The Electrostatic Interactions of Protein-Binding Ligands.

Robert Patrick Apaya

A thesis submitted in partial fulfilment of the requirements of the University of London for the degree of Doctor of Philosophy,
July 1996.

University College London,
Department of Chemistry,
Christopher Ingold Laboratories,
20 Gordon Street,
London, WC1H 0AJ.
Personal Acknowledgements

I would like to thank my Mother and my Father for their unending support during all of my university studies. This has been particularly important during the last year and a half, which brought changes and challenges for my family which have affected us all.

During my first two years at UCL I worked in the G25 Theoretical Chemistry Lab. Many thanks to Sandra Lipari for making the first six months of those two years so memorable, *no ho mai riso tanto da quando*...

Special thanks Katy Bostock, also of G25 fame, who has been a constant source of encouragement in the face of all my complaints.

It is a great pleasure to thank the ‘Chemists...in Theory?’ of Lab 124 who made life in the lab so...*indescribable* during my final year at UCL, in their own unique ways. I would like to thank David Coombes, and to thank Irene Nobeli and Jasbir Bachoo - whom I never would have met if all had gone according to the ‘official’ timetable. However, they both helped to turn the lab into a better place during the last year ... mainly through the addition of soft furnishings, coffee tables and pastel hints... Thanks also to Giovanni Di Gregorio; *cartoonista, filosofo e amante internazionale delle donne che amano gli uomini calvi*. ‘Eos’ sarà molto grata per la tua ammirazione.

It is a pleasure to thank Maria Bondi, who worked with me as an Erasmus project student during my final year, and some of whose results have been included in this thesis. *Grazie alla ‘bella ragazza siciliana’ che* (in her own words) *ha portato un po’ di sole nel nostro laboratorio!*

*Me gustaría agradecer Juan Frau Munar su amistad y su apoyo en innumerables situaciones, a pesar de las circunstancias personales que hicieron 1994-95 un año tan difícil para mí. Gracias por tu paciencia, aunque no siempre la mereciera. Gracias por tu ayuda con mi castellano y los caracoles!*

Finally, I would also like to thank all the friends who listened with wry smiles when I told them that I would finish writing ‘next week’. Obviously *this* is the week I was talking about...!
Academic Acknowledgments

It is a great pleasure to thank my supervisor Dr. Sally Price, for her guidance and enthusiasm over the last three years, and for her patience over the last six months. Her comments and criticisms have served to improve the content and structure of this dissertation.

I would like to thank Dr. Andy Vinter, whose early work along with that of his co-workers, served as the springboard for some of the ideas developed in this dissertation.

I would also like to acknowledge the help of Dr. Graham Sexton at Zeneca, who suggested Glycolate Oxidase as the enzyme system studied in chapter 5, and who provided useful guidance and comments on the results obtained.

Finally I would like to thank Dr. Nora van der Wenden and Dr. Keith Wibley, who both worked as Postdoctoral researchers in our lab, and who helped to clarify some of my ideas about computational biological modelling.
The Electrostatic Interactions of Protein-Binding Ligands. Robert Apaya

A method for determining the relative binding orientation of different ligands within an unknown binding site is developed. By overlaying the maxima and minima in the electrostatic potential outside ligands, it is possible to match regions where strong electrostatic interactions, including hydrogen bonds, with binding site residues may be possible. This method of matching the potential extrema is in line with the belief that electrostatic complementarity plays a vital role in some molecular recognition processes.

A key feature of this approach is the accurate calculation of the electrostatic potential, derived from a distributed multipole analysis (DMA) of an \textit{ab initio} charge density of the molecule, which correctly represents the effects of anisotropy in the molecular charge density, eliminating the uncertainties arising from other approximate methods of representing molecular charge distributions.

The method has been applied to the phosphodiesterase (PDE) III substrate adenosine-3'-5'-cyclic monophosphate (cAMP) and a range of inhibitors. Despite the structural variation between cAMP and the inhibitors, a plausible relative binding orientation can be found for each inhibitor, in which they are sufficiently sterically and electrostatically similar to the natural substrate to account for their affinity for PDE III.

This method has also been applied to other systems, including the substrate and inhibitors for the enzyme Glycolate Oxidase, for which information about the binding site structure is available. The method has generally predicted binding orientations obtained by optimizing the electrostatic interactions of each inhibitor within the binding site.

The use of electrostatic extrema in predicting the positions of H and O in hydrogen bonded complexes has been examined for various small molecules modelling biologically important N-H--O=C hydrogen bonding interactions. Optimising the electrostatic interactions of pairs of these molecules in hydrogen bonded geometries has shown that the electrostatic extrema correlate with the positions of hydrogen bond donor and acceptor atoms. Thus, the theoretical basis of the approach is justified, and the method has been validated for use when the binding site structure is unavailable.
# Contents

Chapter 1. Introduction ................................................................. 1

References for Chapter 1 .............................................................. 5

Chapter 2. Intermolecular Forces ............................................... 6

2.1 Introduction ............................................................................. 6
   2.1.1 The Intermolecular Potential ............................................... 6
   2.1.2 Contributions to the Intermolecular Pair Potential ................. 7

2.2 Electrostatic Energy ............................................................... 9

2.3 Induction Energy ..................................................................... 9

2.4 Dispersion Energy .................................................................. 10

2.5 Short Range Contributions ..................................................... 11
   2.5.1 Exchange-Repulsion Energy ............................................... 11
   2.5.2 Charge Transfer Energy ..................................................... 11
   2.5.3 Penetration Effects ........................................................... 12

2.6 Models for Molecular Charge Distributions ......................... 12
   2.6.1 Atomic Point Charge Models ............................................. 13
      2.6.1.1 Mulliken Population Analysis ........................................ 14
      2.6.1.2 Potential-Derived Point Charges ................................... 15
   2.6.2 Other Charge Partition Procedures .................................... 20
      2.6.2.1 Extended Electron Distributions .................................... 20
      2.6.2.2 Hirshfeld’s Procedure ................................................. 21
      2.6.2.3 Bader Partitioning ....................................................... 21
   2.6.3 Central Multipole Model .................................................... 23
   2.6.4 Distributed Multipole Models ............................................ 25
      2.6.4.1 Cumulative Atomic Multipole Moments ....................... 26
      2.6.4.2 Stone’s Distributed Multipole Analysis ....................... 27
      2.6.4.3 Advantages of the DMA .............................................. 28

2.7 Hydrogen Bonding ............................................................... 30
   2.7.1 Van der Waals Dimers ..................................................... 30
   2.7.2 Hydrogen Bond Interaction Energy .................................... 31
   2.7.3 Electrostatic Directionality of Hydrogen Bonds ................... 31
   2.7.4 Modelling Hydrogen Bonds ............................................. 32

2.8 Summary .............................................................................. 33

References for Chapter 2 ............................................................ 34

[continued overleaf]
# Chapter 3. Molecular Similarity

3.1 Introduction .................................................. 37

3.2 Calculation of Molecular Similarity .................................................. 39

3.2.1 Superposition of Atom Positions .................................................. 39

3.2.1.1 The OVID Algorithm .................................................. 39

3.2.1.2 Superposition of Flexible Molecules .................................................. 40

3.2.2 Pattern Matches on Molecular Surfaces .................................................. 41

3.2.2.1 Surface Pattern Matching .................................................. 42

3.2.2.2 ‘Blind’ Surface Searching .................................................. 42

3.2.2.3 The SUPER Algorithm .................................................. 43

3.2.3 Molecular Similarity Indices .................................................. 44

3.2.3.1 The Carbó Index .................................................. 44

3.2.3.2 The Hodgkin Index .................................................. 46

3.2.3.3 Applications of Molecular Similarity Indices .................................................. 47

3.2.4 Superposition of Hydrogen Bonding Partners .................................................. 48

3.2.5 Molecular Electrostatic Potential Similarity .................................................. 49

3.2.6 Molecular Electrostatic Field Comparisons .................................................. 52

3.3 Summary of Similarity Methods .................................................. 53

3.4 Similarity of Electrostatic Extrema .................................................. 53

3.4.1 Region for Comparison .................................................. 55

3.4.2 Experimental Support .................................................. 56

3.4.3 Similarity Using Electrostatic Extrema .................................................. 58

3.5 Methods .................................................. 59

3.5.1 Ab Initio Calculations .................................................. 59

3.5.2 Electrostatic Calculations .................................................. 59

3.5.3 Ligand Superposition and Relative Binding Orientations .................................................. 61

3.5.4 Assessment of Relative Orientations .................................................. 61

3.6 Development of Similarity of Electrostatic Extrema .................................................. 63

References for Chapter 3 .................................................. 65

# Chapter 4. Phosphodiesterase III Inhibitor Modelling

4.1 Introduction .................................................. 68

4.2 Background .................................................. 69

4.2.1 Biological Inhibitor Data .................................................. 70

4.2.2 cAMP Conformational Preference .................................................. 70

4.2.3 Previous Electrostatic Modelling .................................................. 71

[continued overleaf]
Chapter 1. Introduction

What factors determine the ability of molecules to enter, and occupy a binding site within a protein? Shape - steric accessibility - is perhaps the most basic consideration, and this is popularly embodied in the 'lock and key' analogy of enzyme-ligand binding processes. This is a reasonable model, if we assume that the protein and ligand structures remain rigid during the binding process. The analogy of the binding process has been extended from the 'lock and key' to that of the 'hand and glove', to symbolize the ability of the protein receptor to alter its conformation in response to the ligand. However, the situation is further complicated when many 'hands' may wear the same 'glove'. An open binding site could comfortably accommodate ligands of varying sizes which appear structurally unrelated. All three of the molecules shown in figure 1 are able to bind to the enzyme Phosphodiesterase III.

![Figure 1. Three ligands for Phosphodiesterase III, a) adenosine- 3', 5'-cyclic monophosphate (cAMP), b) adenine, c) SK&F93741.](image)

The first structure, adenosine- 3', 5'-cyclic monophosphate (cAMP) is the enzyme's natural substrate, whilst adenine and SK&F93741 both act as phosphodiesterase inhibitors. From a purely structural comparison, we might reasonably assume that the adenine inhibitor will occupy the same region of the phosphodiesterase binding site as the adenine moiety of cAMP. However, the binding orientation of SK&F93741 with respect to cAMP is considerably less obvious from a structural point of view, and yet all three molecules are capable of occupying the same binding site. Hence these molecules must have some features in common which are important to the ligand binding process;
that is to say they must be *similar* in some way. As we have seen from the examples above, the definition of similarity depends on the properties which are being compared. Structural similarity alone cannot adequately explain the ability of cAMP, adenine and SK&F93741 to bind to Phosphodiesterase III.

Ligand-receptor binding depends on the interplay of many different chemical and physical factors. Any one, or combination, of these factors could be used as a measure of similarity. For systems involving polar ligands, or systems in which hydrogen bond interactions are important, one critical factor during ligand binding processes is the electrostatic energy. Even when an array of polar binding site residues and complementary ligand functional groups are identified, it is still necessary to orient the ligand in three dimensions, with respect to the binding site, to maximize the binding interaction.

Given the importance of the electrostatic contribution to determining relative binding orientations, it is necessary to consider the theoretical treatment of electrostatic interactions. The work presented in this dissertation will begin with a review of the principal contributions to intermolecular forces and their physical origins. The electrostatic component, and the different levels of approximation used to model molecular charge distributions, will be examined in detail. The implications of these approximations for hydrogen bonding, a critically important stabilizing interaction in many chemical systems, will be considered, highlighting the need for accurate charge density models. Traditionally, atom-centred point charge models have been used to model molecular charge distributions. However this fails to represent non-spherical features of charge density, such as lone pair and π electron density, important in hydrogen bond interactions. A key feature of the work presented here is the accurate calculation of molecular electrostatic properties, derived from a distributed multipole analysis (DMA) of an *ab initio* molecular charge density. The DMA correctly represents the effects of anisotropy in the molecular charge density, and so eliminates the uncertainties which arise from other more approximate methods of representing molecular charge distributions.

In this work, a method for determining the relative binding orientation of different ligands within an unknown binding site is developed. By overlaying the maxima and
minima in the electrostatic potential outside ligands, it is possible to match regions where strong electrostatic interactions, including hydrogen bonds, with binding site residues may be possible. This method of matching the potential extrema is in line with the belief that electrostatic complementarity plays a vital role in some molecular recognition processes. A selection of methods currently used in the calculation of molecular similarity will be reviewed, and the approach used in this work is presented in this context. The existing methods broadly fall into three categories; those which attempt to match structural features between molecules, those which measure similarity in terms of the properties of molecular surfaces, and those which combine information about the chemical structure and properties of the molecules. The approach developed here, for determining the position of maximum electrostatic similarity between two ligands common to a receptor site, is introduced based on an accurate DMA model for the molecular charge distribution. The relative orientations are determined by matching electrostatic potential extrema between molecules. The electrostatic extrema are determined at fixed distances beyond the molecular van der Waals surface, and represent points where strong binding interactions with a receptor are possible, and so consider both the steric and chemical aspects of similarity.

The method has been applied to the phosphodiesterase (PDE) III substrate adenosine-3', 5'-cyclic monophosphate (cAMP) and a range of inhibitors. Despite the structural variation between cAMP and the inhibitors, it will be shown that a plausible relative binding orientation can be found for each inhibitor, in which they are sufficiently sterically and electrostatically similar to the natural substrate to account for their affinity for PDE III. The approach has been developed further by application to the substrate and inhibitors of the enzyme Glycolate Oxidase. For this system, information about the binding site structure is available, hence it has been possible to compare the relative binding orientations of the inhibitors produced by matching electrostatic extrema, with those found by optimizing the electrostatic interactions of each inhibitor within the binding site. This method has also been applied to other systems, including ligands for the adenosine A<sub>1</sub> receptor, and a selection of lactam compounds.

To investigate the relationship between the positions of minima and maxima in the electrostatic potential and hydrogen bond donor and acceptor groups, calculations
have been carried out using small test molecules and amino acid sidechains containing groups which can form the biologically important N-H·O=C hydrogen bonding interactions, and other hydrogen bond interactions such as N-H·N, C-H·O. Their ability to predict the positions of H and O in hydrogen bonded complexes found by optimising the electrostatic interactions of pairs of such molecules constrained by the molecular shapes has been assessed. Optimising the electrostatic interactions of pairs of these molecules in hydrogen bonded geometries demonstrates that the electrostatic extrema correlate with the positions of hydrogen bond donor and acceptor atoms. Thus, the theoretical basis of the approach is justified, and the method has been validated for use when the binding site structure is unavailable.
References for Chapter 1.


Chapter 2. Intermolecular Forces

In this chapter, we shall look at the principal contributions to intermolecular forces, their physical origins, and the theoretical treatment of these forces. The electrostatic component and the different levels of approximation used to model molecular charge distributions are emphasized. The implications of these approximations for hydrogen bonding, a critically important stabilizing interaction in many chemical systems, are examined, highlighting the need for accurate charge density models.

2.1 Introduction

The entire range of chemical and physical behaviour observed for associated molecular species is controlled by the forces between molecules. Many of the forces which govern the behaviour of molecules in association arise from interactions between molecular charge distributions. Phenomena from crystal structure packing arrangements to protein folding and molecular recognition are essentially due to the interaction of one molecule with another. To be able to understand these forces is to be able to predict their behaviour and therefore the behaviour of an entire range of molecular systems. We need to start from a knowledge of the physical origins of intermolecular forces and an accurate description of their principal components.

2.1.1 The Intermolecular Potential

To predict the properties of molecular solids, liquids and gases, an accurate expression for the forces between associated molecules and their total interaction energy, $U$, is needed. The total energy of a group of associated molecules can be expressed as

$$U = \sum_i U_i + \sum_{i<j} U_{ij} + \sum_{i<j<k} U_{ijk} + \ldots .$$

(2.1)

The first term, $U_i$, is the energy of an isolated molecule $i$, while $U_{ij}$, the pair potential, describes the energy of $i$ and a second molecule $j$, in isolation. $U_{ijk}$ is a three body term which, together with the higher-body terms, represents the non-additivity of the pairwise interactions. The pair potential and all higher-body terms depend on the separations and relative orientations of molecules $i, j, k, \ldots$. The precise energy of $N$ interacting molecules would involve the sum over the true pair potential, plus the sum
over the three-body potential, plus four-body terms, up to N-body terms. However, it is usual to assume that this equation converges rapidly because the effects of higher-body terms will be negligible, and that the interaction potential is primarily pairwise additive. For liquid and solid systems we use the pairwise additive approximation, in which the total energy of a system of associated molecules is taken as the sum of the pair potentials for all pairs of atoms or molecules. This gives an effective intermolecular pair potential, used to calculate the interaction energy of a system of $N$ molecules, as a function of their intermolecular separation $R$, and relative orientation $\Omega$

$$U = \sum_{i<j} U_{ij}(R_{ij}, \Omega_{ij}) . \quad (2.2)$$

### 2.1.2 Contributions to the Intermolecular Pair Potential

The primary forces between atoms and molecules are electrical in origin. At long range, where the intermolecular separation is much greater than the molecular dimensions, the interaction energy, $U_{AB}$, between two molecules $A$ and $B$ is much smaller than the total energy of the two isolated molecules, $E_A^0$ and $E_B^0$, at infinite separation. Under these conditions, where significant overlap of the charge densities of two interacting molecules does not occur and consequently the effects of electron exchange can be ignored, Rayleigh-Schrödinger perturbation theory can be applied. Through the use of a perturbation expansion the interaction energy is divided into physically meaningful terms.

In perturbation theory the difference between the Hamiltonians $\hat{H}_A$ and $\hat{H}_B$ of the unperturbed monomers $A$ and $B$, and the Hamiltonian of the perturbed system is the perturbation $\hat{\mathcal{H}}'$. Since the origin of the intermolecular interaction is electrical, this perturbation can be expressed as the Coulombic interaction between the undistorted charge distributions of the two molecules, and can be written (in atomic units) as

$$\sum_{ij} \frac{e_i^A e_j^B}{r_{ij}} \quad (2.3)$$

where $r_{ij}$ is the separation between electron $e_i$ of molecule $A$ and electron $e_j$ of molecule $B$. The perturbation can be applied incrementally to provide a gradual change from the unperturbed system to the perturbed state. The interaction energy $U_{AB}$ can be expanded
in the series

\[ U_{AB} = U_{AB}^0 + U_{AB}^1 + U_{AB}^2 + \ldots + U_{AB}^n. \] (2.4)

The long range pair potential is then expressed as integrals over \( \psi_A^0 \) and \( \psi_B^0 \) (the ground state wavefunctions) and \( \psi_A^n \) and \( \psi_B^n \) (the excited state wavefunctions) of the isolated molecules. The energy of the zeroth-order state (the ground state) is simply the energy of the isolated molecules

\[ U_{AB}^0 = \langle \psi_A^0 | H_A | \psi_A^0 \rangle + \langle \psi_B^0 | H_B | \psi_B^0 \rangle = E_A^0 + E_B^0 \] (2.5)

Higher powers in the expansion give successive corrections \( U_{AB}^n \) to the zeroth-order energy resulting from the \( n \)-orders of perturbation theory. The higher terms represent the different physical effects which contribute to the intermolecular interaction energy. The first-order perturbation represents the change in energy due to the electrostatic interaction between the permanent multipole moments of the molecules.

\[ U_{AB}^1 = \langle \psi_A^0 \psi_B^0 | \mathcal{H} | \psi_A^0 \psi_B^0 \rangle \] (2.6)

This term can be represented as \( U_{AB}^{\text{elec}} \), the electrostatic interaction energy between molecule \( A \) and \( B \). The second-order perturbation \( U_{AB}^2 \) represents two physical effects: the induction interaction, where molecule \( A \) polarizes molecule \( B \),

\[ U_{AB}^{\text{Ind}} = - \sum_{n_B \neq 0} \frac{|\langle \psi_A^0 \psi_B^n | \mathcal{H} | \psi_A^0 \psi_B^n \rangle|^2}{E_B^n - E_B^0} \] (2.7)

and conversely molecule \( B \) polarizes molecule \( A \),

\[ U_{AB}^{\text{Ind}} = - \sum_{n_A \neq 0} \frac{|\langle \psi_A^0 \psi_B^n | \mathcal{H} | \psi_A^0 \psi_B^n \rangle|^2}{E_A^n - E_A^0} \] (2.8)

and the dispersion interaction of \( A \) and \( B \)

\[ U_{AB}^{\text{Disp}} = - \sum_{n_B \neq 0} \frac{|\langle \psi_A^0 \psi_B^n | \mathcal{H} | \psi_A^0 \psi_B^n \rangle|^2}{(E_A^n - E_A^0) + (E_B^n - E_B^0)} \] (2.9)
The terms in the perturbation expansion for the interaction between $A$ and $B$ to second order are then

$$U = U_{estat}^{AB} + U_{ind}^A + U_{ind}^B + U_{disp}^{AB}. $$

(2.10)

The physical origins and theoretical treatment of each of these terms will be discussed briefly in the following sections.

2.2 Electrostatic Energy

The first-order term (equation (2.6)) in the perturbation expansion, corresponds to the electrostatic energy of interaction between two charge distributions. This interaction may be attractive or repulsive. It can be calculated exactly by integration over the ab initio charge distributions. Because the charge distributions are not distorted in the first-order term, the electrostatic contribution to the potential is strictly pairwise additive.

$$U_{estat} = \int_{\text{all space}} \rho^A(r_1) \rho^B(r_2) \frac{d_1 d_2}{|r_1 - r_2|} d_1 d_2 d_2$$

(2.11)

The electrostatic contribution is the result of Coulombic interactions between molecules whose charge distributions are not spherically symmetrical. This interaction is dependent on the separation and relative orientation of the two molecules. Indeed, the electrostatic energy usually dominates the orientation dependence of the intermolecular potential for molecules which contain heteroatoms.

2.3 Induction Energy

Under the influence of electric fields that result from the presence of neighbouring non-spherical molecules, some modification of charge density in a molecule will occur, and this effect gives rise to one type of interaction, the induction energy.

The field arising from a molecule $A$ will distort the charge distribution in a neighbouring molecule $B$ such that molecule $A$ polarizes molecule $B$. Molecule $B$ is polarized by a total vector sum of the electric fields from all neighbouring molecules, so induction is highly non-pairwise additive. The induction energy between ground state molecules is always attractive. It is difficult to model and for most purposes is usually
ignored, though this will depend greatly on the molecules being modelled.

2.4 Dispersion Energy

For spherical electrically neutral molecules, that consequently have no permanent dipole or higher multipole moments, the electrostatic and induction contributions to the intermolecular interaction energy are absent. The only attractive interaction between such molecules at long range is the dispersion interaction, first identified by London in the 1930s.

While the average dipole moment of a spherically neutral atom is zero, at any given time the electrons may not be distributed symmetrically around the nucleus, creating an instantaneous temporary dipole moment on the atom. This temporary dipole moment can induce the formation of a dipole moment on a neighboring atom by polarizing the neighboring atom's electron density, producing an attractive interaction. Hence, any fluctuations in the electron density distribution of the first atom are followed in the second atom. This instantaneous correlation between the motions of the electrons in different atoms is the origin of the dispersion force. This effect is purely quantum mechanical, resulting from the zero-point energy of the electrons, and cannot be described in classical terms.

An expression for the leading term in the dispersion energy is

\[ U_{\text{disp}} = -\frac{3\alpha^2 E_i}{4(4\pi\varepsilon_0)^2 R^6} \approx \frac{C_6}{R^6} \]  (2.12)

where \( \alpha \) is the polarizability and \( E_i \) is the ionization energy of the molecule, and the \( C_6 \) coefficient is approximate. This leading term, resulting from the interaction of an instantaneous dipole and an induced dipole, is the dominant contribution at long range and diminishes as \( R^{-6} \). Contributions from higher-order instantaneous multipoles diminish more rapidly with distance. Although at large separations the dipole-dipole term dominates, interactions between the higher-order instantaneous multipole moments make significant contributions to the total dispersion energy.

In the presence of a third molecule the dispersion interaction is no longer pairwise additive, and a tripole-dipole correction term for the dispersive interaction of three molecules \( A, B \) and \( C \), identified by Axilrod and Teller, must be introduced.
\[ U_{_{AB}}^{ABC} = \frac{C_{AB}^{6}}{R_{AB}^{6}} + \frac{C_{BC}^{6}}{R_{BC}^{6}} + \frac{C_{AC}^{6}}{R_{AC}^{6}} + \nu_{ABC} \frac{(3\cos\theta_{A} \cos\theta_{B} \cos\theta_{C} + 1)}{(R_{AB} R_{BC} R_{AC})^{3}}, \]  

(2.13)

where \( R_{AB} \) is the separation of \( A \) and \( B \), \( \theta_{A} \) is the angle \( B-A-C \).

2.5 Short Range Contributions

Contributions to the intermolecular potential considered to be short range result from situations where non-negligible overlap of charge distributions occurs. In such situations Rayleigh-Schrödinger perturbation theory is no longer applicable since exchange of electrons between molecules can occur.

2.5.1 Exchange-Repulsion Energy

Exchange-repulsion is a combination of two effects arising from the interaction of the overlapping charge densities, and the Pauli exclusion principle. The prevailing interaction at short range is a repulsive force. In accordance with the Pauli exclusion principle, electron density in the charge overlap region decreases, increasing repulsion between the incompletely shielded nuclei. Electrostatic repulsion also occurs between the overlapping charge distributions. This repulsion effectively defines the contact distance between molecules, and is approximately additive. For practical applications it is usually modelled by an exponential function of the separation \( R_{AB} \).

The exchange energy is due to the fact that electrons are in motion over both molecules at short range. The possibility of exchange leads to a lowering of energy. The exchange energy is an attractive, pairwise additive term. Although smaller in magnitude than the repulsion term and opposite in sign, it is observed to vary numerically in the same manner as the repulsion term. In many calculations the two terms are usually considered together, where the repulsive term dominates.

2.5.2 Charge Transfer Energy

Charge transfer from one molecule to another will also take place when there is significant molecular charge overlap. This can be regarded as the extreme of polarization, where the distortion of a first molecule's charge density leads to a transfer of electrons to the second molecule. This is a highly non-additive effect whose
importance can be over-emphasised as a result of basis set superposition error\(^3\). Consequently it is not treated in model potentials for weak interactions between closed shell molecules.\(^4\)

### 2.5.3 Penetration Effects

The repulsion experienced by an electron at a point outside a spherical charge distribution is equivalent to that generated by a point charge of equal magnitude at the centre of the distribution. An electron which penetrates within the space occupied by an atom core, and is a short distance from the nucleus, will experience a reduced repulsion from the inner electrons compared to an electron outside the core. This effect produces an error in the multipole expansion at short range (The consequences of this error will be discussed in section 2.6.3).

### 2.6 Models for Molecular Charge Distributions

In order to calculate correctly the electrostatic interaction energy between two molecules, the charge distributions must be represented as accurately as possible. Any charge distribution within a molecule can be described by a sum of multipoles: charge, dipole, quadrupole etc. The central multipole model (which will be described in section 2.6.3) was frequently used in early models of molecular charge distributions. The electrostatic potential resulting from a charge distribution decreases more rapidly with distance for higher multipoles, so that at very large separations the potential is dominated by the lowest multipole moments. While a complete representation of a charge distribution would require the inclusion of all multipoles, the lowest non-zero moment is usually treated as the most important. At long range the nucleus of an atom and its associated electron density can be viewed as a single point charge. This is the basis of the atomic point charge model for molecular charge distributions. A major justification of the atomic point charge model lies in its relatively simplicity. This model for molecular charge distributions is so enduring, that the question of how to represent the electrostatic interaction between two molecular charge distributions has often become a question of how to assign partial atomic charges to model the molecular charge distributions. Some different methods for representing the charge distribution are reviewed in the following sections.
2.6.1 Atomic Point Charge Models

This is the simplest approach for representing the charge distribution, and it is commonly used for organic molecules. The electron density and nuclear charges are represented by a set of point charges at the centres of the atoms in a molecule. This description of the molecular charge distribution has often been said to have an ‘immediate intuitive appeal’ because the molecular property (i.e., the charge) is centred on the atoms which we consider to compose the molecule. However, a point-charge model gives poor results for a number of reasons.

The molecular charge distribution is represented as a set of spherical point charges. Such an isotropic model of intermolecular interactions ignores the rearrangement of electrons resulting in the formation of chemical bonds. The overlap of orbitals of unequal size in a molecule (for example a large $p$-orbital and a small $s$-orbital) creates an accumulation of charge in the overlap region, which will lie closer to the smaller atom. The assignment of charge density to one or other of the nuclei in this case will not accurately represent the nature of the bonding interaction. The formation of a chemical bond involves the distortion of atomic charge density. This model does not account for non-spherical features of charge density such as lone pairs and $\pi$-electron density. In such bonding features involving hybridized atomic orbitals the centre of nuclear charge no longer coincides with the centre of electronic charge. As a result, additional point charge sites are needed to describe adequately the electrostatic potential due to such anisotropic features of charge distributions. Atomic point charges give a spherically averaged description of the electrostatic potential around a nucleus.

The partial atomic charges used in point-charge models can be obtained from \textit{ab initio} wavefunctions by a number of methods, including Mulliken population analysis\textsuperscript{5} and fitting to the molecular electrostatic potential.\textsuperscript{6} These two methods are outlined in the following sections.
2.6.1.1 Mulliken Population Analysis

A Mulliken analysis\(^5\) of the wavefunction begins from the expansion of normalized molecular orbitals \(\Phi_i\) as a linear combination of normalized atomic orbitals \(\chi_i\)
\[
\Phi_i = c_1 \chi_1 + c_2 \chi_2 + c_3 \chi_3 + \cdots .
\]
(2.14)

The probability density for each electron in the molecular orbital can be expressed as
\[
|\Phi_i|^2 = c_1^2 \chi_1^2 + c_2^2 \chi_2^2 + \cdots + 2 c_1 c_2 \chi_1 \chi_2 + 2 c_1 c_3 \chi_1 \chi_3 + \cdots .
\]
(2.15)

Integrating over all space gives the probability density expressed as net populations (\(c_j^2\)) and overlap populations (\(2 c_j c_k S_{jk}\))
\[
1 = c_1^2 + c_2^2 + c_3^2 + \cdots + 2 c_1 c_2 S_{12} + 2 c_1 c_3 S_{13} + \cdots ,
\]
(2.16)

where \(S\) is the overlap integral
\[
S_{jk} = \int \chi_j \chi_k \, d\tau .
\]
(2.17)

In order to assign charge density to the atomic orbitals, and consequently to the atoms, each electron in the molecular orbital \(\Phi_i\) contributes \(c_j^2\) to the population of \(\chi_j\) plus an artificially divided sum of contributions to overlap populations to give a gross atomic population
\[
c_j^2 + \frac{1}{2} \sum_{k \neq j} 2 c_j c_k S_{jk} .
\]
(2.18)

The net charge on an atom \(A\) is then given by
\[
q_A = Z_A - \sum_i n_i \sum_j^A \left[ c_j^2 + \sum_k c_j c_k S_{jk} \right]
\]
(2.19)

where \(Z_A\) is the nuclear charge, \(n_i\) is the number of electrons in the molecular orbital \(\Phi_i\), \(j\) is the number of atomic orbitals on atom \(A\) and \(k\) is the number of atomic orbitals which are not on atom \(A\).

This has been one of the most widely used method for deriving partial atomic charges, and it has proven useful in a whole range of studies. Nevertheless, there are a number of well known drawbacks to the Mulliken population analysis. This method
contains the assumption that the overlap population density $2c_{ij}^2S_{jk}$ can be divided equally between contributing atomic orbitals $\chi_j$ and $\chi_k$, even if one atom is more electronegative than the other. This is inconsistent with the formation of chemical bonds, which result from the rearrangement of atomic charge density. Atomic orbitals, such as $p$, $d$ and $f$ orbitals, can extend away from their nuclei as far as neighbouring nuclei, so that contributions from this part of charge density should not be assigned to the original nucleus. Unsurprisingly, polar bonds and lone pair electrons, both anisotropic features of charge distributions, are not well represented by a Mulliken analysis. Lone pair density is averaged over the spherical atomic point charge, and contributions from polarized atomic charge densities to the molecular charge density are divided equally between atoms. There is no theoretical justification, nor is any claimed, for the allocation of charge density in the atomic overlap populations. This is part of the simplicity of the population analysis method and it has certainly been useful in identifying trends within groups of molecules. However, the overlap terms can produce ‘apparently unrealistic results’.

Another drawback is that the magnitude of the Mulliken charges is basis set dependent. Even the earliest studies found that the variations in charge densities calculated from Mulliken charges with different basis sets could be as large as those involved in the formation of chemical bonds. One study compared net atomic charges derived from a population analysis of three different basis sets (STO-3G and 6-31G ab initio and CNDO/2 semi empirical wavefunctions) for a series of amides, carboxylic acids and peptides. The charges were found to be ‘considerably different’ in all cases. The larger basis set produced net atomic charges which were almost twice the magnitude of those derived from the STO-3G and CNDO/2 calculations. In some cases the three basis sets produced equivalent charges which differed in sign. Despite such results, the conceptual simplicity and theoretical usefulness of the atomic point-charge model fuelled its development based on other, more physically justified means.

2.6.1.2 Potential-Derived Point Charges
One of the frequently cited failings of the Mulliken population analysis is that the net atomic charge is not a rigorously defined quantum mechanical property, due to the nature of the electron distribution around the nuclei in a molecule. A more promising
route for the derivation of representative point charges is to relate them to more rigorously defined molecular properties, such as the molecular electrostatic potential.

Potential-derived charges are the result of defining atom-centred point charges within the molecule to fit the electrostatic potential outside the molecule, since it is the electrostatic properties in this region which often determine any intermolecular interactions. The charges are fitted in order to reproduce the electrostatic potential calculated at a set of points outside the molecule's van der Waals surface. Hence, there are two stages in the development of a potential-derived point charge model of a molecule: to calculate the electrostatic potential at points outside of the molecule, and to calculate point charges to fit to this electrostatic potential data by means of a least squares fitting procedure.

The electrostatic interaction energy $U_{\text{stat}}$ between a molecular charge distribution, and a unit positive charge $q$ is

$$U_{\text{stat}}(i) = qV(i) \quad (2.20)$$

where $V(i)$, the molecular electrostatic potential, is the net potential experienced by a unit positive charge at a point $i$ at a distance $r_i$ outside the molecule. $V(i)$ and $U_{\text{stat}}$ have the same magnitude when $q$ is a unit positive charge. $V(i)$ is defined quantum mechanically as the interaction between a unit positive charge and the unperturbed molecular charge distribution

$$V_{i}^{\text{QM}} = \sum_{A=1}^{N_A} \frac{Z_A}{|r_i - R_A|} - \int \frac{\rho(r')}{|r_i - r'|} dr' \quad (2.21)$$

where $Z_A$ is the nuclear charge on atom $A$ located at $R_A$, $\rho$ is the electron density and $N_A$ is the number of atoms in the molecule. Unlike Mulliken charges, the electrostatic potential is a rigorously defined quantum mechanical property. It can be calculated directly from the molecular wavefunction using the electrostatic potential operator.

By using atom-centred point charges, the electrostatic potential at the point $i$ can be written in terms of the classical Coulombic potential as

$$V_i^{\text{Coul}}(q_1, q_2, ..., q_{N_A}) = \sum_{j=1}^{N_A} \frac{q_j}{r_{ij}} \quad (2.22)$$
where \( q_j \) is the net charge on atom \( j \) and \( r_{ij} \) is the separation between the point \( i \) and atom \( j \). Thus, the least-squares fit of the charges to the quantum mechanical electrostatic potential is performed by minimizing the function

\[
Y (q_1, q_2, \ldots, q_{N_A}) = \sum_{i=1}^{N_p} \left[ V_i^{QM} - V_i^{Coul}(q_1, q_2, \ldots, q_{N_A}) \right]^2
\]  

(2.23)

where \( N_p \) is the total number of points at which the electrostatic potential is being fitted. It is implicit that \( N_p > N_A \) for a good fit.

The least-squares fitting method introduced by Momany\(^9\) required an initial test set of atomic point charges which converged to their final values through an iterative procedure. Cox and Williams\(^10\) produced minimized net atomic charges by partial differentiation of the function \( Y \). An alternative method involves the use of Lagrange multipliers\(^11\) to solve for the charges directly, with certain constraints added to the function \( Y \). One necessary condition is that the net atomic charges must sum to the total molecular charge. If the net atomic charges produced are physically reasonable, then they should also be able to reproduce other rigorously defined molecular properties calculated quantum mechanically such as molecular dipole moments.\(^13\) However, the least-squares fitting of the atomic charges is usually carried out with the single constraint that they reproduce the total molecular charge.

The number and location of the points outside the molecular van der Waals surface which are used to fit the model electrostatic potential \( V_i^{Coul} \) to the quantum mechanical electrostatic potential \( V_i^{QM} \) are chosen on a three dimensional grid,\(^9\)\(^10\)\(^14\) or on a surface.\(^6\)\(^12\) The grid typically extends between the van der Waals surface and an outer limit of between 1.0 and 3.0 \( \text{Å} \) beyond the surface. The points are chosen around each atom to give between 100-300 points per molecule. Points chosen significantly beyond 3.0 \( \text{Å} \) sample the region where the electrostatic potential becomes increasingly weaker. Inclusion of such points improves the RMS fit of the model electrostatic potential to the quantum mechanical electrostatic potential (as the number of these points increasingly dominates the comparison) without affecting the calculated net atomic charges.

Surface point-selection procedures aim to produce an isotropic distribution of points on the surface of a sphere surrounding the atoms, to ensure the surface points are

17
rotationally independent. Points calculated at the intersections of regular lines of latitude and longitude on a spherical surface produce increased point densities in the polar regions. The Connolly surface algorithm produces an approximately equally-spaced point distribution on the surface of a sphere. Kollman and co-workers use this algorithm on fused spherical surfaces at distances which are scaled multiples of the atomic van der Waals radii. The multiple surfaces reduce the dependence of the charges on the particular surface sampled. However, a recent careful analysis of the Connolly algorithm by Spackman has revealed that, for any requested value \( N_p \), the array of surface points produced has a point density less than or equal to \( N_p \), and that the point distribution contains significant anisotropies. These anisotropies are only eliminated at very high values of \( N_p \). Spackman suggested the use of a geodesic point selection scheme (consisting of tessellations of the triangular faces of the icosahedron with regular triangular or hexagonal grids) to produces highly isotropic surface point arrays. Potential-derived charges calculated for methanol displayed a lower rotational dependence for the geodesic surface when compared to the Connolly surface, and no translational dependence.

The magnitudes of charges derived from fitting to the different surfaces were similar, except for ‘buried’ atoms (those atoms furthest from the molecular surface). For these ‘buried’ atoms, other ‘surface’ atoms (closer to the points \( i \) at which the molecular electrostatic potential is determined) make more significant contributions to the electrostatic potential at those points. Consequently, the values of potential-derived charges for ‘buried’ atoms are often imprecise (they may fluctuate over a wide range) and inaccurate (producing chemically unreasonable values), relative to small changes in the distribution and number of points \( i \).

Recently some doubt has arisen over whether it is desirable, or even correct, to fit charges derived from the molecular electrostatic potential to all atoms in a molecule, if ‘buried’ atoms do not make significant contributions to the electrostatic potential. Stouch and Williams examined the effect of the fitting procedure on the resultant potential-derived charges. They had noted previously that charges on neighbouring atoms were highly correlated and that changes in atomic charge with molecular conformation were reflected in changes in the neighbouring atomic charges. This correlation indicates that ‘...the number of variables that can be fit (the charges) in a
As the degree of correlation increases more of the variance in the data can be explained by a smaller number of variables. It was found that fixing the values of 'buried' atoms to 'chemically reasonable' values had little effect on the quality of the overall fit, while (in some cases) helping to decrease the range of variance of other non-fixed charges. Francl et al\textsuperscript{22} attempted to determine systematically which charges can be fixed and which cannot during fitting (using points \(i\) generated at random within a rectangular box surrounding the molecule). The larger the molecule (and greater 'buried' volume) the greater the dependence of the fitting procedure on fewer atomic charges. Increasing the point density did not have a significant effect beyond a threshold value. Attempts to 'tighten' the electrostatic surface, by bringing the points at which the potential is calculated closer to the atomic charge sites, slightly increased the number of statistically significant charge sites. Bayly et al\textsuperscript{19} employed a restrained electrostatic potential (RESP) model, in which certain charges were restrained to 'target' values in order to conserve the first two non-zero molecular multipole moments. This had little affect on the statistically well-determined values, but restricted the 'buried' atomic charges to assume more chemically reasonable and intuitive values.

Cox and Williams,\textsuperscript{10} following Momany's work,\textsuperscript{9} found that the magnitudes of potential-derived charges do not suffer from the same dependence on basis set which was observed for Mulliken charges.\textsuperscript{8} Chirlian and Francl\textsuperscript{11} have demonstrated a good correlation among atomic charges calculated using 3-21G, 6-31G, 6-31G* and 6-31G** basis sets, with only slight differences seen between the basis sets. Calculations with the larger basis sets without polarization functions only gave limited improvements while the STO-3G minimal basis set gave lower correlation. The 3-21G basis set was found to balance computational time saving compared to the larger basis sets with greater accuracy compared to the minimal basis set.

Only slight differences are reported by Woods et al\textsuperscript{13} between net atomic charges fitted with and without the dipole constraint. This is consistent with the results of Cox and Williams\textsuperscript{10} that molecular dipole moments derived from potential-fitted charges (fitted, of course, without use of a dipole constraint) gave close agreement with SCF-calculated dipole moments, whereas dipole moments calculated with Mulliken charges have been found to be 'notably erratic'.\textsuperscript{11} The most significant effect of the
The potential-derived point charge model has been extended to include charge sites at locations other than nuclear positions, with the intention of modelling anisotropic features of charge distributions. Singh and Kollman\(^6\) studied the inclusion of ‘extra’ lone pair charge sites in water. They found that an improved fit was noted for a 5-point model of water (3 atom-centred charges and 2 lone-pair site charges) over a 3-point charge model (atom-centred charges only). However, to derive a 5-point model that simultaneously described the \(\text{H}_2\text{O}\) dipole and quadrupole moments required the lone pair charges to be inverted from the traditional tetrahedral directions. This placed the lone pair charges on the *same side* of the oxygen as the hydrogen atoms. This inversion was not seen for \(\text{CH}_3\text{OH}\), \((\text{CH}_3)_2\text{O}\) or \(\text{H}_2\text{CO}\). While the inclusion of the lone pair charge sites improved the fitting of the model to the electrostatic potential, the authors state that there is ‘no compelling qualitative reason’ to include lone pairs in simulations including N and O lone pairs.

### 2.6.2 Other Charge Partition Procedures

#### 2.6.2.1 Extended Electron Distributions

Many schemes for partitioning the charge distribution within a molecule exist. Vinter\(^23\) has introduced a new methodology which employs ‘extended electron distributions’ (XEDs) to model molecular charge density. They are designed to simulate the way molecules are expected to polarize as they approach each other. Because the XEDs are only designed to simulate interaction configurations of polar groups, they are not expected to reproduce ground-state properties such as dipole moment.

The different hybridization states of an atom type (*eg.* \(\text{sp}^2\) or \(\text{sp}^3\) O) cannot be represented by a single common orbital arrangement, so a ‘minimum valence basis set’ of six octahedrally distributed ‘orbital points’ (or XEDs) is used as an initial pattern for each atom. These XEDs can be extended, contracted or eliminated according to the hybridization state of each atom. Thus, each atom type is assigned a XED atom type. The XED lengths and XED charges for each type are parameterized using one experimentally well-characterized complex. Integer positive charges are assigned to the nuclei, and negative partial charges are calculated for each atom and distributed equally.
amongst the XEDs. This creates neutral atoms, so that the nuclear charge is the absolute sum of the XEDs for any ensemble. Electronegativity trends and polarization effects are included by adding partial charge to each XED pattern, leaving the nuclear charge unchanged. Atoms with no XEDs (non-polar atoms whose Pauling electronegativity is less than or equal to carbon) are assigned atom-centred charges. Although the method still includes atom-centred point charges, the inclusion of XED sites is intended to reduce their weakness due to isotropy, while accounting for electronegativity and polarization effects.

The XEDs were tested for a range of example complexes and compared favourably with atom-centred point charges and AM1-derived natural atomic orbitals. Unlike atom-centred point charges, the XEDs reproduced the T-shaped benzene dimer found using high quality ab initio DMA calculations, and the lone pair directionality of acetone-water complexes. None of the results using XEDs were grossly in error, whereas the atom-centred point charges were seen to give the incorrect geometries for some intermolecular interactions.

