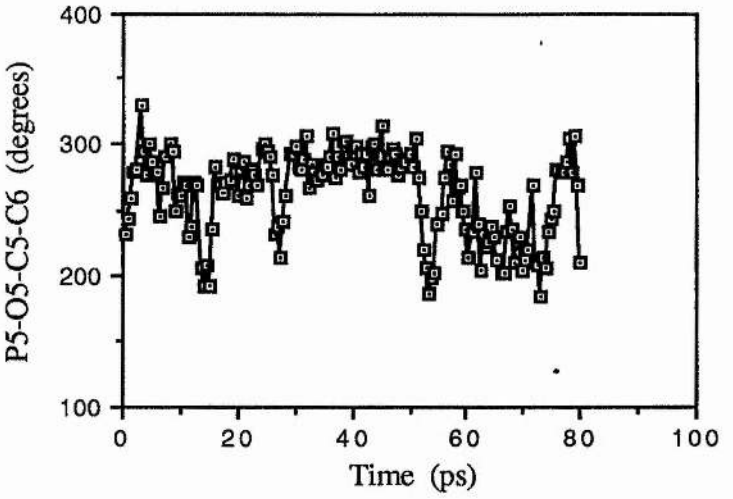


Fig 5.12 Phosphate Dihedrals and Ring Conformation for Water-Bath Dynamics on IP_3 in a Boat Conformation.

5.9 Properties of *Myo* -Inositol (1,4,5) Trisphosphate.

As a further means of comparison between the different parametrisations, single-point calculations were carried out on the "I" minimum energy conformation to determine thermodynamic properties of the molecule along with some one-electron properties and a more detailed description of the HOMO (Highest Occupied Molecular Orbital) and LUMO (Lowest Unoccupied Molecular Orbital) and the results are described below. There are no experimental values for these properties in the literature so it is not possible to compare the accuracy of the three parametrisations. The minima in the "I" conformation was chosen as it had a considerably lower energy with AM1 and MNDO and was only around 1 kcal mol^{-1} higher energy than the lowest with PM3.

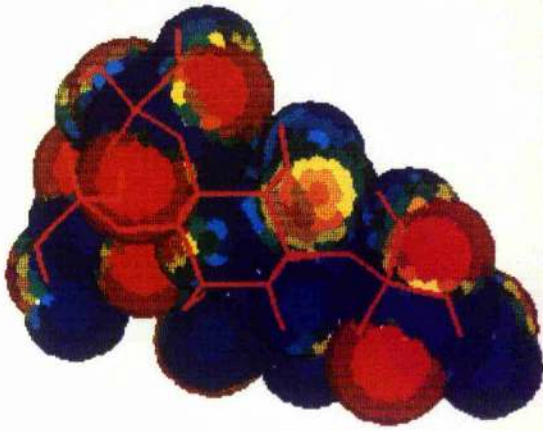
Table 5.11 *Myo* -Inositol (1,4,5) Trisphosphate Properties.

	AM1	PM3	MNDO
Total Energy (eV)	6179.5	5722.7	6286.281
Ionisation Potential (eV)	11.305	10.631	11.068
Dipole Moment (Debye)	6.061	5.407	3.887
Mean Phosphorus Charge	2.638	2.206	1.369
Heat of Formation (kcal mol^{-1} at 300 K)	-967.228	-882.896	-711.291
Enthalpy (kcal mol^{-1} at 300 K)	17.866	19.313	18.329
Heat Capacity (kcal K^{-1} mol^{-1} at 300 K)	0.1005	0.1088	0.1007
Entropy (kcal K^{-1} mol^{-1} at 300 K)	0.1811	0.1983	0.1909

The effect of the different phosphorus charges can be seen in the electrostatic potentials drawn in Fig.5.13 and is largely responsible for the parallel differences in the dipole moment. Apart from this relationship, there are no visible trends in the differences in the calculated properties and the greatest variation is found in the heat of formation and dipole moment which are used in the parametrisation rather than in the thermodynamic properties.

An understanding of the electrophilic / nucleophilic behaviour of a molecule can be obtained from the Hartree-Fock calculation by comparing the atomic orbital coefficients of the highest occupied molecular orbital (the HOMO) and the lowest unoccupied molecular orbital (the LUMO). The electron lost in the first ionisation is lost from the HOMO and its location on the molecule suggests the sites of potential electrophilic attack, while the LUMO is first to gain an electron and so is a site of potential nucleophilic attack. The location of orbitals with significant contributions to the HOMO and LUMO for each of the three parameter sets are given in the following tables:

a/ AM1



b/ PM3

c/ MNDO

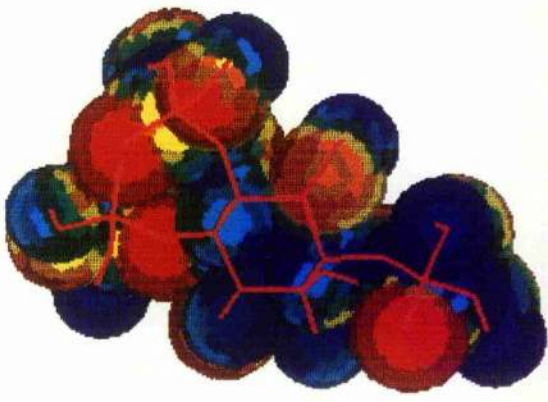


Fig 5.13. EPM's of Myo-Inositol (1,4,5) Trisphosphate.

Table 5.12 Atomic Orbitals with Significant Eigenvectors for the HOMO and LUMO.

a/ AM1

HOMO			LUMO		
Atom	Orbital	Eigenvector	Atom	Orbital	Eigenvector
C2	p	0.250	C1	p	0.108
H2	s	0.172	P1	s	0.214
O2	p	0.163	C4	p	0.237
C3	p	0.297	O4	s	0.147
O3	p	0.486		p	0.256
C4	p	0.362	P4	s	0.592
O4	p	0.322		p	0.184
C5	p	0.248	O=P4	p	0.247
O5	p	0.278	O-P4a	s	0.110
C6	p	0.174		p	0.283
O6	p	0.245	HO4a	s	0.142
			O-P4b	s	0.110
				p	0.202

b/ PM3

HOMO			LUMO		
Atom	Orbital	Eigenvector	Atom	Orbital	Eigenvector
C2	p	0.144	C1	p	0.172
H2	s	0.123	O4	s	0.144
O2	p	0.326		p	0.308
C3	p	0.233	P4	s	0.568
O3	p	0.313		p	0.191
C4	p	0.326	O=P4	p	0.375
H4	s	0.117	O-P4a	s	0.134
O4	p	0.370		p	0.354
C5	p	0.320	HO4a	s	0.124
O5	p	0.439	O-P4b	s	0.136
O=P5	p	0.125		p	0.273
C6	p	-0.237	HO4b	s	0.177

c/ MNDO

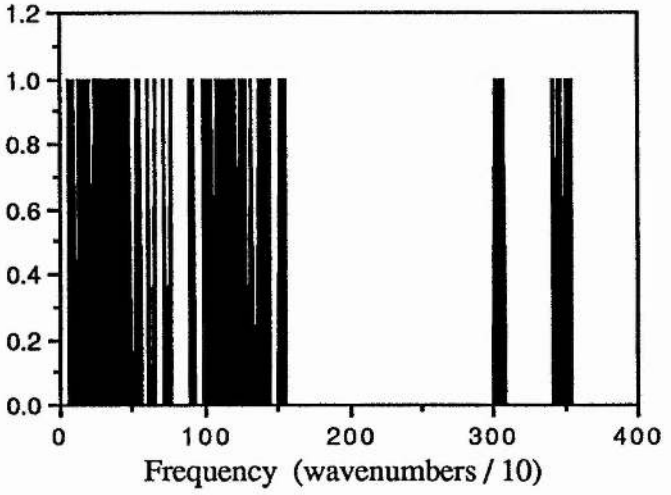
	HOMO			LUMO	
Atom	Orbital	Eigenvector	Atom	Orbital	Eigenvector
C1	p	0.267	C1	p	0.182
O1	p	0.244	O1	s	0.165
C2	p	0.233		p	0.329
O2	p	0.212	P1	s	0.540
C3	p	0.139		p	0.240
O3	p	0.367	O=P1	p	0.333
C4	p	0.260	O-P1a	s	0.119
O4	p	0.291		p	0.334
C5	p	-.265	HO1a	s	0.125
O5	p	0.281	O-P1b	s	0.124
O=P5	p	0.137		p	0.337
C6	p	0.253	HO1b	s	0.133
O6	p	0.364	P5	s	0.102

The HOMO's are essentially the same with all three parameter sets in that they are composed of P atomic orbitals centred on the carbon atoms and the oxygens bonded directly to them, the only difference is that the MNDO orbital is equally distributed across the whole ring while the AM1 and PM3 orbitals do not extend appreciably over C_1. On account of its position, the HOMO is largely inaccessible to external attack, but the large value of the ionisation energy for these molecules suggest that they are unlikely to lose an electron.

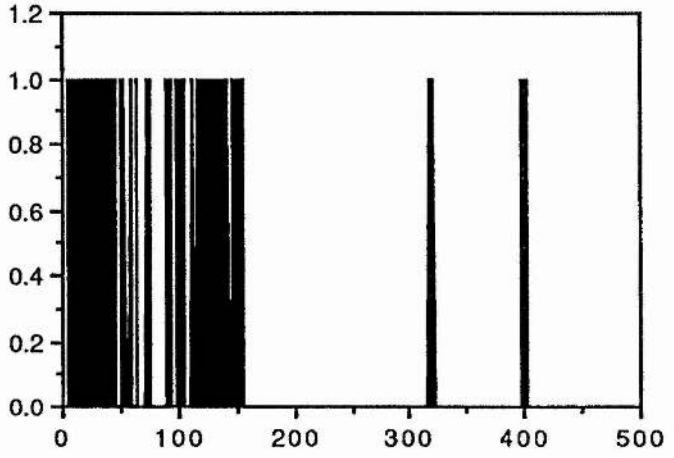
By contrast, the location of the LUMO is different for MNDO than for AM1 and PM3 as it is centred on a different phosphate group. However the three lowest unoccupied orbitals for all three methods form a nearly degenerate set centred on each

of the three phosphate groups and are otherwise identical. MNDO merely orders the energies of these orbitals differently to AM1 and PM3 without noticeably altering their character. The location of the LUMO suggests that the binding of this molecule to its receptor is driven by a electrophilic attack by the phosphate groups, though this would be tempered by the tendency of the molecule to lose protons from these groups.

Calculation of the second derivatives for the internal coordinates gives the frequencies for internal vibrations and hence a prediction of the I.R. spectra. The intensities of the bands cannot be predicted only their positions and these are reproduced below for each of the parameter sets.



a/ AM1



b/ MNDO

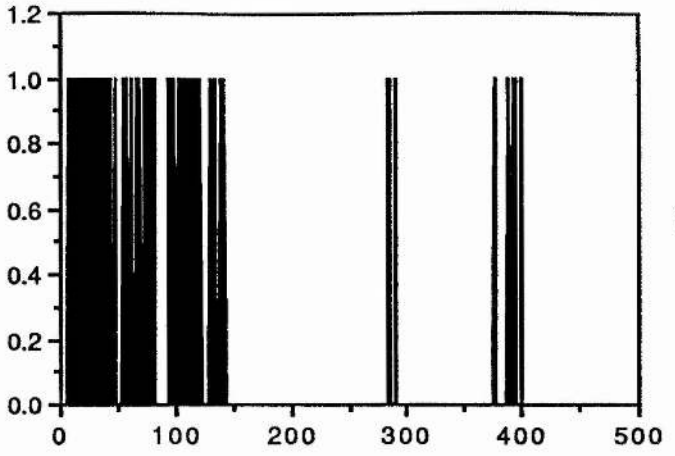


Fig 5.14 I.R. Spectra for *Myo* -Inositol (1,4,5) Trisphosphate.

The high frequency vibrations are bond stretching while bond flexing takes place at lower frequencies and the frequencies of the main groups of stretching vibrations are listed below along with common experimental values taken from measurements over a wide range of molecules [126]. The calculated spectra are compared with typical experimental values as there spectra of the inositol phosphates have not been measured, the only relevant spectrum published being that of inositol [126]. Frequencies are given in cm^{-1}.

Bond	AM1	MNDO	PM3	Experimental
O-H	3425-3542	3985-4022	3765-3951	3335-3635
C-H	3011-3061	3173-3205	2833-2907	2855-2940
C-O	1383-1447	1471-1558	1297-1345	1000-1250
P=O	1392-1410	1207-1332	1082-1084	1250
P-O	922-1055	906-1056	737-829	1000-1110
C-C	530-1314	513-1184	478-1052	910-1250

The wide range for the C-C stretching comes from a low frequency mode involving C_3 and C_4, but most of the C-C stretching frequencies lie at the top of the range, which corresponds well for all three parameter sets with the experimental range of values. None of the parameter sets give frequencies that overlap with the experimental values for all types of vibrations, but the AM1 results are less likely to diverge sharply from the expected values though the C-O and P=O frequencies are overestimated by about 10%. MNDO generally overestimates the frequencies, 10% for O-H and 15% for C-O while the errors for PM3 are less consistent ranging from a small overestimate for O-H and C-O to a nearly 20% underestimate for P=O. However the general appearance of the spectra is consistent with what would be expected for this type of molecule and the divergence between the different parameter sets is less than for the heat of formation.

5.10 Discussion.

From an understanding of the biology of the inositol phosphates and semi-empirical Hartree-Fock theory, the conformational analysis and properties described above would be expected to show several features namely some degree of parameter dependence in the position of minima, a favouring of the "I" ring conformation over the "E" conformation that decreases with adding phosphate groups and some differentiation between (1,4,5) trisphosphate and the other derivatives.

The first of these, parameter dependence of minima position, comes from the different treatment of the core-core repulsion of the tree methods and from the different results they have for hydrogen bonding. MNDO is reported to over-estimate the core-core repulsion [125] and does not reproduce hydrogen-bonds at all, so would be expected to favour conformations with the largest separation of phosphate and, to a lesser extent, hydroxyl groups. Thus MNDO has the largest energy advantage for the "I" conformation over the "E" conformation and is the only one where this is consistently the lowest energy conformation. The effects of the different core-core

repulsion terms and reproduction of hydrogen bonding is best seen in the relative energies of the inositol geometries, especially comparing the relative stabilities of the "2A" and "2E" conformations as the former possesses a strong hydrogen bond 2'-5' while the latter maximises the distance between the ring substituents. With AM1 and MNDO the lowest energy boat conformation is found at "2E" while with PM3 the lowest is at "2A". The lack of hydrogen bonding in MNDO means that the energy of the minimum at "2A" is actually higher than for the neighbouring "1E / 2A" and "2A / 3E" minima.

A further difference between the parameter sets is that there is much less difference in the energy of the minima with PM3 especially for the higher phosphates. There is no clear reason for this either from the calculations presented here or from the Stewart papers that deal with the initial PM3 parametrisation [63, 64]. One dramatic consequence of this is the reversal of the relative energies of the the two chair conformations with PM3 for the (1,4) bisphosphate and the (1,4,5) trisphosphate.

