LIGAND-MACROMOLECULE INTERACTIONS

R.C. WADE

Thesis submitted for the degree of

Doctor of Philosophy in the University of Oxford

Laboratory of Molecular Biophysics

and Wadham College, Oxford

Trinity Term, 1988

ABSTRACT

LIGAND-MACROMOLECULE INTERACTIONS

Rebecca Claire Wade
Wadham College

D. Phil. Thesis
Trinity Term, 1988

The optimisation of ligand-macromolecule interactions is fundamental to the design of therapeutic agents. The GRID method is a procedure for determining energetically favourable ligand binding sites on molecules of known structure using an empirical energy potential. In this thesis, it has been extended, tested, and then applied to the design of anti-influenza agents.

In the GRID method, the energy of a hydrogen-bond is determined by a function which is dependent on the length of the hydrogen-bond, its orientation at the hydrogen-bond donor and acceptor atoms, and the chemical nature of these atoms. This function has been formulated in order to reproduce experimental observations of hydrogen-bond geometries. The reorientation of hydrogen atoms and lone-pair orbitals on the formation of hydrogen-bonds is calculated analytically.

The experimentally observed water structures of crystals of four biological molecules have been used as model systems for testing the GRID method. It has been shown that the location of well-ordered waters can be predicted accurately. The ability of the GRID method to assist in the assignment of water sites during crystallographic refinement has been demonstrated. It has also been shown that waters in the active site of an enzyme may be both stabilized and displaced by a bound substrate.

Ligands have been designed to block the highly conserved host cell receptor site of the influenza virus haemagglutinin in order to prevent the attachment of the virus to the host cells. The protein was mapped energetically by program GRID and specific ligand binding sites were identified. Ligands, which exploited these binding sites, were then designed using computer graphics and energy minimization techniques. Some of the designed ligands were peptides and these were synthesised and assayed. Preliminary results indicate that they may possess anti-influenza activity.

ACKNOWLEDGEMENTS

Many people have helped me over the past three years and I would like to express my gratitude to all of them. In particular, I wish to thank:

Dr. Peter Goodford, my supervisor, for providing encouragement, guidance and many stimulating ideas;

Professor Sir David Phillips for the opportunity to work in the Laboratory of Molecular Biophysics;

the Science and Engineering Research Council for the provision of a research studentship;

Jenny Martin for proof-reading and numerous helpful discussions;

my GRID coworkers: Kevin Clark, David Boobbyer, Gary Alderson and Douglas Campbell;

Dr. Garry Taylor for coping with many computer problems and queries;

Pam Batchelor for her organisational role;

Dr. Louise Johnson for the use of her wordprocessor;

Dr. Graham Richards and all members of his group for including me in their activities and, in particular, Dr. Chris Reynolds for assistance with AMBER and AMPAC;

Professor George Brownlee for agreeing to the synthesis and testing of peptides in the Sir William Dunn School of Pathology, Dr. Keith Gould for his collaboration in carrying out the experimental work, and Rita Anand and Hilary Scotney for their patient instruction to a physicist in unfamiliar surroundings;

Jane and John for many sustaining telephone conversations and the provision of a roof over my head, Amyas for assistance with graphics, and lastly, Dharmendra for his friendship and support.

SYMBOLS AND ABBREVIATIONS

Å	Ångström units ($1Å=10^{-10}m$)
ß-CD	ß-cyclodextrin
cAMP	cyclic adenosine monophosphate
DHFR	Dihydrofolate reductase
DNA	Deoxyribonucleic acid
ε_0	Permittivity of a vacuum
HA	Influenza haemagglutinin
HL	Human lysozyme
HEWL	Hen egg white lysozyme
kcal	Kilocalorie (1kcal=4.184kJ)
MEP	Molecular electrostatic potential
NA	Influenza neuraminidase
nmr	Nuclear magnetic resonance
QSAR	Quantitative structure activity relationship
rms	Root mean square
RNA	Ribonucleic acid
TEWL	Tortoise egg white lysozyme

Program GRID

r	Interatomic distance
t	Angle of deviation of a hydrogen-bond from linearity at a target atom
p	Angle of deviation of a hydrogen-bond from linearity at the probe
t_o	Angle of deviation of a hydrogen-bond from the plane of the lone pair orbitals of a target atom

p_o	Angle of deviation of a hydrogen-bond from the plane of the lone pair orbitals of the probe
p_b	Angle of deviation of a hydrogen-bond from the bisector of the lone pair orbitals of the probe
p_{ad}	Angle subtended by a target hydrogen-bond acceptor atom and a target hydrogen-bonding hydrogen at the probe
E_{tot}	Total interaction energy of the probe with the target
E_{lj}	Lennard-Jones energy
E_{el}	Electrostatic energy
E_{hb}	Hydrogen-bond energy
Er	Hydrogen-bond term dependent on r
Et	Hydrogen-bond term dependent on t
Ep	Hydrogen-bond term dependent on p

<u>The analysis of the predictions of water made by program GRID</u>

δd	Distance between the experimentally observed water position and the water position predicted by program GRID
E_{obs}	Energy calculated by program GRID at the experimentally observed water position
E_{pred}	Energy calculated by program GRID at the predicted probe position
δE	$= E_{obs} - E_{pred}$
B_{obs}	Temperature factor of the water at the experimentally observed position
B_{pred}	Calculated temperature factor for the water at the position predicted by program GRID
$\langle u^2 \rangle^{\frac{1}{2}}$	Observed root mean square displacement of the water from its mean rest position

CONTENTS

CHAPTER 1

INTRODUCTION

1.1 LIGAND-MACROMOLECULE INTERACTIONS

Interactions between ligands and macromolecules are vital to the function of biological systems [1,2,3]. Examples are the interactions between an enzyme and its substrate; a receptor and a drug; and an antibody and an antigen. These interactions are often highly specific enabling molecules to recognise each other [3,4,5,6]. Many ligands bind reversibly and non-covalently, and non-covalent interactions also contribute to the recognition between molecules that bind covalently [7]. It is these non-covalent interactions that are studied in this thesis.

1.1.1 Non-covalent interactions

The non-covalent interaction energy may be considered to consist of the following five terms.

1. The repulsive electron overlap energy. Electrons obey the Pauli exclusion principle which states that electrons of the same spin cannot occupy the same space simultaneously and thus they repel each other. This energy may be modelled by a distance dependent C/r^{12} [8,9] or exponential term [10].

2. The weak attractive dispersion energy. This is due to the correlated motion of electrons on neighbouring molecules resulting in instantaneous induced dipole-induced dipole interactions. It may be

given by a term proportional to $\alpha_1\alpha_2/r^6$ where α_1 and α_2 are molecular polarizabilities [11].

3. The electrostatic energy. If the charge distribution of the molecules is represented by partial point charges at the atomic centres, this may be given by a Coulomb term $q_1q_2/4\pi\varepsilon r^2$ where q_1 and q_2 are atomic charges and ε is the dielectric constant of the surrounding medium. It is repulsive for like charges and attractive for unlike charges.

4. The attractive charge transfer energy. This is due to the partial migration of electrons between interacting molecules.

5. The attractive polarization or induction energy. This is due to permanent dipole-induced dipole interactions. It may given by a term proportional to $-\mu^2\alpha/r^6$ where μ is the permanent dipole moment inducing a dipole on a molecule of polarizability α.

All these terms contribute to the energy of a hydrogen-bond [12] which makes a particularly important contribution to the specificity of intermolecular interactions [5,6] and is therefore discussed in detail in Chapter 3.

The electromagnetic forces giving rise to non-covalent interactions are distance dependent and are only effective in biological systems for interatomic separations up to about 12Å [7]. The approach of a ligand to within this range of a macromolecule occurs by diffusion and is governed by Brownian motion and the Boltzmann distribution. The subsequent binding of the ligand is dependent on the steric accessibility of the binding site on the macromolecule, the magnitude of the stabilization energy of the ligand-macromolecule complex and entropic factors.

The solvent plays an important role in ligand-macromolecule interactions [13] and has therefore been studied in this thesis (see Chapter 4). Water has a high dielectric constant and consequently shields the electrostatic interaction between the ligand and the macromolecule. However, it is also responsible for the hydrophobic effect [14]. This causes water to make a favourable contribution to the free energy of ligand binding because of the positive entropic change which occurs on the displacement by the ligand of ordered waters to bulk water. It gives rise to favourable interactions between non-polar solutes surrounding which water molecules may form cage-like hydrogen-bonded networks. Thus, hydrophobicity can be considered to be a major factor in stabilizing molecular associations [3,14,15].

1.1.2 Theoretical studies of the ligand-macromolecule interaction

Three models of the ligand-macromolecule interaction may be considered.

1. The 'lock and key' model [16] in which the ligand and macromolecule are rigidly complementary.
2. The 'zipper' model [2] in which the ligand is flexible in solution and adopts a conformation complementary to a rigid receptor site on binding. Segments of the ligand bind sequentially resulting in less entropic loss and faster binding than in the 'lock and key' model.
3. The 'induced-fit' model [17] in which the binding of the ligand causes a change in the conformation of the macromolecule. This can occur by the 'zipper' action. The conformation of the ligand may

also be altered on binding. The ligand may act allosterically triggering an action at another site on the macromolecule.

The latter model is generally accepted as being the most realistic.

Theoretical methods [18] of studying ligand-macromolecule interactions include model building, using such techniques as computer graphics and distance geometry, and energy calculations. The latter may be performed using ab initio molecular orbital theory, semi-empirical quantum mechanics and classical molecular mechanics in energy minimization, Monte Carlo simulation and molecular dynamics.

In this thesis, study is directed primarily at ligand-macromolecule interactions of pharmacological interest.

1.2 THE DISCOVERY OF THERAPEUTIC AGENTS

1.2.1 Introduction

There are, perhaps, 30,000 diseases worldwide, and for most of these there is no available drug treatment [19]. Disease may be caused by external agents such as viruses, bacteria, protozoa, helminths or other parasites; by damage e.g. cancer; or internally by genetic inheritance e.g.sickle cell anaemia [20].

A drug may be defined as any natural or synthetic chemical substance that produces a biological response of therapeutic value. The traditional methods of drug discovery are [20,21]:

1. the large scale screening of compounds, e.g. tetracyclines were discovered after systematic screening for antibiotic-producing micro-organisms in the soil;

2. the extraction of the active substances in natural products, e.g. salicylic acid from willow bark;

3. the follow-up of unexpected results in modern medical, physiological and biochemical investigations e.g. penicillin;

4. the exploitation of biochemical information on key metabolic pathways or other biochemical activities in man and pathogens e.g. the inhibition of the biosynthesis of tetrahydrofolic acid by methotrexate in cancer chemotherapy.

These methods result in only about five novel drugs being brought on to the market each year [20].

Recently, methods of designing therapeutic agents have been developed which exploit the concept of the drug-receptor interaction.

1.2.2 Receptors

A firm definition of a receptor has not been agreed on [22], but it may be considered to be a cellular entity with which ligands interact eliciting a biological response [23,24,25]. These ligands may be endogenous agonists causing activation of the receptor e.g. a neurotransmitter; or they may be exogenous acting, for instance, as antagonists causing the blocking of the receptor binding site to agonists.

The concept of a receptor was first proposed by Langley in 1878 [26]. However, it was Ehrlich who, while working independently of Langley, first introduced the term 'receptor' and applied it to the interaction of drugs [27]. He also went on to propose the existence of a 'pharmacophore' consisting of the essential atoms in a drug necessary for it to bind to a receptor. Following these early

theories, receptor theory was given a quantitative basis in the 1920s and 1930s by Clark who applied the law of mass action to the drug-receptor interaction [28,29] and Gaddum who investigated the competitive action of antagonists with agonists at the receptor [30].

The functions of a receptor are to recognise and distinguish biologically active molecules by specific, high affinity binding and to mediate the transduction of the signal generated by ligand binding either directly or via a secondary messenger such as cAMP to an effector e.g. an enzyme or an ion channel. This triggers a cascade of biochemical reactions leading finally to a characteristic physiological or pharmacological response [25].

Receptors may be found in the cell membrane, the cytosol and the nucleus [25]. Most receptors are lipoproteins embedded in the plasma or cell-organelle membrane as intrinsic proteins. These receptors are difficult to isolate as their structure and function may be maintained by the surrounding membrane. Receptors may also be nucleic acids, lipids, glycoproteins or pure proteins [21]. Although there is no rigid distinction between the receptor-drug and the enzyme-substrate interaction, the basic difference is that in the former, it is the receptor while in the latter it is the substrate that is changed.

1.2.3 The design of therapeutic agents

The strategy employed in the design of therapeutic agents depends on whether the identity and structure of the receptor is known.

1.2.3.1 Unknown receptor structure

If the receptor structure is unknown, a 'lead' compound possessing some of the required activity may be obtained by one of the methods described in Section 1.2.1 or it may be generated by theoretical methods. Its properties are then optimised in order to produce a potent and specific therapeutic agent. The major theoretical techniques that may be used are now considered.

1.2.3.1.1 Quantitative Structure Activity Relationships (QSAR)

These relate the biological activity of a series of compounds to their physical, chemical and structural properties [31,32].

The principal method used is Hansch analysis [33,34] in which physicochemical parameters of a group of congeneric structures are correlated with their observed biological activities. Multivariate regression analysis is used giving rise to a linear or parabolic relationship, known as an extrathermodynamic or linear free-energy relationship. Three parameters which commonly show a high correlation to the biological response are the substituent hydrophobicity constant π, the Hammett substituent σ and the steric Taft E_S value.

Structural rather than physicochemical parameters are used in the Free-Wilson or de novo method [35] where biological activity is related directly to the chemical nature of substituents at particular positions in the molecule by linear regression. It is assumed that a substituent at one position does not affect the response to a substituent at another point in the molecule.

These two methods may be used in conjunction with each other. In addition, other parameters such as molecular connectivity [36] or those obtained from molecular orbital calculations [37], may also be correlated to biological response. Parameters for which a statistically high degree of correlation with activity is observed can then be identified as important in determining ligand binding. The optimisation of these parameters can be used to design more active compounds.

1.2.3.1.2 <u>Pharmacophore and receptor mapping</u>

A 'pharmacophore' may be defined as the three dimensional arrangement of the chemical groups in a ligand which are essential for recognition and activation [24,38]. It may be mapped by studying, (e.g. by the Active Analogue Approach [39]), the correspondence between pharmacophoric groups in a set of active compounds. For each compound, all the energetically available conformations, rather than just the minimum energy conformation, are explored in a systematic search and the relative positions of the pharmacophoric groups recorded for each possible conformation. A pharmacophoric pattern common to all active compounds may then be identified. To reduce the extent of conformational search necessary, distances within the active molecules can be constrained using the distance geometry technique [40].

Pharmacophoric patterns may also be identified by searching databases for molecules containing common substructures by statistical pattern recognition techniques [41].

The identified pharmacophore may be used to generate novel

ligands e.g. by searching a database for molecules satisfying the properties of the pharmacophore [42].

The receptor site volume and shape may be mapped by examining inactive as well as active compounds. An inactive compound may fit the identified pharmacophore but it may make sterically unfavourable interactions with the receptor. The active drugs may be superimposed to define a 'receptor-excluded' volume. The difference between this and the volume occupied by the inactive drugs defines the 'receptor-essential' volume where steric overlap occurs between the receptor and the inactive drugs [39].

Molecular electrostatic potentials (MEP), calculated by molecular orbital techniques may be used to map an 'interaction' pharmacophore. The MEP of a molecule is the interaction energy between a unit positive charge and the molecule. This approach has been applied in order to discriminate between agonists and antagonists of the 5-hydroxytryptamine receptor [43].

The molecular similarity of active compounds may be examined in order to identify the pharmacophore [44]. In assessing molecular similarity, one must choose variables to characterise the molecules e.g.MEP; weighting schemes for the variables; and similarity coefficients to determine their resemblance, e.g. an index of molecular similarity based on the electronic properties of two molecules may be computed [46]. Molecular similarity may be used to compare two or more compounds and a series of compounds can be grouped automatically according to similarity by cluster analysis [45].

1.2.3.2 Known receptor structure

If the structure of the receptor is known, it may be possible to design therapeutic agents by the method of receptor fit [47,48]. The emergence of this as a feasible method has been reliant on the recent advances in the techniques of determining the three-dimensional structures of macromolecules.

1.2.3.2.1 Methods of determining the structure of the receptor

The principal technique used is x-ray crystallography. The first protein structure, that of sperm whale myoglobin, was observed in 1958 by x-ray crystallography [49] and details of its structure were revealed in 1960 [50]. The first enzyme structure followed in 1965 when the structure of hen egg-white lysozyme was published [51].

There are now more than 350 macromolecular structures [52] which have been solved by x-ray crystallography and the rate at which structures are being determined is ever increasing. This recent escalation can be ascribed to the following factors [20,53].

1. The availability of powerful computers with the ability to perform large calculations and the development of computer graphics.
2. The use of the high intensity, broad spectrum x-ray synchrotron source. This permits both small crystals and crystals with large unit cells to be studied. In addition, the fast collection of diffraction data by the Laue method using the white x-ray beam is now possible for protein crystals [54].
3. The development of rotation cameras and area detectors enabling fast collection and processing of data.

4. New methods of solving the phase problem such as molecular replacement and direct methods.

5. Improved crystallization techniques aided by the availability of pure preparations of cloned and expressed proteins. However, crystallization still remains a very unpredictable and problematic stage in the process of structure determination.

6. The development of gene and protein sequencing techniques allowing faster interpretation of electron density maps.

7. The development of site-directed mutagenesis allowing the detailed examination of intermolecular and intramolecular interactions.

Neutron diffraction may be used to determine receptor structures with hydrogen coordinates but it is not often applicable to large molecules because large crystals are required due to the low flux currently available from neutron sources, and there are a large number of hydrogens in crystals of biological macromolecules [55].

The structure and dynamics of biological molecules may also be studied by the newer technique of nuclear magnetic resonance (nmr). This has the great advantage that the molecule may be studied in solution and does not have to be crystallized. Experimental observations of spin-spin coupling constants, nuclear Overhauser effects and hydrogen-bonding enable interatomic distances to be defined which, with the application of distance geometry, molecular dynamics and database searching, may be used to determine the structure of small proteins, as well as details of ligand binding and molecular motion [39].

In addition, receptor structures may be modelled. Two methods of modelling proteins are as follows [39]:

1. Homologous proteins may be modelled when the crystal structure of one is known and that of a related one can be deduced by building in side chains and subjecting the protein to energy minimization.

2. The secondary and tertiary structure of proteins may be modelled from their amino acid sequence. However, statistical methods are able to predict secondary structure with only approximately 60% accuracy [56]. Heuristic approaches may give better predictions e.g. turns can be predicted with about 90% accuracy [57].

With these advances in structure determination, it is now possible to choose receptors as targets for molecular design because of their therapeutic relevance rather than because their crystal structure is already available, e.g. renin (see Section 1.2.3.2.3).

1.2.3.2.2 Methods of studying the properties of the receptor

In investigating the ability of the receptor to bind to ligands, the steric and energetic properties of the receptor may be studied.

1. **Steric properties**

Two methods of sterically examining the receptor are as follows.

a. The construction of the solvent accessible surface area [58]. This is given by the locus of the centre of a sphere of radius 1.4Å, representing a water molecule, as it is rolled over the surface of the receptor. Alternatively, the solvent accessible molecular surface [59] can be displayed using the Connolly algorithm [60]. This surface is the locus of the surface of a spherical water probe of

radius 1.4Å rolled over the receptor, and consists of the part of the van der Waals' surface of the atoms accessible to the probe sphere, termed the contact surface, connected by a network of concave and saddle-shaped surfaces, known as the reentrant surface. The surface of a ligand can be similarly defined and the ligand docked interactively by matching the surfaces of the two molecules [61].

b. The construction of receptor spheres to represent the volume of the binding site [62]. The fit of these spheres and spheres representing the shape of a rigid ligand can be maximized in designing a ligand to fit complementarily into the receptor site.

Both these methods have been applied to the study of the binding of thyroxine analogues to prealbumin [62,63]. High affinity analogues of thyroxine were found to fill more of the hormone-binding site than low affinity analogues.

2. **Energetic properties**

The energetics of the receptor-ligand interactions may be studied by the following methods:

a. The calculation of the MEP which may be displayed by means of contours around the receptor [64] or by colour-coding the receptor surface [65].

b. The energetic mapping of the receptor site by the calculation of the binding energy of small chemical probe groups. The calculated energies may be displayed as three dimensional energy contours as in the GRID method [66] studied in this thesis or interpreted by picking out peak binding energies [67]. Alternatively, hydrogen-bonding regions may be identified using geometrical constraints alone [65,68].

c. The calculation of the binding energy of particular ligands using quantum mechanics [69,70,71] or classical molecular mechanics [72,73,74,75,76]. Energy minimization may be performed to optimise the conformation of the ligand and the receptor (e.g. see Chapter 5.). This may be followed by molecular dynamics simulations of the complex. Thermodynamic cycle perturbation and integration techniques [77] may be used to calculate the relative binding affinities of different ligands for a receptor.

All the above methods of studying receptors are complementary and a combination may be used in the design of a ligand to bind to a receptor.

1.2.3.2.3 <u>Design strategies in the method of receptor fit</u>

Therapeutic agents may be designed to act in a variety of ways.

1. They may bind to a biomolecule which is functioning incorrectly.

This strategy was adopted in the design of an anti-sickling agent using the known three-dimensional structure of haemoglobin [78]. Novel left-shifting compounds were designed which bound preferentially to the oxy conformation of haemoglobin between the N-terminal residues of the α-subunits at a site where no naturally occurring ligand had been observed. These compounds caused a reduction in the concentration of the deoxy conformation of haemoglobin and thus in the number of haemoglobin molecules able to undergo sickling. One of these compounds is a potent inhibitor of the sickling of erythrocytes in sickle cell disease and is now undergoing clinical trials [79].

2. They may block the action of infectious organisms by binding to their proteins.

In this thesis, this strategy has been applied to the design of anti-influenza agents (see Chapter 5). If a homologous protein exists in humans, a therapeutic compound must bind selectively to the exogenous protein. For example, inhibitors of the enzyme dihydrofolate reductase (DHFR) have been investigated which have different affinities for the bacterial and mammalian enzymes and hence have antibacterial properties [80,81].

3. They may modify the action of human enzymes or other molecules by binding to them.

For example in order to treat proliferative diseases, drugs may be targeted at DHFR [82]. When blocking enzymes, they may be designed as transition-state analogues in order to bind more tightly than the natural substrate. An example of this is the design of renin inhibitors as antihypertensive agents [83]. The structure of renin was modelled from the crystal structures of homologous aspartic proteases. Peptide inhibitors were then designed which contained substitutions chosen to mimic the transition-state conformation of the scissile bond. The structure of renin has now been solved by x-ray crystallography [79] and should facilitate the design of further antihypertensive agents.

Not only must a drug bind selectively to its receptor, it must also possess appropriate physicochemical properties in order to be clinically useful [47]. It must be non-toxic, stable and readily absorbed. In addition, it should be patentable and easy to synthesise. The method of receptor fit may also be used to optimise

these properties. For example, in the first application [84] of the method, the solubility of compounds designed to stabilise deoxy haemoglobin and produce a right-shift in the oxygen saturation curve, was increased by adding a hydrophilic side chain at a position where steric hindrance with the receptor would not occur and where additional favourable contacts could be made.

The first and only clinical drug that can be said to have been designed by the method of receptor fit is Captopril. This is an inhibitor of angiotensin converting enzyme used in the treatment of hypertension. It was designed to bind to the active site of the enzyme which was modelled on the basis of its similarity to carboxypeptidase A whose crystal structure was well-defined [85]. However, the method of receptor fit is complementary to the many other methods of drug discovery and is now applicable to the design of therapeutic agents against many serious diseases. It can also be used in the design of herbicides, fungicides and pesticides.

In this thesis, the GRID method [66] of designing ligands by receptor fit is extended, tested and applied. This is a method of identifying energetically favourable ligand binding sites on molecules of known structure. It may be used to design novel compounds which are able to bind specifically to particular receptors.

1.3 OUTLINE OF CHAPTERS

In the next chapter, the GRID method of predicting ligand binding sites is described.

In Chapter 3, hydrogen-bonding and its treatment in the

GRID method are discussed. The modelling of the directional properties of hydrogen-bonds is described.

In Chapter 4, the prediction of the water structure in crystals of biological molecules by program GRID is described. Water has been studied because it serves as a model with which to test the ability of the GRID method to reproduce experimental observations and because of the important influence of water on ligand binding.

In Chapter 5, an account is given of the application of the GRID method to the design of anti-influenza agents. A number of other theoretical methods were also used in the design of these compounds and some of these compounds were then synthesised and assayed for activity.

Chapter 6 provides a summary of the results obtained and indicates possible directions for the future extension of the work presented in this thesis.

CHAPTER 2

THE GRID METHOD OF DETERMINING FAVOURABLE LIGAND BINDING SITES

2.1 INTRODUCTION

Program GRID [66,86] has been developed as a tool for use in the design of novel therapeutic compounds. It is concerned with one aspect of the design of drugs, namely the prediction of how a ligand binds to a macromolecule. It computationally determines energetically favourable ligand binding sites on macromolecules using an empirical energy function.

A molecular mechanics method is employed rather than a quantum mechanical one because molecular orbital calculations for large systems can be time-consuming and are usually performed 'in vacuo', whereas molecular mechanics calculations are quicker, and can therefore take some account of the solvent present in biological systems either explicitly or by means of a dielectric constant.

2.2 THE GRID METHOD

In order to determine ligand binding sites, interaction energies are calculated between a small fragment of a ligand, called a probe, and a known molecular structure which is called the target, at different positions of the probe around the target. These energies are then displayed as energy contours over the chosen molecular target, using three-dimensional computer graphics. Different energy contours

may be obtained for different probes enabling specific binding sites to be identified.

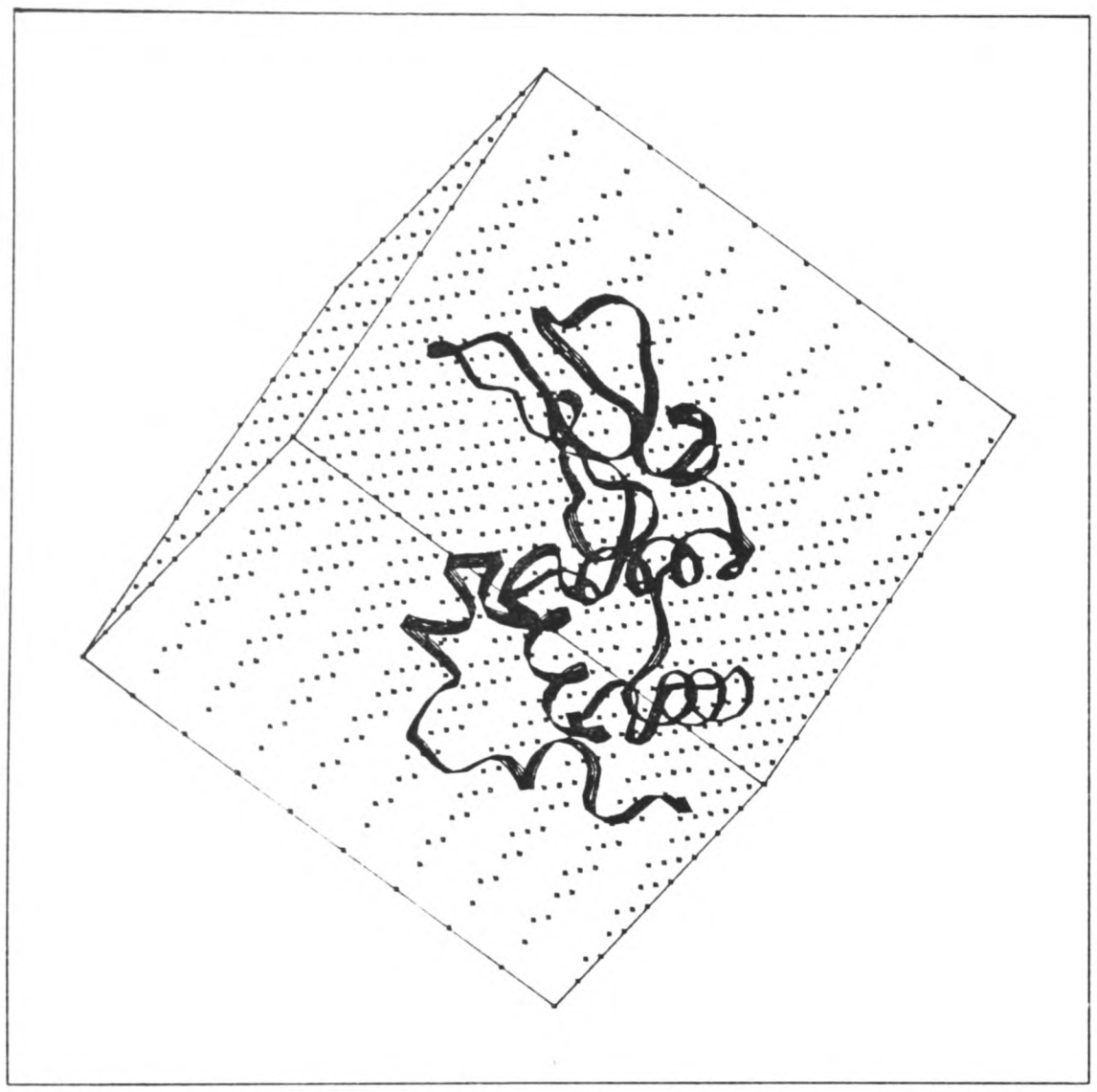
2.2.1 The probe

This may either be a single atom or a small recognisable structural unit such as a methyl group, a water molecule, or an amine nitrogen. Thus, a ligand may consist of a combination of different probe groups each considered separately by program GRID. This treatment neglects mesomeric and inductive effects between the different probe groups in the ligand and also does not take into account conformational restrictions imposed by the structure of the target on the arrangement of the probe groups in the ligand.

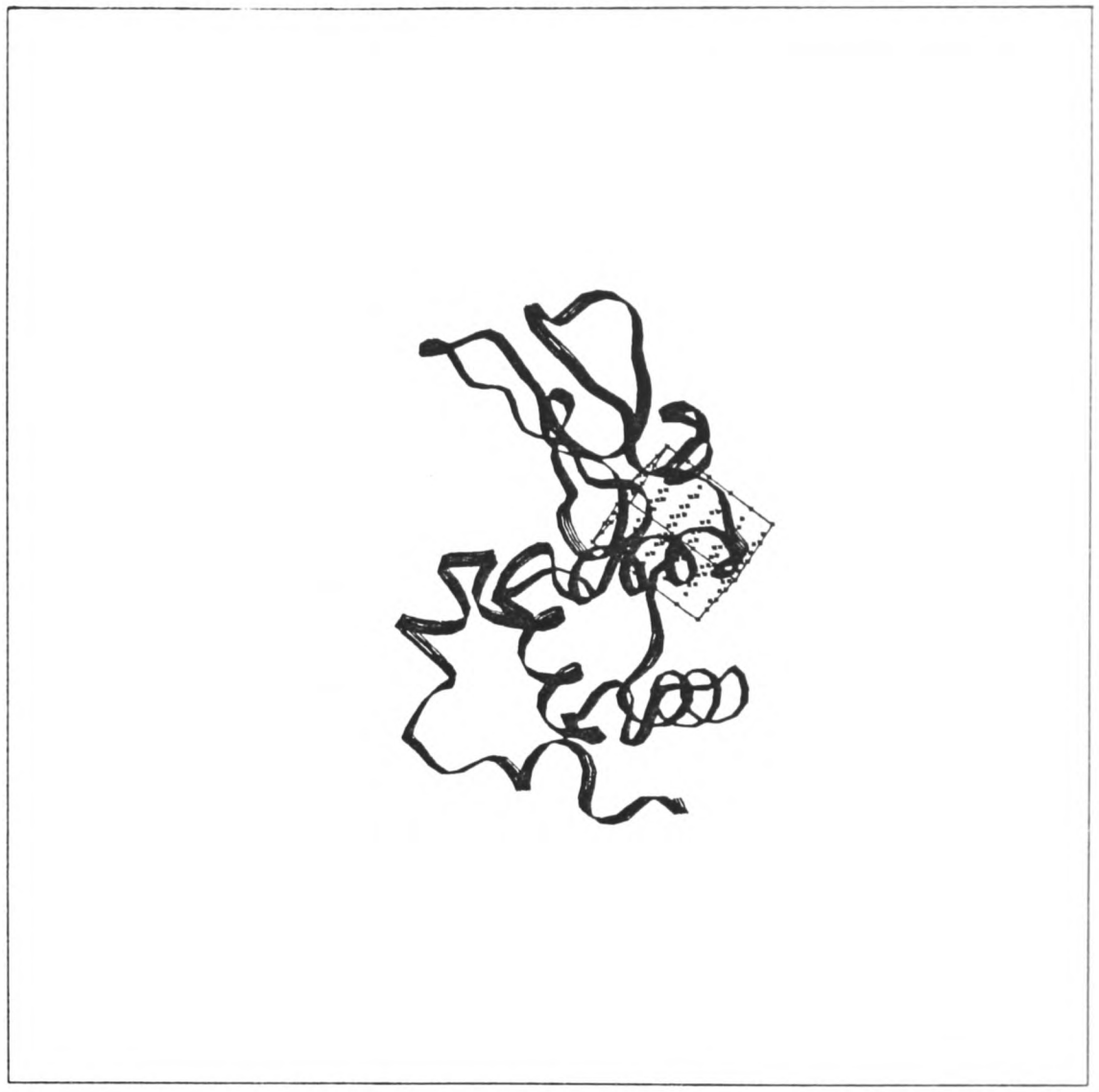
The probe is represented as an extended atom, i.e. as a sphere centred on a point charge consisting of one atom heavier than hydrogen and an appropriate number of hydrogen atoms which are not considered explicitly. The probe is assigned nine parameters which are used to calculate its interaction energy. These are given for a water molecule and a nitrate ion by way of example in Appendix III.5.

2.2.2 The target

This may be a macromolecule e.g. a protein, a nucleic acid, a glycoprotein, or a polysaccharide; a molecular array e.g. a membrane; or a small organic molecule e.g. a drug. It may be considered in an aqueous environment. The target is defined by its atomic coordinates and by parameters describing the chemical nature of its constituent atoms. For large molecules, an 'extended atom'



a.



b.

Figure 2.1 A target molecule, Human Lysozyme (HL), and the spatial array of points at which the probe is positioned and its interaction energy is calculated by program GRID. An array is shown covering a. the whole target molecule and b. just a small region (the cavity of HL containing the internal waters). Calculations using these arrays are described in Section 4.5.10.

representation is adopted in which hydrogen coordinates are only used in order to determine the geometry of hydrogen-bonds. For small molecules, an 'all atom' representation may be used in which the interactions of hydrogen atoms are calculated explicitly. The 'extended atom' representation is advantageous because it significantly reduces the complexity of the target resulting in much faster calculations. It is also useful when studying macromolecular structures determined by x-ray crystallography for which hydrogen positions are not usually available. However, it results in the loss of dipole and quadrupole moments and of steric effects due to the hydrogen atoms. The potential energy function used must therefore include a hydrogen-bond term in order to account for these effects as well as for the effects due to the lone pair electrons which are not treated explicitly in either of these representations and are not observed by x-ray crystallography in biological molecules.

2.2.3 The calculation

The probe is passed through a regular spatial array around the target, and the interaction energy between the probe and the target is calculated at each point in the array. The array may cover the whole target or just a small region of interest as shown in Figure 2.1. The distance between the sampling points is chosen according to the accuracy required and takes values up to 1Å.

Once the interaction energies have been calculated at all points in the array, the location of energy minima can be identified. These minima correspond to favourable binding sites for the probe on

the target. The interaction energies can be converted into energy contours and displayed using computer graphics.

Selectivity contours [87] can also be displayed. These are derived by calculating the difference in interaction energy at each point in the array for two different probes interacting with the same target or for one probe interacting with two similar targets. They enable the relative binding properties of different molecules to be assessed.

The programming and operation of the GRID method are described in Appendix III.

2.2.4 The energy function

The energy function consists of Lennard-Jones (E_{lj}), electrostatic (E_{el}) and hydrogen-bonding (E_{hb}) terms. The total energy (E_{tot}) is computed as the sum of the pairwise interactions of the probe group with each atom in the target.

$$E_{tot} = \Sigma E_{lj} + \Sigma E_{el} + \Sigma E_{hb}$$

2.2.4.1 The Lennard-Jones interaction

This is given by:

$$E_{lj} = A/r^{12} - B/r^{6}$$

where $A = \frac{1}{2}BR_0^{6}$

$R_0 = R_p + R_t$

$B = (1\frac{1}{2}e\sqrt{(h/m)\alpha_p\alpha_t})/C$

$C = \sqrt{(\alpha_p/N_{effp})} + \sqrt{(\alpha_t/N_{efft})}$

where

r is the distance between the probe and an atom in the target in Å;

R_p and R_t are the van der Waals' radii of the probe and the target atom respectively;

α_p and α_t are the polarizability of the probe and the target atom respectively;

N_{effp} and N_{efft} are the number of effective electrons of the probe and the target atom respectively.

A cutoff distance of 8Å is imposed on E_{lj}. At this distance E_{lj} is about -0.01 kcal/mol and the resultant error in the total energy when summed over all atoms is of the order of 0.1 kcal/mol [66].

2.2.4.2 <u>The electrostatic interaction</u>

This is given by:

$$E_{el} = q_p q_t / 4\pi\varepsilon_0 \gamma \{1/r + [(\gamma - \delta)/(\gamma + \delta)]/(r^2 + 4 s_p s_t)^{\frac{1}{2}}\}$$

where

r is the distance between the probe and an atom in the target in Å;

q_p is the electrostatic charge of the probe;

q_t is the electrostatic charge of the target atom;

δ is the dielectric constant of the solvent phase;

γ is the dielectric constant of the target phase;

s_p is the nominal depth of the probe in the target phase in Å;

s_t is the nominal depth of the target atom in the target phase in Å;

In this expression, the Coulomb electrostatic energy for a homogeneous medium, $E_{el} = q_p q_t / 4\pi\gamma\varepsilon_0 r$, has been modified [66] to take account of the heterogeneous nature of the system. An infinite planar interface is assumed to separate a homogeneous target phase of

a.

b.

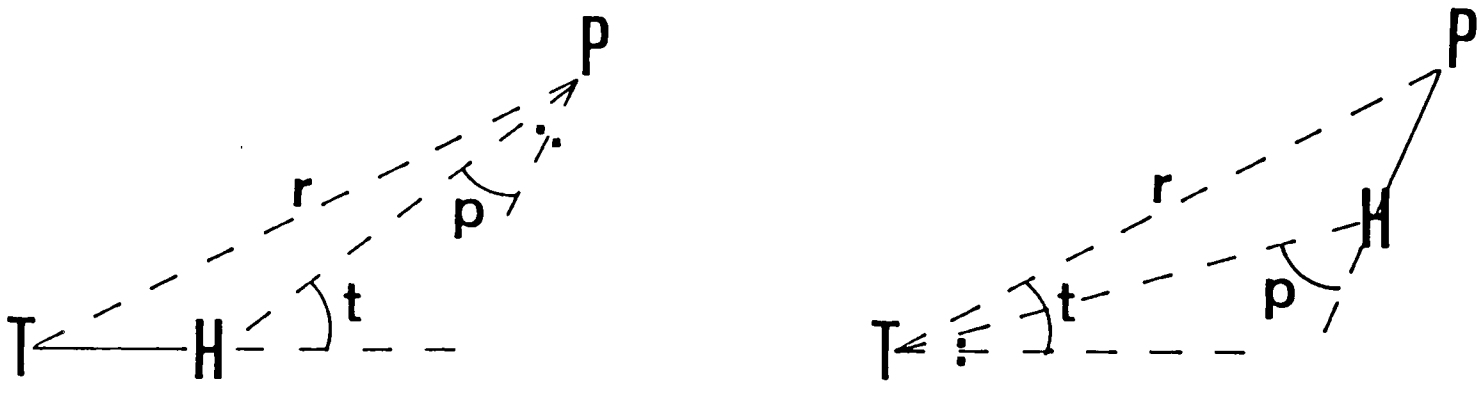


Figure 2.2 The definition of angles t and p when a hydrogen-bond is a. donated and b. accepted by a target atom T. P represents the probe.

dielectric constant γ from a homogeneous solvent phase of dielectric constant δ. For a target protein in aqueous medium, values of γ=4 and δ=80 are used. The electrostatic interaction is calculated using the method of images [88].

The nominal depth s_t of each target atom in the target phase is estimated by counting the number of neighbouring target atoms within a distance of 4Å and translating this into an equivalent depth using a calibrated scale. The nominal depth s_p of the probe is calculated in the same way [66].

The electrostatic interaction is evaluated for every probe-target atom interaction as its magnitude does not diminish rapidly with distance.

2.2.4.3 The hydrogen-bond interaction

This is given by:

$$E_{hb} = Er \times Et \times Ep$$

where Er, Et and Ep are functions of r, t and p respectively. t and p are the angles made by the hydrogen-bond at the target atom and the probe respectively as defined in Figure 2.2.

The modelling of hydrogen-bonds is described in detail in the next chapter.

2.3. THE DEVELOPMENT OF THE GRID METHOD

During the development of a predictive method, it is vital to test the method by examining the agreement between predictions and experimental observations. For this reason, a comparison between the

predictions made by program GRID of the binding of water to biological molecules and crystallographic observations was made and this is described in Chapter 4. As a result of the examination of the binding of water, the modification and extension of the program were found to be necessary. Most of the required improvements were concerned with the treatment of hydrogen-bonds and this is considered in Chapter 3.

Three versions of program GRID were used for the work presented in this thesis: GRID0, GRID1 and GRID2; each with increasingly sophisticated treatments of hydrogen-bonding. The differences between these are outlined in Appendix III. The current version of the program, GRID2, is able to produce results which are consistent with experimental observations and can therefore be used with some confidence for the prediction of ligand binding although the program is naturally still undergoing improvement.

CHAPTER 3

THE ROLE OF HYDROGEN-BONDS IN LIGAND BINDING

3.1 INTRODUCTION

Hydrogen-bonds are important in determining the conformation and intermolecular interactions of proteins, polysaccharides, nucleic acids and organic molecules [89]. The strength of such bonds varies in a complex manner, and a wide variety of empirical energy functions [90] have been devised in order to compute their strength. In general terms, they may be considered as intermediate range interactions between an electron-deficient hydrogen atom and a nearby region of high electron density [91]. Both the donor atom, which is covalently bonded to the hydrogen atom, and the acceptor atom are highly electronegative. When strong hydrogen-bonds are formed, the donor and acceptor atoms approach closer to each other than the sum of their van der Waals' radii [89].

The physical origins of the hydrogen-bond are subject to debate. In 1928, Pauling [92] described the hydrogen-bond as being electrostatic and due to ionic forces between a partially positively charged hydrogen atom and a negative lone pair on the acceptor atom. More recently, quantum mechanical explanations, involving Valence Bond [93,94], Charge Transfer [95], or Molecular Orbital theories [96,97,98,99] have been proposed. These suggest that, at small separations, the hydrogen-bond is governed by a balance between attractive electrostatic, charge transfer, polarization and dispersion components and repulsive electron exchange terms, while at large

distances the hydrogen-bond is almost completely electrostatic. Charge transfer is the second largest attractive component and it enables stabilization of hydrogen-bonds such as O-H--O, where part of the electron charge occupies the anti-bonding orbital causing weakening and lengthening of the O-H bond [95]. Most of the charge redistribution, though, is due to the movement of the σ-electrons from the proton acceptor to the donor atom [99]. The large electron exchange repulsion prevents bond bending and favours the occurrence of linear hydrogen-bonds.

The treatment of hydrogen-bonds in molecular mechanics potentials is shown in Table 3.1. These potentials may be divided into two categories: those which do contain an explicit hydrogen-bonding term in addition to van der Waals' and electrostatic terms [72,73,74,104,105,108,110,111], and those which do not [75,112,113]. The former often employ an exponential [100,101,114] or 12-10 [72,73,76,110] function to describe the interaction between the hydrogen-bonding atoms. The latter generally rely on the explicit representation of hydrogen atoms and sometimes lone pair electrons to account for the hydrogen-bond directionality. However, the lone pair electrons and hydrogen atoms are not normally observed in macromolecules studied by x-ray crystallography and their coordinates must therefore be estimated using standard known geometries.

In the GRID method [66,86], an 'extended' atom representation is used and an explicit hydrogen-bond term is incorporated in the energy potential. In this chapter, the formulation of a set of hydrogen-bond functions for use in the GRID method is described.

Table 3.1 Hydrogen-bond terms in some molecular mechanics energy potentials.

Distance dependence f(H-Y)	Energy potential	Angular dependence Θ=X-H-Y	Angular dependence Φ=H-Y-Z	Atom representation[a] UA/UAC/AA	Application	Notes
EXPONENTIAL						
Morse	MM2 [100]	-	-	AA	Small molecules	Angular function for bifurcated hydrogen-bonds only. Lone pairs on O and N
Morse	MM1 [101]	$\cos^m\Theta$	-	AA	Small molecules	
INVERSE POWER						
-6	Gibson [102]	$\cos^4 2\Theta$	-	UAC	Peptides	
12-6	LEV [103]	$f(\exp\Theta^2)$		UA	Proteins	
12-10	ECEPP [76]	-	-	AA	Proteins	
5-4	Momany [104]	-	-	AA	Carboxylic acids and alcohols	Also increase H charge and decrease O charge
12-10	MCMS [105]	-	-	AA	Proteins	
12-10	RKR [106]	-	-	AA	Peptide in solvent	
12-6	EMPWI [107]	-	-	AA	Peptide-water interactions	Explicit lone pairs
6-4 12-10	CHARMM [72,108,109]	$\cos^m\Theta$ m=0,2,4	$\cos^n\Phi$ n=0,2	UA/UAC/AA	Proteins, Nucleic Acids	Angular functions only for AA and UAC representation
12-10	AMBER [73,74]	-	-	UAC/AA	Proteins, Nucleic Acids	Well-depth =0.5 kcal/mol
12-10	YETI [110,111]	$\cos^2\Theta$	$\cos^4\Phi$ Z=lone pair	UAC	Protein-substrate docking	Geometry defined for each atom type

6-4	GRID0 [66]	$\cos^4\Theta$	-	UA	General intermolecular interactions	Hydrogen-bonding hydrogen coordinates determined in order to calculate Θ
8-6	GRID1	$\cos^m\Theta$ m=4-10	$f(\Phi)$	UA	"	m and $f(\Phi)$ chosen according to atom type
8-6	GRID2 [86]	$\cos^m\Theta$ m=2,4,6	$f(\Phi)$	UA/AA	"	"
GEOMETRIC						
1 r=2.5-3.1Å 0 otherwise	GREEN [65]	1 Θ=0-30° 0 otherwise	1 Φ=0-30° 0 otherwise Y=lone pair	AA	Proteins	All H-bonds=-3kcal/mol
NONE						
-	CCF [75]	-	-	AA	Proteins	
-	GROMOS [112]	-	-	UAC	Proteins Nucleic Acids	
-	EPEN [113]	-	-	UAC	Proteins	

[a] UA=United or extended atom representation.
UAC=United atom representation for carbons, all atom representation for other atoms.
AA=all atom representation.
X=Hydrogen-bond donor.
Y=Hydrogen-bond acceptor.
Z=Atom covalently bound to Y unless lone pairs are explicitly included.
In potentials with a UA representation, the hydrogen-bond distance dependence = f(X-Y) rather than f(H-Y).

3.2 THE HYDROGEN-BOND POTENTIAL Ehb

The hydrogen-bond potential is given by:

E_{hb} = Er x Et x Ep

It was formulated to give predicted structures which agreed with experimental data. This required that the hydrogen-bond energy was the product of three terms: Er, dependent on the separation of the hydrogen-bonding atoms, and Et and Ep dependent on the angle made by the hydrogen-bond at the target or probe atom respectively. Et and Ep take values between 0 and 1. They have different functional forms because there are different constraints acting on hydrogen-bonding atoms in the target and in the probe.

It should be noted that the hydrogen-bond potential was developed for use as part of the GRID energy function and was not designed to be transferable to other energy potentials.

The subroutines used to calculate the hydrogen-bond potential are listed in Appendix III.

3.3 THE DISTANCE DEPENDENCE Er OF THE HYDROGEN-BOND

For the distance dependence Er of the hydrogen-bond, a 12-10 function has been used in a number of other potential functions [72,73,76,110] and this was tested first. However, when used as a term in the GRID energy function, it proved to give too narrow a range of hydrogen-bond lengths when compared with the range of lengths observed experimentally [115]. A 6-4 function, which had also been used previously [72,108], was then tried [66], but this could produce hydrogen-bonds of such a great length that it would be possible for

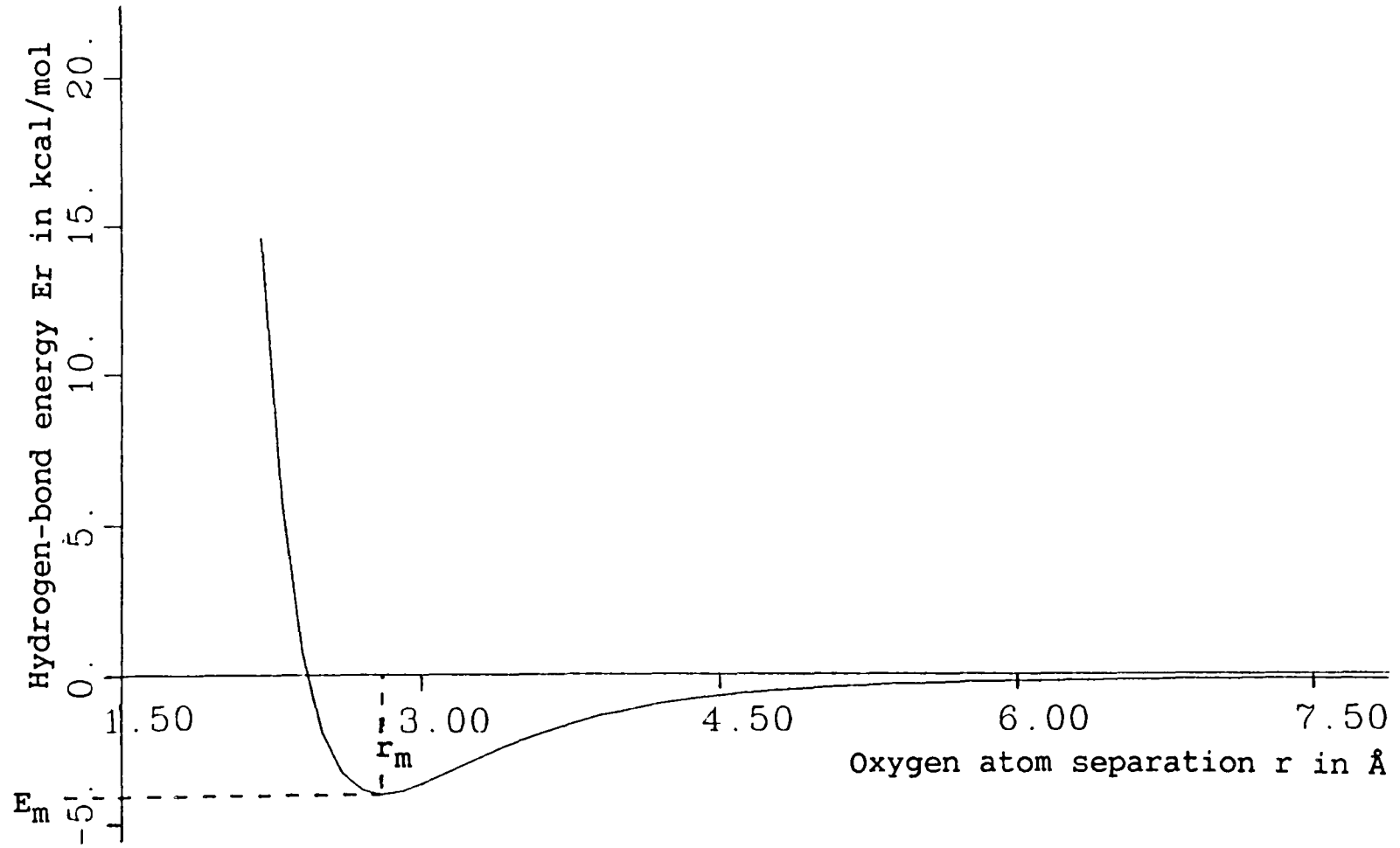


Figure 3.1 The variation of the hydrogen-bond energy Er, in kcal/mol, with the separation of the donor and acceptor atoms r, in Å. See text for details.

another non-hydrogen atom to position itself between the hydrogen-bonding atoms. Eventually an 8-6 function, as suggested by Reiher [116], was adopted [86] and found to give the most satisfactory results and the closest agreement with experimental observations. This is given by the following equation:

$$Er = C/r^8 - D/r^6$$

where $C = -3E_m r_m^8$ kcal Å^8/mol

$D = -4E_m r_m^6$ kcal Å^6/mol

where

r is the separation of the acceptor atom and the donor heavy atom in Å;

E_m is the optimum hydrogen-bond energy for the particular hydrogen-bonding atoms considered in kcal/mol;

r_m is the optimum hydrogen-bond length for the particular hydrogen-bonding atoms considered in Å (see Figure 3.1).

E_m and r_m vary according to the chemical type of the hydrogen-bonding atoms and were chosen in order to reproduce experimentally observed hydrogen-bond lengths and strengths.

A cutoff energy of -0.25 kcal/mol is imposed on long hydrogen-bonds in order to exclude large numbers of weak hydrogen-bonds from further consideration by program GRID.

For short hydrogen-bonds, E_{lj} for the interaction between the hydrogen-bond donor and acceptor atoms would be repulsive and so it is set equal to zero if a hydrogen-bond is formed.

3.4 THE DEPENDENCE Et OF THE HYDROGEN-BOND ON ITS ORIENTATION AT THE TARGET ATOM

3.4.1 Experimental data used for the formulation of Et

The function Et is dependent on the chemical nature of the hydrogen-bonding atoms in the target molecule. It was chosen by fitting to experimental data on hydrogen-bond geometries in crystalline structures observed by x-ray or neutron diffraction. The observed structures were of organic small molecules and proteins as shown in Table 3.2. Neutron studies tend to be slightly more accurate than those by x-ray because neutrons can easily detect hydrogen atoms. However, this difference is small when x-ray results have been 'normalised' to account for the fact that the length of bonds to hydrogen atoms is underestimated by 0.1-0.2Å [134,135]. Spectroscopic studies also give some information about hydrogen-bond geometries [136].

Data were collected from 'good' crystal structures. For small molecules, these can be defined [125,130] as structures in which the crystallographic R-factor is less than 10%, hydrogen coordinates are available, and there is no disorder. Therefore, only molecules which crystallise well were studied. As a result, the sample of hydrogen-bonds may not be representative of those formed between molecules in solution. The hydrogen-bonds in a crystal are likely to be shorter than in solution due to long-range attractive forces but this is unlikely to exert a detectable influence on bond angles [119].

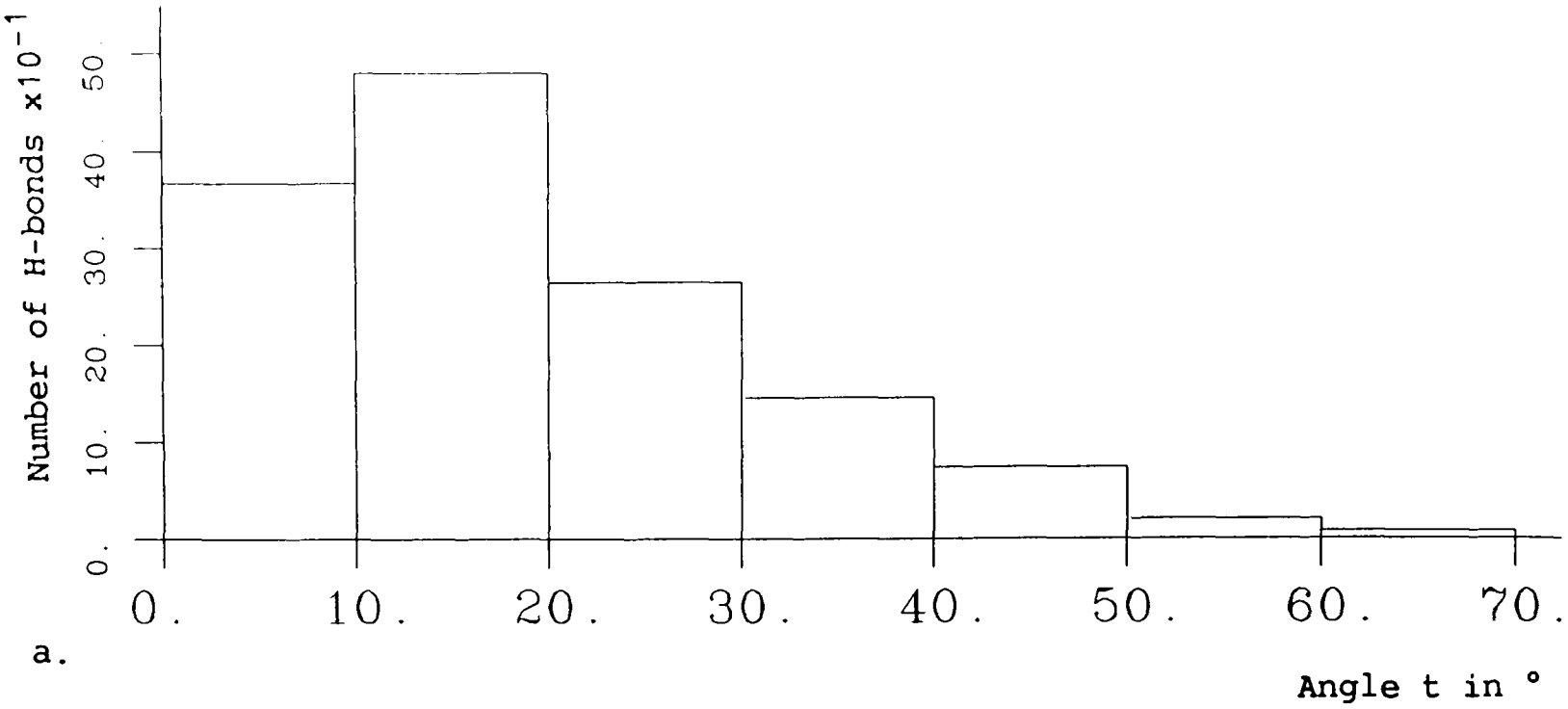
The hydrogen-bond function was designed to be applicable to general intermolecular interactions but not to intramolecular

Table 3.2 Summary of the experimental data on hydrogen-bonding which was used for modelling the angular dependence of hydrogen-bonds.

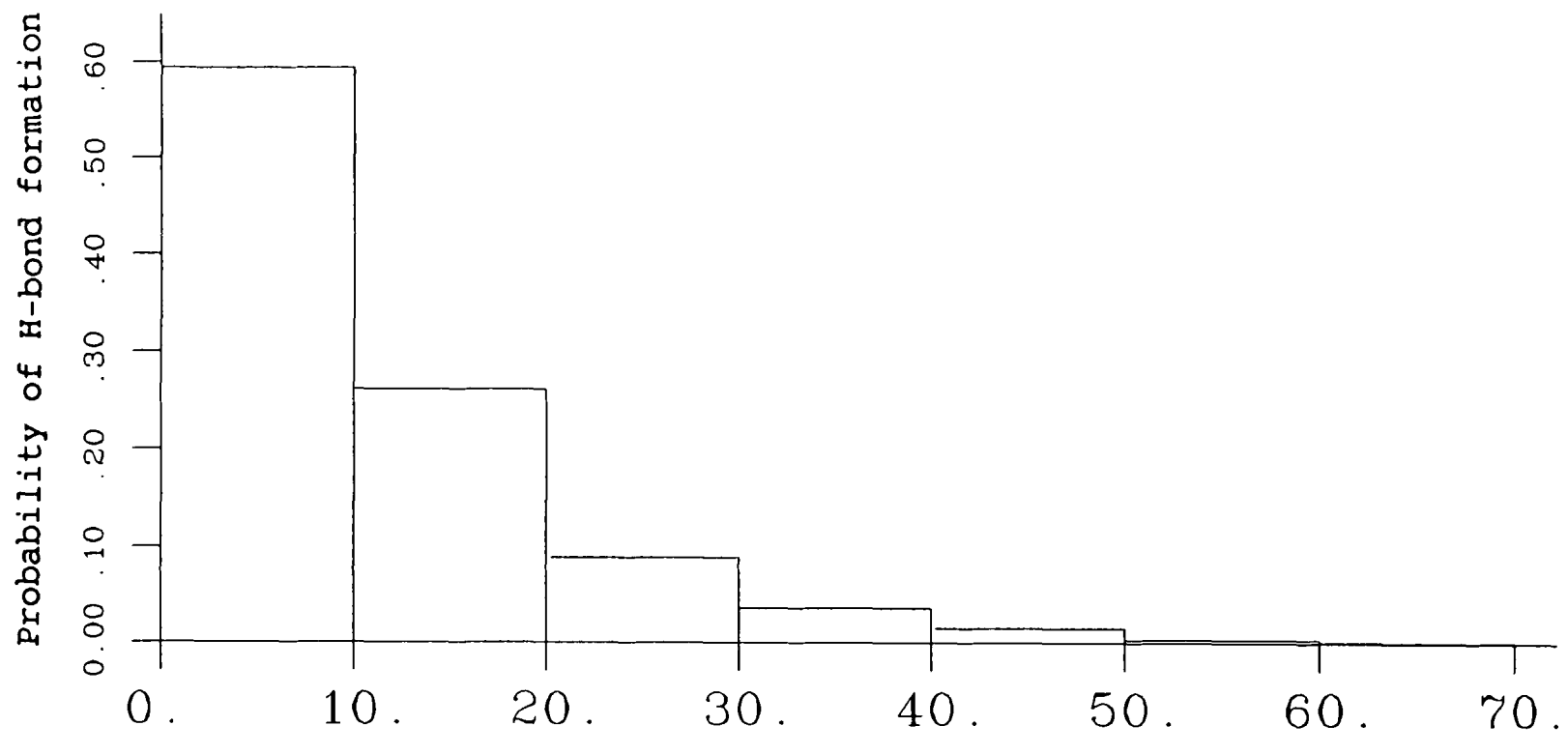
Reference Number	Publication Date	Method of Study N = neutron X = x-ray Diffraction	Number of Structures Studied	Total Number of Hydrogen-Bonds Studied	Hydrogen-Bond Type Studied	Composition of Crystals
[117]	1971	X	40	241	N-H--O	Amino-acid Peptide Polypeptide
[118]	1972	N	41	190	water H-O-H--A[a]	Organic Inorganic
[119]	1974	N	41	190	water H-O-H--A[a]	Organic Inorganic
[120]	1975	X	45	196	O-H--O	Saccharide Polyalcohol and related compounds
[121]	1977	X	90	356	O-H--O	Amino-acid Peptide Oligosaccharide
[122,123]	1979	N,X	95	439	N-H--O O-H--O N-H--X[b] O-H--X	Amino-acid, peptide and related molecules
[124]	1981	N	24	100	O-H--O	Organic Carbohydrate

[125]	1982	N	113	661	C-H--O C-H--N C-H--Cl	Organic
[126]	1982		32	168	N-H--O	Amino-acid
[127]	1983	N,X	58	277	O-H--O	Carbohydrate
[128]	1983	N,X	889	1509	N-H--O	Organic
[129]	1984	N,X	889	1509	N-H--O	Organic
[130]	1984	N,X	Whole Cambridge Structural Database[131]		N-H--O O-H--O	Organic
[132]	1984	X	15	Unknown	N-H--O O-H--O	Protein
[133]	1985	X	86	529	N-H--O O-H--O	Nucleoside Nucleotide

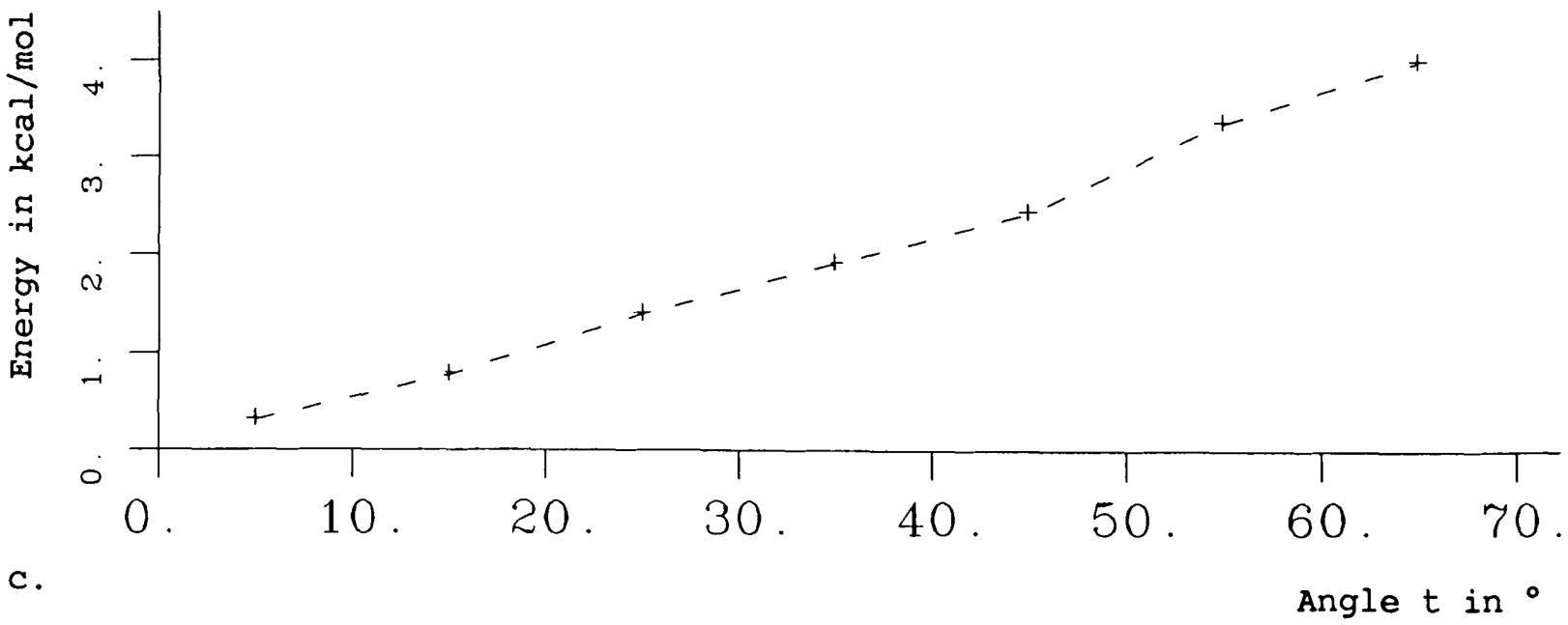
[a] A is any atom accepting an hydrogen-bond.
[b] X^- is a halide ion.



a.



b.



c.

Figure 3.2 a. The observed angular distribution of hydrogen-bonds donated by the N-H group [139]. b. The probability of hydrogen-bond donation by the N-H group determined from the observed distribution shown in a. c. The relative hydrogen-bond energy variation with deviation from bond linearity derived from the probability shown in b. See text for derivation.

interactions which may have a different spatial distribution. However, experimental observations of intra- and intermolecular hydrogen-bonds were both studied. In addition, some molecular systems show hydrogen-bonding patterns which are peculiar only to the particular molecule studied. For example, crystals of amino-acids contain a large number of three-centred hydrogen-bonds due to a proton deficiency and the presence of charged ammonium and carboxylate groups [137], while proteins have secondary structures, such as the α-helix, with characteristic hydrogen-bond orientations which are not typical of intermolecular hydrogen-bonds in general [115]. Therefore, due consideration of the source of the experimental data had to be taken when using the data as a basis for determining the terms occurring in the GRID energy function.

3.4.2 The method of derivation of Et

The observations in Table 3.2 were analyzed to give the number of hydrogen-bonds formed per 10° angular segment. This number was then converted into a probability by multiplying by a geometric factor [119,135,138], since 10° segments of angle t do not occupy the same solid angle. For example, for a target amide nitrogen, the correction factor is 1/(cost'-cost") where the 10° segment is between angles t' and t". This alters the apparent distribution of hydrogen-bonds as a function of angle t. For an N-H--O hydrogen-bond, the most frequently observed value of angle t is about 15° [139]. However, when the experimental data are converted into probabilities, the most probable conformation of the N-H--O hydrogen-bond is linear with angle t=0°. This is illustrated in Figure 3.2.

Table 3.3 The derived functions Et which define the hydrogen-bonding characteristics of target atoms.