2.6.2.2 Hirshfeld’s Procedure

In Hirshfeld’s procedure a hypothetical ‘promolecule’ is constructed by superimposing the spherically symmetrized charge distributions of the isolated atoms (a set of spherically symmetrical ‘proatoms’). The charge density $\rho$ of the real molecule at a point in space is then divided between the constituent atoms in the same proportion as the charge density from the hypothetical proatoms at that point. The atomic populations are obtained by integrating these charge density distributions over all space.

2.6.2.3 Bader Partitioning

Bader’s Theory of Atoms in Molecules is a topological approach. A molecule is divided into a number sub-systems separated by boundary surfaces which satisfy the condition that ‘The surface bounding the sub-system shall not be crossed by any gradient vectors of charge density’. The unit vectors $\mathbf{n}$ are normal to the proposed surface. $\nabla \rho$ is the gradient vector of $\rho$. The condition is expressed mathematically as

$$\nabla \rho \cdot \mathbf{n} = 0$$

(2.24)
where $\nabla \rho$ has no component through the surface charge density contours. The surface which satisfies this condition is a ‘zero-flux surface’. Where $\nabla \rho = 0$, there are ‘critical points’ which are connected by ‘gradient paths’. These gradient paths partition the molecular charge distribution into sub-systems bounded by zero-flux surfaces. Do the resulting sub-systems correspond to atoms? Maxima in $\rho$ exist only at nuclei, so all gradient paths in the vicinity of a nucleus originate (or conversely, terminate) at the nucleus. Therefore the system is partitioned into a number of distinct and non-overlapping ‘basins’, which each contain a single nucleus. The nucleus and its associated basin correspond to an atom. The presence of a ‘saddle point’ indicates electronic charge is accumulated between the nuclei. This is found between every pair of neighbouring atoms. ‘Atomic interaction lines’ link the nuclei, along which $\rho$ is a maximum with respect to any neighbouring line. In a bound system an atomic interaction line corresponds to a chemical bond and is called a ‘bond path’. Thus, the topology of $\rho$ allows for definition of atoms and the network of lines that shows how the atoms are linked together.

In the Hirshfeld and Bader approaches the atoms are non-spherical and the charges are the first term in a Taylor expansion that includes dipole, quadrupole and higher terms. Wiberg and Rablen\textsuperscript{29} examined several procedures for producing atomic charges from \textit{ab initio} molecular orbital calculations. These included charges derived from population analyses, charge density distributions (Hirshfeld’s procedure and Bader partitioning), and the molecular electrostatic potential. The strategies for deriving atomic charges were compared according to four criteria; their basis set dependence, their ability to reproduce electronegativity trends in a molecular series, molecular dipole moments, and molecular electrostatic potentials. Most methods could not satisfactorily reproduce the dipole moment and the molecular electrostatic potentials for the set of chosen molecules. The charges derived from fitting to the molecular electrostatic potential did not reproduce electronegativity trends. This may be because the charge distribution around each nucleus is considered as spherically symmetrical. The authors conclude that molecular charge distributions are ‘...much too anisotropic to be modelled by any single set of atom-centred charges...’ at separations other than long range, without including ‘...at least atomic dipole terms, and possibly higher terms as well.’

Potential-derived point charges give the most accurate representation of the electrostatic
potential possible within the limitations and assumptions of the point-charge method. However, atom-centred charges will always suffer from attempting to model effects which are more accurately accounted for by the inclusion of higher multipoles.

**2.6.3 Central Multipole Model**

When evaluating the electrostatic interaction between two molecules $A$ and $B$, rather than attempting to describe the molecular charge distribution using a number of contributing atomic point-monopoles, the use of higher multipoles can overcome the limitations of the spherically-averaged description of charge density given by point-charge models. The central multipole expansion describes the charge distribution using a single multipole series (charge, dipole, quadrupole etc.) expanded at the molecular centre. Only the lowest non-zero multipole moment is independent of the expansion origin. This is conventionally chosen to be the centre of mass, although another origin may be chosen. The multipole series is obtained by expanding $r_i^{-l}$ in the perturbation operator $\mathcal{H}'$ (equation (2.3)), as an infinite series of spherical harmonic functions to obtain the expression

$$\sum_{l_1 l_2 k_1 k_2} (-1)^{l_1 + l_2} \left( \binom{l_1 + l_2}{l_1} \right) Q_{l_1 k_1}^A Q_{l_2 k_2}^B \mathcal{S}_{l_1 l_2}^{k_1 k_2}(\Omega) R_{AB}^{-l_1 - l_2 - 1} \tag{2.25}$$

$Q_{ik}$ are the different multipole moments of the isolated molecular charge distributions of molecules $A$ and $B$, defined by the expectation values calculated from the molecular wavefunction

$$\hat{Q}_{ik} = \sum e_i r_i^{l} C_{ik}(\theta_i, \phi_i) \tag{2.26}$$

where $C_{ik}$ is a modified spherical harmonic function. $\mathcal{S}(\Omega)$ is a function of the orientation of the two molecules. The integers $l_i$, $r$, $k_j$, and $k_2$ define the different terms in the expansion in inverse powers of the intermolecular separation $R_{AB}$, associated with the different multipole moments $Q_{ik}$ of the isolated molecules. In general each $Q_{ik}$ multipole moment can have $(2l + 1)$ components, with $k = -l$, $(-l + 1)$, $(-l + 2)$...$l$. So for the dipole moment $Q_{ik}$, there are three components, $Q_{0}$, $Q_{1}$, and $Q_{1,-1}$. In practice, symmetry reduces the number of non-zero multipoles. For linear
molecules $k = 0$, therefore a linear molecule has a $Q_{10}$ dipole component only. For numerical calculations it is often more convenient to use the real components when $k 
eq 0$

$$Q_{ikc} = \left(\frac{1}{2}\right)^{\frac{1}{2}}(-1)^k Q_{ik} + Q_{ik}, \quad k > 0 \quad (2.27)$$

$$iQ_{iks} = \left(\frac{1}{2}\right)^{\frac{1}{2}}(-1)^k Q_{ik} - Q_{ik}, \quad k > 0 \quad (2.28)$$

$Q_{ikc}$ and $Q_{iks}$ are the real components of the multipole moments, where $c$ and $s$ indicate cos and sin. Substitution of the expression for the perturbation into the first-order expression from perturbation theory for the electrostatic interaction energy (equation (2.6)) gives the electrostatic interaction energy terms of the complex multipole moments $Q_{ik}$ as

$$U_{\text{estat}}(R, \Omega) = \sum_{l_1 l_2 k_1 k_2} (-1)^{l_1 l_2} \binom{l_1 + l_2}{l_1} Q_{l_1 k_1}^A Q_{l_2 k_2}^B S_{l_1, l_2, k_1, k_2}^A (\Omega) R_{AB}^{-l_1 - l_2 - 1} \quad (2.29)$$

This expression can be simplified in two ways: by expression in terms of the real components of the multipole moments instead of the complex components $Q_{ik}$, and through use of the labels $t$ and $u$ to abbreviate the multipole components $00, 10, 11c, 11s, 20, 21c$ etc. The interaction functions $T_{iu}^{AB}$ contain the radial dependence $R$ and the binomial coefficient, as well as the orientation dependence given by the $S$ functions. We then have equation (2.29) as

$$U_{\text{estat}} = \sum_{iu} Q_{iu}^A T_{iu}^{AB} Q_{iu}^B \quad (2.30)$$

Hence the electrostatic interaction energy, calculated from a multipole model, between molecule $A$ and molecule $B$ is determined by the permanent multipole moments of molecule $A$, $Q_i^A$ and the permanent multipole moments of molecule $B$, $Q_i^B$ as a function of centre of mass separation of the molecules, $R_{AB}$ and their relative orientations, given by the function $S$. 

24
The multipole expansion is valid under the formal condition that $R_{AB}$ is larger than the sum of the distances from the centre of each molecule to the limit of its charge distribution. Under this condition, for the expansion to converge spheres containing the charge distribution and multipole expansion, centred at the origin of each molecule, must not overlap. Since the charge distribution formally extends to infinity, this situation can never occur. For molecules which are non-spherical, overlap of convergence spheres can still occur in orientations where there is little significant overlap of charge density.

The charge distribution within a sphere containing all the nuclei in a molecule can be represented exactly by a multipole expansion. This implies that the charge distribution outside the sphere is zero, whereas in fact the charge distribution decays exponentially towards infinity. At short intermolecular separations this exponential decay causes the molecular charge densities to penetrate each other, producing an error between the multipole representation of the charge distribution and the true charge distribution which cannot be treated as a power series in $R$. However, if penetration effects are treated separately (usually as part of the short-range interaction) then in practice the convergence sphere for the multipole expansion need only encompass the nuclei in the molecule. This is a much less restrictive condition than enclosing the complete molecular charge density, but it is still unsuitable for calculations involving markedly non-spherical molecules in close contact. As a result, the central multipole expansion suffers poor convergence for grossly non-spherical molecules and is invalid for such molecules in condensed phases. The convergence sphere is not compatible with the (typically) non-spherical nature of a molecule, and so the central multipole model is more commonly used for small molecules.

### 2.6.4 Distributed Multipole Models

The central multipole model cannot be applied, for markedly non-spherical molecules, at separations where intermolecular interactions of chemical interest occur, whilst atom-centred charges average out the anisotropic features of charge density which are often integral features of such interactions. Distributed multipole models are essentially an extension and combination of atom-centred point charges with the central multipole expansion. Each atom in a molecule is approximately spherical, therefore it is possible
to use a multipole expansion for each atom to model the charge distribution. A variety of multicentre multipole models have been developed, including Stone's Distributed Multipole Analysis\textsuperscript{35, 36} (which is used throughout this work), and Cumulative Atomic Multipole Moments,\textsuperscript{37} which will briefly be considered here.

2.6.4.1 Cumulative Atomic Multipole Moments

Cumulative Atomic Multipole Moments\textsuperscript{37} (CAMMs) are one attempt to address the problems associated with atomic point charges and central multipole models. In the same way that the molecular charge distribution can be represented by a single multipole expansion, the molecular charge distribution can be divided into atomic multipole moments derived from the molecular multipole moments. The additive atomic multipole moments describe the local charge distribution around each atom and include, at each level, all contributions already described by lower atomic multipole moments. All but the lowest non-zero atomic multipole moments are dependent on the choice of origin. In order to avoid this dependence, contributions from lower moments already described are removed so that the description of charge density requires the cumulative set of atomic multipole moments.

As we have seen, the atomic charge can be defined in a number of ways. In this work the atomic multipole moments are defined by building on the Mulliken population analysis, so that the monopole term in the multipole expansion corresponds directly to the Mulliken charge. The problems associated with Mulliken charges are overcome by the addition of the higher atomic multipole moments.\textsuperscript{38} It has been shown that even the inclusion of only atomic dipole terms significantly reduces the basis set dependence which has limited the use of Mulliken atomic charges. CAMMs have been successfully applied in many studies, including molecules in crystal structure environments\textsuperscript{39} and construction of databases of atom-centred multipoles for nucleic acid bases.\textsuperscript{40}
2.6.4.2 Stone’s Distributed Multipole Analysis

The Distributed Multipole Analysis (DMA) utilises the fact that modern wavefunctions represent atomic orbitals in terms of Gaussian type functions. Each atomic orbital is represented as a sum of several Gaussian functions. The total electron density can be written in terms of Gaussian functions as

\[ \rho(r) = \sum_{\mu} \rho_{\mu} \eta_{\mu}^A(r) \eta_{\mu}^B(r), \]  

(2.31)

where \( \eta_{\mu}(r) \) and \( \eta_{\mu}(r) \) are primitive Gaussian functions. The advantage of this is that the product of two Gaussian functions with exponents \( \alpha^A \) and \( \alpha^B \), based at different sites A and B is another Gaussian centred at a point P between the two functions

\[ P = \frac{\alpha^A + \alpha^B}{\alpha^A + \alpha^B}. \]  

(2.32)

If the Gaussian primitives, \( \eta_{\mu}(r) \) and \( \eta_{\mu}(r) \) have angular momenta \( l_i \) and \( l_j \), then the overlap charge distribution can be represented by a finite multipole series up to rank \( (l_i + l_j) \) at each ‘overlap centre’ P. So the overlap resulting from two \( s \)-orbitals is equivalent to a point charge at P, whilst any charge distribution described by an \( s \) and \( p \) basis set can be represented completely by a multipole series up to quadrupole at many overlap points.

The rank k multipole moment of this overlap density with respect to an origin at P is then

\[ Q_{kq}(P) = -\int R_{kq}(r - P) \rho_{\mu} \eta_{\mu}^A \eta_{\mu}^B d\nu, \]  

(2.33)

where \( R_{kq} \) is a regular solid harmonic. Each atom is chosen as the multipole expansion site for its own contribution to the molecular charge density. The overlap multipoles can be shifted onto neighbouring atoms. DMA expansion sites S are chosen to ensure that no overlap multipole is shifted far from its overlap centre P. Distributed multipole models have an expansion site at the centre of each of a number of regions that the molecule is divided into. While these sites are usually chosen at the nuclei, additional sites can also be chosen at the centre of bonds if required. Each overlap multipole is
represented by a multipole expansion at a nearby site, or shared between sites. Multipole expansions at unwanted sites $\mathbf{P}$ can be represented as a multipole expansion shifted to the nearest chosen site $\mathbf{S}$ by using the equation

$$Q_{lm}(\mathbf{S}) = \sum_{k=0}^{l} \sum_{q=-k}^{k} \left[ \binom{l+m}{k+q} \binom{l-m}{k-q} \right]^{1/2} Q_{kq}(\mathbf{P}) R_{l-k, m-q}(\mathbf{S} - \mathbf{P}) .$$  \hspace{1cm} (2.34)

If $|\mathbf{S} - \mathbf{P}|$ is small, the series is rapidly convergent after $Q_{20}$ for an sp basis set. The convergence sphere for each site is large enough to enclose all overlap multipoles that are assigned completely or partially to that site.

The electrostatic potential at a point $\mathbf{r}$ due to a multipole series at a site $\mathbf{S}$ is given by

$$V(\mathbf{r}) = \sum_{lm} |\mathbf{r} - \mathbf{S}|^{-l-1} C_{lm}(\mathbf{r} - \mathbf{S}) Q_{lm}(\mathbf{S}) .$$  \hspace{1cm} (2.35)

The point where the potential is determined must be further from the multipole expansion site than any of the overlap populations partially or completely contributing to it. The expansion will only converge to the correct potential at points outside the charge distribution, because penetration effects due to overlap of charge density are not accounted for.

2.6.4.3 Advantages of the DMA

The DMA method has several advantages over the central multipole expansion. The method is valid for any molecular configuration where there is no charge overlap. The multipole moments are limited only by the quality of the wavefunction, so use of the DMA method essentially allows calculation of $U_{\text{e}}$ to the accuracy of the wavefunction. Distributed multipoles, unlike a single site multipole expansion, give a better qualitative picture of the charge distribution in a molecule and inherently take account of anisotropic features of charge density such as $\pi$-electron density and lone pair orbitals. Such an accurate model for the electrostatic features of a molecule is practically, as well as theoretically useful.

A comparison of the electrostatic potential resulting from a Mulliken analysis and potential-derived point charges (taken from the AMBER molecular mechanics
force-field of Kollman et al.\textsuperscript{45} for a cyclosporin derivative has been carried out by Price et al.,\textsuperscript{41} using a distributed multipole analysis as a standard. The Mulliken analysis was carried out as described, and the potential-derived charges were transferred from calculations on smaller molecules in standard configurations. The potential surface resulting from Mulliken charges was found to give a deceptive and unrealistic description of the true molecular potential surface. The Mulliken charges were of larger magnitude than the potential-derived charges, therefore bonds in the Mulliken model showed exaggerated polarity. Mulliken charges were originally intended to examine the bonding interactions between small molecules, so the use of a Mulliken population analysis to study the intermolecular interactions of large molecules has proven dubious on many occasions. The potential-derived charges gave a reasonable qualitative agreement with the DMA. However their derivation is computationally more time consuming than either a Mulliken or distributed multipole analysis. The differences between the potential-derived charges and the DMA were due to the basic limitations of point charges and the assumption that charges determined for small molecules in standard geometries can be transferred to larger, more strained molecules.

The problems associated with the transferability of atom-centred point charges between different molecular conformations and different molecular systems are well known. Studies on the transferability of DMAs for peptides and amides\textsuperscript{44} have shown that ‘transferable fragment’ electrostatic models can give good predictions of the electrostatic interactions in peptides. Atomic multipole moments are reasonably transferable if the directly bonded functional groups in each ‘fragment’ are the same. Within this condition, DMAs from entire peptide residues can be transferred in order to include polarization effects from nearby functional groups. For the test molecules the transferable DMA models produced much smaller errors in the molecular charge distribution than Mulliken charges, when compared to ‘exact’ energies calculated from the DMA of the molecular wavefunction.

A comparison of different schemes for partitioning charge density which includes Hirshfeld’s approach, CAMMs and DMAs, was carried out for the hydrogen fluoride dimer by Spackman.\textsuperscript{45} The methods demonstrated excellent agreement for dimer geometries and total energies, when terms up to quadrupole were included in the charge density descriptions.
2.7 Hydrogen Bonding

Most of the important processes occurring in living organisms are the result of binding interactions between molecules. Hydrogen bonding is a critically important structural interaction in biological molecules. Much of the secondary structure of proteins such as α-helices and β-sheets, arises from and is stabilised by hydrogen bonding between peptide links, i.e., N-H \( \cdot \) O=C. Hydrogen bond formation is also important in molecular recognition processes. A detailed understanding of hydrogen bonding in inhibitor recognition processes is important for successful drug design efforts.

Hydrogen bonding occurs between a proton donor group A-H (where A is an electronegative atom) and an electronegative acceptor group B: a site of high electron density such as lone pairs or \( \pi \)-electron density. Hydrogen bonding interactions are distinctly directional and specific. The interaction is weak compared to more familiar chemical bonds, usually between 10 and 40 kJ mol\(^{-1}\).

2.7.1 Van der Waals Dimers

Buckingham and Fowler\(^{46,47}\) predicted the structures of a wide range of van der Waals dimers using a hard-sphere/electrostatic model. The model consists of sets of point multipoles at the centre of van der Waals hard-spheres (to model the short range repulsion). The hydrogen atom radius is considered to be zero for hydrogen bonding interactions, i.e., the proton lies within the radius of the heavy atom it is bonded to. The lack of repulsion between the proton and its acceptor is a characteristic of hydrogen bonds.\(^{45}\) The multipoles were calculated from a distributed multipole analysis of an SCF wavefunction. The structure of each complex was modelled by minimization of the electrostatic interaction in sterically accessible orientations, under the condition that two heavy atoms cannot approach each other more closely than the limit of their van der Waals radii.

In nearly all cases this electrostatic model was in agreement with experimentally available structures. These results agreed with the observation that hydrogen bonds will form to regions of high electron density in the receptor, as stated by the Legon-Millen rule\(^{48}\) for predicting the angular geometries of hydrogen bonded dimers. The success of the Legon-Millen rule and the Buckingham and Fowler model indicates the importance of the electrostatic component to the orientation characteristics of hydrogen
2.7.2 Hydrogen Bond Interaction Energy

Mitchell and Price\textsuperscript{49} used Hayes-Stone intermolecular perturbation theory\textsuperscript{3} to study the distance and orientation dependence of different contributions to the interaction energy of N-H--O=C hydrogen bonds, for a trans- formamide/formaldehyde complex. The energy of the system was minimised with respect to the distance N--O and the angle $\alpha$ (the in-plane angle between the N-H and C=O bond vectors), subject to the constraint that the angle $\beta$ (the elevation of the hydrogen bonding proton above the carbonyl plane) equals zero for planarity.

The electrostatic and exchange-repulsion terms dominated the energy, but polarization, charge transfer and charge transfer correlation terms had a pronounced effect on the value of $\alpha$ and on overcoming repulsion to reduce the N--O distance. The exchange-repulsion energy was almost equal in magnitude to the sum of the penetration energy and other non-electrostatic attractive contributions, so that the electrostatic term was very similar to the total interaction energy. This cancellation of terms means that variations in the DMA electrostatic energy reflected variations in the total interaction energy. Their results showed a marked energetic preference for the N-H donor group to lie in the lone pair plane of the acceptor oxygen, with the N-H--O interaction slightly non-linear. The potential surface for planar near-linear hydrogen bonds was flat, therefore implying no significant preference for hydrogen bonds in lone pair directions in variance to the Legon-Millen rules. This apparent inconsistency is the result of considering isolated hydrogen bonds in the absence of the rest of the molecule.

2.7.3 Electrostatic Directionality of Hydrogen Bonds

Additional studies have been carried out on the electrostatic directionality of N-H--O=C hydrogen bonds by Mitchell and Price.\textsuperscript{50} They determined the electrostatically preferred geometries of 29 van der Waals complexes with hydrogen bonds of the form N-H--O=C, using a Buckingham and Fowler type model.

The complexes favoured near-linear hydrogen bonds (i.e., the hydrogen lies close to the N--O line) and planar hydrogen bond geometry, on electrostatic grounds. Their results suggested a different interpretation of the Legon-Millen rule. The
preference for hydrogen bonds in lone pair directions is not a result of the electrostatic potential due to the lone pair electron density *per se*, but rather to the formation of two N-H•O=C (or more commonly one N-H•O=C and C-H•O=C) interactions which this geometry allows. This second beneficial electrostatic interaction between neighbouring groups leads to a reduction in energy. The lone pair directionality assumed in the Legon-Millen rules, derived for complexes with single hydrogen bond contacts, will not dominate if the possibility for multiple electrostatic interactions exists.

Further calculations\(^1\) have used intermolecular perturbation theory to explore the validity of the Buckingham and Fowler model for the van der Waals dimers examined by Buckingham and Fowler. For each of the molecular systems studied, including those whose structures were not predicted by the Buckingham and Fowler model, the electrostatic energy was the largest term and displayed the same angular variation as the total energy. This demonstrated that the electrostatic contribution to the interaction energy is vital to the directionality of hydrogen bonding and dominates the orientation dependence of the intermolecular potential.

### 2.7.4 Modelling Hydrogen Bonds

The main component of hydrogen bond interaction is the electrostatic interaction energy, so a reliable model for molecular charge density is required for an accurate theoretical treatment of hydrogen bonds. Atomic point charge models are an attempt to represent the charge density information present in an *ab initio* wavefunction in a simplified fashion. Such models are used as standard in molecular mechanics force fields, and have been successful in modelling a variety of biomolecular systems. However, the inherent weakness of point charge representations in regard to the anisotropic features of charge density means that they are unable to represent accurately hydrogen bond interactions. As a result, the empirical lone pair directional characteristics of hydrogen bonds are reproduced in many force fields by the addition of special 'hydrogen bond' terms.\(^5\)\(^2\)\(^3\)\(^4\)\(^5\) These terms are intended to compensate for the deficiencies of the point charge model. An accurate model of charge density should account for all the features of hydrogen bond interactions without the addition of special terms.

The use of distributed multipole models in the improvement of molecular
mechanics force fields has been investigated by Koch and Egert.\textsuperscript{56} They introduced the
formulae for the calculation of the interaction energy between atom-centred multipole
moments into a force field program. Their results for a series of test systems
demonstrated once again the importance of the non-spherical features of atomic charge
densities. The established transferability of the atomic-multipoles from smaller
functional groups to model larger systems, and the success of DMA models transferred
between model peptide conformations\textsuperscript{57} both follow the rationale of force fields, and
point towards their use in the next generation of molecular mechanics force fields.

2.8 Summary
The intermolecular potential is often dominated by its electrostatic component, which
arises from the interaction of ground state molecular charge distributions. Therefore the
accurate modelling of many intermolecular interactions requires a realistic theoretical
description of molecular charge distributions. To obtain a convenient representation of
a molecular charge distribution, atom-centred point charges are often employed. These
may be constrained to reproduce the molecular electrostatic potential or known
molecular multipole moments. However, such a model gives a spherically averaged
description of the electrostatic properties of a molecule, whereas the electrostatic
interaction depends on a realistic treatment of the anisotropic characteristics of charge
density. This becomes particularly apparent with respect to hydrogen bond interactions,
for which lone pair and π electron density are important. Atomic point-charge models
can be extended to include additional charge sites such as lone pairs, or even extended
electron distributions - a ‘distributed monopole analysis’\textsuperscript{23} - to model the anisotropy of
molecular charge density. Such anisotropic modelling is also inherent in distributed
multipole models. Through the use of atom-centred multipole expansions the
Distributed Multipole Analysis gives more detailed information about molecular charge
distributions than atom-centred point charges or a one-centre multipole expansion. Once
a description of molecular properties is obtained (without the added confusion of
approximate charge density models), the molecular electrostatic properties can be used
as one determinant of the requirements critical to an intermolecular interaction, such as
receptor-ligand binding. Identifying the binding criteria of existing ligands is the first
step in the long process of attempting to design protein-binding ligands.
References for Chapter 2.


Chapter 3. Molecular Similarity

In this chapter, several methods which have been proposed for calculating the degree of similarity of two molecules will be reviewed. These methods broadly fall into three categories: those which attempt to match structural features between molecules, those which measure similarity in terms of the properties of molecular surfaces, and those which combine information about the chemical structure and properties of the molecules. A novel method for determining the position of maximum electrostatic similarity of two ligands common to a receptor site is introduced, based on an accurate Distributed Multipole Analysis (DMA) model for the molecular charge distribution. The relative orientations are determined by matching electrostatic potential extrema between molecules. The electrostatic extrema are determined at fixed distances beyond the molecular van der Waals surface, and represent points where strong binding interactions with a receptor are possible, and so consider both the steric and chemical aspects of similarity.

3.1 Introduction

How similar is one molecule to another? The answer to this question is fundamental to understanding molecular recognition processes such as ligand-receptor binding. The definition of the similarity of two molecules depends upon the qualities which are being compared, the method by which they are compared and the region of comparison. If we are discussing the similarity of molecules which share a common protein binding site, then the requirements for binding can be used as the parameters of a similarity comparison. A process as complex as ligand-receptor binding is subject to a number of effects, so it becomes necessary to decide what to include and what to exclude from the similarity comparison in order to produce a clear picture of the similarity requirements. For example, if the molecular functionalities vital for effective binding are known, then non-essential sidechains can be excluded from the comparison.

Steric accessibility is the most basic consideration of any molecular recognition process. A ligand must be able to ‘fit’, at least partially, within a receptor site for a binding interaction to occur. We might assume that two molecules which can both occupy a common receptor site cannot have receptor atoms within either of their van der Waals volumes. Even if the two ligands are of greatly differing sizes, we would
expect the smaller molecule to be contained largely or wholly within the volume of the larger one. This assumption holds unless the receptor site is large and open: in this case different ligands could satisfy binding interactions at different parts of the receptor, and any similarity comparison of such ligands would be rendered invalid. If this is not the typical case, and ligands occupy the same volume within a receptor from which receptor atoms are excluded, then the implicit assumption is that the binding site is rigid. If the binding of different ligands is accompanied by significant structural changes in a binding site (the induced fit model of enzyme action), then a similarity comparison of such ligands is, again, invalidated. Therefore, within the constraints of a rigid binding site model, steric similarity is one possible definition of molecular similarity.

While the steric requirements at a receptor site are a function of the ligand nuclear positions, any interactions are determined by the electron distribution. The electrostatic forces are the main interaction between two approaching molecules. The ligand must be electrostatically complementary to the receptor; regions of opposite electrostatic potential should coincide to produce a stabilizing interaction between the ligand and receptor. Complementarity between a receptor and a series of ligands implies similarity between the ligands. Thus, another aspect of molecular similarity is electrostatic similarity. The region of focus for steric similarity is clearly defined by the molecular van der Waals volume, however the choice of region for electrostatic similarity is more debatable. Electrostatic similarity will only be highlighted if the electrostatic properties are examined in the critical regions. As a further refinement of electrostatic complementarity we might focus on regions where specific ligand-receptor interactions, such as hydrogen bonding, are possible.

There remains the question of how similarity is measured. For a steric similarity comparison a volume overlap or distance-dependent measure, between corresponding atoms, could be employed. Electrostatic similarity can be expressed, either on molecular surfaces or within molecular volumes, as the difference between potentials for two molecules at a series of grid points. However, such an approach requires a regular and comparable distribution of grid points. Another option is to use specific points at regions surrounding the molecules where strong intermolecular interactions are possible. Alternatively, integrals over the electrostatic potential outside the entire region around two molecules could provide a single measure of similarity. Such a quantitative
measure of similarity is perhaps the pragmatic aim of ligand design processes. This would allow a great number of possible drug candidates to be screened and ranked rapidly according to their similarity.

3.2 Calculation of Molecular Similarity

3.2.1 Superposition of Atom Positions
When comparing the similarity of two molecules, their chemical structures are the obvious starting point. The medicinal chemistry concept of optimizing a lead compound is based on the assignment of useful pharmacological properties to specific structural features, and attempting to maximize desirable properties through structural changes. Hence, when comparing two known ligands for a common receptor site and attempting to quantify their similarity, one might begin by comparing structural similarities, i.e., matching atomic positions.

3.2.1.1 The OVID Algorithm
Two related methods for exploring the similarity of molecules using atomic positions have been proposed by Hermann and Heron. Their comparison is made in terms of an overlap integral of two molecules, with a weighting function to make the overlap contribution in certain regions more important than in others. Two separate definitions for the weighting factor are used, employed in two different similarity algorithms. (The second algorithm will be discussed in section 3.2.2.3.)

The first algorithm, OVID, only considers atoms in the ligand molecules which are known to be important to ligand-receptor binding from structure-activity relationships. These atoms are treated as van der Waals hard-spheres and given a weighting factor of 1. All other atoms are given a zero weighting factor. OVID performs a least-squares fit to position the molecules in their relative orientations, and the maximizes the total volume overlap as a sum of atomic volume overlap integrals over the selected atom pairs. Hence, the overlap of the two ligands is determined by the intersecting volume of the selected atom pairs only. The result is expressed as a percentage.

The assignment of weighting is useful, because it identifies certain regions of the molecules as more important than others in receptor binding. However, although the
‘important’ atoms are identified from experimental data, this method requires an inherent presumption (when assigning the atom-atom correspondences) about the relative binding orientations of the two molecules. More importantly, the method requires this explicit assignment of corresponding atom pairs between the two ligands, since the comparison is performed on the basis of hard-sphere overlap. This assignment may not be immediately obvious, or even practical when comparing ligands of vastly differing chemical structures. Structural overlay methods are obviously biased towards molecules which show broad structural similarity in the first instance. A determination of similarity which relies on a structural overlay will fail when there is a large amount of dissimilarity between two molecular structures. Such dissimilarity is further complicated as the number of possible molecular conformations increases.

3.2.1.2 Superposition of Flexible Molecules

A new method for the atom-based superposition of flexible molecules, which avoids the need for the initial assignment of atom-atom correspondences, has been developed. For any pairwise comparison of two \( n \)-atom molecules there will be \( n! \) possible results. As larger and more flexible molecules are encountered, the number of possible superpositions increases beyond that possible by manual superposition. To limit this combinatorial explosion, this method uses simulated annealing and cluster analysis.

Simulated annealing locates the global minimum energy conformation of each molecule. A random sub-set of energetically accessible conformations (those within 33.3 kJ mol\(^{-1}\) of the global minimum) along the trajectory are classified (via cluster analysis) into a small number of significantly distinct clusters. This clustering is based on the assumption that many such energetically accessible conformations will be largely similar in structure, therefore structures which are significantly different from each other can be identified by locating ‘conformational clusters’ in torsional space. A single conformer, representative of each cluster, is selected, so that only a relatively small number of conformers are needed to give a reasonable random sampling of the entire molecular conformational space. These conformers are matched pairwise by minimizing their difference-distance matrix. The conformation of each molecule which gives an optimum superposition can be determined in one of two ways. First, each molecule in
turn is used as a reference for the comparison, searching all conformations of the reference and the superimposed molecules to find the overlay of maximum similarity. Second, the best consensus match can be found by comparing all the molecules in their representative conformations, independent of a reference molecule.

In order to assign similar conformations to the same cluster, the measure of similarity used is a ‘distance measure’ between conformations, the Euclidean distance $d^i_j$ in torsional angle space

$$d_{jk}^i = \sum_{i=1}^{i} (\Delta \tau_i^j)^2$$

is taken as the sum of the squares of the differences in torsion angle $i$ for any two conformers $j$ and $k$ with $t$ torsion angles. The ‘most representative’ structure is chosen as that with the smallest sum of elements in the square distance torsional space, \textit{i.e.}, the smallest ‘distance’ between conformations in a cluster. The representative conformers are matched as rigid conformers. The objective function which is minimized during the subsequent simulated annealing of the superimposed conformers is dependant on the interatomic distances of the two molecules being matched.

This atom-based superposition method avoids the need for specific atom-atom overlay assignments, and also examines a selected range of accessible conformations of each molecule. The use of simulated annealing and cluster analysis helps to limit the number and range of conformations under consideration. While the number of conformational clusters is determined by the data, the authors point out that conformations midway between the representatives from each cluster could be ignored.

### 3.2.2 Pattern Matches on Molecular Surfaces

Since a receptor site effectively can only ‘see’ the van der Waals surface of a ligand, the region of importance is the molecular surface and beyond, rather than the absolute positions of the ligand atoms. Similarity could be more usefully measured in terms of properties on a molecular surface.
3.2.2.1 Surface Pattern Matching

A method for searching for pattern matches on molecular surfaces has been introduced by Dean et al.\(^1\) The similarity of two molecules is not measured in terms of correspondence of atom positions, but in terms of matching surface properties such as the molecular electrostatic potential.

A patch, of any predetermined shape, is fixed on the surface of molecule A and this surface is characterized in terms of the desired properties to be matched. A window of this patch is then projected onto molecule B, and molecule B is allowed free rotation in space until the surface characteristics of A and B in the windowed region exhibit maximum similarity. Highly scoring orientations are chosen by the *objective function* of the pattern matching procedure. The objective function is a measure of the difference between the surface parameter (for the patch and the molecule) at each point on the surface. This difference can be calculated for several different surface parameters. For this comparison to be made, the arrangement of points must be exactly comparable and well distributed. A gnomonic projection of an icosahedral tessellation is used to provide a regular array of points. During the search routine, the ‘pierce points’ of the gnomonic projection, together with the value of the surface parameter are computed.

This method is dependent on the size, shape and position of the patch which is characterized on the surface of molecule A. The description of the characterized region as a ‘patch’ may be slightly misleading, as the most commonly used shape in this study is a hemispherical surface covering almost half of the molecular surface. This is an attempt to simulate the surface matching requirements of a binding site cavity. In order to make a sensible choice, information such the orientation of molecule A in its receptor site (and hence, the region on the surface corresponding to the molecule's active face) is needed.

3.2.2.2 ‘Blind’ Surface Searching

To overcome this difficulty, this method has been extended to a ‘blind searching’ technique,\(^4\) where the relative receptor-bound orientations of both molecules are unknown. This extension is designed for a situation where two molecules are thought to bind at the same receptor site, but with no information on their binding orientations. In this case both molecules A and B are given random initial orientations.
and both are free to rotate, producing six degrees of freedom.

Both methods are sensitive to the number of surface points chosen for comparison. The authors point out that matching electrostatic potentials at each pair of gnomonically projected points gives some poor structural overlays, because the matching is performed on the projected surface without explicit consideration of the overlap of the molecular volumes. Hence, no account is taken of the accessible molecular surface. The particular advantage of the blind searching method is that it does not require any preconceptions regarding which areas of the molecular surface correlate with biological activity.

3.2.2.3 The SUPER Algorithm

SUPER is a more general program developed by Hermann and Heron in concert with the OVID program already described. This method does not consider atom positions or weight their relative importance in the similarity calculation. The points for comparison are taken from a grid of points on the molecular van der Waals surface. The electrostatic potential is computed at each surface point from a point charge model. Two points are considered to correspond if the distance between them is less than half the distance between two points on the same molecular surface (so the distribution of the surface points does not need to be regular or comparable for the two molecules). Rather than weighting the potential in different regions for matching, the matched pair of surface points are disregarded if the difference between their electrostatic potentials is greater than a certain value. An appropriate value is chosen for each pair of molecules being compared.

To search different relative orientations of the molecules, the atom at the geometric centre of each molecule is defined as an origin. Molecule A is then fixed and molecule B is rotated incrementally about its origin atom.Translations are done as a successive superposition of pairs of atoms, so that each atom in each molecule is eventually paired. This atom pairing routine is only used to generate the relative orientations, and does not feature in the similarity calculation. The results are ranked according to the number of corresponding surface points.

The SUPER method was compared with OVID, and demonstrated for the examples chosen, that the best match of surface properties does not correspond to the
best match of atomic centres. The particular advantage of SUPER is that it performs a search of all relative orientations for two molecules without prejudice. However, the use of a cut-off to compare the value of the electrostatic potential at surface points does not allow for situations where the potential at either point being compared is just above or below the cut-off value. This would result in such a pair of corresponding points to be disregarded as unmatched. Ranking the resulting relative orientations according to the greatest correspondence of surface points (within the matching criteria) takes no account of regions of the molecular surface which are significant to receptor binding. Hence, this method does not take any measure of the hydrogen bonding capability of the molecules under examination.

3.2.3 Molecular Similarity Indices

Similarity indices have developed as a method of providing a quantitative measure of molecular similarity. This idea was introduced by Carbó et al., and has been extensively developed by Richards and co-workers.

3.2.3.1 The Carbó Index

Carbó defined molecular similarity using a formula measuring the electron density overlap between two superimposed molecules A and B,

\[ R_{AB} = \frac{\int \rho_A \rho_B dv}{\left( \int \rho_A^2 dv \right)^{1/2} \left( \int \rho_B^2 dv \right)^{1/2}} \]  

(3.2)

where \( \rho_A \) is the electron density of molecule A. The electron density of each of the superimposed molecules is calculated at all points in space and then integrated over all space. \( R_{AB} \) has values in the closed range 0 to 1, where 0 indicates complete dissimilarity, and 1 signifies total similarity. Optimization of the original Carbó molecular similarity index was carried out using charge densities calculated with a CNDO-like approximation of the density functions.

A subsequent comparison of the Carbó index using CNDO-like and \textit{ab initio} charge densities was performed by Bowen-Jenkins et al. Since the value of \( R_{AB} \) is also dependent on the relative position of A and B, three criteria for superimposing the pairs
of molecules were also studied, since a direct optimization of the index $R_{AB}$ was not possible. These calculations were performed separately on a series of isoelectronic pairs of molecules, and a series of molecular isosteres (molecules similar in size and shape).

For the isoelectronic molecules there was good agreement between both the *ab initio* and CNDO similarity indices, and the three superposition methods. The predictions of the two methods of charge density calculation and the three superposition methods were quite different for the isosteric molecules. A general observation (for both the isosteric and isoelectronic pairs) was that the *ab initio* similarity index values decreased more rapidly than the CNDO index, as the molecules were displaced from their positions of maximum $R_{AB}$. For the isosteres, the CNDO index values were insensitive to the method of superposition. The *ab initio* charge densities were therefore more reliable and better discriminators in the calculation of molecular similarity indices. These authors were the first to suggest the use of molecular electrostatic potentials and electrostatic fields as more useful quantities for calculating molecular similarity indices, rather than charge densities.

Some consideration was given by Bowen-Jenkins et al.\(^7\) to the calculation of the Carbó index using the total electron density or valence electron density only, derived from *ab initio* wavefunctions. Results for a series of prostaglandins and histamine H2 antagonists showed that the similarity comparison was not affected by the use of STO-3G basis sets. Chemically invariant but orientationally distinct sidechains on the prostaglandin ring system were truncated as methyl groups. The axial and equatorial positions of the methyl groups were found to be adequate for representing different sidechain orientations without affecting the calculated similarity index values. The use of valence-only electron density gave results which ranked the similarity of the molecules in line with their observed experimental activity, emphasizing the importance of valence electrons as the region of interest when calculating molecular similarity.

The use of *ab initio* molecular similarity indices is computationally exhaustive, requiring the calculation of many four centre integrals. In the same year, a semi-empirical method for calculating molecular similarity was introduced,\(^8\) which gave considerable time saving compared with the *ab initio* calculations. By characterizing the electron density of a series of simple molecules, these can be used as 'standardized' molecular fragments, to build up more complex systems. The electron density is
represented by a series of spherical Gaussian functions placed at maxima in the \textit{ab initio} electron density. Molecular similarity calculations using the Gaussian approximations gave good agreement with \textit{ab initio} valence electron density results, but poorer agreement for total electron density. The three methods gave similar relative positions of maximum similarity for the molecules being compared.

3.2.3.2 The Hodgkin Index

An alternative definition of molecular similarity, based on the Carbo formula was introduced by Hodgkin and Richards\textsuperscript{9} which defines similarity in terms of molecular electrostatic potentials (or electrostatic fields):

\[ H_{AB} = \frac{2\int \rho_A \rho_B dv}{\int \rho_A^2 dv + \int \rho_B^2 dv} \]  

(3.3)

Substitution in the Carbo index of $\rho_A=n\rho_B$ (where $n$ is a constant) gives a value for the index of $R_{AB}=1$, or complete similarity. Hence, whilst the Carbo index would account for similarity between the shape of molecular electrostatic potentials, it does not account for the magnitude of the potentials. In the alternative definition of the index given by Hodgkin and Richards, substitution of $\rho_A=n\rho_B$ gives

\[ H_{AB} = \frac{2n}{1+n^2} \]  

(3.4)

therefore this similarity index also considers the absolute values of the electrostatic potential. $H_{AB}$ ranges from -1.0 to +1.0. To compute the Hodgkin molecular similarity index, the molecular electrostatic potentials were calculated at a grid of points around the two molecules under comparison and the integration performed numerically at each grid point. The grid extends 10.0 Å around the molecules, with a 1.0 Å increment between grid points. Calculations performed with different grid extensions and increments found this specification to be computationally efficient for reproducing results obtained with much larger and ‘finer’ grids. A crude grid will give faster calculations, but with a consequent loss of accuracy. Later calculations\textsuperscript{10} verified molecular similarity ranking does not depend on the grid extent, but is sensitive to the size of the grid increment, with ‘coarser’ grids giving misleading results. Grid points
which lie within the van der Waals volume of both molecules are excluded from the similarity calculation by being assigned a MEP value of zero. The remaining points which lie within the volume of one molecule only are included in the calculation, contributing to the denominator of equation 3.3. However, these points do not contribute to the numerator and so should not be included in the similarity calculation. The only true region for comparison is that in which molecules A and B have coincident grid points outside of their combined van der Waals volumes. Further discussion about on the appropriate region for comparison in molecular similarity studies is made in section 3.4.1.

The electrostatic potential calculated at the grid points around each molecule replaces the electron density in the similarity index (equation 3.3). The molecular similarity index is optimized by fixing the lead molecule A, and rotating and translating molecule B in space in order to maximize its similarity with respect to molecule A. Molecules A and B are given a random initial orientation, and molecule B is subsequently given complete freedom in three dimensions to move towards its position of maximum similarity with A. The optimization proceeds via the SIMPLEX method.

A comparison of results for the Carbó and Hodgkin indices showed that the Carbó index overestimates similarity, based on shape of the electrostatic potential alone. In cases of genuine electrostatic potential similarity, the Carbó and Hodgkin indices were in agreement. This represents a significant improvement over the original definition of the similarity index by Carbó.

### 3.2.3.3 Applications of Molecular Similarity Indices

From the introduction of molecular similarity indices, and the use of molecular electrostatic potentials and fields, this idea has been extended in a number of directions. In one sense, the idea of molecular similarity has come full circle, with the application of molecular similarity indices to molecular shapes defined by hard-spheres. However, a steric cutoff is a discreet quantity whereas electron density is continuous and therefore a more subtle measure of similarity. Molecular similarity indices have found further applications in QSAR studies, and the screening of 3D molecular datasets. All previous molecular similarity indices were calculated using rigid molecular geometries. Burt and Richards introduced the idea of flexibility into the calculation of molecular
similarity indices, \textit{i.e.}, that the flexibility of a molecule may allow it to modify its conformation in order to produce a better fit (and hence greater similarity) at a receptor site. Molecular similarity indices are calculated, as described, for molecules A and B and additionally at all possible rotations about available torsional bonds in the molecules. In order to exclude energetically unfavourable conformations from the calculation, a weighting function is used. Thus, the index which is optimized is weighted to a Boltzmann factor so that only an energetically accessible conformation can be produced as the optimum. The value of the Boltzmann factor is chosen to produce a meaningful distinction between the molecular similarity index values of different conformations, relative to their energies.

The particular feature of the similarity indices is the quantitative measure of similarity which they provide. The usefulness of such a measure, and its relationship to biological data is less apparent.