The difference in the phosphate dihedrals in the minima produced using the different parameter sets also reflect the differences in the core-core interactions. The crowded nature of the trisphosphates and tetrakisphosphate makes steric considerations paramount in dictating the values of the dihedrals in the minima, so differences in the magnitudes (and rate of decay) of the core-core repulsions tend to have a lesser effect than in other situations with the result that several dihedrals are similar with the different parameter sets: e.g.P_5-O_5-C_5-C_6 in (1,3,4,5) tetrakisphosphate and both P_4-O_4-C_4-C_5 and P_5-O_5-C_5-C_6 in (1,4,5) trisphosphate. For all of the molecules studied, the 1' phosphate has considerable steric freedom and suffers from the largest parametrisation dependent variation in the minima position. The 1' minima dihedrals can divided into two groups; large-angle minima (dihedral >200°) and small-angle minima (dihedral <200°). The difference between the two groups is clear in any given graph. When the 3' phosphate is present, the small-angle minima has the lowest values in the PM3 calculations, reflecting the increased hydrogen-bonding forces from the 3' to 1' phosphates via the 2' hydroxyl, in its absence the angle is increased balancing hydrogen bonding with 2' against a similar bond to 6'. By way of contrast, MNDO

leads to the opposite effect. This parameter set does not reproduce hydrogen bonds and overestimates the core-core repulsion. Thus the presence of a 3' phosphate has a repulsive effect on the 1' group leading to an increase in the 1' dihedral. In the case of the large-angle minima, the nearest ring substituent is the 6' hydroxyl and the other phosphate groups are further away than for the small-angle minima and so their presence of absence has a much smaller effect. There is no noticeable difference in the mean value of the 1' dihedral in the minima of different parameter sets but there is some variation in the spread of minima. It is perhaps dangerous to read too much into this difference as differences in the range on minima can be produced by many effects that have nothing to do with the width of the valley containing them.

The relative energies of the two chair conformations is an interesting question. Crystal structures for inositol itself and for lower phosphates (2' monophosphate) show the molecule in the 5 equatorial chair ("I" conformation), while the hexakisphosphate (also called phytate) crystal structure is in the "E" conformation. The reason for this change lies in the problems of accommodating an increasing number of large substituents around the ring. Eventually, the "I" conformation suffers from too many close contacts especially between the phosphate hydrogens and oxygens (or from electrostatic repulsion if ionised) giving it a higher energy then the "E" conformation. The question is how many phosphate groups are required to change the balance between the conformations and whether the position of the groups is important. The calculations in this chapter are concerned primarily with a conformational analysis of the (1,4,5) trisphosphate and its important metabolites rather than answering this question *per se*.

The energy difference between the two chair conformations for the different inositol phosphates with the three parameter sets is shown below, a positive energy difference (termed energy advantage) means that the "I" form is the lower energy and the molecule numbers are as follows:

1: Inositol

2: 1' Monophosphate

3: 1',4' Bisphosphate

4: 1',3',4' Trisphosphate

5: 1',4',5' Trisphosphate

6: 1',3',4',5' Tetrakisphosphate.

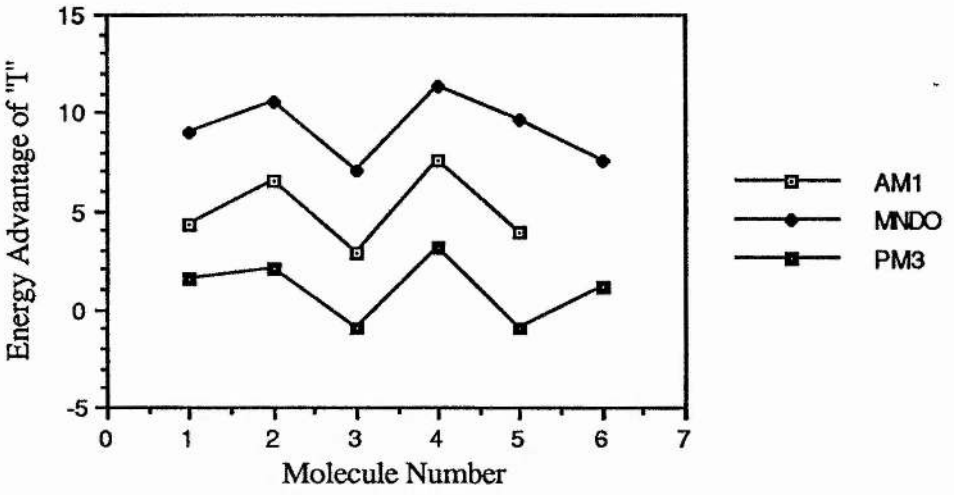


Fig 5.15 Stability of the "I" Ring Conformation over the"E" Ring Conformation for *Myo* -Inositol Phosphates.

With the exception of the tetrakisphosphate, the lines are virtually parallel but show no clear trend relating conformation to phosphorylation state. The parallel nature of the graph reflects the general narrowing of the difference in energy between conformations found with PM3, though it would not in itself be sufficient to change the sign of that difference. An examination of the phosphate dihedrals lead one to expect that the "I" conformation should be stable for all the derivatives studied. The steric crowding at 3', 4' and 5', which is most likely to lead to the conformational change can be partially relieved by allowing the 5' dihedral to increase (towards 300° from 240°) while that of the 3' group decreases from 240°. No such relaxation is possible for the

hexakisphosphate as positions 2 and 6 are also occupied and would be expected to constrain the 3' and 5' dihedrals to 240°, increasing their steric interaction with the 4' phosphate. This relaxation is observed in the higher phosphates studied., while there is no such consistent variation for the "E" conformation. Thus it is more likely that the "I" conformation would be the more stable of the two chair conformations for all of the molecules studied, casting doubts on the veracity of the PM3 energies for this type of molecule. No minima with an"E" ring conformation could be found using AM1 on the tetrakisphosphate.

The final piece of information obtained from this study, namely the differences between the (1,4,5) trisphosphate and the other inositol derivatives (especially the (1,3,4) trisphosphate and the (1,3,4,5) tetrakisphosphate) that are responsible for its calcium releasing activity is the primary reason for this study. The two trisphosphates can potentially have the same relative positioning of the three phosphate groups so can not be distinguished on this basis unless there is a restriction on ring conformation and the rotation of the phosphate groups, or alternatively there may be a requirement for a hydroxyl group also on the binding site which could distinguish between them. Against this, the tetrakisphosphate possesses the same functionalities as the active (1,4,5) trisphosphate except that the 3' hydroxyl is replaced by a phosphate, which leads to the conclusion that some form of geometrical constraint is responsible for the difference in activity. The main difference between the two trisphosphates is the relative stability of the two chair conformations in the PM3 calculations and the the comparative energies of the "2E / 3A" and "2A" boat conformations. The former is the most stable of the boat structures using all three parameter sets on the (1,4,5) trisphosphate and of lower energy than one of the two chair conformations in each case. Both AM1 and MNDO give a low energy minima in the adjacent regions ("2E" and "3A" respectively) for (1,3,4) trisphosphate that is also lower in energy than one of the chair conformations but PM3 does not and for both AM1 and PM3, the energy of the "2A" conformation is more than 4 kcal mol^{-1} lower. If the arrangement of the phosphate groups around the ring is taken into account, the "2E / 3A" conformation of (1,4,5) trisphosphate corresponds to the "2A / 3E" conformation for (1,3,4)

trisphosphate. Only the AM1 calculations produced a minimum in this region and that was of high energy, though the most stable boat minima ("2A") mentioned above is of course in the adjacent region. However, the transition to this region leads to an increase P_1-P_4 separation comparable with a transition from from "2E / 3A" to "3A" for (1,4,5) trisphosphate which is accompanied by a 4-5 kcal mol^{-1} increase in energy. The "2A" conformation is also the most stable boat conformation for the (1,3,4,5) tetrakisphosphate but with AM1 the minima is moved into the "2A / 3E" region though the presence of the extra phosphate may count against this conformation being able to mimic the "2E / 3A" conformation of (1,4,5) trisphosphate. The AM1 calculation also gives a low energy minima in the "2E" region within 2 kcal mol^{-1} of the lowest energy boat conformation but none in the "2E / 3A" region itself. The discussion on the effects on P_1-P_4 distance caused by moving the ring conformation from one region to another also apply to this derivative and so the energy cost for taking up the exact conformation may be too large for it to occur to any great extent.

The relative positions of the phosphate groups can also be affected by differences in the phosphate dihedrals, though there is greater difference between different parameter sets for the same molecule than between different molecules. This stems partly from the considerable restriction in locations of the minima with respect to the phosphate dihedrals for the higher phosphates so much of the difference between the molecules will lie in the ring conformation rather than the phosphate positions.

Without more information about the binding site itself it is impossible to say whether this suggestion that the (1,4,5) trisphosphate binds in a "2E / 3A" ring conformation is correct, but it is a conclusion that is supported by the calculations presented here. In addition it is suggested that all three phosphate groups are required so accounting for the inactivity of the lower phosphates. The possibility of binding involving only two phosphate groups can be discounted because of the flexibility of the molecules. There are several conformations for the (1,4,5) trisphosphate which possess the same phosphate-phosphate distances as another (when considering a phosphate pair not all three) and can be reproduced by one of the bisphosphates (1,4 or 3,4) or (1,3,4) trisphosphate.

A further problem with these calculations is that they lie on the limits of accuracy of the semi-empirical methods. Problems with the core-core repulsions and hydrogen bonding of MNDO could be critical in these highly polar and crowded molecules and they are also well outside the range of molecules used in the parametrisation of all three methods. This is a problem particularly for AM1 as the phosphorus parameters were determined separately from those for hydrogen, carbon, oxygen and nitrogen These difficulties show up in the large differences in the heats of formation (particularly for the higher phosphates) and in the radically different phosphorus charges obtained from the density matrix. The work in chapter 7 was carried out to examine these differences and to produce some measure of the accuracy of the methods for this type of molecule.

Chapter 6

Model Lipid Bilayers.

6.0 Introduction.

The tumour promoting phorbol esters are amphipathic molecules with an extremely low aqueous solubility [6], thus their place of action is likely to be at the surface of the cell membrane. This is supported by evidence of translocation of PKC on activation so that it sits bound to the membrane surface in its active form [146] and is thus accessible to phorbol esters or its endogenous activator DAG. With no evidence that the enzyme is actually inserted into the membrane, it is therefore important that the parts of the phorbol esters involved in binding be exposed at the surface of the membrane. The idea of the enzyme sitting on the surface rather than embedding in the membrane is supported by affinity labelling of phorbol ester derivatives [147-149], that have only succeeded in binding to lipid and not to the enzyme. In all such cases, the labelling groups have been placed on the phorbol ester chain not on the head group.

For the proposed binding site to be correct therefore, the A and B rings of the phorbol esters must be exposed at the surface of the membrane, so with the arrival of the fast minisupercomputers in St. Andrews (SCS-40 and FPS-500), it was decided to study these tumour promoters in a phospholipid bilayer. In addition to the need for exposure at the membrane surface, there is no guarantee that the conformations obtained from gas-phase calculation on isolated molecules are the same as those favoured by the molecule in its natural environment. The cell membrane itself is an extremely complex and variable environment consisting of rather more than just a lipid bilayer. The amount of protein embedded within the membrane varies considerably [150] and the lipid-mosaic model [151] frequently used to describe the appearance of the membrane may only actually be true for a small number of cases where the protein content is quite low, elsewhere the protein content being too high to allow more than a layer or two or phospholipids around each enzyme [149]. In addition many metal ions and ion transport systems are associated with the membrane making its detailed electrostatic properties rather complex. However, to a simple approximation, the phospholipid bilayer gives an adequate representation of the immediate local environment for an amphipathic molecule not involved in any specific lipid or protein

contacts. This had lead several groups to study the dynamic behaviour of various lipid systems, including Prof. Berendsen in Gröningen [152-154] (stability and order of chain conformations) as well as the work presented here from St. Andrews (conformations and accessibility of specific molecules in bilayers). In addition, work on the minimum energy conformations of amphipathic molecules in a phospholipid environment has been carried out by Brasseur *et al* [155].

6.1 Computational Strategy.

Regardless of the particular system under the study, the computational approach is essentially identical and can be divided into two sections each with their own problems. The initial positions and conformations for the molecules in the bilayers must be set and then the minimisation / dynamics calculations performed.

The initial conformations of the molecules will normally be local, low energy minima from calculations on isolated molecules though alternatively crystal structures can be used. Provided an adequate time for equilibration is allowed in the dynamics calculation, there should be no difficulties with any of the source of conformations though setting the initial positions of the molecules within the bilayer is rather more troublesome. The molecules in the bilayer must be sufficiently close packed that water is not placed within the hydrophobic region during the solvation step (EDIT module of AMBER) and also that the water penetration during the dynamics is not to such an extent the the integrity of the bilayer is lost. For homogeneous bilayers, a close packed structure can be obtained from repeating the crystal structure unit cell, though this method is not entirely reliable as the crystal structure repeats in all three dimension rather than the two of a bilayer, thus head group-head group interactions serve to distort the geometry from that found in the bilayer. An alternative is to place energy minimised structures in a regular array, which should eliminate unnatural distortions due to the close packing of the crystal environment. If each face of the bilayer is not homogeneous, neither of these two methods will produce sufficiently close packing to

eliminate water from the hydrophobic region during the solvation stage of AMBER, instead the program HYDRO (see chapter 3) is used to pack the bilayer components into a pseudo-random close-packed structure. The random nature of the structure is closer to that found in natural bilayers than is the regular array found in crystal structures so it is often preferable to use HYDRO to pack even homogeneous systems.

The packed bilayer structure produced using HYDRO is fed into AMBER in the EDIT module as a PDB file (see chapter 3) at which stage it is also solvated. Periodic boundary conditions are imposed for the minimisation and dynamics to avoid edge effects on the interior of the bilayer. Though periodicity is only required for the plane of the bilayer itself, the nature of the AMBER code means that boundary conditions must be applied simultaneously for all three dimensions so it is essential that a sufficient layer of solvent is allowed above and below the bilayer to prevent interaction between neighbouring bilayers. This solvent layer must therefore be at least twice the cutoff used for non-bonded interactions, though in practice a larger separation was used. One consequence of the convergence criteria used in HYDRO is that the packing of the bilayer is more likely to halt when the molecules are in Van der Waals contact than when the surface area threshold is met, thus the surface area of the bilayer would be expected to be a little lower than that for lipids at 300 K while the water coordinates are obtained from Monte Carlo results at room temperature (the default included within the AMBER package). No solvent molecules are allowed within a specified distance of the solute, in the case of bilayers this has been set rather larger than it would normally be to prevent spaces within the hydrophobic portion of the bilayer becoming solvated. Thus there is a small gap between the bilayer and solvent which is removed by minimising the energy of the water molecules while fixing the position of the bilayer. This reduces the pressure of the solvent by an unpredictable degree, leaving it lower than that of the bilayer, which is in turn somewhat above 1 atmosphere from the convergence criteria of HYDRO. Thus the minimisation and dynamics are performed at constant pressure (with anisotropic pressure monitoring) rather than constant volume to avoid unnatural distortions of the structure and to allow these differences in pressure to relax. Constant pressure calculations are normally performed in conjunction with constant temperature

and number of particles, the so called Isobaric / Isothermal Ensemble, so these conditions are also used for all the calculations described in this chapter (pressure = 1 atm., temp = 300 K).