Description of target atom	Number of hydrogen-bonds donated	accepted	Hybridisation	Variation of hydrogen/lone pair position	Angular function Et for an atom which donates[a]	accepts[a]	Angular range of Et
OXYGEN TYPES							
carbonyl, carboxy and phenolate O	0	2	sp^2	two alternative		$(0.9+0.1\sin 2t_i)\cos t_o$ $K_1(K_2-\cos^2 t_i)^3\cos t_o$[b] 0	$0<t_i<90°$ $90°<t_i<110°$ $t_i>110°$
phenolic hydroxyl O	1	1	sp^2	two alternative[c]	$\cos^4 t$	$(0.9+0.1\sin 2t_i)\cos t_o$ $K_1(K_2-\cos^2 t_i)^3\cos t_o$[b] 0	$0<t_i<90°$ $90°<t_i<110°$ $t_i>110°$
				fixed	$\cos^4 t$	$\cos^2 t$	
serine and threonine hydroxyl OH	1	2	sp^3	rotating[d]	$\cos^6 t$	$\cos^4 t$	
	1	2	sp^3	fixed	$\cos^6 t$	$\cos^2 t_o$ $\cos^2 t$	$t_i<55°$[e] $t_i>55°$
general case hydroxyl OH	1	2	sp^3	rotating	$\cos^4 t$	$\cos^2 t$	
	1	2	sp^3	fixed	$\cos^4 t$	$\cos^2 t_o$ $\cos^2 t$	$t_i<55°$[e] $t_i>55°$
ether O	0	2	sp^3	two alternative		$\cos^2 t_o$ $\cos^2 t$	$t_i<55°$[e] $t_i>55°$
water H_2O	2	2	sp^3	infinite	1	1	
				fixed	$\cos^4 t$	$\cos^2 t_o$ $\cos^2 t$	$t_i<55°$[e] $t_i>55°$
NITROGEN TYPES							
nitrogen N	0	1	sp, sp^2, sp^3	fixed		$\cos^2 t$	
nitrogen N	0	2	sp^2	two alternative		$(0.9+0.1\sin 2t_i)\cos t_o$ $K_1(K_2-\cos^2 t_i)^3\cos t_o$[b] 0	$0<t_i<90°$ $90°<t_i<110°$ $t_i>110°$

NH	1	0	sp, sp^2, sp^3	fixed	$\cos^2 t$		
NH	1	1	sp^3	two alternative	$\cos^2 t$	$\cos^2 t$	
histidine N	1[f]	1[f]	sp2	fixed	$\cos^2 t$	$\cos^2 t$	
NH	1	1	sp^2	two alternative[c]	$\cos^2 t$	$(0.9+0.1\sin 2t_i)\cos t_o$ $K_1(K_2-\cos^2 t_i)^3 \cos t_o$[b] 0	$0<t_i<90°$ $90°<t_i<110°$ $t_i>110°$
				fixed	$\cos^2 t$	$\cos^2 t$	
NH_2	2	0	sp^2, sp^3	two alternative	$\cos^2 t$		
amine NH_2	2	1	sp^3	rotating or fixed	$\cos^2 t$	$\cos^2 t$	
ammonium NH_3^+	3	0	sp^3	rotating or fixed	$\cos^2 t$		
Fluorine and Chlorine	0	3	sp^3	infinite	1		

[a] t is the angle made by the hydrogen-bond at the target atom as defined in Figure 2.2. t_i is the deviation of the probe hydrogen atom from the bisector of the lone pair orbitals within the plane of the lone pair orbitals. t_o is the deviation of the probe hydrogen atom from the plane of the lone pair orbitals. See Figure 3.3.

[b] This is a smoothing function with $K_1=0.9/\cos^6 110°$ and $K_2=\cos^2 110°$.

[c] The hydrogen atom is at 64° to the C-O or C-N bond in the plane of the ring rather than the ideal trigonal angle of 60° adopted by hydrogens on other sp^2 target groups and by all lone-pairs as experimental evidence suggests the suitability of this angle [132].

[d] The hydrogen atom is at 67° and the lone-pair is at 58° to the C-O bond rather than the ideal tetrahedral angle of 70° at which hydrogens and lone-pairs are assumed to rotate for other target groups. This difference is due to the particular angular dependence observed experimentally for this target atom [132].

[e] For sp^3 target atoms, the lone-pairs are assumed to lie at $t_i=55°$ and different hydrogen-bonding functions are adopted depending on whether the probe is between or outside the lone-pair orbitals.

[f] Both imidazole nitrogens are assumed able to donate or accept a hydrogen-bond although they can only make one hydrogen-bond at a time. The charge distribution of the histidine residue must be adjusted according to which nitrogen the hydrogen is assumed to be attached to.

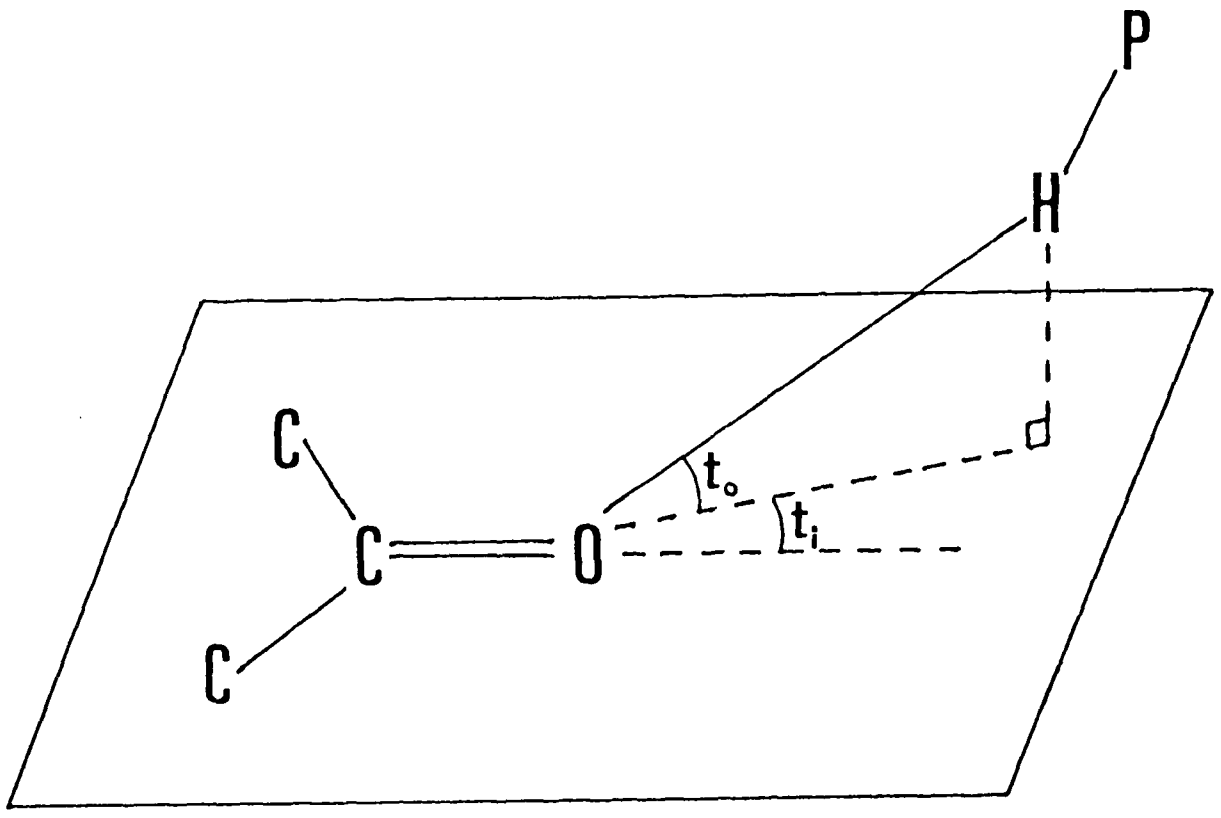


Figure 3.3 The geometry of the target carbonyl group. Angle t_o is the angle subtended at the oxygen atom by the probe's hydrogen atom and the plane of the lone pair orbitals. Angle t_i is the angle subtended by the probe hydrogen atom and the C-O axis within the plane of the lone pair orbitals.

The probabilities were then converted into relative energies assuming a Boltzmann distribution:

$$E_i = -RT(\ln\{P_i\}-\ln\{P_0\}) + E_0$$

where P_i is the probability of bond formation at energy E_i and P_0 is the probability of bond formation at reference energy E_0. These energies were plotted against angle t and a suitable analytical function was chosen to fit them. The function was required to be continuous and of a simple form.

3.4.3 The function Et

The orientational dependence Et of the hydrogen-bond energy varies according to the chemical nature of the target atom. This can be seen in Table 3.3 which lists the functions used to describe the hydrogen-bonds made by different types of target atom. These functions are now described in detail for specific cases.

3.4.3.1 Oxygen

3.4.3.1.1 Carbonyl and carboxyl oxygen atoms

A carbonyl or carboxyl oxygen atom can accept one or two hydrogen-bonds. A definite tendency for these to be formed in the direction of the lone pairs at ±60° to the C-O bond and in the lone pair plane has been reported [117,121,135,139]. The orientation of a hydrogen-bond accepted by a carbonyl or carboxyl oxygen atom is described in terms of two angles: t_o, which is the angle subtended at the target oxygen atom by the hydrogen atom which is donated by the probe, and the lone pair plane of the target oxygen; and t_i, which is

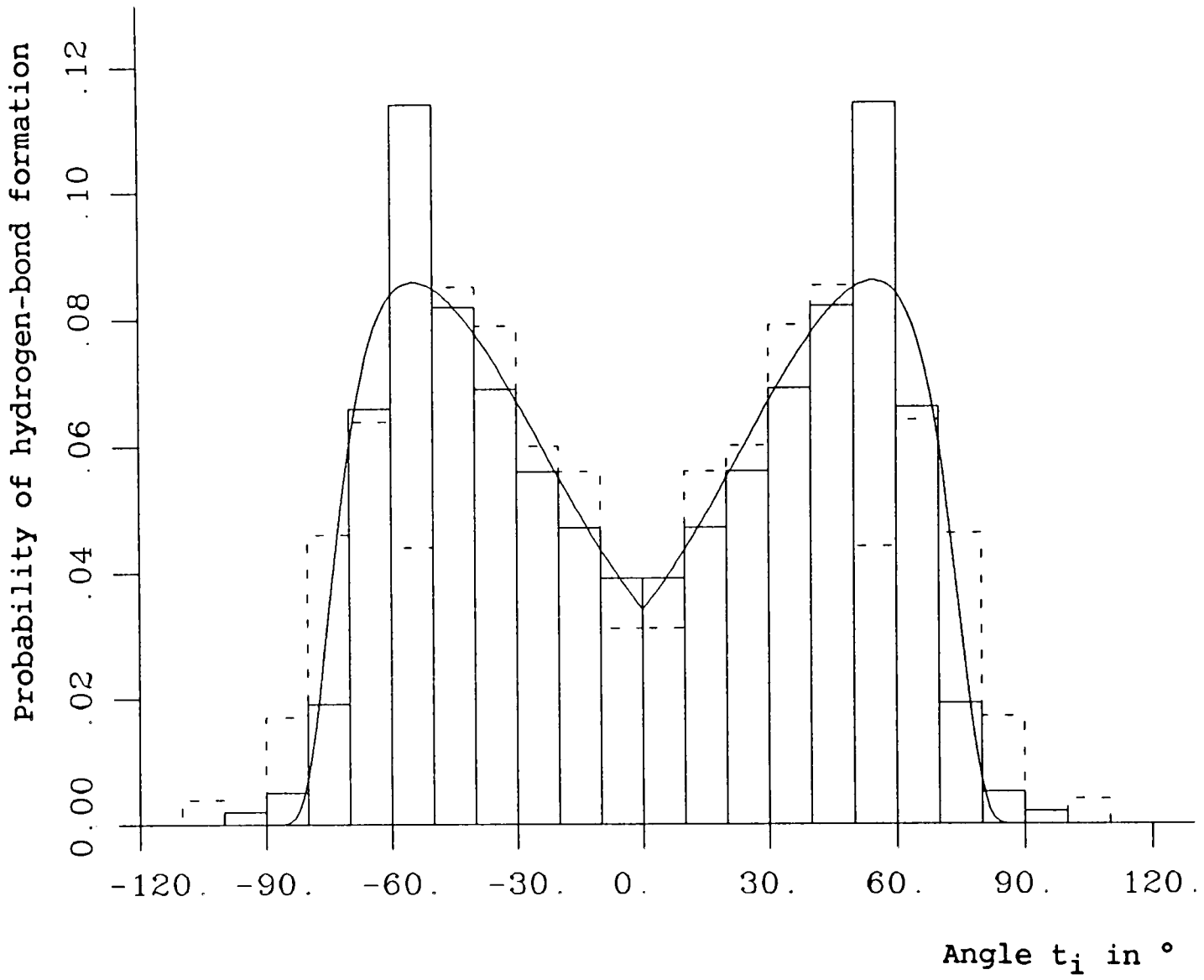


Figure 3.4 The two histograms show the observed probabilities of a hydrogen-bond to a carbonyl oxygen atom derived from the experimental data which was used for modelling the hydrogen-bonding function, (dashed line - data from reference [117], solid line - data from reference [139]). The variation between the histograms is due to the different experimental samples studied. The graph shows the modelled variation of the probability of hydrogen-bond formation with angle t_i derived from the total interaction energy calculated for a target carbonyl oxygen atom interacting with a probe water molecule. The water molecule lies in the lone pair plane, i.e. t_o=0°, at a distance of 2.8Å from the oxygen.

the angle of deviation within the oxygen lone pair plane of the probe hydrogen atom from the direction of the C-O bond. This is illustrated in Figure 3.3. In observed hydrogen-bonds, angle t_o tends to be less than 40° although larger angles are not unusual [134,138]. Within the lone pair plane, hydrogen-bonds along the directions of the lone pair orbitals are favoured and more hydrogen-bonds occur between the orbitals than outside them as can be seen from the histograms in Figure 3.4. The directionality is more marked if two hydrogen-bonds are accepted than if only one hydrogen-bond is accepted, when both lone pairs tend to attract the single donor atom to their bisector, suggesting that both electrostatic and steric factors influence the hydrogen-bond geometry. There is no correlation between angle t_o and angle t_i.

The chosen hydrogen-bonding term Et for a carbonyl or carboxyl oxygen atom [140] has maxima at $t_i = \pm 45°$ and this causes energy E_{hb} to be most attractive at this angle. However, when the Lennard-Jones interactions of the probe with the atoms covalently bonded to the target oxygen are also taken into account, the probability of hydrogen-bond formation, (derived from the total energy calculated by program GRID), is found to be highest in the lone pair orbital directions at $t_i = \pm 60°$ as shown by the graph in Figure 3.4. The functional form of Et causes a variation in E_{hb} of about 10% over the range $t_i = 0\text{-}90°$ and this is not incompatible with molecular orbital calculations [98] for a formamide dimer where the variation in hydrogen-bond energy is about 15% over the range $t_i = 0\text{-}80°$. When angle $t_i > 80°$, the molecular orbital calculations show that the total interaction energy rises rapidly as exchange repulsion effects increase.

3.4.3.1.2 <u>The phenolic hydroxyl group</u>

In proteins, the tyrosine residue has been observed to donate hydrogen-bonds more often than it accepts them [132]. It can be considered able to donate one hydrogen-bond and to accept only one hydrogen-bond, because there is a significant back donation of the oxygen lone pairs which gives partial double bond character to the C-O bond restricting rotational motion around it. The hydroxyl oxygen resonates between sp^2 and sp^3 hybridizations and can thus be considered to have its hydrogen atom and one lone pair in the plane of the aromatic ring at ±60° to the C-O bond direction, thus accounting for the observed hydrogen-bonding geometry. In calculating hydrogen-bond energies, the hydrogen atom is allowed to occupy the energetically most favourable of the two positions available to it, according to the position and chemical characteristics of the interacting probe. Similarly, the lone pair orbital always takes up the most favourable of the two possible positions available to it. Hence, account is taken of the influence of the incoming probe on the conformation of the phenolic hydroxyl group.

The experimental data studied were limited to that on 150 hydrogen-bonds to tyrosine residues from a survey of fifteen protein crystal structures [132]. The geometries are described in terms of one angle only, the deviation from the C-O bond direction, and there are no data for when this angle is greater than 90°. Data are available as a plot of the variation of hydrogen-bond occurrence with angle t, the deviation from the O-H direction, for when tyrosine donates. $Et=\cos^4 t$ was found to be the most suitable function fitting this plot.

The experimental data on hydrogen-bonds accepted by tyrosine are more limited because tyrosine is found to accept only

about 20% of its experimentally observed hydrogen-bonds [132]. However, where it is observed, tyrosine accepts hydrogen-bonds preferentially at 59±16° to the C-O bond. The same angular function is used for accepted hydrogen-bonds as for a carbonyl oxygen atom because the available experimental data on tyrosine are inadequate for formulating a separate function specifically for this group. If better data on observations become available in the future, it may be possible to assign a more suitable function for this hydrogen-bond.

An anionic phenolate oxygen is treated in the same manner as a phenolic hydroxyl group accepting a hydrogen-bond. It may accept up to two hydrogen-bonds but is unable to donate.

3.4.3.1.3 The sp^3 hybridized hydroxyl groups

Crystallographic data show that hydrogen-bonds accepted by the sp^3 hydroxyl group are strongly constrained to the group's lone pair plane which can be defined if the position of the hydroxyl hydrogen is known. This effect is more marked than the corresponding constraint for hydrogen-bonds to form in the lone pair plane of carbonyl oxygen atoms. The angle t_o for a sp^3 hydroxyl group is generally less than 20° [138,139] compared to less than 40° for a carbonyl oxygen atom as mentioned above. Within the lone pair plane, hydrogen-bonds occur more often between the lone pair orbitals than outside them. They have a fairly uniform distribution over the range $-60° < t_i < 60°$ if one hydrogen-bond is formed, but are more strongly constrained to the lone pair directions if two hydrogen-bonds are accepted [124,137].

However, the observed lone pair directionality does vary according to the method of study. Microwave spectroscopy of

hydrogen-bonded dimers [136] shows an energetic preference for hydrogen-bonds to occur in the lone pair directions for sp^3 hybridized oxygens although this is smaller than for sp^2 hybridized oxygens. In charge density studies [12], there is some small deformation in electron density towards lone pairs for sp^2 hybridized oxygens whereas for sp^3 hybridized hydroxyl groups and water, the lone pairs appear to exist as one broad peak in the electron density.

A large amount of crystallographic data is available on the geometry of hydrogen-bonds to hydroxyl groups in proteins. The protein hydroxyl groups in serine and threonine are therefore described by angular functions which differ somewhat from those for general hydroxyl groups. The protein functions define hydrogen-bonds which are more constrained to the lone pair or hydrogen directions than the functions for general hydroxyl groups which make allowance for the incompleteness of the data from which the function for a general hydroxyl group was chosen.

In proteins, hydroxyl groups are more likely to donate a hydrogen-bond than accept one. In a survey of fifteen protein crystals [132], serine was observed to donate 67% of its hydrogen-bonds and threonine 75% of its hydrogen-bonds. Hydrogen-bonds in proteins were found to be accepted at 58°±35° to the C-O bond and donated at 67°±35°. It appears that the donated hydrogen-bonds are closer to the tetrahedral angle because of the greater directional control of the hydrogen atom when compared to the lone pair orbital.

In the calculation of hydrogen-bond energy, the hydrogen atom and the lone pairs of a hydroxyl group are normally allowed to rotate at the tetrahedral angle around the projected C-O axis. The hydrogen atom and lone pairs are assumed to position themselves in the

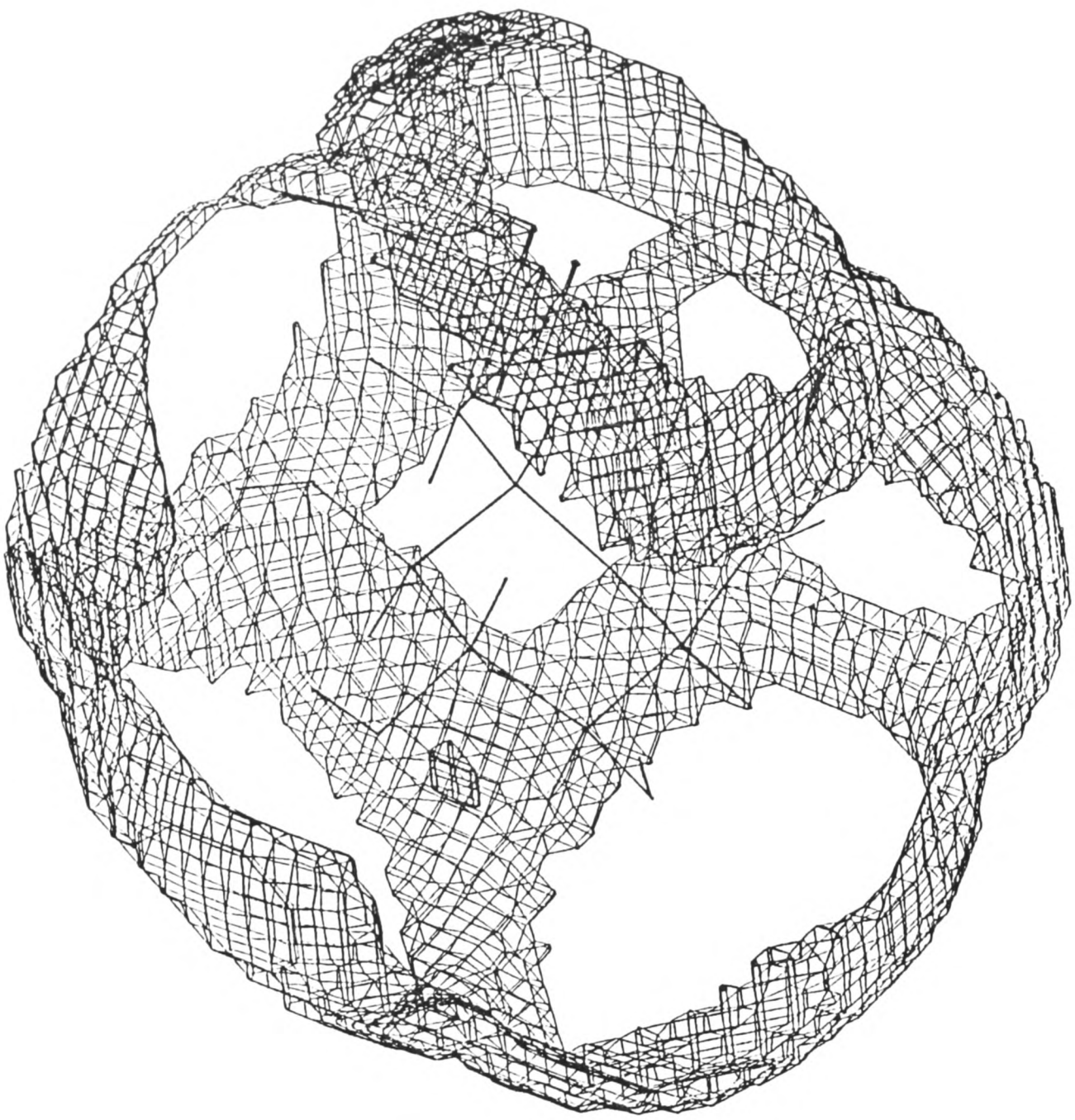


Figure 3.5 Energy contours at -3.2 kcal/mol for the interaction between an amide nitrogen probe able to donate one hydrogen-bond, and a glucose molecule. The hydroxyl hydrogens and lone pairs are allowed to rotate around the C-O axis giving rise to circular regions of favourable energy where a strong linear hydrogen-bond can be made.

most energetically favourable positions according to the nature and position of the probe group. Angle t is then the deviation from this direction when the hydroxyl group donates. In Figure 3.5, the attractive energy contours obtained for the interaction of an amide nitrogen probe with glucose can be seen. The toroidal energy contours show the energetically favourable positions of the nitrogen following the circular locus of the hydroxyl lone pairs as it donates a linear hydrogen-bond to the glucose molecule.

In some circumstances, such as if the hydroxyl group is already making an intramolecular hydrogen-bond to another target atom, the explicit position of the hydrogen atom may be known and can therefore be fixed. The hydrogen-bonds of the hydroxyl group are then calculated differently with donated hydrogen-bonds favoured in the predefined O-H direction, while accepted hydrogen-bonds are favoured in the lone pair plane where they are treated similarly to the hydrogen-bonds of an ether oxygen atom (see below).

3.4.3.1.4 The ether oxygen atom

This is sp^3 hybridized and is able to accept two hydrogen-bonds. Its two lone pairs are in the plane perpendicular to the C-O-C plane and symmetrically placed with respect to it. The hydrogen-bonds that are formed show a definite tendency to be in this lone pair plane but have an approximately constant probability of occurrence between the lone pair orbitals [130]. Unfortunately, there is very little experimental data on the hydrogen-bond geometry of the ether oxygen atom so a relatively weak angular dependence was chosen to fit this data.

3.4.3.1.5 **Water**

This can donate two and accept two hydrogen-bonds. Usually, the coordinates of the hydrogen atoms are unknown so the water is assumed to be able to rotate freely and no angular factor is applied to its hydrogen-bonds. However, sometimes hydrogen atom coordinates are known. These can then be fixed and the same angular constraints applied as for a general target sp^3 hydroxyl group with its hydrogen atom coordinates fixed.

3.4.3.2 **Nitrogen**

Target nitrogen atoms may be able to donate up to three hydrogen-bonds and may be able to accept one or two hydrogen-bonds. Linear hydrogen-bonds are favoured [132,134,139] especially for shorter bonds because as the hydrogen-bond becomes shorter, the non-bonded interaction becomes less favourable and so the hydrogen-bond straightens out and thereby increases the distance between the heavy atoms. An $Et=\cos^2 t$ function centred on the N-H direction was adopted for all hydrogen-bonds donated by these target nitrogen atoms.

For accepted hydrogen-bonds, the same angular functions are applied as for oxygen atoms of the same geometry. However, E_m and r_m are different for the two types of atom and so the resultant hydrogen-bond energy for a target nitrogen atom differs from that for a target oxygen atom. The movement of the nitrogen's hydrogen atoms and lone pairs is also allowed for in the same way as for oxygen. For example, the three hydrogens in an ammonium group may rotate about the C-N bond, or may be fixed at specified coordinates.

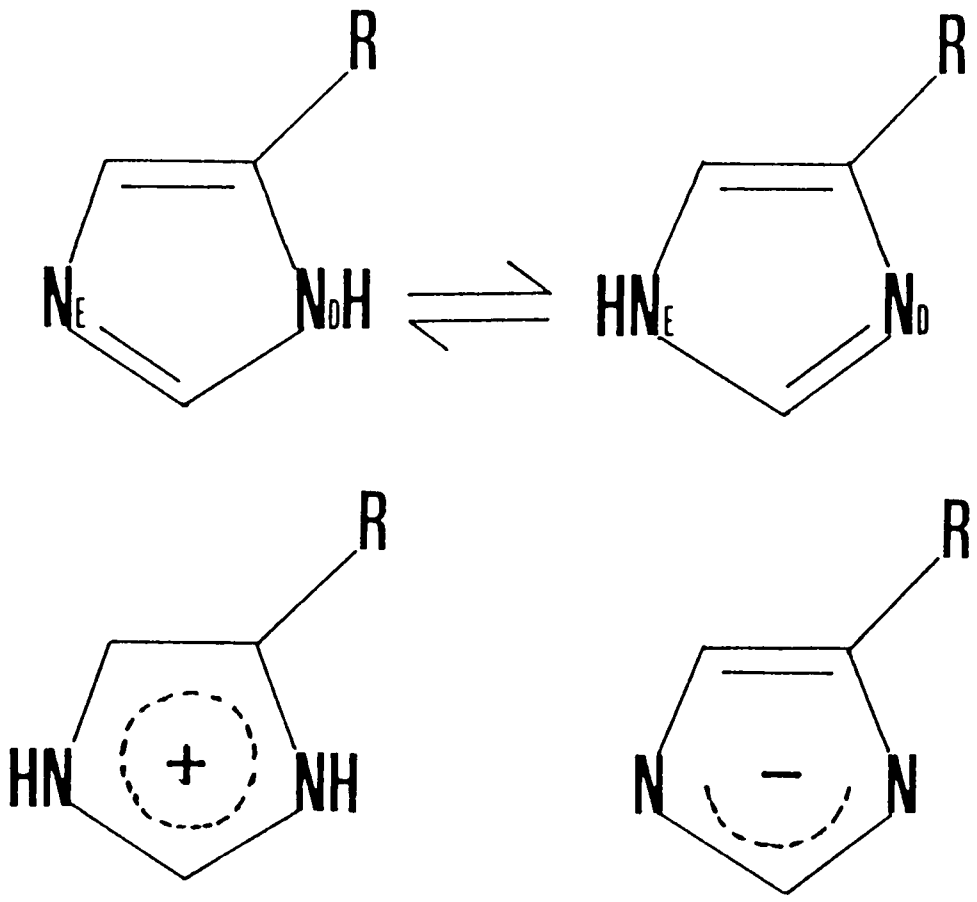


Figure 3.6 The two tautomers, and the protonated and anionic forms of the histidine side chain.

3.4.3.2.1 **Histidine nitrogen**

The side chain of histidine can exist in four forms: two neutral, one protonated and one anionic as shown in Figure 3.6. The protonated form normally exists in significant concentrations only at low pH while the anionic form is generally found only at high pH. According to experiment [141,142,143,144] and theory [145,146], the tautomer with NE protonated should predominate over that with ND protonated. This is because the R-group is electron withdrawing and the effect of the R-group is more marked for ND than NE [147,148].

The Protein Data Bank [52] files of thirty proteins were searched [149] to assess the position of tautomeric equilibrium by examining the atoms neighbouring the histidine residues. However, it was not possible to detect a significant preference for one or other of the tautomeric forms. The hydrogen-bonding function for histidine therefore had to be devised [149] to allow for the existence of both tautomers at physiological pH, and to select the tautomer which gives the most favourable energetic interaction with the probe. It does this by allowing both nitrogens of the histidine residue to donate or accept a hydrogen-bond, and making the way in which the hydrogen-bond is actually formed depend on the nature and position of the probe group. If the probe group cannot make a hydrogen-bond to either of these nitrogens, the hydrogen is assumed to be attached to NE. If a hydrogen-bond is donated by NE or accepted by ND, the electrostatic charge of the histidine residue is unchanged. However, if a hydrogen-bond is donated by ND or accepted by NE, there is an alteration in the electronic charge distribution of the histidine residue and therefore, a compensating term must be applied to the electrostatic energy calculated. This allows not only the hydrogen

position, as for a phenolic hydroxyl group, but also the electronic distribution of the histidine residue to be influenced appropriately by the nature of the probe group.

For calculations at high pH, both histidine nitrogens are assumed to be unprotonated while at low pH they are both assumed to be protonated. The histidine charge is adjusted accordingly but its distribution at these pHs is assumed to be independent of the nature of the probe group.

There are insufficient observations of histidine unequivocally accepting a hydrogen-bond from which to construct a function Et. However, as the histidine nitrogen is sp^2 hybridized, it was assumed that it had the same characteristics as a carbonyl oxygen which is accepting two hydrogen-bonds simultaneously, although only one hydrogen-bond can be accepted by the histidine nitrogen. For carbonyl oxygen atoms with their full hydrogen-bonding capacity satisfied, most hydrogen-bonds are observed [117,121] within 20° of the lone pair orbital directions and their hydrogen-bond distribution can be modelled using the angular function $Et=\cos^2 t$. Nitrogen is less electronegative than oxygen so the lone pair cloud is likely to be more diffuse. However, the same angular function is used for the histidine nitrogen when it accepts a hydrogen-bond, although E_m and r_m in the distance dependent function Er are different.

The geometry of hydrogen-bonds donated by histidine nitrogens has been observed [132]. This data and general data on N-H--O bonds were fitted by the function $Et=\cos^2 t$ centred on the direction of the N-H bond.

3.4.3.3 Halogens

Fluorine and chlorine may be able to accept up to three hydrogen-bonds [150] but as insufficient data were available on their geometry, no angular function was adopted. The hydrogen-bonds to chloride ions have been observed to be generally longer than those to other acceptor atoms suggesting that they are weak [122].

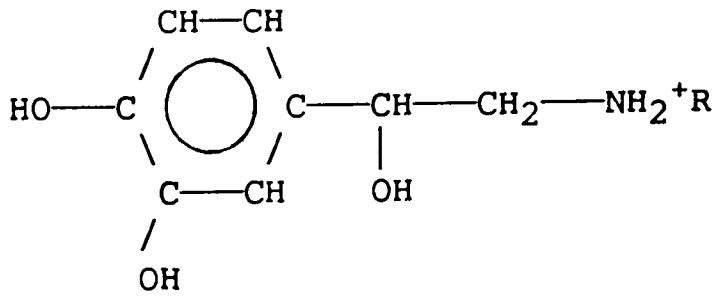
3.4.3.4 Other atoms

Hydrogen-bonds to carbon atoms have been found to be significant in small organic molecules containing nitrogen and in proteins where a carbon atom is adjacent to a protonated nitrogen [125]. Hydrogen-bonds have also been observed to be accepted by aromatic rings perpendicular to the ring plane [151]. Hydrogen-bonds accepted by sulphur atoms in the directions of the lone pair orbitals have also been reported [73,132,152]. However, hydrogen-bonds to all these atoms are not calculated by the present version of program GRID.

3.4.4 Examples of the application of the hydrogen-bond functions Er and Et

3.4.4.1 The catecholamines

The catecholamines act on a number of different receptors producing a variety of effects which are of great pharmacological interest. They are described here both because of their pharmacological importance and because they demonstrate the



I R = -H Noradrenaline

II R = $-CH_3$ Adrenaline

properties of the hydrogen-bond functions used for target hydroxyl groups and nitrogen atoms in program GRID.

The interaction energy of an uncharged nitrogen NH probe, which is able to donate one hydrogen-bond, with L-noradrenaline (I) in a crystallographically observed conformation [153] was calculated. The energy contours obtained at -3 kcal/mol are shown in Figure 3.7.a. The three small contours show where a hydrogen-bond from the NH probe can be donated to the phenolic hydroxyl groups. This hydrogen-bond is favoured in the plane of the aromatic ring at the trigonal angle to the phenolic C-O bonds. The phenolic hydrogens are shown in their crystallographically observed positions, but program GRID allows them and the hydroxyl oxygen lone pairs to exchange positions in the plane of the aromatic ring so that they can make the most energetically favourable interaction with the nitrogen NH probe. Thus, the target molecule is able to adapt to changes in its local environment caused by the approach of the probe group. The presence of both phenolic hydroxyl groups is necessary for full ß-agonist activity [154], so these energy contours may indicate regions where interaction with the ß-adrenergic receptor occurs.

The dominating circular region shows where a linear hydrogen-bond can be made to the alkyl-hydroxyl group as its lone pairs rotate at the tetrahedral angle around the C-O bond. The ring of contours is thickest near the benzene ring and the alkyl-side chain, where the attractive Lennard-Jones and hydrogen-bond interactions of the probe summate. This hydroxyl group plays a very important role in determining the activity and the binding affinity of the ß-adrenergic agonists and antagonists [154]. If it is absent or in a different stereochemical configuration, potency is generally reduced.

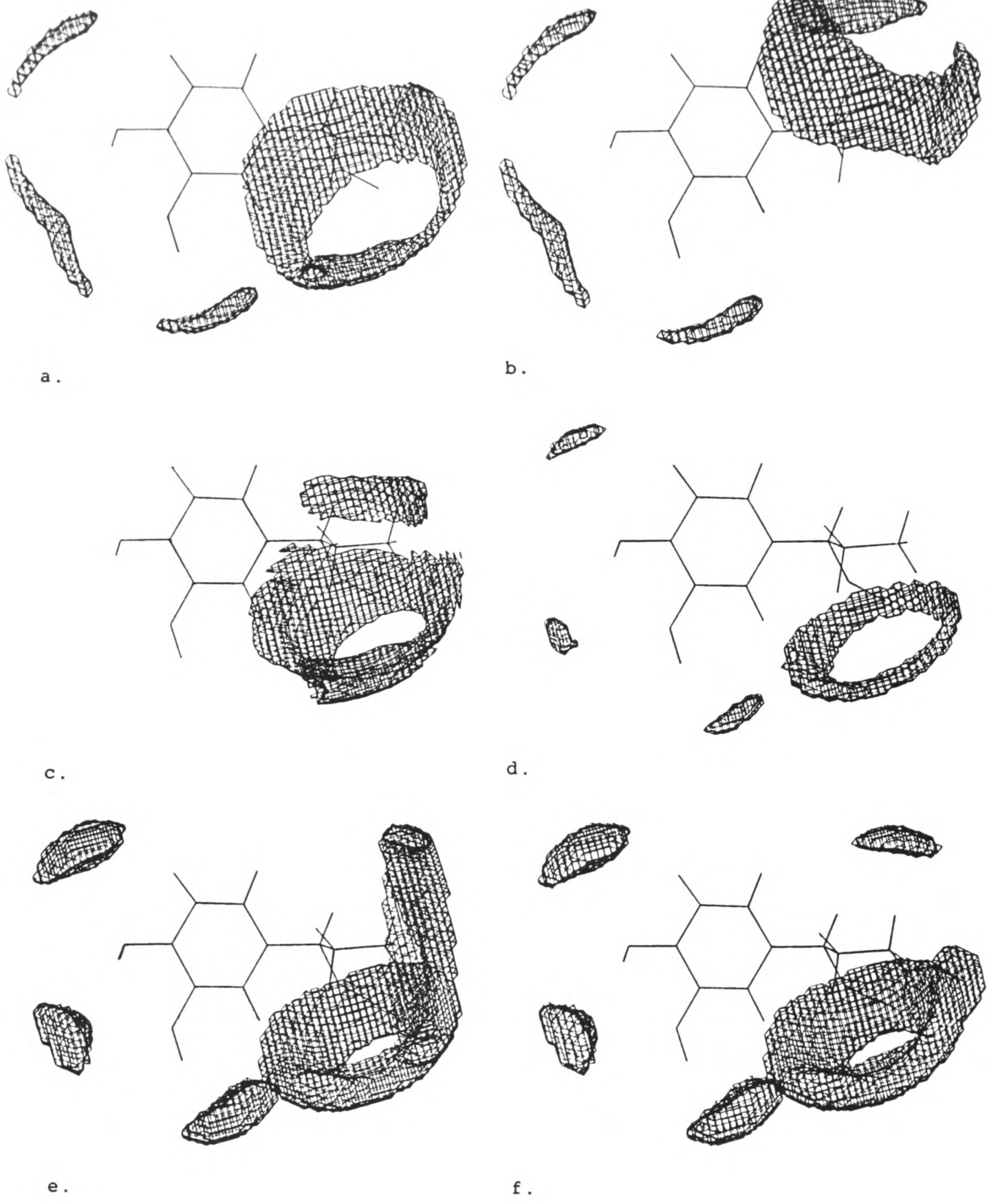


Figure 3.7 a. Energy contours at -3 kcal/mol for a nitrogen NH probe interacting with L-Noradrenaline. b. Energy contours at -3 kcal/mol for a nitrogen NH probe interacting with D-Noradrenaline. c. Selectivity map with energy contours at -2 kcal/mol showing where the interaction of a nitrogen NH probe is more energetically favourable for L-Noradrenaline than for D-Noradrenaline. d. Energy contours at -3 kcal/mol for a nitrogen probe which can accept one hydrogen-bond, interacting with L-Noradrenaline. e. Energy contours at -2 kcal/mol for a nitrogen probe which can accept one hydrogen-bond, interacting with L-Noradrenaline. f. Energy contours at -2 kcal/mol for a nitrogen probe which can accept one hydrogen-bond, interacting with L-Adrenaline. See text.

Figure 3.7.b shows D-noradrenaline arranged in the same overall conformation as L-noradrenaline, so that the catechol and nitrogen moieties of both molecules are in the same spatial relationship to each other. The energy contours at -3 kcal/mol for the same probe group do not form a full ring around the alkyl-hydroxyl group since there is now a repulsive steric interaction between the NH probe and the amine group of noradrenaline, which is closer to the alkyl-hydroxyl group in this conformation. The interaction of the two isomers can be compared in a selectivity map as shown in Figure 3.7.c. This is constructed by calculating, as described in Appendix III, the difference in interaction energy of the probe with the L-isomer and the D-isomer at each point in the grid surrounding noradrenaline, and contouring the difference energies at -2 kcal/mol. The two contoured regions in Figure 3.7.c show where the probe interacts at least 2 kcal/mol more favourably with the L-isomer than the D-isomer. The uncontoured channel between the two regions shows where the two circular rings seen in Figures 3.7.a and 3.7.b overlap and mutually cancel each other out in the difference map. The volume excluded from the lower left-hand side of the main circular difference contour is a region where a hydrogen-bond is made to the alkyl-hydroxyl group in the L-isomer, but in the D-isomer in this conformation a hydrogen-bond cannot be made to the alkyl-hydroxyl group, and is instead made to the adjacent phenolic oxygen, thus diminishing the effect on the binding energy of moving the alkyl-hydroxyl group. Such details are not immediately obvious when the conformations of the molecules are compared, but stand out clearly in appropriately contoured maps calculated by program GRID.

Figure 3.7.d shows the interaction at -3 kcal/mol with L-noradrenaline of an uncharged nitrogen probe which has one lone pair and is therefore able to accept one hydrogen-bond. The small contoured regions show where the phenolic oxygens tend to donate a hydrogen-bond in the plane of the aromatic ring at ±60° to the C-O bond. The phenolic hydrogens are allowed to adopt the position in the plane of the aromatic ring which gives the most energetically favourable interaction with the nitrogen probe. The contours at the phenolic oxygen are of a different shape to those in Figure 3.7.a because they were calculated using an angular dependence of a different functional form to that used when the phenolic oxygen accepts a hydrogen-bond.

As before, the hydrogen of the alkyl-hydroxyl group on the target is able to rotate at the tetrahedral angle to the C-O bond giving rise to the circular contour. The energy contours are at the same energy level as in Figure 3.7.a but the regions enclosed by the contours are seen to be smaller. This is because the hydroxyl hydrogen-bonding hydrogens of the target molecule have more influence on the direction of hydrogen-bonds than the hydroxyl lone pairs do. A hydrogen-bond is thus constrained to occupy a smaller range of angles when the target atom is donating.

Figure 3.7.e again shows the interaction with L-noradrenaline of an uncharged nitrogen N probe able to accept a hydrogen-bond, but this time the map is contoured at -2 kcal/mol. This naturally makes all the original contours around the target oxygen atoms appear larger, but in addition to these contours there is an extra circular contour showing hydrogen-bonding to the amine group which is assumed to exist in its ionised form as NH_3^+ at physiological

pH [155]. These contours are circular because the three hydrogens of the charged amine group may rotate at the tetrahedral angle around the C-N bond. The contours are seen in this figure but not in Figure 3.7.d, because the hydrogen-bond to a nitrogen atom is weaker than that to an oxygen atom and does not show when a contour level of -3 kcal/mol is used. This demonstrates that it is important to choose a suitable contour level to display the interaction energies calculated by program GRID.

In Figure 3.7.f, the interaction with L-adrenaline (II) of an uncharged nitrogen able to accept one hydrogen-bond is shown contoured at -2 kcal/mol. Adrenaline differs from noradrenaline in having a methyl group substituted at its nitrogen, and its conformation was altered from that observed in the adrenaline crystal [156] in order to superimpose it on the previous noradrenaline conformations. The catecholamine nitrogen is again treated as being positively charged, and has two hydrogens whose locations are now assumed to be explicitly determined by the two bonds from the nitrogen to two carbon atoms. This causes the energy contours around the target nitrogen atom to form two discrete regions rather than a continuous ring: one of these regions can be distinctly seen; the other is adjoining the circular contours around the alkyl-hydroxyl group.

These energy maps of noradrenaline and adrenaline demonstrate the way in which a small change in the structure or conformation of a molecule can have a large impact on the molecule's capacity to interact favourably with a ligand. This is due to the specificity of the hydrogen-bonds formed, which have a geometry which is sensitively dependent on the chemical nature and orientation of the hydrogen-bonding atoms. The GRID contours show both the overall

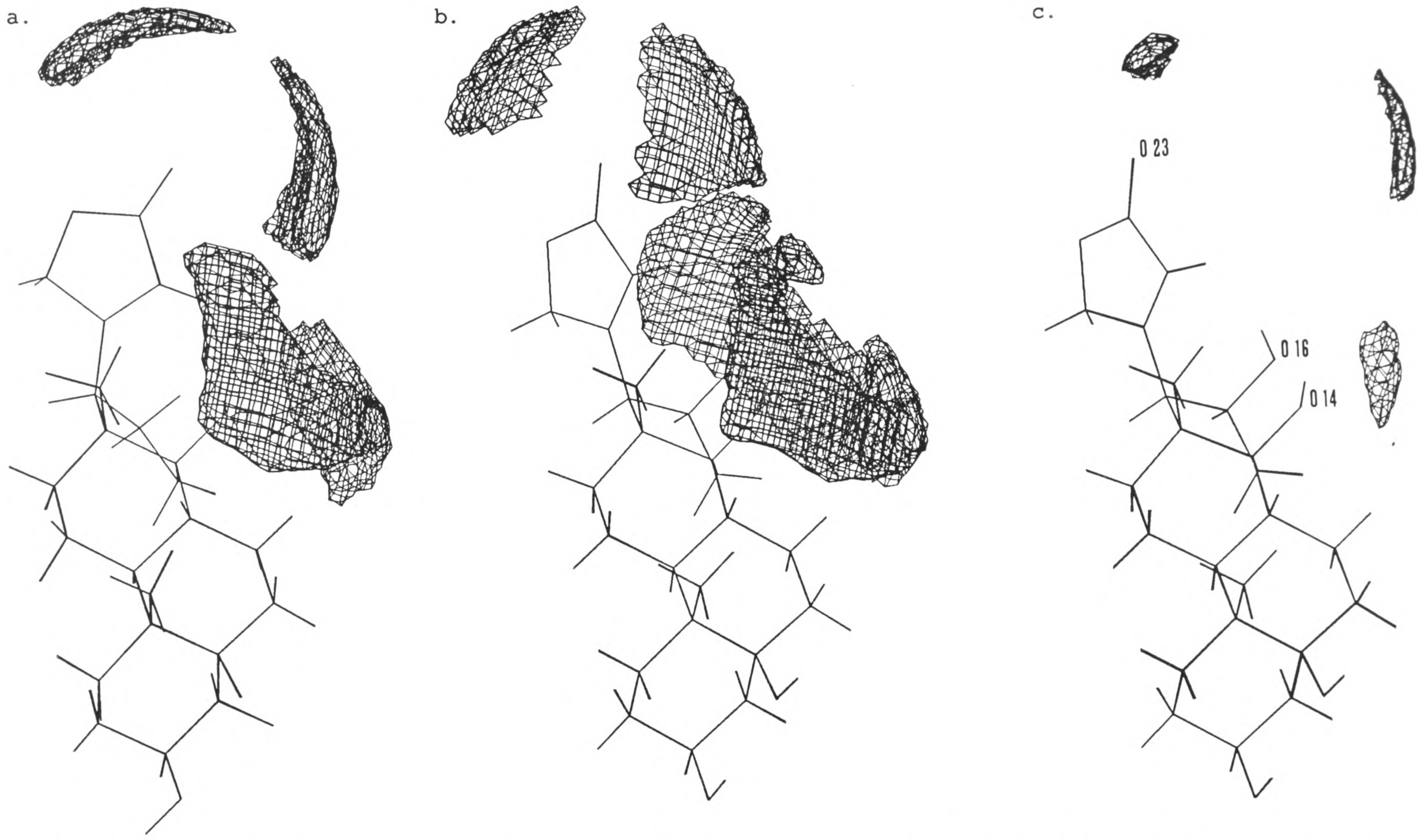


Figure 3.8 Energy contours at -3 kcal/mol for an amide nitrogen NH probe interacting with a. digitoxigenin and b. gitoxigenin. Figure 3.8.c shows the gitoxigenin molecule with a selectivity map with difference energy contours at -3 kcal/mol for an amide nitrogen NH probe showing regions where binding to digitoxigenin is favoured by comparison to binding to gitoxigenin. For clarity, only energy contours near the lactone ring region of the molecules are shown. See text.

character of the interactions and their fine detail in a clear and meaningful way.

3.4.4.2 Cardiac glycosides

The cardiac glycosides are used in the treatment of heart failure as they cause positive inotropic action. Their activity varies according to their structure and alternative hypotheses have been proposed regarding their mode of binding. These suggest that binding occurs either by interactions with specific sites [158], or via a general charge interaction with the molecular dipole of the steroid [157], or by means of hydrophobic interactions with the steroidal nucleus [159]. Two compounds, digitoxigenin and gitoxigenin, consisting of the portion of the cardiac glycoside without the sugar residues, were studied because of the possible importance of hydrogen-bonding in their action, which is of therapeutic interest, and because they demonstrate the properties of the GRID hydrogen-bond functions for a target carbonyl oxygen atom. Their structures, which were obtained from the Cambridge Structural Database [131], are in minimum energy conformations [160] and were first superimposed by doing a least squares fit of atoms C1-C17, O3 and O4 (see Figure 3.8).

Gitoxigenin is much less active than digitoxigenin [157,158]. The directionality of a hydrogen-bond to the carbonyl oxygen O23 on the lactone ring of these two compounds can be seen clearly by using an amide nitrogen probe which is able to donate one hydrogen-bond. The energy contours at -3 kcal/mol near the lactone carbonyl oxygen atom are quite different for digitoxigenin (Figure 3.8.a) and gitoxigenin (Figure 3.8.b). This is highlighted by the

selectivity map (Figure 3.8.c), contoured at -3 kcal/mol and shown with the structure of gitoxigenin, which indicates where a stronger interaction with the probe could occur for digitoxigenin than for gitoxigenin. In addition to the two contours which occur at the anticipated position near the lactone oxygen on this map, there is a third contour due to the extra hydroxyl group O16 in gitoxigenin. This hydroxyl group is able to form an intramolecular hydrogen-bond to O14 thus preventing the probe from hydrogen-bonding to O14 when it is positioned near the coordinates of O16 and giving rise to a region where digitoxigenin can interact more favourably than gitoxigenin with the probe. In fact, the intramolecular hydrogen-bond between O14 and O16 in gitoxigenin is thought to cause the lactone ring to be tilted and rotated so that the carbonyl oxygen O23 is 2.4Å from its position in digitoxigenin [160], and the selectivity map explicitly draws attention to this interesting feature. The decreased binding energy of gitoxigenin indicated by program GRID may contribute to its reduced activity compared to digitoxigenin.

3.4.4.3 Cytochrome P450-cam

Cytochrome P450-cam catalyses the hydroxylation of camphor to 5-exo-hydroxycamphor using molecular oxygen. Its structure has been observed by x-ray crystallography with [161,162] and without [163] the substrate bound. The use of program GRID on this enzyme is described in order to demonstrate how the GRID energy potential with the new hydrogen-bond terms can be applied to macromolecules and can produce results in accordance with known experimental observations.

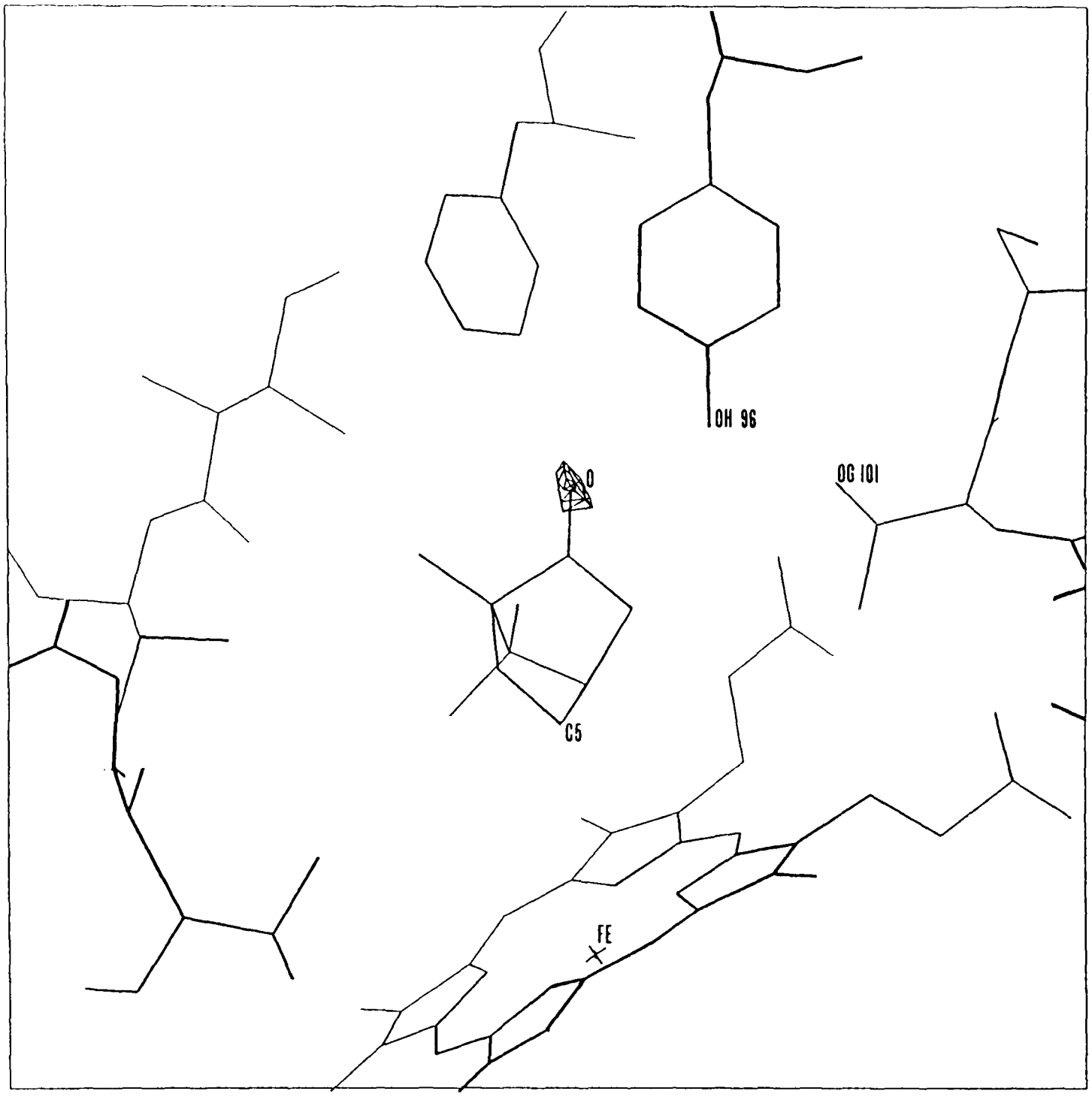


Figure 3.9 An energy contour map at -2.5 kcal/mol for a carbonyl oxygen probe at the substrate binding site of cytochrome P450-cam. See text.

It is thought [161] that the formation of a hydrogen-bond between the camphor carbonyl oxygen and the hydroxyl oxygen of tyrosine 96 may hold the camphor molecule in the correct orientation in the binding site of cytochrome P450-cam with its 5-carbon exposed to the molecular oxygen and to the haem group. The experimental structure shows that the camphor oxygen is in the plane of the tyrosine ring at 65.8° to the C-O axis, and at a distance of 2.65Å from the tyrosine oxygen atom. This arrangement would indicate favourable hydrogen-bond formation. However, it has also been suggested [162] that the interaction of the keto group with the tyrosine 96 hydroxyl group, may only make a modest contribution to the free energy of binding because solvent may bind to tyrosine 96 in the absence of camphor. The role of solvent in this site is considered in Chapter 4.

Program GRID demonstrates the importance of the directionality of the hydrogen-bonds made by the tyrosine 96 residue in this protein. The most energetically favourable position for a carbonyl oxygen probe in the active site is shown by the contours at -2.5 kcal/mol in Figure 3.9. These show where the carbonyl probe can hydrogen-bond to the tyrosine 96 hydroxyl group and are at precisely the observed position of the carbonyl oxygen of the camphor substrate. An energy contour is not seen on the other side of the tyrosine C-O bond because the hydroxyl group of threonine 101 is positioned to make an intramolecular hydrogen-bond with tyrosine 96. Thus, program GRID distinguishes correctly between the hydroxyl groups of residues 96 and 101, and uniquely defines the tyrosine 96 binding site as the position where a favourable and specific interaction can be made.

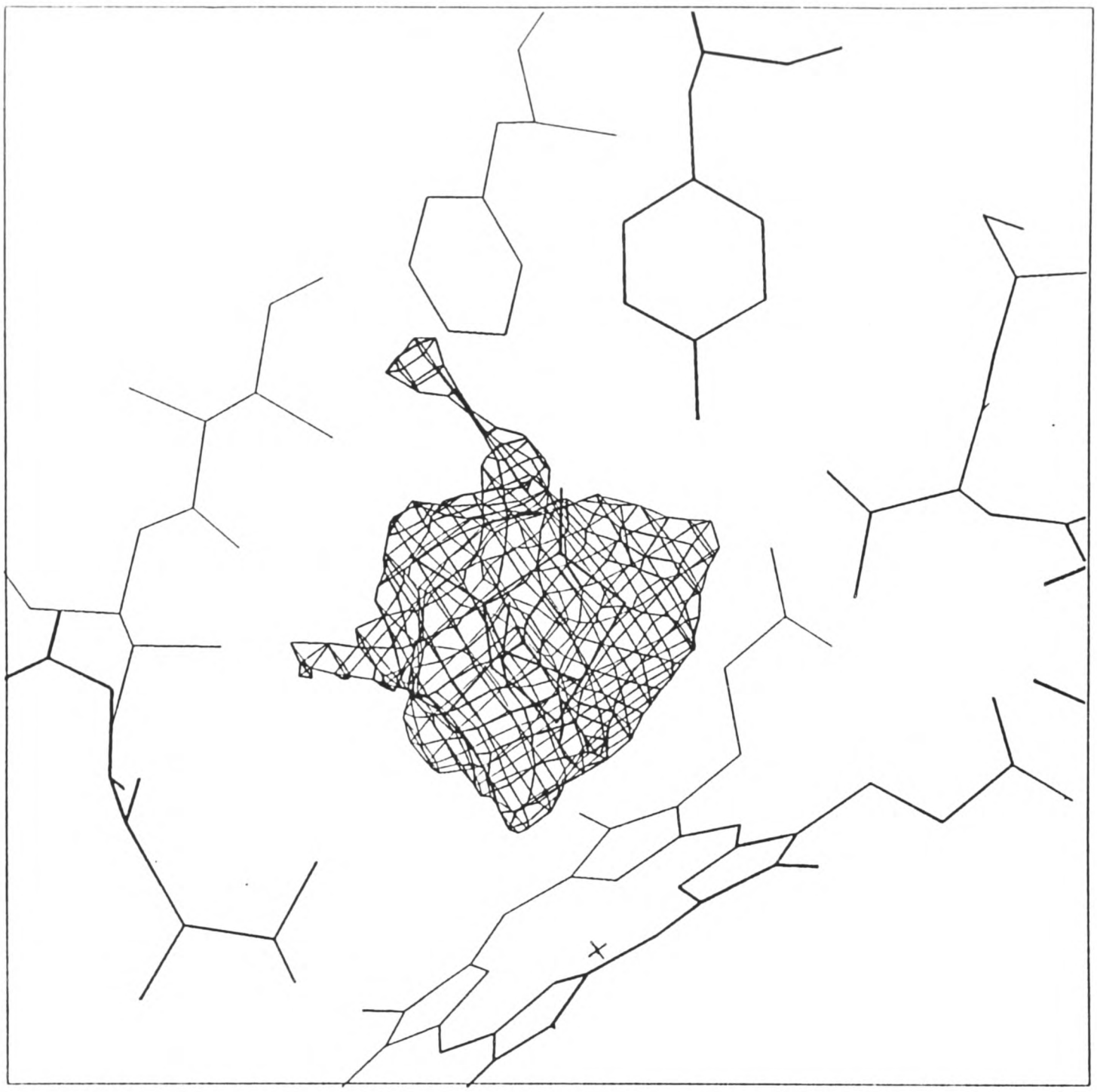


Figure 3.10 The energy contours at -2.0 kcal/mol for a methyl group probe interacting with cytochrome P450-cam. These energy contours surround the camphor substrate except for its polar carbonyl oxygen atom, and they define the position of the camphor molecule much more clearly than the van der Waals' surface of the protein does.

The camphor molecule is hydrophobic apart from its carbonyl oxygen. In fact, the whole hydrocarbon skeleton of camphor is found within a region where a methyl probe can interact energetically favourably with the protein, and an energy contour at -2.0 kcal/mol for this methyl probe, as shown in Figure 3.10, defines the camphor position much more clearly than does the van der Waals' surface of the protein itself. This shows the way in which program GRID can be used in sterically fitting a small molecule to a receptor site. It can be seen that the camphor oxygen atom alone is outside these methyl contours showing that it is in a less hydrophobic region of the active site than the rest of the substrate molecule. The orientation of the camphor is apparently controlled by the precisely defined geometry of the hydrogen-bond between the camphor oxygen and tyrosine 96, while the position of the hydrophobic part of the substrate is determined by steric interactions with the surrounding protein.

3.5 THE DEPENDENCE Ep OF THE HYDROGEN-BOND ON ITS ORIENTATION AT THE PROBE

There are more factors involved in taking account of the geometry of hydrogen-bonds at the probe than at the target as the probe has more degrees of freedom than a target atom of the same chemical type. This is because, in program GRID, the probe is completely free to rotate at each position at which the interaction energy is calculated, and the positions and chemical types of the atoms covalently bound to it are undefined. The repulsive van der Waals' interactions of these adjacent ligand atoms have an important steric effect on the observed hydrogen-bond geometry but this cannot

be considered directly for the probe in the way that it can for a target atom.

There are two distinct aspects to be addressed when determining the dependence of the strength of hydrogen-bonds on their orientation at the probe:

1. The functional form Ep of the angular dependence of the hydrogen-bonds made to the probe. This was modelled to fit experimental data in the same way as Et.

2. The determination of the combination of hydrogen-bonds which gives the most attractive total interaction energy E_{tot}. This varies in complexity according to the hydrogen-bonding capacity of both the probe and the target molecule.

If only one hydrogen-bond is formed to the probe, the probe may be assumed [164] to orient itself so that the hydrogen-bond is optimally aligned [139] along the vector from the centre of the probe to a lone pair orbital or to a hydrogen atom of the probe with angle p=0° and Ep=1 (see Figure 2.2). However, if more than one hydrogen-bond is formed, the probe geometry may be such that not all hydrogen-bonds will be optimally oriented at the probe. The energies of the weaker hydrogen-bonds must then be multiplied by a factor Ep<1 to account for the less favourable geometry. In addition, it may be that at a certain position of the probe, the hydrogen-bonding capacity of the target or of the probe itself cannot be satisfied, and then program GRID must choose which hydrogen-bonds are formed and which are not, such that the total interaction energy E_{tot} is optimised.

The case in which the probe has the capacity to make only two hydrogen-bonds is considered in the following section. Treatment

of probes able to make three or four hydrogen-bonds is described in Section 3.7.

3.6 PROBES CAPABLE OF FORMING TWO HYDROGEN-BONDS

3.6.1 The derivation of function Ep

Function Ep was derived by fitting to experimental data from neutron and x-ray diffraction studies as described for the derivation of Et. Ep is only applied when the probe is making more than one hydrogen-bond, and therefore, where possible, only experimental observations of atoms making more than one hydrogen-bond were used for formulating Ep because the hydrogen-bond geometry is influenced by the number of hydrogen-bonds a particular atom makes. For instance, for a carbonyl oxygen atom as noted before, it is quite common for a hydrogen-bond to occur near the bisector of the lone pair orbitals if the oxygen accepts only one hydrogen-bond. However, if the carbonyl oxygen accepts two hydrogen-bonds, a hydrogen-bond in this direction is unlikely [117,121,165].

An $Ep=\cos^2 p$ function is applied to the energy of the weaker hydrogen-bond for probes which can form two hydrogen-bonds (See Table 3.4). It reproduces the experimental data while being continuous and simple to compute.

It has been experimentally observed that stronger, shorter hydrogen-bonds in general show less deviation from linearity than longer, weaker hydrogen-bonds [129,135,139]. An $Ep=\cos^2 p$ function is able to model this relationship because it produces a greater

Table 3.4 The hydrogen-bond function Ep for different types of probe.

Probe	Coordination geometry[a]	Number[b] of Hydrogen-bonds donated	accepted	Hydrogen-bond number and type[c]		Angular function Ep[d]	Angular range of Ep[e]
TWO HYDROGEN-BONDS							
carbonyl, carboxyl and phenolate O	PT	0	2	2	A	$\cos^2 p$	$0<p<90°$
phenolic hydroxyl OH	PT	1	1	2	A	$\cos^2 p$	$0<p<90°$
ether O	T	0	2	2	A	$\cos^2 p$	$0<p<90°$
N	PT	0	2	2	A	$\cos^2 p$	$0<p<90°$
NH	T,PT	1	1	2	A or D	$\cos^2 p$	$0<p<90°$
amide NH_2	PT	2	0	2	D	$\cos^2 p$	$0<p<90°$
amine NH_2	T	2	0	2	D	$\cos^2 p$	$0<p<90°$
THREE HYDROGEN-BONDS							
hydroxyl OH	T	1	2	2,3	A or D	$\cos^2 p$	$0<p<90°$
	T	1	1	2	A or D	$\cos^2 p$	$p_{ad}<110°$
	?			2	A or D	1	$110°<p_{ad}<128°$
	PT			2	A	$\cos^2 p_b$	$p_{ad}>128°$
				2	D[f]	$\cos^2 p$	$p_{ad}>128°$
amine NH_2	T	1,2	1	2,3	A or D	$\cos^2 p$	$0<p<90°$
ammonium NH_3^+	T	2,3	0	2,3	D	$\cos^2 p$	$0<p<90°$
fluorine and chlorine		0	2,3	2,3	A	1	
FOUR HYDROGEN-BONDS							
water H_2O	T	1,2	2	2,3,4	A or D	$\cos^2 p$	$0<p<90°$
	T	1	1	2	A or D	$\cos^2 p$	$p_{ad}<110°$
	?			2	A or D	1	$110°<p_{ad}<128°$
	PT			2	A	$\cos^2 p_b$	$p_{ad}>128°$
				2	D[f]	$\cos^2 p$	$p_{ad}>128°$
		2	1[g]	2	D	$\cos^2 p$	$0<p<90°$
	T			3	A	$\cos^2 p$	$p_b>55°$
	PT			3	A	$\cos^2 p_o$	$p_b<55°$
	T	2	1[h]	2	A or D	$\cos^2 p$	$p_{ad}<118°$
	PT			2	A	$\cos^2 p_b$	$p_{ad}>118°$
	PT			2	D[f]	$\cos^2 p$	$p_{ad}>118°$
	T,PT			3	D	$\cos^2 p$	$0<p<90°$

[a] T=tetrahedral. PT=planar trigonal. ? =undefined coordination geometry (see note e). The classified coordination geometry is assumed to be that at which optimumly oriented hydrogen-bonds can be made and angle p is determined by the deviation of the positions of the target atoms from this geometry.

[b] This is the number of hydrogen-bonds that the probe can make at the particular probe position considered.

[c] The possible hydrogen-bonds are numbered according to the order in which their orientation at the probe and function Ep is computed. The strongest of the hydrogen-bonds formed is considered first and is always optimumly oriented at the probe with Ep=1. It is not listed in this table. A=accepted by the probe. D=donated by the probe.

[d] p is the angle made by the hydrogen-bond at the probe as defined in Figure 2.2. p_o is the angle of deviation of the hydrogen-bond accepted by the probe from the plane of the lone pair orbitals. p_b is the angle of deviation of the hydrogen-bond accepted by the probe from the bisector of the lone pair orbitals.

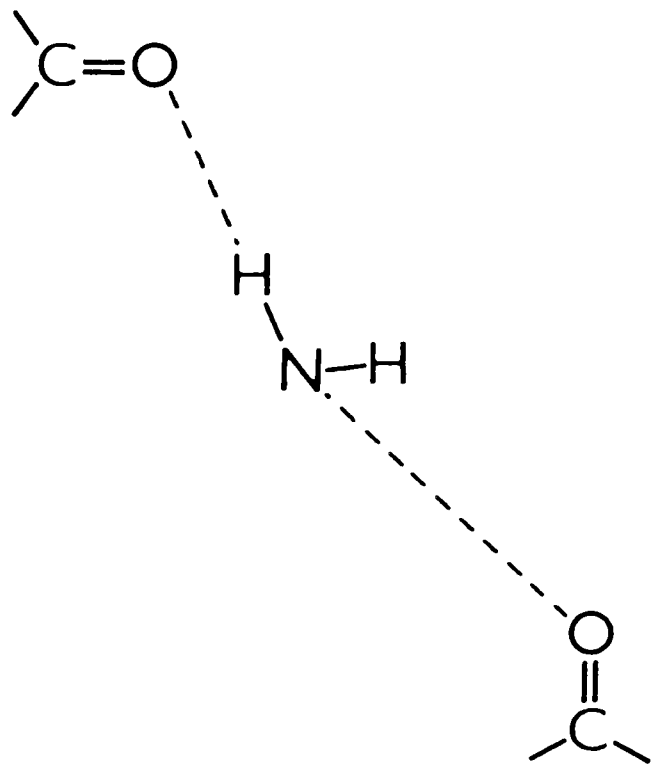
[e] p_{ad} is the angle subtended at the probe by a target acceptor atom and a target hydrogen-bonding hydrogen atom forming the first and second hydrogen-bonds to the probe. For tetrahedral coordination geometry p_{ad}=110°. For trigonal coordination geometry, p_{ad}=128°(=180-52) allowing for the internal angle X-O-H of the probe, where X=H or C, to be as small as 104°. If $110° < p_{ad} < 128°$ and only two hydrogen-bonds are made, the coordination geometry of the probe cannot be defined as trigonal or tetrahedral and the probability of hydrogen-bond formation between these angles is assumed to be constant. If $110° < p_{ad} < 128°$ and three hydrogen-bonds are made, then the coordination geometry of the probe must be determined in order that Ep can be calculated for the third hydrogen-bond. If $p_{ad} < 118° (=\frac{1}{2}[108+128])$, then the coordination geometry is assumed to be tetrahedral, if $p_{ad} > 118°$, it is assumed to be trigonal.

[f] If the coordination geometry is trigonal and the strongest hydrogen-bond is accepted, it is assumed to lie along the bisector of the probe lone pair orbitals. The probe hydrogens may then lie on a circular locus around this bisector.