\textbf{3.2.4 Superposition of Hydrogen Bonding Partners}

It is the physical and chemical properties of a molecule, rather than its chemical structure, which determine its biological activity. A new function has recently been developed\textsuperscript{18, 19} based on a novel least-squares method which superimposes the expected positions and orientations of hydrogen bonding partners in the receptor that are deduced from both molecules. A search for all possible correspondences of hydrogen bonding functional groups between molecules is performed. The correspondence means that a pair of hydrogen bonding heteroatoms of the two ligand molecules can interact with a common receptor hydrogen bonding heteroatom, but this does not necessarily mean matching of their atom positions.

The expected positions of hydrogen bonding heteroatoms in the receptor are calculated as dummy atoms from all possible hydrogen bonding functional groups in both molecules, considering the positions and orientations of X-H or X-lone pair electrons. The program calculates the positions of dummy atoms automatically according to predefined types of hydrogen bonding heteroatoms. For each member of the heteroatom correspondences, different conformers of both molecules are generated by systematic rotation of torsional bonds. A weighting factor is used to favour hydrogen bonding sites with a good fit. If a hydrogen bonding pair are mismatched (\textit{e.g.} donor-
donor or receptor-receptor) they are weighted zero. Different pairs of corresponding heteroatoms are systematically paired, and for each combination and its conformers an iterative least-squares fitting procedure is applied. The quantity minimised in the least-squares calculation is

$$F = \sum_{i=1}^{m} W_i \left\{ \sum_{j=1}^{3} [Y_{Ai(j)} - Y_{Bi(j)}]^2 + \sum_{j=1}^{3} [V_{Ai(j)} - V_{Bi(j)}]^2 \right\}$$

(3.5)

where $Y_{Ai}$ and $Y_{Bi}$ are the positions of the expected receptor hydrogen bonding sites, and $V_{Ai}$ and $V_{Bi}$ are the direction vectors deduced from the positions $X_{Ai}$ and $X_{Bi}$ of the heteroatoms of both molecules. $W$ is the weighting factor, $m$ is the number of pairs of hydrogen bonding sites being considered and $j$ is set to 1, 2, or 3 for the x-, y-, and z-coordinates respectively.

This method has been shown to give good results when compared to conventional atom matching methods for systems where hydrogen bonding is important, but essentially ignores other strong intermolecular interactions. The inclusion of conformational freedom in the method gives a more complete model but makes the method computationally time consuming. A particular advantage of this method is that the chemical nature of the molecule, as generated by its structure, rather than the structure itself, which is being compared. Hence, the method is suited to the comparison of structurally dissimilar molecules.

### 3.2.5 Molecular Electrostatic Potential Similarity

Sanz et al. have produced an integrated computational system for the comparison and analysis of various aspects of molecular electrostatic potentials, and their application to molecular similarity. The MEPSIM program is in modular form, and allows calculation of MEP at several levels of theory, from semi-empirical to ab initio.

The MEPSIM package developed from early observations on a series of monoamine oxidase (MAO) substrates and inhibitors. MAO substrates are classified as type A or type B according to their susceptibility to MAO-A or MAO-B ‘selective’ inhibitors. The electrostatic potential (calculated at the STO-3G ab initio level) was used as one criterion to try to discriminate between ligands with type A and type B activity. Four specific electrostatic potential minima (calculated at 3.0 Å from the
ligand aromatic ring plane) related to different structural features in the substrates were identified and used in the comparison. These authors noted that the presence (or absence) of certain of these minima for known ligands correlated with their experimental classification as MAO-A or MAO-B substrates. Quantitative predictions about experimentally unclassified ligands were made based on the distribution of these four electrostatic potential minima. A subsequent study of β-carbolines (ligands with MAO-A activity) further refined this model by using the geometric relationship between these minima to compare different ligands.

To search for electrostatic potential similarity using MEPSIM the two molecules, A and B, are placed in a random relative orientation. A common 3D grid is defined for both molecules. The electrostatic potential is calculated at grid points which extend up to 3.0 Å beyond the molecules greatest dimension along each Cartesian axis with a grid increment of 0.5 Å. Grid points with a lower MEP than surrounding grid points are identified and a gradient searching method can be performed around these points to determine the precise positions of minima in the electrostatic potential. However, the similarity is assessed by computing the Spearman rank correlation coefficient between the electrostatic potential values of each molecule at each grid point. The search for the position of maximum similarity between the molecules is performed by a gradient method. In order to avoid local maxima in similarity, several random starting positions and orientations are generated. From each of these random positions, any repeated convergence trajectories, or trajectories outside of the specified translation criteria are rejected.

This method has been used to study the inhibition of caffeine by other xanthines, where similarities between regions in the electrostatic potentials were qualitatively highlighted. A later module of the MEPSIM program allows for conformational freedom in searching for maximum similarity. Molecule A is fixed as the reference, and molecule B undergoes internal conformational changes, as well as changes in its relative position. The reference molecule is chosen to be the most rigid molecule, the molecule of known active conformation, or the largest molecule. The initial positions of the molecules are determined by a least-squares fitting of atomic positions for similar molecules, or superposition of charge centres with alignment of molecular dipole moment vectors for dissimilar molecules.
Sanz et al tailored their software package for situations where a receptor interaction depends on a specific region of the ligand electrostatic potential, rather than the entire molecular electrostatic potential. Using this module\textsuperscript{26} the similarity comparison is carried out in a specified molecular plane. Electrostatic similarity is maximized by rotating and translating molecule B, within the specified plane, relative to the reference molecule A, via a gradient driven maximization process. For a set of indole derivatives\textsuperscript{26} which act as ligands for the 5-hydroxytryptamine receptor, this plane was defined 1.6 Å above the indole plane, in order to include the electrostatic potential minima generated by the π electron density of the indole ring. By focusing on this particular region of the ligand electrostatic potential, the gradient maximization was able to reproduce the alignment of orientation vectors previously observed by Weinstein et al\textsuperscript{27,28,29} for 5- and 6-hydroxyindole.

The optimization method employed not only detects absolute maximum similarity, but also other local maxima. However, the translations and internal deformations, which the molecules are subject to, results in the loss of superposition of the precomputed coincident grid points of the two molecules. Each step of the similarity maximization process thus requires the calculation of a new coincident grid, because the array of grid points are not regularly distributed or projected onto a regular surface.

The Spearman rank correlation coefficient

\begin{equation}
    r_{(S)} = 1 - \frac{6 \sum_{i} [r(V_{A_i}) - r(V_{B_i})]^2}{n(n^2-1)} \tag{3.6}
\end{equation}

measures the electrostatic potential similarity at the grid points. The MEP values of the two molecules are sorted in increasing numerical order, so that $r(V_{A_i})$ is the position (rank) of the electrostatic potential of molecule A at grid point $i$. Hence, $[r(V_{A_i}) - r(V_{B_i})]$ is a measure of the difference in ranks between molecules A and B at grid point $i$ for $n$ pairs of points. However, because this equation is calculated for the ranks rather than the actual values of the electrostatic potentials, some authors\textsuperscript{30} have criticised it as a relatively insensitive measure of similarity. Whilst this is index is a useful measure of the shape of the electrostatic potential, it fails to take account of the magnitude of the potential when comparing two molecules.
3.2.6 Molecular Electrostatic Field Comparisons

The molecular electrostatic field has not been as widely studied as the molecular electrostatic potential within the context of molecular similarity. Because the electrostatic field is a vector quantity, its representation around molecules was somewhat problematic until the advent of modern computer graphics. However, the vector characteristics of the molecular electrostatic field also present an advantage for molecular similarity calculations, since both the magnitude and the direction of the field vectors can be used in the calculation of similarity. Electrostatic fields have been used in the calculation of molecular similarity indices, and recently in the study of structure-activity relationships with an emphasis on hydrogen bond interactions. However, these latter studies relied on electrostatic fields calculated using atomic charges derived from a Mulliken analysis. It has been shown that errors in the electrostatic field calculated using Mulliken charges are of the same order as the total field strength, and the largest errors are associated with the non-spherical features of charge density implicated in hydrogen bonding.

Further use of electrostatic field vectors in drug design was investigated by Blaney et al. Their program used a 3D array of arrows to represent the electrostatic field vectors, radiating from a user-defined ‘centre of projection’. The components of the electrostatic field are stored at the positions where the vectors intersect with the molecular surface (this surface is defined as the van der Waals surface multiplied by a constant term). The length of the vector is proportional to its magnitude. As one molecule is rotated in real time against a fixed molecule, the difference in their field vectors is represented as a third display. Shorter vectors indicate better match between the field strengths for each molecule: a point indicates an exact match.

The correlation between the electrostatic field vectors is calculated as

\[
\text{Correlation} = \frac{\sum (v_i \cdot v_2)}{\left( \frac{\sum (v_1 \cdot v_1) + \sum (v_2 \cdot v_2)}{2} \right)}
\]

where \((v_1 \cdot v_2)\) is the dot product of the vectors \(v_1\) and \(v_2\). The result is displayed on a scale from \(-1.0\) to \(+1.0\). However, a test of the field similarity program on a group of benzodiazepine agonists produced low correlation values. To improve the correlation, the field in areas of positive electrostatic potential was ignored, following the belief that
such regions do not represent hydrogen bond acceptor sites on a ligand, and are less useful in drug design. However, results presented in the following chapters will show that it is necessary to consider both the negative and positive regions in the electrostatic potential in order to arrive at a realistic picture of electrostatic similarity.

3.3 Summary of Similarity Methods
Molecular similarity is an active and expanding field of research, and so it would be impossible to discuss all the methodologies which have been proposed in the literature. The ideas presented here represent the main directions of the continuing research efforts focusing on molecular volumes, surface properties and chemical properties. Many of the latest ideas in this field represent technical rather than conceptual developments, as the methods and techniques are subject to continual refinement with the benefit of increased computing power. A summary of the methods discussed here is shown in table 1.

3.4 Similarity of Electrostatic Extrema
This idea was first used by Davis et al. to explain the inhibitory action of a series of compounds on the activity of a phosphodiesterase enzyme. This work will be reviewed more extensively in the next chapter. Davis et al. compared the electrostatic similarity of a range of Phosphodiesterase III inhibitors on the basis of the positions of their electrostatic potential minima, in order to identify possible common binding sites. The electrostatic potentials were calculated using a point charge model derived from semi-empirical calculations. The method of comparison used in this work is an extension of this approach.

Two molecules which bind at a common receptor site are likely to have many of their ligand-receptor interactions, particularly hydrogen bonding, in common. By identifying hydrogen bond donor and acceptor groups in two different ligands which interact with common hydrogen bond acceptor and donor groups in the receptor, it would be possible to determine the relative binding orientations of the two ligands. This rationale is followed by Kato et al., but what of the role played by other strong electrostatic interactions involved in ligand binding?

In this method, all maxima and minima in the electrostatic potential around each
ligand molecule are located at a set distance outside the molecular van der Waals surface, as described. Maxima in the potential can result from hydrogen bond donor groups in the molecule, while electrostatic potential minima may correspond to hydrogen bond acceptor groups. In an ideal situation optimum hydrogen bonding would occur for a receptor which had hydrogen bond protons at minima around the ligand molecule, and hydrogen bond acceptor atoms at maxima in the ligand electrostatic potential. Hence, if the electrostatic potential extrema are calculated at an appropriate distance from the molecular van der Waals surface, we might expect the locations of potential maxima to correspond to the positions of hydrogen bond acceptor atoms, and minima to correspond to hydrogen bond protons.

Sanz and co-workers\textsuperscript{31} noted that the pattern of electrostatic potential minima for a series of monoamine oxidase ligands served as a qualitative measure of their electrostatic similarity. Pépe et al\textsuperscript{37} have also examined the use of the molecular electrostatic potential in drug design. Molecular orientation depends on the electrostatic interaction between the unperturbed charge densities of sufficiently separated approaching molecules. They suggest that the relative positions of electrostatic extrema are more important than their relative magnitudes and they conclude that the use of extrema as a qualitative tool is unaffected by the method used to generate the electrostatic potential. However, it will be shown here that without an accurate representation of molecular charge density, many minima are sufficiently shifted, or even apparently absent, that the true picture of similarity is clouded. This is a common failing of many atom-centred point charge methods, however a ZDO derived semiempirical molecular electrostatic potential correctly identifies electrostatic minima associated with $\pi$ electron density.\textsuperscript{38} Through the use of DMAs relative orientations can be generated without added uncertainties in the calculated electrostatic properties.

Since we propose maxima and minima in the electrostatic potential as identifying possible hydrogen bonding interactions for a ligand molecule, when two ligands, which bind at a common receptor site, are superimposed in their relative binding orientations, we would expect the positions of many of their maxima and minima to coincide. Orientations with the greatest the number of extrema which coincide are assumed to be the most favourable. We therefore attempt to define the relative binding orientations for ligands at a common receptor site on the basis of the
similarity of their electrostatic extrema.

3.4.1 Region for Comparison

It is the electrostatic interaction between ligands and receptors which often dominates molecular recognition. The electrostatic properties in the region outside the van der Waals surface of a ligand can determine its interaction with the receptor site. When two molecules are being compared in their correct relative orientations within a binding site, the receptor atoms cannot be within the van der Waals volume of either molecule, assuming the receptor conformation is rigid. Hence, the points for comparison of the ligands must lie outside the van der Waals volume of both molecules.

The outer limits of the region of comparison are more debatable, and the choice will be important for some comparisons in cases where there is a large discrepancy in the size of the ligands being compared, or in situations where the overlay of ligands results in large side chains protruding in different directions. If all points up to 3.0 Å from either molecule are included, there will be large regions around the protruding residues where the electrostatic similarity is almost certain to be poor, since these regions will only be occupied by one of the molecules. These points will increase in number, and increasingly dominate the comparison as the steric overlay of the molecules gets worse. With the understanding that overlays which result in side chains pointing in different directions are only possible in the unlikely situation where there are two suitably orientated pockets in the receptor, it is appropriate to ignore the regions around side chains in the electrostatic comparison. A suitable region for comparison therefore includes all areas around both molecules which are not within either van der Waals volume, and are not more than 3.0 Å from the surface of both molecules (see figure 1).
Figure 1. The areas of overlap for molecule A (black lines) and molecule B (red lines). For each molecule the inner surface defines the van der Waals volume, and the outer surface is defined by a constant × van der Waals radius. This shell defines the region where intermolecular interactions could occur around molecule A (GA) and molecule B (GB). The appropriate region for comparing the molecular properties of A and B is therefore GA ∩ GB (solid red).

3.4.2 Experimental Support

In the determination of similarity it is not the matching of atomic positions which is important, but the similarity of molecular shape and functionality. The recognition step, given steric access, is controlled by intermolecular forces. Often in biological systems these are predominantly hydrogen bonds. Hydrogen bonding is by no means the only factor in molecular recognition and receptor binding, but it is an essential one. Hydrogen bonding is a highly directional and specific interaction, therefore the position of donor and acceptor groups in the ligand-receptor complex is important. The importance of hydrogen bonding in ligand-receptor complexes is due to their complementarity. A receptor site may be thought of as a three dimensional array of hydrogen bond donor and acceptor sites, and a suitable ligand will consist of a complementary set of acceptor and donor sites.49

The link between the molecular electrostatic potential and the hydrogen bonding ability of a molecule has been studied experimentally by Politzer and co-workers.40
They have identified a correlation between the degree of positive electrostatic potential (measured as the magnitude of the electrostatic maximum) around a series of naphthalene derivatives, and their relative susceptibilities for nucleophilic attack.\textsuperscript{41,42} They also studied the relationship between the solvent hydrogen-bond-acceptor parameter $\beta$, and the magnitude of the relevant electrostatic minima associated with oxygen and nitrogen atoms.\textsuperscript{43} $\beta$ is a measure of a solvent's ability to accept a proton in a solute to solvent hydrogen bond. For a series of nitrogen atoms in primary amines and azines, and oxygen atoms within alkyl ethers and molecules containing carbonyl oxygens, the general tendency was for the magnitude of $\beta$ to decrease with decreasing strength of the electrostatic potential minima (within several different families of solvents). In a similar fashion, for a series of alcohols, ethers and molecules in which alkyl groups were the proton donors, the hydrogen-bond-donating parameter $\alpha$ decreased with decreasing strength of the electrostatic potential maxima.\textsuperscript{44} Hence, the electrostatic potential of a gas phase molecule can be quantitatively related to its ability in solution to donate or accept protons in solvent to solute hydrogen bonds.

This work has been extended by Kenny\textsuperscript{45} to use the gradient of the molecular electrostatic potential, that is the electrostatic field, as a 'useful descriptor of hydrogen bond basicity'. In this work, electrostatic potential minima are shown to predict hydrogen bond basicities very well for a range of heterocycles capable of forming hydrogen bonds. By varying the position at which the electrostatic potential was calculated, the variation in its correlation with the experimental hydrogen bonding data was observed. In order to fit the experimental data it is shown that the molecular electrostatic field must be calculated further away from the molecule. Hence, the positions which correspond to the best fit are markedly different for the electrostatic potential and the electrostatic field, and 'at greater distances from the molecule, the electric field strength becomes a more effective descriptor of hydrogen bond basicity than electrostatic potential'.\textsuperscript{45}

Given the importance of electrostatic interactions, and particularly hydrogen bonding, in ligand-receptor binding, it is possible to define a new approach to assessing similarity on the basis of regions where the possibility of strong electrostatic interactions with a receptor exists. In this work the use of electrostatic potential minima and maxima in the determination of molecular binding orientations is developed,\textsuperscript{46} with
a much more accurate treatment of the molecular charge distribution, the Distributed Multipole Analysis.\textsuperscript{47, 48}

### 3.4.3 Similarity Using Electrostatic Extrema

This approach assumes that two molecules which bind strongly to the same receptor site must have the majority of their electrostatic interactions, especially hydrogen bond donor and acceptor sites, in common. This method locates all maxima and minima in the electrostatic potential around each molecule at a fixed distance from the van der Waals surface. The region for comparison for this method is dependent on the distance from the van der Waals surface at which the potential is calculated.

Around possible hydrogen bond acceptor sites the proton will lie approximately 0.5 Å outside the van der Waals surface of the acceptor atom. Similarly for hydrogen bond donor groups, the electronegative acceptor group more accurately lies between 1.4-1.5 Å beyond the van der Waals radius of the heavy atom the hydrogen is bonded to. Strong hydrogen bonding would exist for any receptor which had positively charged atoms (\textit{e.g.} polar hydrogen atoms) at the sites of all the minima around the molecule, and negatively charged atoms (\textit{e.g.} hydrogen bond acceptors) at the sites of all maxima in the potential. Realistically steric constraints at the receptor will mean that not all possible interaction sites will be occupied, \textit{i.e.}, there will be no polar receptor atoms at some potential extrema, and some polar receptor atoms will not be located exactly at the maxima or minima.

The problem of defining the area of interest when the van der Waals volume overlay is poor does not exist for this method. The maxima and minima are all close to the molecule, and so the method automatically focuses on possible common binding sites. There are unlikely to be any significant extrema around protruding hydrocarbon side chains, which may either be in steric conflict with the receptor, or fit into a hydrophobic pocket.
3.5 Methods

3.5.1 Ab Initio Calculations

Throughout this work, Self-Consistent Field (SCF) wavefunctions were obtained for each molecule using the CADPAC suite of ab initio programs, and a 3-21G basis set. This basis set was found to give essentially the same locations (to within 0.01 Å) and relative strengths of the electrostatic extrema as a 6-31G** basis set for adenine and theophylline (B. Lucchese, unpublished). This is in agreement with the observation by Price et al. that the electrostatic potential around a dipeptide approximately only changes by a scaling factor with basis set, within the range of reasonable quality basis sets. The Distributed Multipole Analysis (DMA) of each wavefunction was calculated using CADPAC, to represent the charge density by a charge, dipole, quadrupole, octupole and hexadecapole moment at every nuclear position. The inclusion of the anisotropic multipole moments gives a more accurate representation of the non-spherical features of charge distributions. Hydrogen bonds form to regions of lone pair and π electron density, and it is precisely these features of charge density which are accurately represented by distributed multipole models, unlike atomic charges.

3.5.2 Electrostatic Calculations

The electrostatic properties of the molecules studied have been calculated from the DMA model of the charge distribution using the program ORIENT, using all terms in the multipole series up to R^4 between each atom in the molecule and a point charge. The molecular van der Waals volume was defined using the Pauling radii, i.e., C: 2.0 Å, O: 1.4 Å, N: 1.5 Å, Cl: 1.8 Å, P: 1.9 Å. All hydrogens were treated as having a zero van der Waals radius, due to the lack of repulsion between the proton and its acceptor atom in hydrogen bonds. Non-polar hydrogens, such as methyl hydrogens, are included in a 'united atom' carbon radius.

The positions and strengths of minima (maxima) in the electrostatic potential energy were determined by minimising the interaction energy of a single positive (negative) point charge with the molecule under examination, using pseudo hard-sphere repulsion between sites with non-zero van der Waals radii. Initially, all work was performed using point charges of 1.0 Å van der Waals radius. It was found that most
of the extrema in the molecular electrostatic potential were located near potential hydrogen bonding groups, with others arranged above and below the plane of aromatic systems. The extrema close to hydrogen bonding groups were generally of greater magnitude than the other extrema.

When considering maxima and minima in the electrostatic potential as likely sites for hydrogen bond donor and acceptor groups, it is not strictly accurate to examine the potential at a uniform distance from the van der Waals surface. The receptor atoms most likely to be near the polar hydrogen atoms in the ligand are hydrogen bond acceptors, which would be sampling the electrostatic potential about 1.4 Å from the van der Waals surface of the ligand. Similarly, any receptor protons hydrogen bonded to the ligand acceptors will be approximately 0.5 Å from its surface (see figure 2). We wish to study the electrostatic similarity of the ligands in the regions most likely to be occupied by the protein atoms in van der Waals contact with the ligand, particularly those which hydrogen bond to it. Thus, the positions and strengths of the potential maxima and minima were found by minimising the electrostatic interaction energy of a negative charge radius 1.4 Å and a positive charge radius 0.5 Å respectively.

![Diagram of hydrogen bond and van der Waals distances](image)

**Figure 2.** A hydrogen bond proton will be approximately 0.5 Å outside the oxygen van der Waals surface and conversely, an acceptor oxygen will be 1.4 Å outside the van der Waals surface of the donor heavy atom. In order to sample the electrostatic potential in these regions, electrostatic minima have been calculated at 0.5 Å and electrostatic maxima at 1.4 Å outside the molecular van der Waals surface.
3.5.3 *Ligand Superposition and Relative Binding Orientations*

This identifies the points around the substrate and each ligand where there could be the strongest interactions with polar binding site atoms. The next step is to find relative orientations of the substrate and the ligand where the maxima around the ligand are close to maxima in the substrate. This implies that both molecules could interact favourably with any electronegative atom in the binding site at the position of the maxima. Minima are similarly overlaid. Orientations with the greatest number of approximately overlaid pairs of maxima or minima have the greatest electrostatic similarity and so are most likely to represent the relative binding orientation. However, steric constraints will ensure that the polar binding site atoms will not be located exactly at the maxima or minima and that not all possible interaction sites will be occupied. Indeed, since many inhibitors can be much smaller than the appropriate substrate, it is likely that different inhibitors will interact with different binding site residues, in which case we would expect their electrostatic extrema to overlay with a different subset of the substrate extrema.

The superposition of the ligand molecules and their maxima and minima was carried out using the 3D molecular graphics software packages MacroModel\textsuperscript{34} version 3.0, running on an Evans and Sutherland PS330 terminal hosted by a MicroVAX II computer, and Insight\textsuperscript{II}\textsuperscript{35} running on a Silicon Graphics Indigo2 workstation. The electrostatic minima and maxima of each compound are input as dummy atoms. For each inhibitor three, or four, extrema are chosen which provide a visually plausible overlay with the natural substrate or chosen reference molecule. The sets of extrema are superimposed using the standard MacroModel (or Insight\textsuperscript{II}) options for rigid least squares molecular superpositions using atomic positions, but in this case the superpositioned dummy atoms represent electrostatic extrema.

3.5.4 *Assessment of Relative Orientations*

The strength of most hydrogen bonds is quite sensitive to the relative positions of the participants. The quality of the three or four point overlay which results from the superposition is quantified by the minimised root-mean-square separation (RMSD) of the chosen pairs of extrema. The RMSD error in the position of the overlay extrema only indicates the degree of similarity at the chosen points, and so the overall
electrostatic similarity is also qualitatively assessed. A probable overlay model would be expected to align areas of similar electrostatic properties at more than just the three or four pairs of points chosen. With this in mind, the separations of all maxima and minima between the superimposed molecules, which are less than 3.0 Å apart, are calculated. Shorter separations of maxima of the same sign between molecules are assumed to be more favourable, since this makes it more probable that a binding site could form optimum hydrogen bonds with both molecules. It is assumed that two extrema do not correspond to a single hydrogen bond site if they are more than 3.0 Å apart. ‘Clashes’ are defined as situations where a maximum of one molecule is within 3.0 Å of a minimum of the other molecule. These are assumed to be unfavourable as a polar residue in the binding site could not stabilize a region of similar potential in a ligand. Any polar atoms in such a region would improve the binding of one molecule and destabilize the binding of the other.

The effect of any clashes which occur will depend not only on the separation of the corresponding extrema, but also on their relative strengths. A ‘strong’ correspondence is defined as one where both extrema have an electrostatic potential of magnitude greater than 50 kJ mol\(^{-1}\). It is assumed that such an overlay of extrema is highly favourable if the electrostatic potentials are of the same sign, since regions of high potential may be involved in strong electrostatic interactions with the binding site. If the potentials are of opposite sign, then such a clash is considered highly destabilising. An overlay is classed as ‘moderate’ when only one extremum has a potential of magnitude greater than 50 kJ mol\(^{-1}\). ‘Weak’ overlays are those where both extrema have a magnitude less than 50 kJ mol\(^{-1}\). Thus, the degree of similarity is assessed on the basis of the distance between corresponding maxima and minima and their relative strengths.

For each inhibitor, a variety of superpositions relative to the reference ligand are attempted, involving different combinations of extrema. These are assessed by eye, and through the use of RMSD separations. Any relative superpositions which give poor fitting at the chosen maxima and minima (and therefore high RMSD errors), or which result in poor steric overlay of the superimposed molecules are disregarded. Visual inspection is required as, in some cases it is possible to match two sets of extrema that are close to each other (i.e., only focusing on a small region around each of the
molecules), and to generate a very low RMSD value for a superposition with little steric overlap of the molecules. However, when a match is obtained using three or more extrema that are widely spaced around the inhibitor, then a sterically plausible overlay is automatically created. The reported overlays correspond to the lowest RMSD with a sterically feasible overlay.

3.6 Development of Similarity of Electrostatic Extrema
This method for assessing the relative binding orientations of receptor ligands has been applied to, and developed using several different biological systems. In the next chapter a comparison of the natural substrate and inhibitors for the enzyme Phosphodiesterase III is presented. This system was first studied by Davis et al.\textsuperscript{36} and their work will be extended here through the use of a DMA model for the ligand charge distributions and through the inclusion of both electrostatic minima and maxima in the similarity comparison. In chapter five an electrostatic comparison of ligands for the enzyme Glycolate Oxidase is presented. For this system, crystal structure coordinates of the enzyme binding site are available. This binding site will be used to test the hypothesis that electrostatic extrema can be used to locate binding site residues which interact with the ligands. Relative binding orientations produced by matching electrostatic extrema will be compared with binding orientations produced by optimizing the binding interactions of the ligands within the receptor site. Finally, a study of hydrogen bonded complexes between a series of small molecules, containing hydrogen bonding functionalities, and amino acid sidechains is presented in chapter six. These will be used to determine how closely electrostatic extrema predict hydrogen bond geometries for these model systems. The results of these separate studies are followed by a discussion and assessment of the method used here.
<table>
<thead>
<tr>
<th>Similarity Method</th>
<th>Molecular Property Compared</th>
<th>Region of Focus</th>
<th>Conformational Flexibility</th>
<th>Similarity Measure</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>OVID Algorithm¹</td>
<td>Atomic positions</td>
<td>van der Waals volume</td>
<td>No</td>
<td>Weighted overlap integral of atomic van der Waals volumes.</td>
<td>Requires explicit atom-atom assignments for overlay.</td>
</tr>
<tr>
<td>Superposition of Flexible Positions²</td>
<td>Atomic Positions</td>
<td>beyond van der Waals surface</td>
<td>Yes</td>
<td>Least-squares-fitting of atomic positions.</td>
<td>No need for explicit atom-atom assignments. Considers most significantly different molecular conformations.</td>
</tr>
<tr>
<td>Surface Pattern Matching³</td>
<td>Electrostatic Potential</td>
<td>predefined patch on gnomonically projected surface</td>
<td>No</td>
<td>Difference between surface parameters for patch and molecule at points on projected surface.</td>
<td>Requires a regular array of surface points. Needs to define active face of one molecule. Depends on patch size and shape.</td>
</tr>
<tr>
<td>'Blind' Surface Searching⁴</td>
<td>Electrostatic Potential</td>
<td>gnomonically projected surface</td>
<td>No</td>
<td>Difference between surface parameters for molecules at points on projected surface.</td>
<td>Requires a regular array of surface points. Depends on patch shape.</td>
</tr>
<tr>
<td>SUPER Algorithm⁵</td>
<td>Electrostatic Potential</td>
<td>beyond van der Waals surface</td>
<td>No</td>
<td>Weighted overlap integral of electrostatic potentials at points on van der Waals surface.</td>
<td>Searches all possible relative orientations.</td>
</tr>
<tr>
<td>Carbo Molecular Similarity Index⁶</td>
<td>Electron Density</td>
<td>All space</td>
<td>No</td>
<td>Integrals of charge density over all space.</td>
<td>Measures electron density shape similarity, but insensitive to magnitude.</td>
</tr>
<tr>
<td>Hodgkin Molecular Similarity Index⁷</td>
<td>Electrostatic Potential</td>
<td>beyond van der Waals surface</td>
<td>Yes</td>
<td>Integrals of electrostatic potential at grid points.</td>
<td>Measures similarity of shape and magnitude of the electrostatic potential.</td>
</tr>
<tr>
<td>Superposition of Hydrogen Bonding Partners⁸</td>
<td>Hydrogen Bonding Atoms</td>
<td>van der Waals surface</td>
<td>Yes</td>
<td>Least-squares-fitting of expected H-bonding receptor atom positions.</td>
<td>Compares the chemical nature of the molecules. Particularly useful for systems where hydrogen bonding is important.</td>
</tr>
<tr>
<td>Electrostatic Potential Similarity⁹</td>
<td>Electrostatic Potential</td>
<td>beyond van der Waals surface</td>
<td>Yes</td>
<td>Spearman Rank Correlation Coefficient at grid points.</td>
<td>Coincident grid must be recalculated throughout similarity optimization. Considers relative ranks of potential, not the potential itself.</td>
</tr>
<tr>
<td>Electrostatic Field Similarity¹⁰</td>
<td>Electrostatic Field</td>
<td>beyond van der Waals surface</td>
<td>No</td>
<td>Sum of dot product of electrostatic field vectors at grid points.</td>
<td>Ignores area of positive electrostatic potential.</td>
</tr>
<tr>
<td>Similarity of Electrostatic Extrema¹¹</td>
<td>Electrostatic Potential</td>
<td>fixed distance beyond van der Waals surface</td>
<td>No</td>
<td>Root mean-square separation of chosen extrema.</td>
<td>Focuses on specific points where strong ligand-receptor interactions are possible.</td>
</tr>
</tbody>
</table>
References for Chapter 3.


Chapter 4. Phosphodiesterase III Inhibitor Modelling

In this chapter, the method of matching electrostatic extrema has been applied to the phosphodiesterase (PDE) III substrate adenosine-3'5'-cyclic monophosphate (cAMP) and a range of inhibitors. Despite the structural variation between cAMP and the inhibitors, a plausible relative binding orientation can be found for each inhibitor, in which they are sufficiently sterically and electrostatically similar to the natural substrate to account for their affinity for PDE III.

4.1 Introduction

This approach to electrostatic similarity, first used by Davis et al., focuses on the regions where particularly strong electrostatic interactions with the unknown binding site are possible, namely the maxima and minima in the electrostatic potential in the accessible region outside the van der Waals surface of the molecule. Davis et al. found that the overlay of potential minima between various inhibitors and the natural substrate of the phosphodiesterase (PDE) III system was more important than steric factors in determining the biological activity of the inhibitors. This confirms the picture of the binding site which emerges from the experimentally observed effects of substitution on the binding of inhibitors, namely that the binding site is very open, and the major binding force is electrostatic. Hence, this is a suitable system for developing methods for testing for electrostatic similarity.

The aim of this study is to establish whether focusing on the relative disposition of electrostatic extrema is a useful method of assessing possible relative binding orientations of ligands. We use a rigorously defined accurate electrostatic model, that is a distributed multipole (DMA) representation of an ab initio wavefunction of each molecule, to test whether the similarity observed by Davis et al. is an artefact of their approximate method of generating electrostatic properties. If electrostatic similarity is important in determining binding, the similarity should be more obvious from accurate electrostatic properties. We also examine all extrema to assess overall similarity. Davis et al. used semi-empirical point charges to find the minima and maxima, but the maxima did not appear to play a role in defining the binding because they were inaccurately determined. A detailed analysis of the electrostatic properties of the substrate and a
diverse sample of inhibitors is provided, to establish the degree of similarity that is found in this system. The complete range of inhibitors is not studied, since Davis et al. have already established sufficient matching of extrema, and the qualitative conclusions will not be affected by a larger sample.

To be compatible with the original work1 we compared the electrostatic potential around a selection of known inhibitors of PDE III (figure 1) with that of the natural substrate, adenosine 3’5’-cyclic monophosphate (cAMP) (figure 2). For each molecule, we located all the maxima and minima in the electrostatic potential, as calculated from a distributed multipole analysis of an ab initio wavefunction, at fixed distances outside the van der Waals surface of the molecule. In each case it was possible to find a relative orientation of the inhibitor and substrate in which four of the extrema of the inhibitor were within an Angstrom or so of an extremum of the same sign of the substrate. These overlays produced an overall match of the electrostatic and steric properties which are consistent with the inhibitor binding to the enzyme in the same site as cAMP. The distribution of the potential extrema around the substrate is sufficiently complex and sensitive, that it is highly unlikely that such a matching would occur by chance. Thus, empirically, a good matching of the extrema appears to pick out molecules which can bind to the same site, and indeed there are some qualitative correlations with the experimental inhibitory activity. These observations could be readily rationalised by assuming that the matched extrema are close to polar binding groups of the opposite sign in the binding site. However, the positions of the binding site atoms may not be that crucial to the success of the method, as the matching of the extrema does produce a general similarity in the steric and electrostatic properties of the ligands.

4.2 Background
Phosphodiesterase (PDE) enzymes catalyse the hydrolysis of cyclic 3’5’-nucleotides such as adenosine-3’5’-cyclicmonophosphate (cAMP), to the corresponding 5’-monophosphates. Several PDE enzymes have been identified, differing with species and tissue type, and in any tissue there may be several PDE isoenzymes differing in their catalytic requirements and their selectivity for cAMP.

This work involves modelling studies on the natural substrate cAMP, non-specific
PDE inhibitors (compounds which inhibit the activity of more than one PDE isoenzyme), and specific PDE type III inhibitors. PDE inhibitors have potential applications in the treatment of congestive heart failure, and such conditions as asthma. These studies build on previous work by Davis et al\(^1\) which will be reviewed first.

### 4.2.1 Biological Inhibitor Data
Phosphodiesterase isoenzymes were isolated and assayed as described by Reeves et al.\(^4\) IC\(_{50}\) values (the minimum concentration of inhibitor needed to cause a 50% drop in the activity of PDE on cAMP) were obtained by incubation of PDE III from cat ventricle (cardiac muscle) at 1\(\mu\)M cAMP and a wide range of inhibitor concentrations. In the original study by Davis et al.\(^1\) 54 specific PDE III inhibitors were considered, based on 21 common structural templates. The inhibitor compounds chosen for this study are based on three structural templates which represent a range of IC\(_{50}\) values in the original set and includes compounds classified as sterically as ‘short’ and ‘long’ inhibitors (compounds with a longer substituent attached through a nitrogen atom or through a carbon atom).

### 4.2.2 cAMP Conformational Preference
Variations in the conformation of cyclic nucleotides are possible with rotation at the purine-ribose bond and flexibility in the ribose-3’-5’-cyclic phosphate ring system. Davis et al\(^1\) studied compounds containing the ribose-3’5’-cyclic phosphate ring system taken from the Cambridge Crystallographic Database\(^5\) and they found a common conformation for the ribose-3’ 5’-cyclic phosphate group, but significant structural variation due to rotation at the purine-ribose torsional bond angle. The COSMIC\(^6\) force field, allowing for a small adjustment of the minimised coordinates, was able to reproduce the crystal structure of cAMP by molecular mechanics minimisation.

Three elements of conformational change were investigated with the COSMIC SPIN01\(^6\) program by constraining certain torsional angles and, excluding the particular fixed torsional angle, allowing the structure to relax. Protonated (phosphoric acid) and anionic (phosphate) forms of cAMP were both investigated, with the assumption that the true charge state of the phosphate anion (solvated or enzyme-bound) will be between
these extreme states. The torsional angles investigated corresponded to rotation of the
purine-ribose bond, flexibility in the ribose ring, and flexibility in the phosphate ring.

Rotations around the purine-ribose bond of the anionic form of cAMP indicated
a preference for the \textit{C4-N9-ribose-O'} torsional angle between 180° and 250°, the anti-
conformation. This conformation was favoured over the syn-conformer by approximately
9.6 kJ mol\(^{-1}\). For the protonated form the anti-conformation was found to be higher in
energy, but only by 3.3 kJ mol\(^{-1}\). Physical evidence confirms that cAMP prefers to adopt
an anti-conformation.\(^7\) This rotation about the purine-ribose bond is the most
energetically accessible conformational change for cyclic nucleotides. Deviations from
the ribose ring crystal structure caused an increase in energy. A secondary minimum
'twist boat' form for the phosphate ring was found above the minimum 'chair' form for
the anti-form of cAMP. Although this form is energetically accessible to cyclic
nucleotides, it is not supported by crystallographic evidence. In this work the protonated
form of cAMP has been used, in the anti-conformation.

\textit{4.2.3 Previous Electrostatic Modelling}

In their work, Davis et al\(^1\) modelled a range of PDE inhibitors using the computational
software COSMIC.\(^6\) Atom-centred point charges were calculated by the CNDO method
and the electrostatic potential and associated minima were calculated through the
program NBMAP.\(^1\) This molecular mechanics procedure calculates the interaction energy
of a probe group at a series of 68,921 grid points of a cube centred on the test molecule.
The potential plots were uniform despite azinone ring structural differences between the
inhibitors, and the site of deepest electrostatic potential minimum was always associated
with the carbonyl of the cyclic amide group.

The cAMP electrostatic potential was similarly calculated, and the phosphate
group was found to be the centre of electronic charge. Moreover, the isopotential in this
region was similar in extent and magnitude to that found in the azinone rings. This
suggests that the important phosphate group is mimicked at the active site by the
carbonyl ring of the inhibitors. The high affinity PDE III inhibitors were found to share
a common electrostatic potential minimum associated with substituents on the phenyl
ring, with the largest electrostatic minimum of the N1 adenine moiety of cAMP. Other
inhibitor compounds could be positioned to overlay a second common electrostatic potential minimum associated with the N3 adenine moiety. The ‘short’ inhibitors were found to overlay this N3 minimum, while the ‘long’ inhibitors overlaid the adenine N1 minimum, and displayed very high affinities. It was suggested that this implied that an optimal interaction probably occurs through a centre at a greater distance from the cyclic amide group. The significant conclusion of their work was that for this system, structural overlay was less important (and sometimes misleading) than a comparison of the electronic properties of the molecules.

In this chapter further work on PDE inhibitors is reported using high quality ab initio wavefunctions and Distributed Multipole Analysis\(^2\,^3\) to obtain a very accurate model of the molecular charge distribution. This work presented here is an extension of the method used by Davis et al.\(^1\) Further to the method of Davis et al,\(^1\) we define relative orientations not only in terms of matching electrostatic potential minima, but matching electrostatic potential maxima as well. This work has been carried out using a smaller sub-set of the original PDE III inhibitors studied by Davis et al.\(^1\)

4.3 Methods

4.3.1 Geometry of cAMP and PDE Inhibitors

Davis et al.\(^1\) concluded that electrostatic similarity was more important in determining inhibitory action than structural similarity, based on their study of 5 non-specific and 54 specific PDE III inhibitors. To confirm the degree of similarity in this work, four non-specific and eight specific PDE III inhibitors have been chosen. The specific inhibitors are a sub-set of the original data set which includes pairs of molecules which are structurally similar, but have very different inhibitory effectiveness. The structure of each specific and non-specific PDE III inhibitor is given in figure 1.

The assumed anti-geometry of cAMP and the geometry of the specific and non-specific PDE III inhibitors studied here have been obtained by energy minimisation with the software package MacroModel\(^8\) using the AMBER\(^9\) molecular mechanics force field, running on a Silicon Graphics workstation. Initial modelling and minimisation of the specific PDE III inhibitors showed that most had essentially planar conformations. This is in agreement with previous studies\(^10\) which propose a generally flat topology for
specific PDE III inhibitors, and the molecular geometry in the crystal of many inhibitors of phosphodiesterase III.¹¹

Figure 1. Molecules used in this study defining the numbering system for the non-specific PDE III inhibitors, a) adenine, b) guanine, c) theophylline, d) papaverine; and the specific PDE III inhibitors, e) anagrelide, f) dazonone, g) SK&F95800, h) SK&F93741g, i) SK&F93741ac, j) SK&F93505, k) milrinone, l) amrinone.
specific PDE III inhibitors, and the molecular geometry in the crystal of many inhibitors of phosphodiesterase III.$^{11}$

**Figure 1.** Molecules used in this study defining the numbering system for the non-specific PDE III inhibitors, a) adenine, b) guanine, c) theophylline, d) papaverine; and the specific PDE III inhibitors, e) anagrelide, f) dazonone, g) SK&F95800, h) SK&F93741g, i) SK&F93741ac, j) SK&F93505, k) milrinone, l) amrinone.
related compound SK&F93741, the minimum energy conformer has the C1′-N7′ and C8′=O8′ bonds of the acetyl group in a cis conformation, with an inter-ring torsion angle of 11°. An alternative conformation, with the C1′-N7′ and C8′=O8′ bonds in a trans arrangement, and the acetyl group rotated at 50° to the aromatic ring plane has also been studied. The barrier to rotation for the acetyl C=O group was approximately 30 kJ mol⁻¹. The acetyl-rotated structure of SK&F93741 mimics the structure of a third pyridazinone inhibitor, SK&F95800, in which the conformation of the C8′=O8′ bond is fixed as part of a ring closure. To investigate the effect of the acetyl conformation on the electrostatic similarity of these inhibitors to cAMP, both the global energy minimum, labelled SK&F93741g, the acetyl-rotated structure, SK&F93741ac, and SK&F95800 have been used in these calculations.

4.3.2 CADPAC Ab Initio Calculations
For this work SCF wavefunctions were obtained for each molecule using the CADPAC¹⁴ suite of ab initio programs, and a 3-21G basis set¹⁵ in all cases. This basis set was found to give essentially the same location (to within 0.01 Å) and relative strengths of the electrostatic extrema as a 6-31G** basis set for adenine and theophylline.¹⁶ This is in agreement with the observation by Price et al.¹⁷ that the electrostatic potential around a dipeptide approximately only changes by a scaling factor with basis set, within the range of reasonable quality basis sets. The Distributed Multipole Analysis²³ of each wavefunction was calculated to represent the charge density by a charge, dipole, quadrupole, octupole and hexadecapole moment at every nuclear position. The inclusion of the anisotropic multipole moments gives a more accurate representation of the non-spherical features of charge distributions. Hydrogen bonds form to regions of lone pair and π electron density, and it is precisely these features of charge density which are accurately represented by distributed multipole models, unlike atomic charges.

4.3.3 ORIENT2 Electrostatic Calculations
The electrostatic potential surface was calculated from the DMA model of the charge distribution using the program ORIENT2,¹⁸¹⁹ using all terms in the multipole series up to R⁻⁵ between each atom in the molecule and a point charge. The electrostatic potential
surface was first calculated on a grid of points at a distance of 1.0 Å from the molecular van der Waals surface for visualisation. The molecular van der Waals volume was defined using the Pauling radii, *i.e.*, C: 2.0 Å, O: 1.4 Å, N: 1.5 Å, Cl: 1.8 Å, P: 1.9 Å. All hydrogens were treated as having a zero van der Waals radius, due to the lack of repulsion between the proton and its acceptor atom in hydrogen bonds. Non-polar hydrogens, such as methyl hydrogens, are included in a 'united atom’ carbon radius.