The EDIT module of AMBER, in the original form, rotates the solute molecule so as to minimise the number of solvent molecules required within the specified cutoff. A similar rotation of the bilayer would lead to solvent molecules penetrating the hydrophobic region near to the edges of the periodic box and would destroy the nature of the bilayer, so it was necessary to override this. In doing so however, the usual coordinate resetting to place the centre of gravity at the origin is not performed, so the coordinates for the bilayer from HYDRO must be centred at the origin to allow correct solvation. Also the energy of the bilayer is minimised prior to the dynamics calculations to avoid gross distortions of the structure or large velocities appearing in the initial steps of the dynamics calculations.

To test the validity of the packing procedure and of the minimisation / dynamics calculations, several test calculations were performed on homogeneous lipid bilayers before moving on to systems involving phorbol esters in mixed lipid bilayers, Also, as the full solvated bilayer involves a large number of atoms, model system with smaller bilayer components and a smaller solvation layer were considered. The number of atoms required for the calculations were further reduced by using the united-atom approach for minimisation and dynamics.

6.2 Model Systems.

There are several reports in the literature of the work in Gröningen of Prof. Berendsen and coworkers concerning dynamics of decanoic acid bilayers [152, 153], so the first calculations performed in St. Andrews were based on this system. Decanoic acid possesses a single lipid chain and a very small headgroup with a surface area at the bilayer surface of 25 $Å^2$ [156], compared with 60-70 $Å^2$ for full sized phospholipids [149, 156] and though it possesses distinct polar and non-polar regions,

it does not naturally form bilayers, preferring instead to form strong headgroup-headgroup hydrogen bonds. However, as it only possesses 12 heavy atoms and 1 hydrogen not bound to carbon, it is a useful model for lipid systems.

The packing using HYDRO produced an extremely small surface area, calculated from the average nearest neighbours distance, using a hexagonal packing approximation[1], of only 19.4 $Å^2$ per molecule, though the area is difficult to estimate accurately for a small randomly packed structure such as this. 50 molecules of decanoic acid with an initial hydrophobic surface area of 9800 $Å^2$, were packed to yield a final hydrophobic surface area of 4300 $Å^2$. From a rectangular box, dimensions 17.25 Å x 12.92 Å, containing twelve molecules was chosen to provide the initial position for the dynamics calculation. A solvation layer 12 Å deep was allowed on each side of the bilayer and a non-bonded cutoff of 10 Å used. As it is the hydrophobic region and the lipid-solvent interface that is of primary concern, a dielectric constant (ε) of 4 is used for all of the calculations described in this chapter, thus allowing only for minimal polarisation of the molecules in the system.

At 300 K, the bilayer is extremely unstable with the fatty acids noticeably departing from the bilayer after 8 ps of the equilibration stage and were evenly distributed throughout the periodic box after only 24 ps. As the molecular weight of the fatty acids is considerably less than that of the phospholipids, they will have a much greater velocity at a given temperature allowing them to overcome the weak non-bonding forces holding them together while the lower value for the reduced mass for the lipid chains should lead to an increase in disorder and fluidity of the chains in the hydrophobic phase were the lipids to remain in the bilayer. The fatty acids also possess a much smaller headgroup than the phospholipids making it rather easier for them to slip between neighbouring water molecules and so escape from the bilayer. At the lower temperature of 250 K, the fatty acids remained in a bilayer for upwards of 40 ps, retaining disorder of the headgroups, but the chains in the lipid bilayer adopted the "chevron" conformation characteristic of lipid crystal structures. It was felt therefore that the fatty acids were not a viable system in which to model the effect of a lipid environment on the phorbol esters as the fine-tuning required to keep the temperature

such that the bilayers remained stable without forming a "crystal-like" phase, would be too complex.

The results for the stability of a fatty acid based model system, presented here, differ from those reported by Berendsen *et al* [152, 153]. In their work on this type of system however, Berendsen *et al* have employed artificial constraints to maintain the fatty acids in a bilayer. Initially a half-harmonic potential was used to prevent headgroups leaving the bilayer for the aqueous phase [152] while in the latter example the potential is replaced by an extremely large electrostatic repulsion, produced as a consequence of an insufficient solvation layer (only 2 Å deep) between neighbouring periodic bilayers [153]. Such constraints can only distort the effects of the environment on the lipids and any inserted amphipathic molecules in a largely unpredictable manner and thus unsuitable for the systems considered here.

[1] The bilayer surface area per molecule, Å, is calculated from the average nearest neighbour distance by:

$$A = \frac{r^2\sqrt{3}}{2}$$

which approximates the positioning of the molecules to a regular hexagonal array. This method gives a smaller value than a square array approximation but a greater value than assuming the molecules have a circular cross-section, which ignores the intermolecular spaces.

6.3 Full-Sized Lipid Systems.

In view of the failure of the model system discussed above, it was decided not to continue with developing small molecule model systems and instead to examine the stability of bilayers made up from full-sized phospholipids rather than either short chain phospholipids or full-chain, small headgroup molecules as both parts of the molecule appeared to be required for different reasons. The size of the headgroup appeared critical to retain bilayer stability as molecules with large headgroup surface areas would need to break a larger number of solvent-solvent hydrogen bonds and thus be less likely to escape from the bilayer, while the length of the acyl chains of phorbol esters appeared important to their activity and thus potentially to their orientation, so it was essential that the bilayer was wide enough to accommodate the entire phorbol acyl chain. The polar functional groups on the phorbol moiety are spread rather further apart than the equivalent groups of the phospholipids, especially along the Z-axis (perpendicular to the plane of the bilayer), so the effect of shortening the chain would be expected to be greater than for equivalent modifications to the phospholipids. By using the united atom approach, the number of atoms required for each phospholipid could be reduced from 131 (PE), 133 (PS), 140 (PC) to 54 (PE, PC) and 57 (PS).

Fifty PC molecules were packed into a monolayer with HYDRO, the solvent accessible hydrophobic surface area decreasing from 68381 $Å^2$ to 32118 $Å^2$ in 86 cycles taking 16152 seconds CPU time on the FPS-500. From this, a 40 x 40 Å box containing 21 molecules was taken and copied to produce a bilayer containing 42 phospholipids. A solvation layer 12.0 Å deep, containing a total of 1035 water molecules, was placed above and below the bilayer with no water molecules within 5.0 Å of the nearest bilayer atom to ensure no penetration of solvent into the bilayer. After minimisation of the solvent (3203 cycles in 14289 seconds) the potential energy of the system was -3.928×10^3 kcal mol^{-1} and the gap between the bilayer and solvent had disappeared. The potential energy was reduced further to -5.463×10^3 kcal mol^{-1} after minimisations of the whole structure (10108 cycles and 100310 seconds). The calculation times could have been reduced by using a smaller solvation layer as the non-

bonded cutoff was limited to 9 Å by the available memory. After the failure of the decanoic acid model, the dynamics calculations involving the complete lipids were approached rather cautiously. The temperature for the calculation was allowed to increase first to 250 K over 4 ps (2000 steps) and then to 300 K over a further 1 ps (500 steps). No further equilibration was performed as the calculations were primarily concerned with the stability of the bilayer and calculations on a system of this size are extremely time consuming. The coordinates were sampled every 2 ps over the following 20 ps, from which data concerning the dimensions of the bilayer and the degree of disorder were obtained and are presented below. The dynamics calculation was carried out at a constant pressure of 1 atmosphere and a temperature of 300 K, allowing deviations of up to 10 K, though the RMS deviation turned out to be only 3.5 K.

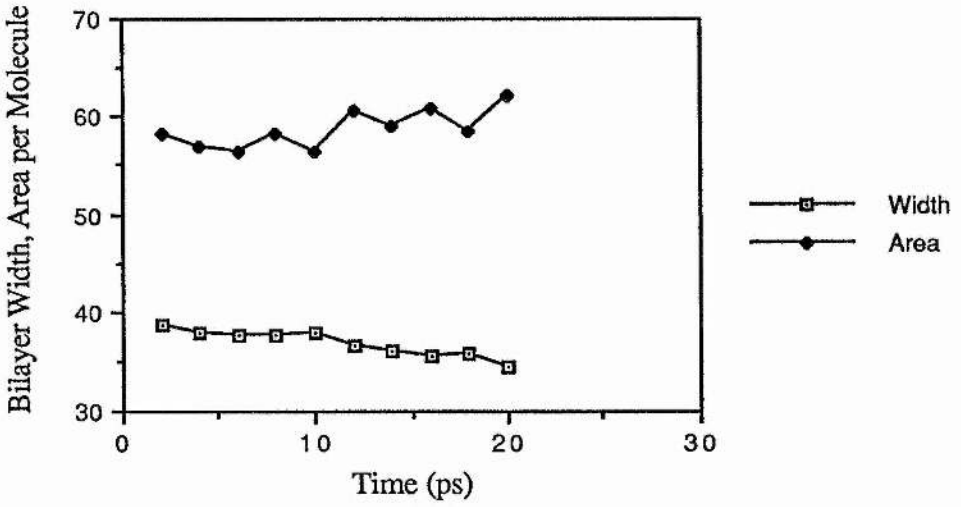


Fig 6.1 Width (Å) and Area per Molecule (Å^2) vs. Time for a PC Bilayer.

Solvent penetration into the bilayer was at a minimum with never more than 8 water molecules within the hydrophobic region. On occasions when a molecule did penetrate relatively deeply into the bilayer, the comparatively large long-range electrostatic force, brought about by the low value of the dielectric, was able to draw it

out again towards the headgroups and solvent layer. The dimension of the bilayer during the dynamics simulation compares well with experimental values (40 Å wide [149] and an area of 60-70 $Å^2$ per lipid molecule [149, 156]) though the latter are rather imprecise due to the heterogeneous nature of physiological bilayers. The surface area per molecule is at the very bottom of the experimental range and is reflected by the distribution of angles for the P -> N vector. Experimentally, this is believed to lie almost parallel with the membrane surface [157] and as it is the major factor in determining the bilayer surface area, any deviations in its angle will lead to a reduction in surface area. The lipid conformation used in the packing has a P -> N vector very close to the horizontal and produces a surface area of 73.45 $Å^2$ per molecule (the area from the packing is from the hexagonal approximation, that from dynamics from the dimensions of the periodic box). Several factors contribute to the change in the position of the P -> N vector and reflect deficiencies in the nature of the model. A low value has been chosen for the dielectric to give the most accurate representation of the interior of the bilayer and serves to keep the solvent away from the interior but also has the side effect of overestimating the electrostatic attraction of the solvent for both phosphate and quaternary nitrogen groups and also overestimating the electrostatic repulsion between neighbouring quaternary nitrogens. The former serves to pull the nitrogens out into solution (large positive P -> N angle), while the latter pushes some deeper into the bilayer (large negative P -> N angle). After the first couple of picoseconds, there is no change in the distribution of the angle of the P -> N vector (see fig 6.2) as this part of the bilayer stabilises very rapidly and the surface area only shows a low-frequency "breathing" motion. The reason behind the reduction in bilayer width from experimental values is rather harder to determine. The differences between the composition of cell membranes and this simple bilayer model would have a considerable effect on bilayer width. A greater degree of unsaturation tends to make the lipids less flexible and more "bulky" restricting the extent of interleaving of the two sides of the bilayer increasing its width, which would be augmented by the protein content of the cell membrane. Thus it is difficult to tell whether the reduction in bilayer width is real or apparent and whether there ought to be some modification to the non-

bonding forces within the hydrophobic region or to the compressibility of the system.

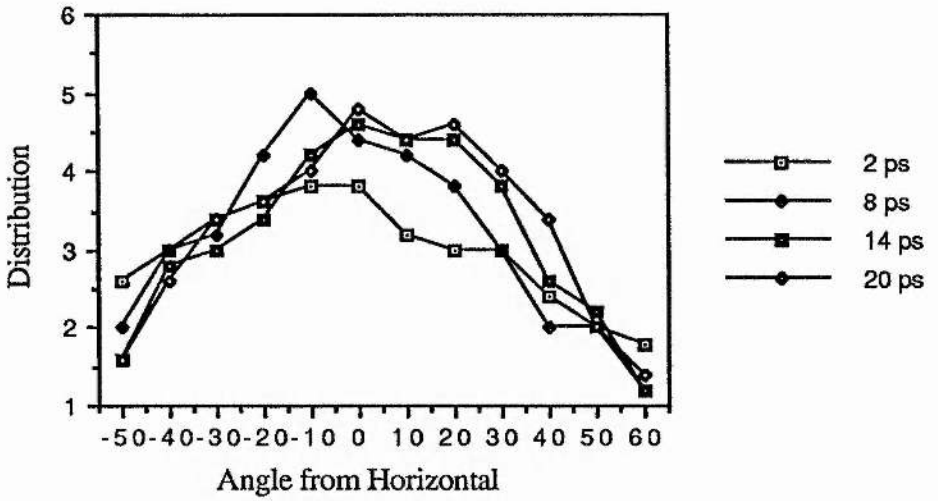


Fig 6.2 Distribution (in Blocks of 10) of Angle of the P -> N Vector to the Horizontal.

Fig 6.3. shows the dynamics simulation at the end of the run (20 ps) giving some idea of the degree of disorder in the lipid chains and in the headgroup orientation.

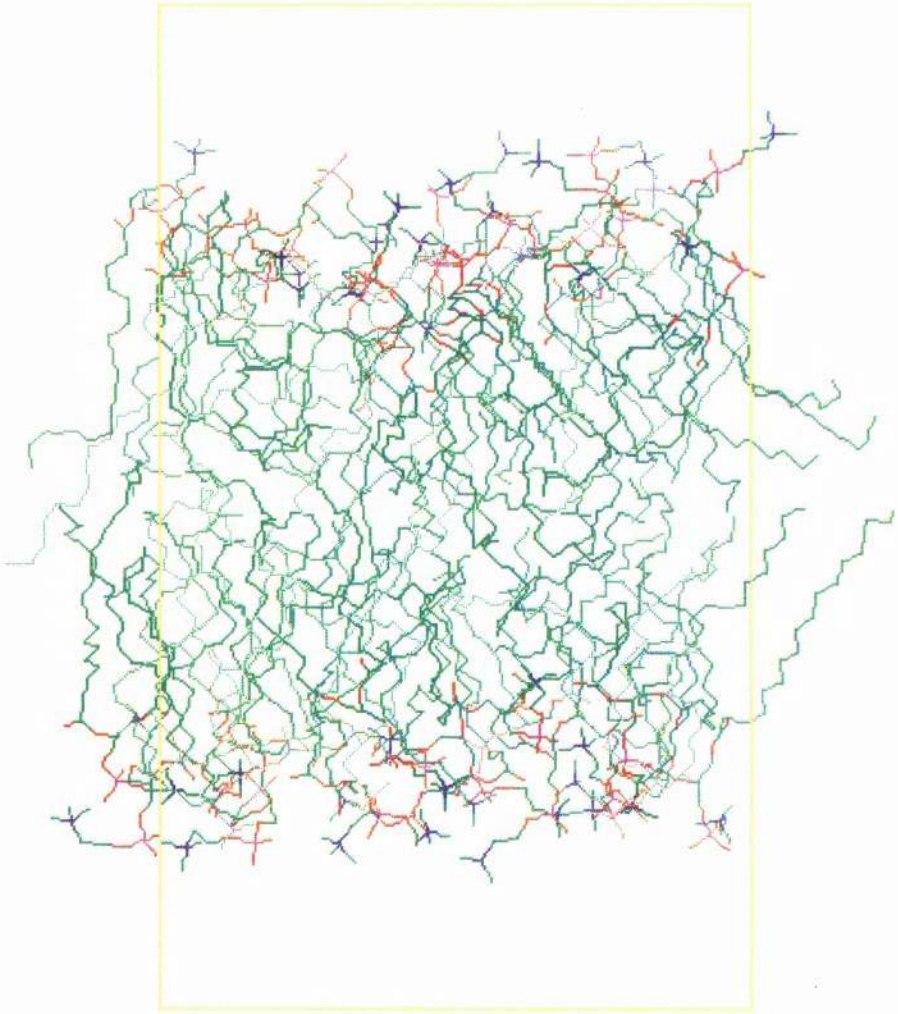


Fig 6.3. Phosphatidyl Choline Bilayer.