[g] The two hydrogen-bonds donated by the probe are stronger than the accepted one and therefore the positions of the probe hydrogen atoms and lone pairs can be assigned before Ep is calculated for the accepted hydrogen-bond. If $p_b > 55°$, the hydrogen-bond lies outside the lone pair orbitals and the water probe is assumed to be tetrahedrally coordinated. If $p_b < 55°$, the hydrogen-bond is within the lone pair orbitals and the water may be trigonally coordinated.

[h] The accepted hydrogen-bond is stronger than one or both of the donated hydrogen-bonds. The assigned positions of the hydrogen atoms and lone-pairs depends on the angle p_{ad} and so the lone pair plane is not defined when Ep is calculated for the accepted hydrogen-bond.

a.



b.

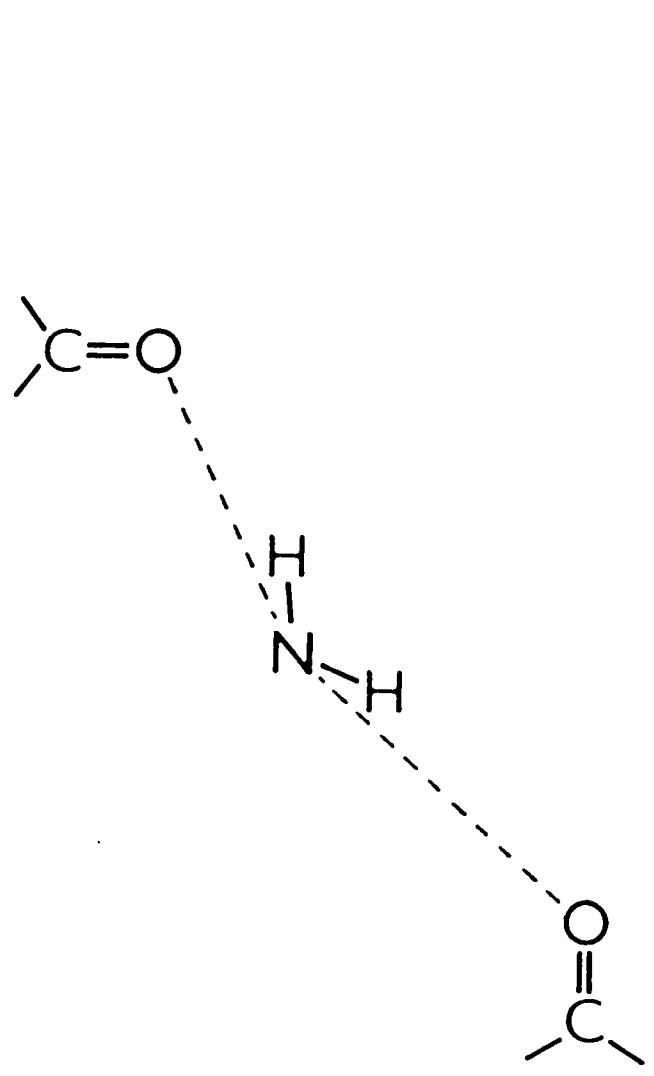


Figure 3.11 The geometry of the hydrogen-bonds donated by an amine nitrogen probe: a. asymmetric with an amine hydrogen pointing directly at the closest target oxygen atom. b. symmetric with the bisector of angle O-N-O coinciding with the bisector of angle H-N-H. From reference [164].

variation in hydrogen-bond energy with angle p for strong hydrogen-bonds than for weak hydrogen-bonds.

3.6.2 The determination of the combination and strengths of the hydrogen-bonds made to the probe

It has been shown [164], for an amine and an amide nitrogen probe hydrogen-bonding to two target carbonyl oxygens, that an asymmetric configuration (as defined in Figure 3.11), of the two hydrogen-bonds results in a more favourable total hydrogen-bond energy than a symmetric configuration unless both the hydrogen-bonds are very short. In the latter case, the difference in the hydrogen-bond energy calculated for the two configurations is very small. This result can be explained by the fact that, in the asymmetric case, $Ep<1$ only for the weaker hydrogen-bond and the full strength of the stronger hydrogen-bond is always retained, whereas in the symmetric case, $Ep<1$ for both hydrogen-bonds. The results of this theoretical calculation are supported by nmr investigations [166] which suggest that, for a water molecule hydrogen-bonded to a protein, one hydrogen-bond is stronger than the others so that the water has an asymmetric arrangement of hydrogen-bonds and its motion is rotationally anisotropic.

Accordingly, it is assumed in program GRID that the hydrogen-bonds adopt an asymmetric configuration in which the strongest hydrogen-bond is aligned optimally at the probe with $Ep=1$, and the weaker one deviates from linearity at the probe by an angle p and is therefore modified by an appropriate angular factor $Ep<1$.

The procedure adopted by program GRID in determining the hydrogen-bonds made by the probe is as follows.

Step 1. Orient the probe so that its hydrogen atom, if donating, or its lone pair, if accepting, points directly at the target atom to which it can make the strongest hydrogen-bond so that p=0° and Ep=1. (The target atom is a hydrogen atom when the hydrogen-bond is accepted by the probe. See Figure 2.2).

Step 2. Calculate $Ep=\cos^2 p$ for the other weaker hydrogen-bonds which the probe could make and identify the most energetically favourable one which results in the optimum total interaction energy, E_{tot}, (taking account of the dependence of E_{lj} on the formation of hydrogen-bonds).

Step 3. Orient the probe so that p=0° for the second strongest hydrogen-bond. Repeat Step 2. to see if a more favourable value of E_{tot} can be obtained by assuming that this hydrogen-bond is linear.

Step 4. Repeat Step 3. for the third strongest hydrogen-bond and so on until the combination of hydrogen-bonds yielding the most favourable value of E_{tot} is found.

This procedure is illustrated for a carbonyl oxygen probe in Figure 3.12 and Table 3.5.

3.6.3 An example of the application of the hydrogen-bond function Ep for probes capable of forming two hydrogen-bonds

L-noradrenaline (I) serves as a suitable target molecule with which to demonstrate the effect on the calculated hydrogen-bond energy of taking the geometry of the probe into account. Calculations were performed with program GRID for three different probes, each

Table 3.5 The determination by program GRID of the hydrogen-bonds accepted by a carbonyl oxygen probe interacting with different targets.

Figure[a]	Target atom	Distance r in Å	Er in kcal/mol	Angle t in °	Et (=$\cos^2 t$)	Angle p in °	Ep (=$\cos^2 p$)	E_{hb} in kcal/mol	ΣE_{hb}[b] in kcal/mol
a	N1	3.00	-3.00	0	1.00	0	1.00	-3.00	-3.00
	N2	3.25	-2.68	0	1.00	30	0.75	-2.01	-2.01
									-5.01
b	N1	3.00	-3.00	0	1.00	0	1.00	-3.00	-3.00
	N2	3.25	-2.68	0	1.00	30	0.75	-2.01	-
	N3	3.00	-3.00	20	0.88	20	0.88	-2.32	-2.32
									-5.32
c	N1	3.00	-3.00	0	1.00	0	1.00	-3.00	-3.00
	N2	3.25	-2.68	0	1.00	30	0.75	-2.01	-2.01
	N3	3.00	-3.00	20	0.88	30	0.75	-1.98	-
									-5.01
d[c]	N1	3.00	-3.00	0	1.00	-	-	-	-
	N2	3.25	-2.68	0	1.00	0	1.00	-2.68	-2.68
	N3	3.00	-3.00	20	0.88	0	1.00	-2.64	-2.64
									-5.32

[a] The carbonyl oxygen probe surrounded by different arrangements of target nitrogens is shown in Figure 3.12.

[b] The hydrogen-bonds predicted to be formed are listed and their sum ΣE_{hb} gives the total hydrogen-bond interaction energy between the probe and the target.

[c] If the strongest hydrogen-bond is accepted from N1 then ΣE_{hb} will be the same as in Figure c. However, a greater hydrogen-bonding energy can be achieved if the probe accepts hydrogen-bonds from N2 and N3 and not from N1.

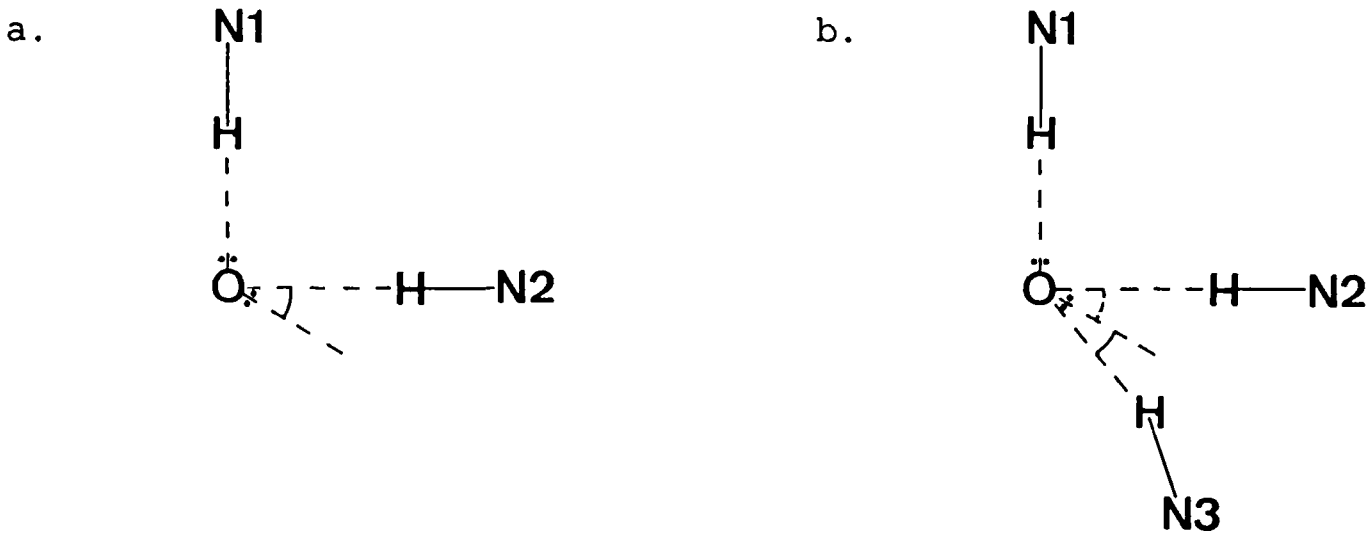
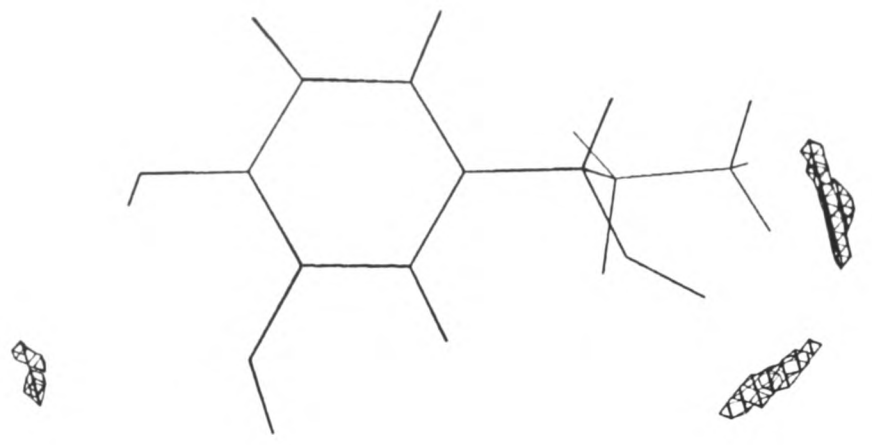
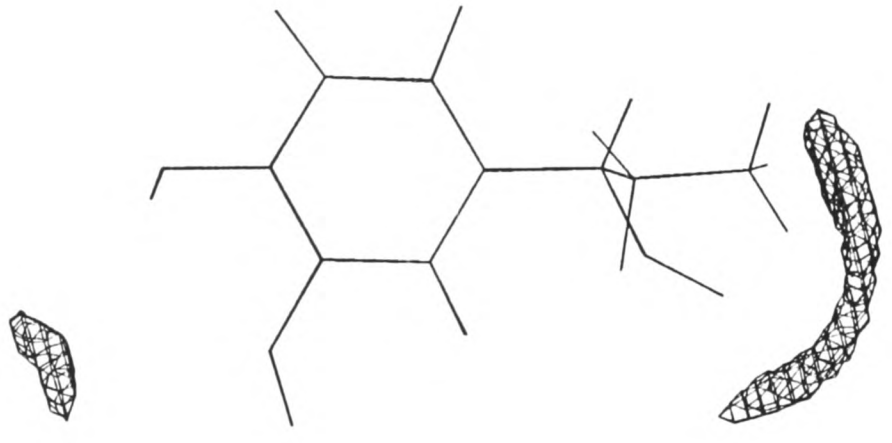


Figure 3.12 The determination by program GRID of the hydrogen-bonds accepted by a carbonyl oxygen probe from different target molecules containing nitrogens able to donate hydrogen-bonds. The hydrogen-bonds formed and their computed energies are listed in Table 3.5. The plane of the paper corresponds to the plane of the probe's lone-pair orbitals. Angle p is marked by a solid line for hydrogen-bonds predicted to be formed and by a dashed line for those not predicted to be formed. a. The target contains two hydrogen-bonding atoms only. b.c. and d. The target contains three hydrogen-bonding atoms at different orientations giving rise to different combinations of hydrogen-bonds being formed. The hydrogen-bond to N3 deviates from linearity at the target with t=20°. In d. the hydrogens of N2 and N3 subtend an angle of 120° at the probe and so it is more favourable for hydrogen-bonds to be formed to both these atoms than to N1.

a.



b.



c.

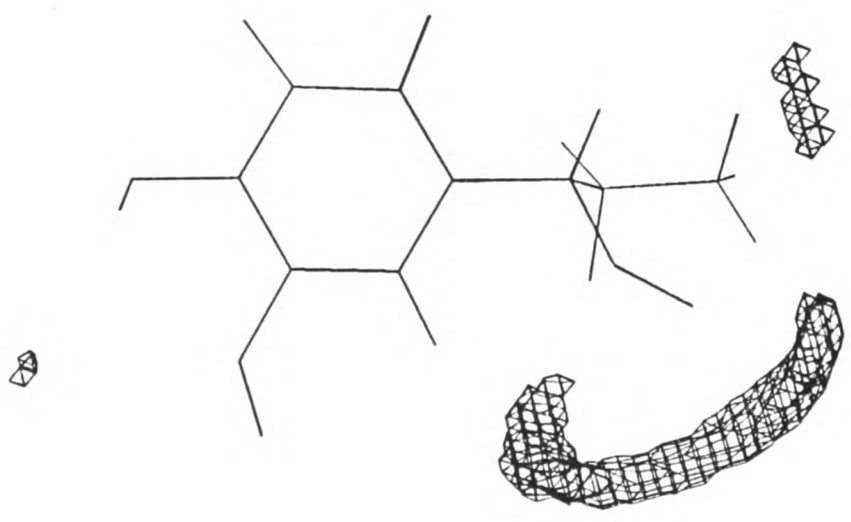


Figure 3.13 a. and b. L-noradrenaline with energy contours at -5.3 kcal/mol for the interaction with a. a phenolic hydroxyl probe and b. an sp^3 hybridized hydroxyl probe. Both probes have a charge of -0.1e which is approximately equal to that assigned to such groups in the side-chains of amino acid residues by program GRID. c. L-noradrenaline with energy contours at -5.6 kcal/mol for the interaction with a carboxyl oxygen probe with a charge of -0.575e.

capable of making two hydrogen-bonds, interacting with a target of L-noradrenaline in a crystallographically observed conformation [153].

In Figure 3.13.a, energy contours at -5.3 kcal/mol are shown for the interaction with L-noradrenaline of a phenolic hydroxyl probe with a charge of -0.1e, which is the charge assigned to the tyrosine hydroxyl group in program GRID. The contours show regions where the probe can form two hydrogen-bonds to the target.

In Figure 3.13.b, energy contours are shown at the same energy with the same target molecule for an sp^3 hybridized hydroxyl probe with the same charge, which is approximately equal to the charge assigned to a serine or threonine hydroxyl group in program GRID. The probe was only permitted to accept one hydrogen-bond so that a direct comparison with the calculations for the phenolic hydroxyl probe could be made. In fact, the contours obtained if the sp^3 hybridized hydroxyl probe is allowed to accept two hydrogen-bonds are almost identical. The energy contours show regions where the probe accepts one hydrogen-bond and donates one hydrogen-bond.

The differences between the contours in Figure 3.13.a and 3.13.b are due solely to the difference in the geometry of the two probes. The contoured regions are smaller for the phenolic hydroxyl probe, (for which the optimum angle subtended by the hydrogen-bonds at the probe is assumed to be 120°), than for the sp^3 hybridized hydroxyl probe, (for which the optimum subtended angle is assumed to be 110°), because of the close proximity of the atoms of the target molecule which hydrogen-bond simultaneously to the probe i.e. the two oxygens on the aromatic moiety of the target, and the nitrogen and alkyl hydroxyl oxygen on the ethanolamine moiety. Thus at a given position of the probe, angle p is larger, and therefore Ep is smaller,

for the weaker hydrogen-bond to the phenolic hydroxyl probe than for that to the sp^3 hybridized hydroxyl probe.

It is thought that the positively charged nitrogen of noradrenaline may play a dominant role in determining binding to the adrenergic receptor by interacting with a negatively charged group on the receptor such as a phosphate or a carboxylate group [167,168]. Therefore, the interaction of a carboxyl oxygen probe with L-noradrenaline was calculated using program GRID. A charge of -0.575e was placed on the probe as this charge is assigned to an oxygen atom in the side chain of Asp and Glu residues in program GRID. The energy contours obtained are shown in Figure 3.13.c. The largest contour, which also contains the most energetically favourable binding site, is at a position where the probe can accept hydrogen-bonds from the nitrogen and the alkyl-hydroxyl group and can also make a strong attractive electrostatic interaction with the positively charged nitrogen. This contour thus indicates a region where the adrenergic receptor could interact strongly with noradrenaline. The adrenergic receptor has been modelled by formate [169] and the favourable binding region for the carboxyl probe identified by program GRID is consistent with the position adopted by formate on energy minimization with noradrenaline. The contours near the aromatic moiety of noradrenaline in Figure 3.13.c are smaller than those for the phenolic hydroxyl group because there is a stronger repulsive interaction between the target phenolic oxygens and the carboxyl oxygen probe.

These calculations of the interaction of three different probes with noradrenaline demonstrate the significant impact on molecular interactions of small changes in the geometry and charge of the probe.

3.7 PROBES CAPABLE OF FORMING MORE THAN TWO HYDROGEN-BONDS

3.7.1 The derivation of function Ep

Function Ep for probes which are able to make more than two hydrogen-bonds was derived from experimental data in the same way as for probes which can only make two hydrogen-bonds. It is given in Table 3.4. The orientation of hydrogen-bonds at water molecules was not considered in detail when function Et was formulated and therefore, its hydrogen-bonding properties are discussed now.

3.7.1.1 Water

The water molecule is able to donate two and accept two hydrogen-bonds. On hydrogen-bonding, the H-O-H angle opens from its water vapour value of 104.5° by an average of 3.6±0.4° and the O-H bond elongates from 0.9572±0.0003Å to 0.98±0.006Å [170]. In crystal structures, the water molecule has been observed to be both tetrahedrally and planar trigonally coordinated with angle H-O-H =106.9±0.6° and 109±0.5° respectively [170]. Therefore a water probe may reasonably be modelled in program GRID, for the purpose of determining the optimum orientation of its hydrogen-bonds, with a perfectly tetrahedral angle H-O-H and an O-H bond length of 1Å.

3.7.1.1.1 Hydrogen-bonds donated by the water probe

Water has been found to donate hydrogen-bonds more often than it accepts them. In small molecule hydrates, water nearly always donates two hydrogen-bonds [118,119,120]. In proteins, bound waters

are more frequently observed donating hydrogen-bonds than accepting them [132,171]. This is because there are more oxygens than nitrogens in proteins and these can generally accept more hydrogen-bonds with fewer geometrical constraints than nitrogens. In the peptide bond, the carbonyl oxygen can stick out further into the solvent than the amide nitrogen and has a greater hydrogen-bonding capacity. However, it has been postulated that water may also have an innate tendency to donate hydrogen-bonds [132].

Linear hydrogen-bonds donated along the direction of the water O-H bond are favoured. Bending of these hydrogen-bonds has been found to be isotropic and to follow a Gaussian distribution for angle $p<20°$ [119]. No preference for the hydrogen-bond acceptor to lie in the plane of the donor water molecule has been found [119].

In program GRID, and $Ep=\cos^2 p$ function is applied to hydrogen-bonds donated by the water probe.

3.7.1.1.2 <u>Hydrogen-bonds accepted by the water probe</u>

Water tends to accept hydrogen-bonds in the lone pair plane with the angle of deviation p_o from this plane generally less than 20° [121]. The hydrogen-bond distribution shows less angular dependence between the lone pair orbitals than outside them. The variation in angle p is smaller if two hydrogen-bonds are accepted than if only one is [121].

In small molecule crystal hydrates, water has been classified according to its lone pair coordination into the following categories [118,172]:

1. Tetrahedrally coordinated with two hydrogen-bonds accepted along the lone pair orbitals.

2. Tetrahedrally coordinated with one hydrogen-bond accepted along a lone pair orbital approximately 45° out of the plane of the water molecule.

3. Planar trigonally coordinated with one hydrogen-bond accepted along the bisector of the lone pairs within 25-30° of the plane of the water molecule. In a study of 663 waters observed by neutron diffraction [170], more than half of the waters coordinated to a cation and almost all those coordinated to highly charged cations were planar trigonally coordinated. Of the 70 waters that had only one donor OH or NH neighbour, eight were planar trigonally coordinated. Several of these made a hydrogen-bond with an acidic hydroxyl group which was very short (≈2.55Å). A planar trigonal coordination is taken into account by program GRID even though it is probably less common than a tetrahedral coordination in biological systems.

4. Bipyramidally coordinated with three hydrogen-bonds accepted. This is very rare [170] and is therefore not modelled in program GRID.

5. No hydrogen bonds accepted.

In program GRID, the water probe is modelled as an extended atom with a coordination geometry which is dependent on the hydrogen-bonding capacity of the target molecule and may be classified as either tetrahedral or trigonal.

For a tetrahedrally coordinated water probe, an $Ep=\cos^2 p$ function is applied to the accepted hydrogen-bonds. For waters with a planar trigonal coordination geometry for which the lone pair plane can be defined, $Ep=\cos^2 p_o$ for an accepted hydrogen-bond which lies within the lone pair orbitals. Thus the hydrogen-bond is constrained

to the lone pair plane but has a uniform probability of occurrence between the lone pair orbitals.

3.7.2 The determination of the combination and strengths of the hydrogen-bonds made to the probe

This is a more involved process than for a probe making two hydrogen-bonds and is as follows.

Step 1. Orient the probe so that Ep=1 for the strongest hydrogen-bond.

Step 2. Calculate Ep for each of the combinations of possible weaker hydrogen-bonds and find the combination resulting in the most attractive total energy E_{tot}. Water and hydroxyl probes may be classified as either trigonally or tetrahedrally coordinated and their coordination geometry must be determined for each hydrogen-bond combination and the appropriate functional form of Ep computed (see Table 3.4).

Step 3. Orient the probe so that Ep=1 for the second strongest hydrogen-bond. Repeat Step 2. to see if a more favourable value of E_{tot} can be achieved.

Step 4. Repeat Step 3. for the third strongest hydrogen-bond and so on until the most favourable combination of hydrogen-bonds is found.

This procedure is illustrated for a water probe in Figure 3.14 and Table 3.6.

Table 3.6 The determination by program GRID of the hydrogen-bonds made by a water probe interacting with different targets.

Figure[a]	Target atom[b]	Distance r in Å	Er in kcal/mol	Coordination geometry[c]	Angle p in °	Ep (=$\cos^2 p$)	E_{hb} in kcal/mol	ΣE_{hb}[d] in kcal/mol
a	N1	3.00	-3.00		0.0	1.00	-3.00	
	N2	3.25	-2.68	T	20.0	0.88	-2.36	-5.36
b	O1	3.00	-3.67		0.0	1.00	-3.67	
	O2	3.25	-2.90	?	28.4	0.77	-2.24	-5.91
c	O1	3.00	-3.67		0.0	1.00	-3.67	
	$N2_1$	3.25	-2.68	T	20.0	0.88	-2.36	-6.03
	$N2_2$	3.25	-2.68	?	?	1.00	-2.68	-6.35
	$N2_3$	3.25	-2.68	PT	p_b=20.0[e]	0.88	-2.36	-6.03
d	N1	3.00	-3.00		0.0	1.00	-3.00	
	$O2_1$	3.25	-2.90	T	28.4	0.77	-2.24	-5.24
	$O2_2$	3.25	-2.90	?	?	1.00	-2.90	-5.90
	$O2_3$	3.25	-2.90	PT	28.4	0.77	-2.24	-5.24
e	N1	3.00	-3.00		0.0	1.00	-3.00	
	N2	3.25	-2.68		20.0	0.88	-2.36	
	O3	3.50	-2.18	T	28.4	0.77	-1.69	-7.05
f	O1	3.00	-3.67		0.0	1.00	-3.67	
	O2	3.25	-2.90		28.4	0.77	-2.24	
	$N3_1$	3.50	-2.14	T	20.0	0.88	-1.89	-7.80
	$N3_2$	3.50	-2.14	PT	p_o=10.0[e]	0.97	-2.08	-7.99

g	O1	3.00	-3.67		0.0	1.00	-3.67	
	N2	3.25	-2.68		20.0	0.88	-2.36	
	N3	3.50	-2.14	T	20.0	0.88	-1.89	-7.92
	O3	3.50	-2.18	T	28.4	0.77	-1.69	-7.72
h	O1	3.00	-3.67		0.0	1.00	-3.67	
	N2	3.25	-2.68		p_b=20·0e	0.88	-2.36	
	O3	3.50	-2.18	PT	28.4	0.77	-1.69	-7.72
i	N1	3.00	-3.00		0.0	1.00	-3.00	
	O2	3.25	-2.90		28.4	0.77	-2.24	
	N3	3.50	-2.14	T	20.0	0.88	-1.89	-7.13
	O3	3.50	-2.18	T	28.4	0.77	-1.69	-6.93
j	N1	3.00	-3.00		0.0	1.00	-3.00	
	O2	3.25	-2.90		28.4	0.77	-1.89	
	O3	3.50	-2.18	PT	52.0	0.38	-0.83	-5.72

[a] The water probe and the different targets are shown in Figure 3.14.

[b] Subscripts 1, 2 and 3 denote different positions of one of the target atoms.

[c] T=tetrahedral coordination geometry. PT=planar trigonal coordination geometry. ?=coordination geometry cannot be classified as tetrahedral or trigonal.

[d] ΣE_{hb} is the total hydrogen-bond interaction energy between the probe and the target.

[e] Angle p_b is the deviation from the lone pair bisector and $Ep=\cos^2 p_b$. Angle p_o is the deviation from the lone pair plane of the probe and $Ep=\cos^2 p_o$.

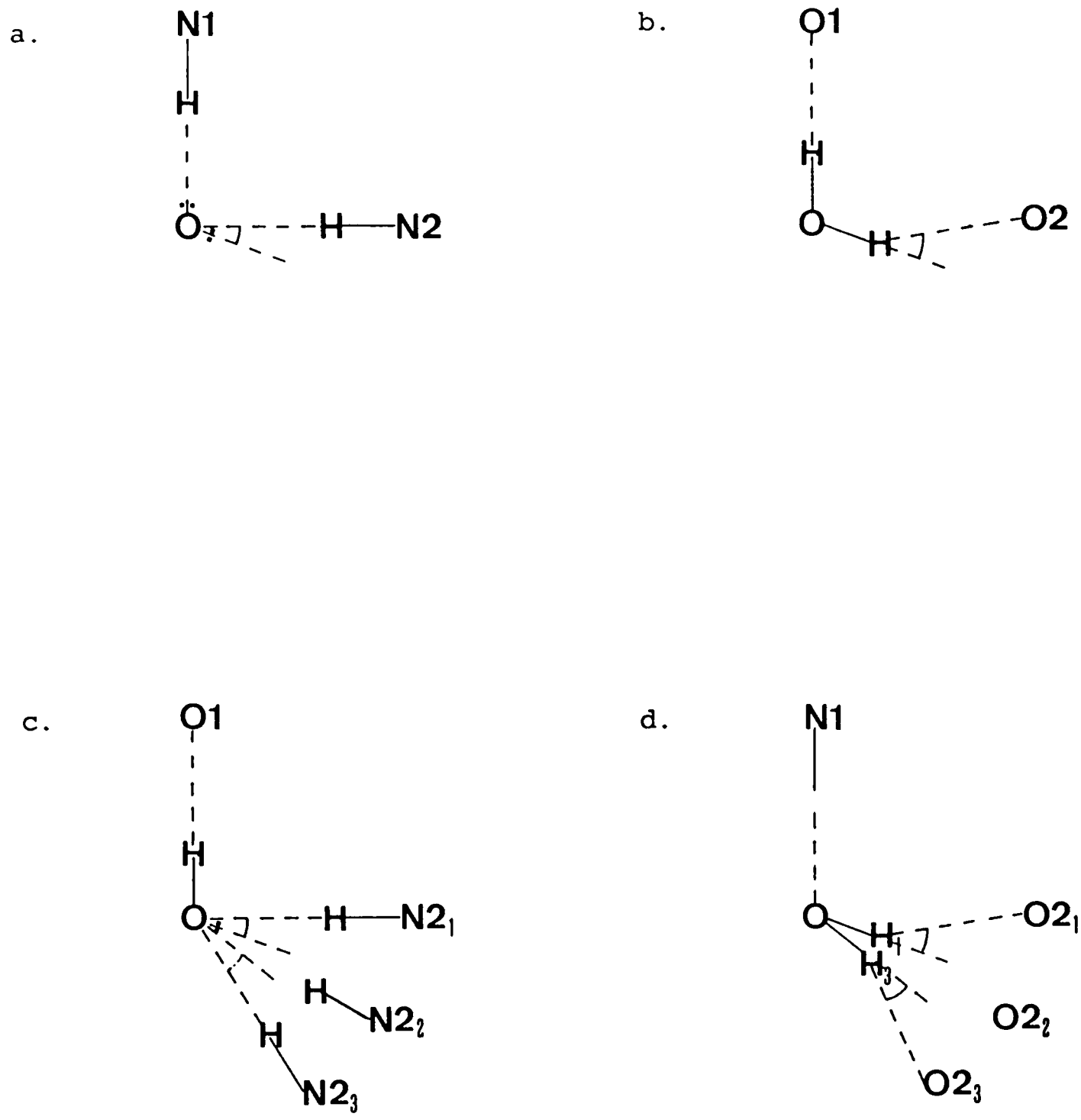
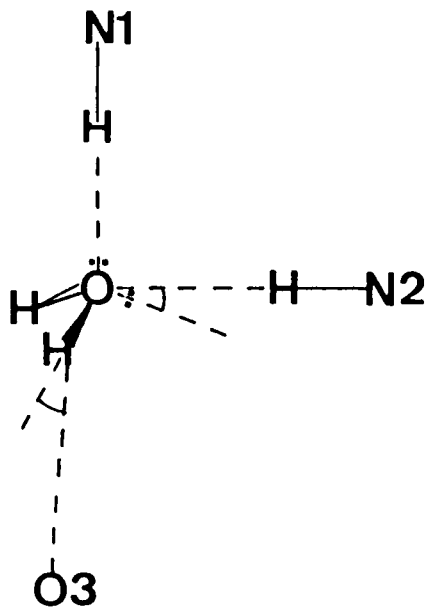
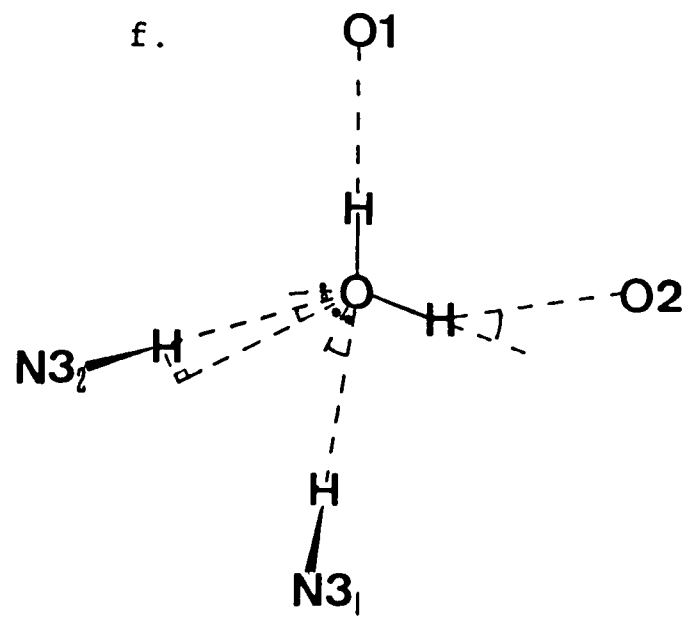


Figure 3.14 The determination by program GRID of the hydrogen-bonds made by the water probe. The hydrogen-bond energies evaluated are listed in Table 3.6. The plane of the paper corresponds to that of the water oxygen atom and the two target atoms making the strongest hydrogen-bonds to the probe. Angle p is marked by a solid line and angle p_b is marked by a dashed line. Where the water is planar trigonally coordinated, the lone-pair bisector is given by a dashed line and lone-pairs are not shown. a-d. The target contains two hydrogen-bonding atoms. In c. and d., three different orientations (1-3) of the target atom making the weaker hydrogen-bond are shown. e-j. The target contains three hydrogen-bonding atoms. If the probe is tetrahedrally coordinated, it could make a fourth hydrogen-bond if a suitable target atom were present. The orientation of the probe lone-pair orbital or hydrogen atom participating in a fourth hydrogen-bond, and hence angle p, could be deduced from the geometry of the three other hydrogen-bonds In f. two different orientations of the third hydrogen-bond are shown, the second to $N3_2$ corresponding to a hydrogen-bond near the bisector of the water's lone-pairs. This hydrogen-bond deviates from the lone-pair plane by angle p_o. In h. and j. the water probe is planar trigonally coordinated.

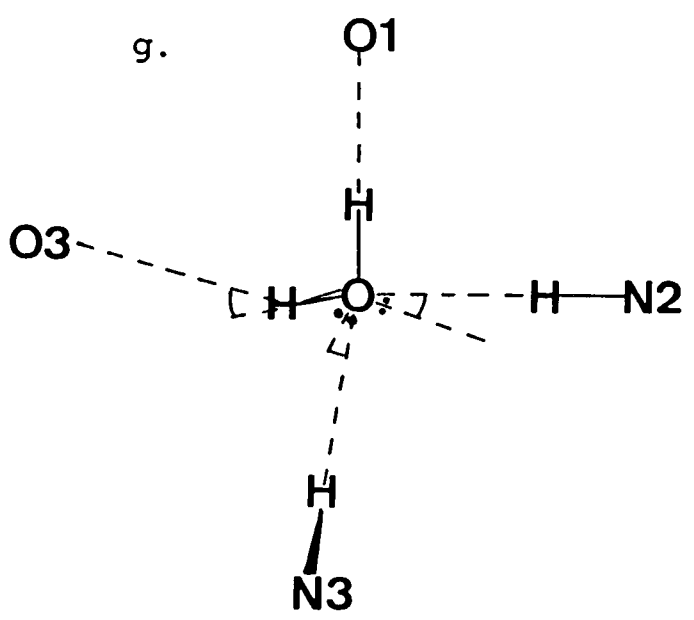
e.


N1
H
H
O
H
N2
H
O3

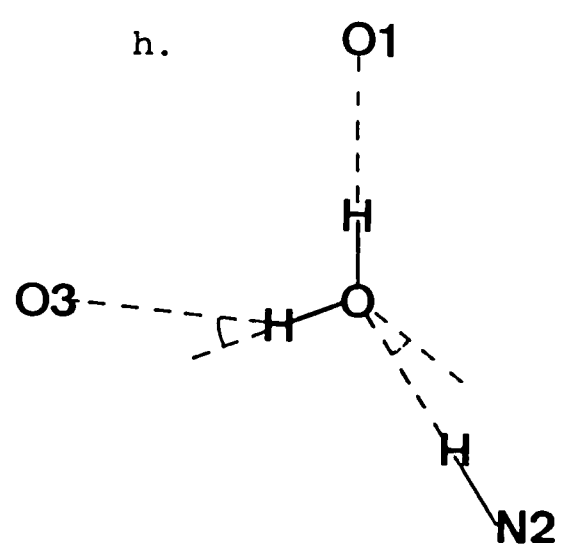
f.


O1
H
O
H
O2
N3
H
H
N3

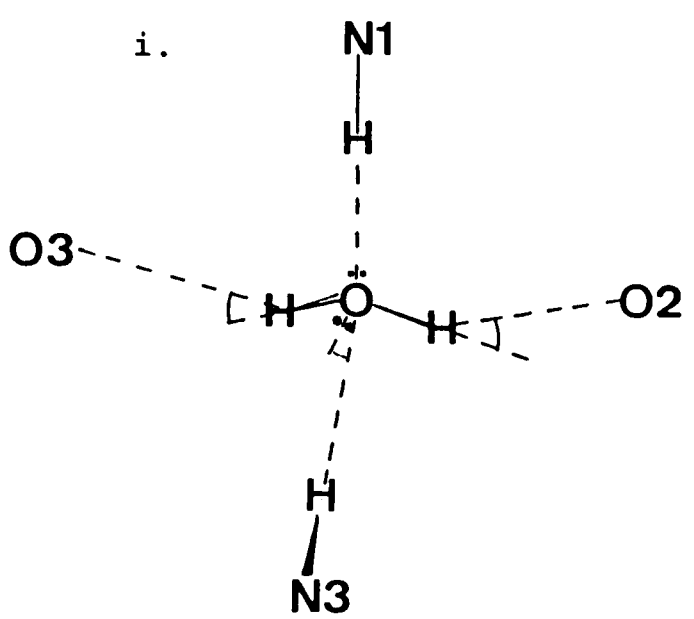
g.


O1
H
O3
H
O
H
N2
H
N3

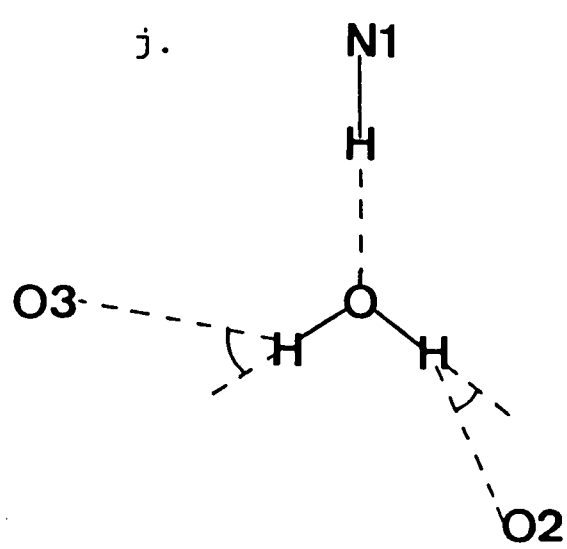
h.


O1
H
O3
O
H
N2

i.


N1
H
O3
H
O
H
O2
H
N3

j.


N1
H
O3
O
H
H
O2

3.7.3 The application of the hydrogen-bond function Ep for probes able to form more than two hydrogen-bonds

The function Ep was tested for probes which are able to form more than two hydrogen-bonds by predicting the positions of the ordered waters in biological crystal structures. This is described in the next chapter.

3.8 CONCLUSION

An energy function has been developed for use in calculating the interaction between a molecular probe group and a target molecule. The energy function consists of a Lennard-Jones, an electrostatic and a hydrogen-bond term. There are sufficient experimental observations of hydrogen-bonds in the literature to enable a set of hydrogen-bond functions to be modelled and fitted to experimental data. These hydrogen-bond functions are specific to different types of atom, and model the variation in strength and geometry of the hydrogen-bond according to the chemical nature of the donor and acceptor atoms as well as their position and orientation in the system under study. However, current experimental data for some types of atom are sparse and so functions to describe the hydrogen-bonding characteristics of certain atoms are still poorly defined. As more experimental data become available in the future, it should be possible to refine and improve such hydrogen-bond functions.

When appropriate, the hydrogen-bond term Et takes account of the mobility of the target hydrogens and lone pairs analytically.

Consideration of this property is vital in order that realistic predictions may be made. When a tautomeric change occurs on the approach of a ligand molecule, the functions allow for the appropriate charge redistribution in the target molecule. This treatment of molecular polarization has only been applied to histidine residues, but it exemplifies an important principle which could be extended to other target systems.

Program GRID allows the probe to orient itself so that the most energetically favourable interaction with the target is obtained. This is achieved by determining the optimum combination and strengths of the hydrogen-bonds to the probe while taking account of the fact that there may be more than one preferred coordination geometry for some probes, e.g. waters have been observed to be both trigonally and tetrahedrally coordinated.

However, a ligand may not be able to occupy a site at which strong binding is predicted by program GRID because of steric constraints. For instance, a carbonyl oxygen is covalently bound to a carbon atom. The algorithm for determining the optimum orientation of the carbonyl oxygen probe takes no account of the position of the carbon atom and its interaction with the target. It may make repulsive van der Waals' interactions with the target causing the predicted probe binding site to be unfavourable. Similarly, for a carboxyl oxygen probe, only the interaction of the oxygen itself is considered, but the binding site should be able to accommodate a whole carboxylate or amide group in a suitable orientation. These factors must at present be considered by examination of the target molecular structure and the GRID energy maps using computer graphics. However, the extended atom model of the probe used in program GRID could be

modified to take account of these factors automatically and this modification is currently being undertaken for a carbonyl group [173].

A few examples of the resultant energy maps have been shown, giving predictions which agree with experimental observations and which are chemically meaningful. The new energy function (in version GRID2) produces better results than the functions previously used in program GRID, (versions GRID0 and GRID1), which had only simple and general hydrogen-bond terms. The present findings confirm that the quality and accuracy of the predictions is improved when more carefully constructed hydrogen-bonding terms are used. The extensive testing of the new energy function is described in the next chapter and the energy function is applied to the design of novel therapeutic agents in Chapter 5.

CHAPTER 4

THE STUDY OF WATER BY THE GRID METHOD

In developing a theoretical method for predicting ligand binding, it is essential to test the method in order to ensure that there is good agreement between predictions and experimental observations. Program GRID was tested by using it to predict the interactions of water with a variety of molecules. The predicted properties of the water were compared both qualitatively and quantitatively with experimental observations and the extent of correspondence assessed. Program GRID was then used to examine the role of water in ligand binding.

4.1 WATER

4.1.1 The choice of water for this study

Water serves as a model ligand for a number of reasons.

1. It is ubiquitous, and observed experimentally in large numbers in a variety of environments. This enables statistical analysis of the accuracy of predictions as well as the detailed examination of predictions of particular water molecules. Most data on its three dimensional structure in biological systems come from x-ray and neutron diffraction. For example, the water structure may be observed in protein crystals which contain 30-78% water by volume [174] so the protein can be considered to be in a very concentrated solution rather than in the solid state.

2. The small size of the water molecule makes calculations less complex than those for larger molecules such as drugs. However, the prediction of water is not trivial and many attempts have been made to simulate its properties [175].

3. Water is of intrinsic interest. It makes important contributions to the properties of biological molecules which have evolved in water and are saturated with it. For instance, dry enzymes are inactive but they can be reactivated by hydration [13]. It influences the folding, stability, dynamics and function of biological molecules through its ability to form hydrogen-bonds, fill space and confer entropic advantage on particular molecular conformations [13].

4. It is important to include water in the consideration of the binding of ligands to macromolecules since water molecules may be displaced on binding, or a ligand may bind via water to a target molecule, (see e.g. [176,177]). Knowledge of the properties of water at a binding site is therefore necessary for the correct prediction of ligand binding.

5. Water is a well-studied molecule. There have been many experimental and theoretical studies of bulk water [178] and of the interaction of water with solute molecules [179,180].

4.1.2 The characteristics of the water solvating biological molecules

Water plays a vital role in the properties of biological molecules and they, in turn, influence its nature [179]. Distinctive characteristics of the water structure of biological macromolecules may be identified.

In proteins, water may be classified as follows according to the extent to which it is ordered [181,182,183].

1. **Buried, internal water**

These waters are completely surrounded by the protein and are inaccessible to the bulk solvent. They may exist individually or in small groups and their total number tends to increase with increasing protein size [132]. They have long lifetimes of the order of seconds. They may coordinate metal ions or they may hydrogen-bond to buried hydrophilic groups in the protein bridging peptide bonds or mediating salt bridges [184]. These waters can thus stabilise the protein and can in many cases be considered as an integral part of the protein structure. They may also have a catalytic function in enzymes.

2. **Ordered, surface water**

These waters may be bound in clefts and crevices of the protein structure or on the surface of the protein where they can form hydrogen-bonds to charged and hydrophilic polar groups of the protein. They are particularly found binding at turns which are generally the most hydrophilic parts of proteins and they may act to stabilise these turns [185]. Ordered waters may extend out to the second hydration shell where they may be strongly hydrogen-bonded to well-ordered waters which are in turn hydrogen-bonded to the protein. The ordered waters may be further subdivided [171] into those which are strongly bound making multiple hydrogen-bonds with the protein and those which are weakly bound making one hydrogen-bond or not hydrogen-bonding to the protein at all. Ordered waters are able to exchange with both the solvent and the protein with lifetimes of the order of 10^{-6} to 10^{-10}s

[182]. In high resolution (<2Å) electron density maps, ordered water molecules can be distinguished by discrete peaks and their contribution to the high angle region of the x-ray diffraction pattern [171].

The peptide bond in proteins is very hydrophilic [186] and so waters are observed hydrogen-bonding to main-chain carbonyl oxygens at regular intervals along α-helices [132]. However, side chain atoms make more hydrogen-bonds to water than main-chain atoms because of their greater solvent accessible area [132]. The affinity of the different amino acid side chains for water is variable and can be expressed quantitatively in terms of a 'hydration potential' [187].

Distinct patterns are seen in the structure of ordered water. For example, in crambin, water has been observed [188] to form pentagonal arrays around a hydrophobic region of the protein. This 'cage-like' structure may arise because of the hydrophobic effect [14]. Over hydrophilic regions, the water molecules form chains connecting polar groups on the protein [188]. Such water structures are also observed in carbohydrates and nucleic acids [127,181].

3. Disordered water

Order has not been detected in the majority of water molecules present in protein crystals [171]. Disordered waters may be present around exposed hydrophobic groups as well as at some distance from the protein. They have similar properties to bulk liquid water with lifetimes of the order of 10^{-12}s [182]. In x-ray diffraction maps, they produce a featureless electron density contributing only to the low angle (>4.5Å) region of the diffraction pattern. They may be

modelled as a single-valued smooth electron density continuum [171,189].

The divisions between these classifications are flexible and very dependent on the method by which the solvent is studied [179,180].

4.1.3 The theoretical study of the water around biological molecules

Many models of water have been proposed each of which is able to reproduce some but not all of its experimentally observed properties [183,190,191]. Liquid water may be represented as a continuum with an effective dielectric constant or as a set of discrete molecules. The former representation does not explicitly consider solute hydrogen-bonding effects and water-water interactions. Some of the models of the water molecule used in the latter representation are outlined below.

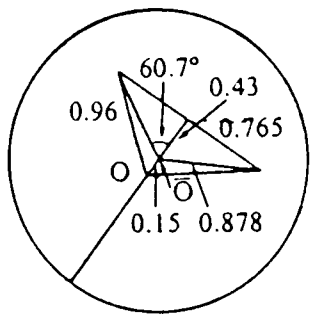
4.1.3.1 Models of the water molecule

1. Pair additive rigid body

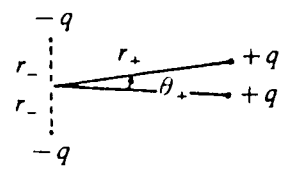
a. Three-point charge

The first model of water was introduced in 1933 [192] and had a positive charge on each hydrogen and a negative charge on the bisector of the HOH angle as shown in Figure 4.1.a. It was modelled to reproduce the experimentally observed dipole moment of water. There have since been many modifications of this model including the SPC [193], TIPS [194,195] and RWK [196] models.

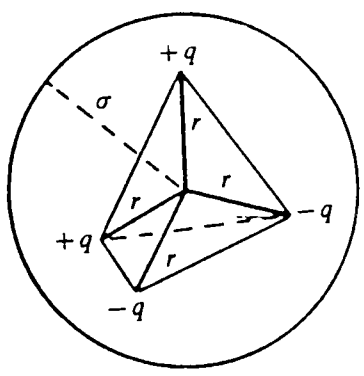
a.



b.



c.



d.

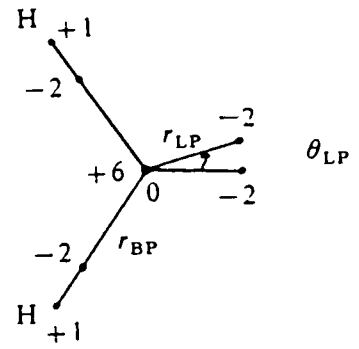


Figure 4.1 Models of the water molecule. See Section 4.1.3.1.
a. Bernal-Fowler model (q_O=-0.49e) [192]
b. Rowlinson model (q=0.33e, r_-=0.25Å, r_+=0.96Å, Θ_+=104.5°) [198]
c. BNS model (q=0.2e, r=1Å, σ=2.82Å) [199]
d. EPEN model (r_{LP}=0.272Å, r_{BP}=0.5672Å, r_{OH}=0.9572Å, Θ_{LP}=104.5°) [201,202].

The Configuration Interaction (CI) model [197] is also a three-point model. It employs an analytical potential function composed of exponential and Coulomb terms which is parameterised to fit high-level quantum mechanical calculations of the interaction energy of the water dimer.

b. Four-point charge

The first model of this type is the Rowlinson model [198] in which the two hydrogen atoms and the two lone pairs are assigned equal and opposite charges as shown in Figure 4.1.b. The lone-pairs are positioned 0.25Å above and below the molecular plane so as to reproduce the experimental quadrupole moment. Modifications of this model are the strictly tetrahedral Ben-Naim-Stillinger (BNS) model [199] with the lone-pairs and hydrogens 1Å from the oxygen (Figure 4.1.c) and the ST2 model [200] which places less emphasis on the tetrahedral shape of water with the oxygen to lone-pair distance at 0.8Å.

c. Seven-point charge

In the EPEN model [201,202], the water has seven charge centres as shown in Figure 4.1.d and is described by nine parameters, used in an effective pair potential, which were determined by fitting to either experimental data or quantum mechanical water dimer energies.

d. Extended Atom

For solvent accessibility calculations [58], water is treated as a sphere of radius 1.4Å. In molecular dynamics simulations

[203], water may be represented as a neutral spherical atom of van der Waals' radius 1.7Å and mass 18.0154 a.u. In program GRID [66], the water probe is modelled as an extended atom. Target waters are also modelled as extended atoms unless the coordinates of the water hydrogens are known, in which case a three-point atom-centred charge model may be used (see Section 4.2.1 for further details).

2. **Pair additive non-rigid body**

The Central Force (CF) model [204] allows intramolecular motion by treating the oxygen and hydrogen atoms separately and assigning point charges to the three atom centres. The geometry of the water molecule is maintained by a strong O-H potential and an H-H potential with a local minimum at the intramolecular H-H distance.

3. **Non-pair additive model**

Several models incorporating polarizability have been proposed [205,206,207,208]. In the Polarizable Electropole (PE) model [207,208], the charge distribution of the water molecule is represented as a multipole expansion containing the experimental dipole moment, the quantum mechanical quadrupoles and the dipole polarizability, and allowing the dipole moment to change under the influence of the surrounding field. The potential also contains repulsion and dispersion terms and the hydrogen atoms are not modelled explicitly. Non-bonded parameters are fitted to experimental data.

4.1.3.2 Theoretical methods of studying water interactions

The principal methods employed to study water interactions are as follows.

1. **Monte-Carlo Simulations**

The first Monte-Carlo simulation of water was performed by Barker and Watts in 1969 [209] using the Rowlinson model of water. Since then, many more simulations have been undertaken [175,210]. In the Metropolis algorithm [211], a starting configuration of atoms is chosen and new configurations generated by the random movement of a randomly chosen solvent molecule. At each configuration, the total potential energy is calculated and compared to that of the previous configuration. If the new energy is favourable then the new configuration is adopted, otherwise the Boltzmann Factor is examined to decide if the new or old configuration is accepted. The calculation is continued until an equilibrium configuration is obtained.

2. **Molecular Dynamics**

This was performed in 1971 for bulk water represented by a box of 216 waters using an effective pair potential [178]. In a simulation of a solvated molecule [203], a starting configuration of solvent atoms on a cubic lattice in a rectangular box with a density of $1g/cm^3$ is created. The configuration is randomised and then the solvated molecule is positioned in the box. The energy of the system is minimized and the dynamics of the system are then studied by integrating the classical equations of motion for all the atoms over a finite period of time.

3. Thermodynamic solvation energy

A method for evaluating the free-energy of solvation of a protein has been developed [212]. Atomic Solvation Parameters (ASP) are defined for each atom and the free energy δG calculated as

$$\delta G = \Sigma\ \delta S.A. \times ASP$$

where δS.A. is the change in atomic solvent accessible area [58] on protein folding or ligand binding. ASP are derived for five classes of atom, (C, N/O, O^-, N^+ and S), by fitting to the experimental free energies of transfer of amino acids and calculating atomic solvent accessible areas for standard residue conformations.

5. Steric and energetic location of water binding sites

The interaction energy of one water molecule with a solute molecule can be calculated and displayed graphically enabling strong water binding sites to be located. Ab initio and semi-empirical molecular orbital methods may be used for small solutes [213,214]. The solute with the identified tightly bound waters can then be treated as a 'supermolecule' in the calculation of conformational energy maps. Alternatively, the interaction of the solute with a water molecule can be calculated using a classical molecular mechanics energy potential. This approach is applicable to large solute molecules and is used in the GRID method as described in the next section.

4.2 METHODS OF STUDYING WATER USING PROGRAM GRID

4.2.1 The model of the water molecule used in program GRID

In the GRID method, water may be treated as a probe or, if it is well-ordered, as part of the target.

In GRID2, a water probe is modelled as an uncharged 'extended atom' with a van der Waals' radius of 1.7Å and a hydrogen-bonding radius of 1.4Å. The directionality of its hydrogen-bonds is accounted for by the hydrogen-bond potential function described in Section 3.7.1.1.

A target water molecule is generally modelled as an 'extended atom' with no directional component in its hydrogen-bonding potential. However, if the orientation of its hydrogen atoms is defined, it may be represented by a three-point atom-centred charge model with a charge of -0.25e on the oxygen and the directionality of its hydrogen-bonds considered as described in Section 3.4.3.1.5. The parameters for the 'extended atom' model of water are listed in Appendix III.5.

4.2.2 Prediction methods using program GRID

When used in the design of therapeutic agents, program GRID is generally run with a target consisting of one molecule in solution. However, when testing the GRID method by assessing the agreement of predictions of water binding sites with crystallographic observations, it is necessary to take into account the factors influencing a water molecule in a crystal. For this reason, two methods of predicting

water binding sites with program GRID were developed: Prediction Method 1 in which there is only one molecule in the target; and Prediction Method 2, which has been designed for the testing of the GRID method, in which the target is composed of the symmetry related molecules of a crystal.

These methods are carried out as follows:

Prediction Method 1

The interaction of a water probe with an isolated target molecule in solution is calculated by program GRID for positions of the probe throughout the target. An energy map is then displayed and favourable water binding sites identified.

Prediction Method 2

The experimentally observed waters in a target crystal structure may be predicted using a target consisting of all crystallographic symmetry related molecules surrounding the region of the crystal where the water prediction is to be made. The target may include waters observed in the crystal structure. Each observed water in the crystal structure is predicted individually, a temperature factor is calculated as shown in Appendix I, and the properties of the predicted water molecule are compared to those of the experimentally observed water, (see the next section for comparison methods). This method is carried out using the series of programs listed and described in Appendix III.4.

This second method enables the objective and quantitative assessment of the accuracy of the predictions. In particular, it does not rely on the potentially subjective visual examination of GRID energy maps using computer graphics.

4.3 THE ASSESSMENT OF THE PREDICTION OF WATER BY PROGRAM GRID

Criteria for assessing the quality of predictions made by program GRID were required. The properties examined for each water molecule predicted using Prediction Method 2 were:

1. The separation, δd, in distance between the observed and the predicted water position.

2. The energy E_{obs} calculated at the observed water position, the energy E_{pred} calculated at the predicted water position and the difference between these energies, $\delta E = E_{obs} - E_{pred}$. The energies E_{obs} and E_{pred} may be compared to the binding energy of the water dimer which has been found, by experimental and theoretical molecular orbital methods, to be about -5.5 kcal/mol [215].

3. The hydrogen-bonds to the probe at the observed and the predicted water position. The identity of the target atoms to which hydrogen-bonds are made at each position may be compared.

4. The thermal motion of the water at the observed and the predicted position. This is given experimentally by a temperature factor or B-value $B_{obs}=(8\pi^2/3)\langle u^2\rangle$ where $\langle u^2\rangle$ is the mean square displacement of the water from its mean rest position [216]. The temperature factor may contain contributions due to the static disorder of the crystal, the partial occupancy of atomic sites, and experimental and systematic errors, in addition to those due to thermal motion [217]. The static and dynamic contributions to the B-values can in principle be distinguished by determining the crystal structure at different temperatures [218]. The thermal motion includes contributions from the rigid-body motion of molecules and from

Table 4.1 The prediction of the water molecule in crystalline L-serine monohydrate by program GRID.

GRID version	Target hydrogen represention[a]	δd in Å	E_{pred} in kcal/mol	δE in kcal/mol	B_{pred} in Å^{2}[b]	Fourth hydrogen-bonding target atom[c]	Energy of fourth hydrogen-bond in kcal/mol
GRID0	-H	0.26	-13.29	0.53	1.86	N 8	-0.20
GRID1	+Hr	0.15	-14.81	0.34	1.99	OXT 62	-1.47
GRID2	-H	0.31	-14.35	0.57	1.86	OXT 32	-1.42
GRID2	+Hf	0.14	-14.43	0.42	1.91	OW1 8	-0.74
GRID2	+H	0.18	-13.75	0.45	1.82	-	-

[a] There are several ways of treating target hydrogen atoms in program GRID:
-H Only hydrogen-bonding hydrogen positions are determined and these are used to calculate E_{hb}. No Lennard-Jones or electrostatic interactions are calculated for hydrogen atoms.
+Hr All observed hydrogen atoms except those on target waters are represented explicitly and their van der Waals' and electrostatic interactions calculated. The hydrogens are allowed to rotate on hydroxyl groups.
+Hf As +Hr except hydrogens on hydroxyl groups are fixed at their experimentally observed coordinates.
+H All observed hydrogen atoms including those on target waters are explicitly represented and their van der Waals' and electrostatic interactions calculated. Their positions are fixed at the experimentally observed coordinates.

[b] B_{pred} is the predicted B-value. The observed B-value B_{obs}=3.08Å^{2} ($\langle u^2 \rangle^{\frac{1}{2}}$=0.34Å). This isotropic B-value was derived from the experimental anisotropic temperature factors. The conversion was made for comparison purposes because for most large molecules, only isotropic B-values are available.

[c] The water is experimentally observed to make three strong hydrogen-bonds: two donated to O 8 at 2.81Å and OG 33 at 2.88Å, and one accepted from N 6 at 2.92Å (see Figure 4.2). All versions of program GRID predicted hydrogen-bonds to these atoms. The prediction of a fourth weaker hydrogen-bond and the identity of the target atom to which it was made depended on the version of program GRID used. See Section 4.4.2 for details.
The atom number denotes the symmetry related molecule:
6=(0.5-x,-y,0.5+z), 8=(0.5-x,-y,-0.5+z), 32=(-0.5+x,0.5-y,-z), 33=(-0.5+x,0.5-y,1-z), 62=(1-x,-0.5+y,0.5-z)

intramolecular vibrations. The latter increase with increasing molecular size [217].

A B-value, B_{pred}, may be calculated for the water molecule from the shape of the energy well predicted by program GRID as shown in Appendix I.

4.4 THE TEST OF THE PREDICTION AND ASSESSMENT METHODS ON L-SERINE MONOHYDRATE

Prediction Method 2 was tested by using it to predict the water binding site in crystalline L-serine monohydrate before going on to use it to predict the water structure of crystals of macromolecules. L-serine monohydrate was chosen as a well-defined test system because its crystal structure has been determined both by x-ray and neutron diffraction to high resolution [219] and it has been used previously as a test system for Monte-Carlo simulations [220].

The crystal structure coordinates were obtained from the Cambridge Structural Database [131] and predictions were made with the three versions of program GRID described in Appendix III. The results are given in Table 4.1.

4.4.1 The test of the prediction method with GRID0

Only version GRID0 was available when these results were used to decide whether the methods of prediction and assessment were appropriate. GRID0 predicted the water molecule with reasonable accuracy 0.26Å from its experimentally observed position. The three experimentally observed hydrogen-bonds were predicted. The calculated

B-value of 1.86Å² was of the correct order of magnitude. The predicted B-value would be expected to be smaller than the observed B-value of 3.08Å² which may contain contributions due to static disorder and experimental error in addition to those due to thermal motion. It will also be dependent on the mobility of the amino-acid atoms which was neglected in the calculations.

These results compare favourably with those obtained by Monte Carlo simulations [220] which had an average value of δd=0.43Å and of B=0.95Å² and were also obtained assuming that the amino acid atoms were stationary.

The prediction made with GRID0 was of adequate accuracy to suggest that Prediction Method 2 may be suitable for testing program GRID and could be applied to the prediction of the water structure of crystals of macromolecules.

4.4.2 The improvement of the predictions with GRID1 and GRID2

Subsequently, predictions were made with versions GRID1 and GRID2 (see Table 4.1). The differences in the predictions with these versions of program GRID were principally due to differences in the treatment of target hydrogen atoms and in the determination of hydrogen-bonds (see Appendix III.3 for the differences between the three versions of GRID).

GRID0 predicted a weak hydrogen-bond of energy -0.20 kcal/mol to be donated to the water probe by N 8. This hydrogen-bond was not observed experimentally [219]. Subsequent versions of program GRID did not predict this hydrogen-bond for two reasons. Firstly, a cutoff energy of -0.25 kcal/mol was applied to weak hydrogen-bonds.

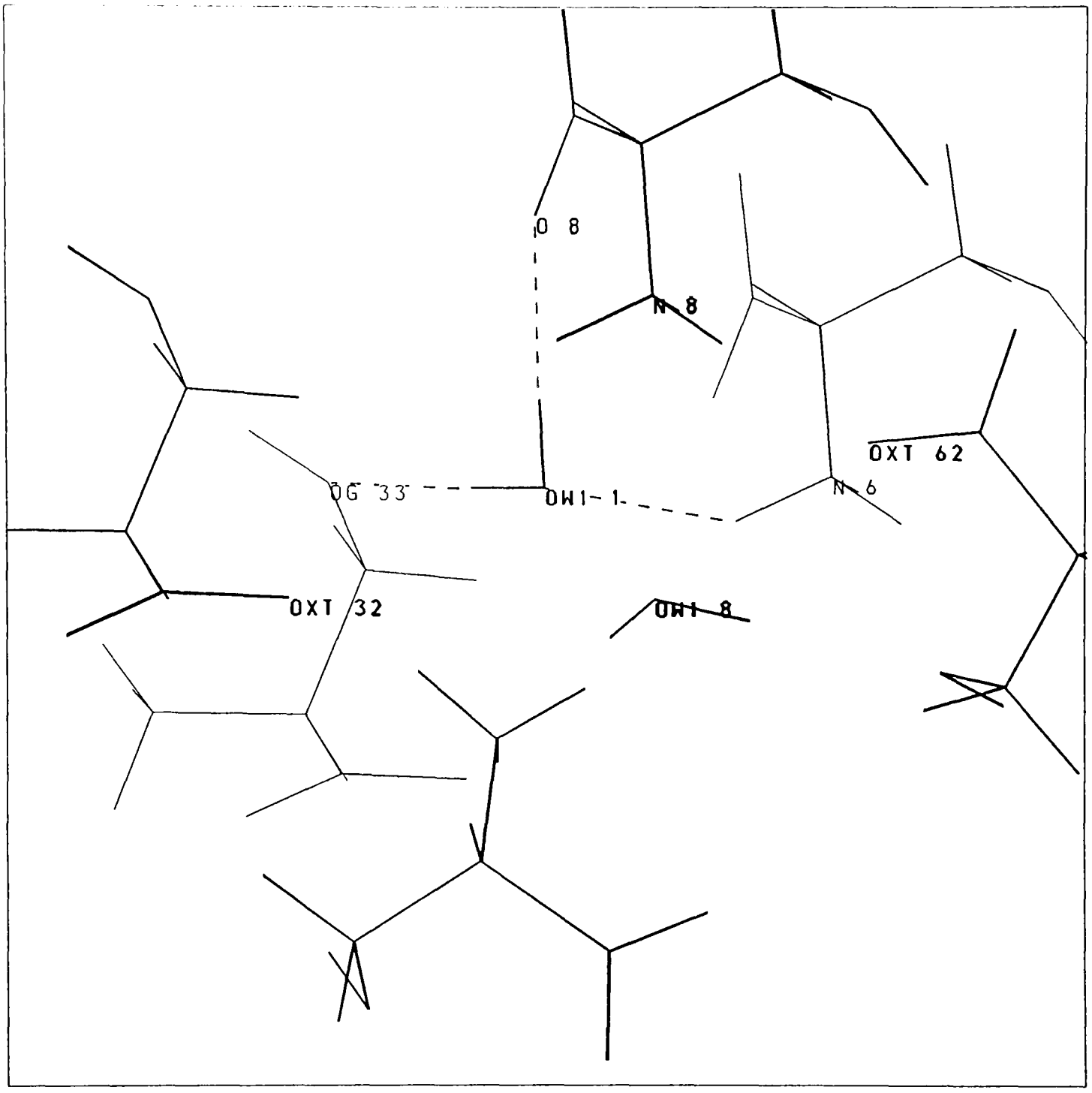


Figure 4.2 The region of the L-serine monohydrate crystal surrounding the predicted water OW1 1. Strong hydrogen-bonds, shown by dashed lines, were predicted by program GRID to O 8, OG 33 and N 6 and were observed experimentally. A fourth weak hydrogen-bond was predicted to N 8, OXT 62, OXT 32 or OW1 8 according to the version of program GRID used (see Table 4.1). See Section 4.4.2 for details.

Secondly, the algorithm used in GRID0 for selecting the hydrogen-bond made by a target atom that has the capacity to both accept and donate a hydrogen-bond always predicted a hydrogen-bond to be accepted from the probe. Thus the water probe was predicted by GRID0 to donate a hydrogen-bond to OG 33 and accept a fourth hydrogen-bond from N 8 despite the presence of oxygen atoms in the target which could accept stronger hydrogen-bonds than that made to N 8. In GRID1 and GRID2, this algorithm was improved to allow the hydrogen-bond formed by such target atoms to be selected on energetic criteria.

In GRID1, a fourth hydrogen-bond was predicted to OXT 62. However, experimentally, OXT 62 was observed to accept hydrogen-bonds from N 6 and N 8 and not to make a hydrogen-bond to the predicted water OW1 1 (see Figure 4.2). The angle subtended by N 6 and OXT 62 at water OW1 1 is only 48° so the simultaneous formation of hydrogen-bonds to both these target atoms is unfavourable. GRID1 did not take into account the geometry of the hydrogen-bonds made to the probe and this prediction clearly shows its importance.

In GRID2, geometrical constraints were applied to the hydrogen-bonds to the probe and so a hydrogen-bond to OXT 62 was no longer predicted. The prediction of a fourth hydrogen-bond by GRID2 was dependent on the treatment of the target hydrogen atoms.

If they were allowed to rotate on hydroxyl groups, OG 33 was predicted to donate a hydrogen-bond to the water and a fourth hydrogen-bond was accepted from OXT 32. This prediction was in disagreement with experimental observations in which water donates a hydrogen-bond to OG 33 and OXT 32 accepts hydrogen-bonds from N 1 and N 3 (see Figure 4.2). The hydrogen atom of OG 33 would be unlikely to rotate because it was experimentally observed to donate a hydrogen-

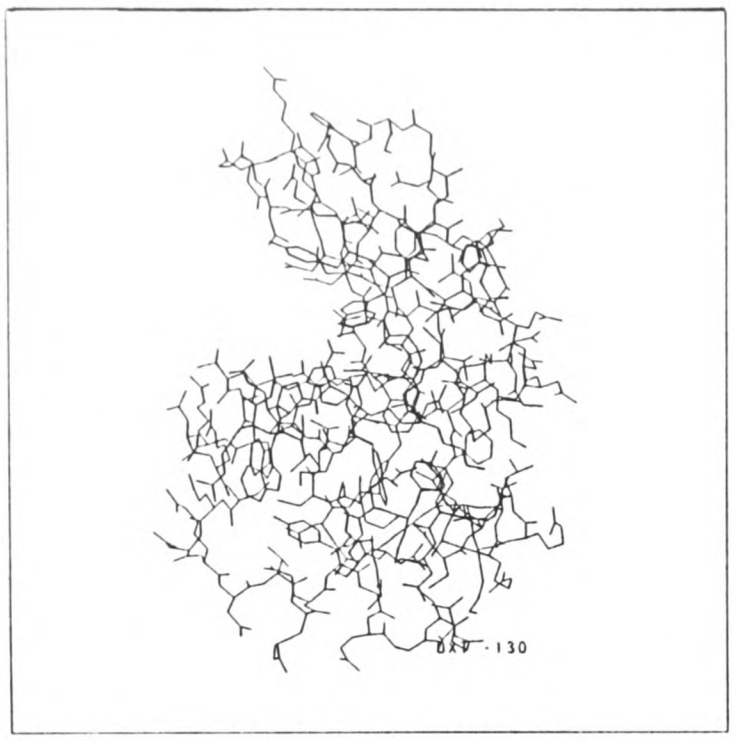
bond to a serine oxygen and therefore, it may be reasonable to fix the positions of the hydroxyl hydrogens in this target.