The positions and strengths of minima (maxima) in the electrostatic potential energy were determined by minimising the interaction energy of a single positive (negative) point charge with the molecule under examination, using pseudo hard-sphere repulsion between sites with non-zero van der Waals radii. Initially, all work was performed using point charges of 1.0 Å van der Waals radius. It was found that most of the extrema in the molecular electrostatic potential were located near potential hydrogen bonding groups, with others arranged above and below the plane of aromatic systems. The extrema close to hydrogen bonding groups were generally of greater magnitude than the other extrema. Thus, to sample the electrostatic potential in the region around such groups, the positions and strengths of the potential maxima and minima were found by minimising the electrostatic interaction energy of a negative charge radius 1.4 Å and a positive charge radius 0.5 Å respectively.

This identified the points around cAMP and each ligand where there could be the strongest interactions with polar binding site atoms. We then sought relative orientations of cAMP and the ligand where the maxima around the ligand are close to maxima in cAMP. This implies that both molecules could interact favourably with any electronegative atom in the binding site at the position of the maxima. Minima were similarly overlaid. Since many of the inhibitors are much smaller than cAMP, it is likely that different inhibitors will interact with different binding site residues, in which case we would expect their electrostatic extrema to overlay with a different subset of cAMP extrema.

The superposition of the ligand molecules and their maxima and minima was carried out using the 3D molecular graphics software MacroModel version 3.0,* running on an Evans and Sutherland PS330 terminal hosted by a MicroVAX II computer. For each inhibitor three, or four, extrema were chosen which would provide a visually
plausible overlay with cAMP. The quality of the resulting three or four point overlay was quantified by the minimised root-mean-square separation (RMSD) of the chosen pairs of extrema.

The RMSD error in the position of the overlay extrema only indicates the degree of similarity at the chosen points, and so the overall electrostatic similarity was also qualitatively assessed by calculating the separations of all maxima and minima (between the superimposed molecules) which were less than 3.0 Å apart. Clashes were defined as situations where a maximum of one molecule was within 3.0 Å of a minimum of the other molecule. A 'strong' correspondence is defined was one where both extrema had an electrostatic potential of magnitude greater than 50 kJ mol⁻¹. An overlay was classed as 'moderate' when only one extremum has a potential of magnitude greater than 50 kJ mol⁻¹. 'Weak' overlays were those where both extrema have a magnitude less than 50 kJ mol⁻¹. Thus the degree of similarity was assessed on the basis of the distance between corresponding maxima and minima and their relative strengths. For each inhibitor, a variety of superpositions relative to cAMP were attempted, involving different combinations of extrema. These were assessed by eye, and through the use of RMSD separations. The reported overlay corresponds to the lowest RMSD with a sterically feasible overlay.

In the following sections the results of comparing PDE III inhibitors and the natural substrate cAMP by the similarity of their electrostatic extrema will be presented. The four point relative binding orientations of the inhibitors with respect to cAMP are presented in figures 3 to 9. The details of the overlay models proposed for each inhibitor are recorded in tables 1 and 2. In each case an overlay based on matching three pairs of maxima and/or minima sites is given, and where possible a subsequent four point match is also included. All pairs of maxima and minima within 3.0 Å of each other are listed. The extrema used as matching points are listed above the dashed line.
4.4 Results

cAMP, in the assumed anti-conformation has a complex electrostatic potential surface (see figure 3b). The deepest negative electrostatic potential minimum occurs near the O6' of the phosphate group. Three other important potential minima are arranged around N1, N3 and N7 of the adenosine ring. The main regions of positive electrostatic potential occur with maxima near N6 of the purine ring and around O7' and O2' of the sugar ring. Despite the complexity of the electrostatic potential surface around cAMP, it was possible to match four extrema of each specific inhibitor very closely with those of cAMP, to produce a small RMSD (tables 1 and 2). This overlay generally resulted in other pairs of extrema (with the same sign) coming within 3.0 Å of each other, though there were some potentially destabilising clashes for the poorer inhibitors.

4.4.1 Specific PDE III Inhibitors

For the specific PDE III inhibitors, the deepest electrostatic potential minimum was generally found near the cyclic amide group. Anagrelide is one exception, where the deepest electrostatic minimum was found between the two chlorine atoms, in the molecular plane. All the overlays of the specific PDE III inhibitors reported in table 1 match the deep cAMP O6'_{\text{min}} (−206.8 kJ mol\(^{-1}\)), this is in agreement with a model of the cAMP binding site proposed by Moos et al.,\(^{12}\) which is based on the ability of the inhibitor cyclic amide function to occupy the cAMP 5'-phosphate region in the binding site. Three of the best inhibitors, anagrelide, SK&F95800, and dazonone, all match the O1'_{\text{min}} (−73.5 kJ mol\(^{-1}\)) of cAMP. In each case the inhibitor compounds match the position of this cAMP minimum with a minimum from the \(\pi\) electron density of the aromatic ring system which these molecules have in common. Such electrostatic minima are generally poorly represented by point charge models.

Anagrelide

This compound is the best inhibitor of the original sample studied by Davis et al.\(^{1}\) The four point relative orientation proposed in table 1, and shown in figure 3a, produces many favourable corresponding pairs of extrema. While the electrostatic minima associated with Cl3 and Cl4 do not indicate possible hydrogen bonding groups, they are
still regions of significant electrostatic potential which can be matched with the
electrostatic potential of cAMP. The region of positive potential above and below the
plane of the cyclic amide of anagrelide matches the large positive region on the
electrostatic potential surface of cAMP (see figure 3b). The activity of anagrelide, and
its analogue dazonone, has been attributed by Venuti et al.\(^{39}\) to the ability of the cyclic
amide group to mimic the cAMP phosphate, and to the aromatic \(\pi\) electron density of
the inhibitors, which provide a further strong interaction in the absence of hydrogen
bonding substituents at C3 and C4. For anagrelide, this interaction is identified by the
overlay of the minimum \(\text{Ar}^\Delta\) with \(\text{O1}'\) of cAMP. Thus the matching of electrostatic
extrema has reproduced the overlay proposed on the grounds that structural elements of
the specific PDE III inhibitors can mimic the anti-conformation of cAMP.

**Dazonone**

This inhibitor is structurally similar to anagrelide, but the electrostatic potential surface
differs in some regions (see figure 4a). Dazonone has a single Cl atom, and the
associated electrostatic minima are arranged symmetrically above and below the
molecular plane, rather than in the plane as in anagrelide. A region of negative
electrostatic potential in anagrelide around the aromatic ring near Cl3 is replaced by a
region of positive potential in dazonone, which results in the weak potential clash seen
in table 1 for this relative orientation. The proposed relative orientation of dazonone is
fairly similar to that of anagrelide, despite the explicitly matched extrema being
somewhat different.

**SK&F95800**

This is the second best inhibitor in this sample and it overlays cAMP with a relatively
good RMSD value. The first four moderate clashes listed in table 1 all occur in one small
region of the overlay. These result from the close SK&F95800 double maxima \(\text{C10}'–\text{H}_{\max}\)
and \(\text{C5}'–\text{H}_{\max}\), which lie near the negative potential of cAMP \(\text{N3}_{\min}\) and \(\text{O2}'_{\min}\), in this
orientation. In addition there are two strongly clashing regions around N7 and O2' in the
cAMP electrostatic potential, which are approximately 2.0 Å from SK&F95800
electrostatic extrema in each case.
SK&F93741

The low energy conformation, SK&F93741g, produces a good RMSD value, with four favourable overlay points separated by >1.0 Å. However a single strong clash appears at the 3.0 Å limit. The best overlay of the alternative conformation (with the acetyl group rotated in a trans conformation, to mimic SK&F95800), SK&F93741ac, matches different points and gives a comparable RMSD value, with only two moderate clashes. SK&F95800 is a better inhibitor than SK&F93741, despite lacking a methyl group which is known to increase inhibitory activity. Hence it is not necessary to assume that SK&F93741 adopts the higher energy conformation, analogous to SK&F95800, in order to have an electrostatically favourable binding orientation.

SK&F93505

This compound is structurally related to SK&F93741, differing in the lack of an acetyl group at N7', but is a much poorer PDE III inhibitor. The relative orientation shown in table 1 satisfies potential hydrogen bonding sites at the four chosen points. However there are three strong and potentially destabilising clashes, all separated by more than 2.0 Å. Not all extrema will have corresponding complementary extrema at the receptor. These clashes may be important.

Milrinone

Milrinone is shown in figure 7a, in its proposed relative binding orientation with the four matching extrema indicated. Milrinone overlays primarily with electrostatic maxima of cAMP, with no clashes of regions of opposite potential. Figure 7b shows the electrostatic potential surfaces of milrinone and cAMP in their proposed relative binding orientations. Similarity between the electrostatic potential surfaces can be seen once the molecules are correctly orientated. The electrostatic matching for this inhibitor is particularly good in the regions of positive electrostatic potential, the explicitly matched extrema are clustered around the sugar residue of cAMP. As a result, in this orientation milrinone does not match at the cAMP N3 minima, as proposed previously for these 'short' inhibitors, but lies slightly forward of the plane of the cAMP purine ring, with the

80
negative potential areas $N_{1_{\text{min}}}$ and $N_{6^A_{\text{min}}}$ coinciding. The relative orientation reported here is in agreement with earlier findings that non-planar conformations in certain molecules allow for more favourable matching with portions of the cAMP sugar regions, by ‘introducing a favourable electrostatic potential in the vicinity of the 2′-hydroxyl to compensate for non-planarity’.\textsuperscript{12} We identify this favourable electrostatic overlap as the coincidence of the cAMP maximum $O_{2}^{'-H_{\text{max}}}$ and milrinone maximum $C_{4-H_{\text{max}}}$. This accounts for the earlier difficulties of Davis et al\textsuperscript{1} in describing the binding of milrinone, as they neglected the potential maxima because of their crude method of calculation.

**Amrinone**

This is the worst overall inhibitor of the original set, where a reasonable match of electrostatic properties is still possible at some points, but at the expense of clashes elsewhere. The lowest RMSD is quite high, and these overlay points result from a string of electrostatic minima arranged symmetrically above and below the amrinone molecular plane. Closer examination shows that these minima are clustered around cAMP $N_{3_{\text{max}}}$ ($\sim 132.1$ kJ mol\textsuperscript{−1}) or $O_{2}^{'-H_{\text{max}}}$ ($\sim 105.0$ kJ mol\textsuperscript{−1}), so they do not represent several potential ligand-binding site interaction points, but rather indicate similar electrostatic properties existing for both molecules across this region. There are several potential clashes, including two strong extrema of opposite signs within 2.0 Å of each other. Such potentially destabilising clashes feature in all plausible superpositions for amrinone, and so it would not be expected to bind to the same site as cAMP very strongly, which correlates with its low inhibitory strength.

### 4.4.2 Non-Specific PDE Inhibitors

The non-specific PDE inhibitors are expected to bind in the receptor region occupied by the purine moiety of cAMP, and indeed many of the non-specific inhibitors are structurally similar to the purine ring system. The results for the non-specific PDE inhibitors can be seen in table 2. The non-specific PDE inhibitors can all overlay at least three electrostatic extrema associated with the purine moiety of cAMP. Given the structural similarity of many non-specific PDE inhibitors to the purine ring system, this might be supposed from a comparison of molecular structures. Matching extrema for
adenine does, trivially, result in a direct overlay of the adenosine ring (figure 8a). However, a direct ring system overlay of the inhibitors onto cAMP does not always necessarily produce the best agreement in electrostatic properties. This is particularly apparent in the case of theophylline. The best overlay reported in table 2 corresponds approximately to the ‘N^9-C8’ model introduced by Peet et al. and investigated by van der Wenden et al. In this overlay the N1, N3 and N9 atoms of theophylline are superimposed onto N9, N3 and N1 of cAMP. We have closely approximated this model by matching the relevant electrostatic maxima and minima, rather than atomic positions (figure 9a). All the non-specific inhibitors studied here can overlay at least three extrema surrounding the purine ring of cAMP with few clashes, in general agreement with Davis et al.

4.5 Conclusions

The use of accurate electrostatic models has demonstrated that there is far more correspondence between the electrostatic properties of the substrate and inhibitors of phosphodiesterase III than was noted in the earlier work, which used minima only. The observation by Davis et al. that matching a couple of electrostatic minima provides a rationale for determining relative binding orientations has been extended into a new method of comparing electrostatic similarity, which could be applied to other systems where such effects are dominant. By optimising the overlay of three or four electrostatic extrema of the natural substrate and each inhibitor, it is possible to find an overlay which is a plausible relative binding orientation, being both sterically feasible and having a close correspondence between potential hydrogen bonding sites and other regions where strong electrostatic interactions with the binding site are possible. This success for such a structurally diverse range of inhibitors increases our confidence that the proposed overlays may correspond to the relative binding orientations within the enzyme. The complexity of the three dimensional arrangement of the 6 maxima and 8 minima around cAMP, suggests that it is unlikely that a plausible overlay could be found for many molecules, and therefore that the method could be a useful screen in designing new inhibitors.

This approach of matching the potential extrema is a distillation of the more
general belief that electrostatic potential fields play a vital role in molecular recognition processes. If the matched extrema are spread out around the inhibitor, then this approach will also ensure that there is reasonable steric overlay. We also observe that such overlays of known inhibitors usually have similar electrostatic properties in other regions, in that extrema not used in the match are generally located in regions of the correct sign. This suggests that the approach is physically reasonable and may be matching points where there is strong interaction with the binding site. Certainly the method is matching potential hydrogen bonding sites. Indeed, the common appearance of some cAMP electrostatic extrema in many overlays, specifically O6'_{min} and O2'_{min}, suggests that there are appropriate polar residues in this region in the binding site.

The uncertainty of how close an overlay of extrema is necessary for binding is less important than the uncertainty as to whether there is a suitable polar group in the vicinity of a potential extrema match or clash. Thus, like all methods of comparing ligands, it is unlikely to provide a good quantitative correlation with even the binding energy of the ligand to the protein, let alone more remote experimental quantities such as inhibitor strengths. Nevertheless, there are qualitative correlations. The method is being automated to consider all possible sets of extrema for matching.

The use of DMA electrostatic models has made a considerable difference to the numbers of extrema that match. Davis et al\(^{1}\) only matched two minima for the specific inhibitors since the their CNDO semi-empirical atomic charge calculations severely underestimated the magnitude of the maxima and did not use any minima above the aromatic rings. Thus, a reliable method of calculating the electrostatic properties which reflects the anisotropy of the atomic electron distributions is essential. This can be done by calculating the electrostatic properties direct from the wavefunction\(^{23}\) but the DMA method is more practical for large systems, and those with a high degree of conformational flexibility.

In applying this method to systems with more structural flexibility, it will be necessary to consider conformations other than the global minimum. It is possible to calculate the extrema in other conformations readily from a DMA, without performing another \textit{ab initio} calculation, by neglecting changes in the atomic charge distribution which occur with conformation. The errors in this approximation are unlikely to be
significant for this type of similarity comparison. However, for large molecules where many conformations or induced movement has to be taken into account, even a DMA \textit{ab initio} based method will be computationally too demanding. Thus, methods of representing the lone pair and \( \pi \) electron density by additional point charges, without requiring the calculation of a wavefunction, are being developed. The anisotropic distribution of electron density in atoms contributes to the exquisite sensitivity of electrostatic extrema to both molecular structure and conformation which helps explain the huge variations in biological interactions.

It has been observed that it is possible to match the electrostatic extrema of a diverse range of inhibitors onto those of the natural substrate, producing a relative orientation of the molecules which could both favourably interact with the same unknown binding site. Thus, the approach could be useful in designing new ligands in a range of systems where the natural substrate or a good lead compound is known. This will be established by the application of the approach to other systems in the following chapters. Chapter 5 presents a study of the intermolecular interactions within the Glycolate Oxidase enzyme binding site.
Figure 3. a) 2D projection of the four-point relative binding orientation of cAMP (grey) and anagrelide (full colour). The four explicitly matched pairs of extrema are highlighted as bold crosses, and their separation is indicated. Key: Minima $+$ (cAMP), $+$ (anagrelide); Maxima $+$ (cAMP), $+$ (anagrelide).

b) The electrostatic potential at 1.0 Å from the van der Waals surface around cAMP and anagrelide in their proposed four-point relative binding orientation. The potential $V$ (kJ mol$^{-1}$) is calculated from the DMA of a 3-21G wavefunction, colour coded as follows:

white $< -120 <$ grey $< -90 <$ magenta $< -60 <$ blue $< -30 <$ cyan $< 0 <$ green $< 30 <$ yellow $< 60 <$ orange $< 90 <$ brown $< 120 <$ red.
Figure 4. 2D projection of the four-point relative binding orientation of cAMP and the specific PDE III inhibitors, a) dazonone, and b) SK&F95800. The four explicitly matched pairs of extrema are highlighted as bold crosses, and their separation is indicated. For clarity, cAMP is rendered in grey only, and each inhibitor is shown in full colour.

Key: Minima \( \bullet \) (cAMP), \( + \) (inhibitor); Maxima \( \bullet \) (cAMP), \( + \) (inhibitor).
**Figure 5.** 2D projection of the four-point relative binding orientation of cAMP and the specific PDE III inhibitors, a) SK&F93741g, and b) SK&F93741ac. The four explicitly matched pairs of extrema are highlighted as bold crosses, and their separation is indicated. For clarity, cAMP is rendered in grey only, and each inhibitor is shown in full colour.

Key: Minima + (cAMP), + (inhibitor); Maxima + (cAMP), + (inhibitor).
Figure 6. The four-point relative binding orientation of cAMP and the specific PDE III inhibitors, a) SK&F93505, and b) amrinone. The four explicitly matched pairs of extrema are highlighted as bold crosses, and their separation is indicated. For clarity, cAMP is rendered in grey only, and each inhibitor is shown in full colour.

Key: Minima + (cAMP), + (inhibitor); Maxima + (cAMP), + (inhibitor).
Figure 7. a) 2D projection of the four-point relative binding orientation of cAMP (grey) and milrinone (full colour). The four explicitly matched pairs of extrema are highlighted as bold crosses, and their separation is indicated. Key: Minima + (cAMP), + (milrinone); Maxima + (cAMP), (milrinone).
b) The electrostatic potential at 1.0 Å from the van der Waals surface around cAMP and milrinone in their proposed four-point relative binding orientation. The potential V (kJ mol⁻¹) is calculated from the DMA of a 3-21G wavefunction, colour coded as follows:
white < -120 < grey < -90 < magenta < -60 < blue < -30 < cyan < 0 < green < 30 < yellow < 60 < orange < 90 < brown < 120 < red.
Figure 8. The three-point relative binding orientation of cAMP and the non-specific PDE III inhibitors, a) adenine, and b) guanine. The three explicitly matched pairs of extrema are highlighted as bold crosses, and their separation is indicated. For clarity, cAMP is rendered in grey only, and each inhibitor is shown in full colour. Note that adenine is almost exactly superimposed on the adenine moiety of cAMP.

Key: Minima + (cAMP), + (inhibitor); Maxima + (cAMP), + (inhibitor).
Figure 9. The three-point relative binding orientation of cAMP and the non-specific PDE III inhibitor, a) theophylline, and b) the four-point relative binding orientation of papaverine. The three (or four) explicitly matched pairs of extrema are highlighted as bold crosses, and their separation is indicated. For clarity, cAMP is rendered in grey only, and each inhibitor is shown in full colour.

Key: Minima + (cAMP), + (inhibitor); Maxima + (cAMP), + (inhibitor).
Table 1. Structure and proposed four point relative binding orientation of the specific PDE III inhibitors.

<table>
<thead>
<tr>
<th>Inhibitor</th>
<th>IC&lt;sub&gt;50&lt;/sub&gt; (µM)</th>
<th>CORRESPONDING MAXIMA AND MINIMA</th>
<th>Same</th>
<th>Strength</th>
<th>Separation</th>
<th>RMSD OF POINTS IN MATCH&lt;sup&gt;a&lt;/sup&gt;</th>
<th>4 point</th>
<th>(3 point)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Nearest Inhibitor</td>
<td>Nearest cAMP</td>
<td>Nearest</td>
<td>Same</td>
<td>Sign</td>
<td>Strength</td>
<td>Separation</td>
</tr>
<tr>
<td>Anagrelide</td>
<td>0.13</td>
<td>O12&lt;sub&gt;an&lt;/sub&gt;</td>
<td>-73.3</td>
<td>O6&lt;sub&gt;an&lt;/sub&gt;</td>
<td>-206.8</td>
<td>✓</td>
<td>strong</td>
<td>0.688</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N7&lt;sub&gt;an&lt;/sub&gt;</td>
<td>-116.4</td>
<td>O2&lt;sub&gt;an&lt;/sub&gt;</td>
<td>-105.0</td>
<td>✓</td>
<td>strong</td>
<td>0.377</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ar&lt;sup&gt;a&lt;/sup&gt;</td>
<td>-88.2</td>
<td>-73.5</td>
<td>✓</td>
<td>strong</td>
<td>0.835</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>C13&lt;sub&gt;an&lt;/sub&gt;</td>
<td>-46.6</td>
<td>N6&lt;sub&gt;an&lt;/sub&gt;</td>
<td>-16.7</td>
<td>✓</td>
<td>weak</td>
<td>0.376</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C14&lt;sub&gt;an&lt;/sub&gt;</td>
<td>-93.6</td>
<td>N7&lt;sub&gt;an&lt;/sub&gt;</td>
<td>-102.2</td>
<td>✓</td>
<td>strong</td>
<td>1.352</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C14&lt;sub&gt;hs&lt;/sub&gt;</td>
<td>+62.5</td>
<td>O7&lt;sub&gt;hs&lt;/sub&gt;</td>
<td>+126.9</td>
<td>✓</td>
<td>strong</td>
<td>2.359</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C13&lt;sub&gt;ha&lt;/sub&gt;</td>
<td>+58.5</td>
<td>C3&lt;sub&gt;ha&lt;/sub&gt;</td>
<td>+86.5</td>
<td>✓</td>
<td>strong</td>
<td>3.013</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ar&lt;sup&gt;a&lt;/sup&gt;</td>
<td>-88.2</td>
<td>C2&lt;sub&gt;ha&lt;/sub&gt;</td>
<td>+56.4</td>
<td>✓</td>
<td>strong</td>
<td>2.723</td>
</tr>
<tr>
<td>Dazonone</td>
<td>1.68</td>
<td>O17&lt;sub&gt;an&lt;/sub&gt;</td>
<td>-143.9</td>
<td>O6&lt;sub&gt;an&lt;/sub&gt;</td>
<td>-206.8</td>
<td>✓</td>
<td>strong</td>
<td>0.423</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N7&lt;sub&gt;an&lt;/sub&gt;</td>
<td>-112.1</td>
<td>O2&lt;sub&gt;an&lt;/sub&gt;</td>
<td>-105.0</td>
<td>✓</td>
<td>strong</td>
<td>0.545</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ar&lt;sup&gt;a&lt;/sup&gt;</td>
<td>-54.3</td>
<td>O1&lt;sub&gt;an&lt;/sub&gt;</td>
<td>-73.5</td>
<td>✓</td>
<td>strong</td>
<td>0.851</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C13&lt;sub&gt;an&lt;/sub&gt;</td>
<td>-40.4</td>
<td>N7&lt;sub&gt;an&lt;/sub&gt;</td>
<td>-102.2</td>
<td>✓</td>
<td>moderate</td>
<td>0.439</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C13&lt;sub&gt;hs&lt;/sub&gt;</td>
<td>+23.1</td>
<td>C2&lt;sub&gt;hs&lt;/sub&gt;</td>
<td>+56.4</td>
<td>✓</td>
<td>moderate</td>
<td>2.049</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N7&lt;sub&gt;an&lt;/sub&gt;</td>
<td>-112.1</td>
<td>N3&lt;sub&gt;an&lt;/sub&gt;</td>
<td>-132.1</td>
<td>✓</td>
<td>strong</td>
<td>2.313</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C13&lt;sub&gt;ha&lt;/sub&gt;</td>
<td>+46.3</td>
<td>C8&lt;sub&gt;ha&lt;/sub&gt;</td>
<td>+86.5</td>
<td>✓</td>
<td>moderate</td>
<td>2.798</td>
</tr>
<tr>
<td>SK&amp;F95800</td>
<td>0.76</td>
<td>O6&lt;sub&gt;an&lt;/sub&gt;</td>
<td>-152.3</td>
<td>O6&lt;sub&gt;an&lt;/sub&gt;</td>
<td>-206.8</td>
<td>✓</td>
<td>strong</td>
<td>1.123</td>
</tr>
<tr>
<td></td>
<td></td>
<td>O8&lt;sub&gt;an&lt;/sub&gt;</td>
<td>-167.0</td>
<td>N1&lt;sub&gt;an&lt;/sub&gt;</td>
<td>-118.1</td>
<td>✓</td>
<td>strong</td>
<td>0.984</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ar&lt;sup&gt;a&lt;/sup&gt;</td>
<td>-49.7</td>
<td>O1&lt;sub&gt;an&lt;/sub&gt;</td>
<td>-73.5</td>
<td>✓</td>
<td>moderate</td>
<td>1.380</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C2&lt;sub&gt;an&lt;/sub&gt;</td>
<td>+63.9</td>
<td>C8&lt;sub&gt;an&lt;/sub&gt;</td>
<td>+86.5</td>
<td>✓</td>
<td>strong</td>
<td>0.425</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N2&lt;sub&gt;an&lt;/sub&gt;</td>
<td>-50.1</td>
<td>O2&lt;sub&gt;an&lt;/sub&gt;</td>
<td>-105.0</td>
<td>✓</td>
<td>strong</td>
<td>2.218</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N7&lt;sub&gt;an&lt;/sub&gt;</td>
<td>+84.7</td>
<td>N6&lt;sub&gt;an&lt;/sub&gt;</td>
<td>+65.8</td>
<td>✓</td>
<td>strong</td>
<td>2.608</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C10&lt;sub&gt;an&lt;/sub&gt;</td>
<td>+33.9</td>
<td>O2&lt;sub&gt;an&lt;/sub&gt;</td>
<td>-105.0</td>
<td>✓</td>
<td>moderate</td>
<td>2.192</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C10&lt;sub&gt;hs&lt;/sub&gt;</td>
<td>+33.9</td>
<td>N3&lt;sub&gt;an&lt;/sub&gt;</td>
<td>-132.1</td>
<td>✓</td>
<td>moderate</td>
<td>2.629</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C5&lt;sub&gt;an&lt;/sub&gt;</td>
<td>+33.9</td>
<td>O2&lt;sub&gt;an&lt;/sub&gt;</td>
<td>-105.0</td>
<td>✓</td>
<td>moderate</td>
<td>1.537</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C5&lt;sub&gt;hs&lt;/sub&gt;</td>
<td>+33.9</td>
<td>N3&lt;sub&gt;an&lt;/sub&gt;</td>
<td>-132.1</td>
<td>✓</td>
<td>moderate</td>
<td>2.585</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N7&lt;sub&gt;an&lt;/sub&gt;</td>
<td>+84.7</td>
<td>N7&lt;sub&gt;an&lt;/sub&gt;</td>
<td>-102.2</td>
<td>✓</td>
<td>strong</td>
<td>1.986</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N7&lt;sub&gt;hs&lt;/sub&gt;</td>
<td>+84.7</td>
<td>N6&lt;sub&gt;an&lt;/sub&gt;</td>
<td>-16.7</td>
<td>✓</td>
<td>moderate</td>
<td>2.333</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ar&lt;sup&gt;a&lt;/sup&gt;</td>
<td>-44.7</td>
<td>C2&lt;sub&gt;an&lt;/sub&gt;</td>
<td>+56.4</td>
<td>✓</td>
<td>moderate</td>
<td>2.359</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N2&lt;sub&gt;an&lt;/sub&gt;</td>
<td>-50.1</td>
<td>O2&lt;sub&gt;an&lt;/sub&gt;</td>
<td>-105.0</td>
<td>✓</td>
<td>strong</td>
<td>2.218</td>
</tr>
</tbody>
</table>

*a:* Additional close extrema
Table 1. (Continued)

<table>
<thead>
<tr>
<th>Inhibitor</th>
<th>IC&lt;sub&gt;50&lt;/sub&gt; (µM)</th>
<th>CORRESPONDING MAXIMA AND MINIMA</th>
<th>Same Sign</th>
<th>Strength</th>
<th>Separation /Å</th>
<th>RMSD OF POINTS IN MATCH&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Nearest Inhibitor Atom</td>
<td>Nearest cAMP Atom</td>
<td>V /kJ mol&lt;sup&gt;-1&lt;/sup&gt;</td>
<td>V /kJ mol&lt;sup&gt;-1&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>SK&amp;F93741g</td>
<td>0.90</td>
<td>O&lt;sub&gt;6&lt;/sub&gt;&lt;sup&gt;3&lt;/sup&gt;</td>
<td>O&lt;sub&gt;6&lt;/sub&gt;&lt;sup&gt;3&lt;/sup&gt;</td>
<td>-162.0</td>
<td>-206.8</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td></td>
<td>O&lt;sub&gt;8&lt;/sub&gt;&lt;sup&gt;5&lt;/sup&gt;</td>
<td>N&lt;sub&gt;7&lt;/sub&gt;&lt;sup&gt;2&lt;/sup&gt;</td>
<td>-164.8</td>
<td>-102.2</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N&lt;sub&gt;2&lt;/sub&gt;&lt;sup&gt;3&lt;/sup&gt;</td>
<td>O&lt;sub&gt;2&lt;/sub&gt;&lt;sup&gt;3&lt;/sup&gt;</td>
<td>-85.3</td>
<td>-105.0</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C&lt;sub&gt;3&lt;/sub&gt;&lt;sup&gt;-3&lt;/sup&gt;</td>
<td>C&lt;sub&gt;2&lt;/sub&gt;&lt;sup&gt;-2&lt;/sup&gt;</td>
<td>+33.4</td>
<td>+56.4</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N&lt;sub&gt;7&lt;/sub&gt;&lt;sup&gt;-2&lt;/sup&gt;</td>
<td>N&lt;sub&gt;7&lt;/sub&gt;&lt;sup&gt;-2&lt;/sup&gt;</td>
<td>-64.6</td>
<td>-132.1</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C&lt;sub&gt;7&lt;/sub&gt;&lt;sup&gt;-2&lt;/sup&gt;</td>
<td>N&lt;sub&gt;7&lt;/sub&gt;&lt;sup&gt;-2&lt;/sup&gt;</td>
<td>+30.6</td>
<td>+126.9</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N&lt;sub&gt;7&lt;/sub&gt;&lt;sup&gt;-2&lt;/sup&gt;</td>
<td>N&lt;sub&gt;7&lt;/sub&gt;&lt;sup&gt;-2&lt;/sup&gt;</td>
<td>+113.5</td>
<td>-118.1</td>
<td>x</td>
</tr>
<tr>
<td>SK&amp;F93741ac</td>
<td>0.90</td>
<td>O&lt;sub&gt;6&lt;/sub&gt;&lt;sup&gt;3&lt;/sup&gt;</td>
<td>O&lt;sub&gt;6&lt;/sub&gt;&lt;sup&gt;3&lt;/sup&gt;</td>
<td>-149.8</td>
<td>-206.8</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td></td>
<td>O&lt;sub&gt;8&lt;/sub&gt;&lt;sup&gt;5&lt;/sup&gt;</td>
<td>N&lt;sub&gt;7&lt;/sub&gt;&lt;sup&gt;2&lt;/sup&gt;</td>
<td>-176.3</td>
<td>-118.1</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N&lt;sub&gt;2&lt;/sub&gt;&lt;sup&gt;3&lt;/sup&gt;</td>
<td>O&lt;sub&gt;2&lt;/sub&gt;&lt;sup&gt;3&lt;/sup&gt;</td>
<td>-99.5</td>
<td>-105.0</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C&lt;sub&gt;3&lt;/sub&gt;&lt;sup&gt;-3&lt;/sup&gt;</td>
<td>C&lt;sub&gt;8&lt;/sub&gt;&lt;sup&gt;-3&lt;/sup&gt;</td>
<td>+61.1</td>
<td>+65.4</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C&lt;sub&gt;5&lt;/sub&gt;&lt;sup&gt;-3&lt;/sup&gt;</td>
<td>O&lt;sub&gt;7&lt;/sub&gt;&lt;sup&gt;-3&lt;/sup&gt;</td>
<td>+44.4</td>
<td>+126.9</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C&lt;sub&gt;5&lt;/sub&gt;&lt;sup&gt;-3&lt;/sup&gt;</td>
<td>C&lt;sub&gt;2&lt;/sub&gt;&lt;sup&gt;-2&lt;/sup&gt;</td>
<td>+44.4</td>
<td>+64.3</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N&lt;sub&gt;7&lt;/sub&gt;&lt;sup&gt;-2&lt;/sup&gt;</td>
<td>N&lt;sub&gt;6&lt;/sub&gt;&lt;sup&gt;-2&lt;/sup&gt;</td>
<td>+65.4</td>
<td>-11.8</td>
<td>x</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C&lt;sub&gt;3&lt;/sub&gt;&lt;sup&gt;-3&lt;/sup&gt;</td>
<td>C&lt;sub&gt;8&lt;/sub&gt;&lt;sup&gt;-3&lt;/sup&gt;</td>
<td>+61.1</td>
<td>+65.4</td>
<td>✓</td>
</tr>
<tr>
<td>SK&amp;F93505</td>
<td>37.00</td>
<td>O&lt;sub&gt;6&lt;/sub&gt;&lt;sup&gt;3&lt;/sup&gt;</td>
<td>O&lt;sub&gt;6&lt;/sub&gt;&lt;sup&gt;3&lt;/sup&gt;</td>
<td>-168.5</td>
<td>-206.8</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N&lt;sub&gt;1&lt;/sub&gt;&lt;sup&gt;-1&lt;/sup&gt;</td>
<td>O&lt;sub&gt;2&lt;/sub&gt;&lt;sup&gt;-1&lt;/sup&gt;</td>
<td>+47.4</td>
<td>+64.3</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N&lt;sub&gt;7&lt;/sub&gt;&lt;sup&gt;-2&lt;/sup&gt;</td>
<td>N&lt;sub&gt;7&lt;/sub&gt;&lt;sup&gt;-2&lt;/sup&gt;</td>
<td>+75.5</td>
<td>+65.8</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N&lt;sub&gt;7&lt;/sub&gt;&lt;sup&gt;-2&lt;/sup&gt;</td>
<td>N&lt;sub&gt;7&lt;/sub&gt;&lt;sup&gt;-2&lt;/sup&gt;</td>
<td>+74.3</td>
<td>+69.2</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N&lt;sub&gt;7&lt;/sub&gt;&lt;sup&gt;-2&lt;/sup&gt;</td>
<td>N&lt;sub&gt;7&lt;/sub&gt;&lt;sup&gt;-2&lt;/sup&gt;</td>
<td>+75.5</td>
<td>+65.8</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N&lt;sub&gt;7&lt;/sub&gt;&lt;sup&gt;-2&lt;/sup&gt;</td>
<td>N&lt;sub&gt;7&lt;/sub&gt;&lt;sup&gt;-2&lt;/sup&gt;</td>
<td>-54.0</td>
<td>-11.8</td>
<td>x</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C&lt;sub&gt;3&lt;/sub&gt;&lt;sup&gt;-3&lt;/sup&gt;</td>
<td>C&lt;sub&gt;8&lt;/sub&gt;&lt;sup&gt;-3&lt;/sup&gt;</td>
<td>+41.7</td>
<td>+86.5</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N&lt;sub&gt;2&lt;/sub&gt;&lt;sup&gt;3&lt;/sup&gt;</td>
<td>C&lt;sub&gt;2&lt;/sub&gt;&lt;sup&gt;3&lt;/sup&gt;</td>
<td>-102.4</td>
<td>+56.4</td>
<td>x</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N&lt;sub&gt;7&lt;/sub&gt;&lt;sup&gt;-2&lt;/sup&gt;</td>
<td>N&lt;sub&gt;7&lt;/sub&gt;&lt;sup&gt;-2&lt;/sup&gt;</td>
<td>+74.3</td>
<td>-118.1</td>
<td>x</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C&lt;sub&gt;3&lt;/sub&gt;&lt;sup&gt;-3&lt;/sup&gt;</td>
<td>C&lt;sub&gt;2&lt;/sub&gt;&lt;sup&gt;2&lt;/sup&gt;</td>
<td>-87.8</td>
<td>+56.4</td>
<td>x</td>
</tr>
</tbody>
</table>
Table 1. (Continued)

<table>
<thead>
<tr>
<th>Inhibitor</th>
<th>IC&lt;sub&gt;50&lt;/sub&gt; (µM)</th>
<th>CORRESPONDING MAXIMA AND MINIMA</th>
<th>Same</th>
<th>Strength&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Separation</th>
<th>RMSD OF POINTS IN MATCH&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Nearest Inhibitor Atom</td>
<td>V /kJ mol&lt;sup&gt;-1&lt;/sup&gt;</td>
<td>Nearest cAMP Atom</td>
<td>V /kJ mol&lt;sup&gt;-1&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Milrinone</td>
<td>2.20</td>
<td>O&lt;sub&gt;6&lt;/sub&gt; -170.0 O&lt;sub&gt;6'&lt;/sub&gt; -206.8</td>
<td>✓</td>
<td>strong</td>
<td>0.674</td>
<td>0.749 (0.497)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C7-H&lt;sub&gt;max&lt;/sub&gt; +88.2 C8-H&lt;sub&gt;max&lt;/sub&gt; +86.5</td>
<td>✓</td>
<td>strong</td>
<td>0.626</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>N1-H&lt;sub&gt;max&lt;/sub&gt; +117.2 O7'-H&lt;sub&gt;max&lt;/sub&gt; +126.9</td>
<td>✓</td>
<td>strong</td>
<td>0.527</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>C4-H&lt;sub&gt;max&lt;/sub&gt; +49.0 O2'-H&lt;sub&gt;max&lt;/sub&gt; +64.3</td>
<td>✓</td>
<td>moderate</td>
<td>1.058</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>N1'-&lt;sub&gt;min&lt;/sub&gt; -128.2 N0'&lt;sub&gt;min&lt;/sub&gt; -11.8</td>
<td>✓</td>
<td>moderate</td>
<td>1.729</td>
<td></td>
</tr>
<tr>
<td>Anrirone</td>
<td>51.00</td>
<td>O&lt;sub&gt;6&lt;/sub&gt; -142.8 O&lt;sub&gt;6'&lt;/sub&gt; -206.8</td>
<td>✓</td>
<td>strong</td>
<td>1.786</td>
<td>1.406 (0.858)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N1'-&lt;sub&gt;min&lt;/sub&gt; -147.4 N1'-&lt;sub&gt;min&lt;/sub&gt; -118.1</td>
<td>✓</td>
<td>strong</td>
<td>0.479</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>N1-H&lt;sub&gt;max&lt;/sub&gt; +94.2 O2'-H&lt;sub&gt;max&lt;/sub&gt; +64.3</td>
<td>✓</td>
<td>strong</td>
<td>0.550</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>X&lt;sup&gt;a&lt;/sup&gt; +40.7 N6'&lt;sub&gt;min&lt;/sub&gt; -11.8</td>
<td>✓</td>
<td>weak</td>
<td>2.045</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>C3'&lt;sub&gt;min&lt;/sub&gt; -29.4 N3'&lt;sub&gt;max&lt;/sub&gt; -132.1</td>
<td>✓</td>
<td>moderate</td>
<td>1.449</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>C4'&lt;sub&gt;min&lt;/sub&gt; -32.2 O'&lt;sub&gt;max&lt;/sub&gt; -73.5</td>
<td>✓</td>
<td>moderate</td>
<td>1.975</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>N1'&lt;sub&gt;max&lt;/sub&gt; +61.1 N3'&lt;sub&gt;min&lt;/sub&gt; -132.1</td>
<td>✓</td>
<td>moderate</td>
<td>2.023</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>C4'&lt;sub&gt;min&lt;/sub&gt; -31.6 N3'&lt;sub&gt;min&lt;/sub&gt; -132.1</td>
<td>✓</td>
<td>moderate</td>
<td>2.422</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>C5'&lt;sub&gt;min&lt;/sub&gt; -33.7 O'&lt;sub&gt;max&lt;/sub&gt; -73.5</td>
<td>✓</td>
<td>moderate</td>
<td>2.848</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>C4'&lt;sub&gt;max&lt;/sub&gt; -32.2 C2'-H&lt;sub&gt;max&lt;/sub&gt; +56.4</td>
<td>✓</td>
<td>moderate</td>
<td>2.422</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>N3'&lt;sub&gt;max&lt;/sub&gt; -58.4 C8-H&lt;sub&gt;max&lt;/sub&gt; +86.5</td>
<td>x</td>
<td>strong</td>
<td>1.360</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>N5-H&lt;sub&gt;max&lt;/sub&gt; +86.0 O'&lt;sub&gt;max&lt;/sub&gt; -73.5</td>
<td>x</td>
<td>strong</td>
<td>1.953</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>C3'-H&lt;sub&gt;max&lt;/sub&gt; +33.4 C2'-H&lt;sub&gt;max&lt;/sub&gt; +56.4</td>
<td>x</td>
<td>moderate</td>
<td>2.066</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>C3'-H&lt;sub&gt;max&lt;/sub&gt; +33.4 O'&lt;sub&gt;max&lt;/sub&gt; -73.5</td>
<td>x</td>
<td>moderate</td>
<td>2.353</td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>The relative strength of each pair of corresponding extrema is classified according to the definition: strong - both extrema have | V | > 50 kJ mol<sup>-1</sup>; moderate - only one extremum has | V | > 50 kJ mol<sup>-1</sup>; weak - both extrema have | V | < 50 kJ mol<sup>-1</sup>.<br>
<sup>b</sup>Relative binding orientations quantified by the minimum root mean square of the separation of the first four (three) corresponding pairs of matched points (above the dashed line). All other pairs of contacts within 3Å are also given. Separation of extrema pairs are given for the four point match.
<table>
<thead>
<tr>
<th>Inhibitor</th>
<th>CORRESPONDING MAXIMA AND MINIMA</th>
<th>RMSD OF POINTS IN MATCH$^4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adenine</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Nearest Inhibitor Atom</td>
<td>$V_{kJ mol^{-1}}$</td>
</tr>
<tr>
<td>Adenine</td>
<td>N7$_{\text{max}}$</td>
<td>-120.2</td>
</tr>
<tr>
<td></td>
<td>N3$_{\text{max}}$</td>
<td>-138.6</td>
</tr>
<tr>
<td></td>
<td>N1$_{\text{max}}$</td>
<td>-131.4</td>
</tr>
<tr>
<td></td>
<td>N6$_{\text{max}}$</td>
<td>-35.9</td>
</tr>
<tr>
<td></td>
<td>N6-H$_{\text{max}}$</td>
<td>-35.9</td>
</tr>
<tr>
<td></td>
<td>N6-H$_{1\text{max}}$</td>
<td>+54.2</td>
</tr>
<tr>
<td></td>
<td>N6-H$_{2\text{max}}$</td>
<td>+45.7</td>
</tr>
<tr>
<td>Guanine</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Nearest Inhibitor Atom</td>
<td>$V_{kJ mol^{-1}}$</td>
</tr>
<tr>
<td>Guanine</td>
<td>O6$_{\text{max}}$</td>
<td>-234.6</td>
</tr>
<tr>
<td></td>
<td>N3$_{\text{max}}$</td>
<td>-79.2</td>
</tr>
<tr>
<td></td>
<td>N2-H$_{1\text{max}}$</td>
<td>+151.2</td>
</tr>
<tr>
<td></td>
<td>N2-H$_{1\text{min}}$</td>
<td>+151.2</td>
</tr>
<tr>
<td>Theophylline</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Theophylline</td>
<td>Nearest Inhibitor Atom</td>
<td>$V_{kJ mol^{-1}}$</td>
</tr>
<tr>
<td>Theophylline</td>
<td>N7-H$_{\text{max}}$</td>
<td>+110.1</td>
</tr>
<tr>
<td></td>
<td>O2$_{\text{max}}$</td>
<td>-151.0</td>
</tr>
<tr>
<td></td>
<td>O9$_{\text{max}}$</td>
<td>-112.0</td>
</tr>
<tr>
<td></td>
<td>O2$_{\text{min}}$</td>
<td>-151.0</td>
</tr>
<tr>
<td>Papaverine</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Papaverine</td>
<td>Nearest Inhibitor Atom</td>
<td>$V_{kJ mol^{-1}}$</td>
</tr>
<tr>
<td>Papaverine</td>
<td>O1$_{\text{max}}$</td>
<td>-105.4</td>
</tr>
<tr>
<td></td>
<td>N1$_{\text{max}}$</td>
<td>-160.0</td>
</tr>
<tr>
<td></td>
<td>C14-H$_{1\text{max}}$</td>
<td>+59.8</td>
</tr>
<tr>
<td></td>
<td>Ar$_{1\text{max}}$</td>
<td>-60.5</td>
</tr>
<tr>
<td></td>
<td>C14-H$_{2\text{max}}$</td>
<td>+46.1</td>
</tr>
<tr>
<td></td>
<td>Ar$_{2\text{max}}$</td>
<td>-57.0</td>
</tr>
<tr>
<td></td>
<td>Ar$_{1\text{max}}$</td>
<td>-60.5</td>
</tr>
<tr>
<td></td>
<td>O13$^B$</td>
<td>-30.2</td>
</tr>
<tr>
<td></td>
<td>O13$^B$</td>
<td>-30.2</td>
</tr>
</tbody>
</table>

$^a$The relative strength of each pair of corresponding extrema is classified according to the definition: strong - both extrema have $|V| > 50$ kJ mol$^{-1}$; moderate - only one extremum has $|V| > 50$ kJ mol$^{-1}$; weak - both extrema have $|V| < 50$ kJ mol$^{-1}$.