The periodic box is shown in yellow and the solvent molecules ignored. The more prominent molecules are nearest to the viewer, with the phosphate groups shown in purple and nitrogen in blue.

6.4 Phorbol Esters in a Lipid Bilayer.

The lipids packed around a central phorbol ester were chosen randomly with probabilities to reflect approximately their natural occurrence, 60% PC, 30% PE and 10% PS. No Z-offset was used for the phorbol ester (see chapter 3). 49 lipids were thus packed around the phorbol ester and the 19 (12 x PC and 7 x PE) closest were chosen for the minimisation and dynamics. A solvation layer 10 Å deep was added either side, some 7.0 Å away from the nearest bilayer atom so as not to solvate the interior of the bilayer. A larger distance cutoff was required than with the homogeneous bilayer as the phorbol ester is smaller than the other lipids and disrupt the efficient packing of the bilayer. A total of 963 water molecules were used in the solvation and after the same two stage minimisation as before, the potential energy was -6.839 $x10^3$ kcal mol^{-1}. The dynamics calculation was split into 20 ps equilibration and 40 ps data collection, sampling every 2 ps, under exactly the same conditions as were used for the PC bilayer.

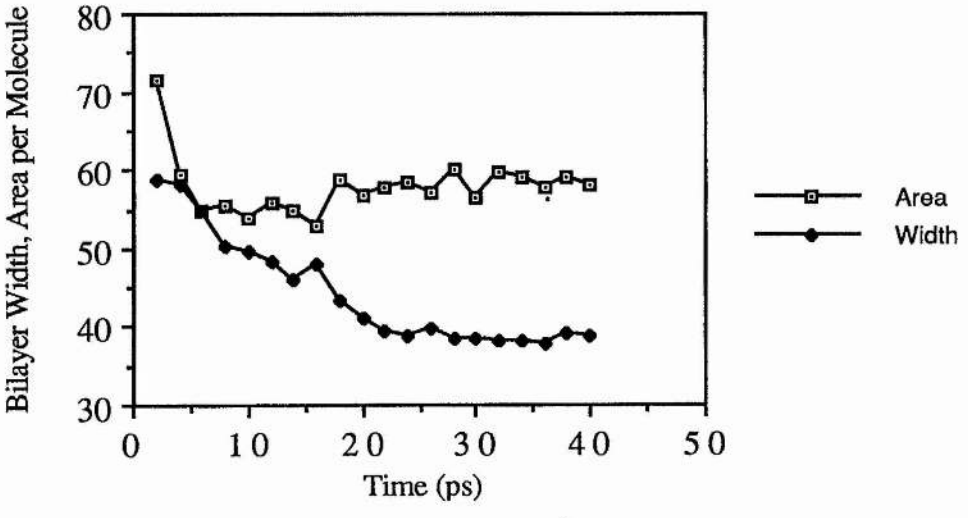


Fig 6.4 Width (Å) and Area per Molecule ($Å^2$) of Bilayer with Phorbol Ester.

Considerable penetration of solvent in to the bilayer was found especially close to the inserted phorbol ester and though the ester would be expected to have a smaller cross-sectional surface area in the plane of bilayer than the lipids, there is a minimal effect on the overall surface area of the bilayer. The surface areas of the two types of molecule are such that packing is still in a 1:1 form so the smaller surface area of the phorbol ester leaves larger interstitial spaces which are then subject to greater solvent penetration. Unlike the lipid headgroup, the phorbol group is not able to rotate to substantially alter its cross-sectional area and would not be expected to have a similarly large response to pressure so at higher pressures, the reduced surface area of the lipids would lead to a closing of the interstitial spaces. The width of the bilayer is greater than was the case for the bilayer composed of PC alone, stabilising after some 30 ps, and is accompanied by a visible reduction in the chain disorder. The result is a bilayer structure that closely resembles that expected from experimental measurements of cell membranes and lipid bilayers.

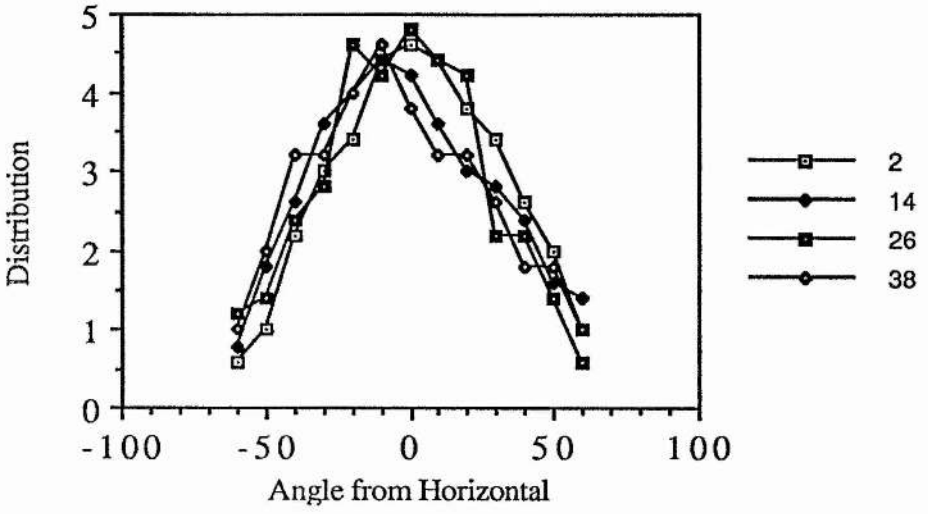


Fig 6.5 Distribution of the angle of the P -> N vector to the Horizontal.

The distribution of the vectors angles again varies little with time except to favour a slight depression of the vector so that the nitrogens are contained within the headgroup region rather than being exposed to solution. There is no reason why there should be a difference from the results for the PC bilayer and this small difference may well come only from the random element of the calculation.

The positions of the molecules after 20 ps of the data run are shown on the previous page (fig 6.6), with the phorbol esters at the centre of each face of the bilayer.

The phorbol ester retains its orientation, with the A and B rings uppermost, but over the course of the dynamics sinks a small distance into the bilayer. This is not sufficient to bury the site groups (see chapter 4) completely, but would require the enzyme to be in close contact with the neighbouring lipids for the ester to enter its binding site. This is consistent with the observations of Ganong *et al* [29] on the stoicheometry of the phorbol ester-lipid-PKC interaction and with the requirement for specific phospholipid, PS, in order to activate PKC [158].

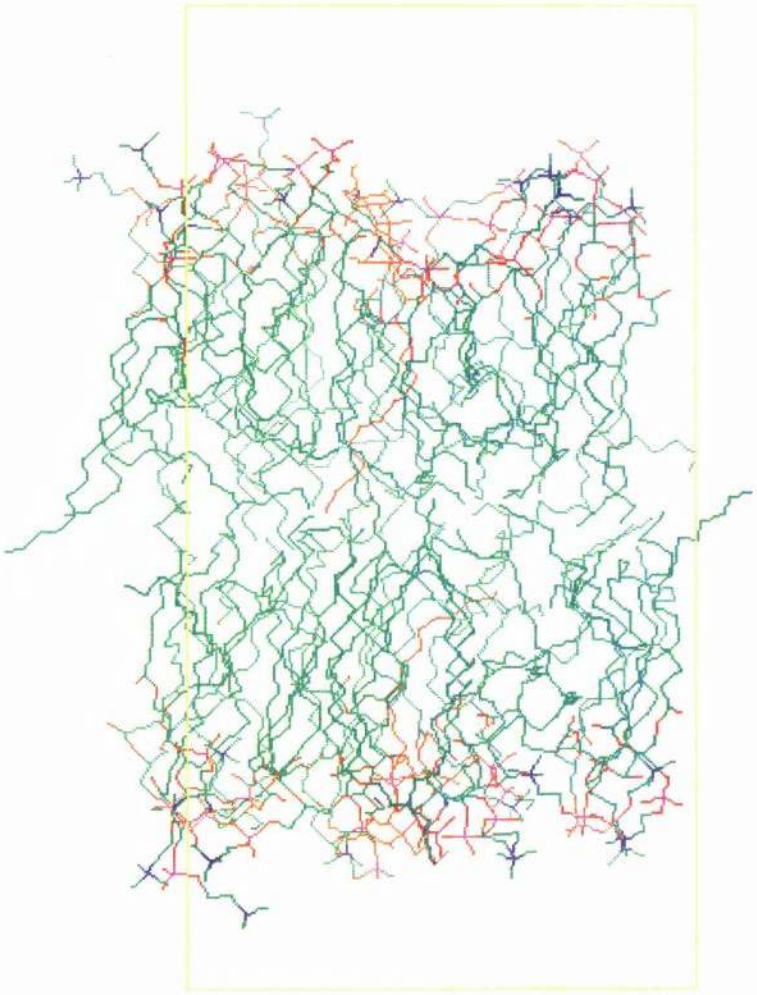


Fig 6.6. TPA in a Mixed Lipid Bilayer.

The colour conventions are the same as for fig. 6.3, except that the two phorbol esters (one in each monolayer) are displayed in orange.

6.5 Discussion.

The simulation of the dynamics of lipid bilayers is not a simple task as the work presented here has shown. However the results of these calculation are extremely promising as it has proved possible to develop model systems that remain stable in a condition similar to that expected from experimental data. In contrast, the small model systems of Berendsen *et al* [152, 153], and that presented here in section 6.2, failed to produce a viable system. The small headgroup, single chain fatty acids are too mobile to remain in a bilayer both in reality and in a dynamics simulation. They can be held in position by the use of an artificial potential [152], but this is not desirable as these potentials distort the system to an unknown degree. The report of the development of a stable bilayer using this fatty acid system [153], failed to take into account the separation of neighbouring bilayers under periodic boundary conditions. The extremely small solvation layer (<2 Å) and the use of charged molecules ensured that the bilayers were kept intact by a very large electrostatic repulsion. The size of the solvent layer is extremely important and must be sufficient, in relation to the non-bonded cutoff used, to prevent interaction of the bilayers. The use of constant pressure conditions especially combined with the HYDRO packing algorithm can cause some problems in this direction as the depth of the two solvation layers tend to decrease dramatically over the course of equilibration (this is caused both by the use of Van der Waals contact to terminate the packing and the large gap between solvent and bilayer to prevent solvation of the interior of the bilayer in EDIT). Berendsen *et al* have subsequently reported long-term stability of dynamics calculations on bilayers built up from phospholipids [154, 159]. More recently, Charifson *et al* [160] have reported dynamics simulation on solvated mixed (PC and PE) lipid monolayers, also using the AMBER molecular mechanics force field. In common with the other simulations discussed here, they used a regular array of lipids and were concerned primarily with the interactions between the head groups and calcium ions at the surface of the monolayer. Their calculation produced a slightly smaller surface area per lipid (53.4 Å after 64 ps), reflecting the smaller contribution to area from PE, and the orientations of

the P -> N vectors are comparable with those described here. As they were concerned only with interactions involving the headgroups, Charifson *et al* used only saturated fatty acid chains fixed in an all trans conformation and thus were able to ignore interactions between the chains and solvent that arose from considering only monolayers. This approach is not appropriate for calculations involving the phorbol esters in the lipid bilayer as the chains differ in both number and length from those of the phospholipids thus, particularly as affinity labelling studies label only lipid, it is essential to consider chain interactions.

The supermolecule approach used in all of these simulations has also been used to study aspects of the solvated headgroup and headgroup-ion interactions at an SCF level by Pullman *et al* [161-165]. Though the calculations do not involve a consideration of the dynamic behaviour, they have shown the minimum energy coordination structures of the groups possessing significant charge and have examined the differences between direct and "through water" interactions [161]. The electrostatic potentials at the surface of the bilayer and at distances above it have been calculated [166], using the PE crystal structure [167].

There are few differences between the simulation of the homogeneous PC bilayer and the simulation of a phorbol ester in a mixed lipid bilayer, most notably the degree of solvent penetration. The Berendsen work with fatty acid bilayers has considerable deep solvation penetration and cites experimental data to support the position and agree with the findings from the phorbol ester calculation, but not with the PC bilayer. The reason behind this may well lie in the relative sizes of the headgroups and number of chains. Phosphatidylcholine as a larger headgroup than the other phospholipids and, per acyl chain, proportionately larger than the fatty acids, thus, at any given temperature, would be less likely to have gaps between the headgroups caused by the flexing of the acyl chains and so be less susceptible to solvent penetration. The effect of headgroup size can also be seen in the disorder of the chains as in bilayers with proportionately smaller headgroups, the flexing of the chains is balanced by the need to separate the headgroups.

Chapter 7

Comparison of Different Theoretical Methods.

7.0 Introduction.

A thorough comparison of the heats of formation, geometric values and ionisation potentials calculated for a wide range of small molecules has been published [125] and includes a summary of the known deficiencies of the three parametric methods. MNDO is described as underestimating the stability of sterically crowded molecules (the core-core repulsion is overestimated) and of hypervalent compounds. Both of these contribute to the comparatively high energy given by MNDO for the inositol phosphates while the failure to reproduce the hydrogen bond introduces further error into the results. AM1 is believed to give significantly better results for all the above cases, but, for phosphorus, introduces a spurious sharp potential barrier at 3.0 Å which serves to distort the calculated geometries. The example of P_4O_6 is cited where AM1 produces differences of 0.4 Å in supposedly equal P-P bonds. By comparison PM3 is an extremely new method and though questions have been raised as to its value [144], it is claimed to represent a general improvement over AM1: however publications comparing its accuracy for systems outside the parametrisation set are, as yet, rare, e.g. [65].

The calculations on the phorbol derivatives, described in chapter 4, showed very little difference in structure, heat of formation and population analysis between PM3 and AM1 for the carbon, hydrogen and oxygen systems. Including nitrogen into the system, as for the teleocidins, produced a significant difference in the heats of formation and population analysis, but not in the final geometries. For the phosphate based systems however, the position is rather different. Substantial differences are noticeable in the heats of formation, structure and population analysis especially at the phosphate groups. It is therefore in the latter system where the accuracy and suitability of the different methods is of greatest concern.