When the hydroxyl hydrogens were fixed in position, GRID2 correctly predicted OG 33 to accept a hydrogen-bond from the water. It now predicted a fourth hydrogen-bond to be donated by water OW1 8 which was assumed to be able to rotate. However, this hydrogen-bond was not observed experimentally (see Figure 4.2).

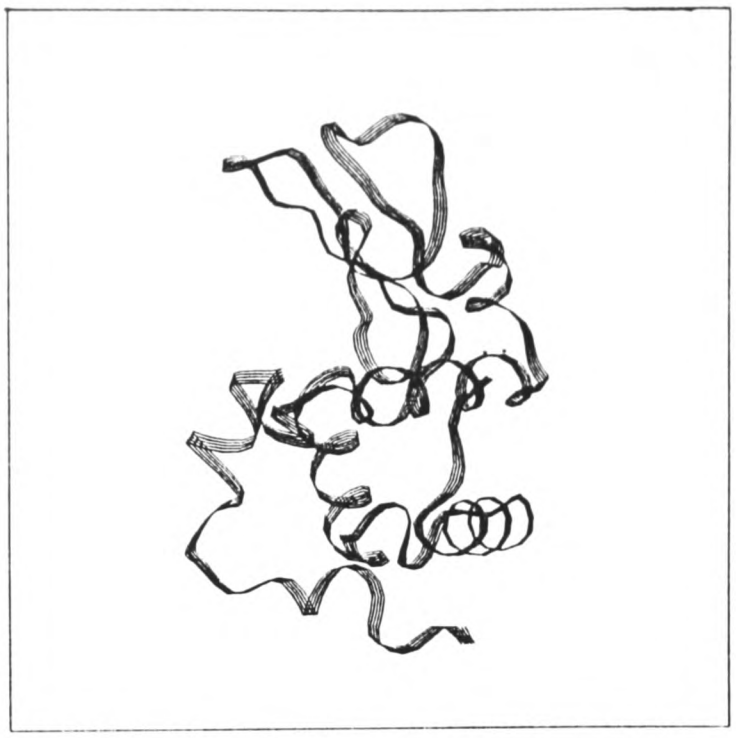
If all target waters were also assumed to be unable to rotate and all target hydrogen atoms were fixed at their experimentally observed positions, the water was predicted to make only three hydrogen-bonds in accordance with experimental observations. It was predicted close to its observed position at a distance of 0.18Å.

These three predictions with GRID2 show the need to take account of the hydrogen-bonds between target atoms when determining the hydrogen-bonds that can be formed to the probe. Program GRID was able to detect these hydrogen-bonds in the L-serine monohydrate crystal if the target hydrogen coordinates were fixed.

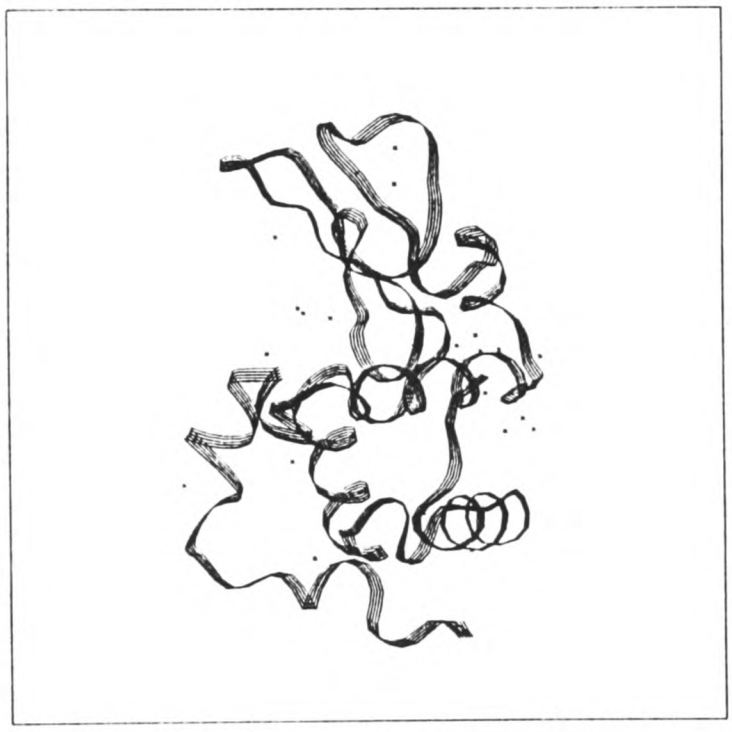
The examination of the predictions of the water in L-serine monohydrate has clearly shown the improvement in predictions obtained when the geometric properties of hydrogen-bonds are taken into account by program GRID as described in Chapter 3.



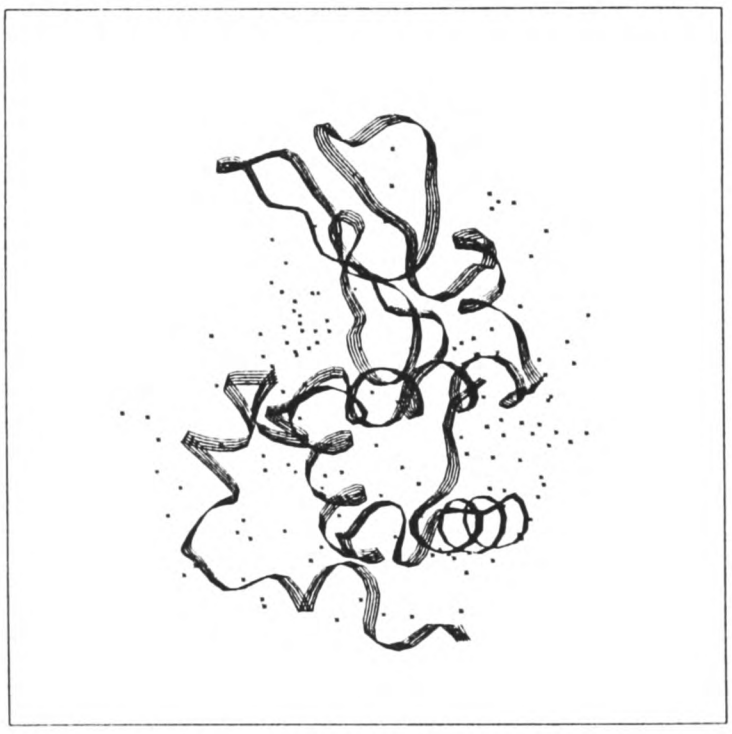
a.



b.



c.



d.

Figure 4.3 a. Human lysozyme (HL) showing the lower helical domain and the upper ß-sheet domain with the active site cleft in between. The domain-linking helix runs horizontally across the molecule. b. A ribbon plot of HL with the positions of the four internal water molecules shown. c. HL with the positions of the 27 water sites conserved in HL, HEWL and TEWL shown. Six are in the active site cleft. d. HL with the positions of the 132 observed ordered water molecules shown.

4.5 STUDIES OF THE WATER IN CRYSTALS OF HUMAN LYSOZYME (HL)

4.5.1 The selection of HL for study

The water structure of crystalline Human Lysozyme (HL) was chosen as a model system for testing the GRID method for several reasons. It has been observed by x-ray crystallography at 1.5Å resolution with an R-factor of 18.7% [115,171]. The structure of HL was determined in Oxford and so the original data and electron density maps were available. Consequently the predicted water positions could be compared not only with the coordinates assigned to the observed waters but also with the electron density map from which their positions were determined. In addition, the water structure of crystals of HL has been analyzed in detail [171] and has been found to contain features typical of globular proteins.

4.5.2 Human lysozyme and the water structure of its crystals

Lysozyme is a hydrolase found in most bodily secretions which destroys certain bacteria by cleaving the mucopolysaccharide chains forming part of the cell wall. It is thought to bind a hexasaccharide substrate, (with sugar rings at subsites A-F), and catalyze the cleavage of the glycosidic link between the residues at sites D and E [221].

Human lysozyme (HL) has 130 residues and a molecular weight of about 14,500. It is a small ellipsoidal molecule with dimensions of about 30Å x 30Å x 45Å (see Figure 4.3.a) and an active site crevice dividing it into two lobes joined by a helix (residues 88-100). One

lobe is predominantly helical and hydrophobic (residues 1-40, 101-130), while the other one is irregular and mainly hydrophilic containing an anti-parallel ß-sheet (residues 41-87). In general, HL is polar on the exterior and non-polar on the interior.

Crystalline HL contains about 350 water molecules per protein molecule [171]. Four of these waters (see Figure 4.3.b) are buried in an internal cavity forming a semi-circular string of well-ordered waters around Ala 92. A further 128 ordered waters and three nitrate ions have been observed covering about 75% of the protein surface and extending out to a distance of 4.5Å (see Figure 4.3.d). There are 27 equivalent water sites present in the crystals of the hen egg white (HEWL), human (HL) and tortoise egg white (TEWL) lysozymes [222] and these are shown in Figure 4.3.c. Four of these are internal and six are in the active site cleft, (one of latter is occupied by the oxygen atom of a nitrate ion in HL). In the active site, many of the positions of the polar hydroxyl, carbonyl and amino groups of the hexasaccharide substrate are occupied by waters in the native protein. These waters are displaced on substrate binding. Other waters are in positions where they may mediate a protein-substrate hydrogen-bond e.g. 200/3. Water may be important in the catalytic action of the enzyme e.g. 204/5 hydrogen-bonds to the catalytic residue Glu 35 and may attack the glycosidic bond [171].

The remaining 60% of the waters have not been individually identified experimentally and may be disordered [171].

Previous attempts have been made to simulate the water structure around HEWL. One of these was by Monte-Carlo simulation with a Rowlinson model of water and a molecular mechanics' force field [223]. In this simulation, 49 of the 80 waters observed by x-ray

diffraction were found in the same 'environmental niche' as the simulated waters. Another simulation [224,225] made use of atom pair potentials derived from ab initio molecular orbital calculations and in this, half the waters observed crystallographically in triclinic HEWL were predicted to within 2.5Å of their observed positions.

4.5.3 Results of the predictions of the water structure of crystalline HL made with GRID0

The structure of water around HL was predicted with GRID0 using Prediction Method 1 by Holloway [226]. By searching the GRID energy map for water binding sites in the region around each experimentally observed water, he was able to predict 88 of the 132 observed water molecules correctly. A correctly predicted water site was defined as being small (so that the minimum energy position could be identified visually), low in energy (i.e. in the range -5 to -10 kcal/mol), and within 1.5Å of an observed water molecule.

Two causes of the absence of predicted water binding sites for some of the remaining water molecules were identified [226]. These were the neglection of water-water interactions, and the neglection of interactions with atoms in neighbouring lysozyme molecules in the crystal. In addition, the need for a quantitative way of assessing the quality of the prediction was identified. This lead to the development, in the present work, of Prediction Method 2 which was designed to overcome these problems.

The coordinates (HUMSY.PDB and HUMB.DAT) and electron density map (HUMAP.BRK) of HL [171] were used in this study.

Table 4.2 The assessment of the predictions by GRID1 and GRID2 of the waters in crystalline human lysozyme according to the separation between the observed and predicted positions δd.

Criterion	Number[a] of waters predicted by GRID1	GRID2[b]
$\delta d < 1.0$Å	56 (42)	64 (48)
$\delta d < \langle u^2 \rangle^{\frac{1}{2}}$	81 (61)	84 (64)
$\delta d < 1.5$Å[c]	87 (66)	85 (64)
$\delta d < 3.0$Å	125 (95)	128 (97)

[a] The total number of experimentally observed waters is 132. Percentages are given in parentheses.

[b] δd is smaller for GRID2 than GRID1 for 68 (52%) waters. The rms value of δd for all 132 ordered waters is 1.52Å for GRID1 and 1.28Å for GRID2.

[c] For GRID0, the number of waters with $\delta d < 1.5$Å was 88 (67%) [226]. For predictions with GRID0, less stringent criteria were applied to the determination of the position of the predicted water. The predicted water was required to have a binding energy greater than 5kcal/mol in magnitude and be located in an energy well but it was not required to be exactly at the position of minimum energy in the energy well as it was for the predictions made with GRID1 and GRID2.

Table 4.3 The assessment of the predictions by GRID1 and GRID2 of the waters in crystalline human lysozyme according to the energy calculated at the observed position E_{obs}.

Range of E_{obs} in kcal/mol[a]	Number[b] of waters predicted by GRID1	GRID2[b]
$E_{obs} > 0$	25 (19)	17 (13)
$-6 < E_{obs} < 0$	11 (8)	22 (17)
$E_{obs} < -6$	96 (73)	93 (70)

[a] The interaction energy for a water dimer has been experimentally observed and theoretically calculated to be about -5.5kcal/mol [215]. Thus if E_{obs} is positive the water is unfavourably positioned. If E_{obs} is negative but less than 6kcal/mol in magnitude then the water can be considered to bind weakly. If E_{obs} is less than -6 kcal/mol, the water is favourably positioned.

[b] The total number of experimentally observed waters is 132. Percentages are given in parentheses.

4.5.4 Results of predictions of the water structure of crystalline HL made with GRID1

A brief summary of the results analyzed according to the four properties described in Section 4.4 is given here.

1. **Distance δd**

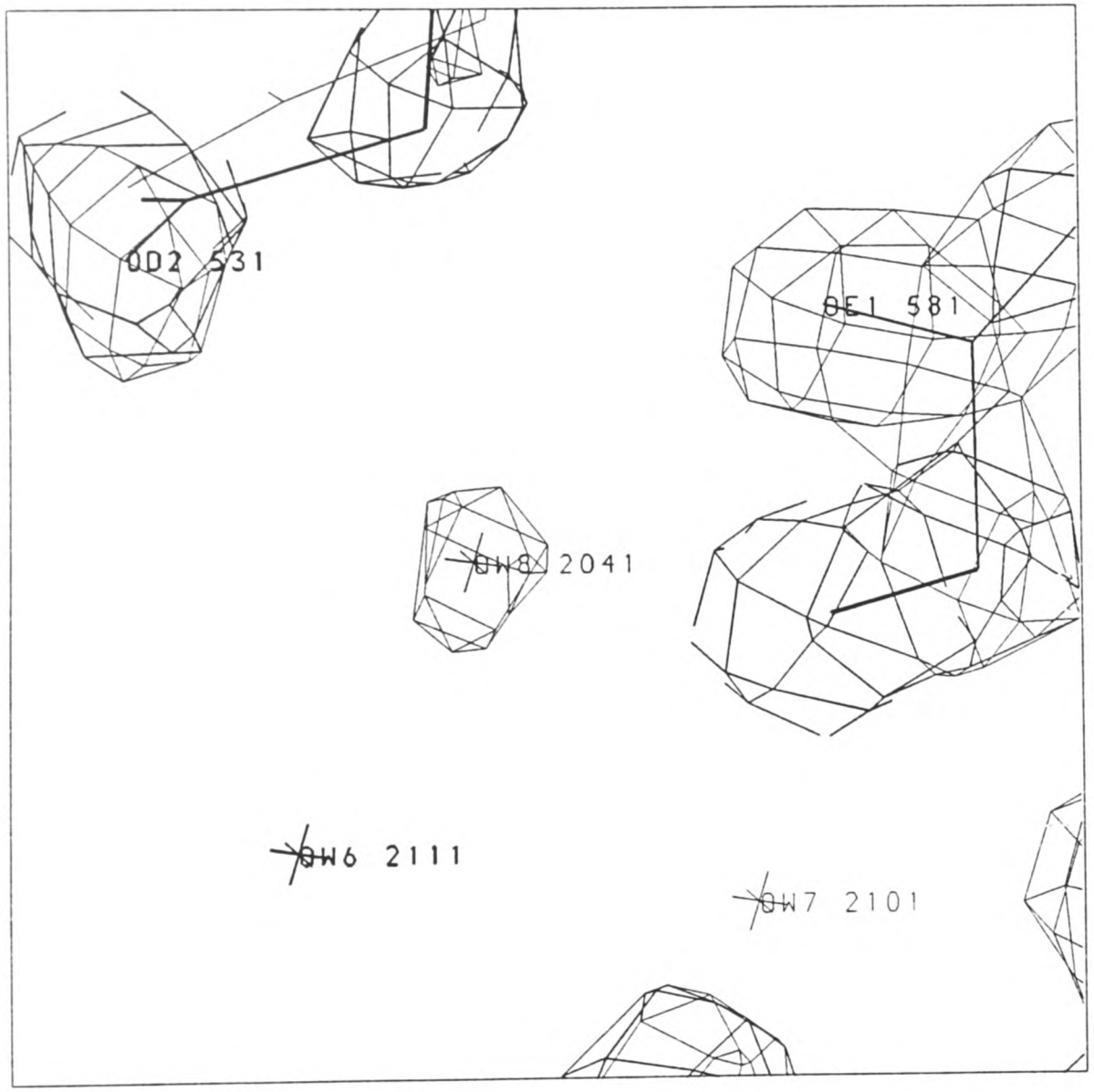
66% of the waters were predicted to within 1.5Å of the observed position and the rms value of δd for all 132 ordered waters was 1.52Å (see Table 4.2). Examples of well predicted waters are 200/1 which was predicted 0.09Å from its observed position, and 204/8, a water in the active site cleft, which was predicted 0.10Å from its observed position (see Figure 4.4).

2. **Energy**

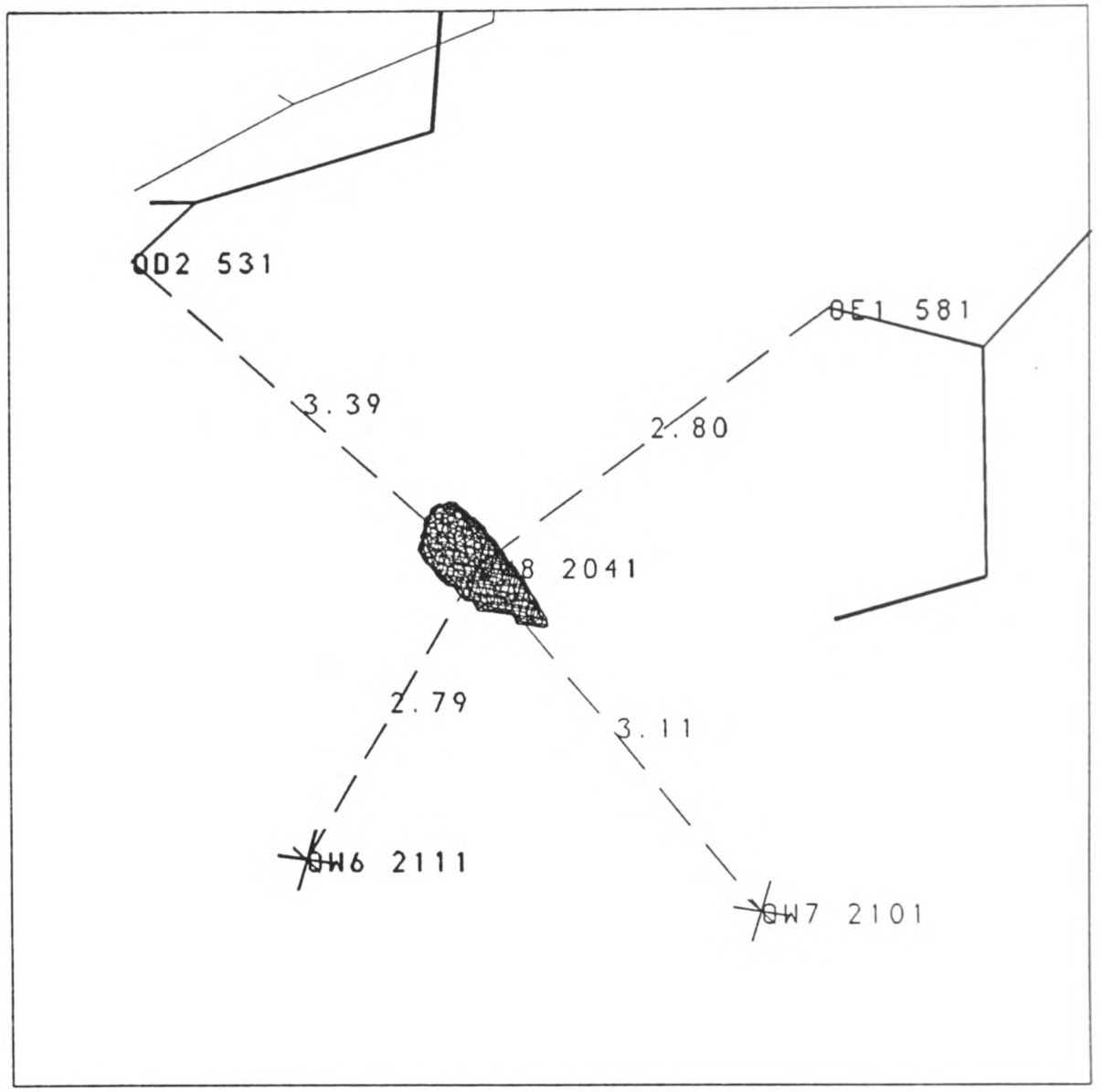
The distribution in the energies calculated at the observed water positions, E_{obs}, is given in Table 4.3. Those with $E_{obs}>0$ kcal/mol were found to be making repulsive contacts with an atom which had adopted a position in very close proximity to the observed position of the predicted water during crystallographic refinement.

3. **Hydrogen-bonds**

The target atoms to which hydrogen-bonds were calculated were the same at the observed and predicted positions of 33 of the waters. For the remaining waters, the difference in the hydrogen-bonds determined at the observed and predicted positions was mostly due to the prediction of more hydrogen-bonds, usually to other waters in the crystal structure, at the predicted water position than at the



a.



b.

Figure 4.4 The region of human lysozyme surrounding the observed position of water 204/8. a. Part of the electron density map with contours at 0.8eÅ^{-3} showing the spherical contours defining the position of water 204/8. b. Energy contours at kT (0.6 kcal/mol) above the energy minimum at -16.3 kcal/mol predicted by program GRID 0.1Å from the experimentally observed position of water 204/8. Their ellipsoidal shape shows the anisotropy of the predicted energy well.

observed water position. This was due to the fact that no angular constraints were applied to hydrogen-bonds to the probe or to target water molecules in GRID1.

4. **Temperature factors**

The calculated B-values for water molecules in HL were small, averaging about 4Å^2 while the observed B-values ranged from 12 to 83Å^2. The calculated B-values were mostly very anisotropic giving rise to ellipsoidal contours as seen in Figure 4.4.b. This anisotropy was not reflected in the shape of the electron density contours around the waters which were approximately spherical in most cases (see Figure 4.4.a). The calculated B-value appeared to be unrelated to the observed B-value.

Negative B-values were calculated for seven waters. These appeared to be caused by the use of abrupt cutoffs in the modelling of some of the hydrogen-bond interactions.

This analysis showed the need for the modification of GRID1. In particular, the geometry of the hydrogen-bonds at the probe required consideration and the angular dependence of the hydrogen-bonds at the target molecules needed alteration. Modifications were therefore made to GRID1 resulting in version GRID2. These are outlined in Appendix III.3.

Three nitrate ions per HL molecule have been observed in the HL crystal because it was crystallised from 7M ammonium nitrate [171]. These were represented as waters in the predictions made with GRID1 but in order to predict them and the surrounding water

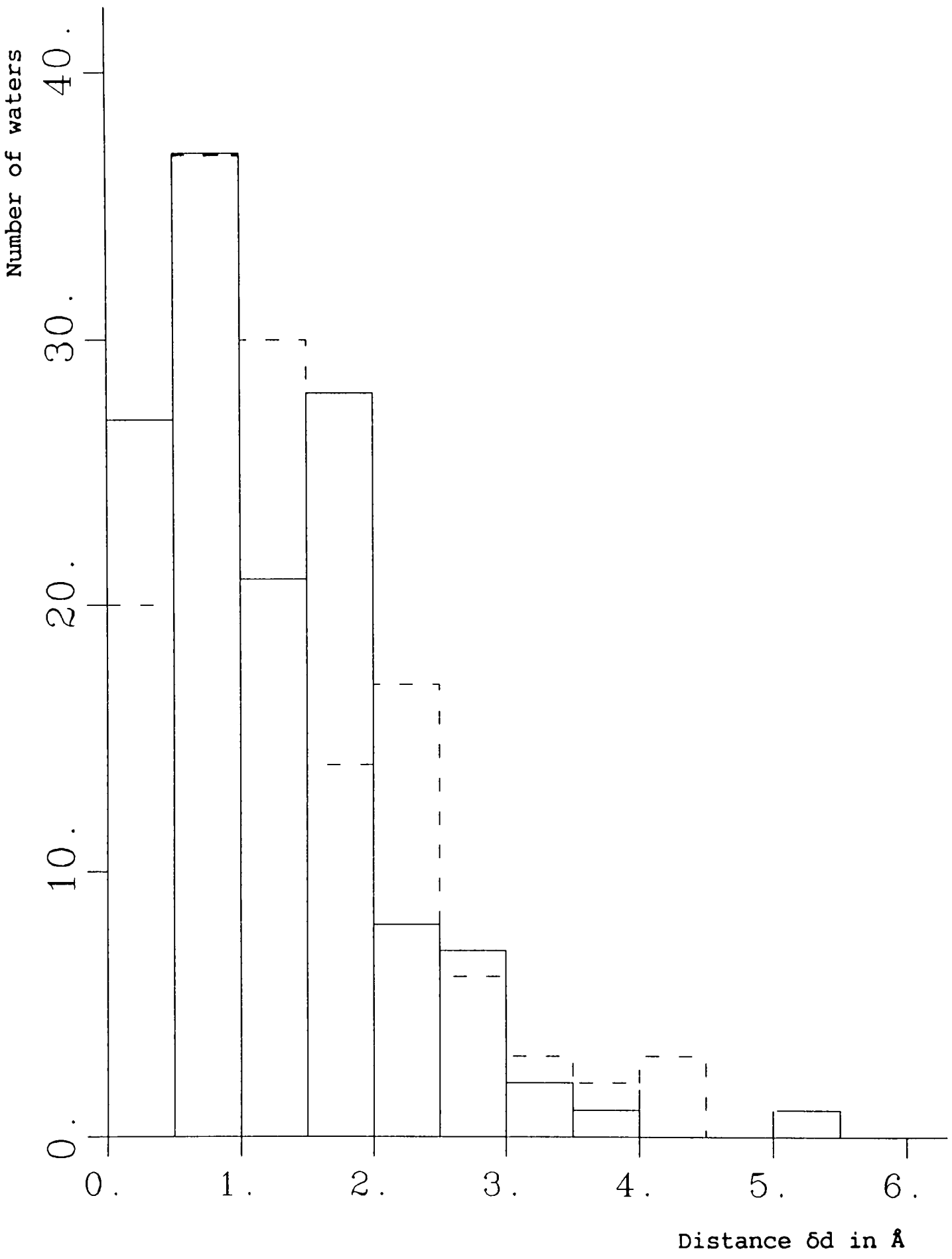


Figure 4.5 Histogram of the distance δd of the predicted positions of the waters in human lysozyme from their experimentally observed positions for predictions with GRID1 (dashed line) and GRID2 (full line). In the predictions made with GRID2 more waters had δd<0.5Å and fewer waters had δd>3.0Å than in the predictions made with GRID1.

correctly, it was necessary to derive appropriate parameters for them. These are listed in Appendix III.5.

The prediction of the water structure of crystals of HL was repeated with GRID2 and with the inclusion of the correctly parameterised nitrate ions.

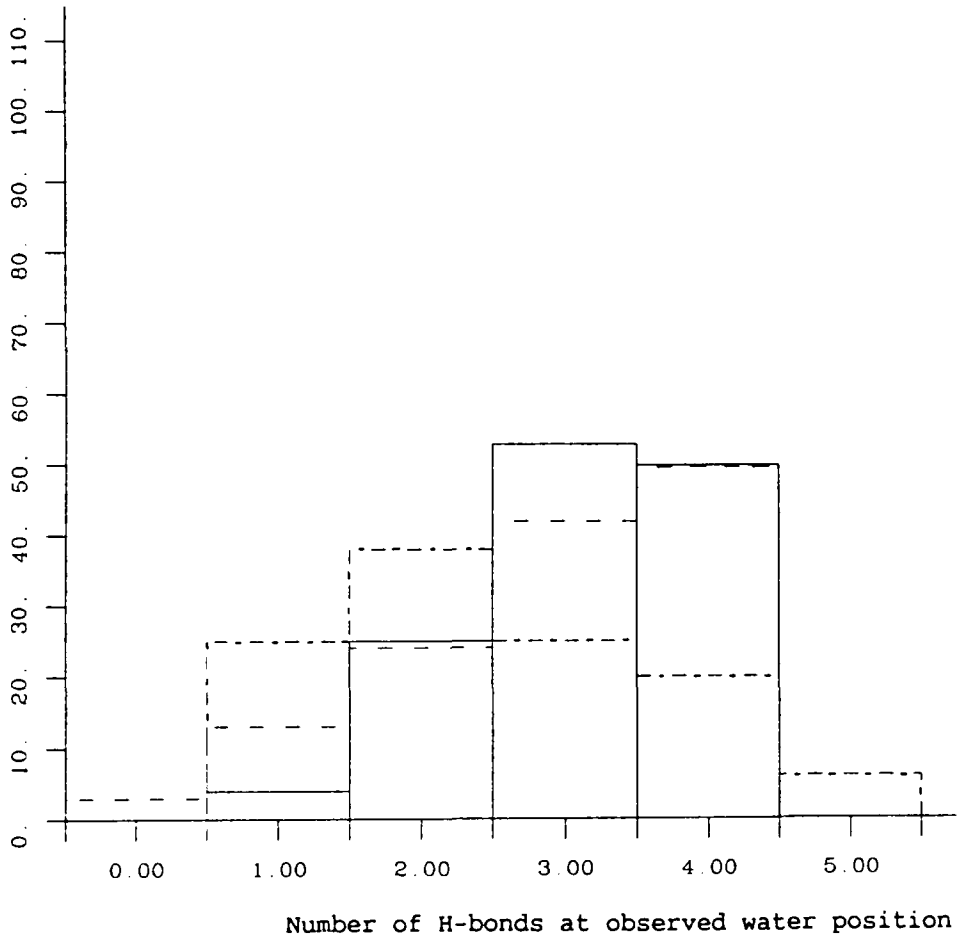
4.5.5 Results of the predictions of the water structure of crystalline HL made with GRID2

These were analyzed according to the four criteria previously described and they were compared to the predictions made with GRID1.

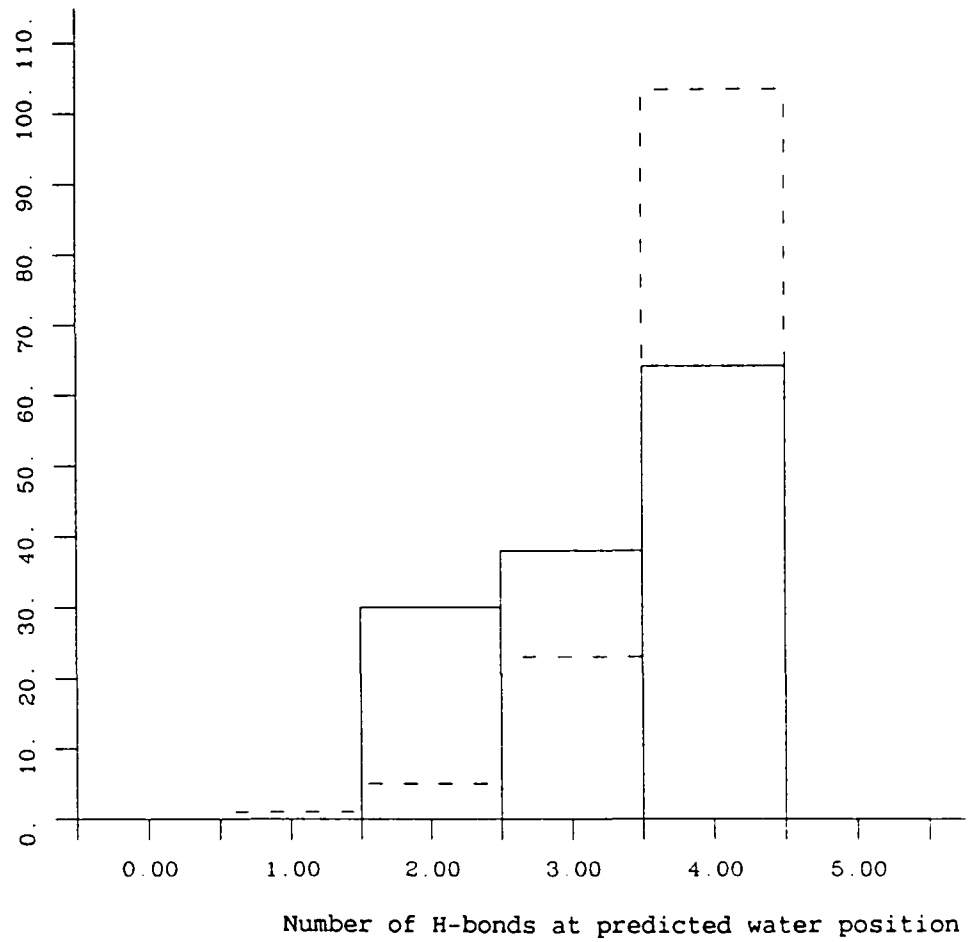
1. **Distance δd**

The distribution of δd is displayed in the histogram in Figure 4.5 for both sets of predictions. Different criteria can be used to assess the predictions on the basis of δd as shown in Table 4.2. Notably, with GRID2, more waters were predicted with δd<0.5Å and there were fewer for which the water site could not be predicted and δd>3Å. The rms value of δd was reduced to 1.28Å for the predictions of all the ordered waters with GRID2.

Program GRID was designed to locate sites on macromolecules at which ligands bind strongly and was therefore not expected to be able to predict the positions of the waters in the second hydration shell which are not tightly bound to the protein. Thus, the four waters with δd>3Å were observed to lie far out in the solvent channel making little contact with the protein and having high temperature factors in the range 60-82Å^2 (See Figure 4.13.b). Program GRID



a.



b.

Figure 4.6 Histograms of the number of hydrogen-bonds calculated to a. each experimentally observed and b. each predicted water position in human lysozyme. Results with GRID1 (dashed line) and GRID2 (full line) are shown. In a, the number of experimentally determined hydrogen-bonds [171] is also shown (dotted line).

predicted their observed position within a large shallow energy well but located far from the position of the energy minimum. Nonetheless, it is shown in Section 4.5.8. that program GRID may be of value in investigating waters which interact weakly with the protein.

2. **Energy**

The energies at the observed positions E_{obs} ranged from -16 to 231 kcal/mol and the distribution of E_{obs} is shown in Table 4.3. As for GRID1, very large repulsive energies were calculated for waters at positions very close to other atoms in the crystal structure. The energies at the predicted positions E_{pred} were all in the range -7 to -17 kcal/mol.

3. **Hydrogen-bonds**

The effect of applying constraints to the geometry of the hydrogen-bonds at the probe can be seen by examining the number of hydrogen-bonds made by the waters.

With GRID1, the number of predicted hydrogen-bonds was far greater at the predicted water positions than at the water observed positions (see Figure 4.6) and 78% of the waters were calculated to make four hydrogen-bonds at their predicted positions. With GRID2, this discrepancy between the number of hydrogen-bonds at the observed and predicted positions was reduced because the geometry of the hydrogen-bonds at the probe was taken into account (see Figure 4.6). There was still a tendency for more hydrogen-bonds to be found at the predicted position than at the observed position. This would be expected because, in order to find the most energetically favourable probe position, program GRID optimises the hydrogen-bonding

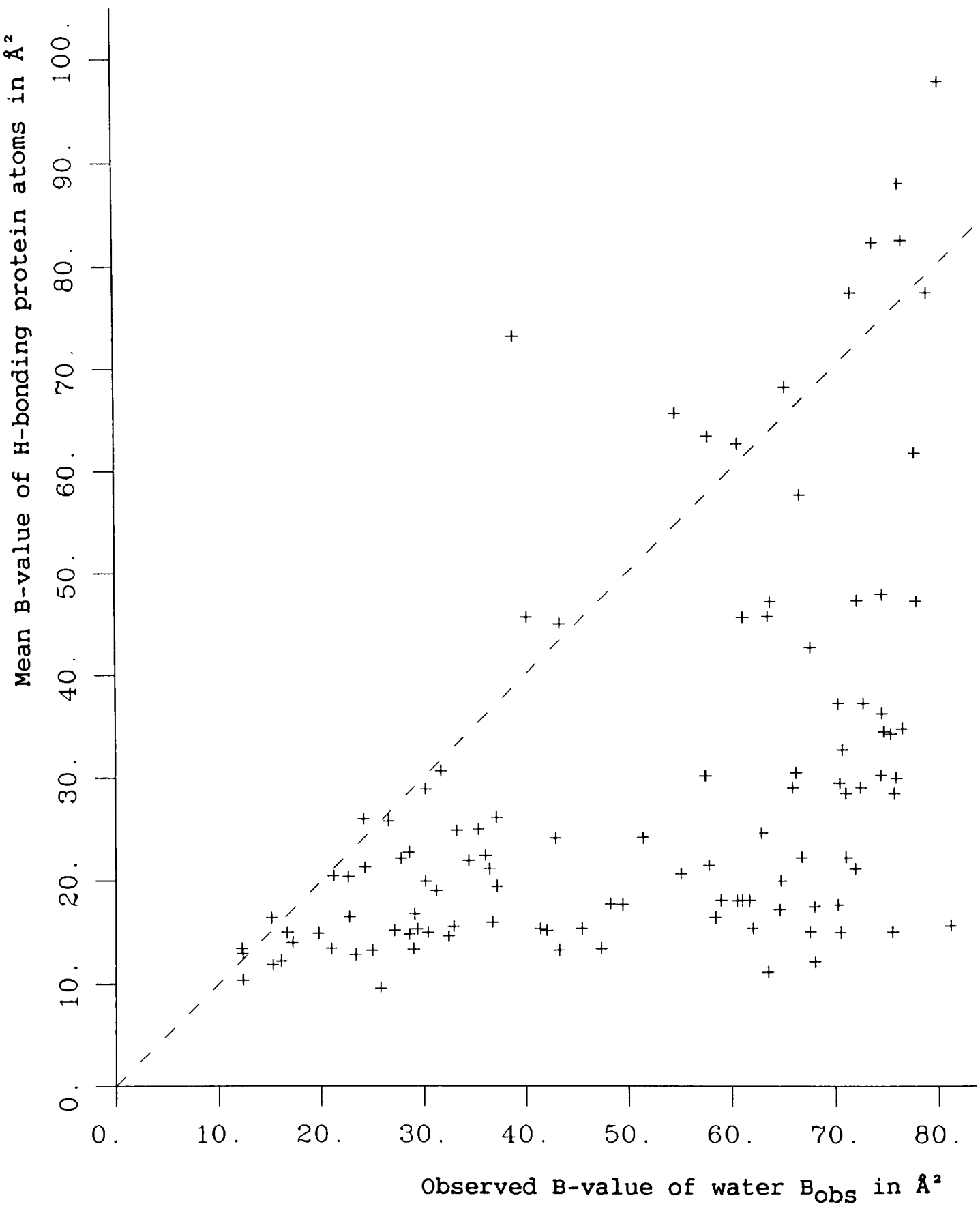


Figure 4.7 A plot of the mean temperature factors of the atoms in human lysozyme to which program GRID predicts that each water hydrogen-bonds when it is at the experimentally observed position against the temperature factors B_{obs} of the water molecules themselves. The dashed line is the line of equivalence. The waters have similar or greater temperature factors than the protein atoms to which they were calculated to hydrogen-bond.

interaction energy subject to the other constraints acting on the probe.

The number of hydrogen-bonds calculated for each water at the observed position was generally greater than that experimentally observed [171] (see Figure 4.6.a) because program GRID calculates all hydrogen-bonds with strengths greater than 0.25 kcal/mol. In the analysis [171] of the experimental observations, hydrogen-bonds were defined by having a length of 2.6 to 3.4Å, and this definition excluded the long, weak hydrogen-bonds calculated by program GRID.

4. Temperature factors

The calculated B-values of the predicted waters were, as in the GRID1 predictions, much smaller than the observed B-values and were very anisotropic. No relationship between B_{pred} and B_{obs} was identified. This may be because program GRID models all the target atoms except hydrogens as being stationary. The mobility of the waters has been found to be very dependent on that of the protein to which it is bound [171]. From the experimental data, waters making multiple hydrogen-bonds to the protein were found to have similar B-values to their protein neighbours while waters making a single hydrogen-bond to the protein were found to be more mobile than the protein atom to which they were hydrogen-bonded. The observed water B-value was of similar magnitude or greater than the B-values of the protein atoms predicted by program GRID to hydrogen-bond to it (see Figure 4.7). This suggests that the mobility of the ordered waters is very dependent on the mobility of the surrounding neighbours and that the ordered waters have little intrinsic motion of their own.

4.5.6 The predictions of the conserved waters with GRID2

Of the 26 conserved waters for which predictions were made, 24 were predicted within 1.5Å of their observed position and 22 were within $\langle u^2 \rangle^{\frac{1}{2}}$ of their observed position. They had an rms value of δd of 0.73Å. All these waters were found to have negative values of E_{obs}. All the waters in the active site were predicted with $\delta d < 0.8$Å and within $\langle u^2 \rangle^{\frac{1}{2}}$ of their observed position.

4.5.6.1 The internal waters

The internal waters lie in a chain with 200/8 furthest into the protein stabilizing a reverse turn [222] and 200/7 hydrogen-bonding to 200/4 leading out to the solvent. They are clearly defined in the electron density map [171] (see Figure 4.8.a). The predictions of the internal waters are given in Table 4.4 and shown in Figure 4.8.b. All the internal waters were predicted close to their observed positions within a distance of $\langle u^2 \rangle^{\frac{1}{2}}$ and with $\delta d < 0.7$Å. Program GRID predicted a consistent hydrogen-bonding pattern between the waters as shown below:

200/8 ‹— 200/6 ‹— 200/5 —› 200/7 —› 200/4

where —› represents the donation of a hydrogen-bond.

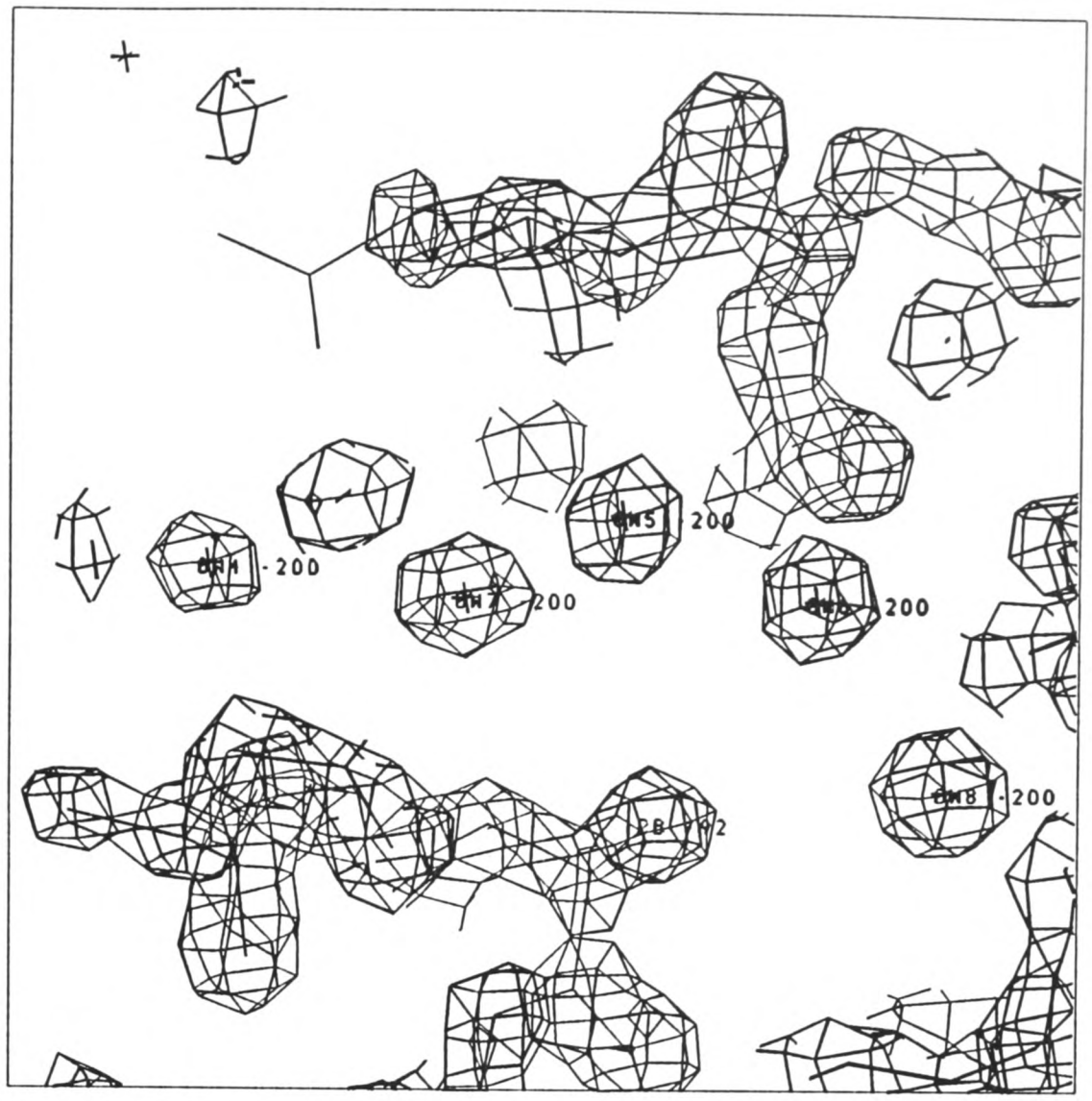
The predictions for these waters were similar to those made with GRID1. However, in three cases, four hydrogen-bonds were calculated to the water at the observed position by GRID1 and only three were calculated by GRID2. In addition, for all the waters, E_{pred} was of smaller magnitude than calculated with GRID1. These differences were due to the application of constraints on the geometry

Table 4.4 The predictions of the internal waters in crystalline human lysozyme by program GRID using Prediction Method 2.

Water	$\langle u^2 \rangle^{\frac{1}{2}}$ in Å	δd in Å	E_{pred} in kcal/mol	δE in kcal/mol	H-bond partners at observed position	H-bond partners at predicted position
200/8	0.68	0.36 (0.49)	-14.06 (-15.61)	0.76 (1.07)	O Tyr 54 N Phe 57 200/6 (O Leu 84)	O Tyr 54 N Phe 57 200/6 (O Leu 84)
200/6	0.78	0.51 (0.53)	-13.31 (-16.65)	3.98 (1.48)	O Leu 84 N Ile 56 200/5 (200/8)	O Leu 84 N Ile 56 200/5 200/8
200/5	0.76	0.66 (0.69)	-15.95 (-17.78)	1.22 (1.05)	O Gln 86 200/6 200/7 (O Asn 88)	O Gln 86 O Ala 83 200/6 200/7
200/7	0.68	0.27 (0.23)	-15.78 (-19.25)	0.78 (0.60)	O Ala 83 N Ala 92 200/4 200/5	O Ala 83 N Ala 92 200/4 200/5
200/4	0.90	0.34 (0.27)	-13.36 (-16.44)	1.80 (0.75)	OD1 Asn 88 N Asp 91 200/7 212/1	OD1 Asn 88 N Asp 91 200/7 212/1

Values in parentheses are for predictions made with GRID1. All hydrogen-bond partners were the same for both predictions except those given in parentheses to which hydrogen-bonds were predicted only by GRID1.

a.



b.

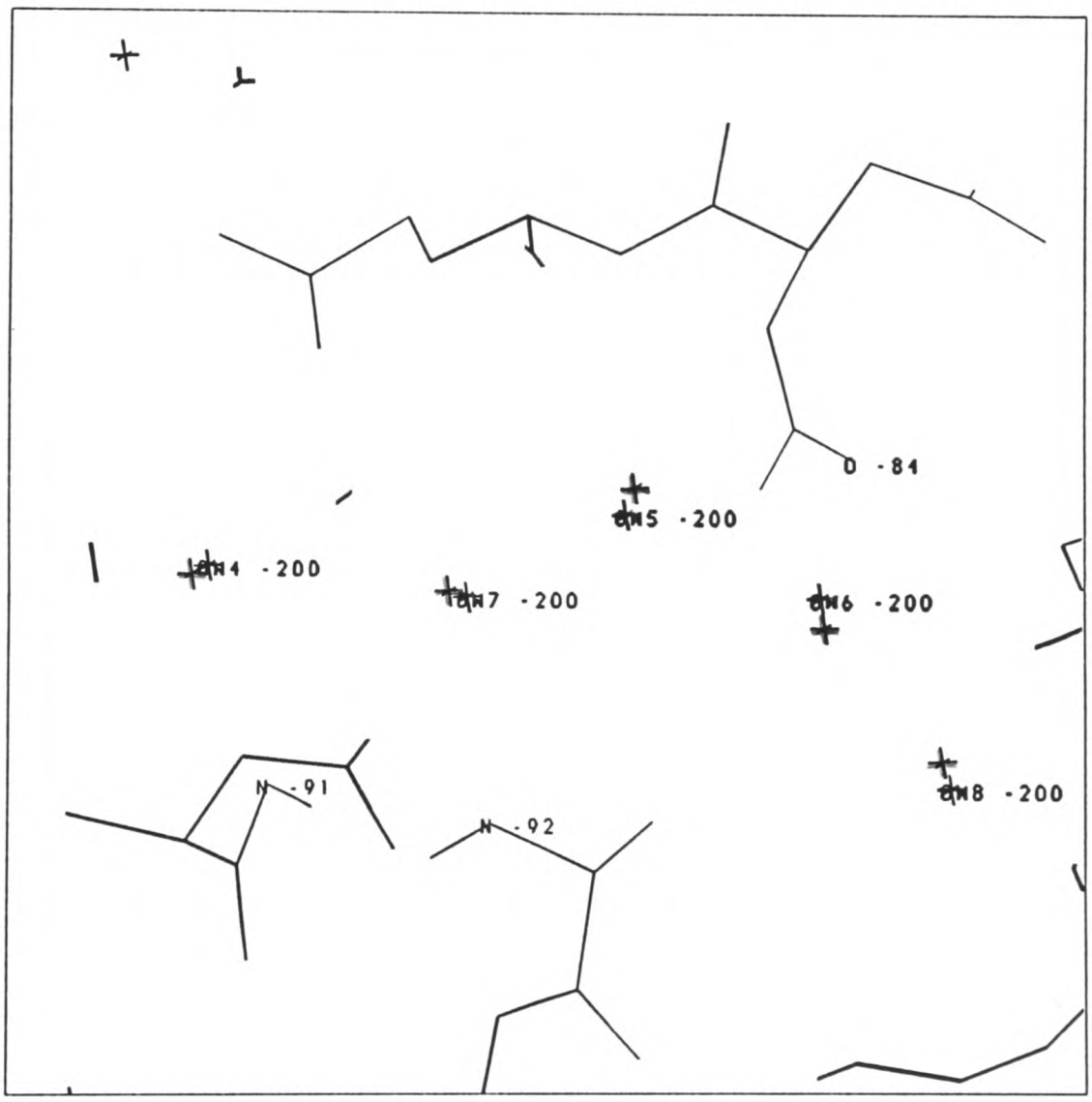


Figure 4.8 a. The electron density map of human lysozyme with contours at 0.8eÅ^{-3} showing the four internal waters 200/5-8 with 200/4 leading out to the solvent. The waters are grouped around Ala 92 and 200/8 is thought to stabilise a reverse turn in the protein. b. The positions of waters 200/4-8 predicted by GRID2 using Prediction Method 2 are shown in red and are all within 0.7Å of the observed positions. See Table 4.4.

Table 4.5 The predictions of the nitrate ions in crystalline human lysozyme by program GRID.

Nitrogen of nitrate ion	$\langle u^2 \rangle^{\frac{1}{2}}$ of the nitrogen atom in Å	δd in Å	E_{pred} in kcal/mol	δE in kcal/mol	H-bond partners at observed position	H-bond partners at predicted position
207/9	1.39	0.11	-18.16	1.09	202/1 204N/3 205N/2 206N/8 OG Ser 24N ND Asn 66	202/1 204N/3 205N/2 206N/8 OG Ser 24N ND Asn 66
218/6	1.50	0.91	-16.24	7.40	206/9 213/9 214/7 NH1 Arg 110	206/9 213/9 214/7 209B/7 N Cys 6 N Glu 7
221/8	1.60	0.59	-11.26	3.79	210/1 209/2 210/5 210/0 N Asn 60	210/1 209/2 210/5 210/0 N Asn 60

of the hydrogen-bond at the probe in GRID2 which prevented the prediction of hydrogen-bonds which were poorly oriented at the probe. For instance, at water 200/8, O Leu 84 and water 200/6 subtend an angle of 37°. GRID1 predicted hydrogen-bonds to be made by 200/8 to both O Leu 84 and 200/6 and this is clearly unfavourable. GRID2 was able to take account of the angle subtended at 200/8 by these atoms and so it predicted a hydrogen-bond to be made only to 200/6. This resulted in an improved prediction for water 200/8 with the magnitude of δd being reduced by 0.13Å.

4.5.7 The predictions of the nitrate ions

By running program GRID with a nitrate probe, the three nitrate ions were predicted with $\delta d < 1$Å and within $\langle u^2 \rangle^{\frac{1}{2}}$ of the observed position of the nitrogen atom of the ion (see Table 4.5). δd was greatest for the nitrate ion centred on 218/6 because program GRID predicted the movement of this nitrate ion in order to satisfy its hydrogen-bonding potential. The satisfactory prediction of the nitrate ions suggested that the parameters derived to describe them were not inappropriate.

4.5.8 The potential application of the GRID method to the interpretation of electron density in x-ray maps

Program GRID may be of use in indicating the assignment of electron density in x-ray maps and it is shown here that it may be of particular value in the interpretation of the less well defined regions of such maps.

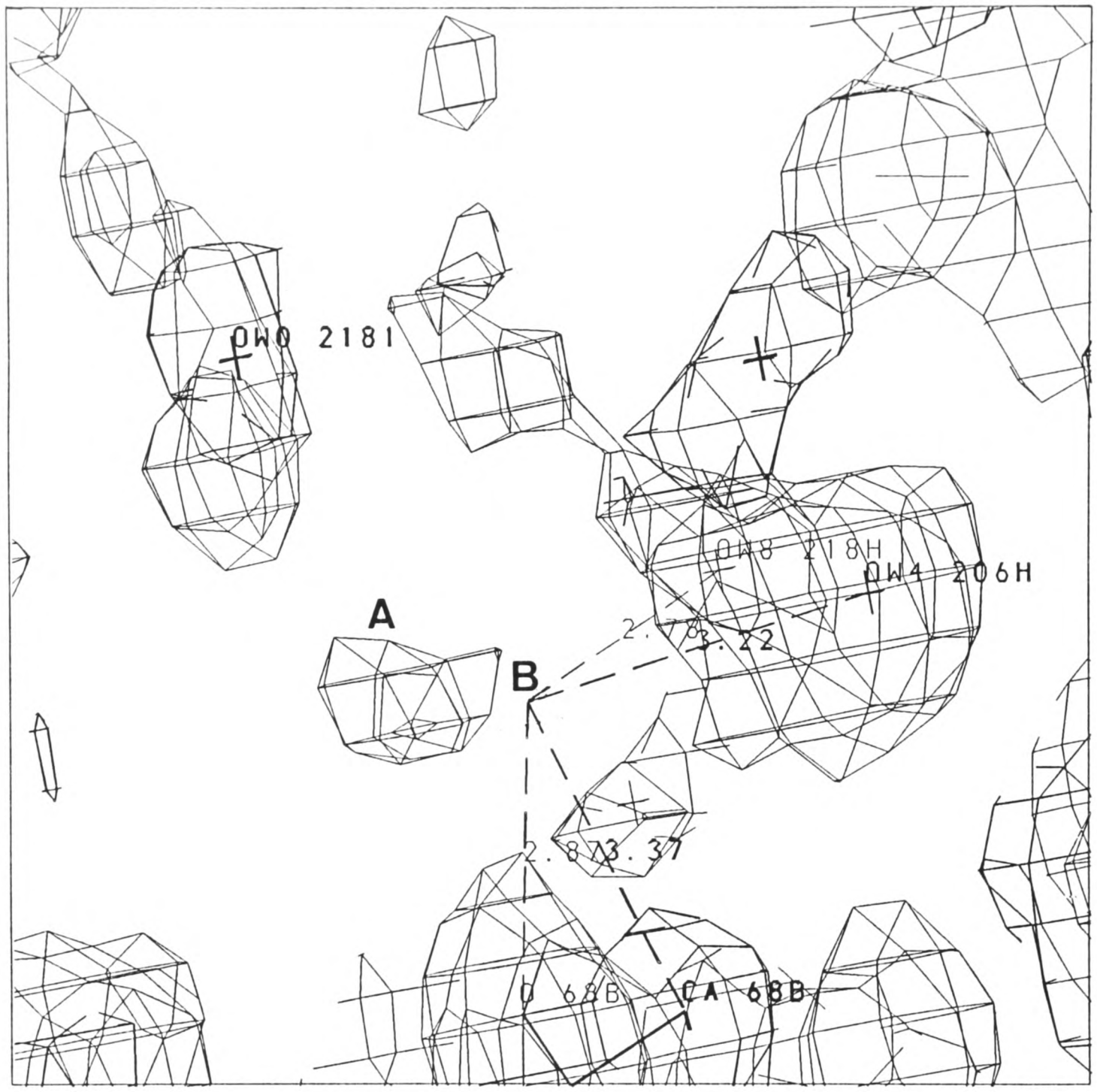


Figure 4.9 Part of the electron density map of human lysozyme with contours at 0.3e$Å^{-3}$. When program GRID was run for a water probe in this region with all observed waters included in the target, an energy minimum was identified at the site labelled B just adjacent to a region of unassigned density labelled A suggesting that this density could be occupied by a water molecule. At the energy minimum at B, a water could form hydrogen-bonds to O 68B, 218H/8 and 206H/4. See text for details.

Examination of the region of the electron density map shown in Figure 4.9 shows an unassigned volume of density in the solvent labelled A. When program GRID was run with a water probe over this region, an energy minimum was predicted at the site labelled B where a water could make hydrogen-bonds to O 68 B, 211/7H and 215/5H. This site is just adjacent to the volume of unassigned density A suggesting that this density may be occupied by a water molecule.

Water 218/0 (see Figure 4.9) could not be predicted using Prediction Method 2 and had the largest value of δd (=5.46Å) of all the waters predicted and a value of E_{obs}=-1.85 kcal/mol. However, if a water molecule occupied the volume of unassigned density A, it could form a hydrogen-bond to 218/0 and this might then enable program GRID to locate a water binding site near the observed position of water 218/0.

17 waters had values of E_{obs}>0 because of their close proximity to another atom less than 2.4Å away which was either part of the protein or another solvent molecule. Further investigations of some of these waters were carried out using program GRID.

4.5.8.1 Observed water position very close to the protein

One such water is 207/4 which is near to the mobile Arg 14 side-chain as shown in Figure 4.10. It is much better defined, having a B-value of about 48Å^2, than NH2 Arg 14 which has a large B-value of 95Å^2. NH2 Arg 14 is only 2.39Å away from 207/4 and was therefore calculated by program GRID to interact repulsively with 207/4. Program GRID predicted an energy minimum for this water at the position marked

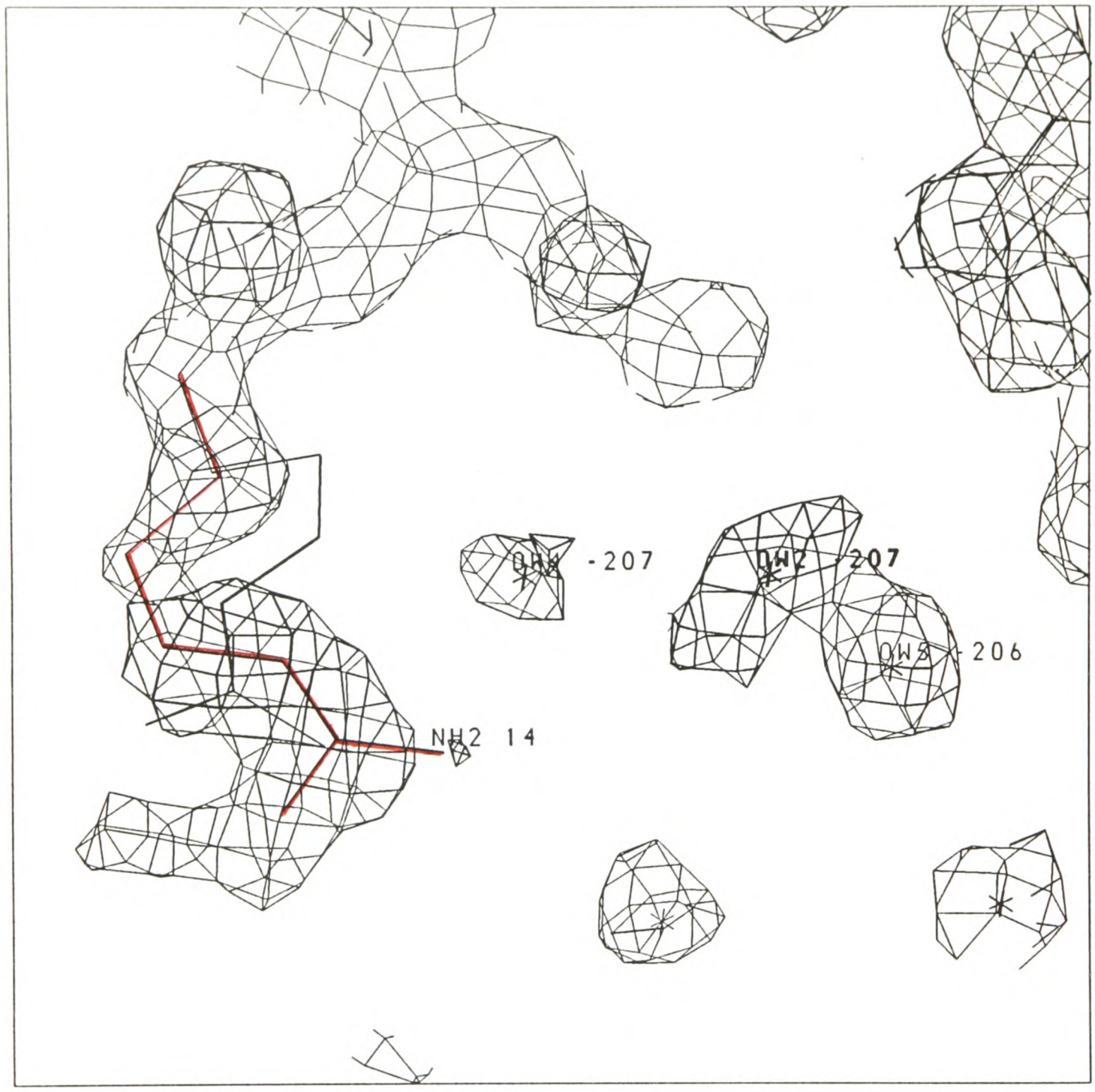


Figure 4.10 Part of the electron density map of human lysozyme with contours at 0.5eÅ^{-3} showing water 207/4 in a better defined position than NH2 Arg 14. Two possible conformations of Arg 14 are shown. The red one was used for the first GRID prediction and appears to fit the electron density better than the other conformation shown in black. However, water 207/4 interacts unfavourably with HL when the Arg 14 side-chain is in the conformation shown in red as it is only 2.39Å from NH2 Arg 14, and it makes much better contacts with Arg 14 when it is in the other conformation. The triangular pointer shows the predicted water position at δd=0.46Å calculated with the conformation of HL shown in red. (A pair of closely spaced waters, 207/2 and 206/5 at a separation of 2.34Å are also shown.)

by the triangular pointer at δd=0.46Å further away from NH2 Arg 14 than the experimentally observed position.

The GRID predictions were made for one conformation of the HL molecule. In fact, the protein is mobile and alternative conformations have been found experimentally for some side-chains including Arg 14 [227] (see Figure 4.10). In the alternative conformation, NH2 Arg 14 is further from 207/4 and so program GRID predicted favourable binding of water 207/4 to the protein in this conformation with E_{obs}=-11.2 kcal/mol. Therefore, the GRID calculations suggest that if the Arg 14 side-chain swings between the two identified conformations, water 207/4 may move complementarily.

4.5.8.2 Observed water position very close to another water

Two examples where waters were very closely positioned were examined.

4.5.8.2.1 Waters 217/1 and 217/9

Waters 217/1 and 217/9 are at positions only 2.14Å apart (see Figure 4.11) and repulsive values of E_{obs} of 15.8 and 18.3 kcal/mol respectively were calculated for them by program GRID. They lie at the interface between two HL molecules and are mobile with B-values of 71.09Å^2 and 73.47Å^2 respectively.

Using program GRID with Prediction Method 2, the nearest energy minimum to 217/1 was found at the position marked OW 2W and the nearest energy minimum to 217/9 was found at the position marked OW 1W. If both observed waters were removed from the target and a

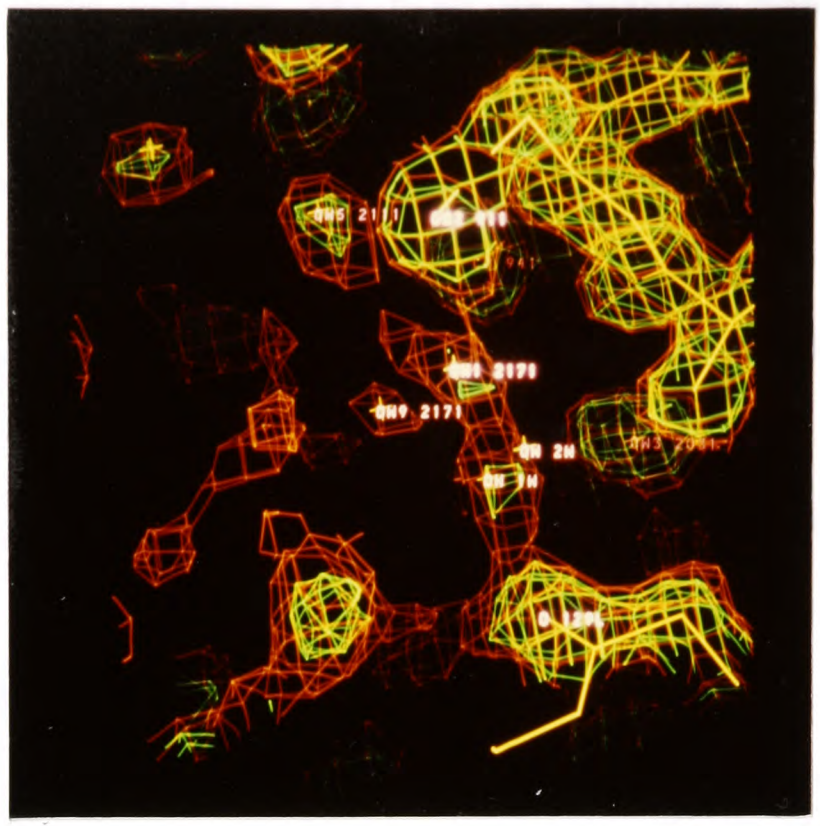


Figure 4.11 The electron density map of human lysozyme with contours at 0.3eÅ^{-3} (red) and 0.5e^{-3} (green) showing two waters, 217/1 and 217/9, 2.14Å apart at the interface between two HL molecules and their predicted positions at OW 2W and OW 1W respectively. When both observed waters were removed from the target, program GRID predicted a water at site OW 1W. See text.

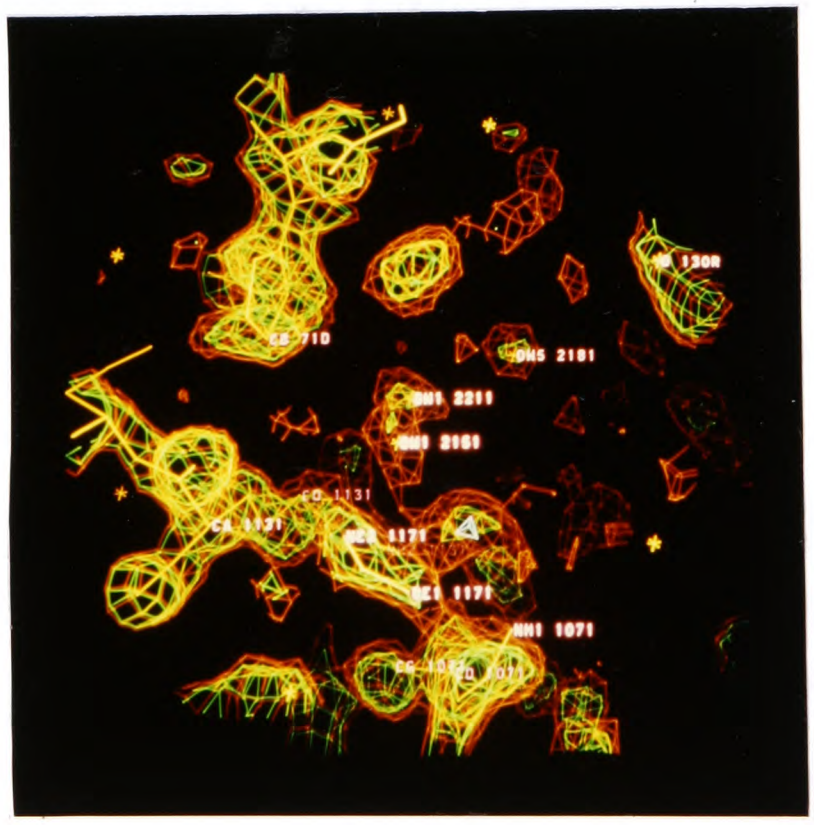


Figure 4.12 The electron density map of human lysozyme contoured at 0.3eÅ^{-3} (red) and 0.5eÅ^{-3} (green). Two waters 215/1 and 221/1 1.93Å apart are shown at the interface between three symmetry related molecules. 215/1 was predicted by program GRID to lie nearer the protein at δd=1.8Å at the neck in the red contours shown in the map at a region of comparable electron density to that at the assigned water position. The white pointer is in a region of unassigned electron density near Asn 117 and Arg 107 within which a water binding site was predicted by program GRID. See text.

prediction made with program GRID, an energy minimum for a water was found at site OW 1W where it could hydrogen-bond to O 129L and 202/3. Site OW 1W is in a region of considerably more electron density than 217/1 and 217/9 and thus it is reasonable to suggest that it may be occupied by a water molecule. Such a water could stabilize the position of water 217/1 by making a favourable hydrogen-bond to it. It would be only 2.2Å from 217/9 but as the latter water is in a region of much weaker electron density, it may be possible to adjust its position in order to accommodate a water site at OW 1W.