Relative binding orientations quantified by the minimum root mean square of the separation of the first (three) or four corresponding pairs of matched points (above the dashed line). All other pairs of contacts within 3 Å are also given. Separation of extrema pairs are given for the three point match, if a four point match was not possible.
References for Chapter 4.


96


Chapter 5. Glycolate Oxidase Inhibitors

In this chapter, the method used for determining relative binding orientations of ligands has been applied to the substrate and inhibitors of the enzyme Glycolate Oxidase. For this system, information about the binding site structure is available, hence it has been possible to compare the relative binding orientations of the inhibitors produced by matching electrostatic extrema, with those found by optimizing the electrostatic interactions of each inhibitor within the binding site.

5.1 Glycolate Oxidase Specific Inhibitors

Glycolate Oxidase (GOX) belongs to the class of enzymes called flavoprotein oxidases, and it can be derived from both plant and mammalian sources. In plant species, GOX is a key factor in photorespiration, causing oxidation of the α-hydroxy acid glycolic acid (anionic form glycolate) to glyoxylic acid (anionic form glyoxylate). During the oxidation of glycolate, Flavin Mononucleotide (FMN) acts as a cofactor. A cofactor is a non-protein compound that is an essential part of an enzyme. Two electrons are temporarily transferred from the electron donor (glycolate) to the cofactor flavin ring, and in a second reaction, transferred to oxygen as the terminal electron acceptor.

The mechanism of photosynthesis is very complex and requires the interplay of a number of proteins and small molecules. GOX is one enzyme involved in a salvage pathway of photosynthesis, recovering part of the carbon skeleton of phosphoglycolate, a metabolite of photosynthesis. A specific phosphatase converts phosphoglycolate to glycolate, which is then oxidized to glyoxylate by GOX. Transamination of glyoxylate (the transfer of an amino group from an α-amino acid to an α-keto acid) yields glycine. Serine is then formed from two glycine molecules, with the release of CO₂ and NH₄⁺. As a result of this pathway for carbon metabolism, three out of the four carbon atoms of two glycolate molecules are recovered, reducing net photosynthesis.

Figure 1. The reaction catalysed by GOX; the oxidation of a) glycolic acid, via b) glyoxylic acid, to c) oxalic acid.
The reaction catalysed by GOX can be summarised as the oxidation of glycolic acid to oxalic acid, via glyoxylic acid (figure 1). Schuman and Massey\(^1\) described the inhibition of GOX by a series of monocarboxylic and dicarboxylic acids. Inhibition by monocarboxylic acids was non-competitive with respect to glycolate (figure 2(a)). Inhibition constants were found to decrease as the number of carbon atoms in the alkyl side-chain increased. Inhibition by dicarboxylic acids was competitive with respect to glycolate. The affinity of the enzyme for oxalate (figure 2(b)) was high compared to monocarboxylic acids, therefore it was proposed that the binding of oxalate involves an electrostatic interaction of both carboxylate groups with two adjacent positively charged groups in the active site. For dicarboxylic acids, inhibition constants increased (i.e., binding became weaker) as the number of carbon atoms between the carboxylate groups increased. Since the electrostatic interaction involving both carboxylate groups is sterically favourable for oxalate, an increasing carbon chain length between the carboxylate groups makes a conformation which could maintain both interactions increasingly unfavourable energetically. On this basis, a model was proposed for the binding site of GOX. The suggested essential features of the binding site consisted of two positively charged groups in close proximity and a hydrophobic binding region large enough to accommodate at least 5 carbon atoms.

Figure 2. Molecules used in this study defining the numbering system for, a) glycolate, the natural substrate for GOX, b) oxalate, c) 4-substituted 3-hydroxy-1\(H\)-pyrrole-2,5-dione (pyrroledione), d) 5-substituted 4-hydroxy-isothiazolinone (isothiazoline), e) 4-substituted methylhydantoin (methylhydantoin), and three conformations of 2,4-dioxo-1-butanoic acid; f) dioxobutanoic acid, g) dioxo_insight, h) dioxo_sybyl.
Prior to the determination of the crystal structure of GOX, several workers studied potential inhibitors for both plant and animal-derived GOX. Early work on GOX inhibitors focused on natural substrate analogues. Randall et al. investigated a series of simple natural substrate mimics. Since inhibitors had previously been postulated to interact with two positively charged groups at the binding site of GOX, electron donating substituents capable of increasing the negative charge at the carboxylate group and the neighbouring keto, hydroxyl or ether oxygen would be expected to increase the strength of the inhibitor binding. For substituted glyoxylic acids, the results of this study emphasize the presence of a nucleophilic group, in close proximity to the α-carbonyl of the glyoxylate group, as being highly significant.

Rooney et al. investigated a series of 4-substituted 3-hydroxy-1H-pyrrole-2,5-dione derivatives (figure 2(c)) as inhibitors of GOX. Since Schuman and Massey had concluded that the GOX binding site contained two positively charged groups with a hydrophobic binding region, a series of diacidic compounds was studied. These compounds proved to be more potent than any of the previous natural substrate mimics studied as mammalian GOX inhibitors. Their work indicated that a range of quite large lipophilic 4-substituents could interact with the hydrophobic region of the enzyme binding site, producing very potent inhibitors derived from the 3-hydroxy-1H-pyrrole-2,5-dione ring system.

Williams et al. examined a series of 4-substituted 2,4-dioxo-1-butanoic acid derivatives (figure 2(f, g, h)). As in previous studies, high lipophilicity of the 4-substituents was highlighted as an important structural feature of effective inhibitors. They then directed their efforts towards substituted phenyl and biphenyl analogues, which had been successful for the 4-pyrrole-2,5-dione inhibitors. Introduction of a bromine at the 4'-biphenyl position gave enhanced potency (this was also seen with the corresponding 4-pyrrole-2,5-dione inhibitors). It was found that very large 4' terminal groups with some hydrophilic character could be accommodated by the enzyme hydrophobic bonding region, and these compounds showed the best potency, equivalent to the 4'-Br compound. The results agreed with previous findings that the presence of two acidic groups in the vicinity of large lipophilic substituent seemed to define the general characteristics of potent GOX inhibitors.
When the absolute potencies of five of the aromatic ring substituted 2,4-dioxo-1-butanoic acid inhibitors were compared to those of the equivalent 4-pyrrole-2,5-dione inhibitors investigated by Rooney et al.\(^5\) a close correlation could be seen. This indicated that the identical substituents could interact with the enzyme binding site in a similar manner. However, the authors point out that the molecules are not exactly superimposable, due to the possibility of tautomeric forms in the 2,4-dioxo-1-butanoic acid inhibitors.

Other work has proposed the 5-substituted 4-hydroxy-isothiazolinone ring\(^5\)\(^6\) (figure 2(d)) and the 4-substituted hydantoin\(^7\) (figure 2(e)) as potential inhibitor 'spearheads', whose diacidic features are an attempt to accommodate the two positively charged groups proposed in the Schuman and Massey\(^1\) binding site model.

5.2 Glycolate Oxidase Crystal Structure

The crystal structure of GOX was determined by Lindqvist and Brändén\(^8\) to 2.2 Å resolution, and was subsequently refined to 2.0 Å resolution.\(^9\) The structure of the apoenzyme (from which the cofactor has been removed) was later reported by Sandalova and Lindqvist.\(^10\) There are 369 amino acid residues in GOX, with FMN as the cofactor. The main features of the enzyme are the eight \(\alpha\)-helices and \(\beta\)-strands which form the barrel structure. The eight-stranded \(\alpha/\beta\)-barrel structure is a common stable structural motif that has been found in a variety of different enzymes of non-homologous amino acid sequences and with quite different functions. The barrel is built up from a common core of eight parallel \(\beta\)-strands and eight \(\alpha\)-helices connected sequentially along the chain by loop regions of variable length and conformation. The strands forming the barrel core are the eight mainly hydrophobic \(\beta\)-strands, and the eight \(\alpha\)-helices face the surrounding solution. Although functionally, GOX belongs to the flavin dependent oxidases enzyme family, it belongs to a different structural family (Trisphosphate isomerase-barrel type) with a novel mode of cofactor binding.

The active site for this class of enzymes is usually at the end of the barrel, at the COOH terminal of the strands. Binding and catalysis residues are found in the loop regions joining the \(\beta\)-strands to the \(\alpha\)-helices. In all previously studied barrel structures, this site is the active site. In GOX however, the binding site created by the barrel is
occupied by the cofactor FMN. Most of the basic residues that are not involved in intra- or inter-subunit salt bridges, are lining the entrance to the active site cleft, or are involved in FMN phosphate or substrate binding.

All residues which hydrogen bond to the cofactor are in the loop regions or at the COOH terminal of the β-strands. A number of amino acid residues surrounding the flavin ring are aromatic. The N5 edge of the flavin ring is exposed to the solvent, and the ribityl side chain of the cofactor is buried inside the barrel. Outside the N5 of the FMN isoalloxazine ring is a pocket which is accessible to the solvent via a channel at the top of the barrel. Residues lining this pocket near the cofactor are: Tyr-24, Trp-108, Tyr-129, Arg-164 (further out in the channel), His-254, and Arg-257. Trp-108 is considered to be part of the active site, although it has been referred to as the ‘doorstep tryptophan’, because it limits the steric space available to ligands in the binding site passage. Practically all of the active site residues are in the loops going from the β-strands to the α-helices. The pocket is narrow, with three water molecules filling the space of the pocket close to the cofactor. Two of these water molecules have an internal hydrogen bond arrangement which mimics the way in which the carboxylic group of the substrate, glycolate, could be bound in the site.

5.2.1 Comparison of Glycolate Oxidase with Flavocytochrome b₂

A comparison of the three dimensional structures of GOX and another FMN-dependent enzyme, flavocytochrome b₂, was carried out by Lindqvist et al. GOX and flavocytochrome b₂ show high structural homology (RMS = 0.93 Å for 311 Cα atoms), but they have different modes of binding FMN. The substrates for both enzymes are α-hydroxy acids, glycolate and lactate, respectively. The active site residues of GOX and flavocytochrome b₂ can be strictly superimposed, except for Trp-108, which is replaced by a leucine residue in flavocytochrome b₂. Additionally, part of the volume of the Trp-108 ring is occupied by the propionate side chain of the heme in flavocytochrome b₂. Therefore, even though the α/β-barrel is a structural motif common to both enzymes, GOX and flavocytochrome b₂ are also grossly similar in all loop regions containing the active site residues. The major difference between the two enzymes involves the loop in GOX which contains the 29 residues covering the active site. In flavocytochrome b₂ this
loop is moved away from the active site. In GOX the FMN ring is tilted away from strand one, creating a pocket on the re-face of the FMN ring where a water molecule is bound.

5.3 Glycolate Oxidase Proposed Model for Glycolate Binding
On the basis of the binding site structure of GOX and observed similarities to flavocytochrome b₂ (where a crystal structure of the bound substrate is available), a model for glycolate binding to GOX was proposed by Lindqvist et al.\textsuperscript{12} It was assumed that there is only one possible binding site for glycolate close enough to the cofactor. Subsequently, glycolate was model built into the substrate pocket, assuming the position of two water molecules in the site as the position of the carboxylic group of glycolate. In this model the carboxylic group of glycolate forms a strong ionic interaction with Arg-257, and a hydrogen bond to Tyr-24. The hydroxyl group of glycolate makes a hydrogen bond to Tyr-129. The C2 of glycolate comes close to N5 of FMN, and to N3 of His-254 (figure 3). This is similar to L-lactate binding in flavocytochrome b₂.\textsuperscript{11} When superimposing the C\_α atoms of GOX and flavocytochrome b₂, the proposed model for glycolate/GOX binding falls on top of the observed binding site position of pyruvate/flavocytochrome b₂ (pyruvate is the oxidation product of lactate).

![Diagram of the Lindqvist model for glycolate binding in the active site of GOX](image)

**Figure 3.** The Lindqvist model for glycolate binding in the active site of GOX, with the stabilizing intra- and intermolecular interactions indicated. (after Lindqvist \textit{et al.}, J.Biol.Chem., 266 (1991) 3198).
The Lindqvist glycolate binding model puts the C2 α-H in a position to be attacked by N3 of His-254 for catalysis. During catalysis, the substrate binds, and the C2 proton is abstracted from the substrate. The carbanion attaches to N5 of FMN, transiently forming a covalent adduct, transferring 2 electrons to FMN. However, the authors state that there is no positive evidence concerning this electron transfer for GOX or flavocytochrome b$_2$. Steric hindrance may occur for formation of a covalent intermediate in flavocytochrome b$_2$. In GOX however, this crowding would be less severe due to a different orientation of FMN. Reduced FMN is then reoxidized by oxygen, but it is not known where on the flavin ring this occurs in oxidase enzymes. The proton added to His-254 can form a hydrogen bond to N5 of FMN, preventing free exchange with the solvent. Subsequent electron transfer to the FMN ring from the carbanion is facilitated by Lys-230–N1/02(FMN) interaction (which stabilizes the reduced FMN anion). Lindqvist and Brändén conclude that similarities in the active sites of GOX and flavocytochrome b$_2$ support the idea of an identical first step in catalysis; carbanion formation. The different FMN orientations in the two enzymes would seem responsible for functional differences in the flavin oxidative half-reaction.

5.3.1 Role of Tyr-129

In the glycolate binding model, the OH group of Tyr-129 was proposed to form a hydrogen bond to the α-OH group of glycolate, and cause breakdown of a putative N5 adduct to yield reduced FMN and glyoxylate (the product). The specific role of the Tyr-129 residue has been investigated using a mutated enzyme structure, in which the tyrosine residue was replaced by a phenylalanine group, and so rendered incapable of forming the proposed hydrogen bond. The results of this work using the [Y129F] mutant GOX (replacing Tyr-129 by Phe) confirmed that Tyr-129 participates in the reductive step of catalysis, but the Tyr-129 hydroxyl group was found to be non-essential for substrate binding. Subsequent [Y129F] oxidation of glycolate was not inhibited by oxalate (even at 12M oxalate). From this loss of oxalate binding on removal of Tyr-129, it was therefore concluded that the Tyr-129 hydroxyl group must be essential for binding oxalate. In [Y129F] the FMN N3 was not deprotonated (as is the wild type GOX). Most other typical flavoprotein oxidase properties were not affected by mutation of Tyr-129.
Schuman and Massey\(^4\) reported a shift in the pka of the N3 FMN proton in the presence of oxalate for pig liver GOX. With wild type spinach GOX a similar pka shift was found in the presence of oxalate. Tyr-129 was assumed to be involved in lowering the pka. Macheroux et al\(^5\) related the protonation of the N3 and lack of oxalate inhibition in the mutant GOX. From the results of pka shift in the presence of oxalate, they conclude that oxalate forms a hydrogen bond with Tyr-129, reducing the interaction between Tyr-129 and the FMN C4=O oxygen. This leads to an increase in pka of the N3 proton.

The earlier work by Schuman and Massey\(^4\) proposed that the binding of oxalate involved an electrostatic interaction of both carboxylate groups with two adjacent positively charged groups in the active site. Arg-257 and Arg-164 are two such residues which could bind oxalate. However, Macheroux et al\(^5\) have suggested that this seems to be too far from FMN for productive substrate binding, and that perhaps these residues form an initial binding site. From their observations it was concluded that Tyr-129 is not involved in binding the substrate in GOX. It appears that oxalate acts as a transition state analogue, and the role of Tyr-129 is to stabilize the transition state, not to bind the substrate. This explains the complete loss of oxalate binding on mutation of Tyr-129 to Phe.

5.3.2 Role of Tyr-24 and Trp-108

More recent work\(^14\) has examined the predicted roles of Tyr-24 and Trp-108 in the binding and catalytic functions of spinach GOX, by replacing these residues with phenylalanine and serine respectively. From the [Y129F]glycolate oxidase study,\(^13\) Tyr-24 has been suggested to be directly involved in catalysis, by protonating the glycolate carboxylate group thereby stabilizing the transiently formed carbanion through delocalization of the negative charge and lowering the pka of the α-carbon hydrogen. Studies on the kinetic properties of the mutant [Y24F]glycolate oxidase\(^14\) suggested that Tyr-24 does not stabilize the transition state. Binding of glycolate was significantly affected. Two important changes were observed in the structure of the active site of the [Y24F]glycolate oxidase mutant compared with the wild-type GOX. The side chain of
Arg-164 moved towards the substrate binding pocket, and the side chain of Trp-108 moved into the position of a water molecule which is bound to Tyr-24 in the wild-type GOX. The effects on substrate binding and catalysis of the substitution of Try-24 suggest that the residue is involved in substrate binding, but does not stabilize the transition state. Earlier mechanisms\textsuperscript{15,16} proposed that Tyr-24 stabilizes a transient carbanion by delocalizing the negative charge. Removal of the Tyr-24 hydroxyl group by replacement with phenylalanine would be expected to reduce the rate of reduction if this mechanistic proposal were true. However, the studies with the [Y24F]glycolate oxidase mutant do not demonstrate this. Stenberg et al\textsuperscript{14} suggest an alternative mechanism, which is consistent with the results obtained from the [Y129F] and [Y24F] mutant GOX enzymes. In this mechanism, His-254 abstracts the more acidic hydroxyl glycolate proton, creating an alcoholate (rather than a carbanion) which could be stabilized by the positively charged His-254 imidazole ring and Tyr-129.

Trp-108 has always been considered to be of importance in limiting the steric space available in the binding site, but not to have any catalytic function. A structural comparison of GOX with other related \(\alpha\)-hydroxy-acid flavoprotein oxidases has shown that Tyr-24 and Tyr-129 are conserved, but Trp-108 (which lies between the two tyrosine residues) is not. The initial interest in the [W108S]glycolate oxidase mutant was to study the effect of the replacement of Trp-108 on hydrogen bonding arrangements in the binding site. It has been noted that the size of the amino acid in this position in \(\alpha\)-hydroxy-acid flavoprotein oxidases can be related to the substrate specificity of the enzyme. Steady-state experiments using the [W108S]glycolate oxidase found glycolate to be a poor substrate for the mutant, whilst substrates with larger side chains were preferably bound. The replacement of Trp-108 by a serine residue therefore favoured larger substrate side chains. A sequence alignment of four \(\alpha\)-hydroxy-acid flavoprotein oxidases showed a correlation between the position of the amino acid residue corresponding to Trp-108, and the size of the substrate side chain. Larger substrate side chains correlated with smaller amino acid side chains in this position, whilst neighbouring residues were conserved in all four enzymes. This led to the conclusion that the amino acid in the Trp-108 position influences the substrate specificity of the enzyme. The clear conclusion is that the role of Trp-108 goes beyond limiting the space available in the
binding site, and is important to the catalytic function of GOX. This effect was not seen with the [Y24F] glycolate oxidase mutant.

Since Schuman and Massey \(^1\) proposed the features of their model of the GOX binding site prior to the elucidation of the crystal structure, various studies have investigated the role of the different residues in binding and catalysis of glycolate. Based on the structure of the empty binding site, the original Lindqvist model \(^12\) for glycolate binding proposed a strong ionic interaction with Arg-257 and hydrogen bonds from the carboxylate and hydroxyl oxygens of glycolate to Tyr-24 and Tyr-129 respectively. Subsequently it has been demonstrated that Tyr-24 binds the substrate, \(^14\) but that Tyr-129 is involved in catalysis, \(^13\) stabilizing the transition state. The function of Trp-108 would seem to go beyond the original suggestion that it simply limits the steric space available in the binding site.

In this work, several GOX inhibitor 'spearheads', where the primary electrostatic interactions between the ligand and binding site residues occur, have been considered, along with the natural substrate glycolate and oxalate (thought to be a transition state analogue). We aim to establish the electrostatic similarity of these ligands, and to see whether matching their electrostatic extrema can produce relative superpositions of the inhibitors which identify common binding site interactions.

5.4 Methods

5.4.1 Geometry of Ligands and Binding Site Residues

The geometries of the molecules studied here have been obtained by energy minimization within the software package InsightII \(^17\), utilizing the CVFF \(^18\) molecular mechanics force field, running on a Silicon Graphics Indigo 2 workstation. The natural substrate, glycolate and one of the inhibitors, oxalate are small rigid molecules (figure 2). The three closed ring inhibitors, 3-hydroxy-1H-pyrrole-2,5-dione, 4-hydroxy-isothiazolinone and 4-substituted hydantoin have essentially fixed conformations. The 2,4-dioxo-1-butanoic acid inhibitor is an open chain compound, whose flexibility allows a variety of low energy conformations to be adopted. For this reason, three separate low energy conformations of this inhibitor have been considered, two obtained by a conformational analysis within InsightII, and a third modelled using the SYBYL \(^19\) software. These three conformers,
labelled dioxobutanoic acid, dioxo_insight and dioxo_sybyl differ at the C3-C2-C1-O1' torsion angle. This angle differs by 140.7° between the two InsightII conformations, effectively exchanging the positions of the C1=O1 and C1-O1'H bonds of the carboxylic group and altering the electrostatic environment of the two molecules in this region. For the dioxo_sybyl conformer, the important differences with the dioxobutanoic acid and dioxo_insight structures were the conformation at the carboxylic group, and the 2.8° difference in the C5-C4-C3 bond angle, which determines the orientation of the methyl group. All three conformations are shown in figure 2(f, g, h).

It should be noted that the 4-substituted hydantoin (figure 2(e)) was proposed as a potential inhibitor on the basis of a structural comparison with existing inhibitors, but was subsequently found not to have any inhibitory effect on GOX. This ligand has been included to determine whether its lack of inhibition was due to electrostatic incompatibility with the existing inhibitors and natural substrate. The primary electrostatic interactions occur at the ‘spearhead’ of each inhibitor, whilst the hydrocarbon substituent participates in hydrophobic interactions between the enzyme and the inhibitor. Therefore, for the purpose of this work, the hydrocarbon R group has been approximated to a methyl group for all of the inhibitors.

The peptide coordinates of the binding site residues were taken from the spinach GOX crystal structure in the Brookhaven Databank. Each sequence of amino acid residues was blocked using N-acetyl and N'-methylamide terminal groups, to approximate the charge distribution of the residues in the peptide sequence. This method has been successfully used to model the electrostatic properties of polypeptides. The N-acetyl, N'-methylamide blocked residues were constructed directly from the crystal structure atomic coordinates, by substituting methyl groups at the coordinates of the Ca atoms of the residues preceding and following the selected residue in the peptide sequence. When a series of sequential residues was selected, N-acetyl and N'-methylamide blocking groups were placed at the ends of the entire peptide sequence.

5.4.2 CADPAC Ab Initio Calculations
SCF wavefunctions were obtained, using the CADPAC suite of ab initio programs, for each ligand, the FMN cofactor, and 16 GOX binding site residues: Tyr-24, Ser-106, Ser-
107, Trp-108, Gln-127, Leu-128, Tyr-129, Thr-155, Val-156, Asp-157, Arg-164, Lys-230, His-254, Gly-255, Ala-256 and Arg-257, using a 3-21G basis set\textsuperscript{23} in all cases. The Distributed Multipole Analysis\textsuperscript{24 25} of each wavefunction was calculated to represent the charge density by a charge, dipole, quadrupole, octupole and hexadecapole moment at every nuclear position. An additional 4 residues, which do not form part of the binding site but which define the channel leading to the binding site, were included as repulsion-dispersion only sites, specifically Thr-158, Leu-161, Asp-167, and Phe-172.

5.4.3 ORIENT3 Electrostatic Calculations

All docking calculations have been performed using the DMA model of the charge distribution of the ligands and binding site residues within the program ORIENT3.\textsuperscript{26} As a test of the sensitivity of the calculations to the form of the repulsion-dispersion potential used, the ORIENT3 docking calculations using the DMAs and pseudo hard-sphere repulsion have been compared to calculations using the DMA with three repulsion-dispersion potentials, for glycolate and oxalate. The repulsion-dispersion parameters used were taken from the CVFF force-field\textsuperscript{18} (the default force-field available within InsightII\textsuperscript{17}), the SYBYL TRIPOS 5.2\textsuperscript{27} force field 6-12 potential parameters, and the WCMP 6-exp repulsion-dispersion potential (the parameters were taken from Williams and Cox\textsuperscript{28} for C, N and non-polar hydrogens, derived from fitting to azahydrocarbons, and the O parameters from using oxohydrocarbons\textsuperscript{29}). For the ORIENT3 docking calculations the pseudo hard-sphere repulsion potential was found to be too insensitive for positioning the larger inhibitors ligands within the enzyme binding site, producing unavoidable high repulsions. The SYBYL TRIPOS 5.2 repulsion-dispersion parameters were found to produce unrealistically short intermolecular interactions when used in combination with the DMA, and were disregarded for the larger inhibitors. For these ligands therefore, the ORIENT3 docking calculations docking calculations were performed using WCMP 6-exp and CVFF force-field repulsion-dispersion potential parameters only.

The electrostatic potential surface was calculated from the DMA model of the charge distribution for the GOX substrate, inhibitors and binding site residues within ORIENT3, using all terms in the multipole expansion up to $R^{-5}$ between each atom in the
molecule and a point charge. For calculation of the positions and energies of the electrostatic extrema, the molecular van der Waals volume was defined using the Pauling radii C: 2.0 Å, O: 1.4 Å, N: 1.5 Å, S: 1.8 Å. All hydrogens were treated as having a zero van der Waals radius, due to the lack of repulsion between the proton and its acceptor atom in hydrogen bonds. Non-polar hydrogens, such as methyl hydrogens, are included in a 'united atom' carbon radius. The positions and energies of the extrema in the electrostatic potential of each of the ligands and binding site residues were determined by minimising the interaction energy of a single point charge with the molecule in question, using pseudo hard-sphere repulsion between sites with non-zero van der Waals radii.

5.4.4 Matching Electrostatic Extrema

The superposition of the ligand molecules was performed using the InsightII modelling software. The electrostatic extrema for each molecule were input as dummy atoms, and these dummy atoms were used in the standard InsightII options for superposition of atomic positions. Superpositions for pairs of molecules were attempted by choosing three or four extrema from each molecule and performing a RMS fit to minimise the separation of the chosen pairs. Extrema were chosen which would maximize the electrostatic similarity of the overlaid molecules, with the condition that the overlay must also produce a plausible steric overlay of the two molecules. For each inhibitor, a variety of overlays relative to glycolate were attempted, involving different combinations of extrema. The potential surface around the charged glycolate is negative, and so only electrostatic minima were found (figure 4(a)). For any system where electrostatic interactions are important, inhibitors for such a system must, by definition, be similar to the electrostatic properties of the natural substrate in regions occupied by both molecules within the binding site. However, use of glycolate as the standard for comparison highlighted a difficulty in any method of matching molecular electrostatic properties, when the natural substrate is smaller than the inhibitors, since the distribution of electrostatic extrema in the potential is partially determined by the size of the molecule. Since glycolate is significantly smaller than any of the inhibitors studied, several different orientations of glycolate relative to each inhibitor are possible. Any
plausible superposition of substrate and inhibitor based on their electrostatic properties would also be expected to produce good steric overlap, when comparing molecules with similar steric volume. Comparing molecules of different sizes leaves large volumes which are only occupied by one molecule, and so it is not possible to assess the quality of the overlay using the additional information available from the degree of steric overlap. For these reasons, the comparison of the GOX ligands has been made primarily with reference to the 4-methyl 3-hydroxy-1H-pyrrole-2,5-dione inhibitor (figure 2(c)), abbreviated here as pyrroledione. This ligand has a fixed conformation, and a varied distribution of electrostatic maxima and minima (figure 4(b)). Relative overlays obtained using extrema which are widely spaced around the molecules automatically ensure a plausible steric overlap of the two molecules.

Relative orientations of pairs of inhibitors have been produced by only considering electrostatic extrema calculated for each molecule in isolation. The separations of all maxima and minima between the superimposed molecules which were less than 3.0 Å apart, have been calculated. Shorter separations of extrema of the same sign between molecules were assumed to be more favourable. Clashes were defined as situations where a maximum of one molecule is within 3.0 Å of a minimum of the other molecule. The correspondence between extrema pairs has been classified as ‘strong’, ‘moderate’ or ‘weak’ as before. Thus, similarity was assessed on the basis of the distance between corresponding maxima and minima and their relative strengths.

5.4.5 Binding Site Docking Calculations

For the docking calculations within the GOX binding site, the starting orientations of glycolate and oxalate have been produced according to the Lindqvist model for glycolate binding. The starting orientation of pyrroledione (the inhibitor used as the reference for the relative superpositions of the other ligands) has been produced by manually positioning the ligand in an orientation which allows favourable binding interactions to occur. The remaining larger ligands have then been positioned in the binding site in orientations relative to pyrroledione which were determined previously by matching electrostatic extrema. Since we propose that locating maxima and minima in the electrostatic potential of the isolated ligands should give some indication of the
relative position of receptor site atoms, we would expect the relative orientations produced in this way to identify possible hydrogen bonds between the ligands and the binding site.

5.5 Results I - Matching Electrostatic Extrema

5.5.1 Pyrroledione

The potential surface around the natural substrate, glycolate is predominately negative, with a simple arrangement of electrostatic minima, and no electrostatic maxima (figure 4). The potential surface of pyrroledione shows more variation. The important features of the electrostatic potential surface of pyrroledione are the minima arranged around O2 and O5, and the maximum related to the N1-H group. These three points, together with the steric restriction imposed in the binding site, based on the orientation of the 4-methyl group, have been used as the main criteria for each attempted overlay. The four point overlays chosen for each inhibitor ‘spearhead’ are detailed in table 1, and are illustrated in figures 5 - 7.

5.5.2 Isothiazolinone

The electrostatic potential surfaces of isothiazolinone and pyrroledione have a high degree of similarity as a result of the structural similarity between the two inhibitors. The relative superposition produced by a comparison of their electrostatic extrema (figure 5a) overlays their common structural features, with a close coincidence of electrostatic potential features. The steric requirement of the direction of the methyl group is also satisfied. For both the three point and four point overlays, a low RMS value is produced, with additional correspondence between electrostatic extrema which were not explicitly matched.

5.5.3 Methylhydantoin

This compound was proposed as an inhibitor but was subsequently found to be biologically inactive. An initial structural comparison suggests that the methylhydantoin functional groups would generate an electrostatic potential similar to that of pyrroledione, by matching the N1 and N3 positions of methylhydantoin to N1 and C3 of
pyrroledione. This is confirmed by the distribution of electrostatic extrema, which produces a very good RMS value for the resulting 3 point relative superposition with pyrroledione. For the four point overlay (figure 5b), the structural overlap of the ring atoms is somewhat poorer, but the electrostatic match is still relatively good. It would seem therefore, that the electrostatic requirements of the binding process are satisfied by this compound, and the lack of activity appears to be the result of some property other than the electrostatic properties of the 'spearhead'. A likely explanation relates to the steric restrictions imposed on the position of the R-methyl group. The C4 carbon of methylhydantoin is $sp^2$ hybridized rather than $sp^3$ hybridized as in pyrroledione. For the 3 and 4 point electrostatic matches, the separation between the methyl group carbons of pyrroledione and methylhydantoin is 1.76 Å and 2.00 Å out of the ring plane. The result is that the orientations of the two methyl groups are not in agreement. For this compound therefore, it would seem that structural factors are responsible for it's biological inactivity.

5.5.4 Dioxobutanoic Acid Conformations

Dioxobutanoic Acid

With the arrangement of electrostatic extrema around dioxobutanoic acid in this conformation, a large number of three and four point overlays with pyrroledione were possible, using different combinations of electrostatic extrema. Many of these relative superpositions were grossly similar sterically, and could be divided into two broad groups depending on the orientation of the 4-methyl group. Using the separation of the two inhibitor methyl groups as an extra steric criterion in addition to the electrostatic RMS value, some of the relative overlays have been disregarded. The dioxobutanoic acid overlays reported here match a similar set of features of the pyrroledione electrostatic potential surface as the previous isothiazolinone overlay. In addition the reported overlay places the 4-methyl R-group of dioxobutanoic acid in the correct orientation relative to the pyrroledione 4-methyl group (figure 6a). The separations between the methyl groups for the three and four point overlays are 0.97 Å and 1.55 Å, but the side chains are collinear. The relative superpositions suggest that the O2, O4 and O1'-H of dioxobutanoic acid could form binding interactions with the same receptor atoms as O2,
O5 and N1-H of pyrroledione.

**Dioxo_insight**

Dioxo_insight differs from dioxobutanoic acid at the C3-C2-C1-O1 torsion angle. This conformation change results in the loss of one electrostatic minimum around O2/O1, shifting of a minimum at O1', and loss of a weak maximum at C3. The O1'-H maximum (57.8 kJ mol⁻¹) is of weaker magnitude compared to the equivalent dioxobutanoic acid maximum (98.1 kJ mol⁻¹). The position of the O1'-H_max is also shifted. Again, a large number of three and four point relative overlays of dioxo_insight with respect to pyrroledione were possible. The occurrence of an extra electrostatic minimum for dioxobutanoic acid means that additional overlays were considered with respect to the pyrroledione electrostatic minima which were not possible for dioxo_insight. After consideration, several of these overlays were dismissed on steric grounds. The three and four point relative superposition reported here are similar to those chosen for dioxobutanoic acid, since they rely on the same set of electrostatic extrema, despite the rearrangement of extrema which results from the torsional angle change. However there is some difference in the four point overlay (figure 6b).

**Dioxo_sybyl**

The conformational change in this 2,4-dioxo-1-butanoic acid structure produced a consequent alteration in the arrangement of electrostatic maxima and minima around this inhibitor. The number of electrostatic maxima located was increased relative to dioxobutanoic acid and dioxo_insight, and a double minimum related to the carboxylic group was also found. These changes proved to be more significant because of the loss of a minimum at O2 and a maximum at O1'-H position. The electrostatic properties around the carboxylic group are effectively exchanged relative to dioxobutanoic acid and dioxo_insight, resulting in different relative superpositions. As for the previous dioxobutanoic acid conformers, a range of electrostatic overlays were found. The overlay presented in table 1 produces the lowest RMS value of the three 2,4-dioxo-1-butanoic acid conformations, however as a result of the conformation at the carboxylic group the dioxo_sybyl conformer does not have any electrostatic maxima near to the N1-H_max of
pyrroledione in any overlay, unlike all of the previous inhibitor spearhead groups (figure 7).

5.6 Conclusions I - Matching Electrostatic Extrema
A comparison of all the four point relative superpositions of the GOX inhibitor spearheads (figure 8) highlights their electrostatic similarity and clearly distinguishes areas of positive and negative electrostatic potential. There is also good agreement in the position of the substituted methyl group, except in the case of methylhydantoin, which correlates with its lack of inhibitory activity.

Based on the matching of electrostatic extrema, the conformation of the 2,4-dioxo-1-butanoic acid inhibitor which produces the best relative overlay with pyrroledione is dioxobutanoic acid. Although the dioxo_sbyyl conformation has the lowest three point RMS value, dioxobutanoic acid has the lowest four point RMS, and for both the three point and four point overlays has more strong correspondences between favourable extrema than either of the other two conformations.

5.7 Results II - Binding Site Docking Calculations
For the natural substrate, glycolate, the initial position within the binding site was produced according to the Lindqvist model for glycolate binding. Due to the small size of the natural substrate glycolate, and one inhibitor oxalate, calculations using hard-sphere repulsion between sites with non-zero van der Waals radii were possible for these two ligands. These hard-sphere positions have been compared to those obtained using the WCMP exp-6 repulsion-dispersion parameters, and parameters taken from the SYBYL TRIPOS 5.2 and CVFF force fields. For glycolate and oxalate the energy details of the initial Lindqvist model position using each potential, and the final position obtained for each ligand within the binding site are listed in tables 2 and 3 for comparison. For the larger inhibitors, results are given using the WCMP repulsion-dispersion parameters only (except in the case of isothiazolinone, for which results using the CVFF parameters are reported) in table 4. The final position within the binding site for each inhibitor is shown in figures 9-13.
5.7.1 Glycolate - Lindqvist Model

The Lindqvist model\textsuperscript{12} position for glycolate with hard-sphere repulsion is slightly repulsive with respect to Tyr-24 and the Ser-106, Ser-107, Trp-108 tripeptide sequence (figure 9, table 2). There is additional large repulsion between glycolate and the cofactor FMN. This repulsion can be attributed to 3 atoms: C4, C5\textsubscript{A} and C6 of FMN, which are each approximately 0.1 Å too close to Cl of glycolate. Using the pseudo hard-sphere potential glycolate converged to a minimum with an RMS shift of heavy atoms from the initial model position of 1.968 Å. The first glycolate binding model proposed by Lindqvist et al\textsuperscript{12} suggested the formation of a strong ionic interaction with Arg-257 and a hydrogen bond to Tyr-24 from the glycolate carboxylic group, and a hydrogen bond between the glycolate hydroxyl group and Tyr-129. Later studies on mutated GOX\textsuperscript{13} (in which Tyr-129 was altered for a Phe group) found Tyr-129 to be non-essential for substrate binding, while another investigation\textsuperscript{14} found Tyr-24 to be necessary for binding the substrate. The final position obtained here for glycolate using hard-sphere repulsion is in agreement with the findings regarding a Tyr-129 interaction. There is a weak electrostatic interaction between the ligand and the sequence Gln-127, Leu-128, Tyr-129 (\(-5.129\) kJ mol\(^{-1}\)). However, this position favours strong interactions with the sequence His-254, Gly-255, Ala-256, Arg-257 (\(-283.9\) kJ mol\(^{-1}\)) and Arg-164 (\(-295.9\) kJ mol\(^{-1}\)) rather than a hydrogen bond to Tyr-24, as in the Lindqvist model.\textsuperscript{12} The ligand/Tyr-24 electrostatic interaction energy is \(-17.172\) kJ mol\(^{-1}\), over a separation of 3.23 Å, which appears to be too long for the formation of a hydrogen bond. This is not supported by experiment evidence from a mutant GOX, in which Tyr-24 was replaced by phenylalanine\textsuperscript{14} (this mutation was found to have a significant effect on substrate binding) suggesting that this binding orientation is wrong.

The final positions obtained after optimization of glycolate within the binding site using DMAs and each set of repulsion-dispersion parameters (SYBYL, WCMP, CVFF) are in good agreement, with a comparable RMS shift in each case between the initial and final positions. As can be seen in table 2, these additional final positions agree closely on the relative total interaction energies between each residue or sequence and the substrate. Compared to the hard-sphere position, the WCMP final position is shifted towards Arg-164 and Tyr-24, with consequent increased interactions. The Tyr-24 interaction at the
WCMP final position for glycolate is double that of the hard-sphere final position. WCMP final position predicts a similar interaction with the Ser-106, Ser-107, Trp-108 tripeptide to that found with hard-sphere repulsion. The CVFF final position favours increased interactions with Tyr-24 and Arg-257, but a reduced interaction with Arg-164 compared to the hard-sphere final position. The SYBYL total interaction is exaggerated compared to the other final position energies. This is due to unrealistically short intermolecular interactions (the O1'--NH2(Arg-257) separation is only 2.01 Å). Consequently, this position appears to be unreliable.

All of the final positions obtained using the different repulsion-dispersion potentials predict a strong interaction with Tyr-24 within an energy range that would be expected for a hydrogen bond (−16.2 to −57.4 kJ mol⁻¹). The WCMP and CVFF final positions for glycolate agree with the findings of Stenberg et al⁴⁴ that Tyr-24 participates in binding the substrate. In addition all of the final positions (excluding the SYBYL result) predict a significant interaction between the substrate and Trp-108, and agree closely on the magnitude of this interaction. This supports experimental observations⁴⁴ which suggest Trp-108 plays a larger role in the GOX binding site, beyond steric constraint.

### 5.7.2 Oxalate - Lindqvist Model

The starting position for the binding site calculation was produced by fitting oxalate to the positions of three binding site water molecules (following the Lindqvist model⁵³) whose distribution mimics an arrangement which could allow a carboxylic group to bind in the GOX site (figure 10, table 3). Two distinct minima have been found for oxalate using the different repulsion-dispersion parameters. The hard-sphere calculation converges to a minimum with an RMS shift of heavy atom positions from the initial position of 1.91 Å. In this position one carboxylate group of oxalate is 'end-on' to Arg-257 and Tyr-24, producing strong interactions with the His-254, Gly-255, Ala-256, Arg-257 sequence (−416.407 kJ mol⁻¹), Arg-164 (−578.682 kJ mol⁻¹) and a weaker interaction with Tyr-24 (−29.826 kJ mol⁻¹). Macheroux et al⁴³ proposed that oxalate is a transition state analogue for glycolate, and their studies on the binding of oxalate indicate that the inhibitor interacts with, and is stabilized by Tyr-129. At the hard-sphere
minimum, the Tyr-129–oxalate separation is 3.81 Å, with a weak electrostatic interaction (−19.613 kJ mol⁻¹). The orientation of the CVFF final position is ‘end-on’ to Arg-257 and Arg-164, in agreement with the hard-sphere final position. The strongest interactions at this position are between oxalate and Arg-164 (−535.4 kJ mol⁻¹), the His-254, Gly-255, Ala-256, Arg-257 sequence (−503.8 kJ mol⁻¹) and Tyr-24 (−61.9 kJ mol⁻¹). This position is separated from Tyr-129 by 4.15 Å. The ‘end-on’ orientation is similar to the orientation of the initial model position, but is shifted away from Tyr-129 in both the hard-sphere and CVFF calculations. In each case, this shift precludes the formation of a significant interaction between Tyr-129 and oxalate.

The position and orientation of the WCMP and SYBYL minima are equivalent, and differ completely from the minimum obtained using the hard-sphere repulsion. The WCMP oxalate position is ‘face-on’ to Arg-164 and Arg-257, rather than ‘end-on’, like the hard-sphere final position. At this minimum for oxalate, both carboxylate groups can interact with the two positively charged residues Arg-257 and (further out in the binding site passage) Arg-164. At this ‘face-on’ minimum position proposed by the SYBYL and WCMP potentials oxalate could not interact with Tyr-129, which contradicts the experimental evidence.¹⁵ The smallest separation between the ligand and Tyr-129 is 6.37 Å.

5.7.3 Pyrroledione - Docking (WCMP Repulsion-Dispersion)

For pyrroledione, the initial position indicates three possible interactions between the ligand and the binding site residues, His 254 NE2/N1-H, Arg-257 NH2-HH2’ and NE-HNE/O5, and Tyr-129 OH/O2. This position also correctly aligns the ligand R-group so that it occupies the steric space created by the binding site channel. After the docking calculation using ORIENT3, a slightly shifted minimum position is found, but the overall orientation is in agreement with the initial position (figure 11a). The three possible interactions between the ligand and the binding site are conserved. There are two attractive interactions between the ligands and the binding site atoms, His 254 NE2/N1-H (−16.3 kJ mol⁻¹), Arg-164 NH2-HH2’, NE-HNE/O5 (−47.6 kJ mol⁻¹) with a weakly repulsive interaction between Tyr-129 OH/O2 (4.6 kJ mol⁻¹). The main difference between the two positions is that the ORIENT3 position has moved out of the binding
site slightly relative to the initial position.