The very large differences in the heats of formation obtained in chapter 5 for the inositol polyphosphates using the three different MNDO-type parametrisations raises the question of their accuracy for this type of molecule especially as simple organic phosphates such as these were not included in the original parametrisations [45, 62-64,

145]. In order to provide some sort of answer to this question, the properties of *myo* - inositol 1' monophosphate calculated using the different semi-empirical methods are compared with *ab initio* results and, where possible, with experimental data. The size of this molecule makes anything larger than a single-point, minimal basis set calculation unfeasible both in terms of disk requirements and processor time, so a smaller three-carbon phosphate ester was also considered. Again optimisation proved too time consuming but it was possible to perform single point calculations using a variety of basis sets before the constraints of disk space came into effect. There is no such problem with semi-empirical methods, so the properties are calculated for fully optimised structures with all three parameter sets.

7.1 Phosphate Ester Fragment.

Fig 7.1 shows the three-carbon fragment of inositol, containing a single phosphate group, that was used in the comparison of results from different *ab initio* basis sets with those of the semi-empirical methods. For the reasons mentioned above, the *ab initio* calculations were performed on a single point, chosen as the PM3 optimised geometry as this method gave heats of formation midway between the other two. As the *ab initio* calculations were not performed at a stationary point (for the particular method), the comparison was based on properties such as population analysis and the electrostatic potential, that are not very sensitive to changes in geometry and the forces on the atoms.

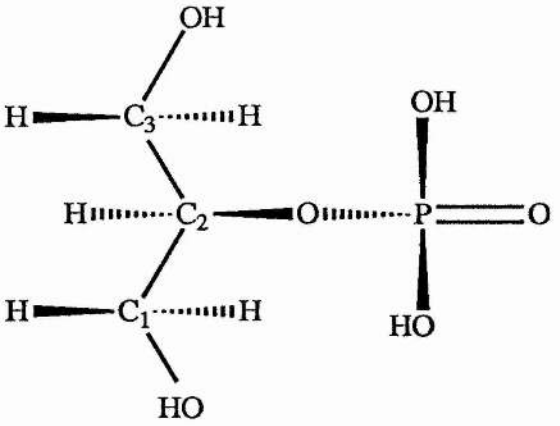


Fig 7.1 Phosphate Ester Fragment.

Table 7.1 Atomic Point Charges.

a/ *Ab Initio* charges from Mulliken Analysis.

	STO-3G	STO-3G*	STO-6G	STO-6G*	3-21G	3-21G*	4-31G
HC	0.060	0.063	0.062	0.065	0.223	0.222	0.194
HO	0.193	0.194	0.201	0.202	0.378	0.378	0.400
C1	0.008	0.009	0.009	0.009	-0.061	-0.063	0.026
C2	0.060	0.072	0.063	0.075	-0.021	-0.023	0.070
C3	0.013	0.012	0.013	0.013	-0.035	-0.037	0.048
OC	-0.313	-0.313	-0.325	-0.325	-0.679	-0.680	-0.726
Ester O	-0.435	-0.331	-0.445	-0.341	-0.868	-0.785	-0.950
O=P	-0.575	-0.321	-0.589	-0.339	-0.800	-0.574	-0.925
O-P	-0.470	-0.355	-0.486	-0.377	-0.853	-0.756	-0.912
P	1.379	0.711	1.414	0.753	2.094	1.625	2.295

b/ *Ab Initio* charges from the Electrostatic Potential.

	STO-3G	STO-3G*	STO-6G	STO-6G*	3-21G	3-21G*	4-31G
HC	0.019	0.019	0.023	0.023	0.026	0.026	0.017
HO	0.301	0.304	0.306	0.310	0.437	0.436	0.469
C1	0.251	0.269	0.249	0.267	0.411	0.417	0.526
C2	0.086	0.070	0.085	0.070	0.053	0.038	0.009
C3	0.122	0.141	0.121	0.140	0.288	0.284	0.399
OC	-0.488	-0.494	-0.499	-0.505	-0.721	-0.721	-0.802
Ester O	-0.563	-0.486	-0.574	-0.499	-0.782	-0.713	-0.862
O=P	-0.707	-0.530	-0.721	-0.546	-0.840	-0.728	-0.883
O-P	-0.704	-0.594	-0.717	-0.608	-0.938	-0.860	-1.001
P	1.790	1.278	1.822	1.316	2.098	1.799	2.264

c/ Semi-Empirical charges from the Density Matrix.

	AM1	PM3	MNDO
HC	0.093	0.049	0.013
HO	0.227	0.186	0.195
C1	-0.042	0.037	0.140
C2	-0.091	0.077	0.170
C3	-0.059	0.053	0.132
OC	-0.345	-0.306	-0.329
Ester O	-0.806	-0.686	-0.524
O=P	-1.080	-0.826	-0.651
O-P	-0.800	-0.669	-0.454
P	2.667	2.194	1.376

All of the basis sets listed above include s-functions on hydrogen and s and p-functions on the heavy atoms. In addition, those marked with an asterisk '*' have six d-functions on phosphorus, a change which is reflected in the substantial reduction of the phosphorus charge calculated using Mulliken population analysis compared with the corresponding s and p-only basis set. As the least squares fit from the electrostatic potential takes the Mulliken charges as a starting point, those values are reflected in the potential-based charges, thus the d-function basis sets still have a lower phosphorus charge, though some idea of the appropriate charges can be gained from the direction and magnitude of the changes in charge brought about by fitting to the potential. Thus the small change in the phosphorus charge at both the 3-21G and 4-31G levels suggests the appropriate value is around 2.15, close to that obtained by the PM3 calculation, and corresponding large negative charges on the phosphate oxygens. The effects of the phosphate ESP can be seen on the carbon charges, which are similar for any given basis set (or MOPAC charges) from Mulliken analysis, but differ when calculated from the potential. PM3 gives charges closer to the *ab initio* values, excluding the 3-21G and 3-21G* results, than either of the other two methods, and also fits with the intuitive expectation that C-H bonds exhibit minimal polarisation. Mulliken analysis on the split-valence wavefunction gives substantial polarisation of the C-H bonds, with negative carbon much as does AM1. By using double zeta functions only for the valence electron, the split valence basis sets distort the balance of functions between the heavy atoms and hydrogens resulting in a shift of charge during Mulliken analysis, and so would be expected to give less accurate charges than either a minimal basis set or a genuinely double zeta basis. The hydroxyl charges from all three semi-empirical methods are consistent with the minimal basis results.

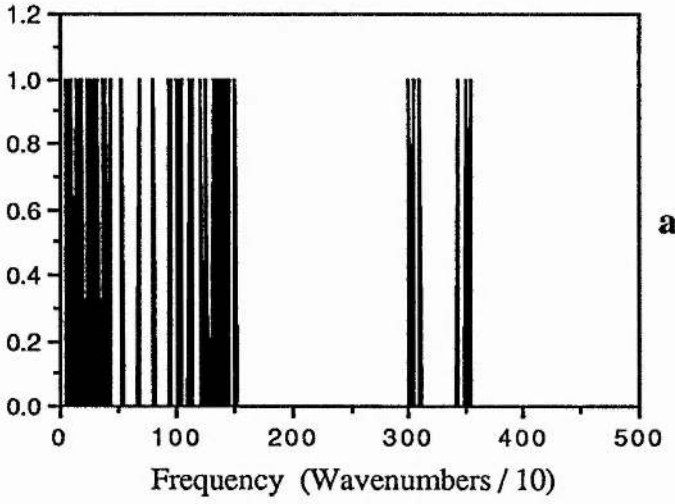
Table 7.2 Properties Determined by *Ab Initio* Calculation.

	Dipole Moment (Debye)	HOMO Energy (Hartree)	LUMO Energy (Hartree)	Total Energy (Hartree)
STO-3G	2.567	-0.3114	0.3567	-897.27
STO-3G*	2.601	-0.3301	0.3813	-897.67
STO-6G	2.579	-0.3173	0.3464	-905.30
STO-6G*	2.611	-0.3362	0.3714	-905.69
3-21G	3.370	-0.4217	0.1642	-903.85
3-21G*	3.332	-0.4192	0.1800	-904.07
4-31G	3.244	-0.4364	0.1324	-907.46

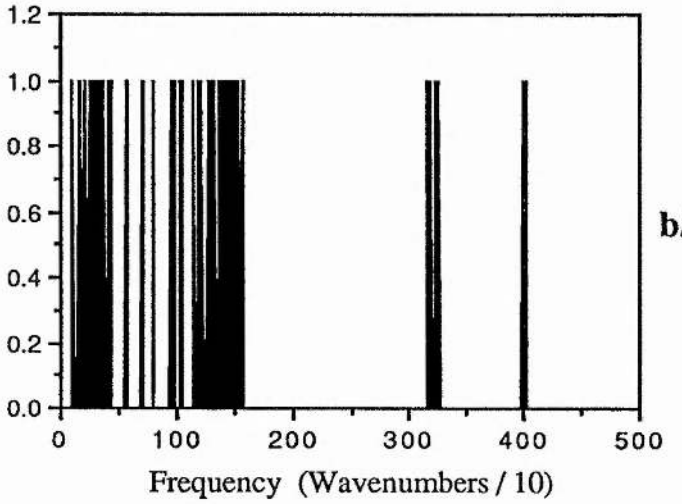
Table 7.3 Properties Determined by Semi-Empirical Calculation.

	AM1	MNDO	PM3
Dipole Moment (Debye)	6.007	4.135	2.627
HOMO Energy (Hartree)	-0.3972	-0.4041	-0.3949
LUMO Energy (Hartree)	0.0294	-0.0239	0.0039
Total Energy (Hartree)	-94.21	-95.66	-87.53
Heat of Formation (kcal mol^{-1} at 300 K)	-374.38	-294.41	-348.22
Heat Capacity (Cal K^{-1} mol^{-1} at 300 K)	44.67	44.32	47.97
Enthalpy (Cal mol^{-1} at 300 K)	8478	8546	9168
Entropy (Cal K^{-1} mol^{-1} at 300 K)	109.9	111.9	119.6

As with the atomic charges, the dipole moment shows some degree of basis sets dependence, with the minimal basis sets giving a value somewhat lower than that obtained from the split valence basis sets. By comparison, inclusion of d-functions on the phosphorus produces a much smaller effect. Of the semi-empirical methods, only PM3 gives a value comparable with the *ab initio* results, being at the top of the range for the minimal basis sets. With the exception of the heat of formation, there is little variation in the properties calculated with the semi-empirical methods, though the LUMO energy and total energy differ from the *ab initio* values. The factor of ten difference in the total energies between the two levels of theory is extremely surprising and no reason can be found for this.



a/ AM1



b/ MNDO

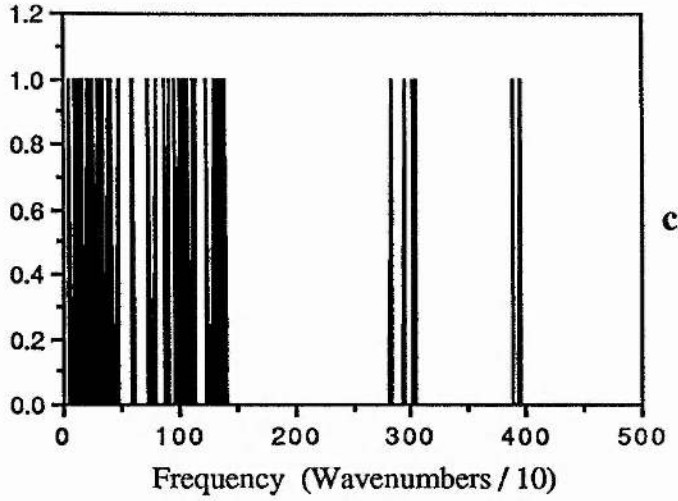


Fig 7.2 I.R. Frequencies for Phosphate Ester Fragment.

Table 7.4 Bond Stretching Frequencies (in cm^{-1}).

Bond	AM1	MNDO	PM3	Experimental
O-H	3421-3539	3999-4026	3883-3951	3335-3635
C-H	3007-3096	3152-3256	2823-3031	2855-2940
C-O	1327-1452	1374-1546	1125-1389	1000-1250
P=O	1398	1331	1083	1250
P-O	967-1120	803-1407	743-972	1000-1110
C-C	1261-1513	1157-1276	879-1035	910-1250

The treatment of the calculated I.R. spectra is the same as in chapter 5, comparing stretching frequencies with typical values obtained from the same source [126] as the spectrum has not been determined experimentally. No comparable analysis of other modes of vibration has been carried out as these are subject to greater variation due to substitution of groups. The results are largely the same as for *myo* -inositol (1,4,5) trisphosphate, except that the low frequency C-C stretching mode is not

present.

The calculated AM1 frequencies for the different bond types all lie within, or very close to, the range of experimental values with the exception of the P=O frequency, which only MNDO predicts as closer to the expected range. Against this MNDO consistently predicts the frequencies to be higher than experimental values, except for P-O (which has a large range of values) and C-C for which the frequencies are comparable with AM1. Contrary to expectation, PM3 does not give better results than AM1 as it substantially overestimates the O-H frequencies while underestimating both type of P-O frequencies. As phosphorus was included in the initial PM3 parametrisation [63, 64], but not that for AM1 [62, 145], PM3 would be expected to give more reliable results for phosphorus compounds.

7.2 Inositol 1' Monophosphate.

As in the previous section, the PM3 optimised geometry was used for the single-point *ab initio* calculations while the geometries were optimised for the other semi-empirical methods. The disk limitations restricted the *ab initio* calculations to a minimal basis which could not be overcome by calculating the 2 electron integrals only when required (direct SCF) as even an STO-3G single point took over 6 days processor time on the μVAX. The same property measurements were made as in the previous section and, in addition, bond lengths and angles were compared with values from X-ray crystallography [126, 139, 142, 143, 168, 169]. As all of these structures, except the hexakisphosphate [169] are in the "I" conformation (see table 5.2 and Fig 5.3), the starting point for all the semi-empirical calculations was chosen as the lowest energy conformation with this particular ring conformation, produced from the PM3 calculations (see chapter 5).

Table 7.5a *Ab Initio* Atomic Point Charges.

	STO-3G		STO-3G*		STO-6G		STO-6G*	
	Mulliken	ESP	Mulliken	ESP	Mulliken	ESP	Mulliken	ESP
HC	0.054	0.060	0.056	0.064	0.055	0.064	0.058	0.067
HO	0.213	0.336	0.212	0.338	0.219	0.344	0.220	0.345
C1	0.068	-0.139	0.078	-0.163	0.070	-0.133	0.080	-0.158
C2 (C6)	0.069	0.186	0.069	0.191	0.071	0.184	0.072	0.190
C3 (C5)	0.068	0.102	0.068	0.098	0.071	0.103	0.071	0.099
C4	0.066	0.206	0.066	0.209	0.068	0.205	0.068	0.207
OC	-0.334	-0.523	-0.333	-0.523	-0.345	-0.533	-0.345	-0.534
Ester O	-0.415	-0.428	-0.309	-0.319	-0.424	-0.440	-0.319	-0.333
O=P	-0.581	-0.698	-0.325	-0.519	-0.595	-0.712	-0.342	-0.535
O-P	-0.470	-0.671	-0.356	-0.565	-0.486	-0.685	-0.374	-0.580
P	1.377	1.714	0.707	1.185	1.412	1.747	0.749	1.225

The Mulliken based charges for both *ab initio* and semi-empirical methods show considerable similarity to those obtained from the model fragment, the only exception being the carbon charges. The position is similar for the ESP based charges but here the magnitude of the charges on the components of the phosphate group are quite noticeably reduced and the attached carbon atom has acquired a negative charge. This is not the case for with any of the semi-empirical charges were it has a larger positive charge than the other carbons regardless of parametrisation, though this is countered by the larger positive charge on the phosphorus.