4.5.8.2.2 <u>Waters 215/1 and 221/1</u>

Waters 215/1 and 221/1 are at the interface between three symmetry related HL molecules (see Figure 4.12). They are very mobile with temperature factors of 67.65Å^2 and 73.09Å^2 respectively. 215/1 and 221/1 are positioned 1.93Å apart and this is shown to be unrealistic by program GRID which calculated very repulsive values of E_{obs} (≈80 kcal/mol) for these two waters.

215/1 was predicted by program GRID to lie nearer to the protein and further away from 221/1 at δd=1.8Å at the neck in the red contours seen in Figure 4.12. The predicted site was at a similar electron density to the observed site suggesting that a water placed at the predicted site could simultaneously satisfy the observed electron density and make energetically favourable interactions with the surrounding atoms.

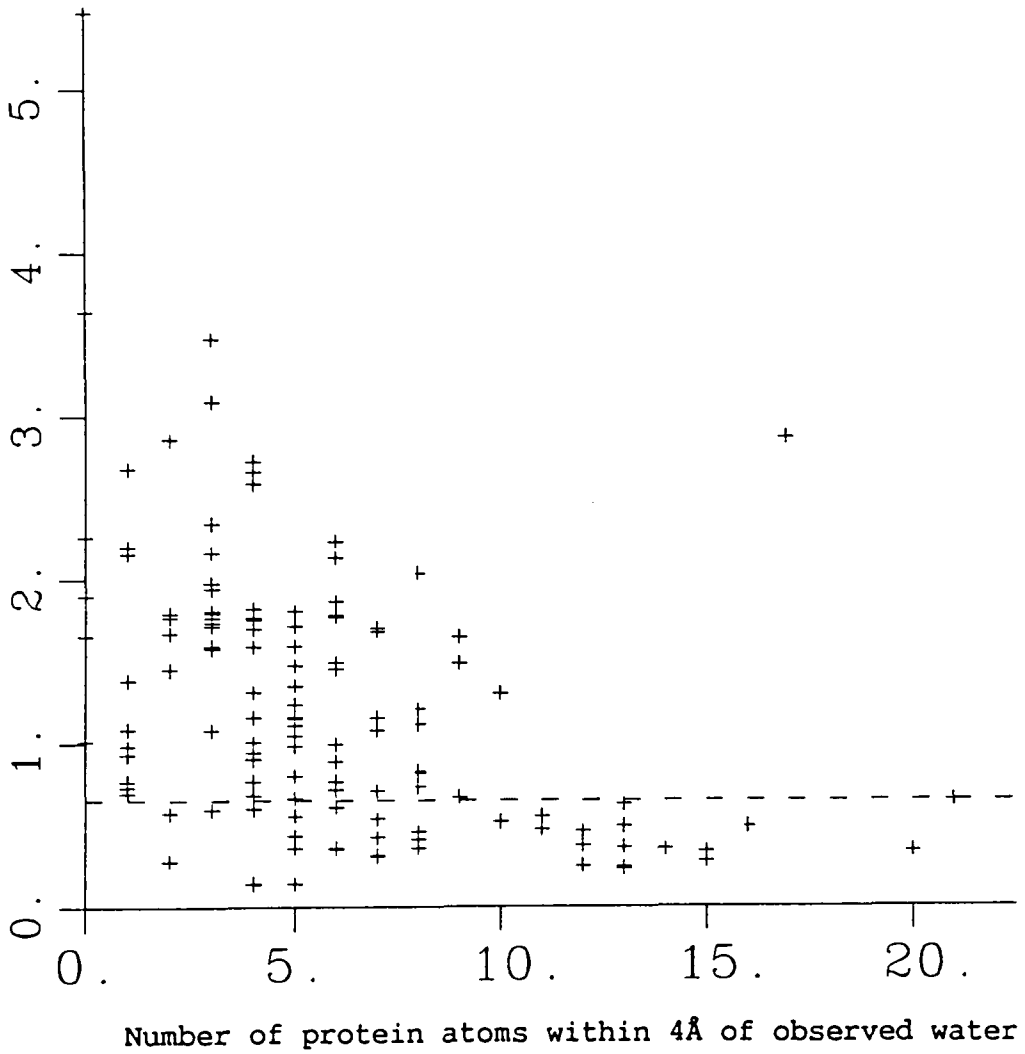
Nearby, there is a region of considerable unassigned electron density indicated by the white pointer (see Figure 4.12). This might be expected to contain a water molecule which could interact with Arg 107 and Asn 117. When program GRID was run over this

region with a water probe, an energy minimum was predicted within the red electron density contours at a position where hydrogen-bonds could be made to 215/1, 221/1 and NH1 107 suggesting that this site could be occupied by a water. However, Arg 107 is very mobile [227] and this may affect the possibility of assigning a water in this region. Two possible conformations have been found for Arg 107, one as shown here and one with NH1 pointing away from these waters and the identified region of unassigned density.

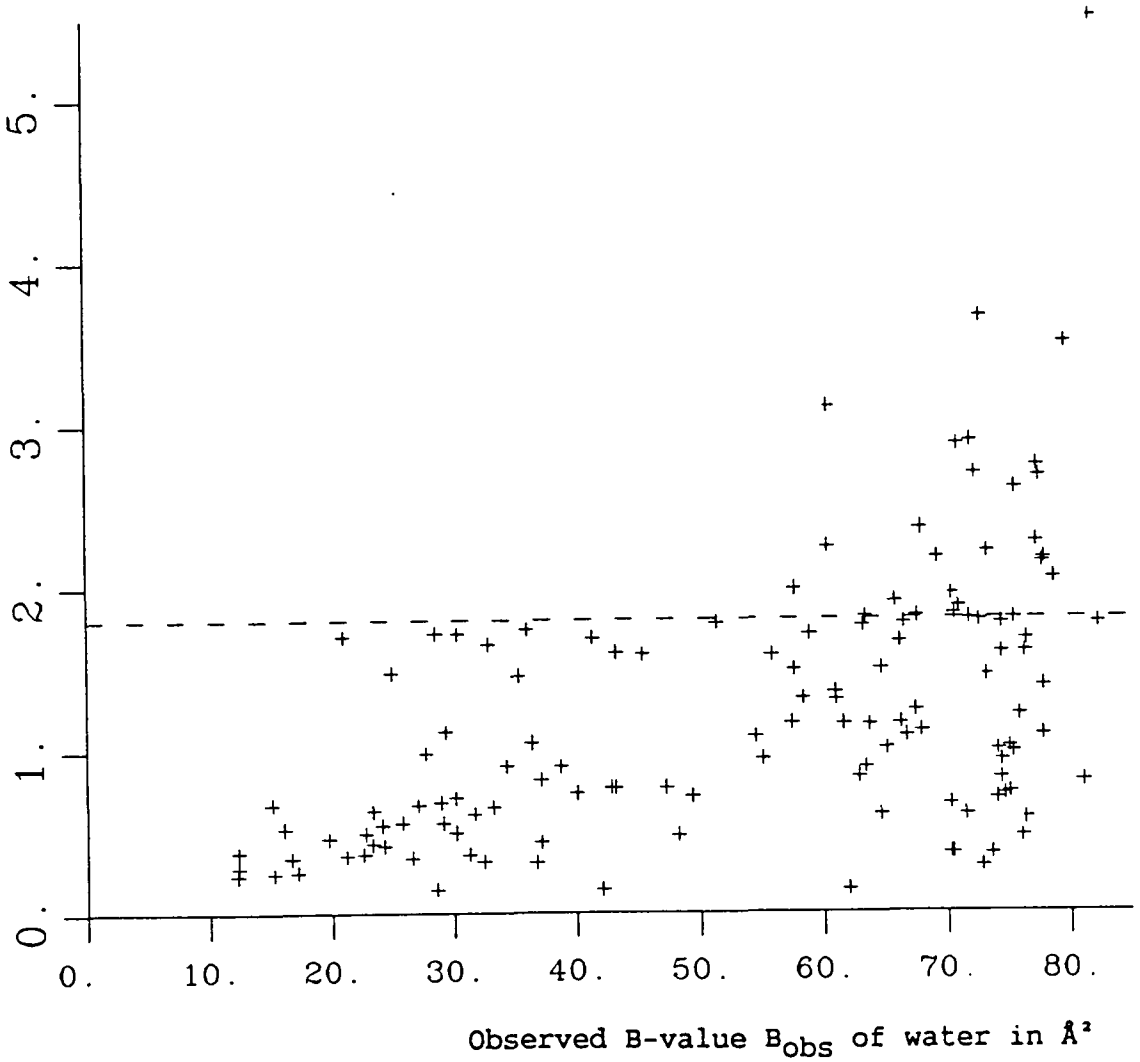
These examples illustrate some of the ways in which program GRID may assist in the assignment of the positions of solvent molecules during crystallographic refinement. Program GRID may be of value in locating both well-ordered waters, such as the internal waters, and waters in more mobile regions of the crystal such as in the solvent channel between two protein molecules.

4.5.9 The estimation of the accuracy of the predictions made with GRID2

It is important to have a measure by which to estimate the accuracy of predictions made for ligands which are not experimentally observed. Therefore, using the predictions of the water structure of HL, a search was made for a suitable benchmark for assessing the accuracy of predictions made by program GRID.



a.



b.

Figure 4.13 a. A plot of distance δd against the number of protein neighbours within 4Å of each experimentally observed water site in human lysozyme. The horizontal dashed line is at δd=0.65Å and it can be seen that all but one of the waters with more than ten protein neighbours within 4Å were predicted with δd<0.65Å. b. A plot of distance δd against temperature factor B_{obs} for each water molecule. The horizontal dashed line is at δd=1.8Å and it can be seen that all waters with B_{obs}<56Å² were predicted with δd<1.8Å.

4.5.9.1 Interaction with the protein

Waters making many contacts with the protein were well predicted. All but one of the ordered waters with more than ten protein neighbours closer than 4Å were predicted within 0.65Å of their observed positions (see Figure 4.13.a). Internal waters, waters bound in clefts and crevices, and conserved waters found in equivalent sites in homologous protein structures were all well predicted by program GRID. Therefore, site conservation and the strength of binding may act as indicators of the reliability of a prediction.

4.5.9.2 Mobility

Using mobility as a criterion, program GRID was able to predict all ordered waters in HL with $B_{obs}<56Å^2$ ($\langle u^2\rangle^{\frac{1}{2}}<1.46Å$) to within 1.8Å of their experimentally observed positions (see Figure 4.13.b). More mobile waters had a smaller probability of being well predicted. The waters with high values of δd were often near very mobile protein side chains or else positioned far out into the solvent channels. In these regions, inaccuracies in the experimental assignment of waters may be comparable to inaccuracies in the GRID predictions.

A similar classification of water molecules may be applicable to other macromolecules.

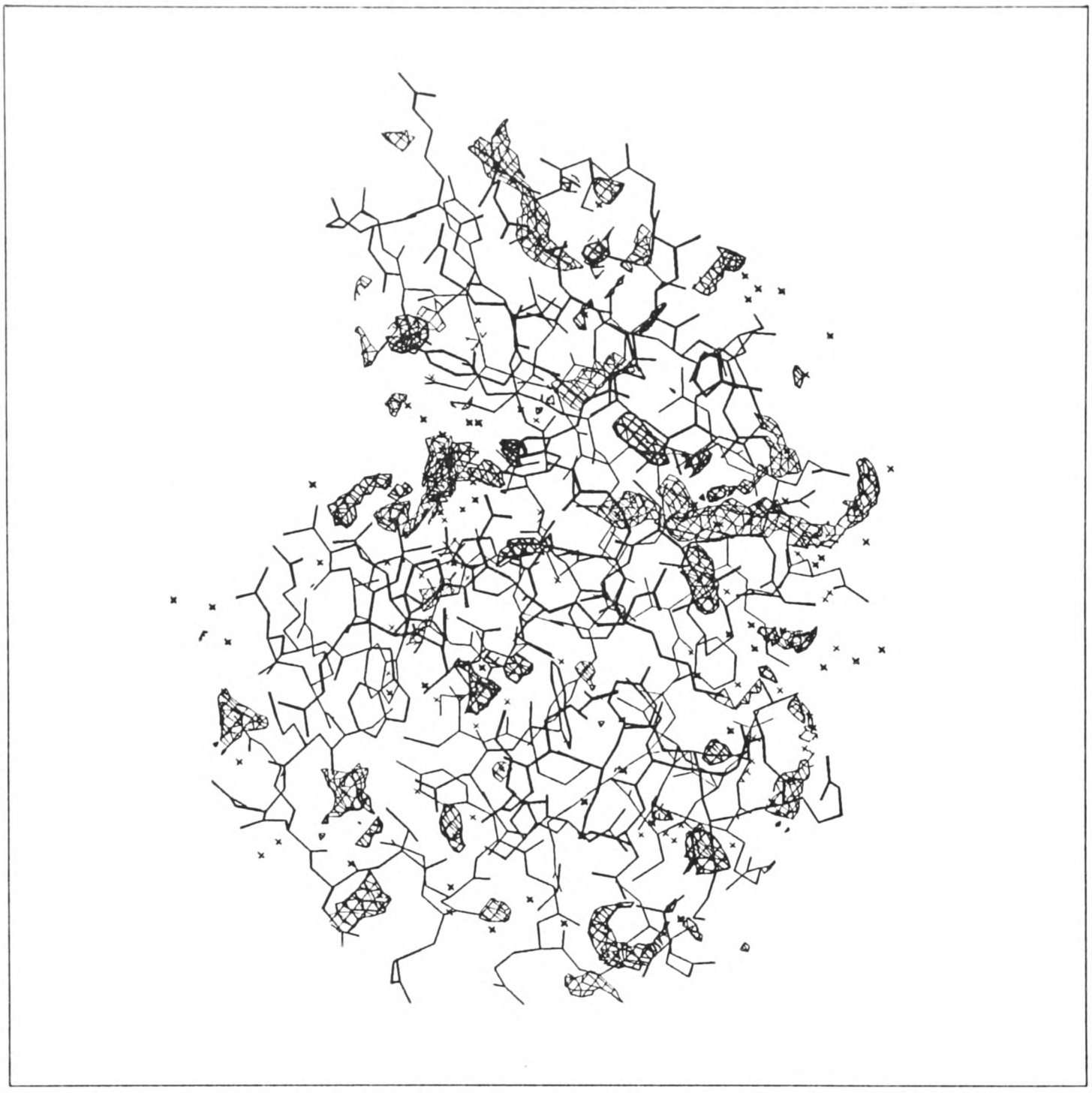


Figure 4.14 Human lysozyme with energy contours at -6 kcal/mol for the interaction with a water probe. 32 of the observed waters were found within these contours. The largest contoured region is at the location of the internal water molecules.

4.5.10 The prediction of water by GRID2 using Prediction Method 1

For proteins of pharmacological interest, experimental water coordinates may not always be available and so it is desirable that program GRID should be able to predict the important water binding sites in such proteins. In order to investigate the ability of program GRID to do this, GRID2 was run for a water probe over the whole isolated HL molecule. The array used for the calculations is shown in Figure 2.1.a. The energy contours obtained at -6 kcal/mol are shown in Figure 4.14. The contoured regions show where water was predicted to bind strongly to the protein and 32 of the observed water sites were within these contoured regions.

4.5.10.1 The prediction of the internal waters by GRID2 using Prediction Method 1

The largest contoured region in the energy map shown in Figure 4.14 is in the internal cavity of HL which contains the chain of water molecules 200/4-8 described in Section 4.5.6.1. Program GRID was used to sequentially position these waters in the cavity without utilising knowledge of their experimental coordinates. This was performed as follows:

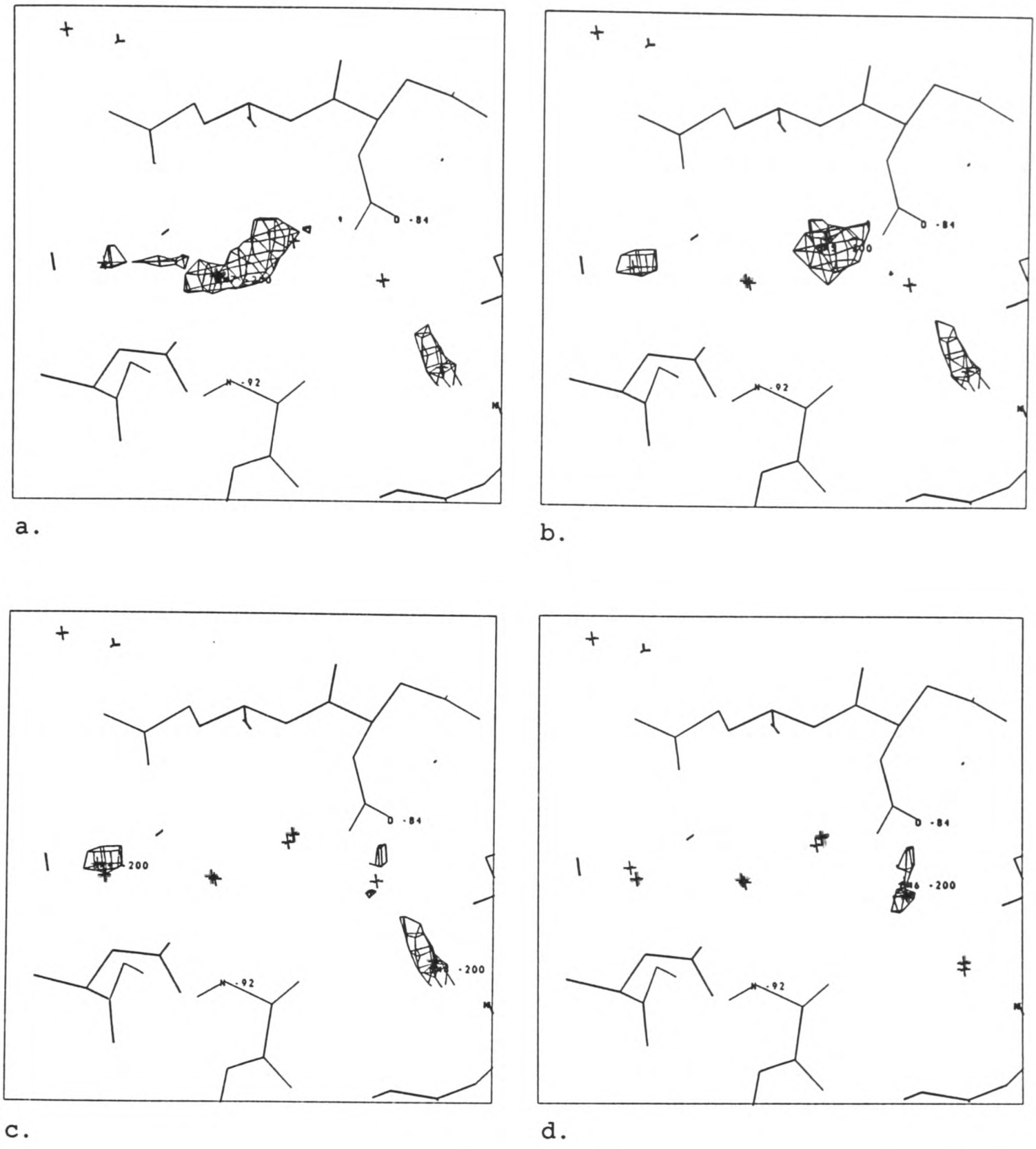
1. Program GRID was run with a water probe and a target consisting of one HL molecule only over the whole internal cavity region. The shape and location of the array used for these calculations is shown in Figure 2.1.b. The energy contours obtained are shown in Figure 4.15.a at a contour level of -9 kcal/mol. An

Table 4.6 The prediction of the internal waters in human lysozyme by GRID2 using Prediction Method 1.

Predicted water	Number of predicted waters included in the target molecule	Closest observed water molecule	δd in Å[a]	E_{pred} in kcal/mol	Hydrogen-bond partners at predicted position[b]
300/1	0	200/7	0.12 (0.27)	-13.80	N Ala 92 O Asn 88 O Ala 83
300/2	1	200/5	0.46 (0.66)	-15.17	O Gln 86 300/1
300/3	2	200/4	0.38 (0.34)	-13.24	N Asp 91 OD1 Asn 88 300/1
300/4	2	200/8	0.19 (0.36)	-11.34	N Phe 57 O Tyr 54
300/5	4	200/6	0.44 (0.51)	-12.79	O Leu 84 N Ile 56 300/2 300/4

[a] Values in parentheses were obtained using Prediction Method 2 and GRID2.

[b] All hydrogen-bond partners were also predicted by GRID2 using Prediction Method 2 except for O Asn 88 for water 300/1(200/7). See text for details.



a. b. c. d.

Figure 4.15 Energy maps for a water probe at a contour level of -9 kcal/mol showing the sequential prediction of waters 200/4-8 in human lysozyme. The observed positions of the waters are marked by black crosses and the predicted positions by red crosses. a. The target consisted of HL only. The red cross marks the position of the predicted energy minimum 0.12Å from the observed position of water 200/7. Water 300/1 was assigned to this energy minimum. b. The target consisted of HL and one predicted water (300/1). Water 300/2 was predicted at the energy minimum near 200/5. c. The target consisted of HL and two predicted waters (300/1, 300/2). Two energy minima could be identified in this map and were predicted to be occupied by water 300/3 near the observed position of 200/4 and water 300/4 near the observed position of 200/8. d. The target consisted of HL and four predicted waters (300/1-4). Water 300/5 was predicted near the observed position of 200/6. No further waters could be fitted into this cavity. See Table 4.6.

energy minimum was found, its position was refined, and a water (300/1) was assigned to it (indicated by a red cross).

2. Program GRID was then run with a target consisting of one HL molecule and water 300/1. A second energy minimum was found and water 300/2 assigned to it (Figure 4.15.b).

3. This procedure was repeated (Figure 4.15.c and d), adding each predicted water to the target, until no suitably deep energy minimum could be found. This occurred when five waters (300/1-5) had been predicted.

The properties of the predicted waters are listed in Table 4.6. The waters were all predicted at positions less than 0.5Å from observed waters and with δd less than $\langle u^2 \rangle^{\frac{1}{2}}$.

Several hydrogen-bonding networks between the waters were possible including the one predicted using Prediction Method 2. All of the hydrogen-bonds predicted except for that between 300/1(200/7) and O Asn 88 were also predicted using Prediction Method 2. The latter hydrogen-bond was not predicted using Prediction Method 2 because the waters 200/4 and 200/5 were included in the target and were calculated to make more favourable hydrogen-bonds to 200/7 than O Asn 88.

The positions of these predicted waters could be refined by repredicting each water in turn with all the other predicted waters included as part of the target until there was no significant variation in the predicted positions and energies.

This prediction of the internal waters illustrates how program GRID may be used to accurately determine well-ordered water sites in crevices of macromolecules. Such water sites are often found at positions where ligands bind to macromolecules and thus the

ability of program GRID to predict their positions is of vital importance.

4.5.11 Conclusion

The water structure of crystalline HL has been studied energetically using program GRID. The accuracy of the predictions of the water molecules has been assessed by examining their position, energy, hydrogen-bonds and mobility.

In summary, the results obtained using Prediction Method 2 are as follows:

1. 64% of the waters were predicted to within 1.5Å of their observed position. The rms distance between observed and predicted water positions was 1.28Å.
2. 90% of the experimentally observed water positions were found to be in energetically favourable regions.
3. The atoms to which hydrogen-bonds were made could be predicted correctly for the more ordered waters but there was a tendency to overestimate the occurrence of water-water hydrogen-bonds. The hydrogen-bonds were predicted better by GRID2 than GRID1.
4. The mobility of the waters could not be predicted by program GRID and this may be because the mobility of the target is not taken into account in the calculations.

Generally, more ordered waters were predicted better than less well ordered waters. Thus the ability to make a good prediction could be estimated from the B-value and the number of protein neighbours of the water molecule.

The ability of program GRID to assist in the assignment of waters during crystallographic refinement has been demonstrated.

Program GRID has also been shown to be able to accurately predict the positions of the strongly bound internal waters from a knowledge of the protein structure alone and thus, program GRID should be of value in determining the solvation of ligand binding sites.

4.6 STUDIES OF THE WATER IN CRYSTALS OF β-CYCLODEXTRIN (β-CD)

4.6.1 The purpose of the study of the water in crystalline β-CD

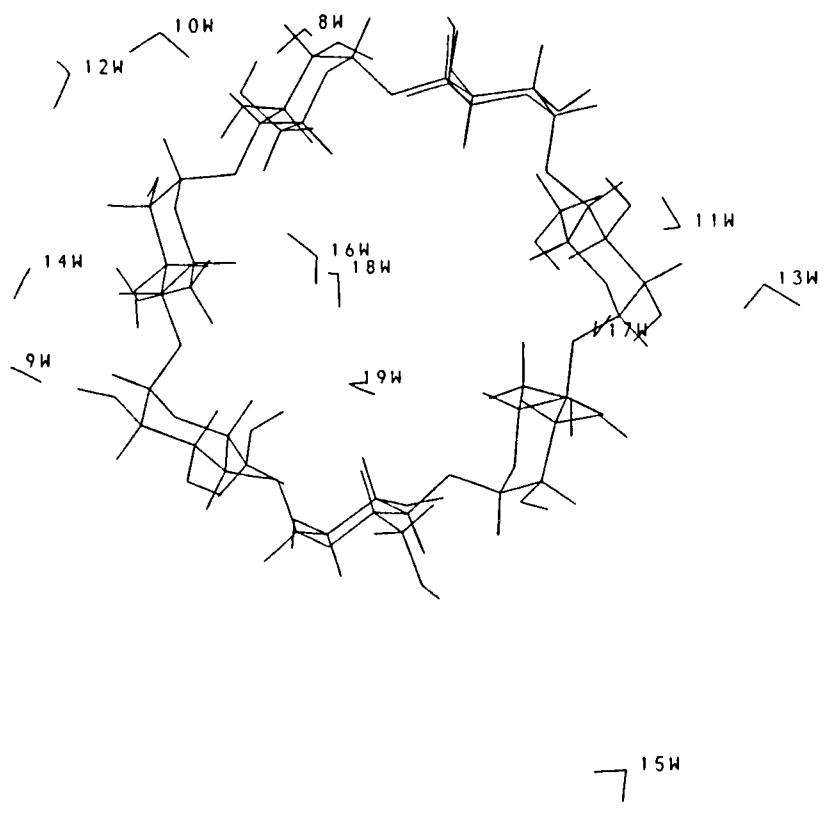
The water structure of this carbohydrate was studied for a number of reasons:

1. It provides a model system for the testing of the functions used in program GRID for determining hydrogen-bonds to target hydroxyl groups and ether oxygens. It has previously been used in order to test the GROMOS force field [228].

2. As β-CD has been studied by neutron diffraction [229,230], the coordinates of the hydrogen atoms have been determined. Predictions with program GRID can be made with and without the explicit representation of all the target hydrogen atoms. β-CD can thus act as a model system for testing the effect of using different representations of target hydrogen atoms.

3. A number of three-centred hydrogen-bonds have been observed in the crystal structure of β-CD [230]. In addition, homodromic chains of waters have been observed in which cooperative interactions are thought to occur polarizing the O-H--O hydrogen-bond and increasing its strength [181]. It is of interest to see how these factors affect

a.



b.

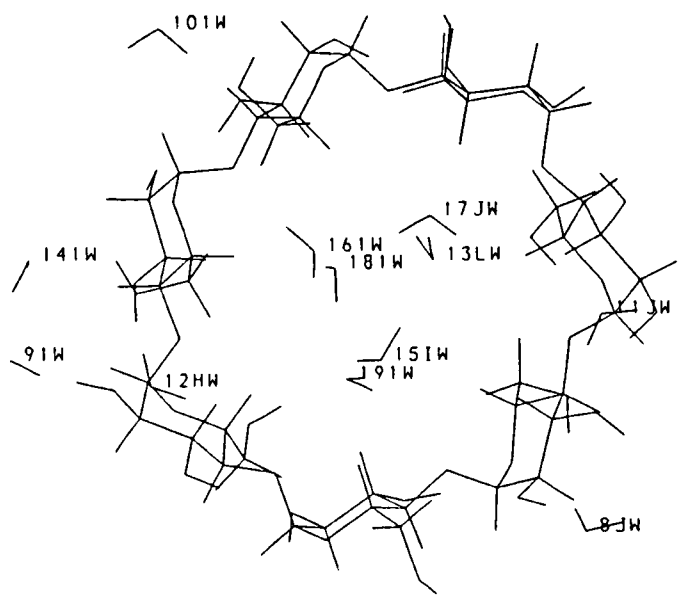


Figure 4.16 a. The structure of ß-cyclodextrin viewed from the side of the ring of larger diameter with the twelve observed water sites in the first subunit shown. Some of these waters are closer to other ß-CD molecules in the crystal and other symmetry related waters are closer to this ß-CD molecule. b. The twelve observed water sites surrounding the subunit 1 ß-CD molecule are shown.

program GRID's calculations and whether correct predictions of water sites can be made in carbohydrates as well as in proteins.

4. The interior of ß-CD is hydrophobic while the exterior is hydrophilic and this gives rise to two different types of water structure within one crystal. Thus, differences in the ability of program GRID to predict the two types of waters may be examined.

5. The structure of ß-CD has been determined at 120K [230], and at this temperature, the motion of the carbohydrate is very small. The possibility that this might enable the temperature factors of the waters to be predicted with program GRID can thus be investigated.

4.6.2 ß-CD and the water structure of its crystals

ß-CD consists of a ring of seven α(1-4) linked glucose molecules (see Figure 4.16). The ring has a smaller diameter at one end than the other thus forming a bucket shape which has a hydrophobic interior and a hydrophilic exterior. ß-CD can form inclusion complexes by accommodating guest molecules within its ring cavity which are thought to bind by non-specific hydrophobic interactions [231,232]. Thus, ß-CD is of therapeutic interest because it may be used to make hydrophobic drugs, e.g. the cardiac glycosides [233], more soluble.

The structure of ß-CD has been determined 120K at high resolution by neutron diffraction [230]. Twelve water sites (8W-19W) occupied by 11.6 waters have been observed (see Figure 4.16). One of these sites (15W) has been assigned an occupancy of 0.64. (In this thesis, the water sites are numbered 8W-19W. These sites correspond to

the experimentally observed water sites named W1-W10, W13, and W14 [230]).

The water structure has been observed to contain a four-membered flip-flop ring (9W, 14W, 024, 035) in which the directions of the O-H--O hydrogen-bonds fluctuate; a pentagonal, homodromic ring (13W, 17W, 18W, 19W, 15W) in the cavity; and several homodromic chains of waters.

4.6.3 Method of prediction of the water structure with program GRID

Predictions were made using Prediction Method 1 for one isolated ß-CD ring. The hydroxyl hydrogens were assumed to be able to rotate and an extended atom representation was used.

Two predictions were then made of the water structure of the ß-CD crystal using Prediction Method 2 with different models of the target hydrogen atoms:

1. The hydrogens were represented explicitly and fixed in their crystallographically observed positions for both ß-CD molecules and target waters.
2. The ß-CD hydrogens and target water molecules were allowed to rotate and an extended atom representation was used.

All predictions were made with version GRID2. The coordinates for ß-CD at 120K [230] were obtained from the Cambridge Structural Database [131].

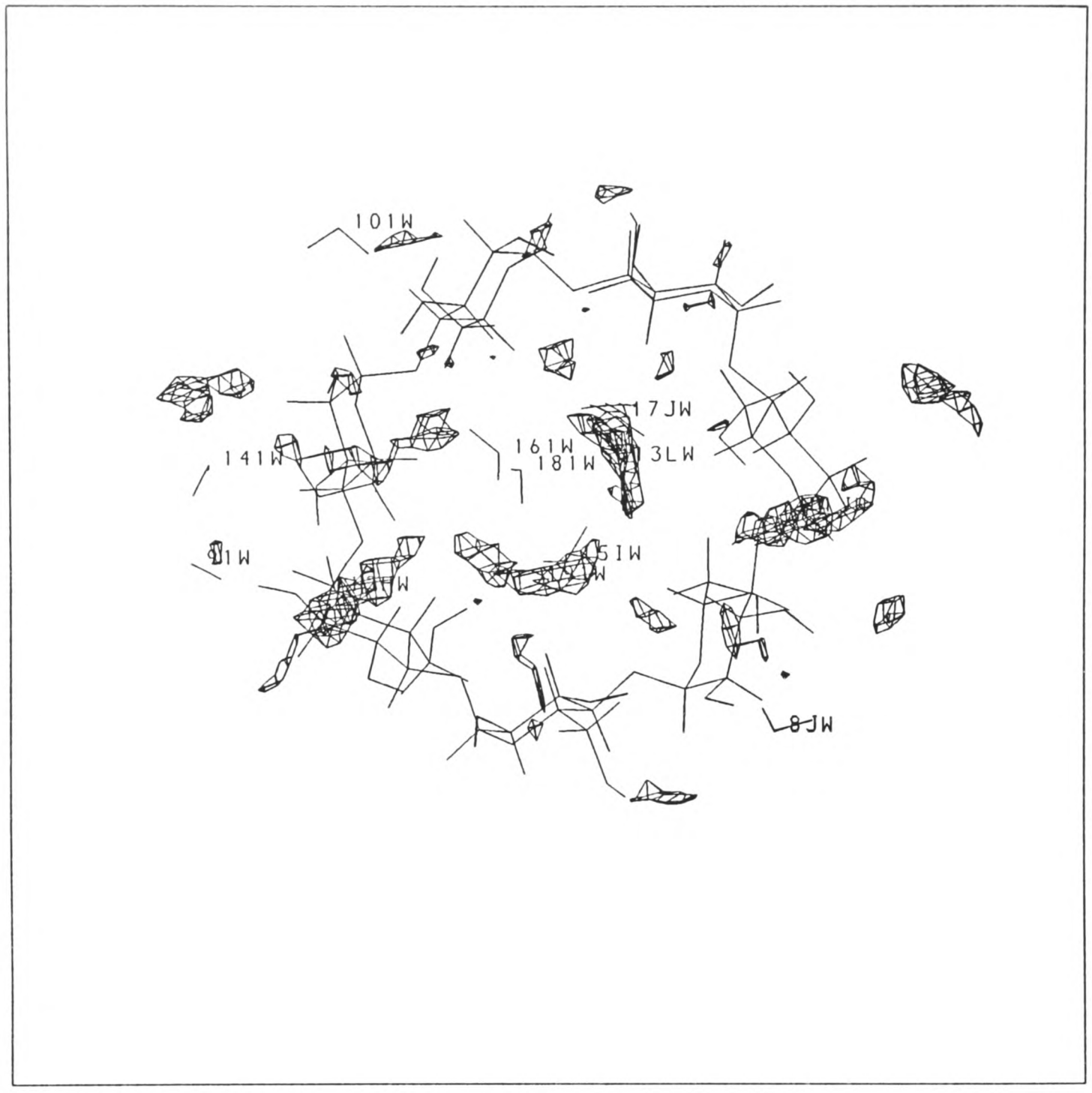


Figure 4.17 ß-cyclodextrin with GRID energy contours at -6 kcal/mol for a water probe calculated using Prediction Method 1. Five waters (15IW, 13LW, 17JW inside the ring and 12HW and 11JW outside the ring) are within these contours. See text.

4.6.4 Results of the predictions using Prediction Method 1

Only five of the observed waters, 15IW, 13LW and 17JW inside the ring and 12HW and 11JW outside, were predicted at energies of less than -6 kcal/mol as shown in Figure 4.17. The 11W site was also found [229] to be fully occupied at room temperature showing that it was a well defined site. The weak binding energy predicted at the other seven observed sites suggests that in crystalline ß-CD, as was found in HL, crystal contacts are important in determining the water structure.

4.6.5 Results of the predictions using Prediction Method 2

The results of the predictions of the twelve observed water sites are listed in Table 4.7. These were analyzed according to the four properties described in Section 4.4.

1. **Distance δd**

The distance δd was in the range 0.0 to 0.42Å and was always less than $\langle u^2 \rangle^{\frac{1}{2}}$ if hydrogens were explicitly fixed and greater only for 8W if hydrogens were not fixed. The rms value of δd was 0.20Å if hydrogens were fixed and 0.22Å if they were not. Thus, program GRID predicted the water binding sites close to their observed positions.

2. **Energy**

For all water sites, $E_{obs} < -7$ kcal/mol showing that all the

Table 4.7 The prediction of the water structure of crystalline ß-cyclodextrin by program GRID.

Water[a]	$\langle u^2 \rangle^{\frac{1}{2}}$ observed in Å	δd in Å	E_{pred} in kcal/mol	δE in kcal/mol	$\langle u^2 \rangle^{\frac{1}{2}}$ predicted in Å	Difference in Hydrogen-bond partners[b]	Hydrogen-bond partners at observed position[c]	Hydrogen-bond partners at predicted position[c]	Hydrogen-bond network[d]
8	0.23	0.04	-16.27	0.03	0.14	0	O1 K 7	O1 K 7	chain
		(0.42)	(-16.51)	(0.50)	(0.20)	(1)	O1 L 2	O1 L 2	chain
							OC2 E 7	OC2 E 7	
								(O1 E 7)	
							O1 1 6	O1 1 6	
9	0.31	0.00	-14.85	0.00	0.16	0	O1 1 4	O1 1 4	flip-flop
		(0.06)	(-15.33)	(0.16)	(0.17)	(0)	W 1 14	W 1 14	flip-flop
							O1 G 6	O1 G 6	
							O1 E 4	O1 E 4	
							(W G 12)	(W G 12)	
10	0.25	0.14	-17.42	0.79	0.10	0	O1 9 3	O1 C 3	chain
		(0.11)	(-17.49)	(0.31)	(0.04)	(0)	O1 E 7	O1 E 7	
							W 1 12	W 1 12	
							O1 1 6	O1 1 6	chain
11	0.28	0.19	-15.05	1.42	0.09	0	W J 16	W J 16	network
		(0.15)	(-15.87)	(1.36)	(0.17)	(0)	O1 L 1	O1 L 1	
							O1 L 2	O1 L 2	chain
							O1 1 1	O1 1 1	chain
12	0.24	0.13	-17.03	2.03	0.10	0	W 1 10	W 1 10	
		(0.12)	(-17.82)	(1.81)	(0.24)	(0)	O1 D 3	O1 D 3	
							O1 G 5	O1 G 5	
							O1 G 4	O1 G 4	
13	0.70	0.41	-11.96	1.22	0.17	0	W B 15	W B 15	ring
		(0.39)	(-15.27)	(4.68)	(0.12)	(1)	O1 1 1	O1 1 1	network
							W A 17	W A 17	ring
								(W J 16)	

14	0.29	0.17 (0.13)	-13.74 (-16.92)	0.71 (2.75)	0.12 (0.43)	0 (1)	O1 G 5 W E 18 W 1 9	O1 G 5 W E 18 W 1 9 (O1 1 5)	 flip-flop flip-flop
15	0.35	0.20 (0.19)	-12.99 (-15.52)	0.76 (4.18)	0.20 (0.17)	0 (1)	W M 19 W F 13 OC2 M 3	W M 19 W F 13 OC2 M 3 (W F 11)	ring ring
16	0.39	0.16 (0.35)	-12.76 (-15.30)	0.95 (3.24)	0.32 (0.06)	0 (1)	W 1 18 W L 11 OC2 1 6	W 1 18 W L 11 OC2 1 6 (W L 13)	network network
17	0.40	0.18 (0.13)	-13.50 (-13.80)	0.15 (0.05)	0.19 (0.25)	0 (0)	O1 J 5 W E 13 W L 18 O1 L 1	O1 J 5 W E 13 W L 18 O1 L 1	chain ring ring
18	0.48	0.09 (0.10)	-16.15 (-17.27)	0.22 (0.53)	0.12 (0.10)	0 (0)	W A 14 W J 17 W 1 16 W 1 19	W A 14 W J 17 W 1 16 W 1 19	 ring network ring
19	0.41	0.33 (0.08)	-13.78 (-15.12)	6.52 (5.46)	0.16 (0.12)	2 (2)	O1 1 4 W I 15 O1A[e]A 4	O1 1 4 W 1 18 W I 15 O1B[e]A 4	 ring ring

Values listed first are for predictions with fixed target hydrogen coordinates and those in parentheses are for predictions with mobile target hydrogens and with target waters represented as extended atoms. Hydrogen-bond partners are the same for the two predictions unless a different hydrogen-bond partner is shown in parentheses. If a fifth hydrogen-bond partner is listed in parentheses, it was predicted instead of the one it is listed below.

[a] 8W, 9W, 10W, 11W, 12W and 14W are in the interstitial space; 13W, 15W, 16W, 17W, 18W and 19W are in the cavity.

[b] This is the number of differences in the hydrogen-bond partners at the observed and predicted water positions.

[c] Subunits correspond to the following symmetry operations:

1=x,y,z A=x,1+y,z B=x,1+y,-1+z C=x,y,-1+z D=x,-1+y,-1+z E=x,-1+y,z F=x,-1+y,1+z G=2-x,-0.5+y,-z
H=2-x,0.5+y,-z I=1-x,0.5+y,1-z J=1-x,0.5+y,-z K=1-x,-0.5+y,-1-z L=1-x,-0.5+y,-z M=1-x,-0.5+y,1-z
O1=hydroxyl oxygen, OC2=ether oxygen, W=water.

[d] These are hydrogen-bond structures identified experimentally [230].

e O1A and O1B are different hydroxyl groups on the same glucose residue

waters were observed in favourable sites. Strong binding energies were predicted with E_{pred} in the range -12.0 to -17.5 kcal/mol.

3. Hydrogen-bonds

All waters were found by program GRID to make three or four hydrogen-bonds at the experimentally observed site and to always make four hydrogen-bonds at the predicted site if the target hydrogens were considered to be mobile. All waters were calculated to donate two hydrogen-bonds and this is in agreement with the general observation that waters tend to donate hydrogen-bonds where possible [132,171]. Two factors were found to influence the predictions of hydrogen-bonds.

a. The treatment of the target hydrogens.

These predictions show differences according to the treatment of the target hydrogens of a similar nature to those seen for L-serine monohydrate. If hydrogen coordinates were fixed, hydrogen-bonds to the same target atoms were calculated at the observed and predicted positions for all waters except 19W. If the hydrogens were not fixed, there were differences in the hydrogen-bond partners at the observed and predicted positions for waters 8W, 13W, 14W, 15W, 16W and 19W. These differences occurred principally because the neglection of the geometry of the hydrogen-bonds to the target waters led to the overestimation of the number of hydrogen-bonds to water made by the probe. At higher temperatures, there would be more flip-flop hydrogen-bonds and the fixing of hydrogen coordinates may not then be justified.

Table 4.8 A comparison of the predictions of the waters in the ring cavity and in the interstitial space of crystalline ß-cyclodextrin.

Property	Mean value for the six cavity waters	Mean value for the six interstitial waters
δd in Å[a]	0.23±0.11 (0.21±0.12)	0.11±0.07 (0.17±0.12)
E_{pred} in kcal/mol	-13.52±1.31 (-15.28±1.01)	-15.73±1.29 (-16.66±0.87)
δE_{pred} in kcal/mol	1.64±2.22 (3.02±2.05)	0.83±0.72 (1.15±0.93)
$\langle u^2 \rangle^{\frac{1}{2}}$ observed in Å	0.46±0.12	0.27±0.03
$\langle u^2 \rangle^{\frac{1}{2}}$ predicted in Å	0.19±0.06 (0.14±0.06)	0.12±0.03 (0.21±0.12)
Number of waters with different H-bond partners at observed and predicted positions	1 (4)	0 (2)
Number of H-bonds to carbohydrate at observed position	1.17±0.69 (1.17±0.69)	2.83±0.90 (2.67±0.94)
Number of H-bonds to carbohydrate at predicted position	1.17±0.69 (1.29±0.70)	2.83±0.90 (2.83±0.69)

Values listed first are for predictions with fixed target hydrogen coordinates and those in parentheses are for predictions with mobile target hydrogens and with target waters represented as extended atoms.

Error values are standard deviations.

[a] rms values of δd were 0.25(0.24)Å for cavity waters and 0.13(0.20)Å for interstitial waters.

b. The determination of the geometry of the hydrogen-bonds at the target atoms and at the probe.

These predictions show that, by determining the dependence of the strength of the hydrogen-bonds on their geometry, program GRID was able to predict hydrogen-bonds in accordance with experimental observations.

Two waters (8W and 16W) have been observed to make hydrogen-bonds to ether oxygens. These hydrogen-bonds were both predicted and so was an additional one from 15W to an ether oxygen. However, if the hydrogen atoms were allowed to rotate, the hydrogen-bond to the ether oxygen was not calculated at the predicted position of 8W which was 0.42Å from the observed water position. This was because a hydroxyl group on the same residue of a neighbouring ß-CD ring as the ether oxygen was predicted to reorient itself and make a stronger hydrogen-bond to the probe than the ether oxygen.

Three-centred hydrogen-bonds have been observed to be donated by 9W and 14W to each other and to an ether oxygen [230]. These hydrogen-bonds to the ether oxygen were not detected by program GRID. This was not unexpected because program GRID does not recognise three-centred hydrogen-bonds. The water-ether oxygen interaction is weaker than the water-water interaction and so neglection of the weaker interaction had little effect on the calculated energies and predicted positions of the two waters.

The effect of the careful determination of hydrogen-bonds by program GRID is shown by its ability to predict the interstitial waters near the hydrophilic surface of ß-CD closer to their experimentally observed positions than the cavity waters surrounded by hydrophobic carbohydrate (see Table 4.8). The interstitial waters

overall made more hydrogen-bonds than the cavity ones did and they were predicted to have a greater binding energy. They were observed to be less mobile having an rms displacement of 0.27Å compared to 0.47Å for the cavity waters.

4. **Temperature factors**

Calculated values of $\langle u^2 \rangle^{\frac{1}{2}}$ were of the correct order of magnitude but were always smaller than the observed values of $\langle u^2 \rangle^{\frac{1}{2}}$ and showed no relationship to them. This may again be due to the fact that the motion of target atoms other than hydrogen was not taken into account in the calculations.

4.6.6 Conclusion

Program GRID has reproduced the experimentally observed water structure of crystalline ß-CD at 120K although it was unable to determine its mobility.

It was able to predict the hydrogen-bonds made by the waters. The quality of the predictions suggest that the hydroxyl and ether oxygen hydrogen-bond functions used in program GRID are reasonable representations of the geometry of hydrogen-bonds to these atoms.

Program GRID was able to predict the water structure despite the fact that it does not explicitly take account of the cooperative interactions that may increase the strength of the hydrogen-bonds in chains of water and hydroxyl groups.

Better predictions for the ß-CD waters were obtained if the hydrogen coordinates were fixed. However, this would be unlikely to be

the case at higher temperatures because of the increased presence of flip-flop hydrogen-bonds. The number of cases in which one can realistically fix target hydrogen coordinates is limited and must be chosen with care.

The predictions were better for water sites in a hydrophilic environment than in a hydrophobic one and this was probably due to the greater strength and number of hydrogen-bonds to the waters in hydrophilic surroundings.

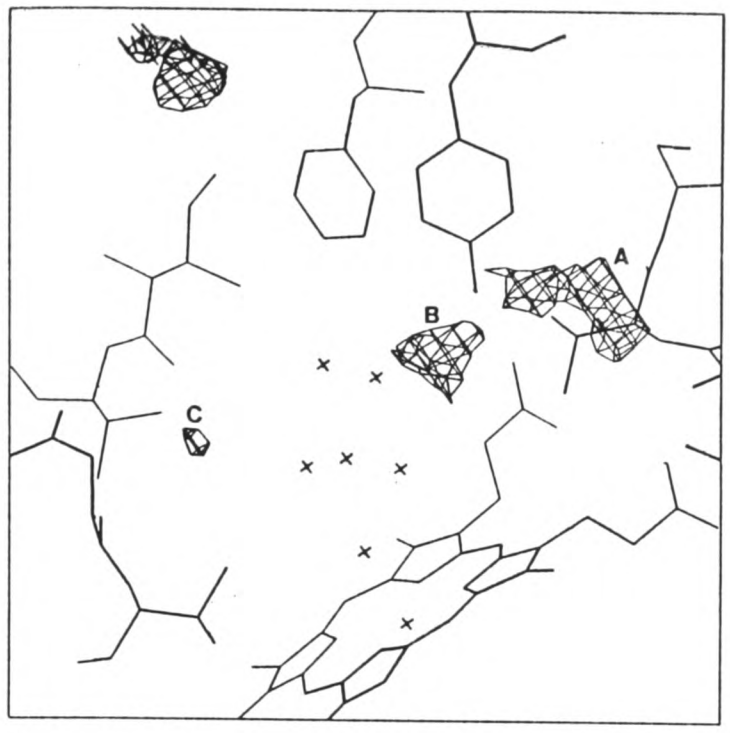
4.7 SOLVATION OF THE ACTIVE SITE OF CYTOCHROME P450-CAM

The interaction of camphor with cytochrome P450-cam was studied in Chapter 3 without considering the effect of solvent. However, the hydrophobic active site of cytochrome P450-cam contains six crystallographically observed waters [163]. Therefore, the study of cytochrome P450-cam has been extended to assess the role of water in the enzyme-substrate interaction. Program GRID has been used to predict the experimentally observed waters, to postulate additional water binding sites and to propose functions for the waters in the active site.

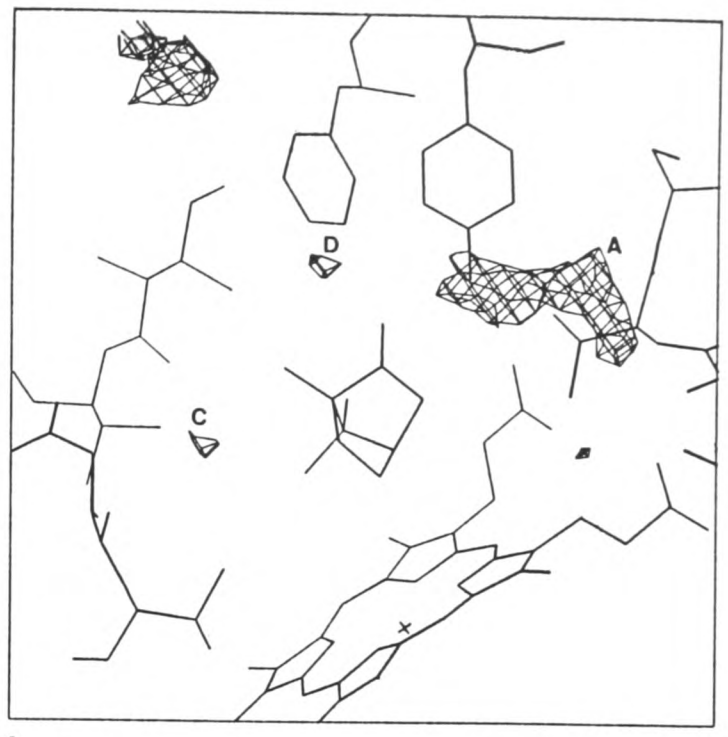
4.7.1 Method of study with program GRID

Using Prediction method 1, program GRID was run with a water probe over the active site for targets of cytochrome P450-cam with and without camphor bound. Only small conformational changes between the substrate-bound and the substrate-free enzyme structures have been observed [163] and therefore, the calculated energy maps

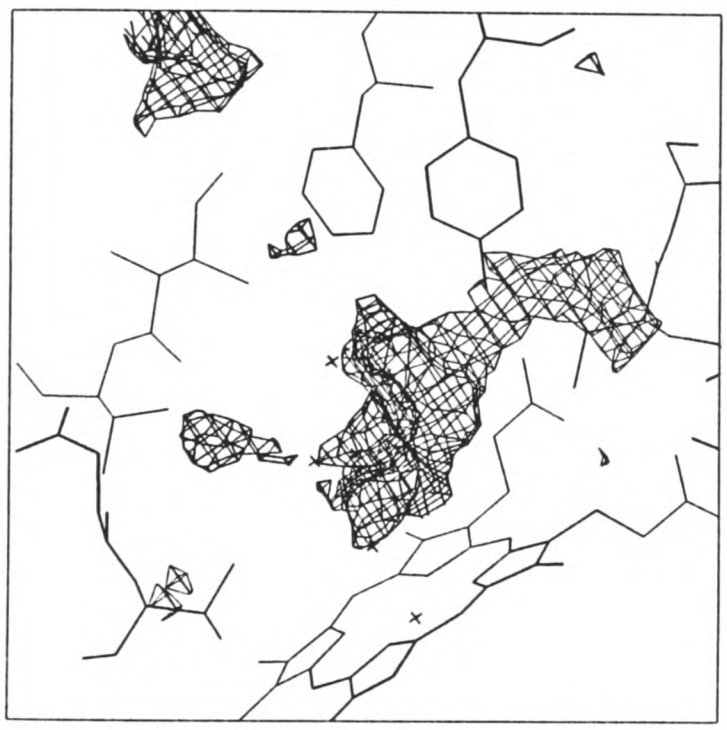
Figure 4.18 a. and b. Energy contours at -6 kcal/mol for a water probe in the active site of cytochrome P450-cam are shown a. in the substrate-free enzyme and b. in the substrate-bound enzyme. Sites A,B,C and D are distinct water binding sites predicted by program GRID. c. and d. Energy contours at -3 kcal/mol for a water probe are shown c. in the substrate-free enzyme and d. in the substrate-bound enzyme. e. Energy contours at 0.5 kcal/mol for a methyl probe in the substrate-free enzyme showing the shape of the active site. f. The six crystallographically determined water sites (W801-806) in the active site of the substrate-free enzyme and water W652, which may act to stabilize some of the active site residues, are shown. See text.


A
B
C

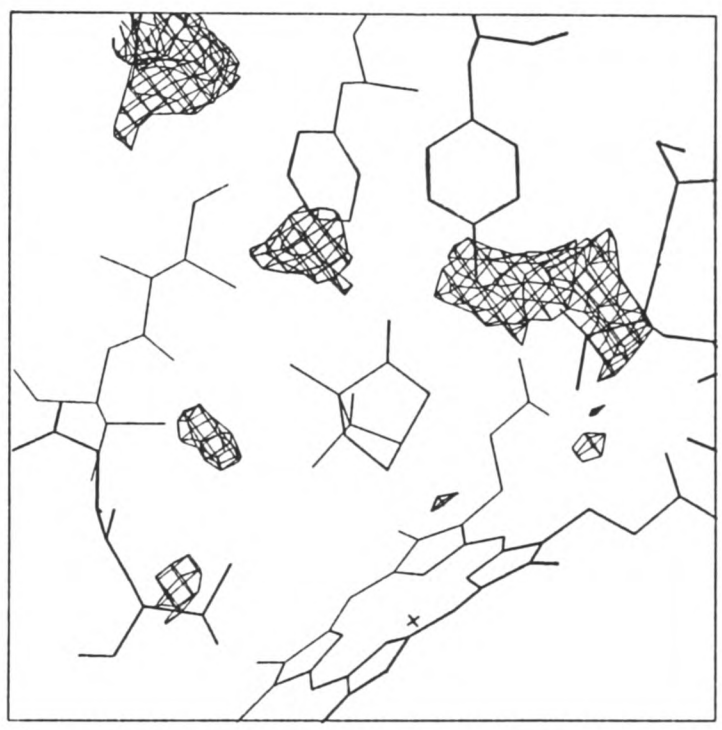
a.


D
A
C

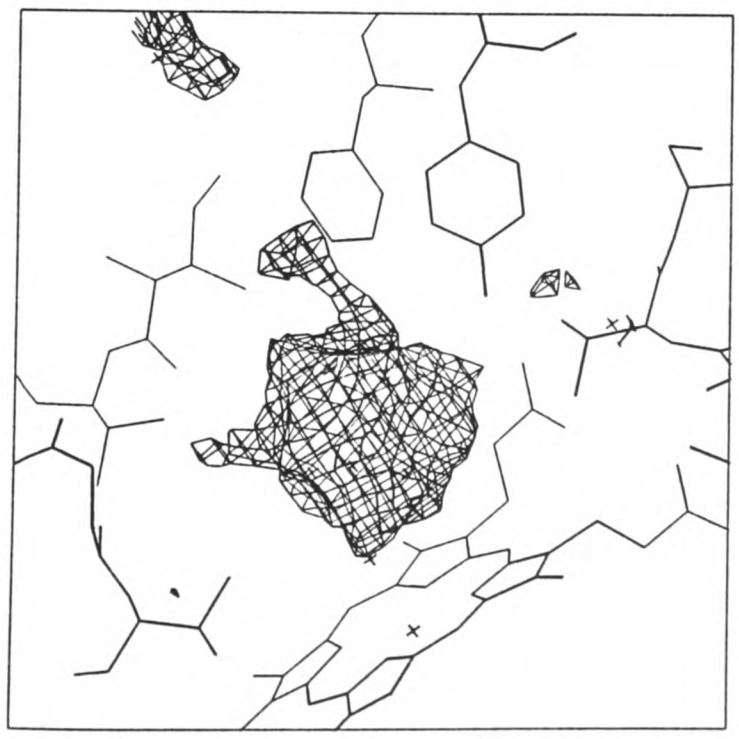
b.



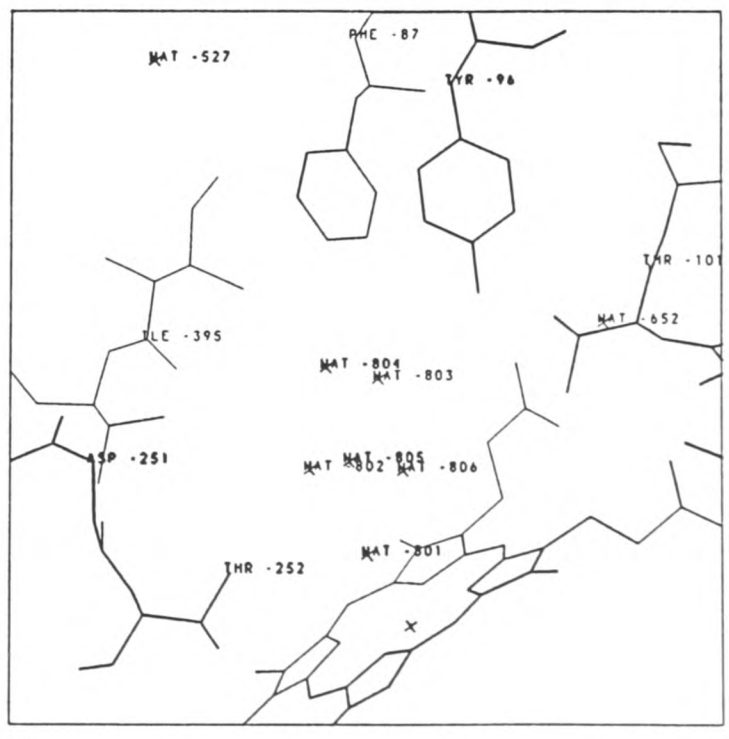
c.



d.



e.


MAT -527
PHE -87
TYR -96
THR -101
MAT -652
ILE -395
MAT -804
MAT -803
ASP -251
MAT -805
MAT -806
MAT -801
THR -252

f.

could be compared to each other and also to the energy map for a methyl probe in the camphor-free enzyme structure which delineated the shape of the active site.

Coordinates for the substrate-bound enzyme were obtained from the Brookhaven Protein Databank [52] and those for the substrate-free enzyme were provided by Dr.T.Poulos. Version GRID2 was used for all calculations.

4.7.2 Results

GRID energy contour maps for water and methyl probes in the region of the active site of cytochrome P450-cam are shown in Figure 4.18. Four distinct water binding sites, (labelled A,B,C and D), were predicted in this region of the protein. These may be occupied by water molecules which perform different functions in the enzyme and would act as follows.

1. **Water which would be firmly bound to the protein adjacent to the active site at Site A.**

A large binding region for water (labelled A in Figure 4.18.a and b) may be seen at -6 kcal/mol near Tyr 96 and Thr 101 in both structures. The experimentally observed water W652 in the substrate-free structure, (see Figure 4.18.f), occupies Site A and program GRID predicts it to be strongly bound with E_{obs}=-9 kcal/mol and able to make hydrogen-bonds to O Thr 101 and OG Ser 83. This particular water may play a structural role in the enzyme.

2. **Water at Site B which would be displaced from the active site on substrate binding.**

The second large contour (labelled B in Figure 4.18.a) seen at -6 kcal/mol near Tyr 96 in the substrate-free structure extends towards the region where camphor binds. If a water occupied this region, it would be displaced by the bound substrate. None of the observed waters are seen within this contoured region.

However, at -3 kcal/mol, there is a large favourable water binding region in the substrate-free active site (see Figure 4.18.c). This contains half of the observed water sites, (at all the observed water sites E_{obs}<-2 kcal/mol).

An energy minimum was predicted at -5.36 kcal/mol 0.23Å from the experimentally determined iron-linked aqua-ligand W801 on the opposite side of it from the haem ring. The experimentally observed aqua-ligand was found 2.28Å from the Fe ion with a low temperature factor (16.5Å^2) and full occupancy. It has been suggested [163] that a hydroxide ion may occupy this observed site instead of a neutral water molecule and would be expected to approach closer to the Fe ion because it could make a favourable electrostatic interaction.

The remaining experimentally determined waters (W802-6) were not predicted in defined energy minima but were found in a large region of weakly attractive binding energy. This is consistent with crystallographic observations [163] in which these five water sites were located within one large region of electron density and could not be individually resolved. On average, they were observed to have higher temperature factors and lower occupancies than the ordered surface waters suggesting that these water sites were not well defined and that the cavity contained a mobile water structure [163].

Displacement of these mobile waters on the binding of camphor would be favourable as it would result in little loss of water-binding energy and a favourable entropic gain. In fact, the binding of camphor is thought to be largely entropically driven due to the desolvation of the camphor substrate and the release of the active site waters allowing camphor to bind in a region of low dielectric constant [163]. However, the GRID energy maps suggest that water may play a more complex role in the binding of camphor as shown by the next two binding sites discussed.

3. **Water which would not be displaced on substrate binding and which would occupy a vacant crevice in the active site at Site C.**

At -6 kcal/mol, there is a distinct water binding site (labelled C in Figure 4.1 8.a and b) near Val 247 and Asp 251 present in both substrate-free and substrate-bound structures. From the methyl contours in Figure 4.18.e, it can be seen that Site C lies in a small crevice leading from the main volume of the active site, so the position of Site C is well defined by van der Waals' contacts. In addition, a water molecule in Site C is predicted to form hydrogen-bonds to O Val 247 and OD1 Asp 251. This suggests that a water could bind strongly in Site C and would not be displaced by the camphor substrate.

On the binding of camphor, the mobility of Asp 251 has been observed to be reduced despite the fact that it does not contact the substrate [163]. This could be explained by the presence of a water at Site C. The motion of this water would be restricted by the binding of

camphor and because this water could donate a hydrogen-bond to OD1 Asp 251, it would constrain Asp 251 and reduce its mobility.

4. **Water which would bind more tightly in the active site on substrate binding and which would serve to orient the substrate correctly and to fill a vacant crevice at Site D.**

In the substrate-bound enzyme, a third binding site (labelled D in Figure 4.18.b) is seen in the active site at -6 kcal/mol. It is only present in the camphor-free structure at a less attractive energy contour level (see Figure 4.18.c with energy contours at -3 kcal/mol). Site D is in a crevice of the active site as shown by the methyl probe contours (see Figure 4.18.e).

In the absence of camphor, Site D is very hydrophobic, containing residues Phe 87, Phe 98, Val 247, Tyr 96, and no hydrogen-bonding atoms. Thus a water may only bind weakly at this site in the substrate-free enzyme.

However, on substrate binding, a water at Site D could be stabilised by a hydrogen-bond which it could donate to the camphor oxygen at a distance of about 2.9Å away. This would be a strong hydrogen-bond because it would be optimally oriented at the camphor oxygen at the trigonal angle to the C-O bond and would lie approximately in the plane of the oxygen lone pair orbitals. Indeed, this hydrogen-bond would serve, in conjunction with that from OH Tyr 96 to the camphor oxygen, to orient the substrate correctly for the highly specific hydroxylation reaction to occur. A water at Site D would thus play an important role in stabilizing the bound substrate.

On the other hand, the presence of a water at Site D when the substrate has bound would be entropically unfavourable. This

suggests that modification of the substrate by the replacement of the carbonyl oxygen moiety by an alkoxy group which could fill Site D and displace water from it may result in a ligand with a greater binding affinity than camphor. Binding of such a ligand would permit the entropically favoured release of water from Site D and the formation of enthalpically favourable van der Waals' contacts with the hydrophobic protein residues surrounding Site D. Indeed such a ligand may bind sufficiently strongly to act as an inhibitor of cytochrome P450-cam by reducing the rate at which product is released.

4.7.3 Conclusion

Several possible roles of water in ligand binding have been identified. The GRID energy maps suggest that when camphor binds cytochrome P450-cam, two waters in the active site are undisplaced and one of these assists in orientating the substrate correctly. In addition, a number of less strongly bound waters are displaced. Consequently, water makes both favourable entropic and favourable enthalpic contributions to the binding energy of the substrate.

4.8 CONCLUSION

The water structure in four different biological systems has been studied with program GRID with the aim of testing the accuracy of the GRID method and of using it to investigate the properties of water. These studies have shown that it can successfully locate well-ordered water molecules but that it cannot reliably predict the positions of all weakly bound waters.

Version GRID2 has been shown to make more accurate predictions than versions GRID0 and GRID1. In particular, the predictions of hydrogen-bonds made by waters are improved in GRID2. Further improvements in the prediction of water by program GRID might be obtained by taking into account the directionality of hydrogen-bonds made by target waters, intratarget hydrogen-bonds and the mobility of the target.

Possible applications of program GRID to the study of water have been demonstrated. Its potential employment in the assignment of water sites during crystallographic refinement has been shown and its ability to predict the positions of well-ordered waters in ligand-binding sites has been established. Program GRID has also been used to investigate the function of water in ligand binding and to propose multiple roles for water in the binding of substrate to cytochrome P450-cam. It has shown that water may be both stabilized in and expelled from the active site of the enzyme on the binding of camphor.

Thus, program GRID may be of value in the investigation of a variety of aspects of the solvent present in biological systems.

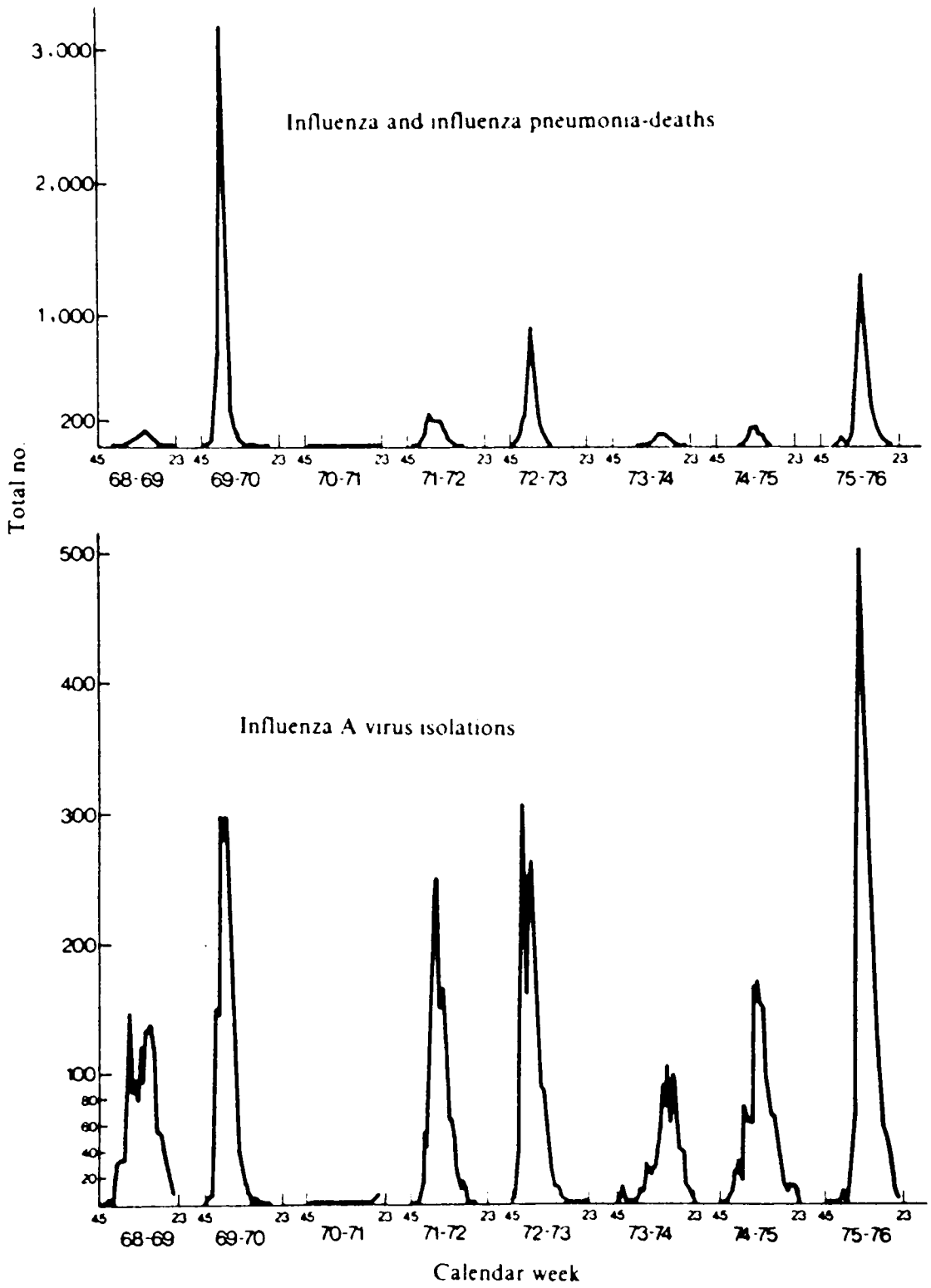


Figure 5.1 Mortality associated with influenza A in England and Wales for the years 1968-1976. The effects are shown of a pandemic in 1968 caused by the A/Hong Kong/1/68 virus and of two epidemics caused by the A/England/42/72 virus and a virus antigenically similar to the A/Victoria/3/75 virus both of which were derived from the Hong Kong virus by antigenic drift. (From reference [236]).