5.7.4 Isothiazolinone - Docking (CVFF Repulsion-Dispersion)

This docking orientation for this inhibitor is only reported using the CVFF repulsion-dispersion parameters, because the WCMP repulsion-dispersion potential is not parameterized for sulphur atoms. The CVFF final position for isothiazolinone favours three strong interactions with Tyr-129, His-254 and Arg-257, with another weaker interaction with Arg-164 (figure 11b). The O1 of isothiazolinone lies between the planes of the guanidinium groups of Arg-164 and Arg-257. The major difference between the initial and final positions is the shift towards Tyr-129.

5.7.5 Dioxobutanoic Acid Conformations - Docking (WCMP Repulsion-Dispersion)

Dioxobutanoic Acid

The initial position has three possible interactions, Tyr-129/O1, His-254 NE2/O1'-H, and Arg-257/O4. The ORIENT3 final position maintains all three of these interactions with Tyr-129/O1 (−19.2 kJ mol⁻¹), His-254 NE2/O1'-H (−49.3 kJ mol⁻¹) although this distance is somewhat long, and Arg-257/O4 (−49.3 kJ mol⁻¹), with a weaker interaction with Tyr-24 (−7.9 kJ mol⁻¹) (figure 12a). There are some slightly repulsive interactions between the ligand and Arg-164 (1.914 kJ mol⁻¹), and the FMN cofactor (16.3 kJ mol⁻¹).

Dioxo_insight

The initial and final ORIENT3 positions are in good agreement for dioxo_insight. The largest individual atom displacements occur at O1'-H, O2 and O4. Two of these atom shifts, at O1'-H and O4, are in favour of strong interactions at His-254 NE2 (−97.8 kJ mol⁻¹) and Tyr-24 OH (−11.7 kJ mol⁻¹). Favourable electrostatic interactions also occur with Arg-164 (−11.9 kJ mol⁻¹). There is an additional, somewhat weaker interaction between O2 and Tyr-129 OH (−9.3 kJ mol⁻¹), and an electrostatic repulsion with Trp-108 (10.0 kJ mol⁻¹) (figure 12b).

Dioxo_sybyl

For this conformation, the initial position favours strong interactions between the ligand
and the two binding site arginine residues, specifically Arg-164/O4 and Arg-257/O1. The Tyr-129/O1′ separation is 5.09 Å. From this initial position, ORIENT3 converges to a minimum which conserves the arginine interactions, Arg-164/O4 (−26.8 kJ mol⁻¹) and Arg-257/O1 (−52.5 kJ mol⁻¹). However, there is only a slight interaction with His-254, and a weakly repulsive electrostatic interaction with Tyr-129 (3.7 kJ mol⁻¹) (figure 13).

5.8 Conclusions II

After the completion of this work, some doubt has arisen regarding the protonation state of the inhibitors studied here. Recent work⁷ has concluded that the GOX inhibitors are more likely to exist in an anionic form within the binding site (except for methylhydantoin), although the active site of a protein is not like the bulk solution and the 'local pH' could be very different. For this reason the manual dockings reported for the neutral forms of pyrroledione, isothiazolinone and dioxobutanoic acid will be somewhat different to the binding orientations which would be found for the anionic inhibitors. However, the results of this study, comparing the electrostatic properties of the isolated neutral inhibitors to the orientation of the neutral inhibitors in the binding site remain internally consistent. The doubt regarding the protonation state of the inhibitors does not affect the comparison between the two sets of relative orientations, or the conclusions drawn from them. A second consideration raised by this recent work⁷ was the apparent lack of interaction between the inhibitor ‘spearheads’ and Arg-164. This has been clarified by the recent resolution of the crystal structure of a designed inhibitor bound within the active site of GOX. This crystal structure revealed substantial movement of the Arg-164 sidechain in order to accommodate and stabilized the designed ‘spearhead’. Such ‘induced fit’ effects, in common with many methods for determining relative binding orientations, are not accounted for by the methods used in this work.

5.8.1 Comparison of Relative Overlays and Optimized Binding Site Positions

A comparison can be made between the final positions for each ‘spearhead’ inhibitor obtained within the binding site and the relative orientations obtained by superposition of the electrostatic extrema of each ligand onto pyrroledione. The overlays were used as starting points for the larger inhibitors (relative to pyrroledione), but in each case there
was movement in favour of specific binding site interactions. In both cases the regions of positive and negative electrostatic potential are in broad agreement. A comparison of the positions reveals enough similarity to show the value of the method of matching electrostatic extrema in cases where the binding site structure is unavailable.

The four point overlay and final orientation obtained within the binding site for isothiazolinone are broadly equivalent, with an RMS shift of 1.03 Å between the two positions (figure 14a). This is not unexpected, since the electrostatic extrema used to produce the relative overlays also pick out the same GOX binding site residues: that is Tyr-129, His-254 and Arg-257. However, given the similarity of the structures of these two inhibitors, a structural overlay of the ring systems of pyroledione and isothiazolinone would also have predicted a similar relative overlay.

For pyroledione and dioxobutanoic acid there is no apparent structural similarity. The RMS shift between the four point extrema overlay and the final binding site position is 1.58 Å (figure 14b). The most apparent difference between the relative orientations is that as a result of the docking, the dioxobutanoic acid has moved further out of the binding site relative to its four point overlay position. The inhibitor has also rotated slightly to increase the interaction between O4 and Tyr-24.

Relative to pyroledione, dioxo_insight has experienced an RMS shift of 1.37 Å between its four point overlay and final binding site position (figure 14c). The main influence of the binding site residues on the docked position is a similar rotation to that seen for dioxobutanoic acid, in favour of the Tyr-24 interaction, and relieving repulsion between Arg-257 and O4, and His-254 and O1'. The final orientations of dioxobutanoic acid and dioxo_insight are very similar, with an RMS difference in the position of equivalent heavy atoms of 0.70 Å.

For dioxo_sybyl there is a larger difference between the four point overlay position and the final binding site position (figure 14d). The RMS shift of 2.24 Å between the two positions reflects the fact that the four point overlay used as the initial orientation within the binding site was unsatisfactory. For this inhibitor, even the best possible four point overlay was relatively poor, with an unconvincing overlay of extrema. The final orientation of dioxo_sybyl within the binding site is rather different to those of the other conformations, as a result of the conformation of the C1-O1'-O1 carboxylic
group. In dioxobutanoic acid, the O1' -H bond is able to interact with His-254 NE2. The effect of this is to cause the structure to pivot around the C3 position, relative to dioxo_sybyl, so that the C5 methyl group lies lower down in the binding site passage. The overall direction of the methyl group along the binding site passage remains the same.

For the dioxobutanoic acid conformations, although the binding site orientations are shifted from the chosen four point overlays relative to pyrroledione, in each case alternative extrema overlays exist which more closely match the final binding site orientations with respect to pyrroledione (figure 15). For dioxo_sybyl there is one three point overlay close to the binding site final position, displaced by an RMS shift in heavy atom positions of 1.91 Å (figure 15c). This position was previously disregarded because of strong clashes between two pairs of extrema, each separated by 2.85 Å. For dioxo_insight, there are two alternative three point overlays, which give a subsequent alternative four point overlay (figure 15b). In the initial assessment of the overlays relative to pyrroledione, these alternatives were discarded due to potential clashes between extrema of the opposite sign. However, these clashes occur close to the 3.0 Å limit defined as the maximum separation between extrema which can be considered to be corresponding. A small movement of the ligand within the binding site could relieve such an unfavourable overlay of extrema. The RMS shifts between these alternative three and four point overlays and the binding site final positions are 0.58 Å and 0.59 Å respectively. This is better than the RMS shift between the superpositioned and final binding site positions for isothiazolinone relative to pyrroledione, where there is a high degree of structural similarity. For dioxobutanoic acid, there is one alternative three point overlay, and one alternative four point overlay (figure 15a). These were, again, disregarded as a result of slight clashes between the overlaid extrema. These results suggest that the conformation of 2,4-dioxo-1-butanoic acid which is most likely to bind to GOX is dioxo_insight. It gives the closest agreement between the bound orientation and relative orientation produced by matching electrostatic extrema.
5.8.2 Comment on the Lindqvist Glycolate Binding Model

The starting position for the substrate was produced by matching the three glycolate oxygen atoms O1, O1', and O2 to the binding site water molecules 408, 419 and 651. This model position proposes the formation of a strong ionic interaction with Arg-257 and a hydrogen bond to Tyr-24. It should be noted that the Lindqvist model starting position does not correspond to a minimum for glycolate (or oxalate) with any of the repulsion-dispersion parameters sets used here. However, despite different assumptions in the different force fields, the final positions obtained for glycolate are displaced from the model position by the same relative shift, and are in broad agreement. These positions vary between favouring interactions with Tyr-24, Arg-164 and Arg-257. Experimental evidence has shown that Tyr-129 is not involved in binding glycolate, and all of the final docking positions obtained here for glycolate are in agreement with this finding. The separation of over 5.0 Å between the substrate and Tyr-129 discounts the possibility that a significant interaction could take place. The final position for each repulsion-dispersion model estimates an interaction energy of between -5.1 and -6.1 kJ mol⁻¹ with Tyr-129.

Gerlt et al.¹⁵ ¹⁶ suggest that Tyr-24 (and Arg-257) participate in catalysis, protonating the glycolate carboxylate group thereby lowering the pka so the α-proton can be more easily abstracted by His-254. However, subsequent experimental work by Stenberg et al.¹⁴ has shown that removal of the Tyr-24 hydroxyl group does not significantly affect the rate of reaction, and therefore that Tyr-24 is not involved in stabilizing the transition state. There is less agreement between the final positions for glycolate, using the different repulsion-dispersion parameters, regarding Tyr-24. The hard-sphere final position is too far away from Tyr-24 for the formation of a hydrogen bond, but both the WCMP and CVFF final positions have stronger interactions with Tyr-24, at plausible hydrogen bonding separations. These positions agree with the experimental evidence on the importance of Tyr-24 in binding the substrate.

The oxalate hard-sphere final position has a slight electrostatic interaction with Tyr-129, over a separation of 3.81 Å, which would seem to exclude the occurrence of any bonding interaction. The significant interaction between oxalate and Tyr-129 which is seen experimentally¹³ could not occur at the final positions obtained here with any of the repulsion-dispersion potentials used. Of the two distinct minima found, the 'end-on'
orientation (allowing one carboxylate group of oxalate to interact with Tyr-24 and Arg-257) for the hard-sphere and CVFF potentials agrees closely with the final orientations found for glycolate. This is consistent with evidence that oxalate acts as a transition state analogue for glycolate. However, the WCMP ‘face-on’ final position (allowing both carboxylate groups of oxalate to interact with Arg-164 and Arg-257) supports the original proposal by Schuman and Massey that both carboxylate groups of oxalate form electrostatic interactions with two positively charged residues in the binding site. The energy difference between the hard-sphere/CVFF and WCMP minima is approximately 100 kJ mol$^{-1}$.

5.8.3 Comment on Matching Electrostatic Extrema

The value of the method of matching electrostatic extrema lies in its ability to quickly distinguishing between probable and improbable relative binding orientations of ligands within a hypothetical binding site. The number of relative orientations of two dissimilar ligands which could be accommodated stericly within a binding site is relatively small, if we assume the binding site is not particularly open. Once steric considerations are accounted for, the choice of acceptable relative orientations becomes manageable enough to be examined by manual superposition of the electrostatic extrema.

Any overlay which directly superpositions extrema of opposite electrostatic potential is clearly unlikely, since a particular region of a binding site could not stabilize two ligands with completely different electrostatic properties. An ideal overlay would consist of extrema of the same sign directly superimposed, and no potential clashes. Between these two extremes cases lies the reality, where relative superpositions may include close favourable correspondences of maxima and minima, at the expense of some potential clashes elsewhere. It may be necessary to consider not only the occurrence of clashes, but also their relative importance. A separation of almost 3.0 Å between an electrostatic minima and maxima could be alleviated by a small shift in the position of one ligand relative to another, without significantly affecting the superposition of the three or four matched extrema. This can be seen in the results for the 2,4-dioxo-1-butanoic acid conformations, where relative overlays which more closely matched the final binding site orientation (but which were excluded due to potential clashes at approximately 3.0
Å) can be accommodated within the binding site. This suggests that the 3.0 Å cut-off, which was originally chosen as a reasonable separation between extrema of opposite electrostatic potential, can be reduced and that shorter separations between maxima and minima in relative binding orientations are acceptable. Reducing the cut-off increases the number of acceptable relative binding orientations of a pair of ligands (by excluding more extrema clashes which occur at or just below the 3.0 Å limit), but it also reduces the number of favourable correspondences between extrema of the same sign in any overlay, by excluding all separations between extrema near the 3.0 Å limit. In practice, since these correspondences of extrema would be widely separated they would be relatively unimportant. This means that the method would increasingly focus on those relative binding orientations which contain more favourable correspondences (i.e., extrema of the same sign closer together), where it would be more probable that both molecules could form a binding interaction.

In some cases the relationship between the ligand extrema and the binding site atoms is not clearly defined due to the poor geometry of the ligand/binding site interactions. The final orientations of the inhibitors obtained within the binding site are subject to the consideration that the enzyme binding site and ligand geometries have been held rigid throughout all calculations. This is a necessary approximation because of the uncertainty regarding the conformational changes induced in the binding site upon ligand binding. The absence of flexibility in the enzyme binding site does not affect the comparison between the relative binding orientations produced by matching electrostatic extrema, and the final orientations optimized within the binding site. A rigid binding site is consistent with the method of matching electrostatic extrema, since we are attempting to match points which are fixed relative to one ligand, and provides a useful test of the method.

It is encouraging to note that the final orientations within the binding site are predicted among the range of superpositions generated by matching electrostatic extrema. An exact match would be unlikely, due to subtleties in the electrostatic environment of the binding site. Although matching electrostatic extrema automatically ensures a good steric overlay with a well chosen superposition, the competing effects within the binding site of relieving repulsion and maximising favourable interactions can
never be fully accounted for by isolated molecules. The general agreement between the isolated relative overlays and enzyme-bound overlays suggests that matching extrema does in fact highlight common binding site interactions for the ligands, although a closer examination of the correlation between the electrostatic extrema, and the positions of hydrogen bond donor and acceptor atoms in the binding site is necessary.
**Figure 4.** Electrostatic potential maxima (minima) calculated by interaction with a single negative (positive) point charge of radius 1.4 Å (0.5 Å) for, **a)** glycolate, the natural substrate for GOX and, **b)** pyrroledione.

Key: Electrostatic potential maxima, +; Electrostatic potential minima, −.
Figure 5. 2D projection of the four-point relative orientation of pyrroledione (grey) with, a) isothiazolinone (full colour) and, b) methylhydantoin (full colour). The four explicitly matched pairs of extrema are highlighted as bold crosses, and their separation is indicated. For clarity the reference inhibitor, pyrroledione, is shown in grey only.

Key: Minima, + (pyrroledione), + (isothiazolinone/methylhydantoin); Maxima, + (pyrroledione), + (isothiazolinone/methylhydantoin).
Figure 6. 2D projection of the four-point relative orientation of pyrroledione (grey) with, a) dioxobutanoic acid (full colour) and, b) dioxo_insight (full colour). The four explicitly matched pairs of extrema are highlighted as bold crosses, and their separation is indicated. For clarity the reference inhibitor, pyrroledione, is shown in grey only.

Key: Minima, + (pyrroledione), + (dioxobutanoic acid/dioxo_insight);
Maxima, + (pyrroledione), + (dioxobutanoic acid/dioxo_insight).
Figure 7. 2D projection of the four-point relative orientation of pyrroledione (grey) with, 
(1) dioxo sybyl (full colour). The four explicitly matched pairs of extrema are highlighted 
as bold crosses, and their separation is indicated. For clarity the reference inhibitor,  
pyrroledione, is shown in grey only.

Key: Minimum, + (pyrroledione), + (dioxo sybyl);  
Maximum, + (pyrroledione), + (dioxo sybyl).

Figure 8. A 2D projection of the four-point relative orientations of the GOX inhibitors  
shown relative to pyrroledione (grey) for comparison.

Key: Electrostatic potential minima, +; Electrostatic potential maxima, +.
Figure 9. Final positions for glycolate optimized in the GOX binding site using: a) pseudo hard-sphere repulsion, b) WCMP repulsion-dispersion and, c) CVFF repulsion-dispersion parameters.
Figure 10. Final positions for oxalate optimized in the GOX binding site using: a) pseudo hard-sphere repulsion, b) WCMP repulsion-dispersion and, c) CVFF repulsion-dispersion parameters.
Figure 11. The ORIENT3 optimized orientation for: a) pyrroleidine and, b) isothiazolinone. The shortest separations (in Ångstroms) between the inhibitors and the binding site residues are shown.
Figure 12. The ORIENT3 optimized orientation for: a) dioxobutanoic acid and, b) dioxo_insight. The shortest separations (in Ångstroms) between the inhibitors and the binding site residues are shown.
Figure 13. The ORIENT3 optimized orientation for dioxob_symbyl. The shortest separations (in Ångstroms) between the inhibitor and the binding site residues are shown.
Figure 14. Comparison of the relative orientations determined for: a) isothiazolinone, b) dioxobutanoic acid, c) dioxo_insight and d) dioxo_sybyl, by matching four electrostatic extrema (green), and by optimization within the GOX binding site (orange). Each orientation is shown relative to pyrroledione (grey).

Figure 15. Possible alternative relative binding orientations produced by matching electrostatic extrema (full colour), which approximately match the inhibitor orientation determined by optimization within the GOX binding site (orange), shown relative to pyrroledione (grey). The RMS difference in the position of the heavy atoms between the optimized orientation and the alternative extrema overlay is given for: a) dioxobutanoic acid, b) dioxo_insight, c) dioxo_sybyl.
<table>
<thead>
<tr>
<th>INHIBITOR</th>
<th>CORRESPONDING MAXIMA AND MINIMA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Nearest Inhibitor Atom</td>
</tr>
<tr>
<td>Isothiazolinone</td>
<td>O(_3)</td>
</tr>
<tr>
<td></td>
<td>O(_4)</td>
</tr>
<tr>
<td></td>
<td>N(<em>2)-H(</em>{\text{mm}})</td>
</tr>
<tr>
<td></td>
<td>O(<em>4)-H(</em>{\text{mm}})</td>
</tr>
<tr>
<td></td>
<td>C(<em>8)-H(</em>{3\text{mm}})</td>
</tr>
<tr>
<td></td>
<td>C(<em>6)-H(</em>{5\text{mm}})</td>
</tr>
<tr>
<td>Methylhydantoin</td>
<td>O(<em>2)-H(</em>{\text{mm}})</td>
</tr>
<tr>
<td></td>
<td>O(<em>5)-H(</em>{\text{mm}})</td>
</tr>
<tr>
<td></td>
<td>N(<em>1)-H(</em>{\text{mm}})</td>
</tr>
<tr>
<td></td>
<td>N(<em>3)-H(</em>{\text{mm}})</td>
</tr>
<tr>
<td>Dioxobutanoic Acid</td>
<td>O(<em>2)-H(</em>{\text{mm}})</td>
</tr>
<tr>
<td></td>
<td>O(<em>4)-H(</em>{\text{mm}})</td>
</tr>
<tr>
<td></td>
<td>O(<em>1)-H(</em>{\text{mm}})</td>
</tr>
<tr>
<td></td>
<td>C(<em>5)-H(</em>{\text{mm}})</td>
</tr>
<tr>
<td>Dioxo_insight</td>
<td>O(<em>1)-H(</em>{\text{mm}})</td>
</tr>
<tr>
<td></td>
<td>O(<em>1)-H(</em>{\text{mm}})</td>
</tr>
<tr>
<td>Dioxo_sybyl</td>
<td>O(<em>1)-H(</em>{\text{mm}})</td>
</tr>
<tr>
<td></td>
<td>O(<em>4)-H(</em>{\text{mm}})</td>
</tr>
<tr>
<td></td>
<td>O(<em>1)-H(</em>{\text{mm}})</td>
</tr>
<tr>
<td></td>
<td>C(<em>5)-H(</em>{\text{mm}})</td>
</tr>
</tbody>
</table>

\(a\) The relative strength of each pair of corresponding extrema is classified according to the definition: strong - both extrema have \(|V| > 50 \text{ kJ mol}^{-1}\); moderate - only one extrema has \(|V| > 50 \text{ kJ mol}^{-1}\); weak - both extrema have \(|V| < 50 \text{ kJ mol}^{-1}\).

\(b\) Relative binding orientations quantified by the minimum root mean square of the separation of the first four corresponding pairs of matched points (above the dashed line). All other pairs of contacts within 3Å are also given.
Table 2. Interaction Energies for Glycolate at Lindqvist model position and final positions for four different repulsion-dispersion parameter sets.

<table>
<thead>
<tr>
<th>Repulsion Model</th>
<th>Interaction Energy at Initial (Lindqvist) Position /kJ mol⁻¹</th>
<th>Interaction Energy at Final Position /kJ mol⁻¹</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hard-Sphere+DMA</td>
<td>Sybyl+DMA</td>
</tr>
<tr>
<td>Thr155-Val156-Asp157</td>
<td>130.731</td>
<td>130.364</td>
</tr>
<tr>
<td>Lys230</td>
<td>-141.488</td>
<td>-141.550</td>
</tr>
<tr>
<td>Blocking Residues</td>
<td>1.275</td>
<td>-3.635</td>
</tr>
</tbody>
</table>

Total Energy           | 5206.015        | -548.531  | -552.706 | -578.021 |
RMS Shift/Å             | 1.968           | 2.002     | 2.345    | 2.134    |

Interaction Energy at Final Position /kJ mol⁻¹

<table>
<thead>
<tr>
<th>Repulsion Model</th>
<th>Interaction Energy /kJ mol⁻¹</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hard-Sphere+DMA</td>
</tr>
<tr>
<td>Tyr24</td>
<td>-16.238</td>
</tr>
<tr>
<td>Gln127-Leu128-Tyr129</td>
<td>-5.129</td>
</tr>
<tr>
<td>Thr155-Val156-Asp157</td>
<td>116.892</td>
</tr>
<tr>
<td>Lys230</td>
<td>-122.776</td>
</tr>
<tr>
<td>His254-Gly255-Ala256-Arg257</td>
<td>-283.997</td>
</tr>
<tr>
<td>Blocking Residues</td>
<td>0.642</td>
</tr>
</tbody>
</table>

Total Energy           | -637.993         | -810.030  | -688.768 | -721.328 |

a Thr 158, Leu 161, Asp 167, and Phe 172: a set of 4 residues included as repulsion-dispersion sites only, with no DMA.
b This repulsion can be attributed to 3 atoms of FMN; C4, C5A and C6, which are each approximately 0.1 Å too close to C1 of glycolate.
c Root-mean-square of the difference between the initial and final position for all heavy atoms of glycolate.
d SYBYL total interaction energy at the final position is exaggerated due to unphysically short intermolecular interactions.
Table 3. Interaction Energies for Oxalate at Lindqvist model position and final positions for four different repulsion-(dispersion) parameter sets.

<table>
<thead>
<tr>
<th>Repulsion Model</th>
<th>Hard-Sphere+DMA</th>
<th>Sybyl+DMA</th>
<th>WCMP+DMA</th>
<th>CVFF+DMA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tyr24</td>
<td>10.221</td>
<td>8.214</td>
<td>7.815</td>
<td>5.472</td>
</tr>
<tr>
<td>Ser106-Ser107-Trp108</td>
<td>128.891</td>
<td>1.616</td>
<td>0.574</td>
<td>-3.854</td>
</tr>
<tr>
<td>Thr155-Val156-Asp157</td>
<td>273.017</td>
<td>272.623</td>
<td>272.480</td>
<td>272.134</td>
</tr>
<tr>
<td>Arg164</td>
<td>-434.722</td>
<td>-435.697</td>
<td>-436.460</td>
<td>-437.952</td>
</tr>
<tr>
<td>His254-Gly255-Ala256-Arg257</td>
<td>-293.877</td>
<td>-436.240</td>
<td>-435.869</td>
<td>-444.986</td>
</tr>
<tr>
<td>FMN</td>
<td>7171.392</td>
<td>21.099</td>
<td>18.746</td>
<td>2.202</td>
</tr>
<tr>
<td>Blocking Residues(a)</td>
<td>0.644</td>
<td>-3.678</td>
<td>-4.280</td>
<td>-5.778</td>
</tr>
<tr>
<td><strong>Total Energy</strong></td>
<td><strong>959410.336(b)</strong></td>
<td><strong>-888.138</strong></td>
<td><strong>-887.671</strong></td>
<td><strong>-925.137</strong></td>
</tr>
<tr>
<td><strong>RMS Shift/Å</strong></td>
<td>1.91</td>
<td>3.58</td>
<td>2.97</td>
<td>2.02</td>
</tr>
</tbody>
</table>

Interaction Energy at Final Position kJ/mol

<table>
<thead>
<tr>
<th></th>
<th>Tyr24</th>
<th>-29.814</th>
<th>-81.895</th>
<th>-72.947</th>
<th>-52.699</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thr155-Val156-Asp157</td>
<td>249.520</td>
<td>230.088</td>
<td>225.599</td>
<td>237.002</td>
<td></td>
</tr>
<tr>
<td>Arg164</td>
<td>-574.943</td>
<td>-801.929</td>
<td>-671.996</td>
<td>-540.821</td>
<td></td>
</tr>
<tr>
<td>Lys230</td>
<td>-258.246</td>
<td>-234.655</td>
<td>-230.489</td>
<td>-245.593</td>
<td></td>
</tr>
<tr>
<td>His254-Gly255-Ala256-Arg257</td>
<td>-413.021</td>
<td>-592.268</td>
<td>-468.614</td>
<td>-495.271</td>
<td></td>
</tr>
<tr>
<td>FMN</td>
<td>21.201</td>
<td>-22.482</td>
<td>-12.996</td>
<td>-1.605</td>
<td></td>
</tr>
<tr>
<td>Blocking Residues</td>
<td>1.080</td>
<td>-3.609</td>
<td>-5.425</td>
<td>-8.074</td>
<td></td>
</tr>
<tr>
<td><strong>Total Energy</strong></td>
<td><strong>-1036.944</strong></td>
<td><strong>-1530.574</strong></td>
<td><strong>-1269.491</strong></td>
<td><strong>-1172.923</strong></td>
<td></td>
</tr>
</tbody>
</table>

\(a\) Thr 158, Leu 161, Asp 167, and Phe 172: a set of 4 residues included as repulsion-dispersion sites only, with no DMA.

\(b\) This repulsion can be attributed to OH of Tyr-129 which is approximately 0.1 Å too close to C1 of glycolate.

\(c\) Root-mean-square of the difference between the initial and final position for all heavy atoms of glycolate.

d SYBYL total interaction energy at the final position is exaggerated due to unphysically short internmolecular interactions.
Table 4. ORIENT3 final positions for inhibitors within the GOX binding site.

<table>
<thead>
<tr>
<th>Contact</th>
<th>Separation* /Å</th>
<th>Electrostatic Interaction Energy /kJ mol⁻¹</th>
<th>Donor Atom</th>
<th>V /kJ mol⁻¹</th>
<th>Distance from Donor H /Å</th>
<th>Nearest Electrostatic Minima</th>
<th>V /kJ mol⁻¹</th>
<th>Distance from Acceptor /Å</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pyrroldione</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>O2–H-OH (Tyr129)</td>
<td>2.66</td>
<td>4.628</td>
<td>O2_e</td>
<td>-102.8</td>
<td>1.77</td>
<td>OH-H_e (Tyr129)</td>
<td>51.1</td>
<td>1.80</td>
</tr>
<tr>
<td>(His254) NE2–H-N1</td>
<td>3.16</td>
<td>-16.330</td>
<td>NE2_e (His254)</td>
<td>-236.2</td>
<td>1.36</td>
<td>N1-H_e (His254)</td>
<td>80.8</td>
<td>1.44</td>
</tr>
<tr>
<td>O5–HE-NE (Arg164)</td>
<td>2.75</td>
<td>-47.686</td>
<td>O5_e</td>
<td>-129.7</td>
<td>0.51</td>
<td>NE1-H_e (Arg164)</td>
<td>452.3</td>
<td>1.51</td>
</tr>
<tr>
<td>Isothiazolinone</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>O2–H-OH (Tyr129)</td>
<td>2.45</td>
<td>-36.882</td>
<td>O2_e</td>
<td>-188.9</td>
<td>1.01</td>
<td>OH-H_e (Tyr129)</td>
<td>51.1</td>
<td>1.27</td>
</tr>
<tr>
<td>(His254) NE2–H-N1</td>
<td>3.13</td>
<td>-108.954</td>
<td>NE2_e (His254)</td>
<td>-236.2</td>
<td>0.97</td>
<td>N2-H_e (His254)</td>
<td>106.8</td>
<td>1.13</td>
</tr>
<tr>
<td>O1–NE-NE (Arg257)</td>
<td>2.89</td>
<td>-108.954</td>
<td>O1_e</td>
<td>-69.5</td>
<td>2.07</td>
<td>NE1-H_e (Arg257)</td>
<td>435.1</td>
<td>0.61</td>
</tr>
<tr>
<td>O1–HE-NE (Arg164)</td>
<td>3.71</td>
<td>-11.043</td>
<td>O1_e</td>
<td>-69.5</td>
<td>2.18</td>
<td>NE1-H_e (Arg164)</td>
<td>452.3</td>
<td>3.10</td>
</tr>
<tr>
<td>Dioxobutanoic Acid</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>O1–H-OH (Tyr129)</td>
<td>2.81</td>
<td>-19.248</td>
<td>O1_e</td>
<td>-131.7</td>
<td>1.46</td>
<td>OH-H_e (Tyr129)</td>
<td>51.1</td>
<td>1.65</td>
</tr>
<tr>
<td>(His254) NE2–H-O1'</td>
<td>3.34</td>
<td>-49.388</td>
<td>NE2_e (His254)</td>
<td>-236.2</td>
<td>0.93</td>
<td>O1'-H_e (His254)</td>
<td>98.1</td>
<td>1.83</td>
</tr>
<tr>
<td>O4–H-NH2 (Arg257)</td>
<td>3.28</td>
<td>-49.388</td>
<td>O4_e</td>
<td>-127.0</td>
<td>1.63</td>
<td>NE1-H_e (Arg257)</td>
<td>435.1</td>
<td>1.48</td>
</tr>
<tr>
<td>O4–OH (Tyr24)</td>
<td>2.71</td>
<td>-7.931</td>
<td>O4_e</td>
<td>-124.7</td>
<td>2.38</td>
<td>OH-H_e (Tyr24)</td>
<td>118.3</td>
<td>1.71</td>
</tr>
<tr>
<td>O4–NH2 (Arg164)</td>
<td>3.14</td>
<td>1.914</td>
<td>O1_e</td>
<td>-127.0</td>
<td>1.29</td>
<td>NE1-H_e (Arg164)</td>
<td>452.3</td>
<td>3.31</td>
</tr>
<tr>
<td>Dioxobutanoic Acid (dioxo_insight)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>O2–H-OH (Tyr129)</td>
<td>3.30</td>
<td>-9.324</td>
<td>O2_e</td>
<td>-131.8</td>
<td>1.56</td>
<td>OH-H_e (Tyr129)</td>
<td>51.1</td>
<td>2.73</td>
</tr>
<tr>
<td>(His254) NE2–H-O1'</td>
<td>2.69</td>
<td>-97.819</td>
<td>NE2_e (His254)</td>
<td>-236.2</td>
<td>0.49</td>
<td>O1'-H_e (His254)</td>
<td>57.8</td>
<td>1.01</td>
</tr>
<tr>
<td>O1–NE-NE (Arg257)</td>
<td>3.90</td>
<td>-97.819</td>
<td>O1_e</td>
<td>-103.7</td>
<td>2.41</td>
<td>NE1-H_e (Arg257)</td>
<td>435.1</td>
<td>1.67</td>
</tr>
<tr>
<td>O4–H-NH2 (Arg257)</td>
<td>4.30</td>
<td>-97.819</td>
<td>O4_e</td>
<td>-130.4</td>
<td>1.40</td>
<td>NE1-H_e (Arg257)</td>
<td>435.1</td>
<td>2.72</td>
</tr>
<tr>
<td>O4–OH (Tyr24)</td>
<td>2.71</td>
<td>-11.765</td>
<td>O4_e</td>
<td>-130.4</td>
<td>2.74</td>
<td>OH-H_e (Tyr24)</td>
<td>118.3</td>
<td>1.68</td>
</tr>
<tr>
<td>O4–NH2 (Arg164)</td>
<td>3.02</td>
<td>-11.858</td>
<td>O4_e</td>
<td>-130.4</td>
<td>2.89</td>
<td>NE1-H_e (Arg164)</td>
<td>452.3</td>
<td>3.31</td>
</tr>
<tr>
<td>Dioxobutanoic Acid (dioxo_sbybx)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>O1–H-NH2 (Arg257)</td>
<td>2.93</td>
<td>-52.561</td>
<td>O1_e</td>
<td>-136.8</td>
<td>1.83</td>
<td>NE1-H_e (Arg257)</td>
<td>435.1</td>
<td>1.25</td>
</tr>
<tr>
<td>O4–H-NH2 (Arg164)</td>
<td>2.71</td>
<td>-26.895</td>
<td>O4_e</td>
<td>-127.9</td>
<td>0.95</td>
<td>NE1-H_e (Arg164)</td>
<td>452.3</td>
<td>2.17</td>
</tr>
</tbody>
</table>

a The separation given is the distance between the heavy atoms involved in each contact.

b Parameters taken from the CVFF force field.
References for Chapter 5.


19. SYBYL. Tripos Associates, 1699 S.Hanley Road, Suite 303, St. Louis, MO 63144, U.S.A.


Chapter 6. Predicting Hydrogen-Bonded Complexes

In this chapter, the positions of maxima and minima in the electrostatic potential at appropriate distances from the van der Waals surface are calculated for various molecules containing groups which can form the biologically important N-H···O=\(\text{C}\) hydrogen bonding interactions, and other hydrogen bond interactions such as N-H···N, C-H···O. Their ability to predict the positions of H and O in hydrogen bonded complexes found by optimising the electrostatic interactions of pairs of such molecules constrained by the molecular shapes is assessed.

6.1 Introduction

The N-H···O=\(\text{C}\) hydrogen bond plays a major role in determining the structure of proteins and DNA, and many of their complexes with small molecules, and so is central to biological activity. However, the detailed 3-dimensional structure of many important binding sites is unknown, and the requirements for binding have to be inferred from known ligands. This process could be facilitated by information on where hydrogen bonds are likely to form around donors and acceptors in the ligand. The aim of this work is to determine whether likely hydrogen bonding sites could be predicted more closely from the electrostatic properties of the isolated molecules, since this includes information about the charge distribution around one of the hydrogen bonding groups.

The steric constraints on an N-H group being the appropriate short distance (N···O ~ 2.9 Å) from the C=O group of the other molecule do not define the relative positions particularly closely. Statistical surveys of the orientational preferences for such hydrogen bonds in small molecule structures show a preference for hydrogen bonding in the lone pair direction, but with a large scatter of orientations.\(^1\) In contrast, within proteins, there is a preference for donor-H···acceptor hydrogen bonds to be linear, even including alpha helix structures where structural constraints usually prevent such linear donor-H···acceptor angles.\(^2\) Protein sidechain C=O groups show the expected tendency for lone pair directionality in the donor···O=\(\text{C}\) angle, but in contrast main chain C=O groups generally have more linear donor···O=\(\text{C}\) angles. This may be due to structural constraints within the protein. While many C=O groups in small molecules can accept two hydrogen bonds, the majority of buried main chain C=O groups only accept one hydrogen bond and so show less pronounced lone pair directionality.\(^2\)
If the unknown hydrogen bond donor could be represented by just a partial positive charge on the H atom, H^+\text{\textsuperscript{d}}, then the hydrogen bond would form with the proton at the minimum in the electrostatic potential at approximately 1.9 Å from the oxygen of the known molecule. Similarly, a partial negative charge model for the unknown hydrogen bond acceptor, O^\text{\textsuperscript{d}}, would result in the oxygen atom being at the site of the potential maximum approximately 2.9 Å from the nitrogen (1.9 Å from the H) of the known molecule. If we refine the model of the hydrogen bonding groups to C^\text{\textsuperscript{d}}=O^\text{\textsuperscript{d}} and N^\text{\textsuperscript{d}}-H^\text{\textsuperscript{d}}, represented as a point dipole on the O and H atoms, then the hydrogen bond would form at the positions of the field maxima at the appropriate distances from the known molecule. Both these molecules are clearly gross simplifications for the electrostatic properties of the unknown molecule, but a more detailed model can only be constructed for known molecules. To investigate whether the simple H^\text{\textsuperscript{d}} and O^\text{\textsuperscript{d}} models are at all useful, we can compare their predictions with the geometries of the N-H...O=C hydrogen bonded complexes corresponding to minima in the electrostatic interaction energy between molecules which represent common biological groups.

This study strictly correlates the electrostatic properties of one molecule with its electrostatic interactions in biomolecular complexes, subject to a hard-sphere model of steric constraints. However, this is expected to be a good guide to the likely hydrogen bonding sites for the pairs of molecules, because the electrostatic interaction often dominates the orientation dependence of hydrogen bonding.\textsuperscript{3}\textsuperscript{4}\textsuperscript{5}

6.2 Methods

6.2.1 Choice of molecules

A range of molecules have been examined which contain N and O groups in a variety of local chemical environments which can form hydrogen bonds; formamide, methylacetamide, adenine, guanine and theophylline. Additionally, four amino acid side chains have been studied; arginine, aspartate, asparagine and histidine (figure 1). The amino acid sidechains have been blocked using N-acetyl and N'-methylamide terminal groups to approximate the charge distribution of the residues in the peptide sequence. The N-acetyl and N'-methylacetamide blocking groups were constructed by substituting methyl groups at the C\textsubscript{t} positions of residues which would precede and follow the selected residue in an \(\alpha\)-helical main chain conformation (\(\phi = -57^\circ\), \(\psi = -47^\circ\)). Finally, an N-acetyl-N'-methylamide blocked diglycine peptide has been considered in two
different conformations, a helical conformation ($\phi_1 = \phi_2 = -57^\circ$, $\psi_1 = \psi_2 = -47^\circ$, labelled ADMc), and an extended conformation, which approximates a $\beta$-sheet structure ($\phi_1 = \phi_2 = 180^\circ$, $\psi_1 = \psi_2 = 180^\circ$, labelled ADM$\beta$) (figure 1). The N-acetyl-N'-methylamide blocked residue is the smallest molecular unit which is adequate for modelling the charge density of the peptide backbone NHCHRCo.$^6$

![Molecules used in this study, defining the numbering system for a) formamide, b) methylacetamide, c) adenine, d) guanine, e) theophylline; and N-acetyl, N'-methylamide blocked residues for, f) arginine, g) aspartate, h) asparagine, i) histidine; and j) N-acetyl, N'-methylamide blocked diglycine shown in the extended $\beta$-sheet conformation with the torsional angles $\phi$ and $\psi$ indicated.](image-url)
6.2.2 CADPAC Ab Initio Calculations

SCF wavefunctions were obtained for each molecule, using the CADPAC\textsuperscript{7} suite of ab initio programs, using a 3-21G\textsuperscript{8} basis set in each case. The Distributed Multipole Analysis\textsuperscript{9,10} of each wavefunction was calculated to represent the charge density by a charge, dipole, quadrupole, octupole and hexadecapole moment at every nuclear position.

6.2.3 Locating Electrostatic Maxima and Minima

Using the program ORIENT3\textsuperscript{11}, the positions and strengths of the maxima (minima) in the electrostatic potential energy were determined by minimising the interaction energy of a single negative (positive) point charge of radius 1.4 Å (0.5 Å) with the molecule under examination, in order to sample the electrostatic potential in the region outside the molecular van der Waals surface which would be occupied by hydrogen bond acceptor and donor groups. Pseudo hard-sphere repulsion was used between all sites with non-zero van der Waals radii. The molecular van der Waals volume was defined using the Pauling radii, \textit{i.e.}, C: 2.0 Å, N: 1.5 Å, O: 1.4 Å. All hydrogens were treated as having a zero van der Waals radius, due to the lack of repulsion between the proton and the acceptor atom in a hydrogen bond.\textsuperscript{12} Non-polar hydrogens, such as methyl hydrogens, are included in a ‘united atom’ carbon radius.

6.2.4 Locating Minimum Energy Structures

To model minimum energy structures for pairs of molecules, the DMA electrostatic model plus a pseudo hard-sphere repulsion, to define sterically accessible conformations, has been used within ORIENT3. During the calculation one molecule is considered to have a fixed position and orientation in space, and the second molecule is given free translation and rotation around the fixed molecule. Minima in the electrostatic interaction energy between the molecules have been located from a variety of starting positions, chosen on a sphere surrounding the fixed molecule.
6.3 Results

The locations of the N-H maxima and C=O minima have been defined according to the angle $\alpha$ (figure 2). The distribution of electrostatic maxima and minima obtained for each molecule under consideration are shown in figures 3 - 5. The electrostatic minima (maxima) have been transformed onto a common orientation, and are shown relative to a ‘reference’ C=O (or N-H) group. In order to standardize this transformation, and examine qualitatively the influence of the surrounding molecular functionalities on the position of electrostatic extrema, all C=O (N-H) groups have been overlaid so that neighbouring hydrogen bond acceptor groups are at $-\alpha$ values, and hydrogen bond donor groups at $+\alpha$ values relative to the C=O (N-H) reference group. The distribution of C=O minima and N-H maxima is shown in figure 6.

\[
\begin{align*}
\text{Figure 2. Definition of the angle } \alpha.
\end{align*}
\]

6.3.1 Distribution of Electrostatic Maxima

The molecules under consideration represent a range of hydrogen bonding functionalities in different steric arrangements. For almost all of the N-H donor groups relevant electrostatic maxima have been located. The electrostatic maxima are normally found along the axis of the N-H donor bond, within a narrow range for the angle at the nitrogen (N-H-maxima = 45.4°, figure 6). There is one outlying value for theophylline N7-H$_{\text{max}}$. This maximum is significantly displaced from the N-H bond axis due to the presence of the neighbouring C6=O6 group. This excludes guanine N2-H$_{\text{max}}$ which is shared between N1-H1 and N2-H1, and arginine NH1/NH2-H$_{\text{max}}$, which is shared between NH1-H2 and NH2-H2. These maxima therefore do not lie close to the axis of
Figure 3. Electrostatic potential maxima (minima) calculated by interaction with a single negative (positive) point charge of radius 1.4 Å (0.5 Å) for, a) formamide, b) methylacetamide and, N-acetyl, N'-methylacetamide blocked diglycine in, c) an α-helical conformation and, d) an extended β-sheet conformation.

Key: Electrostatic potential maxima, +; Electrostatic potential minima, −.
Figure 4. Electrostatic potential maxima (minima) calculated by interaction with a single negative (positive) point charge of radius 1.4 Å (0.5 Å) for, a) adenine, b) guanine and, c) theophylline.

Key: Electrostatic potential maxima, +; Electrostatic potential minima, −.
Figure 5. Electrostatic potential maxima (minima) calculated by interaction with a single negative (positive) point charge of radius 1.4 Å (0.5 Å) for the N-acetyl, N'-methylacetamide blocked residues, a) aspartate, b) asparagine, c) arginine and, d) histidine.

Key: Electrostatic potential maxima, +; Electrostatic potential minima, −.
Figure 6. The distribution of: a) oxygen C=O electrostatic potential minima at 0.5 Å from molecular van der Waals surface and, b) sp² nitrogen N-H electrostatic maxima at 1.4 Å from molecular van der Waals surface, as a function of α. The electrostatic potential minima (maxima) are shown transformed onto a common orientation relative to a reference C=O (N-H) group.
either N-H bond. These are only two examples among the molecules studied here of N-H donor groups separated by two intermediate bonds, so that the N-H groups are parallel and coplanar.

In contrast, the only N-H donor groups for which related electrostatic maxima were not found are N3-H in ADMα, and N1-H and N2-H in the middle of the extended β-sheet ADMβ. The ADMβ donors are both parallel and coplanar with neighbouring C=O groups, separated by two intermediate bonds. Rather than reinforcing the electrostatic potential between the groups (as in the case of the shared maxima) this has lead to a cancellation of potentials of opposite magnitude. The electrostatic potential across this region is relatively featureless and only changes by a small magnitude. Hence, for ADMβ N1-H a shallow displaced maximum relating to the neighbouring methyl group was found, and for N2-H no maximum was located. The lack of a definite electrostatic maximum for the ADMα N3-H is a consequence of the α-helical turn. The electrostatic potential generated by a molecular functionality depends on the conformation of the entire molecule, and can be altered by two effects; significant changes in local electronic structure can occur with changes in the peptide torsional angles φ and ψ (see figure 1), and to a lesser extent, by the polarization of the charge distribution of a peptide residue by the electrostatic fields of more distant parts of the molecule. The α-helical conformation presents two distinct hydrogen bonding aspects to a potential second molecule, characterised by N-H donor groups along the ‘top’ edge of the helical backbone, and C=O acceptors along the ‘bottom’. As a result of the helical turn, N3-H and C=O are brought into close proximity to each other, producing a cancellation of potentials.