Table 7.5b Semi-Empirical Atomic Point Charges.

	AM1	PM3	MNDO
HC	0.111	0.064	0.026
HO	0.228	0.216	0.197
C1	0.055	0.124	0.164
C2 (C6)	0.006	0.044	0.110
C3 (C5)	0.003	0.040	0.109
C4	0.000	0.040	0.110
OC	-0.333	-0.329	-0.324
Ester O	-0.744	-0.633	-0.508
O=P	-1.076	-0.847	-0.645
O-P	-0.789	-0.677	-0.454
P	2.610	2.192	1.356

Table 7.6 Molecular Properties.

	Dipole Moment (Debye)	HOMO Energy (Hartree)	LUMO Energy (Hartree)	Total Energy (Hartree)
STO-3G	1.730	-0.3094	0.3504	-1233.37
STO-3G*	1.768	-0.3279	0.3770	-1233.77
STO-6G	1.731	-0.3342	0.3668	-1244.65
STO-6G*	1.756	-0.3342	0.3668	-1245.04
AM1	2.371	-0.4056	0.0175	-145.73
MNDO	2.262	-0.3952	-0.0014	-135.25
PM3	1.706	-0.3992	-0.0318	-147.38

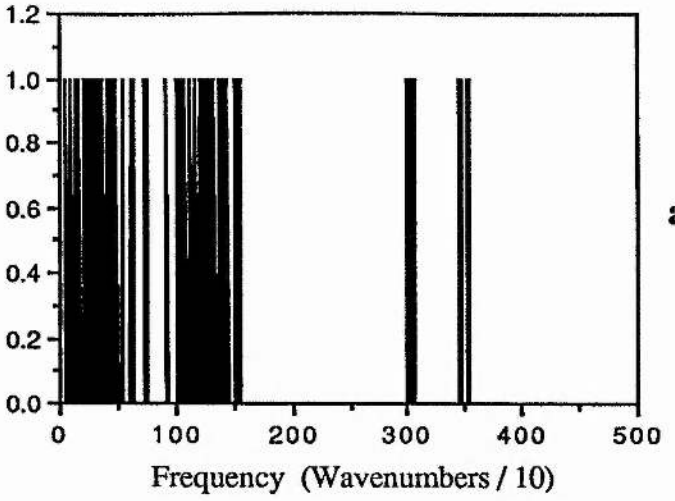
Again the property calculations show a very similar pattern to the equivalent calculations for the model fragment, confirming its validity. For the larger molecule, there is less variation in the semi-empirical dipole moments, with PM3 again closest to the *ab initio* results. The *ab initio* values for the energy of the HOMO and LUMO are very close to the values for the fragment, suggesting that the semi-empirical HOMO energies would be comparable with those of the split-valence basis sets. The same problem of disparity in the total energies is observed with no suggestion as to its cause.

The thermodynamic properties, shown below, also follow the same pattern shown by both the higher phosphate (chapter 5) and the model fragment, in that similar values are obtained with the three parameter sets for all properties except for the heat of formation.

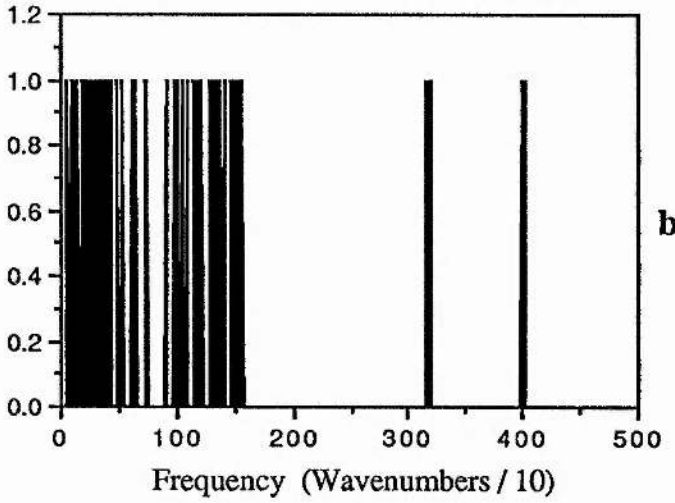
Table 7.7 Semi-Empirical Thermodynamic Properties (at 300 K).

	AM1	MNDO	PM3
Heat of Formation (kcal mol^{-1})	-522.97	-426.16	-473.40
Heat Capacity (Cal K^{-1} mol^{-1})	66.19	65.69	70.12
Enthalpy (kcal mol^{-1})	11.734	11.624	12.037
Entropy (kcal K^{-1} mol^{-1})	0.1351	0.1349	0.1352

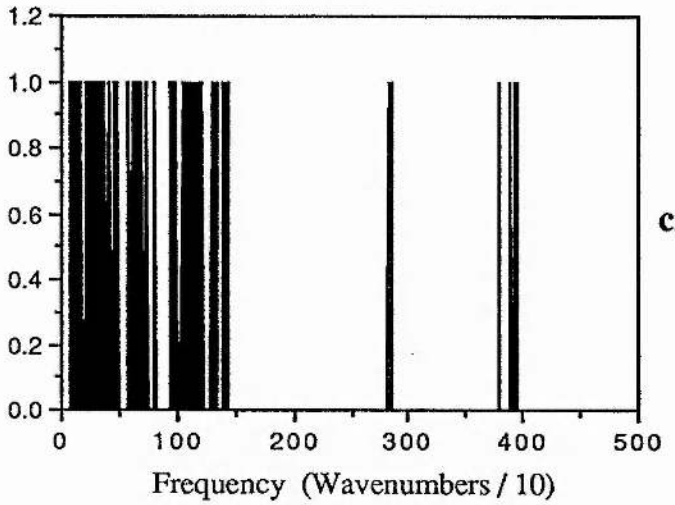
As would be expected, the ranges for the frequencies, shown in table 7.8, is essentially the same as for the model fragment and the trisphosphate. The overall increase in mass is reflected in the lowering of the bottom limit of the C-C stretching frequencies as compared with the fragment. The other significant change is a lowering of the PM3 P=O frequency further from the experimental value. This frequency is essentially unchanged with the other parameter sets, and is too large to be attributed to a change in the reduced mass and so must be taken a further evidence of the unreliability of PM3 for frequency calculation on this type of molecule.



a/ AM1



b/ MNDO



c/ PM3

Fig 7.3 *Myo* -Inositol 1' Monophosphate I.R. Frequencies.

Table 7.8 I.R. Frequencies for *Myo* -Inositol 1' Monophosphate (in cm^{-1}).

Bond	AM1	MNDO	PM3	Experimental
O-H	3456-3544	3993-4023	3793-3959	3335-3635
C-H	3005-3057	3166-3200	2822-2859	2855-2940
C-O	1305-1449	1371-1513	1087-1340	1000-1250
P=O	1409	1338-1342	997	1250
P-O	920-1089	918-1190	747-997	1000-1110
C-C	1028-1559	985-1551	946-1423	910-1250

Table 7.5 Semi-Empirical and Experimental Geometries.

	AM1	MNDO	PM3	Experimental
C-C	1.539 Å	1.585 Å	1.556 Å	1.523 Å
C-0(-H)	1.417 Å	1.397 Å	1.415 Å	1.431 Å
C-O(-P)	1.410 Å	1.395 Å	1.406 Å	1.458 Å
P-O(-H)	1.587 Å	1.599 Å	1.673 Å	1.551 Å
P-O(-C)	1.604 Å	1.609 Å	1.695 Å	1.586 Å
P=O	1.452 Å	1.494 Å	1.444 Å	1.470 Å
C-C-C	109.8°	111.5°	110.1°	111.1°
C-C-O(-H)	109.1°	109.7°	109.2°	109.6°
C-C-O(-P)	109.3°	109.3°	109.5°	109.0°
C-O-P	118.9°	127.5°	122.6°	121.1°
O=P-O(-C)	117.0°	117.5°	120.7°	108.0°
O-P-O(-C)	99.9°	102.6°	97.9°	107.9°

Experimental data is taken from X-ray crystal structures of *myo* -inositol [142, 143, 168] and *myo* -inositol 2' monophosphate [139]. Crystal structure data also exists for myo -inositol hexakisphosphate [169], but this is of the molecule as the duodeca sodium salt, so the three phosphorus oxygens are equivalent. There is also a further X-ray structure of *myo* -inositol [170], but the geometry, especially the bond lengths, differs considerably from the other studies [139, 142, 143, 168].

For all bond lengths, the AM1 results are closer to experiment than are either the MNDO or PM3 results, though advantage over PM3 is not significant for the C-O and O=P values (or over MNDO for P-O). There is no clear advantage for angle measurements, where the semi-empirical values lie within, or close to, the range of experimental values, except in the case of the phosphate oxygens. To ensure correct geometry for pentavalent phosphorus, it is essential to include d orbitals though none of the semi-empirical methods do so, preferring to use the trivalent s^2p^3 configuration rather than the $s^1p^3d^1$ configuration.

7.3 Discussion.

The results given in this chapter did not succeed in revealing which of the semi-empirical parametrisation would give a more reliable value for the heat of formation of the inositol polyphosphates, though certain conclusions as to which is more appropriate for the calculation of different properties. For both frequencies and geometries, especially at the phosphate group itself, AM1 seems to be considerably more reliable than PM3, though the charges, dipole moments of PM3 compare well with those from *ab initio* calculations whereas AM1 values do not. For all the properties calculated, MNDO in at best no better than the other methods, though as it is considerably older, and its deficiencies well known, this is not surprising.

With the exception of the heat of formation, all three methods give similar values for energies (both total and orbital) and for the other thermodynamic properties, which suggests that the large difference in the heats of formation stems largely from

differences in the energies of the isolated atoms.

The heat of formation is calculated according to the following formula:

$$\Delta Hf = E_{TOT} - E_{ATOMS} + \Delta Hf_{ATOMS}$$

Both E_{TOT} and E_{ATOMS} are large relative to ΔHf and ΔHf_{ATOMS} so slight errors in their calculation will result in large deviations in the heat of formation. The calculation of properties of the inositol polyphosphates and other organic phosphates is hampered by the lack of phosphorus d orbitals which will introduce errors into the calcualtion of both the energy and geometry of pentavalent phosphorus [161].

Chapter 8

Conclusions.

8.0 Introduction.

The purpose of this section is to suggest future progress possible in the area described in the chapters containing the calculations and in the final part to assess the effectiveness of PM3, compared with AM1 and MNDO, for the different types of systems dealt with in the project.

8.1 TPA-Like Tumour Promoters.

The promoter binding site, described in chapter 4, is considerably more detailed than any previously published [120, 122-124, 129]. These earlier descriptions of the site did not progress beyond the proposal of a 3 or 4 point hydrogen bonding network, though that of Itai *et al* [129] did suggest a hydrophobic pocket of the acyl chain of TPA and hexamethyl ring of teleocidin which is dealt with in chapter 4. Since the publication of some of the calculations presented in chapter 4 [171] with an initial description of the site, work has been presented [172] based on the study of other, similar tumour promoters, which supports the binding site as presented here. The superiority of the MNDO derived methods over the earlier MINDO/3 can be seen by comparison of the structures obtained, with those presented by Pack [173], which are of significantly lower quality.

The Dynamics simulation of TPA in a lipid bilayer was consistent with the requirements for the ester to be able to enter the binding site as the molecule maintained a stable orientation within the bilayer, presenting the critical A and B ring groups at the surface of the bilayer. It was found, however, that the ester tended to sink slightly into the bilayer and would thus bring PKC into contact with the headgroups of the neighbouring lipids. The composition of the lipid bilayer is known to have an effect on the activity of PKC [158], while Ganong *et al* [29] have suggested that a specific DAG / PS / Ca^{2+} complex is involved in activation of the enzyme.

There is not much potential for future work in this field as it is difficult to refine

the description of the binding site much further from a consideration of the structures of ligands alone, rather some idea of the structure of the enzyme itself is required. However, from the description of the site presented here, it should now be possible to identify potential TPA-like tumour promoters by examining their compatibility with the known features of the binding site, and thus minimise exposure to them and from an experimental point of view, the description of the site lends itself to the design of potential affinity labels. One such label may be to include a strong electrophile in place of the A-ring double bond and thus label the nucleophilic region identified from the LUMO calculation. It is not practicable to design inhibitors of PKC as it is too intimately involved in normal cellular functions. The activation complex of Ganong *et al* [29] deserves more consideration. It would be appropriate to study its stability within a lipid bilayer, using the method presented in chapter 6, though such a calculation would be extremely time consuming which is why it was not done as part of this project.

8.2 Lipid Bilayers.

Though other workers have also succeeded in producing stable solvated phospholipid monolayers and bilayers [159, 160], all have placed their lipids in a regular, square, array. This system will produce closely packed structures only if the surface areas of the components are comparable and can thus cannot deal with the insertion of proteins or radically different amphipathic molecules into the bilayer, while the regular array is a poor representation of the fluid-mosaic membrane model. The random orientations produced by HYDRO allow almost any collection of amphipathic molecules to be placed into a close-packed monolayer and thus the methods described in chapter 6 can, within the limitations of computer resources, be used to study a wide range of bilayer based systems. The study of effects of cholesterol in the bilayer has been suggested, however, the chain disorder found in the dynamics simulations is rather different from that described experimentally [156] and so further work needs to

be done to investigate this problem, and perhaps to alter the rotational parameters for the acyl chain.

8.3 Inositol Polyphosphates.

The considerable flexibility of the inositols and the dominance of the phosphate groups on the electrostatic potential made it impossible to carry out a similar structure / activity analysis to that carried out for the tumour promoters. By building up the different phosphate structures sequentially, the number of starting points were kept to a minimum, while the comparisons of the *tris* and *tetrakis* phosphates suggest that this method did not lead to significant minima being missed. In the absence of a structure for the receptor protein, the requirement for a specific ring conformation, suggested by this work, could be tested by producing structure / activity relationships for rigid systems.

The wide range in calculated heats of formation for the different semi-empirical parameters suggest that these methods are not particularly suitable for this system. In particular, the lack of phosphorus d functions is likely to introduce errors for pentavalent phosphorus [161], though not for trivalent compounds. Against this, the phosphate geometries compare well with those obtained from X-ray data.

8.4 An Assessment of PM3.

The need for the reparametrisations of MNDO as embodied in PM3 has been questioned by Dewar [144], and these results shown here tend to support that view. Its failure to deal correctly with amide nitrogen is acknowledged with the option to include a molecular mechanics correction in MOPAC, though its effectiveness has been questioned [128]. For the carbon, oxygen, hydrogen systems such as the phorbol esters, the heats of formation, geometries and the ESP's from point charges

obtained with both PM3 and AM1 are often almost identical, only the dipole moment differs substantially between the three methods. AM1 predicted I.R. frequencies are closer to typical experimental values than are those from PM3 for both the C,H,O systems and for the phosphates, though for all three methods, the P=O double bond stretching frequency is substantially different from the experimental data.