CHAPTER 5

THE DESIGN OF ANTI-INFLUENZA AGENTS

5.1 INTRODUCTION

In this chapter, the application of the GRID method to the design of anti-influenza agents is described. A target binding site on the influenza virus was first identified. This was mapped energetically using program GRID and ligands were designed to bind tightly to it. Some of these were then synthesised and assayed for inhibitory activity.

5.1.1. Influenza - the disease

Influenza is found worldwide in humans, other mammals, (especially horses and swine), and birds. It is a contagious, acute, respiratory disease whose familiar symptoms, of aches, pains, headaches, coughing and fever, occur within 1-4 days of infection and generally last for 2-5 days. It may be fatal if secondary complications such as pneumonia arise or if it occurs in a very young or old patient [234].

Viral infection occurs in epidemics and these have been recorded since the twelfth century. In the 1919 pandemic, which was the most devastating plague documented, there were more than 20 million deaths worldwide, i.e. more than those due to the first world war, and at least a hundred times this number of people were affected

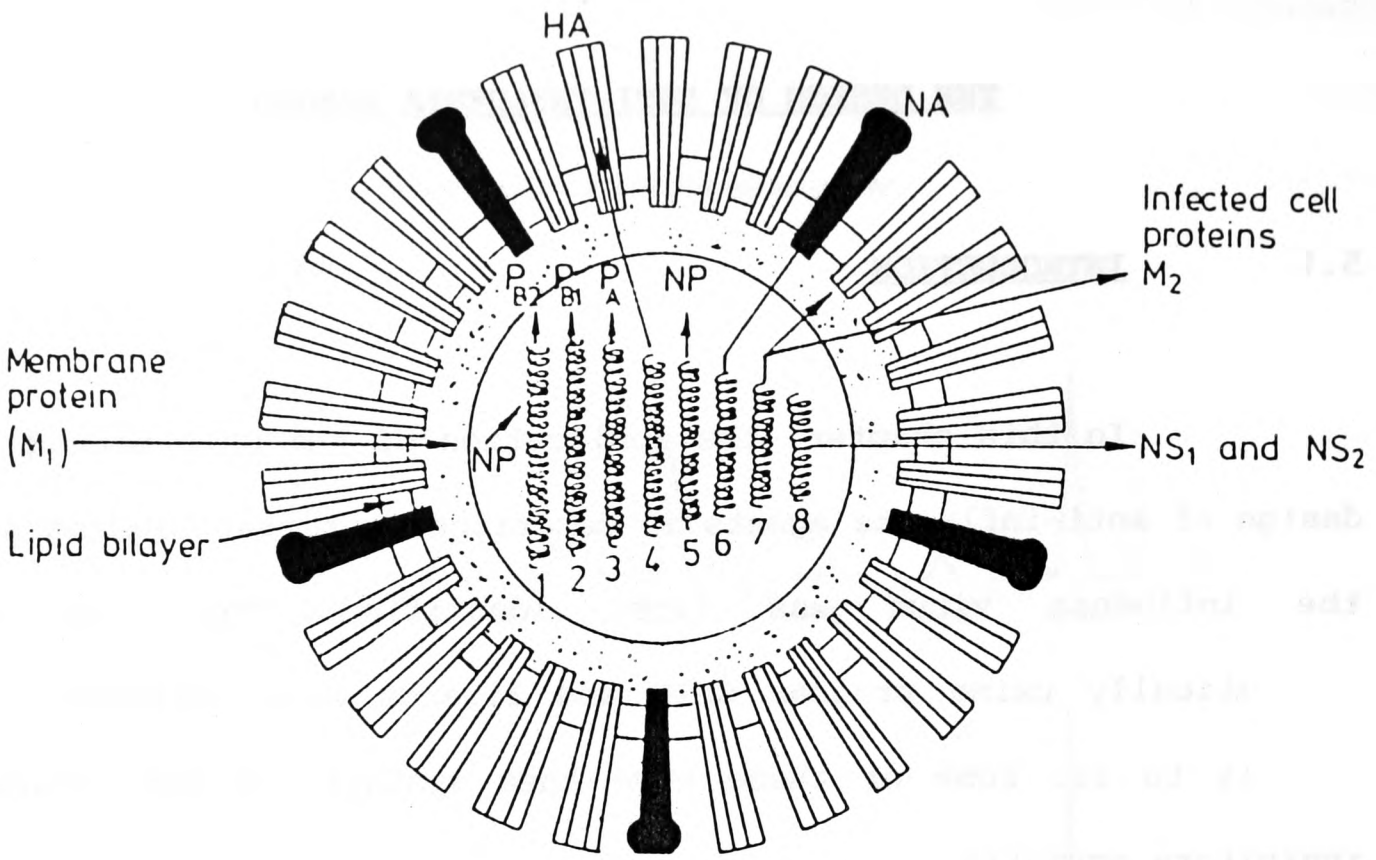


Figure 5.2 Model of the influenza virus showing protein and RNA composition (From reference [237]). See Section 5.1.2.1.

by the virus [235]. Figure 5.1 shows the effect of the more recent pandemic in 1968 and subsequent epidemics in 1972 and 1975.

5.1.2 The influenza virus

5.1.2.1 Structure

Influenza is an orthomyxovirus. The virion is approximately spherical and is 80-100nm in diameter. Its structure is shown in Figure 5.2 [237].

The viral core contains eight strands of single-stranded RNA of negative sense coding for more than ten proteins; a helical nucleocapsid protein (NP); three polymerase proteins (PB_1, PB_2 and PA) which are closely associated with the RNA and are involved in replication and transcription; and three or four virus encoded non-structural proteins of unknown function (NS_1, NS_2).

The membrane is composed of a lipid bilayer, which is acquired on 'budding' through the host-cell plasma membrane, surrounding a matrix or membrane protein (M_1) which is structurally important.

The virus coat is made up of two virally coded integral membrane glycoproteins which take the form of spikes anchored in the matrix protein. There are about 400 spikes on each virion comprising about 30% of the virion protein. The two glycoproteins are haemagglutinin (HA) which has a triangular cross-section and neuraminidase (NA) which has a tetrameric, mushroom-like shape. The HA is present in about four times the quantity of the NA.

5.1.2.2 Virus infection

The virus infects host cells in the epithelium of the upper and lower respiratory tract [234]. After attachment to cell-surface receptors containing sialic acid, the virus may be taken up by the host cells by endocytosis through 'coated pits' which are specialised plasma membrane sites thickened on the cytoplasmic side by clathrin [237,238]. The virion is then enclosed in an endosomal compartment of the cell. Induced by the low pH in the compartment [238,239], the viral membrane then fuses with the endosomal membrane and the viral nucleocapsid is released into the cytoplasm. Replication occurs and the progeny virions are released by budding through the cell membrane.

5.1.2.3 Antigenic variation

The influenza virus has a very large capacity to vary. Three types of influenza virus have been identified on the basis of the antigenic character of the nucleoprotein and the matrix protein [237]:

Type A which causes most epidemics and all pandemics in man as well as affecting animals and birds;

Type B which is less common than type A but occurs in humans and causes local outbreaks of the disease such as occur in schools;

Type C which occurs in humans, mostly children, and also pigs causing mild illness. It is uncommon and little studied. It has only one surface glycoprotein and this coats the virus in a hexagonal array.

The type A influenza virus was studied in this thesis. Many subtypes, and within subtypes, strains, of the type A virus exist due

to the structural mutability of the surface antigens. These give rise to epidemics of the disease as mutations of the viral proteins enable the virus to evade the infected organism's immune defence system.

There are two kinds of variations in the antigenic nature of the surface glycoproteins of influenza A.

1. **Antigenic drift**

This is the gradual accumulation of a series of point mutations resulting in a new strain of a particular subtype of virus. The strains within a single subtype generally differ in sequence by less than 10% [240]. Antigenic drift causes epidemics of varying frequency and impact.

2. **Antigenic shift**

This is the occurrence of a large number of mutations giving rise to a new subtype of virus with completely different antigenic properties. The subtypes show a 20-74% amino acid residue variation in their sequences [240]. They are identified by the nature of their HA and NA as HxNy where x,y=1,2,3... [241]. Antigenic shift causes serious epidemics and occurs every 10-20 years. The cause of antigenic shift is not known but the following proposals have been made [235,237].

1. It is caused by direct mutation of the virus. This is unlikely because it requires that multiple RNA base changes occur simultaneously.

2. A limited number of subtypes exist and 'recycling' of viruses occurs, e.g. the recurrence of the H1N1 subtype in 1977.

3. There is transmission of the virus to humans from non-human sources by genetic reassortment. For example, the 1968 subtype is thought to have arisen in China by the transfer to humans of an avian waterfowl virus which had undergone genetic reassortment in the pig [242].

4. Animal or avian viruses may become infectious to man due to some mutations. The number of mutations required for this to occur and the genes in which they would be located is unknown.

5. The virus may be carried by comets from space [243].

Each strain of virus can be identified according to the WHO system of nomenclature [241] by its type (A, B or C), host of origin (for non-human sources), geographical origin, strain number, year of isolation and subtype (HxNy) e.g. A/Hong Kong/1/68 (H3N2).

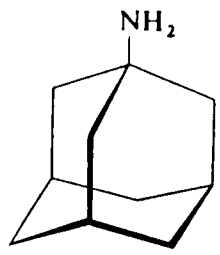
The three influenza A subtypes which have affected humans since the first human influenza virus was isolated in 1933 are the H1N1 'Russian' 'flu (1918), the H2N2 'Asian' 'flu (1957) and the H3N2 'Hong Kong' 'flu (1968) [235]. In humans, subtypes H1N1 and H3N2 of influenza A are currently prevalent and cause disease.

5.1.3 Current therapy of influenza

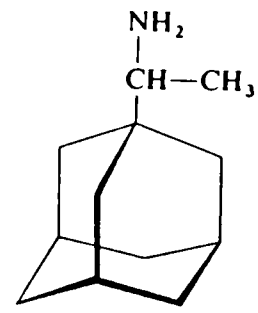
This is either by vaccination or by the use of therapeutic agents.

A vaccine of inactivated influenza virus provides only about 60% protection for a period of a few months [234] and it is recommended only for those who are particularly susceptible to infection. Inactivated vaccines of isolated HA and NA, synthetic

a.



b.



c.

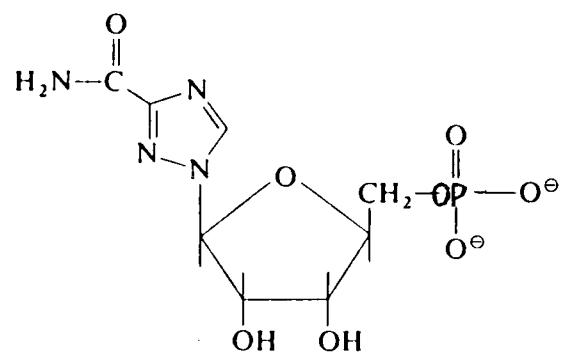


Figure 5.3 Current anti-influenza agents. a. Amantadine b. Rimantadine c. Ribavirin. See Section 5.1.3.

vaccines and live attenuated vaccines are under investigation [237,244].

A few therapeutic agents are available of which the principal ones are:

1. **Amantadine**

Amantadine (see Figure 5.3.a) is effective prophylactically and therapeutically against influenza A [245]. Its mode of action is unknown but it prevents cell penetration [246] leaving the virions attached to the cell surface where they are vulnerable to antibody attack. It is only used in high risk patients because resistance can be acquired, and it causes side-effects involving the central nervous system. It can also be beneficial in the treatment of Parkinson's disease where it acts as a cholinergic blocking agent [247].

2. **Rimantadine**

Rimantadine (see Figure 5.3.b) shows similar antiviral activity to amantadine but produces fewer side-effects [245]. It is thought to affect primary transcription or the process between virus uncoating and the beginning of transcription [248]. It is of use prophylactically and therapeutically and was effective in countering the 1977-78 epidemic caused by the recurrence of the H1N1 subtype. It lacks any effect on Parkinson's disease.

3. **Ribavirin**

Ribavirin (see Figure 5.3.c) is a broad spectrum agent which inhibits RNA and DNA viruses and is effective against influenza A and B. It may inhibit enzymes involved in mRNA synthesis [249].

Interferons may also be of therapeutic value as they have been found to protect cells against viral infection [250].

There are disadvantages to the current methods of therapy. Vaccines are only active against specific viral strains and lose their effectiveness when the virus mutates. Most therapeutic agents lack specificity for the influenza virus and cause side effects. There is thus a need for an effective, specific drug whose action is unaffected by the rapid mutation of the virus.

5.1.4 A proposed strategy for the design of anti-influenza agents

Using the method of receptor fit, it may be possible to design an anti-influenza agent that has the ability to counter all strains of influenza virus regardless of their antigenic variation. This is because a therapeutic agent can be designed so as to bind to a highly conserved region of the viral protein where it may interfere in the lifecycle of the virus. In order that it can be exploited by using the method of receptor fit, a target binding site should satisfy the following requirements.

1. It should have a known molecular structure.
2. It should be involved in a crucial stage of viral infection or replication so that ligand binding will cause inhibition of an important step in the viral lifecycle.
3. It should be composed of well-conserved residues so that binding of a ligand will be unaffected by mutations of the protein. This should enable the designed compound to be effective against as many strains of the influenza virus as possible.

4. It should possess residues capable of binding tightly and specifically to a ligand i.e. charged or polar residues.

5. It should not be highly mobile. If it were, a three-dimensional structure might not provide sufficient information on the real nature of the binding site for the design of a ligand. The GRID method of determining binding sites only takes into account the motion of hydrogen-bonding hydrogens and lone pairs on the target. Therefore, better predictions are likely to be made if the mode of binding is similar to the rigid 'lock and key' model rather than if the target is highly mobile and undergoes a large conformational change on ligand binding.

Having chosen a binding site satisfying these properties, a ligand may be designed using the following procedure:

1. Map the target binding site energetically using program GRID with different chemical probes which could be incorporated into a therapeutic agent.

2. Identify energy minima corresponding to favourable binding sites for the probes. Different energy minima may be identified for different chemical probes enabling specificity in ligand binding to be achieved. The positions of the most favourable energy minima are termed 'ligand points' and indicate the positions at which certain chemical groups should be incorporated in a strongly binding ligand.

3. Model-build a ligand which includes the appropriate probe groups at the ligand points and fits into the whole binding site. It should be constructed so as to optimise all of its properties that are important in the action of a drug. Conformational strain in the model-built ligand may be reduced by energy minimization.

4. Synthesise the designed ligand and assay it for anti-influenza activity.

All the properties of a ligand, including its binding energy, should be optimised in order for effective and efficient therapeutic action to be achieved. Some of the properties required of a therapeutic agent are considered below.

1. It should be very specific otherwise it may interfere with the normal metabolism of the host cells causing toxicity.

2. It should be of suitable rigidity. A rigid molecule would be more likely to bind the receptor site than a flexible one due to entropic factors. However, a rigid molecule would be less able to adapt to small conformational changes in the binding site.

3. It should have an appropriate lipophilicity. This can be assessed by the calculation of the octanol-water partition coefficient P [251].

4. It should have an appropriate pK_a.

5. It should be of appropriate bulk.

6. It should be metabolised and excreted at an appropriate rate. The rate of metabolism may be reduced by the incorporation into the ligand of chemical groups that are not found naturally.

7. It should be easy to synthesise. This facilitates the testing of its binding and activity. It also enables the production costs of a commercial drug to be kept down.

The strategy for the design of anti-influenza agents outlined in this section was employed and its application is now described in detail.

5.2 HAEMAGGLUTININ (HA)

Haemagglutinin (HA) was chosen as a suitable target for the design of a therapeutic agent by the method of receptor fit for the following reasons.

1. The structure of the Hong Kong HA from the A/Aichi/2/68 (H3N2) strain has been solved to 3Å resolution with an R-factor of 20.4% by x-ray crystallography [252,253].

2. It is the most well-characterised and studied viral protein [237].

3. It is important in the function of the influenza virus because it is responsible for the attachment of the virus to the host cell receptors by binding to a terminal sialic acid [254]. Once the virus is inside the host cell endosomal vesicles, HA mediates a low pH induced membrane fusion event [255] leading to the release of the viral nucleocapsid into the cytoplasm and the beginning of the replication cycle. HA also contains important antigenic determinants which stimulate the production of neutralizing antibody following infection or immunization [236].

4. It acts at an early stage of infection so inhibition by a therapeutic agent may prevent invasion of the host cell by the virus.

5. It contains regions of well-conserved residues despite having a higher rate of mutation than any of the other viral proteins.

Neuraminidase (NA) might also be a suitable target because its three-dimensional structure has been determined [256] and it has a large highly conserved catalytic site. However it is a less favourable target than HA for the following reasons.

1. It acts at a later stage in the process of infection and multiplication of the virus. It is responsible for the transport of the virus through the mucin on the cell walls during viral infection and release and also for the destruction of the HA receptor on the host cells allowing elution of the progeny virions from the infected cells and preventing self-aggregation of the virion particles [257].

2. HA is a more important determinant of host resistance than NA [258,259].

3. Antibodies to NA do not prevent infection but do prevent the release and spread of the virus from infected cells. Therefore, a compound designed to bind to NA is less likely to be completely effective against the virus than one binding to HA as it would prevent the spread of the virus but not the initial infection.

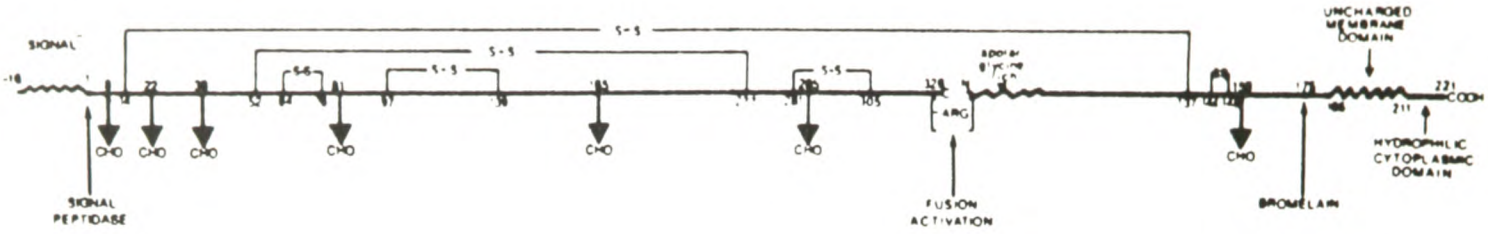
The three-dimensional structures of the other viral proteins have not been determined so they cannot be studied by the method of receptor fit.

5.2.1 The structure of HA

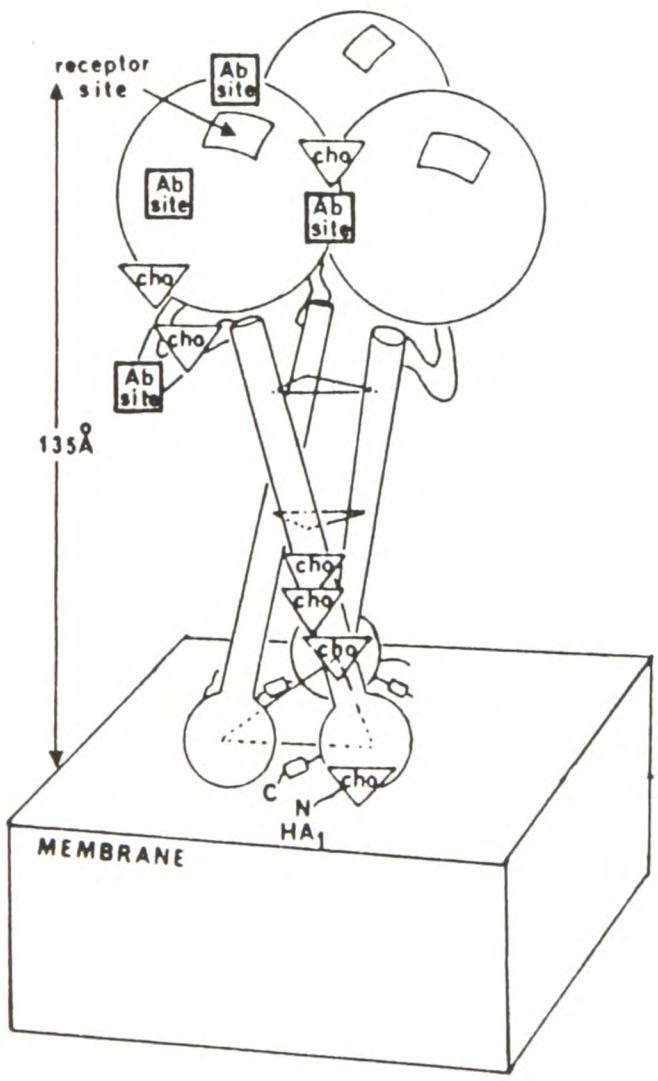
HA is a trimer of molecular weight ≈225kd [252]. Each monomer consists of two glycosylated polypeptides HA_1 (≈47kd) and HA_2 (≈28kd) connected by a disulfide bridge and derived by proteolysis of a single precursor molecule with the loss of a signal peptide (see Figure 5.4.a).

It has a three-domain structure typical of integral membrane glycoproteins [252] (see Figure 5.4). Each monomer consists of the following domains.

a.



b.



c.

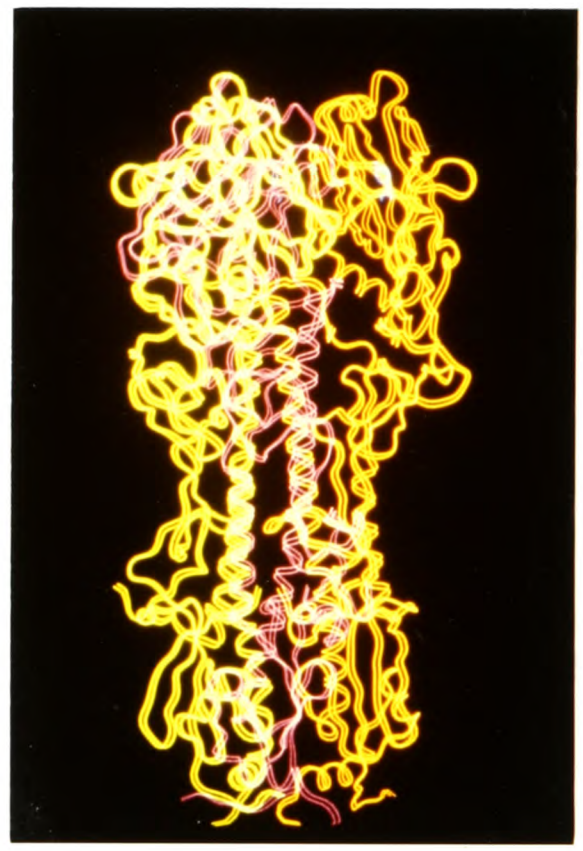


Figure 5.4 a. Schematic diagram of the sequence of the Hong Kong 1968 haemagglutinin showing the external domain (ends 185 HA_2), the transmembrane domain (185-211) and the cytoplasmic domain (212-221). (From reference [252]) b. Schematic diagram of the HA trimer showing the location of the host cell receptor sites, the four antigenic sites (squares) and the sites at which carbohydrate is attached to the protein (triangles).(From reference [252]). c. A ribbon plot of the HA trimer with the location of the host cell receptor site on each monomer indicated by the white tyrosine 98 residue in the 'head' of the protein. See Section 5.2.

1. A large external domain consisting of a globular 'head' containing an eight-stranded anti-parallel ß-sheet which is connected via an α-helical stalk 76Å long to the membrane. In the trimer, these helices form a triple stranded coiled coil. Each monomer has one host cell receptor site on the top of the globular head. There are also at least four topographically distinct antigenic sites on the globular head of each trimer. The carbohydrate residues are attached to the head and stalk regions of the trimer and may modulate the antigenicity of the protein [260].

2. A short transmembrane domain consisting of 24-28 uncharged residues.

3. A small cytoplasmic domain of ten or more residues.

The whole protein trimer has a length of 135Å and a cross-sectional radius of 15-40Å.

5.2.2 The conservation of residues in HA

Four regions of HA are composed of highly conserved residues and may therefore be potential targets for the design of anti-influenza agents. They are as follows.

1. The membrane fusion site at the N-terminus of HA_2. This is very non-polar. Glu 11 is the first charged residue in HA_2 and seven of the first 23 residues are glycines. The first eleven residues have an unusual conformation of sharp bends [252]. The non-polar nature of this site would make the design of a ligand which bound specifically to this region of the protein difficult.

2. The stalk region composed of residues 279-328 of HA_1 (see

Table 5.1 The amino acid sequences of the HA1 peptide of HA from different strains of influenza A virus showing the conservation of the residues in the host cell receptor site (adapted from references [261,262]).

Key: * residues conserved in all strains sequenced.
= residues conserved in H3N2 strains sequenced.
+ residues in the host cell receptor (HCR) site.

```
                        10         20         30         40         50         60         70         80
H1N1 PR34                  DTICIGYHAN NSTDTVDTVL EKNVTVTHSV DILEKTHNGK LCRLKGIAPLQ LGKCNIAGWL LGNPECDPLLP
H2N2 JAP57                 DQICIGYHAN NSTEKVDTNL ERNVTVTHAK DILEKTHNGK LCKLNGIPPLE LGDCSIAGWL LGNPECDTLLS
H3N2 NT68     QDLPGNDNNT ATLCLGHHAV PNGTLVKTIT DDQIEVTNAT ELVQSSSTGK ICN-NPHRILD GIDCTLIDAL LGDPHCD-VFQ
     ENG69    QDLPGNDNST ATLCLGHHAV PNGTLVKTIT NDQTEVTNAT ELVQSSSTGK ICN-NPHRILD GINCTLIDAL LGDPHCD-VFQ
     QU70     QDLPGNDNST ATLCLGHHAV PNGTLVKTIT NDQTEVTNAT ELVQSSSTGK ICN-NPHRILD GIDCTLIDAL LGDPHCD-VFQ
     MEM72    QDLPGNDNST ATLCLGHHAV PNGTLVKTIT NDQTEVTNAT ELVQSSSTGK ICN-NPHRILD GINCTLIDAL LGDPHCD-GFQ
     VIC375   QDLPGNDNST ATLCLGHHAV PNGTLVKTIT NDQTEVTNAT ELVQSSSTGK ICN-NPHRILD GINCTLIDAL LGDPHCD-GFQ
     TEX77    QNLPGNDNST ATLCLGHHAV PNGTLVKTIT NDQTEVTNAT ELVQSSSTGR ICD-SPHRILD GKNCTLIDAL LGDPHCD-GFQ
     BK79     QNLPGNDNST ATLCLGHHAV PNGTLVKTIT NDQTEVTNAT ELVQSSSTGR ICD-SPHRILD GKNCTLIDAL LGDPHCD-GFQ
H7   FPV34                 DKICLGHHAV SNGTKVNTLT ERGVEVVNAT ETVERTNIPK ICS-KGKRTTD LGQCGLLGTI TGPPQCD-QFL
TOTAL                        * * **       * *        *                   *             *        * * **
H3N2           = ====== = ========== ==========  ========= =========  ==   ====== =  ======= =======  ==
HCR SITE

                        90         100        110        120        130        140        150        160
H1N1 PR34      VRSWSYIVET PNSENGICYPG DFIDYEELRE QLSSVSSFER FEIFPKESSWPNH NTTKGVTAACS HAGKSSFYRN LLWL--TEKEGS
H2N2 JAP57     VPEWSYIMEK ENPRDGLCYPG SFNDYEELKH LLSSVKHFEK VKILPKD-RWTQH TTTGGSR-ACA VSGNPSFFRN MVWL--TKEGSD
H3N2 NT68      NETWDLFVER SKAF-SNCYPY DVPDYASLRS LVASSGTLEF ITEGF---TWTGV TQNGGSN-ACK RGPGSGFFSR LNWL--TKSGST
     ENG69     DETWDLFVER SKAF-SNCYPY DVPDYASLRS LVASSGTLEF ITEGF---TWTGV TQNGGSN-ACK RGPGSGFFSR LNWL--TKSGST
     QU70      NETWDLFVER SKAF-SNCYPY DVPDYASLRS LVASSGTLEF ITEGF---TWTGV TQNGGSN-ACK RGPGSGFFSR LNWL--TKSGST
     MEM72     NETWDLFVER SKAF-SNCYPY DVPDYASLRS LVASSGTLEF ITEGF---TWTGV TQNGGSN-ACK RGPGSGFFSR LNWL--TKSGST
     VIC375    NEKWDLFVER SKAF-SNCYPY DVPDYASLRS LVASSGTLEF ITEGF---TWTGV TQNGGSN-ACK RGPGSGFFSR LNWL--TKSGST
     TEX77     NEKWDLFVER SKAF-SNCYPY DVPDYASLRS LVASSGTLEF ITEGF---TWTGV TQNGGSN-ACK RGPGSGFFSR LNWL--TKSGST
     BK79      NEKWDLFVER SKAF-SNCYPY DVPDYASLRS LVASSGTLEF ITEGF---TWTGV TQNGGSN-ACK RGPGSGFFSR LNWL--TKSGST
H7   FPV34     EFSADLIIER REGN-DVCYPG KFVNEEALRQ ILRGSGGIDK ETMGF---TYSGI RTNGTTS-ACR RSGSS-FYAE MEWLLSNTDNAS
TOTAL                  *          ***         *                                      **         *      **
H3N2            = ======= ==== ====== ========== ========== =====   ===== ======= === ========== ====  ======
HCR SITE                          +                                    +++++ +              ++   +
```

```
                      170        180        190        200        210        220        230        240
H1N1  PR34     YPKLKNSYVN KKGKEVLVLW GIHHPSNSKD QQNIYQNENA YVSVVTSNYN RRFTPEIAER PKVRDQAGRM NYYWTLLKPG
H2N2  JAP57    YPVAKGSYNN TSGEQMLIIW GVHHPIDETE QRTLYQNVGT YVSVGTSTLN KRSTPGIATR PKVNGQGGRM EFSWTLLDMW
H3N2  NT68     YPVLNVTMPN NDNFDKLYIW GVHHPSTNQE QTSLYVQASG RVTVSTRRSQ QTIIPNIGSR PWVRGLSSRI SIYWTIVKPG
      ENG69    YPVLNVTMPN NDNFDKLYIW GVHHPSTNQE QTSLYVQASG RVTVSTRRSQ QTIIPNIGSR PWVRGLSSRI SIYWTIVKPG
      QU70     YPVLNVTMPN NDNFDKLYIW GVHHPSTNQE QTSLYVQASG RVTVSTRRSQ QTIIPNIGSR PWVRGQSSRI SIYWTIVKPG
      MEM72    YPVLNVTMPN NDNFDKLYIW GVHHPSTNQE QTSLYVQASG RVTVSTKRSQ QTIIPNIGSR PWVRGQSSRI SIYWTIVKPG
      VIC375   YPVLNVTMPN NDNFDKLYIW GVHHPSTDKE QTNLYVQASG KVTVSTKRSQ QTIIPNVGSR PWVRGLSSRI SIYWTIVKPG
      TEX77    YPVLNVTMPN NDNFDKLYIW GVHHPSTDKE QTNLYVQASG RVTVSTKRSQ QTIIPNVGSR PWVRGLSSRI SIYWTIVKPG
      BK79     YPVLNVTMPN NDNFDKLYIW GVHHPSTDKE QTNLYVRASG RVTVSTKRSQ QTIIPNIGSR PWVRGLSSRI SIYWTIVKPG
H7    FPV34    FPQMTKSYKN TRRESALIVW GIHHSGSTTE QTKLYGSGNK LITVGSSKYH QSFVPSPGTR PQINGQSGRI DFHWLILDPN
TOTAL          *       *       *  * * **       *   *         *           *    * *       *     *
H3N2           ========== ========== = =====  = == === ===  ===== === ====== === ===== ==== ==========
HCR SITE                              +  +   +    ++                              +++++

                      250        260        270        280        290        300        310        320
H1N1  PR34     DTIIFEANGN LIAPRYAFAL SRGFGSGIIT NASMHECNTK CQTPLGAINS SLPFQNIHPV TIGECPKYVR SAKLRMVTGL RNIPSIQS
H2N2  JAP57    DTINFESTGN LIAPEYGFKI SKRGSSGIMKT EGTLENCETK CQTPLGAINT TLPFHNVHPL TIGECPKYVK SEKLVLATGL RNVPQIES
H3N2  NT68     DVLVINSNGN LIAPRGYFKM -RTGKSSIMRS DAPIDTCISE CITPNGSIPN DKPFQNVNKI TYGACPKYVK QNTLKLATGM RNVPEKQT
      ENG69    DVLVINSNGN LIAPRGYFKM -RTGKSSIMRS DAPIDTCISE CITPNGSIPN DKPFQNVNKI TYGACPKYVK QNTLKLATGM RNVPEKQT
      QU70     DVLVINSNGN LIAPRGYFKM -RTGKSSIMRS DAPIDTCISE CITPNGSIPN DKPFQNVNKI TYGACPKYVK QNTLKLATGM RNVPEKQT
      MEM72    DILVINSNGN LIAPRGYFKM -RTGKSSIMRS DAPIGTCISE CITPNGSIPN DKPFQNVNKI TYGACPKYVK QNTLKLATGM RNVPEKQT
      VIC375   DILVINSNGN LIAPRGYFKM -RTGKSSIMRS DAPIGTCSSE CITPNGSIPN DKPFQNVNKI TYGACPKYVK QNTLKLATGM RNVPEKQT
      TEX77    DILLINSNGN LIAPRGYFKI -RTGKSSIMRS DAPIGTCSSE CITPNGSIPN DKPFQNVNKI TYGACPKYVK QNTLKLATGM RNVPEKQT
      BK79     DILLINSNGN LIAPRGYFKI -RTGKSSIMRS DAPIGTCSSE CITPNGSIPN DKPFQNVNKI TYGACPKYVK QNTLKLATGM RNVPEKQT
H7    FPV34    DTVTFSFNGA FIAPNRASFL -RGKSMGIQSD VQVDANCEGE CYHSGGTITS RLPFQNINSR AVGKCPRYVK QESLLLATGM KNVPEPSK
TOTAL          *       *   ***              *         *    *    * *     ** *       * ** **     *   **   * *
H3N2           = = ====== =========   ========== ==== == == ========== ========== ========== ========== ========
HCR SITE
```

All strains listed are human except H7 which is an avian strain.
The three dimensional structure of HA from AICHI68 has been determined [252] and has the same HA1 sequence as NT68 except that N 9 is mutated to S and V 182 is mutated to I.
Residues 279-328 form part of the HA stalk and are highly conserved.

Table 5.1) and much of HA_2 [235,261]. The binding of a ligand to this region may not prevent the normal functioning of HA.

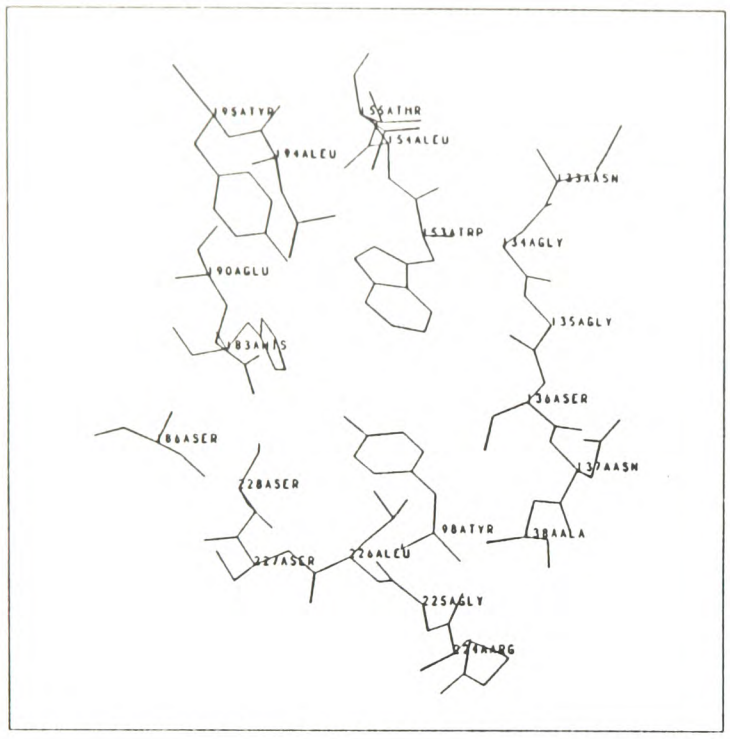
3. The intersubunit contacts at the top of the protein. These may be masked by carbohydrate [252,253] and so this region would probably be inaccessible to a drug.

4. The host-cell receptor site which is the only exposed surface on the top of the HA molecule which has not changed during antigenic drift since 1968 [252,254]. It is composed of residues from the HA_1 peptide whose sequence is given in Table 5.1 for different strains of influenza A. Comparison of the sequences shows that many of the residues in the host cell receptor site exhibit a high degree of conservation over a range of subtypes.

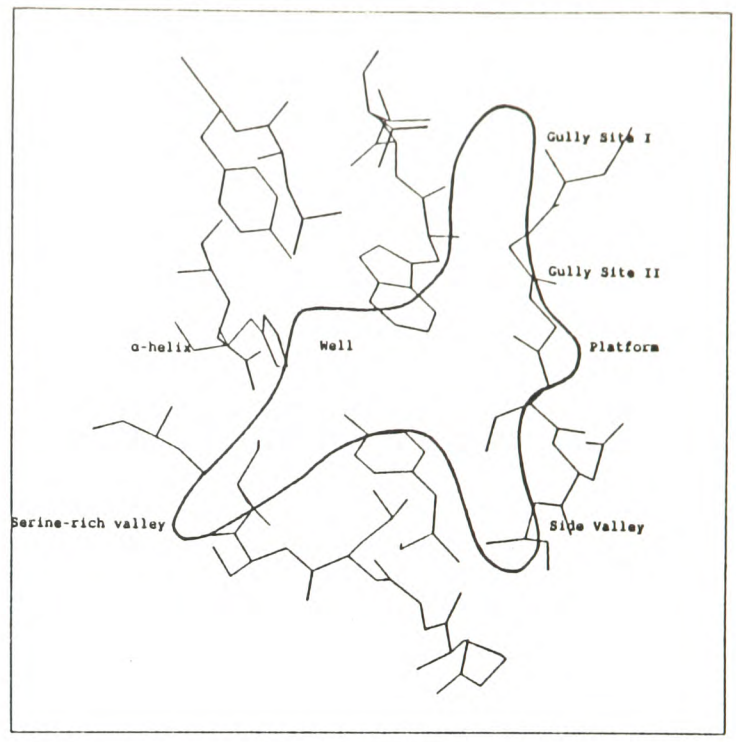
This site satisfies the properties required of a target binding site as described in Section 5.1.4 and it has recently been proposed [254] as a suitable site for the binding of a drug. It is functionally important as it is the site at which the terminal sialic acid of the host oligosaccharide binds. It contains a number of highly conserved residues of low to moderate mobility and it is rich in polar residues. It was therefore chosen as the target for the design of an anti-influenza agent that would act by binding tightly to it and preventing the host-cell receptor, sialic acid, from binding.

5.3 THE HOST CELL RECEPTOR SITE

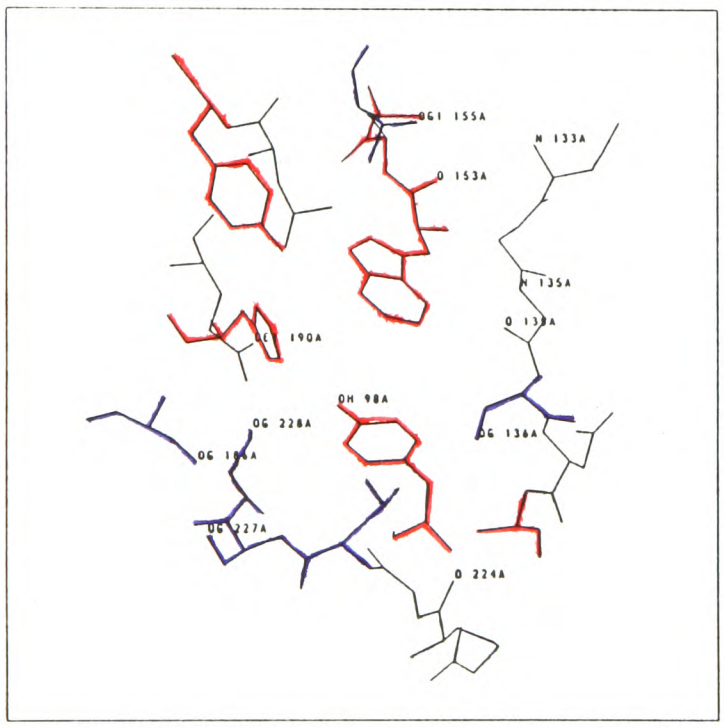
The host cell receptor site is situated in the globular 'head' of the protein as shown in Figure 5.4.c. It is surrounded by a 'rim' of very variable residues with antigenic character [254].



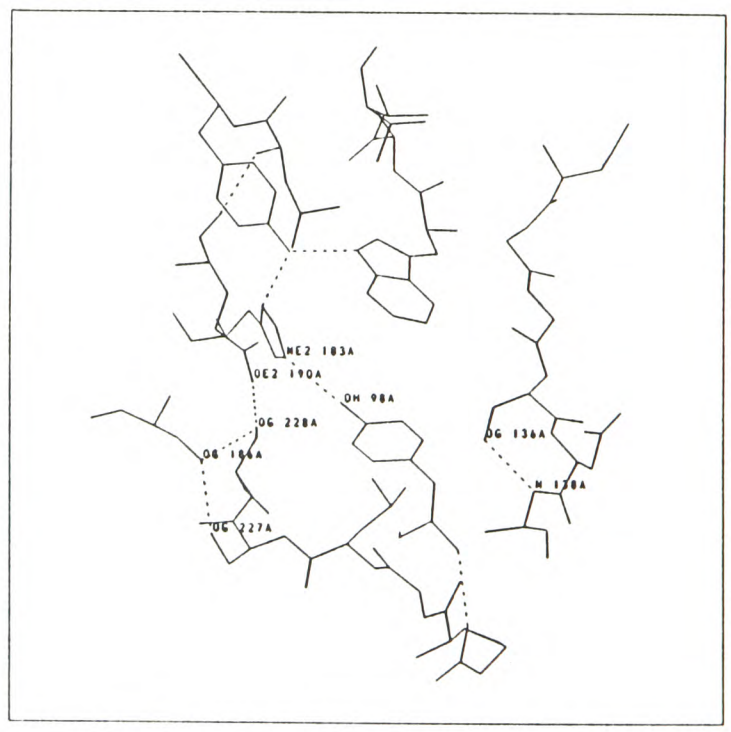
a.



b.



c.



d.

Figure 5.5 a. The haemagglutinin host cell receptor site residues. b. A schematic diagram showing the structure and 'geography' of the host cell receptor site. c. The conservation of the residues of the host cell receptor site is shown. Those in red are totally conserved and provide the basic shape of the receptor site. Those in black undergo some mutations but these should not affect ligand binding. Those in blue undergo significant mutation in some strains. The polar atoms that may participate in ligand binding are labelled. d. Intramolecular hydrogen-bonds in the receptor site are shown. See Section 5.3.

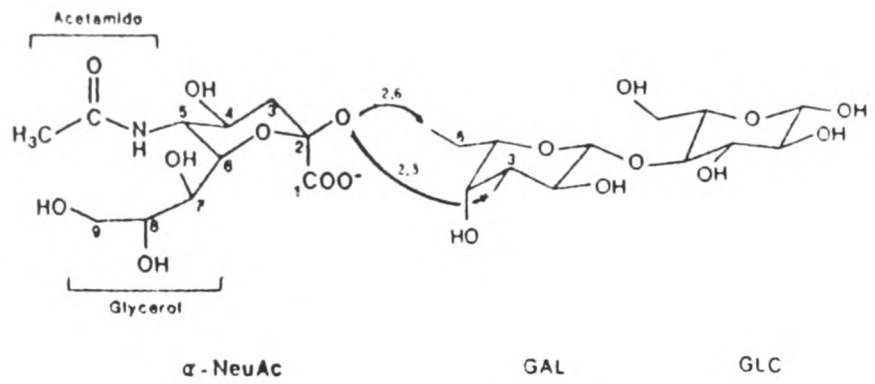
The structure of the host cell receptor site is shown in Figure 5.5. It is viewed from on top with Tyr 98 at the base of the shallow depression. Glu 190 and Leu 194 extend from a short α-helix (see Figure 5.5.b) which is part of one of the antigenic sites of the protein [236]. Its structure can be thought of in geographical terms [263] as a deep well with two valleys and a gully adjoining it and a slight depression or 'platform' to one side as shown in Figure 5.5.b.

The conservation of the residues of the pocket is shown in Figure 5.5.c. The totally conserved residues provide the basic structure of the site. Trp 153 makes an edge-to-face stack with Phe 147 and Phe (or Tyr) 148 (not shown). There is a chain of intramolecular hydrogen-bonds, shown in Figure 5.5.d, linking Trp 153 to Tyr 195 to His 183 to Tyr 98 and then to Glu 190 via a water molecule. These interacting residues may stabilise the architecture of the pocket [254].

There are a number of variable serine residues in the 'serine-rich valley' (see Figure 5.4.b). These serines mutate to residues with hydrocarbon side chains (Gly, Ala, Ile) in other subtypes of influenza A thus losing their functional hydroxyl group (see Table 5.1). Therefore, despite the fact that this region may bind to charged, polar groups strongly, an anti-influenza agent should not be designed to bind to the residues in this region because the extent of ligand binding may be highly susceptible to residue mutation.

Residue 226 is also highly variable and the specificity of the receptor site for certain sialosides is very dependent on the identity of this residue [264]. If residue 226 is a leucine, HA binds preferentially to a NeuAcα2-->6Gal linkage (see Figure 5.6.a), while if it is a glutamine, HA has affinity for a NeuAcα2-->3Gal linkage.

a.



b.

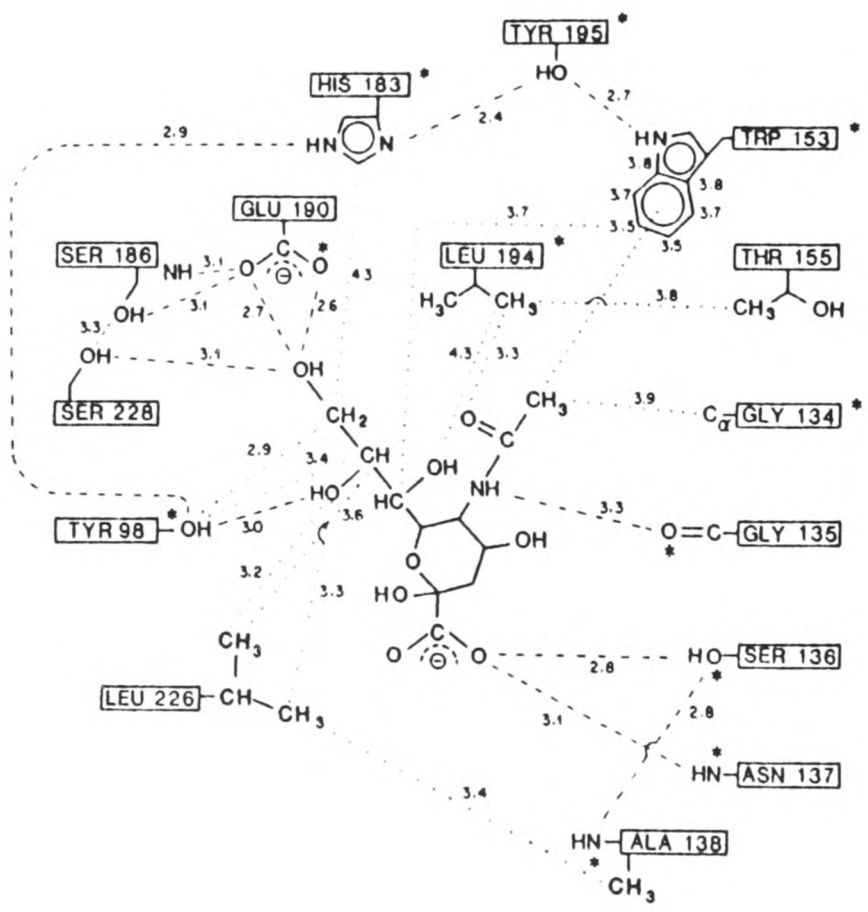


Figure 5.6 a. The structure of sialyllactose (α-N-acetyl neuraminic acid-α(2,3/6)-galactose-ß-(1,4)-glucose (NeuAcα(2,3/6)Galß(1,4)Glc)) showing the position of the α(2,6) and the α(2,3) linkages. b. The potential interactions of sialic acid with HA observed crystallographically. Potential hydrogen-bonds (dashed lines) and van der Waals' contacts (dotted lines) are shown with distances in Å.
* indicates conservation in all known HA sequences from natural isolates. (From reference [254]).

The change in the binding affinity observed on the mutation of residue 226 does not appear to be due to direct interactions with the glycosidic linkage. Instead, it appears to be due to small conformational changes in the protein [254].

Recently, complexes of wild-type HA with α(2,6)sialyl lactose, and of a mutant HA with α(2,3)sialyl lactose (see Figure 5.6.a) have been observed by x-ray crystallography and by nmr [254]. These studies show that the sialic acid binds to the conserved residues of the binding pocket. The shape of sialic acid is complementary to that of the host cell receptor site. The ring substituents specific to sialic acid interact with the protein while the ring atoms and the 4-hydroxyl group do not (see Figure 5.6.b). There is no observed electron density for the lactose moiety suggesting that this is very flexible.

In addition to being well-conserved, the host cell receptor site fulfills the other requirements of a target necessary for investigation by the method of receptor fit. In particular, it contains a number of charged and polar groups, as shown in Figure 5.5.c, so that it should be possible to design a ligand which binds to it strongly and specifically.

In addition, the residues in the site are of moderate mobility. Their temperature factors are mostly in the range 12-23Å^2 while the average for the protein structure is 17Å^2 and 100 residues on each monomer have atoms with B-values greater than 25Å^2 [253]. A notable exception is Glu 190 whose atoms have an average temperature factor of 34Å^2.

5.4 THE IDENTIFICATION OF LIGAND POINTS USING PROGRAM GRID

The interaction energies of a variety of probes with HA were calculated, using program GRID, at positions of the probe at regular intervals of 0.5Å over the whole of the binding site and these were then displayed as contoured energy maps.

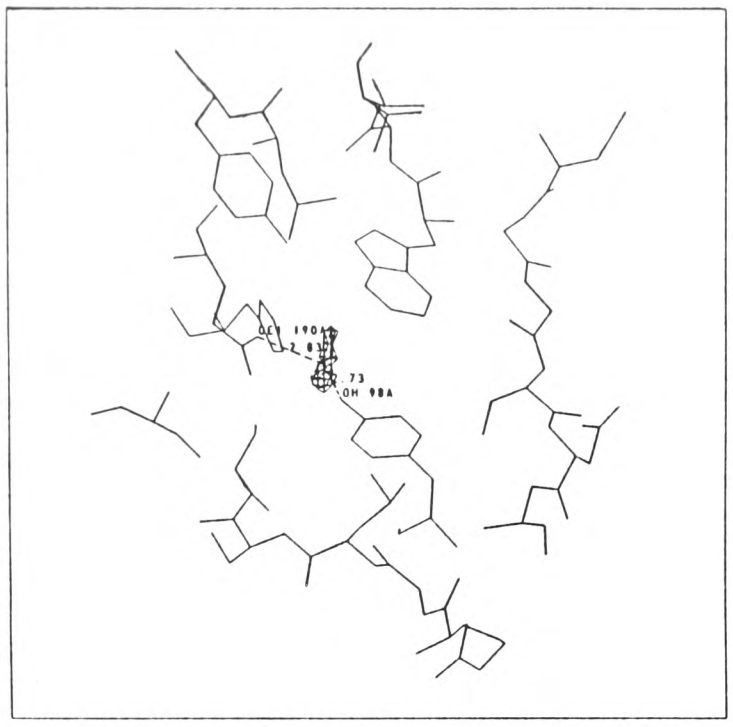
The coordinates of HA were obtained from the Brookhaven Protein Databank [52].

5.4.1 Methyl probe

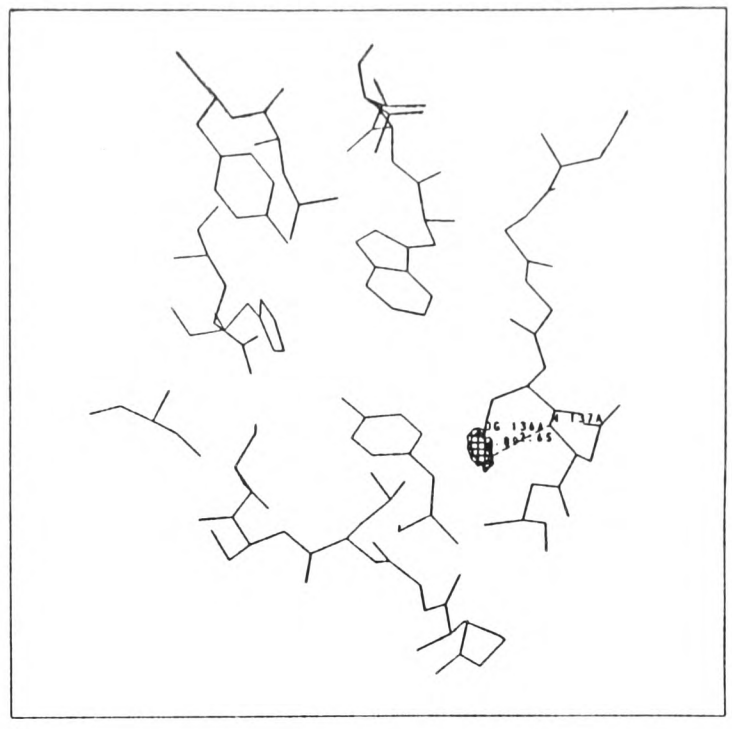
Firstly, the interaction of an uncharged methyl probe with HA was calculated. A GRID energy map contoured at a slightly repulsive energy forms a surface indicating the shape of the receptor binding pocket (see Figure 5.7). This surface is similar to a solvent accessible surface [58] but was determined using energetic rather than steric criteria and was calculated with a probe of van der Waals' radius of 1.95Å.

5.4.2 Other probes

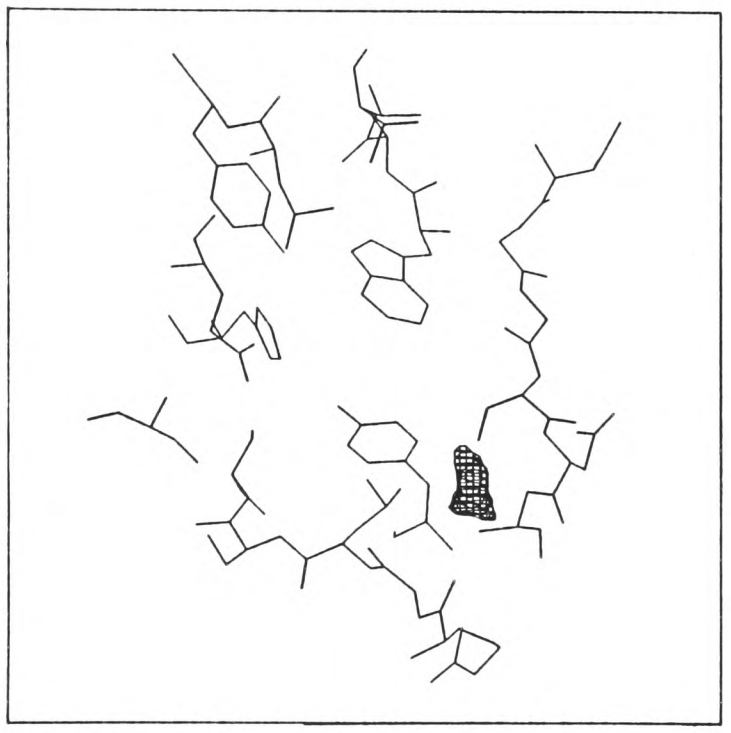
The interactions of the following probes with HA were calculated by program GRID: ammonium nitrogen $-NH_3^+$, amide nitrogen $-NH_2$, amide nitrogen -NH, aromatic nitrogen >N, ether oxygen >O, carbonyl oxygen =O, carboxyl oxygen -O, hydroxyl oxygen -OH, chlorine -Cl, fluorine -F and water H_2O. Regions of the host cell receptor site were identified where a particular probe was found to have a minimum interaction energy and was predicted to bind more strongly than any of



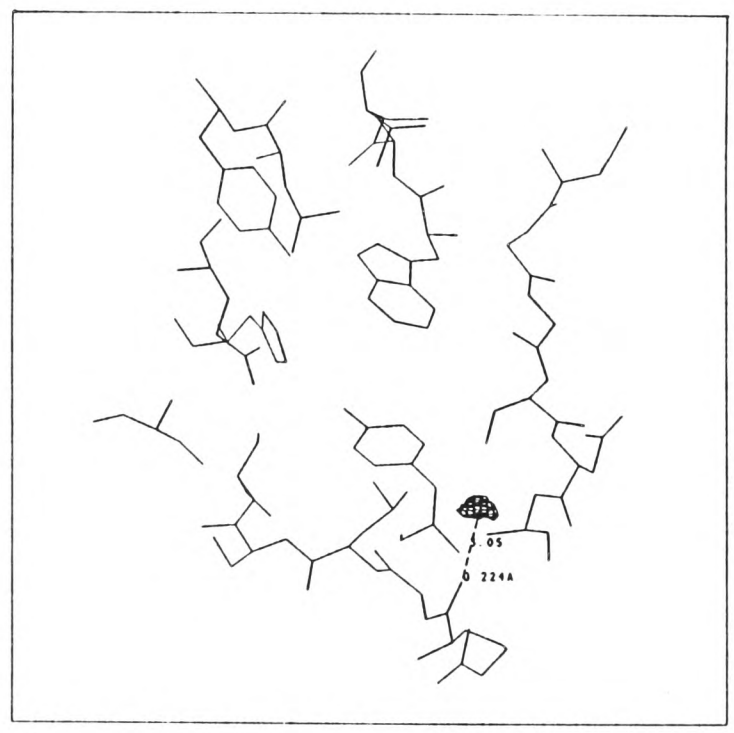
a.



b.

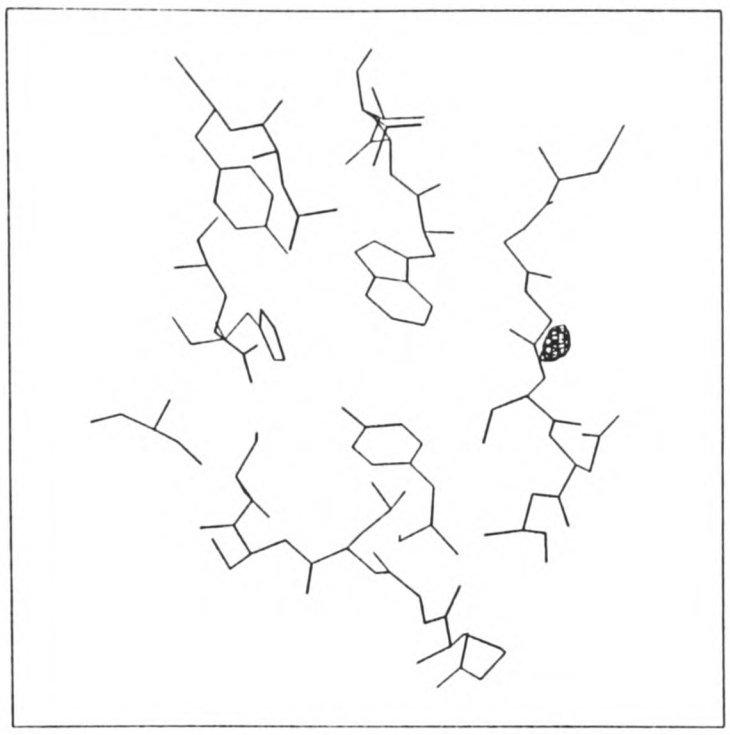


c.

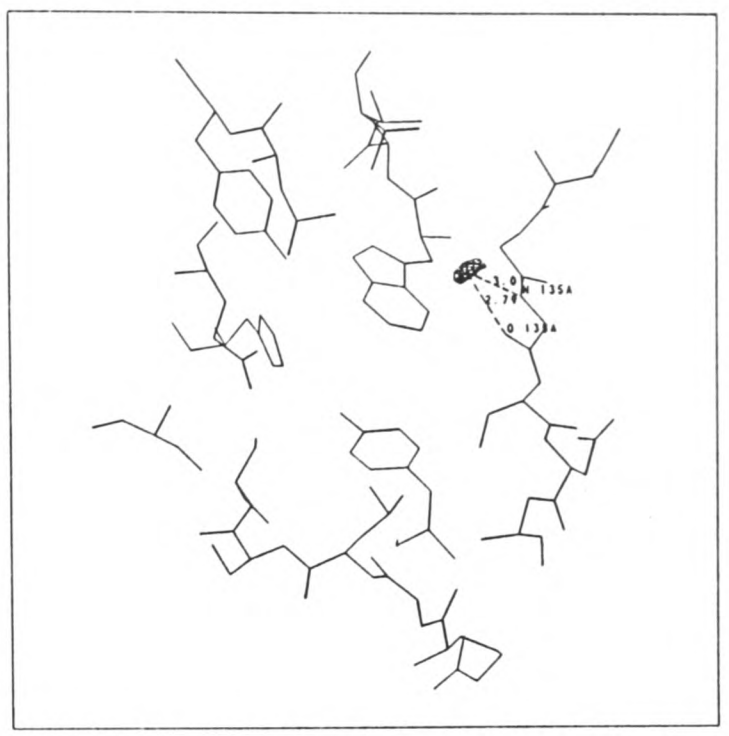


d.

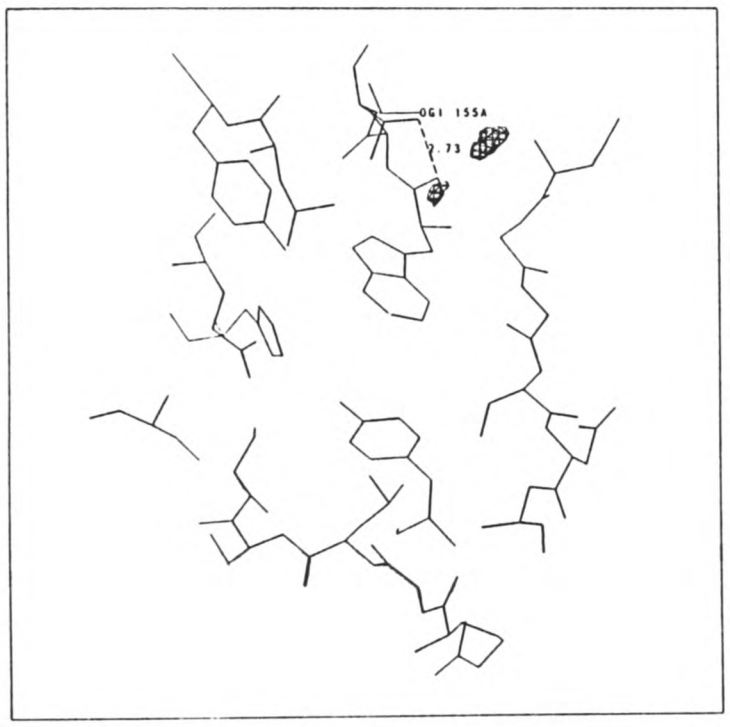
Figure 5.8 The haemagglutinin host cell receptor site with GRID energy contours shown for different probes. a. Energy contours at -10 kcal/mol around the energy minimum for an ammonium nitrogen probe in the well. b. Energy contours at -10 kcal/mol for the interaction of a carboxyl oxygen probe in the side valley. c. Energy contours at -6 kcal/mol for an amide nitrogen probe in the side valley. d. Energy contours at -6 kcal/mol for an amide nitrogen probe in the side valley. In this map, OG Ser 136 is not permitted to hydrogen-bond to the probe because it is assumed that it is already donating one hydrogen-bond to a carboxyl oxygen positioned at the ligand point shown in Figure 5.8.b. e. Energy contours at -5 kcal/mol for a chlorine probe in the platform region. f. Energy contours at -6 kcal/mol for a hydroxyl probe at gully site II. g. Energy contours at -10 kcal/mol for a carboxyl oxygen probe at gully site I. h. Energy contours at -10 kcal/mol for a carboxyl oxygen probe at gully site I. In this map, OG Thr 155 is not permitted to donate a hydrogen-bond as it is assumed that it is already donating one hydrogen-bond to a carboxyl oxygen positioned at the ligand point shown in Figure 5.8.g. i. The seven ligand points identified for the particular probe groups shown labelled. See Section 5.4.2.



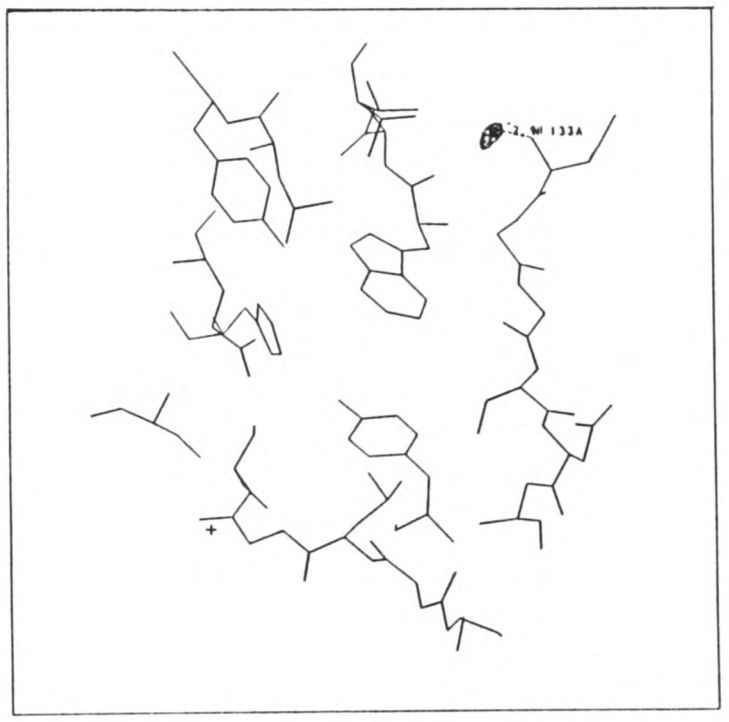
e.



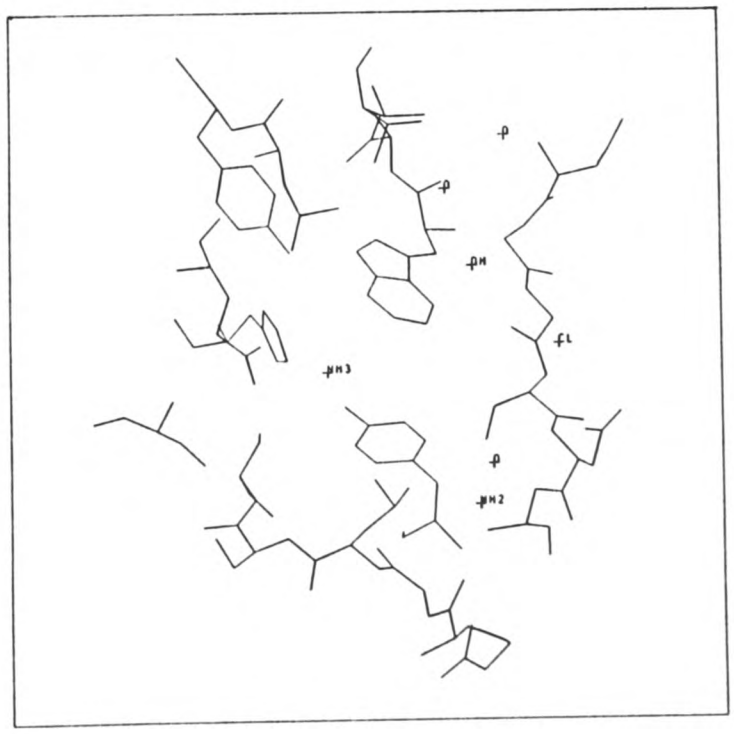
f.



g.



h.



i.