6.3.2 Distribution of Electrostatic Minima

Within the range of molecules studied here, N hydrogen bond acceptors are available in adenine, guanine, theophylline and histidine. In each case (except guanine N7, discussed further below) the electrostatic potential minimum is found close to the axis which bisects the bond angle at the \( sp^2 \) nitrogen, in the direction of the N lone pair orbital. For N minima relative to the line bisecting the \( sp^2 \) bond angle, the spread of α values = ±10.75° (excluding guanine N7).

For the carbonyl oxygen acceptor groups, the relative position of the electrostatic potential minimum is more variable (figure 6). The position of the
minimum on an arc between the lone pair orbitals along the van der Waals surface of the oxygen atom is highly influenced by the local chemical environment, \textit{ie.}, the neighbouring molecular functionalities. This is particularly apparent in the case of guanine C6=O6. Although a distinct minimum was located, its position is closely related to the nearby N7 minimum. The proximity of these two hydrogen bond acceptors to each other creates a broad region of negative electrostatic potential between them. Indeed the electrostatic potential outside the van der Waals surface of oxygen acceptors changes very slowly. Mitchell and Price\textsuperscript{14} have shown that the value of the electrostatic potential is fairly insensitive to the angle at the carbonyl oxygen, \( \alpha \), varying by approximately 4 kJ \text{mol}^{-1} along a 110° arc (the region between the two \( sp^2 \) oxygen lone pairs) for the N-H–O=C hydrogen bonded trans- formamide/formaldehyde complex. As a result, electrostatic minima relating to oxygen acceptors are easily displaced.

Despite this variability of the C=O minimum position, the positions of minima relative to carbonyl oxygen acceptors are clustered in a tighter arc than that defined by the lone pair orbitals. The range for the angle at the carbonyl oxygen C=O-minima = 59.4°, excluding the positions of the minima guanine O6_{min}, and aspartate OD1_{min} and OD2_{min}.

\textbf{Figure 7.} The correspondence between the positions of electrostatic extrema, and the hydrogen bond geometry of the minimum energy complexes, illustrated for formamide and methylacetamide. The quality of the prediction is considered in terms of, \textbf{a}) the separation between the hydrogen bond proton and the acceptor electrostatic minimum and, \textbf{b}) the separation between the hydrogen bond acceptor and the donor electrostatic maximum.

Key: N1-H1\textsubscript{Pem} electrostatic maximum, \(+\); O3\textsubscript{Meh} electrostatic minimum, \(+\).
6.3.3 Minimum Energy Structures

Two types of minimum energy structures have been excluded from further analysis. In some cases, the strict convergence criteria of the ORIENT3 minimization procedure was not satisfied, resulting in the occurrence of ‘doubtful minima’. These frequently contained weak C-H- acceptor interactions, or N-H- acceptor interactions with poor hydrogen bond geometry. In addition, three successfully converged minima, each stabilized by a C-H- acceptor interaction, were found. These ‘doubtful minima’ and C-H- acceptor have not been included in any further analysis. For all of the small molecules and peptide units studied here, a total of 116 minimum energy structures containing N-H-O=C or N-H-N hydrogen bonds have been included.

Among the minimum energy structures stabilized by a single hydrogen bond, a secondary, much weaker C-H-O=C contact could often be identified. This is consistent with the observed formation of C-H-O contacts within the constraints of stronger hydrogen bond contacts. There is crystallographic evidence for the hydrogen bond nature of C-H-O contacts. C-H-O contacts may help to distinguish between alternative hydrogen bonds which are ‘geometrically equivalent but structurally distinct’.

Of the amino acid sidechains studied here, arginine, aspartate and asparagine are capable of forming hydrogen bonds to features of base pairs which allow the sidechains to distinguish between different base pairs. Using these examples, we can see how well the method is able to predict these distinct and biologically important hydrogen bond motifs. For some of the complexes with amino acid side chains, additional hydrogen bonds have formed between the base and hydrogen bond donor and acceptor groups on the main chain. These complexes are generally of lower energy, and deviate from ideal hydrogen bond geometry. In the 3D environment of a polypeptide some of these main chain contacts would be unlikely to form, given the presence of neighbouring residues in the peptide sequence. Such complexes have therefore been excluded from this analysis.

The hydrogen bonding interactions for involved in each of the minimum energy structures are detailed in tables 1-5. The relevant separations between the hydrogen bonded acceptors (protons) and maxima (minima) are defined in figure 7. Since we propose that locating electrostatic maxima and minima identifies the positions of
hydrogen bond acceptors and donors in a bimolecular complex, the smallest separations correspond to the best predictions.

6.3.4 Small Molecule Complexes

Formamide - Formamide
Two different dimers were found, containing a single N-H-O=C contact, and a double N-H-O=C contact. For the single N-H-O=C contact dimer, $\alpha = -37.02^\circ$, compared to the position of the electrostatic potential minimum at $\alpha = -21.95^\circ$.

Methylacetamide - Methylacetamide
Four dimers were found, corresponding to the same hydrogen bond, with the second molecule in different orientations. For the fourth dimer, the electrostatic maximum is in excellent agreement with the position of the oxygen acceptor. The hydrogen bond in this dimer is close to linearity with the carbonyl C=O bond, rather than in the direction of the carbonyl oxygen lone pair orbitals.

Methylacetamide - Formamide
Four complexes were found, in which methylacetamide acts as an acceptor in three. Only two of the dimers are planar. Two of the hydrogen bonds to the methylacetamide carbonyl oxygen form on the same side of the C=O bond (at $\alpha = -43.0^\circ$ and $\alpha = -35.4^\circ$ respectively), in approximate agreement with the relative position of the potential minimum, at $\alpha = -11.5^\circ$. The exception to this is the N-H-O=C interaction at $\alpha = +15.8^\circ$. The complex in which formamide acts as the hydrogen bond acceptor is the least stable.

Adenine - Guanine
Ten complexes were found, six of which are stabilized by two simultaneous hydrogen bonds. Since adenine does not contain any oxygen acceptor atoms, the double hydrogen bond contacts are all of the N-H-O=C and N-H-N type. The remaining four complexes are stabilized by a single hydrogen bond with an additional weak C-H-acceptor van der Waals interaction.
Adenine - Theophylline

The only C=O hydrogen bond acceptors for this molecular pair are on theophylline, so all complexes stabilized by double hydrogen bonds involve at least one N-H--N interaction. The theophylline C2=O2 is bracketed by two methyl groups, and as a result of these steric interactions hydrogen bonding is preferred on steric grounds through C6=O6.

Guanine - Theophylline

Five complexes were found. Because both molecules possess oxygen acceptors, a double N-H--O=C hydrogen bonded complex was found. This complex has the most stable interaction energy. Of the remaining four complexes, three are stabilized by a single hydrogen bond, and the fourth is stabilized by a bifurcated hydrogen bond, in which a theophylline oxygen simultaneously accepts two protons from two guanine N-H donors.

6.3.5 ADMα Complexes

ADMα - Formamide

Four different complexes were found. All of the complexes form unique structures with a single hydrogen bond. In two of the complexes formamide donates a proton to ADMα, and in the other two formamide accepts protons from ADMα. For all but the lowest energy complex, the proton lies at the angle of the lone pair orbitals relative to the carbonyl oxygen, but the hydrogen bond is not in the plane formed by the acceptor O=C-Cα atoms.

ADMα - Methylacetamide

Five successfully converged minimum energy structures were found. In two complexes methylacetamide accepts a hydrogen bond from ADMα N1-H in two different orientations. The different orientations of methylacetamide are the result of the proton being accepted by opposite lone pair orbitals of the methylacetamide carbonyl oxygen. In both orientations, methylacetamide lies perpendicular to the plane formed by the ADMα H1-N-Cα. This situation is repeated for ADMα N2-H. In the cases where pairs of methylacetamide molecules form the same hydrogen bond in different orientations, one orientation was determined to be a doubtful minimum. For the remaining complex
with the weakest interaction energy, ADMα accepts a single hydrogen bond from methylacetamide (N-H•O2=C2).

**ADMα - Adenine**

Seven complexes were found overall. Two of the complexes correspond to the same hydrogen bond, with adenine in a different orientation. In every case except the least stable complex, adenine acts as a hydrogen bond donor, because all of the C=O acceptor groups are found on ADMα. All of these interactions involve the formation of single hydrogen bonds through donation of the N9-H proton of adenine (there is one exception, where the N6-H2 proton is involved in the hydrogen bond). The geometry of the hydrogen bonds at the carbonyl oxygen is consistent with the direction of the lone pair electron orbitals in four cases. In the other cases the hydrogen bond is approximately linear at the carbonyl oxygen. When adenine acts as a hydrogen bond donor, it is orientation is approximately co-planar with the plane formed by the ADMα atoms O=C-Cα. This co-planarity is assisted by the formation of a second C-H•••acceptor atom interaction in every case. In the single example of adenine accepting a hydrogen bond from ADMα, adenine lies perpendicular to the ADMα O=C-Cα plane.

**ADMα - Guanine**

Six different complexes were found. A clear distinction can be seen between the final orientations of guanine which accept hydrogen bonds along the ‘top’ edge of the helix, and those which donate hydrogen bonds along the ‘bottom’ edge of the helix. In three cases ADMα donates a hydrogen bonding proton (two of these complexes correspond to the same hydrogen bond, with guanine in different relative orientations, as a result of the steric effects of the helical turn). In two complexes, guanine forms two simultaneous hydrogen bonds to ADMα.

**ADMα - Theophylline**

Three different complexes were found. In all of the complexes, theophylline forms single hydrogen bonds to ADMα through donation of the N7-H proton.

For adenine, guanine and theophylline, complexes in which the base is coplanar with the local helical plane defined by O=C-Cα are the most stable.
6.3.6 ADMβ Complexes

ADMβ - Formamide

Nine complexes were found. The best agreement with extrema is not the best binding energy. Two complexes are stabilized by double N-H-O=C hydrogen bonds. The remaining complexes are stabilized by N-H-O=C interactions. In all but two of the complexes, ADMβ and formamide are coplanar.

ADMβ - Methylacetamide

The minimization successfully converged for four complexes stabilized by single N-H-O=C hydrogen bonds. In the most stable complex, ADMβ donates the proton and the angle at the carbonyl oxygen is linear. For the remaining complexes methylacetamide acts as the proton donor. In the two examples where the hydrogen bond forms at the lone pair angles of the ADMβ acceptor oxygens, methylacetamide is non-planar with respect to ADMβ.

ADMβ - Adenine

Five complexes were found. All are coplanar, single N-H-O=C interactions, with adenine acting as the proton donor. In each case a second, weaker C-H-O interaction is identifiable.

ADMβ - Guanine

Five complexes were found. The most stable complex is stabilized by twin N-H-O=C hydrogen bonds. Two complexes form single N-H-O=C complexes, whilst two further complexes are stabilized by bifurcated hydrogen bonds shared between guanine N1-H1 and N2-H2.

ADMβ - Theophylline

Five complexes were found, all stabilized by single N-H-O=C hydrogen bonds in which ADMβ and theophylline are coplanar. All of these N-H-O=C contacts are linear at the carbonyl oxygen (i.e., $\alpha = 0.0^\circ$) as a result of the steric crowding of the theophylline methyl groups.
6.3.7 Amino Acid Sidechain Complexes

Amino acid sidechains such as arginine, aspartate and asparagine are capable of forming pairs of hydrogen bonds with nucleic acids which enable them to distinguish between different nucleic acids. The geometric complementarity between the nucleic acid and sidechain hydrogen bonding functionalities allows selective recognition between proteins and nucleic acids. The lowest energy complexes between guanine-arginine, asparagine-adenine and asparagine-guanine correspond to those involved in specific recognition between these nucleic acids and sidechains.

Ippolito et al. analyzed the biological stereochemistry of hydrogen bonds in protein structures, in terms of the electronic structure and chemical reactivity of side chain residues. In enzyme binding sites the reaction coordinate of proton transfer generally coincides with the direction of hydrogen bond formation. Arginine contains three hydrogen bond centres, and aspartate and asparagine each contain two. From their survey of hydrogen bonding in proteins, Ippolito et al. described the observed preference for hydrogen bonding with syn or anti stereochemistry to these sidechains (see figure 8), based on the electronic structure of the acceptor atoms, and the steric accessibility of the proton donors. The hydrogen bonded complexes between these sidechains and adenine, guanine and theophylline will be discussed in terms of these preferences.

(a) Aspartate  (b) Asparagine  (c) Arginine

Figure 8. The stereochemistry of hydrogen bonding interactions for: (a) aspartate, (b) asparagine, (c) arginine. The arrows indicate the observed preferred direction for hydrogen bonding interactions, designated as syn or anti. (after Ippolito et al., J.Mol.Biol., 215 (1990) 457.)
Aspartate - Adenine

Five complexes were found, in which adenine donates through its N9-H proton, except for the most weakly stabilized complex. Hydrogen bonds to aspartate can form with either syn or anti-oriented stereochemistry. For these aspartate-adenine complexes there are three are syn single hydrogen bonds, and two are anti hydrogen bonds. One of the complexes in which adenine is anti-oriented is stabilized by an additional hydrogen bond with the peptide backbone.

Aspartate - Guanine

Two complexes were found, coplanar with the aspartate carboxylate group. The most strongly bound complex is stabilized by two simultaneous hydrogen bonds to the carboxylate sidechain. The second complex has a single N-H···O=C hydrogen bond to the OD1 anti lone pair.

Aspartate - Theophylline

Two complexes were found, which correspond to the same single hydrogen bond formed through theophylline N9-H to OD1 and OD2 of the carboxylate sidechain. Both hydrogen bonds form in the direction of the syn lone pair orbitals.

In the survey of hydrogen bonding preferences, aspartate (and glutamate) showed a preference for hydrogen bonding with syn-stereochemistry. However, the ratio of syn/anti-oriented interactions was smaller for aspartate than for glutamate. It seems that the greater flexibility of the glutamate side-chain places fewer restrictions on achieving optimal hydrogen bond geometry. The clustering for anti-oriented hydrogen bonds was found to be more scattered, reflecting the weaker magnitude of these interactions. The electrostatic minima obtained for aspartate correspond to the syn-stereochemistry, and are consistent with the formation of double N-H···O=C contacts. For these results there is a preference for syn-stereochemical interactions.

Asparagine - Adenine

Five complexes were found. Two complexes correspond to a single hydrogen bond between adenine N3, and the anti and syn protons of the asparagine NH2 sidechain group. The remaining complexes form simultaneous N-H···O=C and N-H···N hydrogen bonds with the asparagine carboxamide. Such double hydrogen bonds can only form
when adenine is syn-oriented with respect to asparagine. All orientations of adenine (except for the single hydrogen bond to the anti-hydrogen of the asparagine NH2) are coplanar with the sidechain carboxamide.

**Asparagine - Guanine**

Four complexes were found. The most stable complex is stabilized by two N-H•O=C hydrogen bonds. Two further complexes form double N-H•O=C and N-H•N hydrogen bonds to the C=O and NH2 asparagine sidechain groups. There is only one example of a single hydrogen bond to the anti-proton of the carboxamide NH2, and no interactions with the anti lone pair of the carboxamide C=O sidechain.

**Asparagine - Theophylline**

Two complexes were found in hydrogen bonding contact with the carboxamide sidechain. One complex forms two syn-oriented N-H•O=C hydrogen bonds, while the second forms a single hydrogen bond with the anti proton of the NH2 asparagine group.

Asparagine shows a preference for anti-oriented hydrogen bonds. This preference for anti-oriented interactions is seen for both the NH2 and C=O groups of the carboxamide. Perrin et al report that the anti-proton of primary amides is more acidic than the syn-proton. This would make the anti-proton a more favourable hydrogen bond donor. Acceptor interactions are more tightly clustered for the NH2 donor, than the distribution of donor atoms to the C=O group (This is consistent with the frequent close proximity of electrostatic maxima to acceptor atoms.) While there is no preference for anti-oriented hydrogen bonds in these results, it should be noted that only syn-oriented stereochemistry will allow the formation of simultaneous double hydrogen bond contacts involving both the C=O and NH2 groups of the carboxamide, and Ippolito et al only examine the distribution of single N-H•O=C hydrogen bonds.

**Arginine - Adenine**

Only one complex was found where adenine is in contact with the guanidinium sidechain of arginine. This is not surprising, since arginine is exclusively a hydrogen bond donor, and adenine (which lacks potential oxygen acceptor atoms) is limited in its ability to accept hydrogen bonds, when compared to guanine or theophylline. In the arginine-adenine complex adenine lies perpendicular to the plane of the guanidinium
group. Since adenine cannot accept two hydrogen bonds simultaneously, any orientations which placed adenine and the guanidinium group coplanar would be highly repulsive.

**Arginine - Guanine**

Five complexes were found between the arginine guanidinium group and guanine. For the three complexes stabilized by a single hydrogen bond, guanine is non-planar with respect to the guanidinium sidechain. The two most stable complexes are doubly hydrogen bonded through NH1 and NH2 of arginine, with guanine in different orientations.

**Arginine - Theophylline**

Four complexes were found, all of which are sidechain contacts. In all of the complexes, theophylline is oriented approximately perpendicular to the plane of the guanidinium group. This appears to be a steric effect, maximising the separation between the guanidinium sidechain and the bulky theophylline methyl groups. Three of the N-H-O=C interactions are single hydrogen bonds, in which the angle at the carbonyl oxygen is linear. The fourth dimer forms a bifurcated hydrogen bond with NE and NH2 of arginine.

In the absence of steric interference, syn-hydrogen bonds are not significantly preferred to anti-hydrogen bonds statistically. Arginine is exclusively a hydrogen bond donor. Arginine contains three chemically distinct hydrogens, whose stereochemical interactions have been designated syn, and anti1 and anti2 by Ippolito et al (figure 8), to distinguish between the two anti positions, which can donate 1 and 2 protons respectively. The survey shows a preference for hydrogen bonds which display anti-oriented stereochemistry. However, this is explained by a simple steric argument. For the anti-oriented interactions there is greater steric accessibility, and more available protons.

A survey of arginine-carboxyl intramolecular protein interactions by Singh et al found a preference for in-plane anti-oriented double N-H-O=C hydrogen bond interactions, with a clear preference for twin NH1 and NE interactions with arginine. Approach to the NH1 and NH2 to form double N-H-O=C interactions was found to be rare. For the approach of arginines towards a reference carboxyl, the possibility of twin
N-H--O=C interactions meant that there was an obvious preference for arginine positions in front of the carbonyl oxygens. Although overall, single N-H--O=C interactions were the most common, since such interactions have more possibility to occur. Twin nitrogen single oxygen interactions were rare. No examples of enzyme inhibitor/substrate complexes favouring NH1/NE interactions over NH1/NH2 were available to the authors.

**Histidine - Adenine**
One complex was found, in which adenine accepts the histidine NE2-H proton.

**Histidine - Guanine**
Only one complex was observed, involving donation of the histidine NE2-H proton to guanine C6=O6.

**Histidine - Theophylline**
One complex which interacts with the imidazole sidechain was found.

Histidine is chemically ambiguous regarding its hydrogen bonding interactions. The near physiological pka of the imidazole side-chain means that the protonation state of histidine is often uncertain in crystallographic studies. The nitrogen atoms of histidine are ambiguous as donors or acceptors. Histidine can only form one hydrogen bond. The survey shows that hydrogen bonds are more tightly clustered around ND1 than NE2. These authors have assumed a neutral imidazole ring, with protonation of NE2 (as in this work). Under these conditions, the clustering at ND1 represents hydrogen bond donor atoms.
6.4 Analysis of Hydrogen Bonding Motifs

The hydrogen bond contacts observed in each of the minimum energy structures can be sub-divided into six types:

- a) N-H--O=C single contacts
- b) N-H--N single contacts
- c) N-H--O=C double contacts
- d) N-H--N double contacts
- e) N-H--O=C, N-HN double contacts
- f) Bifurcated hydrogen bonds

The single N-H--O=C contact occurs the most frequently (71 examples), followed by N-H--N single contacts (14), N-H--O=C, N-H--N double contacts (12), N-H--O=C double contacts (10), and N-H--N double contacts and bifurcated hydrogen bonds (4 examples of each). The ability of electrostatic extrema to predict the formation of these types of hydrogen bonding contacts can be examined. Each of the minimum energy hydrogen bonded structures has been classified according to the type of hydrogen bond contact (from a - e, above) which stabilizes the complex. This information is summarised in figure 9.

![Figure 9](image)

**Figure 9.** Root mean square (RMS) values of the separations between electrostatic minima and hydrogen bond protons (RMS minima), and electrostatic maxima and hydrogen bond acceptors (RMS maxima) for the hydrogen bonded complexes listed in tables 1-5. The RMS values have been calculated by assigning the hydrogen bonds to one of six motifs; a) N-H--O=C single hydrogen bonds, b) N-H--N single hydrogen bonds, c) N-H--O=C double hydrogen bonds, d) N-H--N double hydrogen bonds, e) N-H--O=C and N-H--N double hydrogen bonds, f) Bifurcated hydrogen bonds.
For each type of contact, RMS values are shown for the separation between electrostatic maxima and hydrogen bond acceptor atoms, and electrostatic minima and hydrogen bond protons (as defined in figure 7). An example of each type of hydrogen bonding motif, and the appropriate electrostatic maxima and minima, is given in figure 10.

From figure 9, it can be seen that the positions of hydrogen bond acceptor atoms are generally more closely predicted by electrostatic maxima, when compared to the wider range of values observed for the separations between hydrogen bond protons and electrostatic minima. This is consistent with the observation that the position of an electrostatic minimum is much more variable, particularly in the case of oxygen minima, and subject to the effects of neighbouring molecular functionalities (discussed below). The RMS values for the separations between maxima (minima) and hydrogen bond acceptors (protons) are in much better agreement for N-H-N single bonds, when compared with the predictions for hydrogen bond atom positions for N-H-O=C single bonds.

The separations between electrostatic maxima (minima) and the hydrogen bond acceptor (proton) are lower for single N-H-O=C hydrogen bond contacts, than for double N-H-O=C hydrogen bond contacts. This is not surprising, since the opportunity of forming a second intermolecular contact dominates the hydrogen bond geometry of a complex,^24 making the predictions using electrostatic extrema necessarily poorer.

An exception the general pattern of better predictions of acceptor atom positions using electrostatic maxima is seen for N-H-N double contacts. In this case the trend is reversed. It is possible that this is simply a consequence of the small number of examples of this contact. There are 4 examples, all of which are guanine-adenine complexes. N-H-O=C double contacts, or N-H-O=C, N-H-N double contacts are energetically preferred where possible, and more probable for these molecules, based on the number of available N and O acceptors.

There are 4 examples of bifurcated hydrogen bonds, in which an acceptor C=O group is shared between two N-H donors. For these parallel, coplanar N-H donors, the electrostatic maximum is located between the donors, and consequently gives an excellent prediction of the oxygen acceptor atom position. Conversely, there is only one electrostatic minimum related to the oxygen acceptor. Since this minimum must be ‘shared’ between the two protons, it does not lie particularly close to either of the hydrogen donor atoms, and the predictions are poorer.
Figure 10. Examples of complexes containing the six hydrogen bond motifs illustrated for a) formamide-methylacetamide (N-H-O=C single hydrogen bond), b) guanine-adenine (N-H-N single hydrogen bond), c) asparagine-theophylline (N-H-O=C double hydrogen bonds), d) guanine-adenine (N-H-N double hydrogen bonds), e) adenine-theophylline (N-H-O=C, N-H-N double hydrogen bonds), f) guanine-theophylline (bifurcated hydrogen bond). For each complex, the relevant electrostatic maxima and minima are indicated.

6.5 Variable Position of Oxygen Electrostatic Minima

The examination of the RMS values of the separations between hydrogen bond donors and oxygen or nitrogen minima has highlighted the poorer predictive ability of oxygen electrostatic potential minima. This reflects the more consistent position of the nitrogen electrostatic minima in relation to the nitrogen atom (for N minima the range of $\alpha$ values = 21.5°; for O minima the range of $\alpha = 59.4^\circ$). Figure 11 compares the electrostatic potential surface around the carbonyl oxygen of formaldehyde (whose electrostatic properties were calculated according to the procedure described in the ‘Methods’ section) oxygen and the $sp^2$ nitrogen ND1 of histidine. The electrostatic potential was calculated along an arc 0.5 Å outside of the van der Waals surface of each atom, in the molecular plane of formaldehyde, and in the imidazole ring plane for histidine.
Figure 11. A comparison of the electrostatic potential energy as a function of $\alpha$, for the formaldehyde oxygen and histidine imidazole ND1 atoms. The electrostatic potential energy between a unit positive point charge of radius 0.5 Å has been calculated in the molecular plane of formaldehyde, and in the imidazole ring plane for histidine. For the imidazole ND1 plot, points from 60-90° (which lie within the van der Waals volume of the N-acetyl, N'-methylamide blocked histidine main chain atoms) have been excluded.

The oxygen electrostatic potential is of greater magnitude than the nitrogen potential, which is reflected in the greater magnitude of the oxygen electrostatic potential minimum. However, the potential across the region outside the van der Waals surface of the nitrogen acceptor shows greater change in the magnitude of the potential, changing rapidly across the arc as $\alpha$ varies. Therefore the nitrogen minimum is more sharply defined (in the direction of the $sp^2$ ring nitrogen lone pair), and its relative position is consequently more consistent. The oxygen electrostatic potential is fairly insensitive to $\alpha$, producing a broad featureless minimum. The actual location of the electrostatic minimum is more easily perturbed by neighbouring function groups.

The difficulties caused by this perturbation are illustrated by a complex between aspartate and guanine (see figure 12). Though a single electrostatic minimum is found for each oxygen atom, the oxygen may accept up to two hydrogen bonds from a range of C=O-H angles, so predictions involving C=O minima are poorer. In this example the hydrogen bond contact forms to the anti-lone pair orbital of aspartate OD1, whilst the electrostatic minimum is located at the syn lone pair, due to the proximity of OD2. Thus, the separation between the proton and the minimum is very large (For other examples of this effect, the separation between proton and minimum are listed in parenthesis, in tables 1 to 5).
6.6 Conclusions.

These results demonstrate that there is moderately good correlation between the positions of electrostatic extrema and the location of hydrogen bonded atoms, subject to certain considerations.

One consideration involves the nature of the hydrogen bond or bonds. Complexes stabilized by a single hydrogen bond are more accurately predicted by the positions of electrostatic extrema, than those stabilized by double hydrogen bonds. This is an expected result, since it has been shown that the formation of multiple hydrogen bond contacts dominates over the directional effects of lone electron pair orbitals, in determining the geometry of hydrogen bonded complexes. Lone pair directionality is observed where this fits within the geometrical constraints of multiple contacts. Hence, for doubly hydrogen bonded complexes we might only expect one of the hydrogen bond interactions to be predicted by electrostatic extrema.

For single hydrogen bonds in accessible molecular orientations the electrostatic interactions is the major component of the hydrogen bond. In these cases, the electrostatic potential extrema prove to be good predictors of the geometry of the

![Figure 12.](image)

Figure 12. A hydrogen bonded complex between guanine and aspartate, with the syn and anti lone pairs of OD1 indicated. The hydrogen bond, N2-H22\textsubscript{syn} - OD1\textsubscript{anti}, forms to the anti lone pair of OD1, but due to the presence of the neighbouring OD2, the OD1 electrostatic minimum is found in the direction of the syn lone pair. As a result, the separation between the hydrogen bonded proton and the electrostatic minimum is large.

Key: N2-H22\textsubscript{syn} electrostatic potential maximum, \(+\); OD1\textsubscript{anti} electrostatic potential minimum \(\ast\).

168
hydrogen bond contacts. However, there is a clear distinction between the proximity of electrostatic maxima to acceptor atoms, which is usually less than 0.6 Å (figure 9), and the proximity of electrostatic minima to hydrogen bonded protons, which is quite variable.

The electrostatic minima for the carbonyl oxygen acceptor groups among the molecules studied here vary in their exact position along an arc in the molecular plane between the lone pair orbitals. However, the range of positions do not extended along the entire length of this arc, but are clustered in a tighter range than that defined by the lone pair orbitals. The angle between sp² oxygen lone pair is 120°, but the spread of oxygen minima only covers a 59.4° arc (see figure 6).

N-H maxima are frequently found along the N-H bond axis, but for some of the N-H groups the electrostatic maximum is not found along this axis, but is displaced by up to 20° from the bond axis. This is seen for the N7-H group of theophylline. For guanine, the proximity of the O6 and N7 leads to a reinforcement of negative electrostatic potential between the groups, for theophylline the closeness of O6 and N7-H shifts the electrostatic maximum away from the N7-H bond axis. For sp² N-H donors, a geometric description of a hydrogen bond would predict the angle N-˙H-acceptor to be linear. Moving to a more general view of the N-H maxima, the distribution of maxima does predict a slightly less than linear N-˙H-acceptor angle, as is seen for small molecules. Most electrostatic maxima are located close to linearity, hence the method favours the prediction of near linear hydrogen bonds.

The angle at the acceptor C=˙O-H is more variable, but steric considerations usually mean it is only considered from ±90° relative to the C=O bond axis in the molecular plane. A hydrogen bond aligned with an sp² lone pair orbital would give an angle at the acceptor of α= ±60°. Taylor et al found such lone pair directionality to be less pronounced for C=O groups which only receive one hydrogen bond. For small molecules, studies of lone pair directionality have found geometrical considerations involved in forming the greatest number of electrostatic interactions between molecules to be more important than lone pair directionality. For almost all carbonyl oxygens, a single electrostatic minimum was located. However, each C=O is capable of accepting two hydrogen bonds through its lone pair orbitals. While in practice most C=O groups in protein structures do not express their full hydrogen bonding potential as a result of the protein structural constraints, the hydrogen bond may still form from any direction.
between the lone pair orbitals (allowing for geometric constraints between the interacting molecules). The hydrogen bond energy is fairly insensitive to the angle at the carbonyl oxygen. At 0.5 Å outside the oxygen van der Waals radius (the separation between the O and a hydrogen bonded proton), the proton could be found anywhere along a chord of 3.29 Å between the $sp^2$ lone pair orbitals. Hence, for single N-H--O=C contacts, the RMS value obtained for the separation between minima and proton of 1.30 Å represents 33% of this arc.

The proton is usually less than 0.6 Å from a potential maximum, and the acceptor atom within 1.30 Å of the potential maximum, though these deviations are larger when there are multiple hydrogen bonds. Large separations would only indicate that areas of electrostatic potential of the same sign are overlapping. This observation could be useful in designing new ligands for protein binding sites of unknown structure. From the observation that the location of oxygen electrostatic minima are variable, it may even be necessary to apply stricter criteria to the overlap of electrostatic maxima, and allow a wider range of acceptable values for the coincidence of electrostatic minima. This idea could be extended so that, where possible if it would be preferable to use electrostatic maxima to define and assess relative binding orientations, rather than electrostatic minima, which are more susceptible to the effects of the local chemical environment. However, this will be affected by the uncertainty in the position of protons in x-ray crystallographic structures. These results suggest that, for single N-H--O=C hydrogen bond interactions, electrostatic extrema in the potential around two molecules in their relative binding orientation should be within about 1.5 Å of each other if both are to form a corresponding hydrogen bond to the protein.
Table 1. Hydrogen bonded complexes with Formamide and Methylacetamide

<table>
<thead>
<tr>
<th>Complex</th>
<th>Contact</th>
<th>Separation /Å</th>
<th>Electrostatic Energy /kJ mol⁻¹</th>
<th>DONOR Nearest Donor Atom</th>
<th>Minima Energy /kJ mol⁻¹</th>
<th>ACCEPTOR Nearest Acceptor Atom</th>
<th>Maxima Energy /kJ mol⁻¹</th>
</tr>
</thead>
<tbody>
<tr>
<td>Formamide-Formamide</td>
<td>N1-H₁⁻O₂⁻</td>
<td>2.99</td>
<td>-58.616</td>
<td>N1-H₁⁻ (2.15)</td>
<td>-184.2</td>
<td>O₂⁻</td>
<td>1.15</td>
</tr>
<tr>
<td></td>
<td>N1-H₁⁻O₂⁻</td>
<td>2.99</td>
<td>-34.078</td>
<td>N1-H₁⁻ (2.15)</td>
<td>-184.2</td>
<td>O₂⁻</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td>N1-H₁⁻O₂⁻</td>
<td>2.99</td>
<td>-31.182</td>
<td>N2-H₂⁻</td>
<td>-187.3</td>
<td>O₃⁻</td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td>N2-H₂⁻O₃⁻</td>
<td>3.00</td>
<td>-31.170</td>
<td>N2-H₂⁻</td>
<td>-187.3</td>
<td>O₃⁻</td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td>N2-H₂⁻O₃⁻</td>
<td>3.00</td>
<td>-31.048</td>
<td>N2-H₂⁻</td>
<td>-187.3</td>
<td>O₃⁻</td>
<td>0.19</td>
</tr>
<tr>
<td></td>
<td>N2-H₂⁻O₃⁻</td>
<td>3.00</td>
<td>-30.336</td>
<td>N2-H₂⁻</td>
<td>-187.3</td>
<td>O₃⁻</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>N1-H₁⁻O₃⁻</td>
<td>2.99</td>
<td>-36.298</td>
<td>N1-H₁⁻</td>
<td>-187.3</td>
<td>O₃⁻</td>
<td>0.66</td>
</tr>
<tr>
<td></td>
<td>N1-H₂⁻O₃⁻</td>
<td>2.99</td>
<td>-34.605</td>
<td>N1-H₂⁻</td>
<td>-187.3</td>
<td>O₃⁻</td>
<td>0.29</td>
</tr>
<tr>
<td></td>
<td>N1-H₂⁻O₃⁻</td>
<td>2.99</td>
<td>-34.232</td>
<td>N1-H₂⁻</td>
<td>-187.3</td>
<td>O₃⁻</td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td>N2-H₂⁻O₂⁻</td>
<td>3.00</td>
<td>-31.029</td>
<td>N2-H₂⁻</td>
<td>-184.2</td>
<td>O₂⁻</td>
<td>0.10</td>
</tr>
</tbody>
</table>

For each complex the hydrogen bond interaction(s) is identified. For each hydrogen bond donor (acceptor), the distance to the nearest electrostatic minimum (maximum) is given. Separations in parenthesis indicate complexes in which the hydrogen bond and electrostatic minimum are located at opposite lone pair directions, producing large separations.
### Table 2. Hydrogen bonded complexes with Adenine, Guanine and Theophylline.

<table>
<thead>
<tr>
<th>Complex</th>
<th>Contact</th>
<th>Separation</th>
<th>Hydrogen Bonded Energy</th>
<th>Electrostatic Energy</th>
<th>DONOR</th>
<th>Nearest Donor Atom</th>
<th>Distance from H</th>
<th>Minima</th>
<th>Energy</th>
<th>DONOR</th>
<th>Nearest Donor Atom</th>
<th>Distance from H</th>
<th>Minima</th>
<th>Energy</th>
<th>ACCEPTOR</th>
<th>Nearest Donor Atom</th>
<th>Distance from H</th>
<th>Maxima</th>
<th>Energy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adenine-Theophylline</td>
<td>N9-H\textsubscript{6}</td>
<td>2.99</td>
<td>-72.787</td>
<td>1.88</td>
<td>N9-H\textsubscript{6}</td>
<td>1.00</td>
<td>98.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N7-H\textsubscript{6}</td>
<td>3.09</td>
<td>-63.896</td>
<td>(1.91)</td>
<td>N7-H\textsubscript{6}</td>
<td>0.51</td>
<td>54.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N6-H\textsubscript{6}</td>
<td>3.00</td>
<td>-54.760</td>
<td>1.76</td>
<td>N6-H\textsubscript{6}</td>
<td>0.23</td>
<td>45.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N7-H\textsubscript{1}</td>
<td>3.09</td>
<td>-57.387</td>
<td>0.86</td>
<td>N7-H\textsubscript{1}</td>
<td>0.56</td>
<td>95.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N9-H\textsubscript{1}</td>
<td>3.08</td>
<td>-55.359</td>
<td>0.31</td>
<td>N9-H\textsubscript{1}</td>
<td>0.84</td>
<td>97.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adenine-Guanine</td>
<td>N9-H\textsubscript{6}</td>
<td>2.98</td>
<td>-78.474</td>
<td>3.52</td>
<td>N9-H\textsubscript{6}</td>
<td>1.20</td>
<td>98.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N7-H\textsubscript{6}</td>
<td>3.09</td>
<td>-61.757</td>
<td>0.74</td>
<td>N7-H\textsubscript{6}</td>
<td>0.65</td>
<td>97.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N6-H\textsubscript{6}</td>
<td>3.09</td>
<td>-60.019</td>
<td>0.12</td>
<td>N6-H\textsubscript{6}</td>
<td>1.20</td>
<td>98.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N2-H\textsubscript{6}</td>
<td>3.09</td>
<td>-57.387</td>
<td>0.86</td>
<td>N2-H\textsubscript{6}</td>
<td>0.56</td>
<td>95.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N9-H\textsubscript{1}</td>
<td>3.08</td>
<td>-55.359</td>
<td>0.31</td>
<td>N9-H\textsubscript{1}</td>
<td>0.84</td>
<td>97.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N6-H\textsubscript{1}</td>
<td>3.10</td>
<td>-62.119</td>
<td>0.35</td>
<td>N6-H\textsubscript{1}</td>
<td>0.55</td>
<td>54.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N9-H\textsubscript{7}</td>
<td>3.08</td>
<td>-52.310</td>
<td>1.04</td>
<td>N9-H\textsubscript{7}</td>
<td>1.50</td>
<td>98.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N9-H\textsubscript{7}</td>
<td>3.09</td>
<td>-52.239</td>
<td>1.66</td>
<td>N9-H\textsubscript{7}</td>
<td>1.23</td>
<td>98.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N9-H\textsubscript{6}</td>
<td>2.99</td>
<td>-47.425</td>
<td>1.20</td>
<td>N9-H\textsubscript{6}</td>
<td>0.09</td>
<td>97.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N6-H\textsubscript{6}</td>
<td>3.03</td>
<td>-36.426</td>
<td>1.02</td>
<td>N6-H\textsubscript{6}</td>
<td>0.95</td>
<td>45.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N9-H\textsubscript{6}</td>
<td>3.09</td>
<td>-36.355</td>
<td>1.72</td>
<td>N9-H\textsubscript{6}</td>
<td>0.19</td>
<td>54.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Guanine-Theophylline</td>
<td>N1-H\textsubscript{6}</td>
<td>2.99</td>
<td>-87.959</td>
<td>3.42</td>
<td>N1-H\textsubscript{6}</td>
<td>0.96</td>
<td>95.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N7-H\textsubscript{6}</td>
<td>2.99</td>
<td>-87.959</td>
<td>3.42</td>
<td>N7-H\textsubscript{6}</td>
<td>0.96</td>
<td>95.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N7-H\textsubscript{6}</td>
<td>3.08</td>
<td>-110.12</td>
<td>2.57</td>
<td>N7-H\textsubscript{6}</td>
<td>0.96</td>
<td>95.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N7-H\textsubscript{6}</td>
<td>2.98</td>
<td>-110.12</td>
<td>2.57</td>
<td>N7-H\textsubscript{6}</td>
<td>0.96</td>
<td>95.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

For each complex the hydrogen bond interaction(s) is identified. For each hydrogen bond donor (acceptor), the distance to the nearest electrostatic minimum (maximum) is given. Separations in parenthesis indicate complexes in which the hydrogen bond and electrostatic minimum are located at opposite lone pair directions, producing large separations.
Table 3. Hydrogen bonded complexes with ADMa.