AM1 also gives results that are closer to the experimental values for inositol phosphate geometries and the relative energies of the "E" and "I" ring conformations for the different inositol phosphates are more consistent using the AM1 parameters than with PM3. Against this, the phosphorus charge obtained using AM1 is considerable more positive than that obtained with any other method, including *ab initio* methods. Some trends are noticeable with the phosphorus charges; those obtained from fitting to the ESP are generally larger than those obtained from Mulliken analysis and a larger basis set also favours a higher charge, while adding polarisation functions, not surprisingly, lowers the phosphorus charge. As the potential fitted charges use the Mulliken charges as a starting point, extra functions on the phosphorus atoms lowers the charge here also.

The results presented in this thesis suggest that PM3 gives results comparable with those from AM1 for C,H,O systems, but for systems containing nitrogen or phosphorus, the results are significantly worse. It cannot therefore be described as a "significant improvement over AM1" [115]. The main problem with using the MNDO derived methods for the organic phosphates is that they do not contain phosphorus d functions and so cannot give the correct electron configuration for pentavalent phosphorus.

Appendices.

Appendix 1 New AMBER Parameters.

A1.1 New Atom Types.

CS Aliphatic SP^2 carbon (all atom force field).

Hydrophobic.

Atomic Mass	12.010
Polarisability	1.405
Van der Waals radius	1.85
well depth	0.12

CU Aliphatic SP^2 carbon (united atom force field).

Hydrophobic.

Atomic Mass	13.020
Polarisability	1.405
Van der Waals radius	1.90
well depth	0.15

A1.2 Bond Lengths

Bond	Equilibrium Length	Force Constant
C -OS	1.348	407.0
CS-CS	1.324	648.0
CS-CT	1.505	315.0
CS-HC	1.075	347.0
CU-CU	1.324	648.0
CU-C2	1.505	315.0

A1.3 Bond Angles

Angle	Equilibrium Angle	Force Constant
C2-C -OS	111.9	88.5
O2-C -OS	122.95	70.0
C -CS-CS	121.5	48.3
C -CS-CT	116.9	84.5
CS-CS-CT	124.6	51.0
C -CS-HC	117.0	37.6
CS-CS-HC	121.2	35.6
CT-CS-CT	117.4	84.4
CT-CS-HC	116.3	40.8
HC-CS-HC	114.6	41.3
C -CU-CU	121.5	48.3
CU-CU-C2	124.6	51.0
CU-C2-C2	112.0	46.9
CS-CT-CT	112.0	46.9
CS-CT-HC	110.5	37.3
CS-CT-OH	107.7	82.2
C2-N3-C3	113.0	50.0
C3-N3-C3	113.0	50.0
C -OS-C2	119.6	38.5
C -OS-CH	119.6	38.5
OH-P -OH	98.53	67.26

A1.4 Dihedral Angles

Angle	Periodicity	Phase	Half Height
X -C -CS-X	2	180	1.23
X -CS-CS-X	2	180	28.25
X -C -CU-X	2	180	1.23
X -CU-CU-X	2	180	28.25
CH-OS-C -O2	2	180	4.72
	1	180	1.20
CH-OS-C -C2	2	180	4.72
	1	0	1.20
C2-OS-C -O2	2	180	4.72
	1	180	1.20
C2-OS-C -C2	2	180	4.72
	1	0	1.20
CT-CS-C -HC	2	180	1.23
	1	0	1.50
HC-CS-C -HC	2	180	1.23
	1	0	0.90
CT-CS-C -CT	2	180	6.11
C2-CU-C -C2	2	180	6.11
HC-CT-CS-C	3	180	0.222
	1	180	0.032
HC-CT-CS-CS	3	180	0.344
HC-CT-CS-HC	3	180	0.344
HC-CT-CS-CT	3	180	0.181

A1.4a Dihedral Angles (cont).

Angle	Periodicity	Phase	Half Height
CT-CT-CS-CT	3	0	0.462
	2	180	1.787
	1	0	1.926
CT-CT-CS-HC	3	0	1.413
	2	180	1.525
	1	180	0.641
CT-CT-CS-CS	3	180	1.413
	2	180	1.525
	1	0	0.641
OH-CT-CS-CT	3	0	1.270
	2	180	3.078
	1	0	1.156
OH-CT-CS-CS	3	180	1.270
	2	180	3.078
	1	180	1.156
C2-C2-CU-CU	3	180	1.413
	2	180	1.525
	1	0	0.641
C2-C2-CU-C2	3	0	0.462
	2	180	1.787
	1	0	1.926

Appendix 2 Abreviations.

AAMOM	An Approximate Molecular Orbital Method
ADP	Adenosine Diphosphate
AM1	Austin Method 1
ATP	Adenosine Triphosphate
cDNA	Chimeric Deoxyribose Nucleic Acid
CI	Configuration Interaction
cIP3	Cyclic Inositol (1:2,4,5) Trisphosphate
CNDO	Complete Neglect of Differential Overlap
qDAG	Diacylglycerol
EGF	Epidermal Growth Factor
ESP	Electrostatic Potential
GTP	Guanosine Triphosphate
INDO	Intermediate Neglect of Differential Overlap
IP_3	Inositol 1,4,5 Trisphosphate.
IP_4	Inositol 1,3,4,5 Tetrakisphosphate
MEP	Molecular Electrostatic Potential
MNDO	Modfied Neglect of Diatomic Overlap
mRNA	Messenger Ribose Nucleic Acid
NDDO	Neglect of Diatomic Differential Overlap
PC	Phosphatidyl Choline
PDBu	Phorbol Dibutyrate
PE	Phosphatidyl Ethanolamine
PKC	Protein Kinase C
PKM	Protein Kinase M (catalytic domains of PKC)
PM3	Parametric Method 3
PS	Phosphatidyl Serine
$PtdInsP_2$	Phosphatidyl Inositol 4,5 Bisphosphate

QSAR	Qualitative Structure Activity Data or Quantitative Structure Activity Data
RPA	13-O Retinoyl Phorbol 12-O Acetate
TPA	13-O Tetradecanoyl Phorbol 12-O Acetate

References.

1 P.Pott *facsimile in* Nat. Cancer Inst. Monog. (1963) **10** 7-13.

2 U.Saffiotti. (1985) Preface "Carcinogenesis, A Comprehensive Survey" Vol 8. eds. M.J.Mass, D.G.Kaugman, J.M.Siegfried, V.E.Steele, S.Nesnow. Raven Press (New York).

3 E.Hecker. (1982) Preface "Carcinogenesis, A Comprehensive Survey" Vol 7. eds. E.Hecker, N.E.Fusenig, W.Kunz, F.Marks, H.W.Thielmann. Raven Press (New York).

4 I.Berenblum. (1941) Cancer Res. **1** 44-50.

5 R.E.Albert. (1985) "Carcinogenesis, A Comprehensive Survey". Vol 8. pp443-457. eds. M.J.Mass, D.G.Kaugman, J.M.Siegfried, V.E.Steele, S.Nesnow. Raven Press (New York).

6 E.Hecker, W.Adolf, M.Hergenhahn, R.Schmidt, B.Sorg. (1984) "Cellular Interactions by Environmental Tumour Promoters." pp3-36. Eds. H.Fujiki, T.Sugimura. Japan Scientific Society Press (Tokyo).

7 M.L.M.van de Poll, D.A.M.van der Hulst, A.D.Tates, J.H.N.Meerman. (1990) Carcinogenesis **11** 333-339.

8 J.D.Scribner, N.K.Scribner. (1982) "Carcinogenesis, A Comprehensive Survey". Vol 7.pp13-19 eds. E.Hecker, N.E.Fusenig, W.Kunz, F.Marks, H.W.Thielmann. Raven Press (New York).

9 A.K.Verma. (1982) "Carcinogenesis, A Comprehensive Survey." Vol. 7. pp35-41. Eds. E.Hecker, N.E.Fusenig, W.Kunz, F.Marks, H.W.Thielmann. Raven Press (New York).

10 M.Takahashi, F.Furukawa, K.Toyoda, H.Sato, R.Hasegawa, K.Imaida, Y.Hayashi. (1990) Carcinogenesis. **11** 393-395.

11 E.Hecker. (1978) "Carcinogenesis, A Comprehensive Survey." Vol 2, pp11-48. Eds. T.J.Slaga, A.Sivak, R.K.Boutwell. Raven Press (New York).

12 J.L.Mumford, C.T.Helmes, X.Lee, J.Seidenberg, S.Nesnow. (1990) Carcinogenesis **11** 397-403.

13 K.Goettle, H.Loehrke, J.Schweizer, B.Hesse (1980) Vichows Arch (Pathol. Anat.) **385** 181-186.

14 K.Goettle, H.Loehrke, J.Schweizer, B.Hesse (1980) Cancer Res. **40** 155-161.

15 K.Goettle, H.Loehrke (1977) Virchows Arch. (Pathol. Anat.) **376** 117-122.

16 T.J.Slaga, S.M.Fischer, C.E.Weeks, A.J.P.Klein-Szanto, J.Reiners. (1982) J. Cell Biochem. **18** 99-119.

17 H.Fujiki, M.Suganuma, T.Tahira, A.Yoshioka, M.Nakayasu, Y.Endo, K.Shudo, S.Takayama, R.E.Moore, T.Sugimura. (1984) "Cellular Interactions by Environmental Tumour Promoters." pp37-48. Eds. H.Fujiki, T.Sugimura. Japan Scientific Society Press (Tokyo).

18 M.Hergenhahn, M.Detlefs, M.Fellhaauer, J.Blehl, U.Kloz, E.Hecker Personal communication.

19 C.L.Ashendel. (1986) Biochim. Biophys. Acta. **822** 219-242.

20 Y.Nishizuka. (1984) Nature **308** 693-698.

21 N.A.Sharkey, K.L.Leach, P.M.Blumberg. (1984) PNAS **81** 607-610.

22 J.G.Ebeling, G.R.Vandenbark, L.J.Kulin, B.R.Ganong, R.M.Bell, J.E.Niedel. (1985) PNAS **82** 815-819.

23 C.Cochet, C.Souvignet, M.Keramidas, E.M.Chambaz. (1986) Biochem. Biophys. Res. Comm. **134** 1031-1037.

24 Y.Takai, A.Kishimoto, M.Inoue, Y.Nishizuka. (1977) J. Biol. Chem. **252** 7603-7609.

25 U.Kikkawa, A.Kishimoto, Y.Nishizuka. (1989) Ann. Rev. Biochem. **58** 31-44.

26 M.Inoue, A.Kishimoto, Y.Takai, Y.Nishizuka. (1977) J. Biol. Chem. **252** 7610-7616.

27 Y.Takai, A.Kishimoto, Y.Kawahara, T.Mori, Y.Nishizuka. (1979) J. Biol. Chem. **254** 3692-3695.

28 A.Kishimoto, Y.Takai, T.Mori, U.Kikkawa, Y.Nishizuka. (1980) J. Biol. Chem. **255** 2273-2278.

29 B.R.Ganong, C.R.Loomis, Y.A.Hanun, R.M.Bell. (1986) PNAS **83** 1184-1188

30 Y.Nishizuka. (1986) Science **233** 305-312.

31 U.Kikkawa, Y.Nishizuka. (1986) Ann. Rev. Cell Biology **2** 149-178.

32 Y.Nishizuka. (1988) Nature **334** 661-665.

33 L.Coussens, P.J.Parker, L.Rhee, T.L.Yang-Feng, E.Chen. (1986) Science **223** 859-866.

34 L.Coussens, L.Rhee, P.J.Parker, A.Ullrich. (1987) DNA **6** 389-394.

35 J.L.Knopf, M.H.Lee, L.A.Sultzman, R.W.Kriz, C.R.Loomis. (1986) Cell **46** 491-502.

36 S.Ohno, H.Kawasaki, S.Imajoh, K.Suzuki, M.Inagaki. (1987) Nature **325** 161-166.

37 P.M.Blumberg, S.Jaken, B.Konig, N.A.Sharkey, K.L.Leach, A.Y.Jeng, E.Yeh. (1984) Biochem. Pharm. **33** 933-940.

38 M.J.Berridge. (1984) Biochem. J. **220** 345-360.

39 M.J.Berridge, R.F.Irvine. (1984) Nature **312** 315-321.

40 M.J.Berridge. (1987) Ann. Rev. Biochem. **56** 161-193.

41 T.Furuichi, S.Yoshikawa, A.Miyawaki, K.Wada, N.Maeda, K.Mikoshiba. (1989) Nature **342** 32-38.

42 C.D.Ferris, R.L.Huganir, S.Supattapone, S.H.Snyder. (1989) Nature. **342** 87-89.

43 N.Maeda, M.Niinobe, K.Mikoshiba. (1990) EMBO J. **9** 61-67.

44 C.A.Hansen, S.Mah, J.R.Williamson. (1986) J. Biol. Chem. **261** 8100-8103.

45 M.J.S.Dewar, W.Thiel. (1977) JACS **99** 4899-4907.

46 A.Szabo, N.S.Ostlund. (1982) "Modern Quantum Chemistry." MacMillan (New York).

47 J.A.Pople, J.S.Binkley, R.Seeger. (1976) Int. J. Quantum Chem. **S10** 1-19.

48 J.A.Pople, R.Seeger, R.Krishnan. (1977) Int. J. Quantum Chem. **S11** 149-163.

49 P.W. Atkins. (1983) "Molecular Quantum Mechanics." Oxford University Press (Oxford).

50 R.McWeeny. (1979) "Coulson's Valence, 3rd Edn." Oxford University Press (Oxford).

51 O.Goscinski, E.Brändas. (1968) Chem. Phys. Lett. **2** 299-302.

52 C.C.J.Roothan. (1951) Rev. Mod. Phys. **23** 69-89

53 G.G.Hall. (1951) Proc. R. Soc. London A. **205** 541-552.

54 W.J.Hehre, R.F.Stewart, J.A.Pople. (1969) J. Chem. Phys. **51** 2657-2664.

55 T.H.Dunning, P.J.Hay. (1977) "Methods of Electronic Structure Theory." pp1-27. Ed. H.F.Schaefer. Plenum Press. (New York).

56 Gaussian 90. M.J.Frisch, M.Head-Gordon, G.W.Trucks, J.B.Foresman, H.B.Schlegal, K.Raghavachari, M.A.Robb, J.S.Binkley, C.Gonzalez, D.J.Defrees, D.J.Fox, R.A.Whiteside, R.Seeger, C.F.Melius, J.Baker, R.L.Martin, L.R.Khan, J.J.P.Stewart, S.Topiol, J.A.Pople. (1990) Gaussian Inc. Pittsburg. PA.