Table 5.2 Ligand points in the haemagglutinin host cell receptor site predicted by program GRID.

Region of binding pocket[a]	Probe	E_{pred} in kcal/mol[b]	Hydrogen-bond partners in HA		Hydrogen-bond energy in kcal/mol
Well	Ammonium nitrogen $-NH_3^+$	-11.80	OH	Tyr 98	-1.58
			OE1	Glu 190	-1.86
Side valley	Carboxyl oxygen -O	-11.56	OG	Ser 136	-3.39
	Amide nitrogen $-NH_2$	-6.67	O	Arg 224	-2.68
Platform	Chlorine -Cl	-6.07			
Gully site II	Hydroxyl oxygen -OH	-6.84	O	Gly 135	-3.97
			N	Gly 135	-0.45
Gully site I	Carboxyl oxygen -O	-11.26	OG1	Thr 155	-3.86
	Carboxyl oxygen -O	-11.01	N	Asn 133	-2.57

[a] The location of these regions in the binding pocket is shown in Figure 5.5.b.

[b] E_{pred} is the minimum interaction energy calculated by program GRID between the probe and the protein at the region of the binding pocket considered.

the other probes. The coordinates of the energy minima in each of these regions were refined by rerunning program GRID with an array of spacing 0.2Å over a small volume surrounding the energy minimum and the resultant position of each energy minimum was identified as a ligand point. Some of the probes were predicted to interact with HA more weakly than all the other probes at all points in the binding site and hence, no ligand points were identified for these probes. The positions and binding energies for the ligand points found in different regions of the receptor site are given in Table 5.2 and displayed in Figure 5.8.i. The GRID energy contour maps for these ligand points are shown in Figure 5.8. They are discussed below for each region of the receptor pocket (as defined in Figure 5.5.b).

The well

A binding site for an ammonium nitrogen NH_3^+ group was predicted at the base of the binding pocket and is shown by the contours in Figure 5.8.a. The probe was predicted to make two hydrogen-bonds of length 2.73Å and 2.83Å to the protein at this position. A third hydrogen-bond to NεHis 183 at a distance of 3.3Å was not predicted because it would be unfavourably oriented at the probe in relation to the other two predicted hydrogen-bonds and because the histidine nitrogen can make an intramolecular hydrogen-bond to OH Tyr 98 which is 2.85Å away. The ammonium nitrogen probe was predicted to bind more strongly in the well region than any other probe studied.

A water molecule has been observed experimentally in this region forming a bridge between Glu 190 and Tyr 98 [254]. Program GRID predicted a water binding site in the well with a binding energy

E_{pred}=-9.2 kcal/mol which corresponded to this observed water position. However, program GRID predicted that an ammonium probe group would displace this water molecule because it was calculated to have a more favourable binding energy E_{pred}=-11.80 kcal/mol.

The side valley

A negatively charged carboxyl oxygen was predicted to bind in the side valley more strongly than any other probe. Energy contours for its interaction with HA are shown in Figure 5.8.b. At the energy minimum, it was predicted to accept a hydrogen-bond from OG Ser 136. The GRID carboxyl oxygen probe has the size of only one atom. However, in a ligand, it is part of a larger, planar carboxylate or amide group. Therefore, for this ligand point to be occupied by a carboxyl oxygen, another carboxyl oxygen or amide nitrogen atom must also be able to interact favourably in the side valley at the same time. In addition, the side valley must be large enough to accommodate the whole carboxylate or amide group in an orientation which allows favourable hydrogen-bonds to be made to the protein.

An amide nitrogen probe interacted less favourably than a carboxyl oxygen in the side valley. It had an energy minimum at -7.58 kcal/mol near that of the carboxyl oxygen probe where it could hydrogen-bond to OG Ser 136 and O Arg 224. This was within the energy contours shown in Figure 5.8.c. However, it is not possible for OG Ser 136 to form strong hydrogen-bonds to both the oxygen and the nitrogen of an amide group positioned in the side valley at the same time because the angle subtended by these two hydrogen-bonds at OG Ser 136 would only be about 30-40°. Therefore, program GRID was run for both a carboxyl oxygen probe and an amide nitrogen probe in the side

valley region with no hydrogen-bonds permitted between the probe and OG Ser 136.

The carboxyl probe now had an energy minimum of -8.71 kcal/mol. However, the position of this energy minimum was only 0.75Å from that of the energy minimum predicted for a carboxyl probe which could hydrogen-bond to OG Ser 136, and so both these energy minima could not be occupied at the same time. An amide nitrogen now had an energy minimum of -6.67 kcal/mol, which was less favourable than for a carboxyl oxygen, but this energy minimum was further from the first carboxyl oxygen ligand point at a distance of 1.7Å and at a position where it could donate a hydrogen-bond to O Arg 224 (see Figure 5.8.d). A carboxyl oxygen probe could interact with a similar energy at this position but it was not able to make such a specific interaction with the protein because it could not form a hydrogen-bond to O Arg 224.

Therefore, a ligand amide group could be oriented in the side valley and was predicted to interact favourably with the protein. A carboxylate group could also fit in this valley but it would probably make a less favourable and specific interaction with the protein.

The platform

An energy minimum for a chlorine probe was found in this region and is shown in Figure 5.8.e. A more favourable interaction was predicted for a carboxyl oxygen but there was insufficient space to fit a whole carboxylate group into this region of the binding site. Chlorine bound more strongly at this region than all the other probes because it was able to make large favourable dispersive interactions.

The gully site II

A small favourable binding region (see Figure 5.8.f) was found here for a hydroxyl group where it could make hydrogen-bonds to the nitrogen and the oxygen of Gly 135. Water, (with E_{pred}=-6.52 kcal/mol), had almost the same binding energy as a hydroxyl group at this position because it could make hydrogen-bonds to the same protein atoms.

The gully site I

A carboxyl oxygen interacted more favourably than any other probe at this site. Two minima of very similar binding energy were detected as shown in Figure 5.8.g, but at both, a hydrogen-bond to OG Thr 155 was predicted. However, the two oxygens of a carboxylate group could not hydrogen-bond to OG Thr 155 at once. Therefore, a ligand point was defined at the energy minimum where the shortest and strongest hydrogen-bond was made to OG Thr 155. Then program GRID was run again with a carboxyl oxygen probe over this region of the gully, but this time, OG Thr 155 was not permitted to donate a hydrogen-bond to the probe. This resulted in the predicted binding site shown in Figure 5.8.h where the carboxyl oxygen could accept a strong hydrogen-bond from N Asn 133. This energy minimum was 3Å from the other ligand point in this gully site. The separation of the oxygens in a carboxylate group is about 2.2Å so the ligand carboxylate oxygens cannot both be positioned so they make the most favourable interactions with the probe simultaneously. However, they can be positioned to interact favourably compared to other probes including water (for which E_{pred}=-9.6 kcal/mol in this part of the gully). An amide nitrogen was calculated to have a binding energy of about -6

kcal/mol in this region indicating that it would interact far less favourably at this gully site than a carboxyl oxygen. Therefore, it was not appropriate to include an amide group in the ligand at this position.

The principal results of probing the host cell receptor site with program GRID can be summarized by the seven ligand points identified (see Figure 5.8.i). These were used to design a ligand which would bind strongly to the host cell receptor site.

5.5 THE CONSTRUCTION OF LIGANDS

5.5.1 Method

5.5.1.1 Model-building

Ligands, which incorporated the appropriate probe groups positioned at or near the seven ligand points, were model-built using graphics programs INSIGHT [265] and FRODO [266]. The probe groups were connected by a carbon skeleton which was designed so as not to penetrate the receptor surface shown in Figure 5.7.

Model-building was inevitably a rather subjective procedure and it is likely that many possible molecules that could satisfy the ligand points were overlooked. Automated methods of constructing a molecule to satisfy a set of ligand points are under development. One approach [42] is to search databases of known compounds for those which contain the required chemical groups in a particular geometry. Another [267] is the use of expert systems to generate a planar template of aromatic rings and to adjust its size, shape and ring

substituents to satisfy the ligand points. Methods are also available which would enable ligands to be hand-drawn to incorporate the ligand points and then to be adjusted by computer in order to be chemically correct [268]. Unfortunately, none of these methods was available for use in this work.

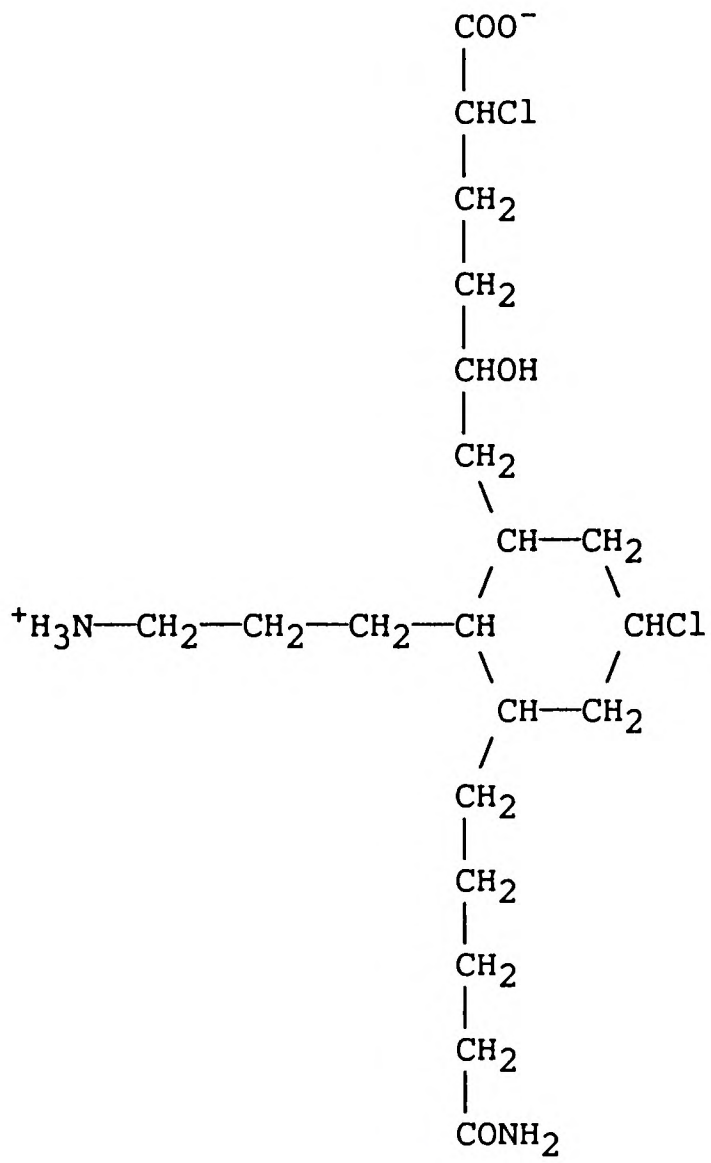
5.5.1.2 Energy minimization

After model-building, the ligands were energy minimized in vacuo using AMPAC [69] (see Appendix II) in order to relieve conformational strain. The energy minimization was stopped when the structure of the ligand was approximately constant, (this was after about twenty cycles of energy minimization).

5.5.1.3 Optimisation of the displacement of water on ligand binding

In order to assess how well the ligand filled the HA binding pocket, program GRID was run with a water probe and a target consisting of HA and the docked ligand. Attractive energy contours between the ligand and the protein indicated regions where water could remain undisplaced in the receptor site on ligand binding. It is preferable that all water is displaced from the binding site by the ligand because undisplaced water may shield the interaction of the ligand with the protein and because the release of bound water makes a favourable entropic contribution to the free energy of ligand binding. The ligands were therefore designed so that the amount of water displaced from the receptor site would be maximized.

a.



b.

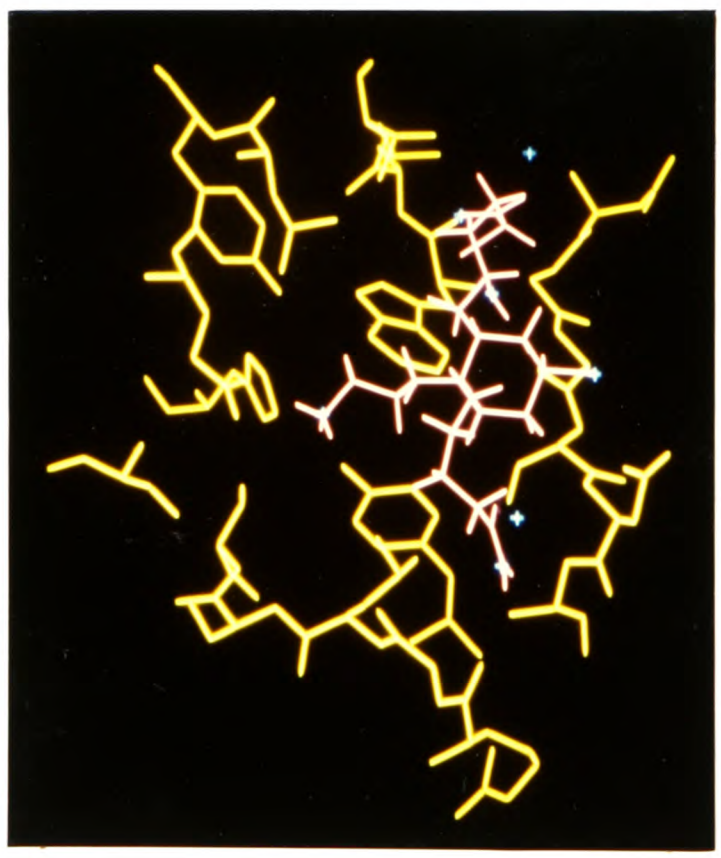


Figure 5.9 a. The structure of the ligand D6 designed to bind to the haemagglutinin host cell receptor site. It contains a cyclohexane ring. b. D6 (shown in pink) docked in the HA receptor site. The ligand points are shown in blue. See Section 5.5.2.

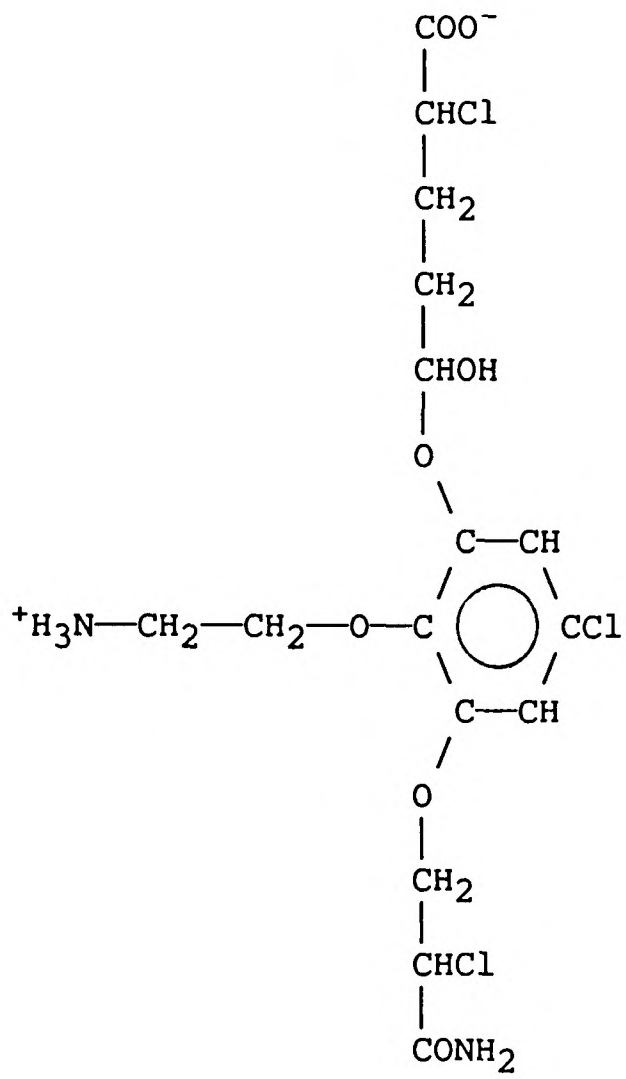
Ligands could be designed to bind to the protein via water bridges by including waters that were predicted to bind strongly to the protein in the target. However, at all the binding sites for water identified by program GRID, other probes were found which could bind more strongly than water and could therefore displace water from the binding site. In addition, the three water molecules observed by crystallography are all in positions occupied by bound sialic acid [254]. This suggests that there are no water molecules in the host cell receptor site which can be considered as an integral part of the target and that a strongly binding ligand should displace all water from the host cell receptor site.

5.5.2 The designed ligands

Sialic acid contains a six-membered sugar ring. Therefore, the approach initially adopted was to append the predicted binding groups to a cyclohexane ring so that the ligand points were occupied by the appropriate probe groups. In Figure 5.9, the structure of the designed ligand, D6, is shown docked in the HA host cell receptor site in the conformation achieved after the reduction of conformational strain by AMPAC.

A chlorine atom was incorporated in this compound in a position where it would bind between gully sites I and II. An energy minimum E_{pred}=-6.86 kcal/mol was predicted by program GRID for a chlorine probe at this site. Energy maps for the other probes did not have minima at this position in the gully. However, the carboxyl oxygen, hydroxyl oxygen and water probes were predicted to bind more strongly than chlorine at this position and so a ligand point was not

a.



b.

Figure 5.10 a. The structure of the ligand D10 designed to bind to the haemagglutinin host cell receptor site. It contains an aromatic ring. b. D10 (shown in pink) docked in the HA receptor site. The ligand points are shown in blue. See Section 5.5.2.

assigned to it. Nonetheless, a chlorine atom at this position may reduce the rate of ß-oxidation of the terminal carboxylate group of the ligand.

A ligand molecule such as D6 containing a cyclohexane ring may be difficult to synthesise because it has optical isomers, only one of which was designed to bind to the receptor site. Therefore, subsequent designed ligands were based on a benzene ring. Another advantage of incorporating a benzene ring is that the ligands would be more amenable to QSAR analysis because the parameters of molecules containing benzene rings generally obey additivity rules [34].

Some of the side chain groups of D6 are also optically active. In order to remove the optical centre at gully site II, a carbonyl oxygen could be substituted instead of the hydroxyl oxygen. However, this would result in the predicted loss of about 2 kcal/mol of binding energy because a carbonyl oxygen could not donate a hydrogen-bond to O Gly 135.

Further ligands were designed containing an aromatic ring. The most satisfactory ligand, D10, which was found to be the most energetically favourable by AMPAC, is shown in Figure 5.10. In this ligand, ether oxygens were positioned immediately adjacent to the aromatic ring in order to release the conformational strain present when carbon atoms are located at these positions and because they might facilitate the synthesis of the ligand, (by allowing the linking of two molecules by a condensation reaction).

Program GRID was run for a water probe with a target of HA with D10 docked in the binding site. It showed that a water molecule could not be fitted between the protein and the ligand suggesting that the shape of D10 was complementary to that of the binding site.

A nitrogen ring substitution could be made at the carbon nearest to the side valley as program GRID predicted a binding energy which was about 0.4 kcal/mol more favourable for an aromatic nitrogen than for an aromatic carbon at this position.

5.5.3 Testing the binding of the designed ligands

Having designed a number of ligands it was appropriate to test experimentally whether they actually bound to the HA host cell receptor site. No facilities were available for the synthesis of the designed compounds, and so a search of chemical databases and Chemical Abstracts was made to see if these compounds or similar ones had been synthesised before. None of the designed ligands were found in this search. In addition, no synthesised compounds incorporating a limited but sufficient number of binding groups were identified. Consequently, it was not possible to test whether the predicted ligand binding actually occurred, and a different approach to designing ligands had to be adopted which would solve the problem of the synthesis of the ligands at the outset. This is described in the next section.

5.6 THE DESIGN OF PEPTIDE INHIBITORS

5.6.1 Introduction

Peptides can be synthesised using standard automated techniques [269,270]. Therefore, by designing peptides to bind to the

host cell receptor site, it was possible to overcome a major hurdle in the development of anti-influenza agents.

However, there are disadvantages in designing peptides. The possible conformations of a peptide are more restricted than for a completely novel ligand because the peptide link is approximately planar. The type and position of the functional groups that can be incorporated into a natural peptide are limited. Thus, the binding energy of a peptide may be less favourable than for a ligand designed with no such restrictions. In addition, peptides are readily metabolised [271] and so modifications to natural amino-acid residues may be required to reduce the rate at which they are broken down.

5.6.2 Method

5.6.2.1 Model-building

Peptides were modelled, using molecular graphics programs INSIGHT [265] and FRODO [266], to incorporate the appropriate probe groups in the energetically favourable positions predicted by program GRID and to sterically fit the receptor site. The structure of the designed ligand D10 was used as a template to assist in assigning the correct shape to the peptides. Peptides were modelled so that their hydrogen-bonding capacity was satisfied, so that the number of hydrophobic surfaces on the peptide facing out to the solvent were minimised and so that hydrophobic regions of the protein were matched by those on the peptide.

5.6.2.2 Energy Minimization

The modelled peptides were energy minimised using program AMBER [73,74] (see Appendix II). The protein was fixed stationary at its crystal structure coordinates and the docked peptide was allowed to relax. A distance dependent dielectric constant was used and solvent was not modelled explicitly.

5.6.2.3 Optimisation of the displacement of water from the binding site by the peptides

A GRID run with a water probe was performed as described in Section 5.5.1.3 for a target of HA with the energy minimised peptide docked in the binding site. The peptide was designed to exclude water from the receptor site by minimizing the volume between the protein and the peptide where water could make attractive interactions.

5.6.2.4 Lipophilicity

The lipophilicity of the designed peptides was assessed quantitatively using ClogP values [251] computed by Dr. K. Kim. In these calculations, the N- and C-terminii of the peptides were assumed to be ionized as would be expected at physiological pH from their pK_a values. If they were unionized, they would be considerably more lipophilic and the ClogP values would be increased.

Table 5.3 The peptides designed to bind to the haemagglutinin host cell receptor site.

Peptide number	Sequence	AMBER intramolecular energy[a] in kcal/mol	AMBER intermolecular energy[b] in kcal/mol	GRID E_{pred} for water[c] in kcal/mol	GRID water binding region	Number of intramolecular hydrogen-bonds[d]	Hydrogen-bond donor	Hydrogen-bond acceptor	ClogP
1	Gly D-Ser Gln D-Ser Leu D-Ser Ser	-62.58	-89.86	-7.5 -7.5	Well Gully Site I and II	6	OG Ser 2 N Gln 3 N Gln 3 N Ser 4 N Leu 5 N Ser 7	O Ser 6 O Gly 1 O Leu 4 O Ser 2 O Gln 3 O Ser 7	-8.84
2	Gly D-Ala D-Gln Gly D-Ser D-Ser Gly	-48.79	-99.27	-4.5	Well	2	N Gly 4 N Ser 5	O Ala 2 O Ala 2	-8.99
3	D-Val D-Ser Gln Pro Ile D-Ser Ala	-32.62	-107.32	-3.0	Gully[e]	2	N Ile 5 OG Ser 6	O Ser 2 O Ile 5	-4.87
4	Gly D-Ser Gln Phe Sta Ala	-33.18	-112.46	-9.5	Well	2	N Phe 4 OB Sta 5	O Ser 2 O Phe 4	-4.86
5	Gly D-Thr D-Gln D-Ser Sta Ser	-40.32	-92.42	-9.0 -5.5	Gully site II Well	2	OG Thr 2 N Ser 4	O Ser 4 O Thr 2	-8.04
6	Ace D-Ser Orn D-Thr Ala	-17.49	-87.83	-5.5 -5.0	Gully site I Well	0	-	-	N/A

[a] This is the energy of the peptide alone in its energy minimized conformation.
[b] This is the interaction energy between the peptide and the protein.
[c] This is the minimum interaction energy of a water probe with a target consisting of the protein with the peptide docked in the binding site.
[d] These are hydrogen-bonds predicted by FRODO [266] between atoms of the peptide.
[e] This water binding region in the gully is continuous with bulk solvent.

5.6.2.5 Metabolism

Peptides are readily degraded by proteases in vivo with a half-life of the order of minutes [271]. Therefore, non-standard and D-amino acids were incorporated in the peptides in order to reduce their rate of metabolism.

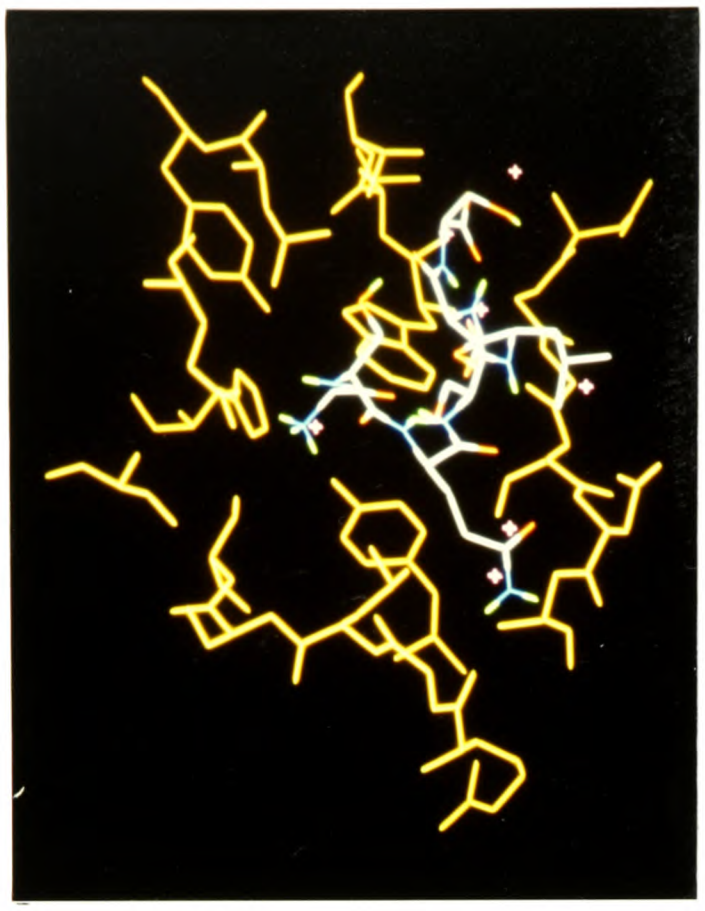
The insertion of a non-standard or D-amino acid at each alternate residue would ensure that the local environment of each peptide link was different from that found naturally and hence, the efficiency of hydrolytic enzymes, which are specific for certain amino acids and which can generally distinguish between optical isomers, might be reduced.

The rate of peptide metabolism could also be reduced by chemical modification of the residues e.g. N-methylation, crosslinking, and cyclisation. Chemical modification might also reduce the flexibility of the peptides. This would result in a decrease in the unfavourable entropic loss that occurs on the binding of short linear peptides which are very flexible in solution but are constrained on binding to a target molecule. However, facilities were not available for chemically modifying peptides experimentally.

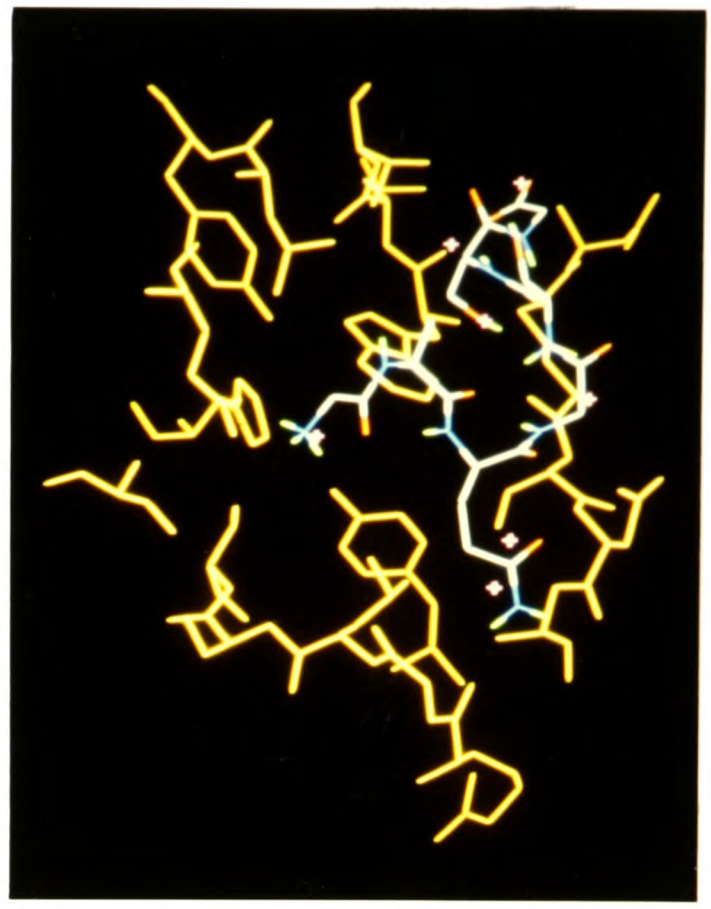
5.6.3 The designed peptides

Of the peptides designed, the six that satisfied the properties required for an effective anti-influenza agent best are listed in Table 5.3.

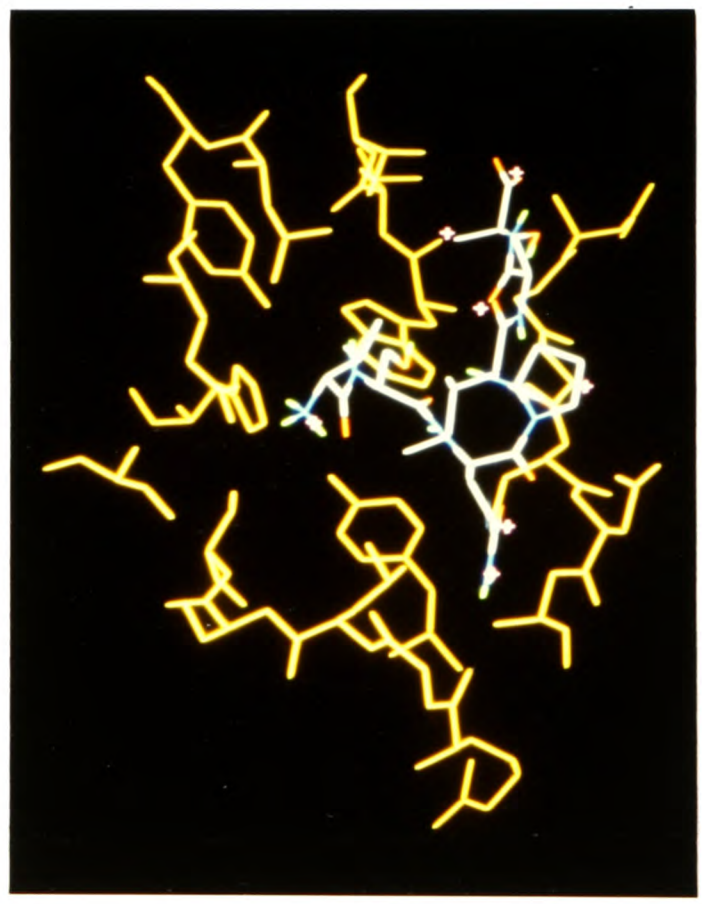
Peptides 1-5 are arranged with their N-terminus in the well and their C-terminus at gully site I and a Gln side-chain positioned



a.

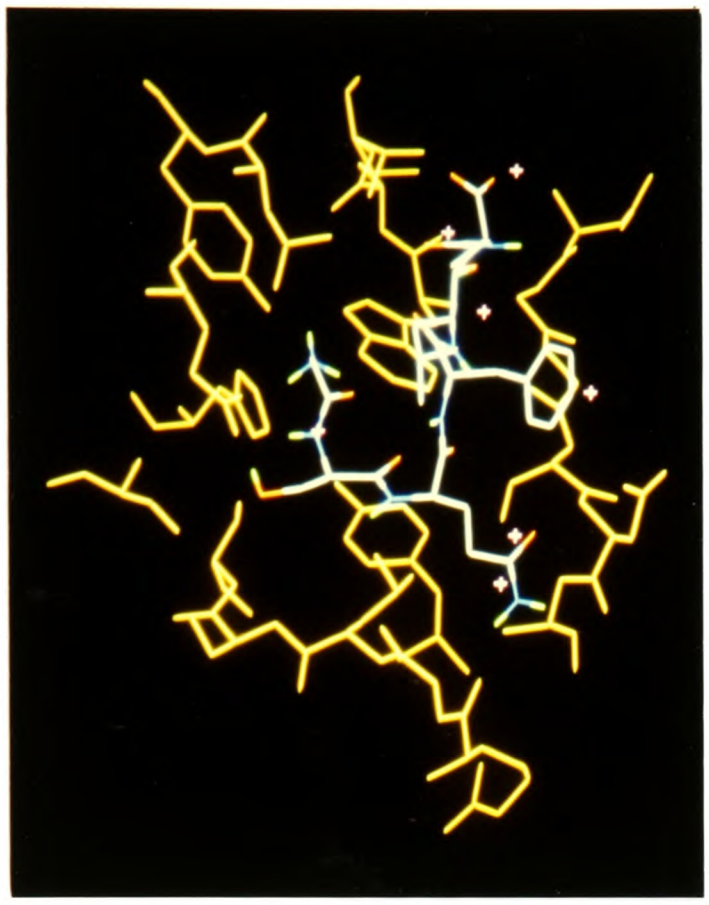


b.

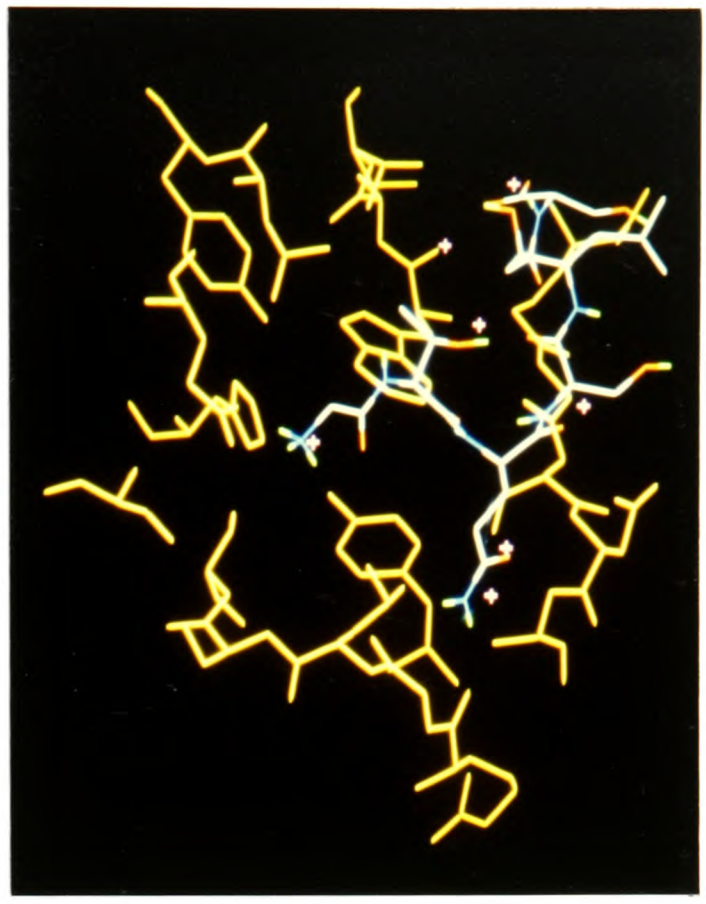


c.

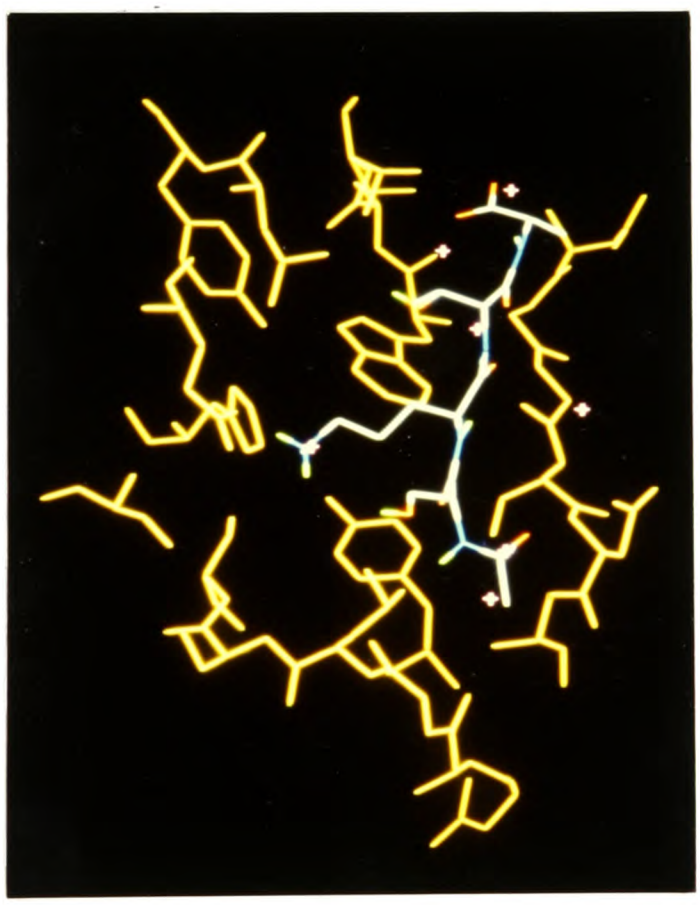
Figure 5.11 The designed peptides docked in the haemagglutinin host cell receptor site in their energy minimized conformations. a-f. Peptides 1-6. The peptides are coloured according to atom type (blue=N, red=O, green=H, white=C). The ligand points are shown in pink. See Section 5.6.3. and Table 5.3.



d.



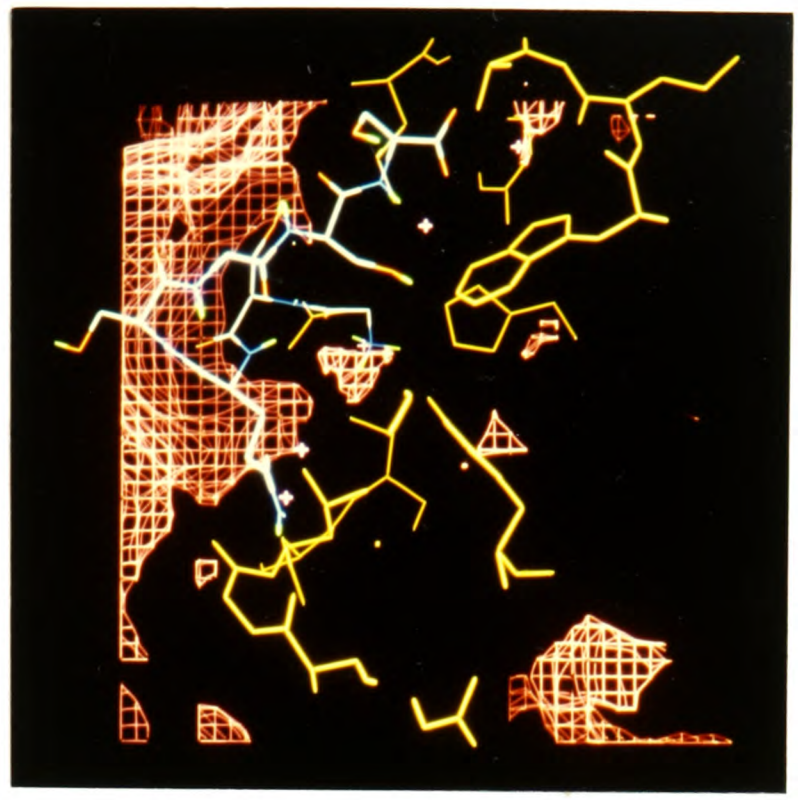
e.



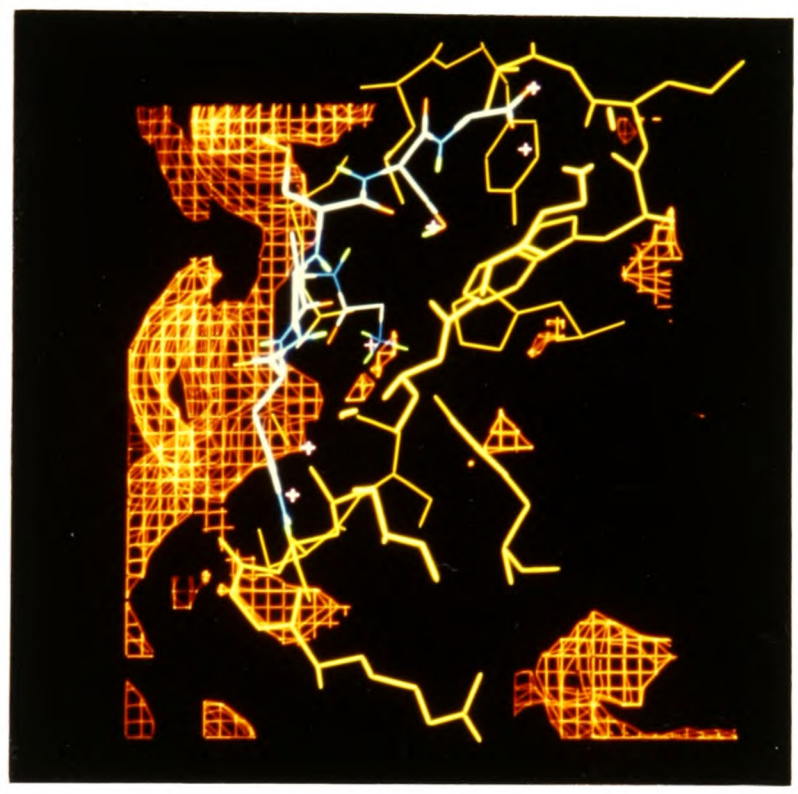
f.

Figure 5.12 The designed peptides docked in the haemagglutinin host cell receptor site in their energy minimized conformations with energy contours (shown in red) at -3 kcal/mol for the interaction of a water probe with a target consisting of HA and the docked peptide. The complexes are seen from a view perpendicular to that in Figure 5.11. The large contoured region corresponds to continuous solvent. The ligand points are shown in pink. a. Peptide 1. The central triangular region between the peptide and the protein corresponds to a region where water can bind favourably at the base of the binding pocket in the well while the peptide is docked. b. Peptide 2. Water can also bind between this peptide and the protein in the well but the contours are smaller showing that the binding of water is less favourable. c. Peptide 3. There are no contours between this peptide and the protein in the well as water is unable to fit favourably between the ligand and HA. This peptide should therefore displace all water from the well region of the receptor binding pocket. See Section 5.6.3. and Table 5.3.

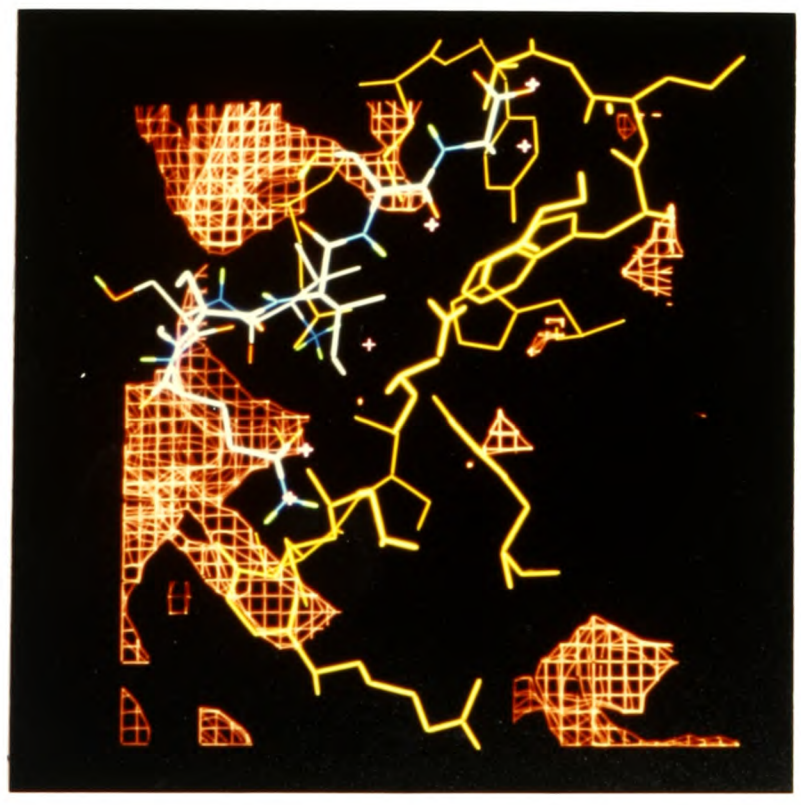
a.



b.



c.



in the side valley. Peptide 6 is linear rather than looped. It has its N-terminus in the side valley, a side chain containing a charged nitrogen in the well, and its C-terminus at gully site I.

The peptides are shown docked in the host cell receptor site in their energy minimized conformations in Figure 5.11. The peptides are now discussed individually.

Peptide 1 Gly D-Ser Gln D-Ser Leu D-Ser Ser

This is a heptapeptide consisting of alternate L- and D-amino acids. The six intramolecular hydrogen-bonds that are predicted to be formed stabilise the peptide in its docked conformation and give rise to the large intramolecular energy calculated for this peptide.

However, there are several disadvantages to this peptide. It does not satisfy all the ligand points. One of the ligand points at gully site I is not occupied because the peptide does not extend far enough up the gully. In addition, OG Ser 6 is not very close to the ligand point at gully site II and can only hydrogen-bond to O Gly 135.

Two polar peptide groups facing the protein do not have their hydrogen-bonding capacity satisfied. These are N Ser 2 and OG Ser 2 which makes only one hydrogen-bond.

The ClogP value is rather negative showing that this is a very polar molecule.

Program GRID shows that water can bind favourably between the docked peptide and the protein at the base of the well (see Figure 5.12.a) and in the gully.

Peptide 2 Gly D-Ala D-Gln Gly D-Ser D-Ser Gly

All ligand points except the one at the platform are satisfied by this heptapeptide so its intermolecular binding energy is of greater magnitude than that of Peptide 1. This peptide fills the binding site better than Peptide 1 and this is shown by the GRID contours for a water probe in Figure 5.12.b. The space a water molecule can favourably occupy in the well when Peptide 2 is docked is smaller because the substitution of D-Gln instead of L-Gln at residue 3 allows Peptide 2 to penetrate deeper into the binding pocket than Peptide 1.

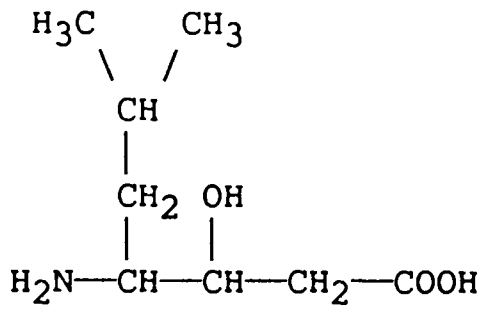
However, there are only two predicted intramolecular hydrogen-bonds, so the magnitude of the intramolecular energy of this peptide is smaller than that for Peptide 1. There are also two unsatisfied hydrogen-bonding groups, (N Ala 2 and O Gly 1), facing the protein.

The ClogP value of this peptide shows that it is also very polar.

Peptide 3 D-Val D-Ser Gln Pro Ile D-Ser Ala

This heptapeptide contains a cis-proline residue. The cis-proline enables the peptide to have an acute turn in it which is necessary in order for the peptide to fit tightly and compactly into the small receptor site displacing all the solvent. On average, only about one in ten proline residues observed in proteins are in the cis rather than the trans conformation [272]. However, cis-prolines occur quite often in proteins at the third position of sharp (type IV) turns reversing the direction of the polypeptide chain. The incorporation of a proline residue is also advantageous because proline occupies

a.



b.

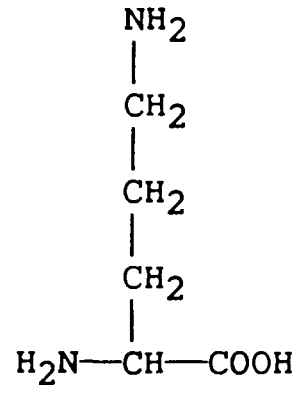


Figure 5.13 The formulae of the non-standard residues incorporated in the peptides designed to bind the haemagglutinin host cell receptor site. a. Statine b. Ornithine. See Section 5.6.3.

less conformational space than any other natural amino acid and therefore the conformation adopted when docked will carry less entropic penalty than for other amino acids.

The GRID water contours (see Figure 5.12.c) show that water, isolated from the bulk solvent, cannot bind favourably between Peptide 3 and the protein, and therefore, the incorporation of a proline residue has enabled the peptide to be buried further into the binding site. The design of Peptides 1-3 shows how the prediction of the binding of water with program GRID can be used to increase the complementarity between a designed ligand and a target molecule.

All the ligand points except that at the platform are occupied by Peptide 3 although the N-terminal nitrogen only hydrogen-bonds to Glu 190.

However, the computed intramolecular energy of the peptide is rather low in magnitude and there are only two predicted intramolecular hydrogen-bonds. Unfavourable interactions may be made by O Val 1 which cannot make a hydrogen-bond to the protein although it could make one hydrogen-bond to the solvent, and also by the non-polar proline ring which faces out to the solvent.

This peptide is less polar than Peptides 1 and 2 although its negative ClogP value indicates its considerable lack of lipophilicity.

Peptide 4 Gly D-Ser Gln Phe Sta Ala

This peptide contains statine (see Figure 5.13.a) which has two more Cα atoms in its backbone than a standard amino acid so that this hexapeptide has the length of a peptide composed of about

6½ standard residues. This enables it to fit into the host cell receptor site in a slightly less bent conformation than Peptides 1-3.

The Phe 4 side chain extends out to occupy the ligand point on the platform which is not occupied by the other peptides. The N-terminal nitrogen donates a hydrogen-bond only to Glu 190 but OG Ser 2 can also donate a hydrogen-bond this residue. The gully site II ligand point is not occupied and there is no hydrogen-bond to N Asn 133 at gully site I.

This peptide is of similar lipophilicity to Peptide 3.

Water can remain undisplaced in the well when the ligand is docked but not in the gully because the hydrophobic side chain of Sta 5 points towards the protein and fills the gully.

Peptide 5 Gly D-Thr D-Gln D-Ser Sta Ser

This peptide also contains statine. It has a D-Gln residue to enable it to fill the well region of the binding site better than Peptide 4. All ligand points except those at gully site II and the platform are occupied.

However, water can remain undisplaced in the well although with a less favourable binding energy than for Peptide 4. Water can also bind at gully site II because the hydrophobic side chain of statine does not fill the gully and instead points unfavourably into the solvent.

This is a much more polar peptide than Peptide 4 as indicated by the ClogP value. This is principally due to the substitution at residues 4 and 6 of non-polar residues with polar residues.

Peptide 6 Ace D-Ser Orn D-Thr Ala

This linear acylated tetrapeptide contains an ornithine residue which has an alkylamine side chain which is one methylene group shorter than in lysine (see Figure 5.13.b). Ornithine was incorporated in this peptide rather than lysine in order to fit the peptide backbone into the gully and the side valley at the same time as the side chain nitrogen occupied the ligand point in the well.

The ligand points are not so well satisfied in this peptide as in the other peptides. At gully site I, one ligand point is not occupied and so no hydrogen-bond is made to N Asn 133. There is no hydroxyl group at the ligand point at Gully site II. However the main chain atoms N Thr 4 and O Thr 4 may form weak hydrogen-bonds to O Gly 135 and N Gly 135 respectively. One ligand point in the side valley is occupied by a methyl group so that no hydrogen-bond is donated to O Arg 224. The ornithine side chain does not occupy all the volume of the well region and its nitrogen donates a hydrogen-bond to Glu 190 but not to Tyr 98. Program GRID shows that water can bind between the peptide and the protein in the well and at gully site I.

Due to its linear conformation, there are no intramolecular hydrogen-bonds and so the structure of this peptide is not well stabilised and has an intramolecular energy of small magnitude. Indeed, this peptide is also likely to be entropically unfavourable compared to Peptides 1-5 because, for flexible chains, linear conformations are less probable than looped ones [273].

5.7 THE SYNTHESIS AND ASSAY OF THE DESIGNED PEPTIDES

The work described in this section was carried out in collaboration with Keith Gould.

5.7.1 The synthesis of the designed peptides

5.7.1.1 Introduction

The synthesis of the peptides was carried out using the Merrifield solid phase peptide synthesis technique [269,270]. In this method, the C-terminal amino acid is attached by a strong covalent linker to an insoluble polymer resin. The peptide is built from the C- to the N-terminus by the sequential addition of protected amino acids to the N-terminus of the peptide attached to the resin. The complete peptide is then cleaved from the resin and all its side chains are deprotected.

The use of an insoluble resin allows the peptide to be filtered and washed free of reagents after the addition of each residue and thus the peptide can be purified by washing away impurities rather than by the much slower process of recrystallization.

The N-terminal ends of the amino acids are protected by an N^{α}-tert-butyloxycarbonyl (t-BOC) group. This is removed from the N-terminus of the peptide attached to the resin just before the addition of the next amino acid. The latter is activated using the coupling reagent N,N'-dicyclohexylcarbodiimide (DCC) for all amino acids except Gln, Asn and Arg for which 1-hydroxybenzotriazole (HOBT) is used.

Double couplings are performed for these residues because their HOBT esters are less reactive than the anhydrides of the other amino acids.

Some amino-acid side chains are protected to prevent them from participating in the coupling reaction. In the peptides synthesised, Ser and Thr were protected by O-benzyl groups.

The repetitive and cyclic nature of this method of synthesis allows it to be automated.

5.7.1.2 **Method**

Peptides 1, 2 and 3 were synthesised on an Applied Biosystems Model 430A Peptide Synthesizer using the standard protocol [274].

All materials were purchased from Applied Biosystems except the t-BOC D-amino acid derivatives which were purchased from Bachem (UK) Ltd. The C-terminal amino acid was supplied linked to a resin of copoly (styrene 1% - divinyl benzene) by a phenylacetoamidomethyl (OCH_2 PAM) linker.

For each peptide, 0.5mmoles of the protected C-terminal resin-linked amino acid derivative was used. Each coupling was performed with the addition of 2.0mmoles (i.e. excess) of the protected amino acid derivative.

The completeness of the coupling reactions was monitored by the ninhydrin assay [275]. Yields were about 99.5% at each coupling giving a total peptide yield of about 97-98%.

The peptides were cleaved from the resin and the protecting groups were removed with trifluoromethanesulphonic acid following the standard protocol [274].

The peptides were then desalted using a P4 Biogel (200-400 mesh) gel filtration column run in 1% acetic acid solution in water. Elution of the peptide was monitored at 228nm.

Amino acid analysis was performed on aliquots with an LKB Amino-Acid Analyzer following hydrolysis according to the standard protocol [274].

Peptides 1 and 2 were dissolved in 0.1% trifluoroacetic acid and Peptide 3 was dissolved in 0.1M ammonium bicarbonate.

5.7.2 The assay of the designed peptides

5.7.2.1 Introduction

The peptides were tested by means of a rapid neutralization test developed by Harmon et al [276]. This test was chosen because it provides a sensitive and quantitative test of viral inhibition. Serial dilutions of the peptide are made and a constant amount of virus added and these are incubated together before adding cells. The virus is allowed to infect the cells and multiply in them before the amount of virus present is measured using an enzyme-linked immunoabsorbance (ELISA) assay. This test measures inhibition of the whole viral infection and replication process. Its sensitivity is achieved by allowing the virus to multiply before measuring the extent of inhibition of virus growth.

5.7.2.2 **Method**

The rapid neutralising antibody assay was performed, with modifications to the protocol [276], as follows.

1. 100µL of 20mM peptide in Dulbecco minimal essential medium was dispensed in 10 2-fold dilutions of 50µl each in a 96-well tissue culture plate. Four control wells of 100µl medium and four of 50µl medium were prepared.

2. 50µl of Dulbecco minimal essential medium containing 10 HAU of A/NT/60/68 (H3N2) influenza virus were added to each well.

3. The virus and peptide were incubated at 37°C for an hour.

4. 100µl of a suspension of trypsinized Madine-Derby Canine Kidney (MDCK) cells in Dulbecco minimal essential medium containing 4% fetal calf's serum (FCS) at a concentration of 3×10^6 cells per ml was added to each well, and the plate was incubated at 37°C for 16 hours in a 5% CO_2 atmosphere.

5. Medium was removed and the cells were fixed with 100µl 80% acetone in phosphate buffered saline (PBS) and held at 4°C for 15 minutes.

6. The acetone solution was aspirated and the plate allowed to air dry at room temperature.

7. The wells were washed three times with 100µl PBS + 10% FCS leaving them to soak 5 minutes at room temperature each time and then removing the PBS with an aspirator.

8. 50µl of a thousand-fold dilution in PBS +10% FCS of monoclonal antibody specific for influenza A nucleoprotein was added to each well, and the plate incubated for an hour at 4°C and washed as before.

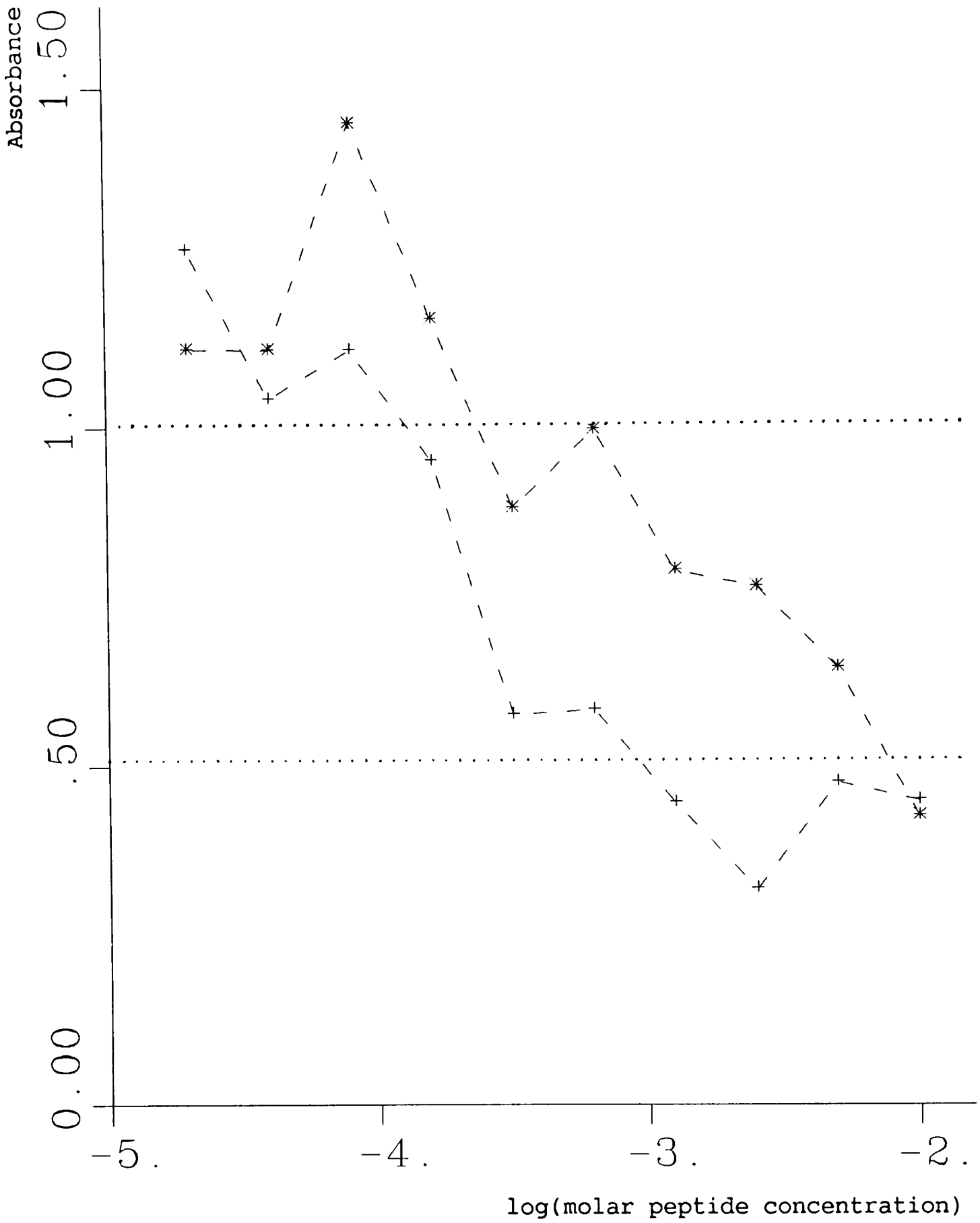


Figure 5.14 The results of the neutralization test used to assay the peptides designed to bind to the haemagglutinin host cell receptor site. Absorbance is plotted against log(peptide concentration) for Peptides 1 and 2. The upper horizontal line is at three standard deviations below the mean absorbance value (1.360 s.d.0.119) for the virus control wells. Absorbance values below this level may be considered to indicate positive inhibition of virus growth [276]. The lower horizontal line is at the mean absorbance value for the control wells with no virus present (0.507 s.d.0.059). The peptides can be seen to have an inhibitory effect which is dependent on the amount of peptide present. Peptide 2 shows greater inhibition than Peptide 1. See Section 5.7.2.3.

9. 50µl of a 500-fold dilution in PBS +10% FCS of antimouse immunoglobulin-peroxidase conjugate was added to each well, incubated for an hour at 4°C and washed as before.

10. Substrate was prepared by adding 10ml citrate pH4 solution to 20mg 2,2'-azinobis (3-ethylbenz-thiazoline sulphonic acid) diammonium salt and then adding 2µl hydrogen peroxide immediately before use.

11. 100µl of substrate was added to each well and allowed to react turning green if peroxidase was present, indicating viral nucleoprotein.

12. After letting the reaction proceed for 1½ hours at room temperature, 50µl of the contents of each well were transferred to a second 96-well plate. The absorbance of the wells in the second plate was measured using an automated microplate reader (Dynatech MR600) at a wavelength of 410nm.

13. Wells with absorbance readings greater than three standard deviations below the mean absorbance value for the virus control wells were considered positive for the inhibition of virus growth.

5.7.2.3 Results

The results are shown in Figure 5.14 for Peptides 1 and 2.

Positive inhibition was obtained with concentrations greater than 0.24mM of Peptide 1 and concentrations greater than 0.13mM of Peptide 2. The inhibitory effect, although weak, increases with increasing peptide concentration.

5.7.2.4 Discussion

Peptide 2 showed greater inhibition than Peptide 1 and this was in agreement with the predictions which suggested that Peptide 2 would bind more strongly to the virus than Peptide 1.

Inhibition by these peptides in solution was observed only at rather high concentrations. One possible reason for this may be that sialic acid is thought to bind the virus cooperatively because of the large number of HA molecules on the surface of the virus interacting with the cell surface [277]. Indeed, it has been suggested [254] that it may be necessary to present a designed ligand either as a polymer or on a microsurface in order that HA has a higher binding affinity for the ligand than for sialic acid.

These are preliminary results and further tests need to be performed. In particular, the following points should be addressed:

1. Alteration of the length of time for which the virus is incubated with peptide may affect the extent of inhibition observed.
2. These peptides have been designed to bind to a conserved region of the virus and therefore their inhibitory activity should be tested on a variety of strains of influenza virus.
3. These peptides may inhibit influenza, not by binding to the HA host cell receptor site, but by hindering some other stage in the viral lifecycle. Therefore, it is necessary to perform experiments which measure solely the attachment of the virus to the cells such as the haemagglutination inhibition test [278] or radio-labelled ligand binding studies with isolated HA. If the peptides were to be shown to bind to HA, they could then be investigated at an atomic level by means of nmr and x-ray crystallography.

The inhibitory activity observed in the neutralization test suggests that further investigation of these peptides is merited.

5.8 CONCLUSION

There is a need for a specific drug which is effective against influenza despite the mutation of the virus which causes periodic epidemics of the disease.

In this thesis, the method of receptor fit has been used to address this problem. This method was suitable because the three-dimensional structure of the haemagglutinin, one of the viral glycoproteins, has been determined. This protein is responsible for the attachment of the virus to the cells on initial infection and therefore ligands were designed to block the HA host cell receptor site. As the residues in this site are highly conserved, the designed ligands should be able to bind to the same site on HA of all strains of influenza A and should thus be effective when new strains of the virus arise through antigenic variation.

In order to design strongly binding ligands, the host cell receptor site was mapped energetically with program GRID and favourable binding sites for particular chemical groups were identified. Ligands were then constructed which incorporated these groups and which were designed so as to have appropriate properties for action as therapeutic agents. The binding of some of the designed peptide ligands was investigated experimentally and preliminary results suggest that they may show inhibitory activity against the influenza virus.

CHAPTER 6

CONCLUSIONS

This thesis has been concerned with the extension, testing and application of the GRID method of predicting favourable ligand binding sites. In this method, the interaction energy between a small fragment of a ligand, (the probe), and a target molecule of known structure is calculated using an empirical energy potential. Energetically favourable ligand binding sites are then identified and displayed using computer graphics. These sites may be utilized in the design of therapeutic agents.

Hydrogen-bonds make an important contribution to the specificity of intermolecular interactions. In this work, the GRID potential function has been revised to enable it to reproduce experimental observations of the geometry of hydrogen-bonds. The variation with length and direction of the strength of hydrogen-bonds is dependent on the chemical nature of the participating atoms and so different hydrogen-bond potential energy functions have been devised for individual atom types. The geometry of hydrogen-bonds to atoms in the target is modelled differently to that of atoms of the same chemical type in the probe because of the different nature of the surrounding environment. A probe may make multiple hydrogen-bonds to the target while each target atom may only make one hydrogen-bond to an 'extended atom' probe. In addition, the positions and chemical nature of the adjacent atoms are available for target atoms but not for the probe. These neighbouring atoms could be considered by extending the size of the probe to include several atoms rather than

one extended atom e.g. the carboxylic group could be modelled by a three atom probe rather than one isolated carboxyl oxygen atom.

The GRID program takes some account of the reorientation of atoms of the target by the incoming ligand and of the ligand by the target. On the target, hydrogen atoms and lone pair orbitals are allowed to move to the position where they can make the most favourable energetic interaction with the probe, e.g. by rotating around the C-O axis in hydroxyl groups or by adopting a different tautomeric conformation as in histidine and tyrosine. The probe is oriented so as to optimise the total interaction energy. For probes with more than one possible preferred hydrogen-bond geometry, the most energetically favourable one is selected according to the nature of the target. The effect of these hydrogen-bond functions in determining ligand binding sites has been demonstrated on molecules of pharmacological interest.

The GRID method has been tested by predicting the experimentally observed water structures of a variety of crystals of biological molecules. It has been shown that the GRID method may be used to determine the positions of well-ordered waters binding to target molecules and that it may also be of use in aiding the assignment of water sites during crystallographic refinement.

The GRID method has also been used to study the role of water in ligand binding. In the region of the active site of cytochrome P450-cam, GRID maps indicate that water may act in four different ways:

1. by stabilizing the protein residues which contribute to the structure of the active site.

2. by interacting weakly with the protein and being displaced

by bound substrate, thus providing a favourable entropic contribution to the free energy of substrate binding.

3. by filling space in the active site which is unoccupied by the substrate and in which it can form hydrogen-bonds to the protein.

4. by filling vacant space in a hydrophobic region of the active site where it is stabilised by the formation of a hydrogen-bond to the substrate. This hydrogen-bond may assist in orienting the substrate correctly for the catalytic reaction.

Water may be able to perform similar functions in other ligand-macromolecule interactions.

The GRID method has been applied to the design of anti-influenza agents. In order that these agents might be effective against all strains of influenza A, they were designed to bind to one of the highly conserved regions of the virus: the host cell receptor site on the haemagglutinin glycoprotein. By binding at this site, the designed compounds should prevent the attachment of the virus to the host cells and should therefore prevent viral infection. Program GRID was used to map the host cell receptor site energetically, thus enabling strong ligand binding sites to be identified. Ligands were then designed to exploit these sites using molecular graphics, molecular mechanics and semi-empirical molecular orbital techniques. Some of these ligands were peptides and these were synthesised and assayed for anti-influenza activity. Preliminary results suggest that they may inhibit the influenza virus.

The GRID method should find increasing application in the design of therapeutic agents as the number of experimentally determined structures of target molecules of pharmacological importance multiplies.

APPENDIX I

THE DERIVATION OF THE THERMAL MOTION OF THE WATER PROBE FROM THE INTERACTION ENERGIES COMPUTED BY PROGRAM GRID

The isotropic temperature factor or B-value, B, of an atom is given by

$$B = (8\pi^2/3)\langle u^2 \rangle$$

where $\langle u^2 \rangle$ is the mean square displacement of the atom from its mean rest position. (The temperature factor can also be defined as $B = 8\pi^2 U$ where U is the mean square displacement along the orthogonal principal axes of the thermal ellipsoid of the atom.) These expressions are obtained by assuming that the atom is moving in a spherically symmetric harmonic energy well [216].

Assuming that program GRID predicts the 'extended atom ' water probe to lie in a harmonic well, its predicted energy, E, can be expressed as a quadratic function of its Cartesian coordinates:

$$E = ax^2 + by^2 + cz^2 + dxy + eyz + fxz + jx + ky + lz + g$$

The constants a-g can be fitted to the interaction energies calculated over the region around the predicted energy minimum.