<table>
<thead>
<tr>
<th>Complex</th>
<th>Contact</th>
<th>Separation /Å</th>
<th>Electrostatic Energy /kJ mol⁻¹</th>
<th>DONOR Nearest Donor Atom</th>
<th>Minima Distance from H-atom /Å</th>
<th>Minima Energy /kJ mol⁻¹</th>
<th>ACCEPTOR Nearest Acceptor Atom</th>
<th>Maxima Distance from Acceptor /Å</th>
<th>Maxima Energy /kJ mol⁻¹</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADMa-Formamide</td>
<td>N1-H₂-O₂</td>
<td>3.00</td>
<td>-42.557</td>
<td>N1-H₂</td>
<td>1.36</td>
<td>-230.0</td>
<td>O₂</td>
<td>0.88</td>
<td>100.6</td>
</tr>
<tr>
<td></td>
<td>N1-H₂-O₂</td>
<td>2.99</td>
<td>-41.882</td>
<td>N1-H₂</td>
<td>0.60</td>
<td>-184.2</td>
<td>O₂</td>
<td>0.43</td>
<td>156.6</td>
</tr>
<tr>
<td></td>
<td>N1-H₂-O₁</td>
<td>3.00</td>
<td>-41.187</td>
<td>N1-H₂</td>
<td>(2.12)</td>
<td>-230.0</td>
<td>O₁</td>
<td>0.36</td>
<td>100.6</td>
</tr>
<tr>
<td></td>
<td>N3-H₁-O₂</td>
<td>2.99</td>
<td>29.905</td>
<td>N3-H₂</td>
<td>1.53</td>
<td>-184.2</td>
<td>O₂</td>
<td>2.94</td>
<td>147.5</td>
</tr>
<tr>
<td>ADMa-Methyacetamide</td>
<td>N2-H₂-O₃</td>
<td>2.99</td>
<td>-42.828</td>
<td>N2-H₂</td>
<td>1.67</td>
<td>-187.3</td>
<td>O₃</td>
<td>0.43</td>
<td>147.5</td>
</tr>
<tr>
<td></td>
<td>N1-H₂-O₃</td>
<td>2.99</td>
<td>42.458</td>
<td>N1-H₂</td>
<td>0.85</td>
<td>-187.3</td>
<td>O₃</td>
<td>0.45</td>
<td>156.6</td>
</tr>
<tr>
<td></td>
<td>N₂-H₂-O₂</td>
<td>2.99</td>
<td>-42.296</td>
<td>N₂-H₂</td>
<td>1.60</td>
<td>-187.3</td>
<td>O₂</td>
<td>0.24</td>
<td>147.5</td>
</tr>
<tr>
<td></td>
<td>N1-H₂-O₃</td>
<td>2.99</td>
<td>42.226</td>
<td>N₁-H₂</td>
<td>1.30</td>
<td>-187.3</td>
<td>O₃</td>
<td>0.44</td>
<td>156.6</td>
</tr>
<tr>
<td></td>
<td>N₂-H₂-O₂</td>
<td>2.99</td>
<td>37.317</td>
<td>N₂-H₂</td>
<td>2.05</td>
<td>128.1</td>
<td>O₂</td>
<td>0.07</td>
<td>94.5</td>
</tr>
<tr>
<td>ADMa-Adenine</td>
<td>N9-H₂-O₁</td>
<td>2.99</td>
<td>-52.188</td>
<td>N9-H₂</td>
<td>(2.59)</td>
<td>-230.0</td>
<td>O₁</td>
<td>0.89</td>
<td>98.1</td>
</tr>
<tr>
<td></td>
<td>N9-H₂-O₁</td>
<td>2.99</td>
<td>47.981</td>
<td>N9-H₂</td>
<td>1.31</td>
<td>-230.0</td>
<td>O₁</td>
<td>0.43</td>
<td>98.1</td>
</tr>
<tr>
<td></td>
<td>N9-H₂-O₉</td>
<td>2.99</td>
<td>47.408</td>
<td>N9-H₂</td>
<td>(2.60)</td>
<td>206.4</td>
<td>O₉</td>
<td>1.02</td>
<td>98.1</td>
</tr>
<tr>
<td></td>
<td>N9-H₂-O₂</td>
<td>2.99</td>
<td>-44.934</td>
<td>N9-H₂</td>
<td>2.99</td>
<td>-230.0</td>
<td>O₂</td>
<td>0.71</td>
<td>98.1</td>
</tr>
<tr>
<td></td>
<td>N₉-H₂-O₉</td>
<td>3.00</td>
<td>44.490</td>
<td>N₉-H₂</td>
<td>0.71</td>
<td>-187.3</td>
<td>O₉</td>
<td>0.71</td>
<td>98.1</td>
</tr>
<tr>
<td></td>
<td>N₁-H₁-O₆</td>
<td>3.09</td>
<td>37.551</td>
<td>N₁-H₁</td>
<td>0.67</td>
<td>206.4</td>
<td>O₆</td>
<td>1.10</td>
<td>156.6</td>
</tr>
<tr>
<td></td>
<td>N₁-H₁-N₃</td>
<td>3.09</td>
<td>-36.004</td>
<td>N₁-H₁</td>
<td>0.92</td>
<td>-138.6</td>
<td>N₃</td>
<td>0.07</td>
<td>94.5</td>
</tr>
<tr>
<td>ADMa-Guanine</td>
<td>N₂-H₁-O₆</td>
<td>3.00</td>
<td>-73.498</td>
<td>N₂-H₁</td>
<td>1.68</td>
<td>-234.6</td>
<td>O₆</td>
<td>0.55</td>
<td>147.5</td>
</tr>
<tr>
<td></td>
<td>N₁-H₁-N₇</td>
<td>3.09</td>
<td>-72.818</td>
<td>N₁-H₁</td>
<td>0.79</td>
<td>-237.4</td>
<td>N₇</td>
<td>0.32</td>
<td>156.6</td>
</tr>
<tr>
<td></td>
<td>N₁-H₁-O₆</td>
<td>2.98</td>
<td>-68.742</td>
<td>N₁-H₁</td>
<td>(2.16)</td>
<td>-230.0</td>
<td>O₆</td>
<td>0.71</td>
<td>151.2</td>
</tr>
<tr>
<td></td>
<td>N₂-H₁-O₁</td>
<td>3.01</td>
<td>68.742</td>
<td>N₂-H₁</td>
<td>1.24</td>
<td>-266.4</td>
<td>O₁</td>
<td>1.90</td>
<td>151.2</td>
</tr>
<tr>
<td></td>
<td>N₂-H₁-O₉</td>
<td>3.00</td>
<td>-72.818</td>
<td>N₂-H₁</td>
<td>1.01</td>
<td>-234.6</td>
<td>O₉</td>
<td>0.30</td>
<td>147.5</td>
</tr>
<tr>
<td></td>
<td>N₁-H₁-N₉</td>
<td>3.10</td>
<td>-48.946</td>
<td>N₁-H₁</td>
<td>0.99</td>
<td>-237.4</td>
<td>N₉</td>
<td>0.63</td>
<td>156.6</td>
</tr>
<tr>
<td></td>
<td>N₉-H₁-O₉</td>
<td>3.00</td>
<td>-41.619</td>
<td>N₉-H₁</td>
<td>1.29</td>
<td>-230.0</td>
<td>O₉</td>
<td>0.20</td>
<td>97.2</td>
</tr>
<tr>
<td>ADMa-Theophylline</td>
<td>N₇-H₉-O₁</td>
<td>2.99</td>
<td>-59.125</td>
<td>N₇-H₉</td>
<td>(2.27)</td>
<td>-230.0</td>
<td>O₁</td>
<td>0.82</td>
<td>110.1</td>
</tr>
<tr>
<td></td>
<td>N₇-H₉-O₁</td>
<td>3.00</td>
<td>51.742</td>
<td>N₇-H₉</td>
<td>0.92</td>
<td>-266.4</td>
<td>O₁</td>
<td>1.67</td>
<td>110.1</td>
</tr>
<tr>
<td></td>
<td>N₇-H₉-O₁</td>
<td>2.99</td>
<td>-47.179</td>
<td>N₇-H₉</td>
<td>(2.40)</td>
<td>-266.4</td>
<td>O₁</td>
<td>0.31</td>
<td>110.1</td>
</tr>
</tbody>
</table>

For each complex the hydrogen bond interaction(s) is identified. For each hydrogen bond donor (acceptor), the distance to the nearest electrostatic minima (maxima) is given.

a Separations in parenthesis indicate complexes in which the hydrogen bond and electrostatic minima are located at opposite lone pair directions, producing large separations.

b This electrostatic minimum is shared between ADMa O₀ and O₁.

c There is no N₃-H₂ maximum. The distance given in the table is to the nearest maximum, N₂-H₂.

d This electrostatic minimum is shared between ADMa O₁ and O₂.
<table>
<thead>
<tr>
<th>Complex</th>
<th>Contact</th>
<th>Separation /Å</th>
<th>Electrostatic Energy /kJ mol⁻¹</th>
<th>Nearest Donor Atom</th>
<th>Minima Distance from H Energy /kJ mol⁻¹</th>
<th>Nearest Acceptor Atom</th>
<th>Maxima Distance from</th>
<th>Energy /kJ mol⁻¹</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ADMβ-Formamide</strong></td>
<td>N1-H1p-O2p</td>
<td>3.00</td>
<td>-44.820</td>
<td>N1-H1p</td>
<td>0.46</td>
<td>-128.1</td>
<td>O2</td>
<td>1.37</td>
</tr>
<tr>
<td></td>
<td>N2-H2-O2p</td>
<td>3.14</td>
<td>-41.190</td>
<td>N3-H3</td>
<td>1.16</td>
<td>-184.2</td>
<td>O2</td>
<td>0.23</td>
</tr>
<tr>
<td></td>
<td>N1-H1p-O1p</td>
<td>3.00</td>
<td>-41.041</td>
<td>N1-H1p</td>
<td>0.45</td>
<td>-186.9</td>
<td>O1p</td>
<td>1.51</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>ADMβ-Methylacetamide</strong></td>
<td>N3-H3-O3p</td>
<td>2.99</td>
<td>-42.210</td>
<td>N3-H3</td>
<td>0.38</td>
<td>-187.3</td>
<td>O3</td>
<td>0.19</td>
</tr>
<tr>
<td></td>
<td>N2-H2-O3p</td>
<td>2.99</td>
<td>-31.137</td>
<td>N2-H2</td>
<td>0.81</td>
<td>-127.8</td>
<td>O2</td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td>N2-H2-O3p</td>
<td>2.99</td>
<td>-24.314</td>
<td>N2-H2</td>
<td>1.47</td>
<td>-186.9</td>
<td>O1</td>
<td>0.56</td>
</tr>
<tr>
<td><strong>ADMβ-Adenine</strong></td>
<td>N9-H9-O0p</td>
<td>2.99</td>
<td>-64.183</td>
<td>N9-H9</td>
<td>1.01</td>
<td>-127.8</td>
<td>O6</td>
<td>0.51</td>
</tr>
<tr>
<td></td>
<td>N9-H9-O2p</td>
<td>2.99</td>
<td>-39.743</td>
<td>N9-H9</td>
<td>0.70</td>
<td>-127.8</td>
<td>O2</td>
<td>0.82</td>
</tr>
<tr>
<td></td>
<td>N6-H6-O2p</td>
<td>2.99</td>
<td>-30.426</td>
<td>N6-H6</td>
<td>0.54</td>
<td>-184.2</td>
<td>O2</td>
<td>0.13</td>
</tr>
<tr>
<td><strong>ADMβ-Guanine</strong></td>
<td>N9-H9-O0p</td>
<td>2.99</td>
<td>-63.851</td>
<td>N9-H9</td>
<td>0.60</td>
<td>-186.9</td>
<td>O6</td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td>N9-H9-O2p</td>
<td>3.00</td>
<td>-64.183</td>
<td>N9-H9</td>
<td>1.18</td>
<td>-127.8</td>
<td>O2</td>
<td>0.14</td>
</tr>
<tr>
<td></td>
<td>N1-H1o-O0p</td>
<td>3.01</td>
<td>-63.773</td>
<td>N1-H1o</td>
<td>2.07</td>
<td>-128.1</td>
<td>O2</td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td>N2-H2-O0p</td>
<td>3.06</td>
<td>-45.523</td>
<td>N2-H2</td>
<td>1.59</td>
<td>-128.1</td>
<td>O2</td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>ADMβ-Theophylline</strong></td>
<td>N7-H7-O0p</td>
<td>2.99</td>
<td>-52.835</td>
<td>N7-H7</td>
<td>1.39</td>
<td>-127.8</td>
<td>O6</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td>N7-H7-O2p</td>
<td>2.99</td>
<td>-45.404</td>
<td>N7-H7</td>
<td>0.13</td>
<td>-186.9</td>
<td>O4</td>
<td>0.12</td>
</tr>
<tr>
<td></td>
<td>N7-H7-O2p</td>
<td>2.99</td>
<td>-39.743</td>
<td>N7-H7</td>
<td>0.20</td>
<td>-128.1</td>
<td>O2</td>
<td>0.60</td>
</tr>
<tr>
<td></td>
<td>N3-H3-O2p</td>
<td>3.00</td>
<td>-71.2</td>
<td>N3-H3</td>
<td>0.57</td>
<td>-137.7</td>
<td>O6</td>
<td>0.07</td>
</tr>
</tbody>
</table>

For each complex the hydrogen bond interaction(s) is identified. For each hydrogen bond donor (acceptor), the distance to the nearest electrostatic minimum (maximum) is given, a / indicates a hydrogen bond with no relevant electrostatic extremum.
Table 5. Hydrogen bonded complexes with amino acid sidechains Arginine, Aspartate, Asparagine and Histidine, and Adenine, Guanine and Theophylline.

<table>
<thead>
<tr>
<th>Complex</th>
<th>Contact</th>
<th>Separation /Å</th>
<th>Electrostatic Energy /kJ mol⁻¹</th>
<th>DONOR Nearest Donor Atom</th>
<th>DONOR Distance from H Energy /kJ mol⁻¹</th>
<th>NEAREST ACCEPTOR</th>
<th>Maxima</th>
<th>ACCEPTOR Nearest Acceptor Energy /kJ mol⁻¹</th>
<th>ACCEPTOR Distance from Acceptor Atom /Å</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arginine-Adenine</td>
<td>NH2-H1a—N1a</td>
<td>3.09</td>
<td>-53.784</td>
<td>NH2-H1a</td>
<td>0.26 -131.4</td>
<td>N1a</td>
<td>0.24</td>
<td>424.8</td>
<td></td>
</tr>
<tr>
<td>Arginine-Guanine</td>
<td>NH1-H2a—O6o—a</td>
<td>2.99</td>
<td>-137.74</td>
<td>NH1-H2a</td>
<td>1.07 -234.6</td>
<td>O6o</td>
<td>1.44</td>
<td>459.2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>NH2-H2a—N7a</td>
<td>3.09</td>
<td>-137.69</td>
<td>NH2-H2a</td>
<td>0.97 -234.6</td>
<td>N7a</td>
<td>1.70</td>
<td>459.2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>NH2-H2a—O6o—a</td>
<td>2.99</td>
<td>-137.69</td>
<td>NH2-H2a</td>
<td>1.08 -234.6</td>
<td>O6o</td>
<td>1.43</td>
<td>459.2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>NH1-H2a—N7a</td>
<td>3.09</td>
<td>-112.536</td>
<td>NH1-H2a</td>
<td>0.95 -234.6</td>
<td>N7a</td>
<td>1.71</td>
<td>459.2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>NH2-H2a—O6o—a</td>
<td>2.99</td>
<td>-112.536</td>
<td>NH2-H2a</td>
<td>1.37 -234.6</td>
<td>O6o</td>
<td>0.25</td>
<td>424.8</td>
<td></td>
</tr>
<tr>
<td></td>
<td>NH1-H1a—O6a</td>
<td>2.99</td>
<td>-108.173</td>
<td>NH1-H1a</td>
<td>1.63 -234.6</td>
<td>O6a</td>
<td>0.28</td>
<td>403.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>NH1-H1a—O6a</td>
<td>2.99</td>
<td>-104.875</td>
<td>NH1-H1a</td>
<td>0.90 -234.6</td>
<td>O6a</td>
<td>0.24</td>
<td>403.5</td>
<td></td>
</tr>
<tr>
<td>Arginine-Theophylline</td>
<td>NH1-H2a—O2o—a</td>
<td>3.05</td>
<td>-73.765</td>
<td>NH1-H2a</td>
<td>1.18 -151.0</td>
<td>O2o</td>
<td>0.07</td>
<td>459.2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>NH2-H2a—O2o—a</td>
<td>3.05</td>
<td>-73.765</td>
<td>NH2-H2a</td>
<td>1.20 -151.0</td>
<td>O2o</td>
<td>0.07</td>
<td>459.2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>NH2-H1a—O2o—a</td>
<td>2.99</td>
<td>-63.700</td>
<td>NH2-H1a</td>
<td>0.68 -151.0</td>
<td>O2o</td>
<td>0.18</td>
<td>424.8</td>
<td></td>
</tr>
<tr>
<td></td>
<td>NH1-H1a—O6o—a</td>
<td>2.90</td>
<td>-57.549</td>
<td>NH1-H1a</td>
<td>0.59 -151.0</td>
<td>O6o</td>
<td>0.27</td>
<td>403.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>NH2-H1a—O6o—a</td>
<td>2.99</td>
<td>-54.649</td>
<td>NH2-H1a</td>
<td>0.57 -137.7</td>
<td>O6o</td>
<td>0.16</td>
<td>424.8</td>
<td></td>
</tr>
<tr>
<td>Asparagine-Adenine</td>
<td>N9-Ho—OD1a—a</td>
<td>2.99</td>
<td>-67.918</td>
<td>N9-Ho</td>
<td>(2.12) -170.8</td>
<td>OD1a</td>
<td>1.25</td>
<td>98.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ND2-H2a—N3a</td>
<td>3.09</td>
<td>-138.8</td>
<td>ND2-H2a</td>
<td>0.67 -138.6</td>
<td>N3a</td>
<td>1.26</td>
<td>90.6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ND2-H2a—N3a</td>
<td>3.09</td>
<td>-170.8</td>
<td>ND2-H2a</td>
<td>0.81 -131.4</td>
<td>N3a</td>
<td>1.38</td>
<td>90.6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ND2-H2a—N3a</td>
<td>3.09</td>
<td>-51.636</td>
<td>N6-H1a</td>
<td>(2.10) -170.8</td>
<td>OD1a</td>
<td>0.53</td>
<td>54.2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ND2-H2a—N7a</td>
<td>3.09</td>
<td>-120.2</td>
<td>ND2-H2a</td>
<td>0.07 -120.2</td>
<td>N7a</td>
<td>1.35</td>
<td>90.6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ND2-H1a—N3a</td>
<td>3.10</td>
<td>-30.917</td>
<td>ND2-H1a</td>
<td>0.51 -138.6</td>
<td>N3a</td>
<td>0.44</td>
<td>144.9</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ND2-H2a—N3a</td>
<td>3.10</td>
<td>-28.974</td>
<td>ND2-H2a</td>
<td>0.21 -138.6</td>
<td>N3a</td>
<td>1.27</td>
<td>90.6</td>
<td></td>
</tr>
<tr>
<td>Asparagine-Guanine</td>
<td>N1-H1o—OD1a—a</td>
<td>2.99</td>
<td>-82.307</td>
<td>N1-H1o</td>
<td>(2.27) -170.8</td>
<td>OD1a</td>
<td>1.24</td>
<td>151.2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ND2-H2a—O6o—a</td>
<td>2.99</td>
<td>-234.6</td>
<td>ND2-H2a</td>
<td>(3.54) -234.6</td>
<td>O6o</td>
<td>1.42</td>
<td>90.6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>N2-H2a—OD1a—a</td>
<td>2.99</td>
<td>-55.882</td>
<td>N2-H2a</td>
<td>(2.30) -170.8</td>
<td>OD1a</td>
<td>0.77</td>
<td>95.4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ND2-H2a—N3a</td>
<td>3.09</td>
<td>-79.2</td>
<td>ND2-H2a</td>
<td>0.19 -170.8</td>
<td>N3a</td>
<td>1.41</td>
<td>90.6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>N9-Ho—OD1a—a</td>
<td>2.99</td>
<td>-54.221</td>
<td>N9-Ho</td>
<td>(2.22) -170.8</td>
<td>OD1a</td>
<td>0.88</td>
<td>97.2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ND2-H2a—N7a</td>
<td>3.10</td>
<td>-79.2</td>
<td>ND2-H2a</td>
<td>0.37 -170.8</td>
<td>N7a</td>
<td>1.18</td>
<td>90.6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ND2-H1a—N7a</td>
<td>3.10</td>
<td>-45.346</td>
<td>ND2-H1a</td>
<td>1.02 -237.4</td>
<td>N7a</td>
<td>0.52</td>
<td>144.9</td>
<td></td>
</tr>
</tbody>
</table>
Table 5. (Continued)

<table>
<thead>
<tr>
<th>Complex</th>
<th>Contact</th>
<th>Separation /Å</th>
<th>Electrostatic Energy /kJ mol&lt;sup&gt;−1&lt;/sup&gt;</th>
<th>DONOR Nearest Donor Atom</th>
<th>Minima Distance from H /Å</th>
<th>Energy /kJ mol&lt;sup&gt;−1&lt;/sup&gt;</th>
<th>Nearest Acceptor Atom</th>
<th>Maxima Distance from Acceptor /Å</th>
<th>Energy /kJ mol&lt;sup&gt;−1&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asparagine-</td>
<td>N7-H&lt;sub&gt;asp&lt;/sub&gt;-OD1&lt;sub&gt;asp&lt;/sub&gt;</td>
<td>2.99</td>
<td>-71.963</td>
<td>N7-H&lt;sub&gt;asp&lt;/sub&gt;</td>
<td>1.87</td>
<td>-170.8</td>
<td>OD1&lt;sub&gt;asp&lt;/sub&gt;</td>
<td>1.07</td>
<td>110.1</td>
</tr>
<tr>
<td>Theophylline</td>
<td>ND2-H&lt;sub&gt;2asp&lt;/sub&gt;-OD1&lt;sub&gt;asp&lt;/sub&gt;</td>
<td>3.00</td>
<td>-37.742</td>
<td>ND2-H&lt;sub&gt;2asp&lt;/sub&gt;</td>
<td>0.70</td>
<td>-151.0</td>
<td>OD1&lt;sub&gt;asp&lt;/sub&gt;</td>
<td>0.52</td>
<td>144.9</td>
</tr>
<tr>
<td></td>
<td>ND2-H&lt;sub&gt;1asp&lt;/sub&gt;-O2&lt;sub&gt;asp&lt;/sub&gt;</td>
<td>3.00</td>
<td>-37.742</td>
<td>ND2-H&lt;sub&gt;1asp&lt;/sub&gt;</td>
<td>0.70</td>
<td>-151.0</td>
<td>OD1&lt;sub&gt;asp&lt;/sub&gt;</td>
<td>0.52</td>
<td>144.9</td>
</tr>
<tr>
<td>Aspartate-Adenine</td>
<td>N9-H&lt;sub&gt;asp&lt;/sub&gt;-OD2&lt;sub&gt;asp&lt;/sub&gt;</td>
<td>2.98</td>
<td>-81.945</td>
<td>N9-H&lt;sub&gt;asp&lt;/sub&gt;</td>
<td>0.43</td>
<td>-590.0</td>
<td>OD2&lt;sub&gt;asp&lt;/sub&gt;</td>
<td>0.27</td>
<td>98.1</td>
</tr>
<tr>
<td></td>
<td>N9-H&lt;sub&gt;asp&lt;/sub&gt;-OD1&lt;sub&gt;asp&lt;/sub&gt;</td>
<td>2.99</td>
<td>-50.345</td>
<td>N9-H&lt;sub&gt;asp&lt;/sub&gt;</td>
<td>0.43</td>
<td>-590.0</td>
<td>OD1&lt;sub&gt;asp&lt;/sub&gt;</td>
<td>0.26</td>
<td>98.1</td>
</tr>
<tr>
<td></td>
<td>N9-H&lt;sub&gt;asp&lt;/sub&gt;-OD2&lt;sub&gt;asp&lt;/sub&gt;</td>
<td>2.98</td>
<td>-78.942</td>
<td>N9-H&lt;sub&gt;asp&lt;/sub&gt;</td>
<td>(3.40)</td>
<td>-590.0</td>
<td>OD2&lt;sub&gt;asp&lt;/sub&gt;</td>
<td>0.65</td>
<td>98.1</td>
</tr>
<tr>
<td></td>
<td>N9-H&lt;sub&gt;asp&lt;/sub&gt;-OD1&lt;sub&gt;asp&lt;/sub&gt;</td>
<td>2.98</td>
<td>-68.414</td>
<td>N9-H&lt;sub&gt;asp&lt;/sub&gt;</td>
<td>(2.74)</td>
<td>-590.0</td>
<td>OD1&lt;sub&gt;asp&lt;/sub&gt;</td>
<td>0.17</td>
<td>98.1</td>
</tr>
<tr>
<td></td>
<td>N6-H&lt;sub&gt;1asp&lt;/sub&gt;-OD2&lt;sub&gt;asp&lt;/sub&gt;</td>
<td>3.00</td>
<td>-44.350</td>
<td>N6-H&lt;sub&gt;1asp&lt;/sub&gt;</td>
<td>0.53</td>
<td>-587.7</td>
<td>OD2&lt;sub&gt;asp&lt;/sub&gt;</td>
<td>0.06</td>
<td>54.2</td>
</tr>
<tr>
<td>Aspartate-Guanine</td>
<td>N2-H&lt;sub&gt;1a&lt;/sub&gt;-OD1&lt;sub&gt;a&lt;/sub&gt;</td>
<td>2.99</td>
<td>-129.468</td>
<td>N2-H&lt;sub&gt;1a&lt;/sub&gt;</td>
<td>0.10</td>
<td>-587.7</td>
<td>OD1&lt;sub&gt;a&lt;/sub&gt;</td>
<td>1.47</td>
<td>151.2</td>
</tr>
<tr>
<td></td>
<td>N1-H&lt;sub&gt;1a&lt;/sub&gt;-OD2&lt;sub&gt;a&lt;/sub&gt;</td>
<td>2.99</td>
<td>-129.468</td>
<td>N1-H&lt;sub&gt;1a&lt;/sub&gt;</td>
<td>0.94</td>
<td>-590.0</td>
<td>OD2&lt;sub&gt;a&lt;/sub&gt;</td>
<td>0.79</td>
<td>151.2</td>
</tr>
<tr>
<td></td>
<td>N2-H&lt;sub&gt;2a&lt;/sub&gt;-OD1&lt;sub&gt;a&lt;/sub&gt;</td>
<td>3.00</td>
<td>-80.302</td>
<td>N2-H&lt;sub&gt;2a&lt;/sub&gt;</td>
<td>(3.08)</td>
<td>-587.7</td>
<td>OD1&lt;sub&gt;a&lt;/sub&gt;</td>
<td>0.13</td>
<td>95.4</td>
</tr>
<tr>
<td>Aspartate-</td>
<td>N7-H&lt;sub&gt;asp&lt;/sub&gt;-OD2&lt;sub&gt;asp&lt;/sub&gt;</td>
<td>2.99</td>
<td>-92.956</td>
<td>N7-H&lt;sub&gt;asp&lt;/sub&gt;</td>
<td>0.52</td>
<td>-590.0</td>
<td>OD2&lt;sub&gt;asp&lt;/sub&gt;</td>
<td>0.51</td>
<td>110.1</td>
</tr>
<tr>
<td>Theophylline</td>
<td>N7-H&lt;sub&gt;asp&lt;/sub&gt;-OD1&lt;sub&gt;asp&lt;/sub&gt;</td>
<td>2.99</td>
<td>-89.944</td>
<td>N7-H&lt;sub&gt;asp&lt;/sub&gt;</td>
<td>0.50</td>
<td>-587.7</td>
<td>OD1&lt;sub&gt;asp&lt;/sub&gt;</td>
<td>0.51</td>
<td>110.1</td>
</tr>
<tr>
<td>Histidine-</td>
<td>NE2-H&lt;sub&gt;1ad&lt;/sub&gt;-N1&lt;sub&gt;ad&lt;/sub&gt;</td>
<td>3.09</td>
<td>-38.316</td>
<td>NE2-H&lt;sub&gt;1ad&lt;/sub&gt;</td>
<td>0.78</td>
<td>-131.4</td>
<td>N1&lt;sub&gt;ad&lt;/sub&gt;</td>
<td>0.82</td>
<td>128.4</td>
</tr>
<tr>
<td>Adenine</td>
<td>NE2-H&lt;sub&gt;1ad&lt;/sub&gt;-O6&lt;sub&gt;ad&lt;/sub&gt;</td>
<td>2.99</td>
<td>-51.257</td>
<td>NE2-H&lt;sub&gt;1ad&lt;/sub&gt;</td>
<td>0.92</td>
<td>-234.6</td>
<td>O6&lt;sub&gt;ad&lt;/sub&gt;</td>
<td>0.12</td>
<td>128.4</td>
</tr>
<tr>
<td>Histidine-</td>
<td>NE2-H&lt;sub&gt;1ad&lt;/sub&gt;-O6&lt;sub&gt;ad&lt;/sub&gt;</td>
<td>2.99</td>
<td>-33.129</td>
<td>NE2-H&lt;sub&gt;1ad&lt;/sub&gt;</td>
<td>0.38</td>
<td>-137.7</td>
<td>O6&lt;sub&gt;ad&lt;/sub&gt;</td>
<td>0.09</td>
<td>128.4</td>
</tr>
</tbody>
</table>

For each complex the hydrogen bond interaction(s) is identified. For each hydrogen bond donor (acceptor), the distance to the nearest electrostatic minimum (maximum) is given. Separations in parenthesis indicate complexes in which the hydrogen bond and electrostatic minimum are located at opposite lone pair directions, producing large separations.
References for Chapter 6.


Chapter 7. Discussion and Conclusions

In the previous chapters, several specific conclusions have been presented which will be drawn together here. In chapter two, the importance of the electrostatic component of the intermolecular potential has been emphasized. In principle the wavefunction of a molecular system contains everything there is to know about that system. The approximate representation of molecular charge distributions inevitably results in a loss of information. This is unavoidable, unless many chemically interesting systems are to remain computationally intractable. There are several degrees of approximation involved in representing charge distributions, the most common of which is the use of atom-centred point charges. Their use has helped to explain many trends in chemical systems, and provided a first approximation for many others. However, their failure to reproduce some empirical observations is due to the spherically-averaged description of charge density which atomic point-monopoles represent.

7.1 Further Applications of Similarity of Electrostatic Extrema

This method for determining relative binding orientations for ligands at a common receptor site has been successfully demonstrated for the systems it has been applied to here; Phosphodiesterase III inhibitors and Glycolate Oxidase inhibitors. In addition, this approach has been applied to two further ligand-receptor systems, Adenosine A₁ receptor ligands¹ and β-lactam ligands.²

7.1.1 Adenosine A₁ Receptor Ligand Modelling

The nucleotide adenosine plays an important role in regulating activities of the central nervous system in animals. All adenosine receptor agonists to date are based on the adenosine ligand, whilst adenosine antagonists belong to a range of chemical classes, including methyl xanthines such as theophylline from tea. Adenosine agonists and antagonists bind at the same receptor site.³ The adenosine and xanthine ring systems both contain groups capable of donating or accepting a hydrogen bond, implying that many hydrogen bonding interactions at the receptor could be in common for both agonists and antagonists.

Three models for the relative orientation of adenosine agonists and antagonists have been proposed. The ‘standard’ model is simply a direct superposition of the four
ring nitrogen atoms in the adenosine and xanthine ring systems. In the 'flipped' model, the xanthine ring is rotated to superimpose distinct steric and electrostatic similarities between the adenosine agonist and xanthine antagonist rings systems. The 'N⁶-C8' model aims to optimize the steric overlay of the N⁶ ring substituents of the agonists with antagonists' C8 ring substituents, rather than the ring systems. The coincidence of the adenosine N⁶-region and the xanthine C8-region produces a larger van der Waals volume overlap of the potent ligands N⁶-cyclopentyladenosine (CPA) and 1,3-dipropyl-8-cyclopentylxanthine (DPCPX), compared to the other two models.

Previous work by van der Wenden et al. investigated a steric and electrostatic comparison of the three models using electrostatic potential grid maps (in the purine ring plane) computed from semi-empirical point charges. While the distinction between the three models was not large, their results suggested the N⁶-C8 model as the more probable model, due to the good steric overlap between the N⁶-substituents and the xanthine C8-substituents. They found that the flipped model showed a greater percentage overlap for positive electrostatic potentials, and the N⁶-C8 model showed a slight degree of higher overlap for regions of negative electrostatic potential.

Further studies of the adenosine A₁ receptor agonists and antagonists have been carried out by van der Wenden et al. using high quality ab initio wavefunctions and distributed multipole analysis (DMA) for a more accurate representation of the molecular charge distribution. The electrostatic similarity was investigated by matching electrostatic maxima and minima determined at 1.0 Å outside the van der Waals surface of each ligand. The comparison of the three models for the relative binding orientations of the adenosine receptor ligands was carried out for two pairs of ligands: adenosine/theophylline and CPA/DPCPX.

The standard model was an extremely poor electrostatic match for both pairs of ligands, with many clashes in regions where a positive potential maximum in one molecule overlays a negative potential minimum in the other. This model can definitely be discarded. Both the flipped and N⁶-C8 models appear possible for both pairs of ligands, but their relative favourability differs for adenosine/theophylline and CPA/DPCPX. The flipped model produces some extrema clashes for both pairs of ligands around the sugar group of the agonists, which could possibly be relieved by conformational flexibility. The N⁶-C8 model for CPA/DPCPX highlights a strong clash.
between electrostatic extrema in the N6/N7 region which appears to be an important hydrogen bonding site. Rotation at the CPA N^6 may relieve some of this potential clash. This extrema clash does not occur for adenosine/theophylline, due to the different substitution and conformation at the N^6 amino group.

Thus, the choice between the flipped and N^6-C8 models depends on the importance of N7/N6 clash for DPCPX/CPA. This will depend on whether there are hydrogen bonding or other strong polar groups of the receptor in this region and whether there is flexibility around the two amine groups. Unlike the previous grid-based comparisons, this strongly unfavourable electrostatic feature of the N^6-C8 overlay for CPA/DPCPX is clearly highlighted by a comparison of their electrostatic extrema. The two pairs of molecules may not necessarily adopt the same overlay in the receptor.

7.7.2 Electrostatic Similarity of Lactam Compounds
Frau and Price^ examined a series of β-lactam, γ-lactam and non-lactam compounds in terms of their electrostatic and structural properties, and compared them with a model of the natural substrate D-alanyl-D-alanine for the carboxy and transpeptidase enzymes. Their aim was to distinguish between known inhibitors and superficially structurally similar lactam non-inhibitors. All previous assessments of anti-bacterial compounds and non-inhibitors had been performed on the basis of structural overlays and other steric criteria. However, this does not provide a sufficiently clear distinction between the inhibitors and non-inhibitors.

β-lactams act as inhibitors of the transpeptidase enzyme because they mimic the active conformation of the terminal D-alanyl-D-alanine of the normal substrate. After an extensive conformational analysis of the substrate, several different substrate conformations were overlaid onto the crystal structures representative of two of the most important classes of β-lactam antibiotics, ampicillin (a penicillin) and cephalotin (a cephalasporin). In the vast majority of β-lactam antibiotics the lactam ring nitrogen is pyramidal, but the equivalent nitrogen in the substrate is planar. This pyramidal distortion occurs to different degrees in antibiotic structures and is thought to be a substrate transition state analogue by some workers. Ampicillin and cephalotin are representative of the degree of pyramidality seen in antibiotic structures, and so were compared to a range of energetically and structurally plausible substrate conformations.
It was found that a good structural overlay of the substrate and antibiotic ligands could be obtained with either a planar or pyramidal nitrogen. Hence, given this range of acceptable substrate conformations, electrostatic similarity appears to be a key factor in distinguishing between biologically active and inactive compounds, rather than subtle differences in structural overlays.

Two different model substrates, differing in the planar or pyramidal nitrogen, were used in the subsequent electrostatic comparison. The substrate and inhibitors are negatively charged at physiological pH, therefore the relative overlays only considered electrostatic minima determined at 0.5 Å from the molecular van der Waals surface. All of the best inhibitors have four minima which can be placed within 1.0 Å of the equivalent minima of the substrate in a sterically plausible orientation. This array of minima corresponds to the expected positions of hydrogen bonding protons in the R61 D-D-peptidase enzyme binding site.9

The electrostatic overlays of compounds structurally similar to the model substrates, but which lack antibacterial activity, were worse than those obtained for the inhibitor compounds. The electrostatic potential surfaces of these non-inhibitors highlighted their dissimilarity with the inhibitors more effectively than a structural overlay. The relative positions of the electrostatic minima were different for the inhibitors and non-inhibitors. This was sometimes due to positional changes in functional groups, but there were also examples of changes in functional groups which conserved the electrostatic similarity and antibacterial activity. Active γ-lactam structures, which contain a 5-membered ring in place of the 4-membered β-lactam ring, still produced the same pattern of electrostatic minima associated with antibacterial activity.

This study has provided the first comparison of the electrostatic properties of lactam and non-lactam compounds. By matching electrostatic potential minima a clear distinction could be seen between successful antibiotics and inactive compounds which might, on the basis of a purely structural comparison, be expected to have some biological activity.
7.2 Similarity of Electrostatic Extrema

The comparison of ligands for a common binding site by locating and matching the positions of extrema in the molecular electrostatic potential was introduced in chapter three. This idea of similarity between electrostatic extrema developed from work by Davis et al\textsuperscript{10} on a series of Phosphodiesterase (PDE) III ligands. They observed that electrostatic potential minima were found near ligand hydrogen bond acceptor groups, and that these minima could be used to identify possible hydrogen bond donor groups in a putative receptor site. This rationale can be justified if it is assumed that, a) the extrema represent points of strongest electrostatic interaction between a ligand and a receptor and, b) that two ligands which are capable of occupying a receptor have many of their ligand-receptor interactions in common. The method does not require a detailed knowledge of enzyme structure, information which is typically unavailable.

Crude comparisons between patterns of electrostatic minima have been noted by other workers.\textsuperscript{11,12} A key feature of the work presented here is the use of anisotropic DMA-based molecular charge density models. It has been stated\textsuperscript{12} that the relative positions of electrostatic minima, rather than their relative energies, could be an important aspect of electrostatic similarity, and that the charge density model is relatively unimportant. The magnitude of the electrostatic potential can vary significantly, even within the range of reasonable quality basis sets.\textsuperscript{13} However the qualitative picture of the electrostatic potential is largely unaffected and the locations of the electrostatic extrema are well predicted. Whilst the relative positions of the electrostatic extrema are certainly more important than their relative energies, we have seen here that the relative positions can be sensitive to the quality of the electrostatic model. The use of DMAs in studying PDE III inhibitors in chapter four has confirmed the initial results of Davis et al\textsuperscript{10} obtained using point-charge models. In each case a sterically and electrostatically plausible relative binding orientation has been found for each inhibitor. By extending their approach to include regions of positive electrostatic potential, a greater degree of similarity than was previously noted has been demonstrated between the inhibitors. The original work only attempted to match two minima for the inhibitors; all of the inhibitors considered here have been matched at four extrema by including maxima, and minima above aromatic rings. The results presented for the PDE III inhibitors clearly demonstrate the importance of accurate
models for charge density and the role of electrostatic maxima. A comparison of electrostatic minima only for neutral ligands can be misleading, and without an accurate picture of the charge distribution many extrema are displaced or absent, obscuring the true picture of similarity.

One of the underlying assumptions of this method is that the electrostatic extrema represent the points of strongest electrostatic interaction between molecules, and as such are a probable location for receptor site atoms interacting with a ligand. Given the importance of the electrostatic component to hydrogen bond interactions, it is assumed that electrostatic extrema can be used to locate hydrogen bonded receptor atoms. The usefulness of electrostatic extrema in predicting the location of hydrogen bond partners was examined for a sample of small molecules and amino acid sidechains. Electrostatic extrema were most successful in prediction of single N-H···O=C contacts, where the electrostatic interaction is the major component of the hydrogen bond energy. For multiple hydrogen bond contacts, geometric considerations compete with electrostatic directionality, and the predictions of electrostatic extrema are necessarily poorer. The location of N-H electrostatic maxima were found to be more precisely and consistently defined than oxygen electrostatic minima, due to the broad, featureless nature of the negative potential region around oxygen acceptors. The proton is usually less than 0.6 Å from a potential minimum and the acceptor atom within 1.3 Å of a potential maximum for single hydrogen bonds. This is consistent with the observed spread of electrostatic maxima around N-H groups and electrostatic minima around C=O groups. Carbonyl oxygens are capable of accepting two hydrogen bonds in the lone pair directions. In practice many C=O groups in protein structures accept only one hydrogen bond, and lone pair directionality for single N-H···O=C hydrogen bonds in crystal structures, and theoretical calculations is less pronounced than for twin hydrogen bonds to C=O groups. Geometric descriptions of hydrogen bond would predict the proton to lie at the lone pair directions or along the intervening arc of 3.97 Å (at 0.5 Å outside the oxygen van der Waals surface). The electrostatic minima predict the proton's location to within one-third of this arc, improving upon a geometric description of the location of hydrogen bond partners.

It has also been assumed that different ligands will have many binding site interactions in common. The validity of this assumption was tested by comparing
ligands for the enzyme Glycolate Oxidase (GOX), for which information about the binding site structure is available. The relative orientations produced by matching electrostatic extrema for the isolated inhibitor ‘spearheads’ generally agreed with the relative orientations of the ‘spearheads’ obtained by optimizing their steric and electrostatic interactions within the enzyme binding site. Matching electrostatic extrema, in common with many methods for generating relative binding orientations, assumes a rigid receptor binding site model. For the GOX system recent crystal structure data for an enzyme-ligand complex reveals some movement of binding site residue sidechains in response to ligand binding. Such ‘induced fit’ effects in ligand-receptor binding are not, and could not, be accounted for by the method described here, and may account for some of the differences between the two sets of orientations.

The binding orientations observed in chapter 4 will be modified by a number of factors. Three of the PDE III inhibitors studied contain an axis of chirality (SK&F95800, SK&F93741g and SK&F93741ac). Rotation around the C3-C4' biphenyl bond allows formation of two enantiomeric forms of these compounds. For SK&F95800 and SK&F93741ac the two primary electrostatic extrema used in the reported overlays, O6_min and O8_min, lie along the axis of chirality and therefore their locations in each enantiomer are unlikely to be perturbed significantly. The effect of the biphenyl rotation on the remaining extrema used in these overlays is less apparent. SK&F93741g is more debatable because one of the groups key to the relative binding overlay (C8'=O8') does not lie along the axis of chirality, therefore overlays produced for the two enantiomers will be likely to differ. Chirality is an important aspect of all enzyme mediated processes. Enzymes serve as enantiospecific templates, and are not only able to recognize their substrates stereospecifically, but inhibitors with sufficient structural similarity can also be recognized in the same way. For these compounds, overlays have been produced using only one enantiomer. It is possible that relative orientations with improved RMS errors could have been produced using the opposite enantiomers.

Through the use of a rigid binding site model, the dynamic properties of protein structures have not been considered. As information obtained from structural biology studies increases, so too does the number of enzyme systems in which ligand binding and solvation effects are seen to alter enzyme conformation. At the extreme, ligand binding processes can involve a complex sequence of conformational changes between both the
ligand and the enzyme. Such conformational flexibility may allow a molecule to enter a binding site which it would otherwise be unable to do, if held rigid. The 'lock and key'
analogy is too rigid to realistically model enzyme-ligand binding processes. Protein structures do not generally change dramatically upon ligand binding, and any large changes that do occur are usually related to activating the function of an enzyme. However, because the binding site is relatively small compared to the total enzyme, specific local adjustments in conformation can occur which increase the enzyme's affinity for a ligand without altering the overall structure of the protein.

While the 3D conformation of proteins is relatively stable, it does allow for some conformational flexibility. Although the primary structure of a protein does not give us direct information about the 3D conformation of a protein, each amino acid residue in a protein can exist in a number of conformations which, together, could lead to a large number of possible protein conformations. Steric restrictions and structural characteristics of the polypeptide will eliminate many of these possible conformers. However, small conformational changes are allowed and occur rapidly, meaning that the probability of finding any unique conformer at any given time is small. A ligand will bind to one conformation of an enzyme, but this does not suggest that an identical arrangement of catalytic and binding residues must be achieved for all ligands. It is now held that different ligands can induced different non-identical forms of the activated enzyme, so that the substrate-dependence of the enzyme conformation persists into the formation of the transition state.

7.3 Molecular Similarity
Many different strategies are used in the search for similarity between molecules, and these approaches employ different molecular properties as determinants of similarity. For a set of molecules which are capable of occupying a common receptor site, any methods which successfully identify positions of optimal molecular similarity should arrive at a consensus conclusion, if the physical properties under comparison are determinants of the binding process. If this is not so, then the ‘similarity’ may be in the eye of the beholder (or the similarity algorithm). Some of the similarity methods discussed here would fail to reach such a consensus because of the strategies they employ in their search for similarity. Hence, their usefulness may not be applicable to
systems beyond those which they have been developed for. The approach of ligands to a receptor site and their orientation within a receptor site are generally dominated by electrostatic interactions, so the importance of the electrostatic interaction energy in molecular recognition processes is evident in many of these approaches to similarity. Many of these similarity calculations emphasize, or rely upon, the electrostatic properties of the molecules under comparison. However, they rely on molecular electrostatic properties derived from Mulliken charges (or other approximate methods for describing the charge density) which are well known to be inadequate for an accurate representation of molecular electrostatic properties. An accurate description of the molecular properties is fundamental to any similarity comparison method. The review of similarity began with the statement that the definition of the similarity of two molecules depends upon the qualities which are being compared. Because these methods focus on different aspects of the chemical and physical nature of molecules to describe similarity, their usefulness may lie in their ability to complement each other.

7.4 Refinements and Improvements
This approach has been shown to provide a rapid method for distinguishing probable orientations. For the molecules studied, the number of maxima and minima in the electrostatic potential of each molecule is relatively few. If we assume that two molecules within a binding site largely occupy the same volume then, when surveying possible correspondences of electrostatic extrema, the number of sterically acceptable orientations diminishes rapidly in practice, so the choice of which maxima and minima sites are to matched is reduced even further. The number of ligands considered in these examples has meant that the superpositions could be chosen effectively ‘by hand’ and the resulting relative orientations can be assessed visually by the use of 3D graphics. In keeping with the need to be able to screen large numbers of compounds rapidly this method could be easily automated to compare systematically all possible overlays. This would serve to automate the acceptance or rejection of superpositions on steric grounds. This is perhaps obvious by visual comparison for smaller molecules, but it would have particular application to large systems with many maxima and minima. It would also make the method more rigorous and less subjective in terms of the choice of which extrema are to be used as overlay points.
The RMS values used in this work provided a semi-quantitative measure of each relative orientation. RMS errors are useful because they tend to zero for very good matches of maxima and minima, and become very large for poor matches. However, RMS errors are not particularly adapted to this method, and there is a need for a more specific measure of quantification than RMS errors. The RMS errors only take account of the error in the points chosen for matching in the overlay. No account in the RMS error is taken of any clashes in relative orientations, where maxima of one molecule are close to minima of the other. Only an approximate quantitative agreement between corresponding maxima and minima is given by division of the overlay points into ‘strong’, ‘moderate’ and ‘weak’. The RMS error could be replaced by a more appropriate method of quantification, which includes (either separately of in an appropriate combination) a measure of the separation of overlay points and a penalty function for relative orientations which produce potential clashes.

Useful information about possible relative orientations of ligands within a receptor site is provided by this approach, but it does not include any information on the strength of such binding orientations. This approach is targeted at systems which have strong electrostatic intermolecular interactions, most importantly hydrogen bonds in a biomolecular system. By calculating the electrostatic maxima and minima at different distances beyond the van der Waals surface (depending on whether the extrema are located near hydrogen bond donor groups or hydrogen bond acceptor groups) improved RMS errors for the inhibitors in their relative binding orientations have been demonstrated in comparison with results obtained by calculating all extrema at a uniform distance from the van der Waals surface. To some extent this focuses the approach on hydrogen bonding interactions. However, this does not diminish its usefulness in highlighting other strong electrostatic interactions.

The results demonstrate that, through similarity in the pattern and distribution of extrema in the electrostatic potential of a range of structurally diverse ligands it is possible to generate sterically and electrostatically plausible relative binding orientations. This method has been successfully applied to a number of different ligand-receptor systems. Anisotropy in charge density is an essential component of the theoretical description of molecular electrostatic properties and interactions, in particular hydrogen bonds. Such anisotropy is inherent in distributed multipole models.
for charge density, such as the distributed multipole analysis. Through the use of the DMA the comparison of molecular electrostatic properties has been made without the added uncertainties related to other more approximate charge density models. The molecular electrostatic properties derived from the DMA model can thus be confidently used as a determinant of molecular similarity. The comparison of electrostatic extrema provides a fast and efficient method for generating plausible relative orientations, and clearly highlights the requirements for electrostatic molecular similarity.
References for Chapter 7.


17. Zeneca Agrochemicals, Private Communication.