57 W.J.Hehre, R.Ditchfield, J.A.Pople. (1972) J. Chem. Phys. **56** 2257-2261.

58 J.B.Collins, P.v.R.Schleyer, J.S.Binkley, J.A.Pople. (1976) J. Chem. Phys. **64** 5142-5151.

59 J.A.Pople, D.L.Beveridge, P.A.Dobosh. (1967) J. Chem. Phys. **47** 2026-2633.

60 J.A.Pople, D.L.Beveridge. (1970) "Approximate Molecular Orbital Theory." McGraw-Hill. (New York).

61 M.Zerner. (1975) J. Chem. Phys. **62** 2788-2799.

62 M.J.S.Dewar, E.G.Zoebisch, E.F.Healy, J.J.P.Stewart. (1985) JACS **107** 3902-3909.

63 J.J.P.Stewart. (1989) J. Comp. Chem. **10** 209-220.

64 J.J.P.Stewart. (1989) J. Comp. Chem. **10** 221-264.

65 P.Scano, C.Thomson. (1990) J. Comp. Chem. In Press.

66 P.Pulay. (1969) Mol. Phys. **17** 197-204.

67 H.B.Schlegel. (1982) J. Comp. Chem. **3** 214-218.

68 R.Bonaccorsi, E.Scrocco, J.Tomasi. (1970) J. Phys. Chem. **52** 5270-5284.

69 M.L.Connolly. (1983) Science. **221** 709-713.

70 M.L.Connolly. (1983) J. Applied Cryst. **16** 548-558.

71 D.R.Higgins. (1988) PhD Thesis, University of St.Andrews, St.Andrews.

72 R.Carbo, L.Leyda, M.Arnau. (1980) Int. J. Quantum Chem. **17** 1185-1189.

73 E.E.Hodgkin, W.G.Richards. (1986) J.Chem.Soc., Chem. Comm. 1342-1344.

74 D.D.Schillady, F.P.Billingsley, J.E.Bloor. (1961) Theoret. Chim. Acta. **21** 1-8.

75 R.S.Mulliken. (1955) J. Chem. Phys. **23** 1833-1840.

76 A.Hinchliffe. (1988) "Computational Quantum Chemistry." John Wiley and Sons. (Chichester).

77 U.Burkert, N.L.Allinger. (1982) "Molecular Mechanics. ACS Monograph." American Chemical Society (Washington).

78 F.A.Momany, R.F.McGuire, A.W.Burgess, H.A.Scheraga. (1975) J. Phys. Chem. **79** 2361-2381.

79 N.L.Allinger. (1977) J. Am. Chem. Soc. **99** 8127-8134.

80 N.L.Allinger, Y.H.Yuh, J.Lii. (1989) J. Am. Chem. Soc. **111** 8551-8582.

81 P.K.Weiner, P.A.Kollman. (1981) J. Comp. Chem. **2** 287-303.

82 S.J.Weiner, P.A.Kollman, N.T.Nguyen, D.A.Case. (1986) J. Comp. Chem. **7** 230-252.

83 L.Verlet. (1967) Phys. Rev. **159** 98.

84 M.P.Allen, D.J.Tildesley. (1987) "Computer Simulation of Liquids." Oxford University Press (Oxford).

85 G.M.Crippen. (1981) "Distance Geometry and Conformational Calculus." Research Studies Press.

86 S.Kirkpatrick, C.D.Gellat, M.P.Vecchi. (1983) Science **220** 671-680.

87 QCPE - Quantum Chemistry Program Exchange. Editor; Richard Counts. University of Indiana, Bloomington, Indiana.

88 P.K.Weiner, P.A.Kollman, J.Caldwell, U.Chandra-Singh. (1987) AMBER version 3.0, UCSF, San Fransisco.

89 G.Nemethy, M.S.Potle, H.A.Scheraga. (1983) J. Phys. Chem **87** 1883-1887.

90 M.J.Sippl, G.Nemethy, H.A.Scheraga. (1984) J. Phys. Chem. **88** 6231-6233.

91 B.R.Brooks, R.E.Bruccoleri, B.D.Olafson, D.J.States, S.Swaminathan, M.J.Karplus. (1983) J. Comp. Chem. **4** 187-217.

92 F.C.Bernstein, T.F.Koetzle, G.J.B.Williams, E.F.Meyer (jr), M.D.Brice, J.R.Rodgers, O.Kennard, T.Shimanouchi, M.Tasumi. (1977) J. Mol. Biol. **112** 535-542.

93 E.E.Abola, F.C.Bernstein, S.H.Bryant, T.F.Koetzle, J.Weng. (1987) "Crystallographic Databases - Information Content, Software Systems, Scientific Applications." pp107-132. Eds. F.H.Allen, G.Bergerhoff, R.Sievers. Data Commission of the International Union of Crystallography. (Bonn, Cambridge, Chester).

94 J.J.P.Stewart. MOPAC. Personal Communication.

95 Cambridge Structural Database, Daresbury Laboratory, Daresbury, Warrington.

96 M.Dupuis, D.Spangler, J.J.Wendoloski. (1980) NRCC Program QG01.

97 R.D.Amos, J.E.Rice. (1987) "CADPAC" The Cambridge Analytical Derivatives Package. Issue 4.0, Cambridge.

98 P.A.Kollman, U.Chandra-Singh. (1987) QUEST v 3.0. UCSF.

99 C.Burt. ASP Automated Similarity Index Package. (1990) Oxford Molecular.

100 J.A.Nelder, R.Mead. (1965) Comput. J. **7** 308-313.

101 CHEM-X. Developed and distributed by Chemical Design Ltd. Oxford, U.K.

102 C.Edge. (1987) PhD Thesis, University of St.Andrews, St.Andrews.

103 M.L.Connolly. (1981) QCPE Bull. **1** 75. **QCPE No.429**

104 M.H.Charlton. Personal Communication.

105 R.Brasseur, E.Goormaghtigh, J.M.Ruysschaert. (1981) Biochem. Biophys. Research Comm. **103** 301-310.

106 R.D.Bonfil, S.Moniki, R.Fridman, R.Reich, R.Reddel, C.C.Harris, A.Klein-Szanto. (1989) Carcinogenesis **10** 2335-2338.

107 R.K.Boutwell, M.Takigawa, A.K.Verma, C.L.Ashendel. (1984) "Cellular Interactions by Environmental Tumour Promoters." pp177-194. eds H.Fujiki, E.Hecher, R.E.Moore, T.Sugimura. Japan Scientific Press (Tokyo).

108 H.Fujiki, M.Mori, T.Sugimura, H.Ohigashi, M.Hirota, K.Koshimizu. (1982) "Carcinogenesis, A Comprehensive Survey." Vol 7. pp65-66 eds. E.Hecker, N.E.Fusenig, W.Kunz, F.Marks, H.W.Thielmann. Raven Press (New York).

109 I.B.Weinstein, J.Arcoleo, J.Backer, A.Jeffrey, W.W.Hsiao, S.Gattoni-Celli, P.Kirschmeier, E.Okin. (1984) "Cellular Interactions by Environmental Tumour Promoters." pp59-74 eds H.Fujiki, E.Hecher, R.E.Moore, T.Sugimura. Japan Scientific Press (Tokyo).

110 E.V.Wattenberg, D.Uemura, K.L.Byron, M.L.Villereal, H.Fujiki, M.R.Rosner. (1989) Cancer Res. **49** 5837-5842.

111 E.Hecker, R.Schmidt. (1974) L. Prog. Chem. Org. Natur. Prod. **31** 377-467.

112 E.Hecker. (1978) "Carcinogenesis, A Comprehensive Survey". Vol 2. pp11-48. eds.T.J.Slaga, A.Sivak, R.K.Boutwell. Raven Press (New York).

113 H.Fujiki, M.Suganuma, M.Nakayasu, T.Tahira, Y.Endo, K.Shudo, T.Sugimura. (1984) Gann. **75** 866-870.

114 T.Horiuchi, H.Fujiki, M.Suganuma, H.Hakii, M.Nakayasu, Y.Hitotsuyanagi, N.Aimi, S.Sakai, Y.Endo, K.Shudo, T.Sugimura. (1984) Gann. **75** 837-840.

115 J.J.P.Stewart. (1990) J. Comp. Chem. **11** 543-544.

116 F.Brandl, M.Rohrl, K.Zechmeister, W.Hoppe. (1971) Acta Cryst. **27B** 1718-1730.

117 H.Harada, N.Sakabe, Y.Hirata, Y.Tomiie, I.Nitta. (1966) Bull. Chem. Soc. Japan. **39** 1773-1775.

118 K.Zechmeister, F.Brandl, W.Hoppe, E.Hecker, H.J.Opferkuch, W.Adolf. (1970) Tet. Lett. 4075-4078.

119 I.Pascher, S.Sundell, H.Hauser. (1981) J. Mol. Biol. **153** 791-806.

120 A.F.Cuthbertson, C.Thomson. (1987) J. Mol. Graphics. **5** 92-96.

121 R.Miyake, Y.Tanaka, T.Tsuda, J.Yamanishi, U.Kikkawa, Y.Nishizuka. (1984) "Cellular Interactions by Environmental Tumour Promoters." pp167-176 eds H.Fujiki, E.Hecher, R.E.Moore, T.Sugimura. Japan Scientific Press (Tokyo).

122 H.Brokerhoff. (1986) FEBS Lett. **2011**-4.

123 P.A.Wender, K.F.Koehler, N.A.Sharkey, M.L.Dell'Aquila, P.M.Blumberg. (1986) Proc. Natl. Acad. Sci. USA. **83** 4214-4218.

124 A.M.Jeffrey, R.M.J.Liskamp. (1986) Proc. Natl. Acad. Sci. USA. **83** 241-245.

125 J.J.P.Stewart. (1990) J. Comput. Aided Mol. Design. **4** 1-105.

126 C.J.Pouchert. (1978) "The Aldrich Library of I.R.Spectra." Aldrich Chemical Company (Wisconsin).

127 Y.Endo, K.Shudo, T.Okamoto. (1982) Chem. Pharm. Bull. **30** 3457-3460.

128 A.S.Phin. Personal Communication.

129 A.Itai, Y.Kato, N.Tomioka, Y.Iitaka, Y.Endo, M.Hasegawa, K.Shudo, H.Fujiki, S.Sakai. (1988) Proc. Natl. Acad. Sci. USA. **85** 3688-3692.

130 S.Supattapone, P.F.Worley, J.M.Baraban, S.H.Snyder. (1988) J. Biol. Chem. **263** 1530-1534.

131 P.F.Irvine, A.J.Letcher, J.P.Hislop, M.J.Berridge. (1986) Nature. **320** 631-634.

132 I.R.Batty, S.R.Nahorski, R.F.Irvine. (1985) Biochem J. **232** 211-215.

133 T.M.Connolly, T.E.Bross, P.W.Majerus. (1985) J. Biol. Chem. **260** 7868-7874.

134 D.J.Storey, S.B.Shears, C.J.Kirk, R.H.Michell. (1984) Nature **312** 374-376.

135 D.B.Wilson, T.M.Connolly, T.E.Bross, P.W.Majerus, W.R.Sherman. (1985) J. Biol. Chem. **260** 13496-13501.

136 P.T.Hawkins, C.P.Berrie, A.J.Morris, C.P.Downes. (1986) Biochem J. **243** 211-218.

137 T.M.Connolly, D.B.Wilson, T.E.Bross, P.W.Majerus. (1986) J. Biol. Chem. **261** 122-126.

138 T.S.Ross, P.W.Majerus. (1986) J. Biol. Chem. **261** 11119-11123.

139 C.S.Yoo, G.Blank, J.Pletcher, M.Sax. (1974) Acta Cryst. **B30** 1983-1987.

140 G.E.Blank, J.Pletcher, M.Sax. (1974) Acta Cryst. **B31** 2584-2592.

141 T.R.Lomer. (1964) Acta Cryst. **16** 264-268.

142 I.N.Rabinowitz, J.Kraut. (1965) Acta Cryst. **17** 159-168.

143 G.Blank. (1973) Acta Cryst. **B29** 1677-83.

144 M.J.S.Dewar, E.F.Healy, A.J.Holder, Y.C.Yuan. (1990) J. Comp. Chem. **11** 541-542.

145 M.J.S.Dewar, C.Jie. (1989) Theochem. **56** 1-13.

146 M.Wolf, H.LeVine, W.S.May, P.Cuatrecasas, N.Sahyoun. (1985) Nature **317** 546-549.

147 M.Hergenhahn, E.Hecker. (1984) "Models, Mechanisms and Etiology of Tumour Promotion" pp165-175. Eds. M.Borzsonyl, K.Lapis, N.E.Day, H.Yamasaki. IARC (Lyon).

148 K.B.Declos, E.Yeh, P.M.Blumberg. (1983) Proc. Natl. Acad. Sci. USA. **80** 3054-3058.

149 R.Schmidt, K.Hech, B.Sorg, E.Hecker. (1985) Cancer Lett. **26** 97-111.

150 M.D.Houslay, K.K.Stanley. (1982) "Dynamics of Biological Membranes" John Wiley. (New York).

151 S.J.Singer, G.L.Nicholson. (1972) Science **175** 720-731.

152 P.Van der Ploeg, H.J.C.Berendsen. (1982) J. Chem. Phys. **76** 3271-3276.

153 E.Egberts, H.J.C.Berendsen. (1988) J. Chem. Phys. **89** 3718-3732.

154 H.J.C.Berendsen, W.F.Van Gunsteren. (1988) Bios Biomolecular Study Centre, Groningen. Annual Report 1988.

155 R.Brasseur, V.Cabiaux, P.Huart, M.Castagna, S.Baztar, J.M.Ruysschaert. (1985) Biochem. Biophys. Res. Comm. **127** 969-976.

156 J.Seelig. (1977) Q. Rev. Biophys. **10** 353-418.

157 J.Seelig, H.U.Gally, R.Wohlgemuth. (1977) Biochim. Biophys. Acta. **467** 109-119.

158 M.Hoshijima, A.Kikuchi, T.Tamimoto, K.Kaibuchi, Y.Taakai. (1986) Cancer Res. **46** 3000-3004.

159 E.Egberts. (1988) PhD. Thesis. University of Groningen.

160 P.S.Charifson, R.G.Hiskey, L.G.Pedersen. (1990) J. Comp. Chem. **11** 1181-1186.

161 B.Pullman, A.Pullman, H.Berthod. (1978) Int. J. Quant. Chem. Quant. Biol. Symp. **5** 79-90.

162 B.Pullman, A.Pullman, H.Berthod, N.Gresh. (1975) Theor. Chim. Acta. **40** 93-111.

163 B.Pullman, N.Gresh, H.Berthod, A.Pullman. (1977) Theor. Chim. Acta. **44** 151-163.

164 A.Pullman, B.Pullman, H.Berthod. (1978) Theor. Chim. Acta. **47** 175-192.

165 N.Gresh, P.Claverie, A.Pullman. (1979) Int. J. Quant. Chem. **S13** 243-253.

166 K.Zakrzewska, A.Pullman. (1982) Int. J. Quant. Chem. Quant. Biol. Symp. **9** 411-424.

167 P.B.Hitchcock, R.Mason, K.M.Thomas, G.G.Shipley. (1977) Proc. R. Soc. London. **A354** 157-170.

168 W.J.Cook, C.E.Bugg. (1973) Acta Cryst. **B29** 2404-2411.

169 G.E.Blank, J.Pletcher, M.Sax. (1974) Acta Cryst. **B31** 2584-2592.

170 T.R.Lomer, A.Miller, C.E.Beevers. (1963) Acta Cryst. **16** 264-268.

171 C.Thomson, J.Wilkie. (1989) Carcinogenesis. **10** 531-540.

172 R.Rein. (1990) Presented at the NFCR Annual Meeting, Washington, September.

173 G.R.Pack. (1981) Cancer Biochem. Biophys. **5** 183-188.