In fact, for predicted waters, the energy contours around an energy minimum are often far from quadratic and so the resultant fitted function is very dependent on the size of the region over which it is fitted. This was chosen by trying to include in the region all points at energies up to kT (0.576 kcal/mol at room temperature) above the predicted energy minimum. If this were to be done accurately, a different size region would be required for each water molecule considered. Therefore, to simplify the calculation, an average size

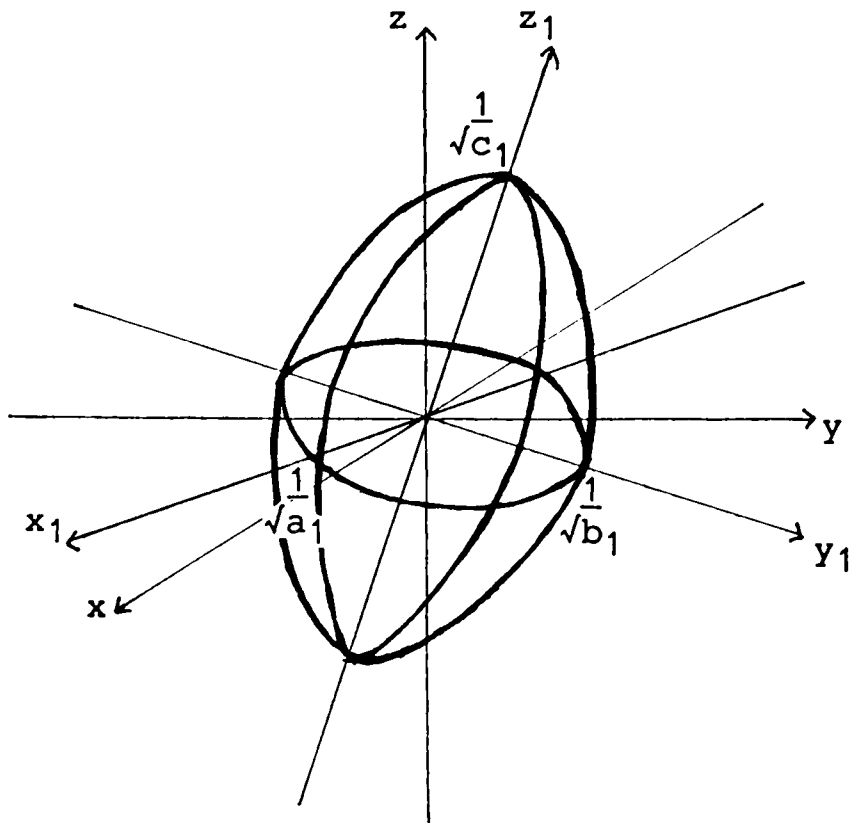


Figure I.1 The ellipsoid of the rate of change of force with distance. See text.

region was found which approximately satisfied this criterion for all the water molecules considered. This region was a cube of side 0.2Å centred on the predicted energy minimum. This required that GRID calculations were performed at points 0.02Å apart over this cube in order to obtain a sufficient number of interaction energies to fit a quadratic function to (see Appendix III.4 for the prediction method).

The restoring force, $\underline{F}$, acting on the water is given by:

$$\underline{F} = -\frac{\delta E}{\delta x}\underline{i} - \frac{\delta E}{\delta y}\underline{j} - \frac{\delta E}{\delta z}\underline{k}$$

$$\underline{F} = -(2ax+dy+fz+j)\underline{i} - (2by+dx+ez+k)\underline{j} - (2cz+ey+fx+1)\underline{k}$$

The rate of change of force with distance is:

$$\frac{d\underline{F}}{du} = -\frac{d^2E}{du^2} = \left[-\frac{\delta^2E}{\delta x^2}, -\frac{\delta^2E}{\delta x\delta y}, \ldots . -\frac{\delta^2E}{\delta z^2}\right]$$

This can be expressed as:

$$[\,x\ y\ z\,]\begin{bmatrix} 2a & d & f \\ d & 2b & e \\ f & e & 2c \end{bmatrix}\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \text{constant}$$

Orthonormalising:

$$[\,x_1\ y_1\ z_1\,]\begin{bmatrix} a_1 & 0 & 0 \\ 0 & b_1 & 0 \\ 0 & 0 & c_1 \end{bmatrix}\begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix} = 1$$

where a_1, b_1, c_1 are eigenvalues of the rate of change of force with distance and x_1, y_1, z_1 are coordinates along the orthogonal axes of the ellipsoid $a_1x_1^2 + b_1y_1^2 + c_1z_1^2=1$ as shown in Figure I.1. A large ellipsoid implies a small rate of change of force with distance and thus a mobile particle with a large B-factor.

Defining space using the ellipsoid axes:

$$\frac{d^2\underline{E}}{du^2} = (a_1, b_1, c_1)$$

$$-\frac{dE}{dx} = F_x = -(a_1x_1 + \alpha_1) = m\ddot{x}_1$$

using Newton's equation of motion with α_1 = constant. This is the equation for simple harmonic motion and has a solution:

$$x_1 = \Gamma\sin\sqrt{(a_1/m)}t + \mu\cos\sqrt{(a_1/m)}t - \alpha_1/a_1$$

where Γ and μ are constants.

Applying the following boundary conditions:

1. E is a minimum at the origin, so dE/du = 0 at $x_1=y_1=z_1=0$

2. Let $x_1=y_1=z_1=0$ at t=0, so $x_1 = \Gamma\sin\sqrt{(a_1/m)}t$

   $\langle x_1^2\rangle = \langle\Gamma^2\rangle/2$ and similarly for y_1 and z_1.

Now, classically for a one-dimensional simple harmonic oscillator:

$$\langle E\rangle = kT \quad (\tfrac{1}{2}kT/\text{degree of freedom})$$

$$\langle E\rangle = 2\langle K.E.\rangle = m\langle\dot{x}_1^2\rangle$$

The velocity $\dot{x}_1 = \Gamma\sqrt{(a_1/m)}\cos\sqrt{(a_1/m)}t$

$$\langle\dot{x}_1^2\rangle = (\langle\Gamma^2\rangle a_1)/2m$$

Therefore, $\langle x_1^2\rangle = \langle\dot{x}_1^2\rangle m/a_1 = kT/a_1$

Similarly, $\langle y_1^2\rangle = kT/b_1$

$$\langle z_1^2\rangle = kT/c_1$$

when the motion is considered as equivalent to three one-dimensional simple harmonic oscillators along the ellipsoid axes. This gives three B-values for motion along the ellipsoid axes which may be averaged to give an isotropic B-value:

$$B=[(8\pi^2kT)/3](1/a_1 + 1/b_1 + 1/c_1)$$

which can be directly compared with the experimentally observed B-value of the same water molecule.

APPENDIX II

COMPUTATIONAL METHODS

II.1 AMPAC

This is a general purpose semi-empirical molecular orbital calculation package [279] in which one of three methods may be used: MINDO/3 [280], MNDO [281] and AM1 [69]. MINDO/3 is the oldest, while AM1 is the most recent and is the method used in this thesis.

II.1.1 Molecular orbital theory

Quantum mechanics is based on the solution of the Schrödinger equation:

$$H\Phi=E\Phi$$

where H is the Hamiltonian operator, Φ is the wavefunction and E is the energy of the system. The probability density given by $\Phi^*\Phi$. Quantum mechanical calculations can determine the energy and the electronic structure of a molecular system, whereas classical mechanics can only be used to evaluate the energy.

The Schrödinger equation can be solved exactly only for a hydrogen atom. For molecules, the Born-Oppenheimer approximation is applied. This states that the nuclei are fixed on the timescale of the movement of the electrons and so the electronic wavefunction can be considered to be independent of nuclear motion. This is a good approximation except for systems with a very flat potential surface

where there is coupling between vibrational and electronic wavefunctions.

For systems containing more than one electron, the orbital approximation is also applied in which each electron is assumed to move independently of the others and to be influenced by the average properties of the other electrons rather than their instantaneous properties. Thus the n-electron wavefunction can be written:

$$\Phi=\phi_1\phi_2..\phi_i..\phi_n$$

where ϕ_i is a one electron spin orbital or molecular orbital which is antisymmetrised so that Pauli's principal is obeyed and which extends over the entire system.

Each molecular orbital (MO) may be approximated as a linear combination of atomic orbitals χ_k (LCAO):

$$\phi_i=\Sigma_k c_{ik}\chi_k$$

A set of χ_k is known as a basis set. c_{ik} are mixing coefficients which are adjusted so that the energy of the molecule is a minimum.

The Schrödinger equation can be solved for molecules with these approximations using ab initio techniques. However, these are computationally expensive and are therefore only applicable to simple molecular systems. For this reason, further semi-empirical approximations are made in order that computations may be performed more quickly and for more complex molecules [282,283].

II.1.2 AM1

There are many levels of approximation in semi-empirical methods. AM1 employs the neglect of diatomic differential overlap (NDDO) approximation which makes the following assumptions.

1. The valence shell approximation is used in which it is assumed that all electrons in fully occupied shells are part of an unpolarizable core of the molecule and do not participate in bonding.

2. The atomic orbitals χ_m form an orthonormal set i.e. $\Sigma_m c_{im} c_{jm} = \delta_{ij}$ and off-diagonal one-electron integrals are neglected.

3. Overlapping charge densities of basis orbitals on different atoms are neglected. i.e. (mn/ls)=0 unless m,n are on the same atom A and l,s are on the same atom B.

AM1 is an improvement on MNDO which overcomes its major weaknesses including its failure to reproduce hydrogen-bonds [69]. This has been done by modifying the core repulsion function in order to lessen the repulsive van der Waals' interaction at short distances.

The terms in the Hamiltonian are determined from experimental data or from semi-empirical expressions which contain numerical parameters which are adjusted to fit experimental data. AM1 produces better results for molecules containing heteroatoms (O,N) than MNDO although the results for these molecules are less accurate than for hydrocarbons.

AM1 can be used to calculate thermodynamic, vibrational and electronic properties including heats of formation, geometries and charges.

In this thesis, it has been used to calculate the charge distribution in a nitrate ion, (see Appendix III.5), and also to reduce energetic strain in the conformations of the ligands designed to bind to the HA host cell receptor site (see Section 5.5.1.2).

II.2 AMBER

AMBER (Assisted Model Building with Energy Refinement) [73,74] is a general purpose molecular mechanics and dynamics program for use in the refinement of the conformations of macromolecules.

It uses a classical analytical potential energy function given by:

$$E_{total} = \sum_{bonds} K_R(R - R_{eq})^2 + \sum_{angles} K_\Theta(\Theta - \Theta_{eq})^2 + \sum_{dihedrals} \tfrac{1}{2}V_n[1 + \cos(n\phi - \gamma)]$$

$$+ \sum_{i<j}\left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} + \frac{q_i q_j}{\varepsilon R_{ij}}\right] + \sum_{H\text{-bonds}}\left[\frac{C_{ij}}{R_{ij}^{12}} - \frac{D_{ij}}{R_{ij}^{10}}\right]$$

The first three terms are components of internal energy describing bond stretching, bond bending and dihedral angle torsion respectively.

Bond stretching is given by Hooke's law where K_R is the bond stretching force constant and R_{eq} is the equilibrium bond length. Bond bending is given by a function of the same form. A harmonic approximation for these two terms is generally adequate as deviations from standard bond lengths and angles are usually small for the systems studied [73].

The torsional term is given by a Fourier series where V_n is the rotational barrier height for a particular set of four atoms, γ is the phase and n is the periodicity i.e. the number of energy minima in a 360° rotation and takes values n=1,2,3,4,6. It is able to give accurate simulations of conformational preferences in simple and complex molecules [73].

The non-bonded interaction (see Section 1.1.1) is given by a Lennard-Jones term, an electrostatic term and a hydrogen-bond term.

In the electrostatic term, a distance dependent dielectric constant $\varepsilon=KR_{ij}$ is used where K is a constant. This form of ε was chosen to mimic the polarization effect in attractive interactions and the damping effect of solvent and because it is computationally convenient. It is of a simpler form than that used in program GRID (see Section 2.2.4.2). Alternatively, solvent may be considered explicitly and a dielectric constant $\varepsilon=1$ used. In this thesis, a distance dependent dielectric constant was used.

The 10-12 hydrogen-bond term is calculated between hydrogen-bonding hydrogens and acceptor atoms and replaces the 6-12 term for these atoms. A standard well-depth of 0.5kcal/mol is used. This term prevents the formation of very short hydrogen-bonds and allows for the matching of hydrogen-bond distances and energies to experimental observations. It generally makes a smaller contribution to the total non-bonded interaction energy than the hydrogen-bond term used in program GRID (see Chapter 3) which may have a well-depth up to 4 kcal/mol in magnitude. It is compared to hydrogen-bond terms used in other force fields in Table 3.1.

AMBER may be used both with a 'united-atom' and an 'all-atom' force field. The 'united atom' force field in which all atoms are included explicitly except hydrogens bonded to carbons was used in this thesis. The lone pairs on sulphur atoms are also included explicitly in order to achieve the directionality suggested by quantum mechanical calculations [73]. The two types of force field produce similar resultant structures but the all-atom force field was introduced to improve simulations of the energetics of nucleic acids [74].

The parameters are derived from experimental results and ab initio molecular orbital and molecular mechanics calculations on small molecules. Partial point charges are fitted to the quantum mechanically calculated electrostatic potential [284].

Energy minimization is performed in cartesian space by steepest descent or conjugate gradient minimization. Molecular dynamics based on solution of Newton's equations at small time intervals along a trajectory may then be performed.

In this thesis, peptides designed to bind to influenza HA were energy minimized by AMBER while docked in the HA host cell receptor site, (see Section 5.6.2.2). Parameter files (.PRP files) were constructed for the unusual amino-acids in these peptides as AMBER was only parameterised for standard residues. These are listed on the appended microfiche.

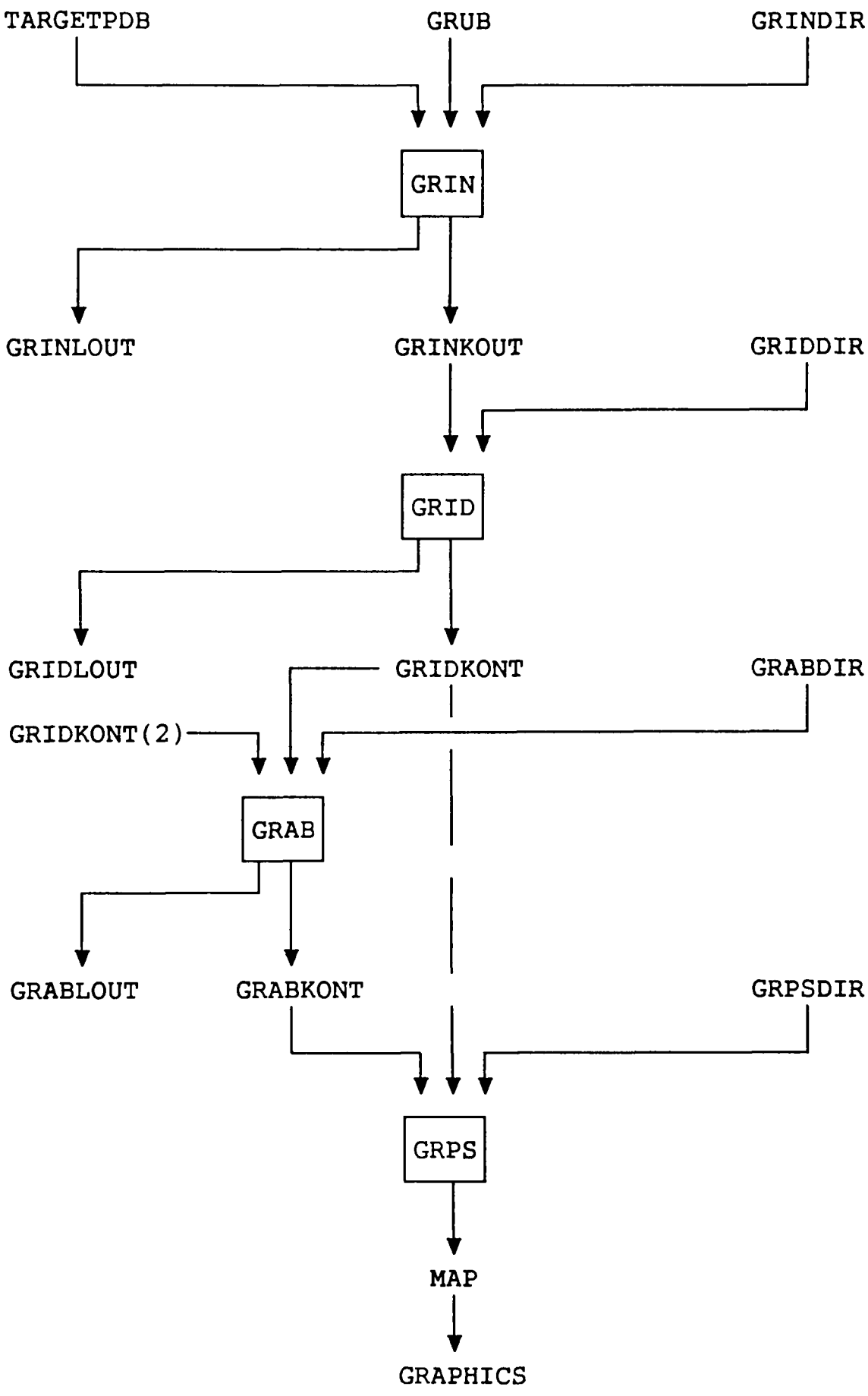


Figure III.1 Flow diagram of the programs used in the GRID method.

The programs are shown in boxes. The contents of the files are as follows:
TARGETPDB contains the coordinates of the target atoms.
GRUB contains atomic parameters.
XDIR files contain directives for program operation.
GRINKOUT contains the target atomic coordinates and parameters.
XLOUT files contain information about the results of the programs.
XKONT files are machine-readable files containing computed energies.
MAP contains computed energies in a format suitable for display on computer graphics.
X = GRIN, GRID, GRAB or GRPS.

APPENDIX III

THE GRID PROGRAM

III.1 IMPLEMENTATION OF THE GRID METHOD

In the GRID method, four programs GRIN, GRID, GRAB, and GRPS, are run sequentially as shown in the flow diagram in Figure III.1.

GRIN prepares coordinate files for the GRID energy calculations by assigning appropriate parameters to all the atoms of the target.

GRID calculates the interaction energies between a probe and the target and identifies the positions of the energy minima.

GRAB computes the difference E_{dif} in the interaction energies (E_1 and E_2) calculated by GRID at each point in an array either for two different probes interacting with the same target or for two different target molecules interacting with the same probe. E_{dif} is computed as follows [87]:

If $E_1<0$, $E_2<0$ $\quad E_{dif} = E_1 - E_2$

If $E_1<0$, $E_2>0$ $\quad E_{dif} = E_1$

If $E_1>0$, $E_2<0$ $\quad E_{dif} = - E_2$

If $E_1>0$, $E_2>0$ $\quad E_{dif} = 0.0$

The first equation cannot be used for all points because it could give rise to large values of E_{dif} in regions of the array which are inaccessible to the probe.

GRPS converts the spatial array of computed energies into a format that can be displayed on computer graphics.

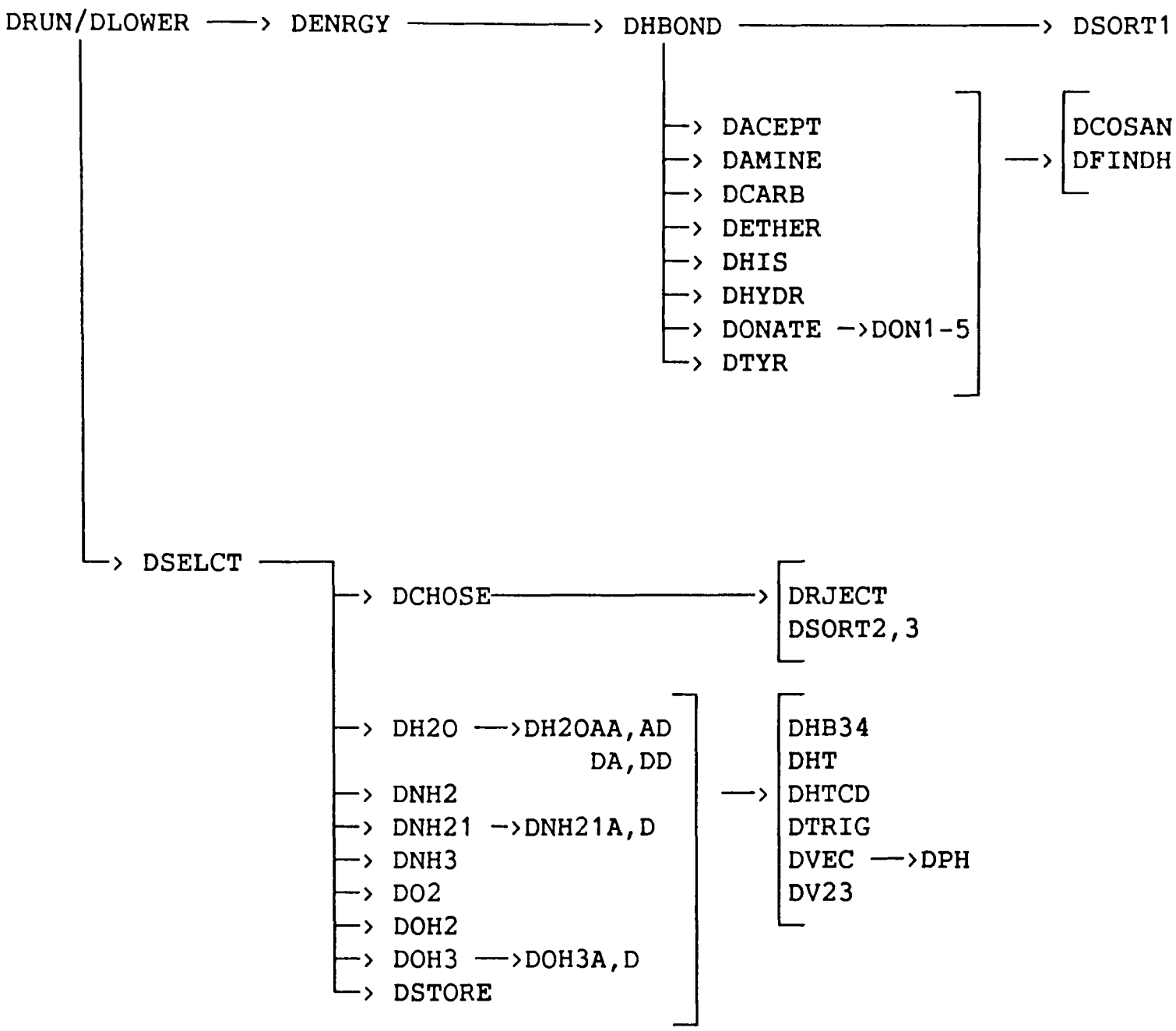


Figure III.2 Flow diagram showing the hydrogen-bonding subroutines in program GRID.

DRUN and DLOWER first call DENRGY for all target atoms and then call DSELCT.

DENRGY calculates E_{el}, E_{lj} and the distance dependent function Er of the hydrogen-bond energy E_{hb} for the interaction of each target atom with the probe.

DHBOND calls subroutines to calculate the dependence Et of hydrogen-bonds on the angle made at the target atom.

DACEPT and DONATE calculate Et for general target hydrogen-bond acceptors and donors respectively.

DAMINE, DCARB, DETHER, DHIS, DHYDR and DTYR calculate Et for the specific target atoms identified in the subroutine name.

DSORT1, DCOSAN and DFINDH perform general geometric and sorting functions.

DSELCT calls subroutines which calculate the dependence Ep of hydrogen-bonds on the angle made at the probe and select the combination of hydrogen-bonds that will give the most favourable binding energy E_{tot}.

DCHOSE is called for probes of undefined hydrogen-bonding geometry.

DH2O, DNH2, DNH21, DNH3, DO2, DOH2 and DOH3 perform calculations for the specific probes identified in the subroutine name.

DRJECT, DSORT2, DSORT3, DHB34, DHT, DHTCD, DTRIG, DVEC, DPH, and DV23 perform general geometric and sorting functions.

DSTORE stores the most favourable hydrogen-bond energies computed.

All calculations were performed on a VAX 11/750. GRID energy maps were displayed using program FRODO [266] on an Evans and Sutherland PS300 colour graphics display system.

III.2 THE HYDROGEN-BONDING SUBROUTINES IN PROGRAM GRID

The use of the subroutines which evaluate E_{hb} in GRID2 is shown in the flow diagram in Figure III.2 along with an outline of their individual functions. These subroutines are listed on the appended microfiche.

Table III.1 The three versions of program GRID used in this thesis.

GRID version	Function Er	Function Et	Treatment of target histidine residues	Function Ep	Selection of hydrogen-bonds made to the probe	Hydrogen-bond energy cutoff
GRID0	$C/r^6 - D/r^4$	$=\cos^4 t$ for all atom types except carbonyl oxygen which has a specific function for Et.	ND always assumed to be protonated.	=1 always.	E_{hb} optimised but target atoms which can donate or accept a hydrogen-bond are always assumed to accept.	None.
GRID1	$C/r^8 - D/r^6$	Varies according to atom type.	Both neutral tautomers of histidine considered.	$=\cos^4 p$ for nitrogen probes which can only donate hydrogen-bonds. =1 for all other probes.	E_{hb} optimised. Target atoms which can donate or accept a hydrogen-bond are considered correctly.	-0.25 kcal/mol except when Ep<1.
GRID2	$C/r^8 - D/r^6$	Varies according to atom type. It takes different functional forms to those in GRID1 and has no discontinuities.	Considered in its low and high pH forms as well as in its two neutral tautomers.	$=\cos^2 p$ for most probes but varies according to the chemical nature of the probe and differs from function Ep used in GRID1.	E_{tot} optimised. Target atoms which can donate or accept a hydrogen-bond are considered correctly.	-0.25 kcal/mol for long hydrogen-bonds

III.3 THE THREE VERSIONS OF PROGRAM GRID

In this thesis, three versions of program GRID have been used. They are numbered 0-2 in chronological order and differ principally in their treatment of hydrogen-bonds as shown in Table III.1.

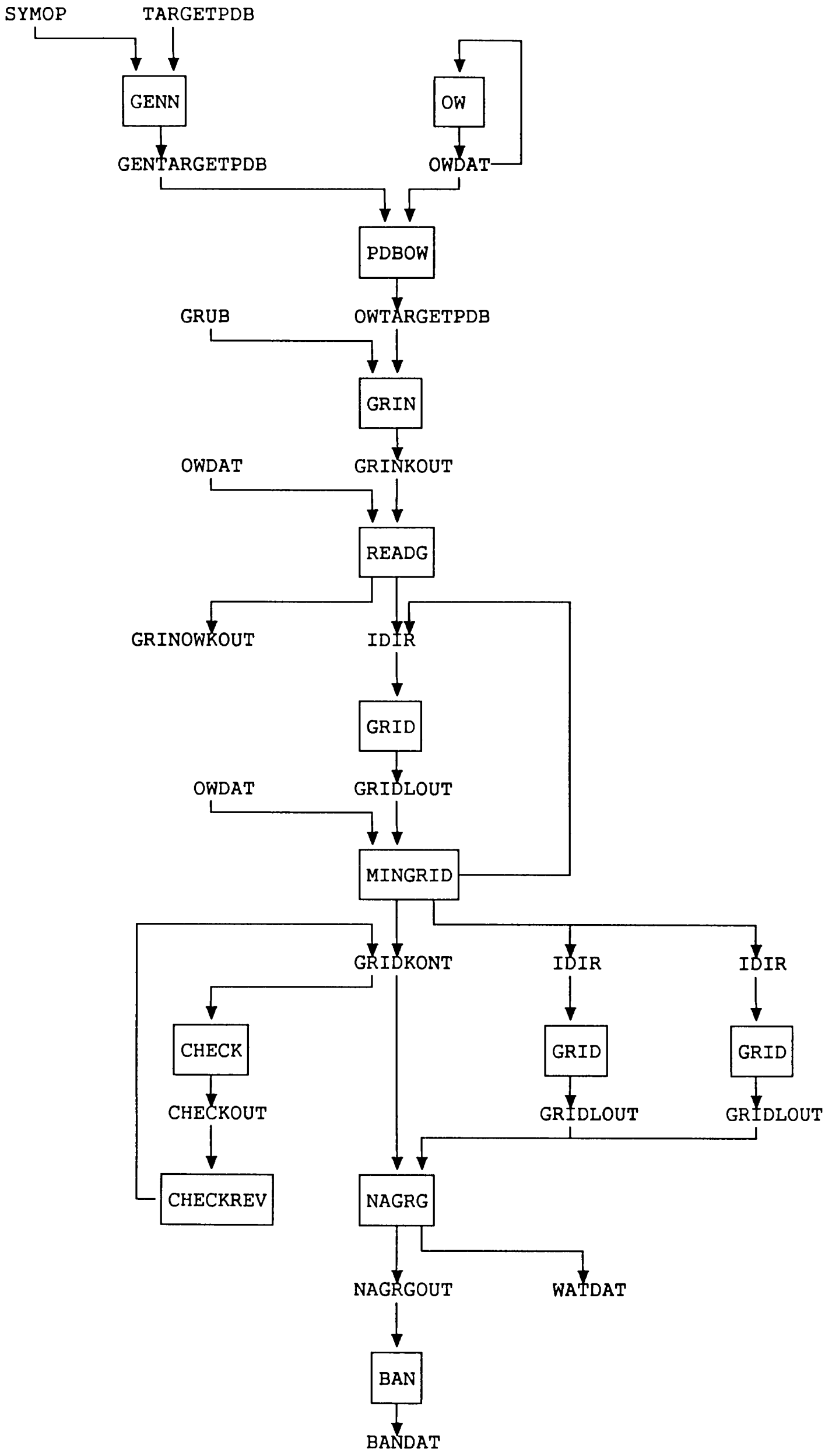


Figure III.3 Flow diagram of the programs used to predict waters with program GRID by Prediction Method 2.

III.4 PROGRAMS FOR THE PREDICTION OF WATER MOLECULES USING THE GRID METHOD AND FOR THE ASSESSMENT OF THE RESULTS

The programs used in Prediction Method 2 are shown in the flow diagram in Figure III.3.

A brief description of the function of each program is given below:

GENN This generates all target atoms in a crystal within a distance of 9Å from any of the water molecules which are to be predicted by program GRID and writes the coordinates to file GENTARGETPDB. The symmetry related target molecules are generated by applying the $P2_12_12_1$ (for L-serine monohydrate and HL) or $P2_1$ (for ß-CD) symmetry operations listed in file SYMOPDAT to the atomic coordinates listed in the TARGETPDB file.

OW OWDAT contains the coordinates of the waters for which predictions are to be made. OW provides the coordinates of the next water to be predicted.

PDBOW This lists to file OWTARGETPDB the coordinates of all atoms up to 9Å from the experimentally observed position of the water to be predicted which is read from the coordinate file OWDAT. It inserts markers to show where fragments of the target molecule begin and end so that GRIN can correctly deduce the connectivity of the atoms and the geometry of the hydrogen-bonding hydrogens.

GRIN See Section III.1.

READG This marks the water molecule to be predicted so that it is not included in the target. It writes directives for the first GRID run to directives file IDIR. These set the spacing of the array points at 0.2Å over a volume of 3.2Å x 3.2Å x 3.2Å centred on the

experimentally observed water coordinates. This volume was chosen because it may be considered to cover the site of one water molecule.

GRID See Section III.1.

MINGRID This reads GRIDLOUT and writes directives file IDIR for the next GRID run according to the following decision tree.

1. If the energy minimum is at the edge of the array used for calculations, IDIR directs program GRID to be rerun with the same array point spacing and the same array volume but centred on the position of the energy minimum. This is because the energy minimum identified may not be a true energy minimum as energies have only been calculated for part of the energy well.

2. If the array spacing is 0.2Å, IDIR directs the next GRID run to be at an array point spacing of 0.05Å on an array of volume 0.5Å x 0.5Å x 0.5Å centred on the energy minimum.

3. If the array spacing is 0.05Å, IDIR directs the next GRID run to be at an array point spacing of 0.02Å on an array of volume 0.12Å x 0.12Å x 0.12Å centred on the energy minimum.

4. If the array spacing is 0.02Å and the energy minimum is not exactly at the centre of the array, IDIR directs the next GRID run to be at an array point spacing of 0.02Å on an array of volume 0.2Å x 0.2Å x 0.2Å centred on the energy minimum. This directs program GRID to produce a GRIDKONT file centred on the refined energy minimum and suitable for calculating a temperature factor.

5. When the energy minimum is fully refined, MINGRID writes directives for two GRID point runs: at the observed water coordinates and at the predicted water coordinates.

CHECK This converts GRIDKONT energies to readable format.

CHECKREV This converts CHECKOUT energies back to machine-readable format.

NAGRG This calculates a temperature factor for the predicted water and writes an analysis of the prediction data to NAGRGOUT (longer version) and WATDAT (summary).

BAN This analyses the B-values of the water and the surrounding atoms.

These programs are listed on the appended microfiche.

Table III.2 Parameters for the water molecule and for the nitrate ion used in the prediction of the solvent of crystals of human lysozyme by program GRID.

Parameter	Water	Probe nitrate ion	Target nitrate nitrogen atom	Target nitrate oxygen atom
van der Waals' radius R in Å	1.70	2.81[a]	1.60	1.60
Number of effective electrons N_{eff}	7	24[b]	6	6
Polarizability α in $Å^3$	1.20	3.80[c]	0.60	2.14
Electric charge q in e	0.000	-1.000	0.704[d]	-0.568[d]
Number of hydrogen-bonds donated	2	0	0	0
Number of hydrogen-bonds accepted	2	6	0	2
Hydrogen-bond energy in kcal/mol	-4.0	-4.0	-	-4.0
Optimum hydrogen-bond radius in Å	1.4	2.61[e]	-	1.4
Hydrogen-bond angle in °[f]	110.0	-	-	-

Unless otherwise stated, parameters for the atoms of the nitrate ion are the same as those already used in program GRID. CHARMM [72] is the original source for many of these.

[a] This is the sum of the covalent bond length N-O (=1.21Å [285]) and the van der Waals' radius of oxygen. The nitrate ion is assumed to be spherical.

[b] In program GRID [79]:
N_{eff} = ½[(number of electrons in the outer shell - 1 electron/bond) + (total number of electrons - 1 electron/bond)]

[c] Average experimental value [286,287].

[d] The atomic charges were calculated using AMPAC [69] (see Appendix II). To check that these charges were suitable for use in program GRID, a comparison of the GRID charges [72] and the AMPAC charges [288] for amino-acids was carried out. This showed reasonable agreement between charges for N and O but less satisfactory agreement for C and S. As the nitrate ion contains only N and O, the AMPAC charges for these atoms did not appear to be incompatible with the GRID charges for other molecules.

[e] The probe hydrogen-bond radius = R - 0.2Å as for a carboxyl oxygen in program GRID.

[f] This angle is not explicitly defined for target atoms but is included implicitly in the angular factor Et. The angle subtended by hydrogen-bonds at the nitrate ion probe is not considered.

III.5 THE PARAMETERISATION OF THE WATER MOLECULE AND THE NITRATE ION FOR THE PREDICTION OF THE SOLVENT IN CRYSTALS OF HUMAN LYSOZYME BY PROGRAM GRID

The parameters assigned to the water molecule and the nitrate ion in program GRID are given in Table III.2.

The parameters for water are for the 'extended atom' representation and are used for both the water probe and target waters in the prediction of the water structure of crystals of HL.

The parameters for the nitrate ion were derived so as to be consistent with the parameters used for other chemical groups in program GRID (see Table III.2). They are given for both probe and target nitrate ions.

REFERENCES

1 Jencks,W.P. 'Catalysis in Chemistry and Enzymology' McGraw-Hill, New York, 1969.

2 Burgen,A.S.V.;Roberts,G.C.K.;Feeney,J. Nature 1975, 253, 753.

3 Chothia,C.H.;Janin,J. Nature 1975, 256, 705.

4 Pincus,M.R.;Scheraga,H.A. Acc.Chem.Res. 1981, 14, 299.

5 Fersht,A.R. Trends Biochem.Sci. 1987, 12, 301.

6 Fersht,A.R. Trends Biochem.Sci. 1984, 9, 145.

7 Burgen,A.S.V. J.Pharm.Pharmacology. 1966, 18, 137.

8 Lennard-Jones,J.E. Proc.Phys.Soc. 1931, 43, 461.

9 Lennard-Jones,J.E. Proc.R.Soc.London A 1924, 106, 463.

10 Buckingham,R.A. Proc.R.Soc.London A 1938, 168, 264.

11 London,F. J.Phys.Chem. 1942, 46, 305.

12 Olovsson,I. Croat.Chem.Acta 1982, 55, 171.

13 Finney,J.L. J.de Physique 1984, 45, C7-197.

14 Tanford,C. 'The Hydrophobic Effect: Formation of Micelles and Biological Membranes' 2nd ed, Wiley, New York, 1980.

15 Kauzmann,W. Adv.Prot.Chem. 1959, 14, 1.

16 Fischer,E. Ber.Deutsch Chem.Ges. 1894, 27, 2984.

17 Koshland,D.E. Proc.Nat.Acad.Sci.U.S.A. 1958, 44, 98.

18 Kollman,P. Acc.Chem.Res. 1985, 18, 105.

19 Austel,V. 'X-ray Crystallography and Drug Action' eds. Horn,A.S.;De Ranter,C.J. Clarendon Press, Oxford, 1984.

20 Hol,W.G.J. Angew.Chem.Int.Ed.Engl. 1986, 25, 767.

21 Nogrady,T. 'Medicinal Chemistry:A Biochemical Approach.' OUP, 1985.

22 Symposium Proceedings, Br. J. Pharmac. July 1988, in press.

23 Gill,E.W. 'Progress in Medicinal Chemistry' eds. Ellis,G.P.; West,G.B. 1965, 4, 39.

24 Gund,P. Ann.Rep.Med.Chem. 1979, 14, 299.

25 Gourley,D.R.H. 'X-ray Crystallography and Drug Action' eds. Horn,A.S.;De Ranter,C.J. Clarendon Press, Oxford, 1984, 95.

26 Langley,J.N. J.Physiol. 1878, 1, 339.

27 Parascandola,J. 'Towards Understanding Receptors' ed. Lamble,J.W. Elsevier, Amsterdam, 1981, 1.

28 Parascandola,J. 'Receptors Again' ed.Lamble,J.W.;Abbott,A.C. Elsevier, Amsterdam, 1984, 10.

29 Clark,A.J. J.Physiol. 1926, 61, 530,547.

30 Gaddum,J.H. J.Physiol. 1937, 89, 7P.

31 Martin,Y.C. 'Quantitative Drug Design' M.Dekker, New York, 1978.

32 Yalkowsky,S.H.;Sinkula,A.A.;Valvani,S.C. 'Physical Chemical Properties of Drugs' M.Dekker, New York, 1980.

33 Hansch,C.;Fujita,T. J.Am.Chem.Soc. 1964, 86, 1616.

34 Hansch,C.;Leo,A. 'Substituent Constants for Correlation Analysis in Chemistry and Biology' John Wiley, New York, 1979.

35 Free,S.M.;Wilson,J.N. J.Med.Chem. 1964, 7, 395.

36 Kier,L.B.;Hall,L.H. 'Molecular Connectivity in Chemistry and Drug Research' Academic Press, New York, 1976.

37 Saunders,M.R.;Tute,M.S.;Webb,G.A. J.Comp.Aided.Mol.Des. 1987, 1, 133.

38 Humblet,C.;Marshall,G.R. Ann.Rep.Med.Chem. 1980, 15, 267.

39 Marshall,G. Ann.Rev.Pharm. 1987, 27, 193.

40 Ghose,A.K.;Crippen,G.M. J.Comp.Chem. 1985, 6, 356.

41 Brint,A.T.;Willett,P. J.Chem.Inf.Comput.Sci. 1987, 27, 152.

42 Sheridan,R.P.;Venkataragharan,R. J.Comp.Aid.Mol.Des. 1987, 1, 243.

43 Weinstein,H.;Osman,R.;Green,J.P.;Topiol,S. 'Chemical Applications of Atomic and Molecular Electrostatic Potentials', Politzer,P.;Truhlar,D.G.eds. Plenum Press, New York, 1981.

44 Danziger,D.J.;Dean,P.M. J.Theor.Biol. 1985, 116, 215.

45 Willett,P. 'Similarity and Clustering in Chemical Information Systems' Research Studies Press, Letchworth, 1987.

46 Hodgkin,E.;Richards,W.G. Int.J.Quant.Chem.Quant.Biol.Symp. 1987, 14, 105.

47 Goodford,P.J. J.Med.Chem. 1984, 27, 557.

48 Beddell,C.R. Chem.Soc.Rev. 1984, 13, 279.

49 Kendrew,J.C.;Bodo,G.;Dintzis,H.M.;Parrish,R.G.;Wyckoff,H.; Phillips,D.C. Nature (London) 1958, 181, 662.

50 Kendrew,J.C.;Dickerson,R.E.;Strandberg,B.E.;Hart,R.G.; Davies,D.R.;Phillips,D.C.;Shore,V.C. Nature 1960, 185, 422.

51 Blake,C.C.F.;Koenig,D.G.;Mair,G.A.;North,A.C.T.;Phillips,D.C.; Sarma,V.R. Nature 1965, 206, 757.

52 Bernstein,F.C.;Koetzle,T.F.;Williams,G.J.B.;Meyer,E.F.; Bryce,M.D.;Rogers,J.R.;Kennard,O.;Shikanouchi,T.;Tasumi,M. J.Mol.Biol. 1977, 112, 535.

53 Stuart,D.I. 'Signal Transduction and Protein Phosphorylation' Heilmeyer,L.M.G. ed. Plenum Press, New York, 1987.

54 Hadju,J.;Machin,P.;Campbell,J.W.;Greenhough,T.J.;Clifton,I.J.; Zurek,S.;Gover,S.;Johnson,L.N.;Elder,M. Nature 1987, 329, 178.

55 Blundell,T;Johnson,L.N. 'Protein Crystallography' Academic Press, London, 1976.

56 Palau,J.;Argos,P.;Puigdomenech,P. Int.J.Pept.Prot.Res. 1982, 19, 394.

57 Cohen,F.E.;Abarbanel,R.M.;Kuntz,I.D.;Fletterick,R.J. Biochemistry 1986, 25, 266.

58 Lee,B;Richards,F.M. J.Mol.Biol. 1971, 55, 379.

59 Richards,F.M. Ann.Rev.Biophys.Bioeng. 1977, 6, 151.

60 Connolly,M.L. Science 1983, 221, 709.

61 Busetta,B.;Tickle,I.J.;Blundell,T. J.Appl.Cryst. 1983, 16, 432.

62 Kuntz,I.D.;Blaney,J.M.;Oatley,S.J.;Langridge,R.;Ferrin,T.E. J.Mol.Biol. 1982, 161, 269.

63 Blaney,J.M.;Jorgensen,E.C.;Connolly,M.L.;Ferrin,T.E.; Langridge,R.;Oatley,S.J.;Burridge,J.M.;Blake,C.C.F. J.Med.Chem. 1982, 25, 785.

64 Hayes,D.M.;Kollman,P. J.Am.Chem.Soc. 1976, 98, 3335.

65 Tomioka,N.;Itai,A.;Iitaka,Y. J.Comp.Aided Des. 1987, 1, 197.

66 Goodford,P.J. J.Med.Chem. 1984, 28, 849.

67 Lesk,A. Acta.Cryst. 1986, A42, 83.

68 Danziger,D.J. J.Mol.Graph. 1985, 3, 102.

69 Dewar,M.J.S.;Zoebisch,E.G.;Healy,E.F.;Stewart,J.J.P. J.Am.Chem.Soc. 1985, 107, 3902.

70 Clementi,E. Lecture notes in chemistry 1980, 19.

71 Pullman,B.;Pullman,A. Prog.Nucl.Acid.Res.(Mol. Biol.) 1969,9,327.

72 Brooks,B.R.;Bruccoleri,R.E.;Olafson,B.D.;States,D.J.; Swaminathan,S.;Karplus,M. J.Comp.Chem. 1983, 4, 187.

73 Weiner,S.J.;Kollman,P.A.;Case,D.A.;Singh,U.C.;Ghio,C; Alagona,G.;Profeta,S.;Weiner,P. J.Am.Chem.Soc. 1984, 106, 765.

74 Weiner,S.J.;Kollman,P.A.;Nguygen,D.T.;Case,D.A. J.Comp.Chem. 1986, 27, 230.

75 Hagler,A.T.;Huler,E.;Lifson,S. J.Am.Chem.Soc. 1974, 96, 5319, 5327.

76 Momany,F.A.;McGuire,R.;Burgess,A.;Scheraga,H.A. J.Phys.Chem. 1975, 79, 2361.

77 van Gunsteren,W.F.;Berendsen,H.J.C. J.Comp.Aided.Mol.Des. 1987, 1, 171.

78 Beddell,C.R.;Goodford,P.J.;Kneen,G.;White,R.D.;Wilkinson,S.; Wootton,R. Br.J.Pharmacol. 1984, 82, 397.

79 Goodford,P.J. personal communication.

80 Kuyper,L.F.;Roth,B.;Baccanari,D.P.;Ferone,R.;Beddell,C.R.; Champness,J.N.;Stammers,D.K.;Norrington,F.E.;Baker,D.T.; Goodford,P.J. J.Med.Chem. 1982, 25, 1120.

81 Matthews,D.A.;Bolin,J.T.;Burridge,J.M.;Filman,D.J.;Volz,K.W.; Kraut,J. J.Biol.Chem. 1985, 260, 392.

82 Reynolds,C.A.;Richards,W.G.;Goodford,P.J. Anti-Cancer Drug Des. 1987, 1, 291.

83 Blundell,T.L.;Cooper,J.;Foundling,S.I.;Jones,D.M.;Atrash,B.; Szelke,M. Biochemistry 1987, 26, 5587.

84 Beddell,C.R.;Goodford,P.J.;Norrington,F.E.;Wilkinson,S.; Wootton,R. Br.J.Pharmacol. 1976, 57, 201.

85 Cushman,D.N.;Cheung,H.S.;Sabo,E.F.;Ondetti,M.A. Biochemistry 1977, 16, 5484.

86 Boobbyer,D.N.A.;Goodford,P.J.;McWhinnie,P.M.;Wade,R.C. submitted 1988.

87 Reynolds,C.A.;Wade,R.C.;Richards,W.G.;Goodford,P.J. submitted 1988.

88 Landau,L.D.;Lifshitz,E.M. 'Course of Theoretical Physics' Englis ed. Pergamon Press, Oxford, 1960, Vol.8.

89 Pauling,L. 'The Nature of the Chemical Bond' Cornell University Press, New York, 3rd ed. 1960, Chapter 12.

90 Hall,D.;Pavitt,N. J.Comp.Chem. 1984, 5, 441.

91 Kollman,P.A.;Allen,L.C. Chem.Rev. 1972, 72, 283.

92 Pauling,L. Proc.Nat.Acad.Sci.U.S. 1928, 14, 359.

93 Coulson,C.A.;Danielsson,U. Arkiv.f.Fys. 1954, 8, 239,245.

94 Tsubomura,H. Bull.Chem.Jap. 1954, 27, 445.

95 Puranik,P.G.;Kumar,V. Proc.Indian Acad.Sci. 1963, 58, 29.

96 Morokuma,K.;Pedersen,L. J.Chem.Phys. 1968, 48, 3275.

97 Morokuma,K.;Win ick,J.R. J.Chem.Phys. 1970, 52, 1301.

98 Dreyfus,M.;Pullman,A. Theor.Chim.Acta. 1970, 19, 20.

99 Umeyama,H.;Morokuma,K. J.Am.Chem.Soc. 1977, 99, 330,1316.

100 Kroon-batenburg,L.M.J.;Kanters,J.A. J.Mol.Struct.(Theochem) 1983, 105, 417.

101 Taylor,R. J.Mol.Struct. 1981, 71, 311.

102 Gibson,C.;Scheraga,H.A. Proc.Nat.Acad.Sci. 1967, 58, 421.

103 Levitt,M. J.Mol.Biol. 1983, 168, 595,621.

104 Momany,F.A. 'Environmental Effects on Molecular Structures and Properties' ed.B.Pullman. D.Reidel, Dordrecht, Holland, 1976, 437.

105 Momany,F.A.;Carruthers,L.M.;McGuire,R.T.;Scheraga,H.A. J.Phys.Chem 1974, 78, 1595.

106 Rossky,P.J.;Karplus,M.;Rahman,A. Biopolymers 1979, 18, 825.

107 Vovelle,F.;Ptak,M. Int.J.Pept.Prot.Res. 1979, 13, 435.
Vovelle,F.;Genest,M.;Ptak,M.;Maigret,B.;Premi,C. J.Theor.Biol. 1980, 87, 85.

108 Nilsson,L.;Karplus,M. J.Comp.Chem. 1986, 27, 591.

109 McCammon,J.A.;Wolynes,P.G;Karplus,M. Biochem. 1979, 18, 937.

110 Vedani,A;Dunitz,J.D. J.Am.Chem.Soc. 1985, 107, 7653.

111 Vedani,A. J.Comp.Chem. 1988, 9, 269.

112 Hermans,J.;Berendsen,H.J.C.;van Gunsteren,W.F.;Postma,J.A. Biopolymers 1984, 23, 1512.

113 Snir,J.;Nemenoff,R.A.;Scheraga,H.A. J.Phys.Chem. 1978, 82, 2497.

114 Lippincott,E.R.;Schroeder,R. J.Chem.Phys. 1955, 23, 1099.

115 Artymiuk,P.J.;Blake,C.C.F. J.Mol.Biol. 1981, 152, 737.

116 Reiher,W.E. Ph.D.Thesis Harvard Univ. 1985.

117 Ramakrishnan,C.;Prasad,N. Int.J.Prot.Res. 1971, III, 209.

118 Ferraris,G.;Franchini-Angela,M. Acta.Cryst. 1972, B28, 3572.

119 Pedersen,B. Acta.Cryst. 1974, B30, 289.

120 Kroon,J.;Kanters,J.A.;van Duijnevelt-Van de Rijdt,J.G.C.M.;van Duijneveldt,F.B.;Vliegenthart,J.A. J.Mol.Struct. 1975, 24, 109.

121 Mitra,J.;Ramakrishnan,C. Int.J.Peptide Prot. Res. 1977, 9, 27.

122 Vinogradev,S.N. Int.J.Pept.Prot.Res. 1979, 14, 281.

123 Vinogradev,S.N. Int.J.Biopolymers 1979, 18, 1559.

124 Ceccarelli,C.;Jeffrey,C.A.;Taylor,R. J.Mol.Struct. 1981, 70, 255.

125 Taylor,R.;Kennard,O. J.Am.Chem.Soc. 1982, 104, 5063.

126 Jeffrey,G.A.;Maluszynska,H. Int.J.Biol.Macromol. 1982, 4, 17.

127 Jeffrey,G.A.;Mitra,J. Acta.Cryst. 1983, B39, 469.

128 Taylor,R.;Kennard,O.;Versichel,W. J.Am.Chem.Soc. 1983, 105, 5761.

129 Taylor,R.;Kennard,O.;Versichel,W. Acta.Cryst. 1984, B40, 280.

130 Murray-Rust,P;Glusker,J.P. J.Am.Chem.Soc. 1984, 106, 1018.

131 Allen,F.H.;Bellord,S.;Brice,M.D.;Cartwright,B.A.; Doubleday,A.;Higgs,H.;Hummelink,T.;Hummelink-Peters,B.G.; Kennard,O.;Motherwell,W.D.S.;Rogers,R.;Watson,D.G. Acta.Cryst. 1979, B35, 2331.

132 Baker,E.N.;Hubbard,R.E. Prog.in Biophys.and Mol.Biol. 1984, 44, 97.

133 Jeffrey,G.A.;Maluszynska,H.;Mitra,J. Int.J.Biol.Macromol. 1985, 7, 336.

134 Taylor,R.;Kennard,O. Acta.Cryst. 1983, B39, 133.

135 Olavsson,I.;Jonssen,P.G. 'The Hydrogen-Bond' eds. Schuster,P.; Zundel,G.;Sandorfy,C. North-Holland, Amsterdam,1976, Volume II, Chapter 8.

136 Millen,D.J. Croat.Chem.Acta 1982, 55, 133.

137 Jeffrey,G.A.;Mitra,J. J.Am.Chem.Soc. 1984, 106, 5546.

138 Kroon,J.;Kanters,J.A. Nature 1974, 248, 667.

139 Taylor,R.;Kennard,O. Acc.Chem.Res. 1984, 17, 320.

140 McWhinnie,P.M. Chemistry Part II thesis, Oxford, 1985.

141 Madden,J.J.;McGandy,E.L.;Seeman,N.C. Acta.Cryst. 1972, B28, 2377.

142 Lehman,M.S.;Koetzle,T.F.;Hamilton,W.C. Int.J.Peptide Protein Res. 1972, 4, 229.

143 Reynolds,W.F.;Peat,I.R.;Freedman,M.H.;Lyerla,J.R. J.Am.Chem.Soc. 1973, 95, 328.

144 Boschcov,P.;Seidel,W.;Muradian,J.;Tominaga,M.;Paiva,A.C.M.; Juliano,L. Bioorganic Chem. 1983, 12, 34.

145 Ramani,R;Boyd,R.J. Can.J.Chem. 1981, 59, 3232.

146 Mezey,P.G.;Ladik,J.J.;Suhai,S. Theor.Chim.Acta. 1979, 51, 323.

147 Black,J.W.;Durant,G.J.;Emmett,J.C.;Ganellin,C.R. Nature 1974, 248, 65.

148 Ganellin,C.R. J.Appl.Chem.Biotechnol. 1978, 28, 183.

149 Boobbyer,D.N.A. Chemistry Part II thesis, Oxford, 1986.

150 Janoscheck,R. Croat.Chem.Acta. 1982, 55, 75.

151 Levitt,M.;Perutz,M.F. J.Mol.Biol. 1988, 201, 757.

152 Nalini,V.;Desiraju,G.R. J.Chem.Soc.Commun. 1986, 13, 1030.

153 Carlstrom,D.;Bergin,R. Acta.Cryst. 1967, 23, 313.

154 Caron,M.C.;Mukherjee,C.;Lefkowitz,R.J. 'Receptors in Pharmacology' eds. Smythies,J.R.;Bradley,R.J. M.Dekker, New York, 1978; 97.

155 Ganellin,C.R. J.Med.Chem. 1977, 20, 579.

156 Carlstrom,D. Acta.Cryst. 1973, B29, 161.

157 Dittrich,F.;Berlin,P.;Kopke,K.;Repke,K.R.H. Curr.Top.Membr.Transp. 1983, 19, 251.

158 DePover,A.;Godfraind,T. Naunyn-Schmiedebergs Arch.Pharmacol. 1982, 321, 135.

159 Dzimin,N.;Fricke,U. Br.J.Pharmacol. 1988, 93, 281.

160 Griffin,J.F.;Rohrer,D.C.;Ahmed,K.;From,A.H.;Hashimoto,T.; Rathore,H.;Fullerton,D.S. Mol.Pharm. 1986, 29, 270.

161 Poulos,T.L.;Finzel,B.C.;Gunsalus,I.C.;Wagner,G.C.;Kraut,J. J.Biol.Chem. 1985, 260, 16122.

162 Poulos,T.L.;Finzel,B.C.;Howard,A.J. J.Mol.Biol. 1987, 195, 687.

163 Poulos,T.L.;Finzel,B.C.;Howard,A.J. Biochem. 1986, 25, 5314.

164 Clark,K.J. Chemistry Part II thesis, Oxford, 1986.

165 Mitra,J;Ramakrishnan,C. Int.J.Peptide Prot.Res. 1981, 17, 401.

166 Bryant,R.C. 'Biophysics of Water' ed. Franks,F. Wiley, New York, 1981, 178.

167 Belleau,B. Ann.N.Y.Sci. 1967, 139, 580.

168 IJzerman,A.P.;van Vlijmen,H.W.T. J.Comp.Aided.Mol.Des. 1988, 2, 43.

169 Mitchell,T.J. Ph.D. thesis, University of Surrey, 1988.

170 Falk,M.;Knop,O. 'Water: A Comprehensive Treatise' ed. Franks,F. Plenum Press, New York, 1973, 2, 55.

171 Blake,C.C.F.;Pulford,W.C.A.;Artymiuk,P.J. J.Mol.Biol. 1983, 167, 693.

172 Chidambaram,P.;Sequira,A.;Sikka,S.K. J.Chem.Phys. 1964, 41, 3616.

173 Campbell,D. Chemistry Part II Thesis, Oxford, 1988.

174 Matthews,B.W. Ann.Rev.Phys.Chem. 1976, 27, 493.

175 Finney,J.L.;Goodfellow,J.M.;Howell,P.L.;Vovelle,F. J.Biomol.Struct.Dyn. 1985, 3, 599.

176 Bolin,J.T.;Filman,D.J.;Matthews,D.A.;Hamlin,R.C.;Kraut,J. J.Biol.Chem. 1982, 257, 13650.

177 Nagy,P. J.Comp.Aided Mol.Des. 1988, 2, 65.

178 Rahman,A.;Stillinger,F.H. J.Chem.Phys. 1971, 55, 3336.

179 Kuntz,I.D.;Kauzmann,W. Adv.Prot.Chem. 1974, 28, 239.

180 Finney,J.L.;Goodfellow,J.M.;Poole,P.L. 'Structural Molecular Biology' eds. Davies,D.B.;Saenger,W;Danyluk,S.S. Plenum Press, New York, 1982.

181 Saenger,W. Ann.Rev.Biophys. and Biophys.Chem. 1987, 16, 93.

182 Finney,J.L. Phil.Trans.Roy.Soc. 1977, B278, 3.

183 Edsall,J.T.;McKenzie,H.A. Adv.Biophys. 1983, 16, 53.

184 Finney,J.L. 'Water:A Comprehensive Treatise' ed.Franks,F. Plenum,N.Y. 1979, 6, 47.

185 Rose,G.D. Nature 1978, 272, 586.

186 Wolfenden, R. Biochem. 1978, 17, 201.

187 Wolfenden,R.;Andersson,L.;Cullis,P.M.;Southgate,C.C.B. Biochemistry 1981, 20, 849.

188 Teeter,M.M. Proc.Nat.Acad.Sci. 1984, 81, 6014.

189 Phillips,S.E.V. J.Mol.Biol. 1980, 142, 531.

190 Finney,J.L. 'Biophysics of Water' ed. Franks,F. Wiley, New York, 1982, 73.

191 Finney,J.L. 'Water Science Reviews' ed. Franks,F. 1985, 1, 93.

192 Bernal,J.D.;Fowler,R.H. J.Chem.Phys. 1933, 1, 515.

193 Berendsen,H.J.C.;Postma,J.P.M.,van Gunsteren,W.F.;Hermans,J. 'Intermolecular Forces' ed. Pullman,B. 1981, 331.

194 Jorgensen,W.L. J.Am.Chem.Soc. 1981, 103, 335.

195 Jorgensen,W.L. J.Chem.Phys. 1982, 77, 4156.

196 Reimers,J.R.;Watts,R.O.;Klein,M.L. Chemical Phys. 1982, 64, 95.

197 Lie,G.C.;Clementi,E.;Yoshimine,M. J.Chem.Phys. 1976, 64, 2314.

198 Rowlinson,J.S. Trans.Faraday Soc. 1951, 47, 120.

199 Ben-Naim,A.;Stillinger,F.H. 'Water and Aqueous Solutions' ed. Horne,R.A. Wiley-Interscience, New York, 1972, 295.

200 Stillinger,F.H.;Rahman,A. J.Chem.Phys. 1974, 60, 1545.

201 Shipman,L.L.;Scheraga,H.A. J.Chem.Phys. 1974, 78, 909.

202 Snir,J.;Nemenoff,R.A.;Scheraga,H.L. J.Phys.Chem. 1982, 82, 2504.

203 van Gunsteren,W.F;Karplus,M. Biochem. 1982, 21, 2259.

204 Lemberg,H.L.;Stillinger,F.H. J.Chem.Phys. 1975, 62, 1677.

205 Campbell,E.S.;Mezei,M. J.Chem.Phys. 1977, 67, 2338.

206 Stillinger,F.H.;David,C.W. J.Chem.Phys. 1978, 69, 1473.

207 Barnes,P.;Finney,J.L.;Nicholas,J.D.;Quinn,J.E. Nature 1979, 252, 459.

208 Gellatly,B.J.;Quinn,J.E.;Barnes,P.;Finney,J.L. Mol.Phys. 1983, 50, 949.

209 Barker,J.A.;Watts,R.O. Chem.Phys.Lett. 1969, 3, 144.

210 Beveridge,D.L.;Mezei,M.;Mehrotra,P.K.;Marchese,F.T.;Ravi-Shankar,G.;Vasu,T.R.;Swaminathan,S. 'Molecular Based Study of Fluids.' eds. Haile,J,M.;Mansoori,G.A. ACS, Washington. 1983, 297.

211 Metropolis,N.;Rosenbluth,A.W.;Rosenbluth,M.N.;Teller,A.H.; Teller,E. J.Chem.Phys. 1953, 21, 1087.

212 Eisenberg,D.;McLachlan,A.D. Nature 1986, 319, 199.

213 Pullman,A;Pullman,B. Quart.Rev.Biophys. 1975, 7, 505.

214 Richards,W.G. 'Water: A Comprehensive Treatise' ed. Franks,F. Plenum, New York, 1979, 6, 123.

215 Curtiss,C.;Fruip,D.;Blander,M. J.Chem.Phys. 1979, 71, 2703.

216 Stuart,D.I.;Phillips,D.C. Met.Enzymol. 1985, 115, 117.

217 Sternberg,M.J.E.;Grace,D.E.P.;Phillips,D.C. J.Mol.Biol. 1979, 130, 131.

218 Phillips,D.C. Biochem. Soc. Symp. 1982, 46, 1.

219 Frey,M.N.;Lehmann,M.S.;Koetzle,T.F.;Hamilton,W.C. Acta.Cryst. 1973, B29, 876.

220 Goodfellow,J.M.;Finney,J.L.;Barnes,P. Proc.R.Soc.Lond. 1982, B214, 213.

221 Imoto,T.;Johnson,L.N.;North,A.C.T.;Phillips,D.C.;Rupley,J.A. 'The Enzymes' ed. Boyer,P.D. Academic Press, London, 1972, 7, 665.

222 Handoll,H.H.G. D.Phil. Thesis, University of Oxford, 1985.

223 Hagler,A.T.;Moult,J. Nature 1978, 272, 222.

224 Clementi,E.;Ranghino,G.;Scordamaglia,R. Chem.Phys.Lett. 1977, 49, 218.

225 Ranghino,G.;Clementi,E. Gazz.Chim.Ital. 1978, 108, 157.

226 Holloway,D.W. Chemistry Part II thesis, Oxford, 1985.

227 Artymiuk,P.J. DPhil Thesis, Oxford. 1979.

228 Koehler,J.;Saenger,W.;van Gunsteren,W.F. Eur.Biophys.J. 1987, 15, 211.

229 Betzel,C;Saenger,W.;Hingerty,B.E.;Brown,G.M. J.Am.Chem.Soc. 1984, 106, 7545.

230 Zabel,V.;Saenger,W.Mason,S.A. J.Am.Chem.Soc. 1986, 108, 3664.

231 Szejtli,J. 'Cyclodextrins and their inclusion complexes' Budapest, 1982.

232 Saenger,W. Angew.Chem.Int.Ed.Engl. 1980, 19, 344.

233 Uekama,K.;Fuijnaga,T.;Hirayama,F.;Otagiri,M. J.Pharm.Sci. 1983, 72, 1338.

234 Timbury,M.C. 'Notes on Medical Virology' Churchill Livingstone, New York, 1978.

235 Webster,R.G.;Laver,W.G.;Air,G.M.;Schild,G.C. Nature 1982, 296, 115.

236 Wiley,D.C.;Wilson,I.A.;Skehel,J.J. Nature 1981, 289, 373.

237 Stuart-Harris,C.H.;Schild,C.G.;Oxford,J.S. 'Influenza: The viruses and the disease.' E.Arnold, 1985.

238 Nir,S.;Stegmann,T.;Wilschut,J. Biochemistry 1986, 25, 257.

239 Maeda,I.;Ohnishi,S. FEBS Lett. 1980, 122, 283.

240 Air,G.M. Proc.Nat.Acad.Sci.U.S.A. 1981, 78, 7639.

241 WHO Memorandum Bull.Wld.Hlth.Org. 1980, 58, 585.

242 Schotissek,C.;Burger,H.;Kistner,O.;Shortridge,K.F. Virology 1985, 147, 287.

243 Hoyle,F.;Wickramasinghe,N.C. Nature, 1986, 322, 509.
Hoyle,F.;Wickramasinghe,N.C. 'Viruses from outer space' Watkins,J. U.C.Cardiff Press, 1986.

244 Tyrell,D.A. Phil.Trans.Roy.Soc.Lond. 1980, B288, 449.

245 Bekimirov,T.A.;Douglas,R.G.;Dolin,R.;Galasso,G.J.;Krylov,V.F.; Oxford,J. WHO Bull. 1985, 63, 51.

246 Skehel,J.J.;May,A.J.;Armstrong,J.A. J.Gen.Virol. 1977, 38, 97.

247 Bowman,W.C;Rand,M.J. 'Textbook of Pharmacology' Blackwell, Oxford, 1980.

248 Indulen,M.K.;Kalninya,V.A. 'Development in Anitviral Therapy' eds. Collier,L.H.;Oxford,J. Academic Press, London, 1980, 107.

249 Smith,R.A.;Kirkpatrick,W. eds. 'Ribavirin: A Broad Spectrum Antiviral Agent' Academic Press, London, 1980.

250 Scott,G.M.;Tyrell,D.A.J. Br. Med. J. 1980, 280, 1558.

251 Leo,A.;Weininger,D. Pomona Coll.Med.Chem.Prog. California, USA 1984.

252 Wilson,I.A.;Skehel,J.J.;Wiley,D.C. Nature 1981, 281, 366.

253 Knossow,M.W.;Lewis,M.;Rees,D.;Wilson,I.A.;Skehel,J.J.;Wiley,D.C. Acta. Cryst. 1986, B42, 627.

254 Weis,W.,Brown,J.H.;Cusack,S.;Paulson,J.C.;Skehel,J.J.;Wiley,D.C. Nature 1988, 333, 426.

255 Wharton,S.A.;Skehel,J.J.;Wiley,D.C. Virology 1986, 149, 27.

256 Varghese,J.N.;Laver,W.G.;Colman,P.M. Nature 1983, 303, 35.

257 Fields,S.;Winter,G.;Brownlee,G.G. Nature 1981, 290, 213.

258 Markoff,L;Lai,C.J. Virology 1982, 119, 288.

259 Bentley,D.R.;Brownlee,G.G. Nuc.Acids.Res. 1982, 10, 5033.

260 Skehel,J.J.;Stevens,D.J.;Daniels,R.S.;Douglas,A.R.;Knossow,M.;
Wilson,I.A.;Wiley,D.C. Proc.Nat.Acad.Sci(USA) 1984, 81, 1179.

261 Both,G.W.;Sleigh,M.J. J.Virol. 1981, 39, 663.

262 Winter,G.;Fields,S.;Brownlee,G.G. Nature 1981, 292, 72.

263 Alderson,G.A. Part II Thesis, University of Oxford, 1987.

264 Rogers,G.N.;Paulson,J.C.;Daniels,R.S.;Skehel,J.J.;
Wilson,I.A.;Wiley,D.C. Nature 1983, 304, 76.

265 Biosym Technologies Ltd.

266 Jones,T.A. J.Appl.Crystallogr. 1978, 11, 268.
Jones,T.A. 'Computational Crystallography' ed. Sayre,D.
Clarendon Press, Oxford, 1982; 303.

267 Lewis,R.A. communication, Soc.Chem.Ind. February 1988.

268 Allen,F.H.;Kennard,O. Persp.in Comp. 1983, 3, 28.

269 Merrifield,R.B. J.Am.Chem.Soc. 1963, 85, 2149.

270 Erickson,B.W.;Merrifield,R.B. 'The Proteins' ed. Neurath,H.;
Hill,R.L. 3rd. ed. 1976, Vol II, 257.

271 Witter,A. 'Stereochemistry and Biological Activity of Drugs'
eds. Ariens,E.J.;Soudijn,W.;Timmermans,P.B.M.W.M.
Blackwell,Oxford, 1983, 151.

272 Jaenicke,R. Prog.Biop.and Mol.Biol. 1987, 49, 117.

273 Gill,E.W. Proc.Roy.Soc.B 1959, 150, 381.

274 Applied Biosystems Peptide Synthesis Handbook.

275 Kaiser,E.;Colescott,R.L.;Bossinger,C.D.;Cook,P.I. Anal.Biochem.
1970, 34, 595.

276 Harmon,M.W.;Rota,P.A.;Walls,H.H.;Kendal,A.P. J.Clin.Microbiol.
1988, 26, 333.

277 Wiley,D.C.;Skehel,J.J. Ann.Rev.Biochem. 1987, 56, 365.

278 Salk,J.E. J.Immunol. 1944, 49, 87.

279 Quantum Chemistry Program Exchange 506.

280 Bingham,R.C.;Dewar,M.J.S.;Lo,D.H. J.Am.Chem.Soc. 1975, 97, 1285, 1294, 1302, 1307.

281 Dewar,M.J.S.;Thiel,W. J.Am.Chem.Soc. 1977, 99, 4899.

282 Richards,W.G.;Horsley,J.A. 'Ab Initio Molecular Orbital Calculations for Chemists' Clarendon Press, Oxford, 1970.

283 Richards,W.G. 'Quantum Pharmacology' Butterworth, London, 2nd. ed. 1983.

284 Singh,U.C.;Kollman,P.A. J.Comp.Chem. 1984, 5, 129.

285 Ketelaar,J. 'Chemical Constitution' Elsevier, New York, 2nd. ed. 1958.

286 Salzmann,J.J.;Jorgensen,C.K. Helv.Chim.Acta. 1968, 51, 1276.

287 Tessman,J.R.;Kahn,A.H.;Shockley,W. Phys.Rev. 1953, 92, 890.

288 Mian,S. personal communication.