I
Introduction

Molecular Interaction Fields. Edited by G. Cruciani

ISBN: 3-527-31087-8

1
The Basic Principles of GRID

Peter Goodford

1.1
Introduction

One cannot go out and buy a computer program in the confident expectation that it will do its job exactly as expected. Of course there are some things, like a lawn mower, where a relatively quick and easy test can be made to discover if it is good enough. Can it cut long grass? Cut wet grass? Does it pick up all the clippings? Will it leave beautiful light and dark stripes on the lawn? However, a molecular interaction field (MIF) is a good deal more complicated than a lawn mower, and it is not at all easy to establish which MIF programs work in a satisfactory way. Each program must be assessed very carefully before deciding what software should be used for any particular task, and many different factors must be taken into account. Some are obvious, like the available computer hardware; its speed; the size of its memory; and the amount of disk space on the user's system. Some are less apparent, such as the objectives, priorities and overall philosophy of the people who wrote the software, and the way in which they devised and calibrated their MIF. The most important factor is to be certain in one's own mind about the precise jobs which one wants the program to do.

1.2
Philosophy and Objectives

Even the most superficial study of molecular interaction fields shows that each MIF has its own particular characteristics. This field may put great emphasis on the accurate computation of the individual atomic charges. A different MIF may give more attention to the way in which those charges are distributed between an atom and its bonds, and a third may place some of each atom's charge onto its lone pair electrons. Another MIF may attempt to make accurate predictions of the pK_a of every polar atom, in order to be certain that each one is appropriately protonated before the MIF computations begin. Some fields may require the system under investigation to have zero overall charge. Other fields will happily do com-

Molecular Interaction Fields. Edited by G. Cruciani

ISBN: 3-527-31087-8

putations on a couple of phosphate ions, for example, with none of the oxygens protonated and no counter cations so that the two anionic phosphates move remorselessly apart until they vanish at the edge of the universe! One field may always compute the local pH, and another may need pH information as part of the input data. This field may give detailed attention to estimating the local dielectric environment and how it changes from place to place, while that one may assume an arbitrary overall dielectric constant.

It is not only the electrostatic treatment which is different in each MIF, but also other molecular characteristics. Many fields require all the hydrogen coordinates to be defined, but some only need the location of hydrogen-bonding hydrogens, and others take no specific account of any hydrogen positions. Some fields use simple harmonic motion to describe bond vibrations, but others attempt to consider deviations from harmonicity. Some have dedicated computations which deal with hydrogen bonds, and others pay no particular attention to hydrogen bond geometry. Most fields do not allow for tautomeric changes, but some can take tautomerism into account and a few can cope with alterations in the hybridisation of an atom. Some deal exclusively with enthalpy, but others can take account of entropy which is a major component of the hydrophobic effect.

Whenever another research group begins to study MIFs, they introduce a new perspective and a new set of ideas, so the extension and improvement of force fields has been a matter of continuously improving approximations. There will never be an absolutely correct MIF, but even the very earliest work was surprisingly valuable. Huggins and Pauling [1] introduced their atomic radii seventy years ago, but they immediately extended the understanding of crystal packing and of many other properties. However no force field is perfect, and one can only hope that the approximations will continue to improve in the years ahead.

1.3 Priorities

The priorities of the people who create any MIF are a concrete manifestation of their scientific philosophy and overall objectives, and seven requirements seemed particularly important when the GRID force field [2] was being designed:

1. This force field was explicitly intended for use with the GRID method.
2. The overall objective was to predict where ligands bind to biological macromolecules, and so gain a better understanding of the factors involved in binding. However that improved understanding should also help in the design of improved ligands.
3. The GRID force field could have been calibrated either by using theoretical calculations, or by studying experimental observations, and after much discussion the experimental approach was adopted whenever possible.
4. The input data must always be thoroughly checked before every computation, and an associated program called GRIN was written to do this job.

5. The equations used for the computation must be reasonably straightforward, so that anybody working in the drug-discovery field (biologist, pharmacologist, medicinal chemist, crystallographer, clinician, statistician, patent expert, administrator etc.) could easily discover exactly how program GRID had calculated any particular result.
6. It must be relatively easy for anybody working in the field to interpret the output from GRID.
7. An annual reappraisal policy was established so that the worst features in the current version of GRID would always be identified, and could be dealt with appropriately in each successive year.

It is conventional to write impersonally about scientific research, but subjective decisions are made when one decides which features are worst, or which objectives are most important for a program. The personal pronoun "we" will therefore be used in this article, whenever it is appropriate to draw attention to subjectivity. "We" are still discussing priorities for the forthcoming release of GRID, but before describing the GRID force field in detail we must first describe how the GRID method actually works.

1.4 The GRID Method

There are many programs which can be used compute the electrostatic potential around a molecule. A computer model is first prepared from the x, y, z coordinates of the atoms, and this model is then surrounded by an imaginary orthogonal grid.

The next step is to compute the work needed to bring a unit electrostatic charge from infinity to the first point on the grid, and the total work required for this job is a measure of the electrical potential at that particular grid point. The same procedure is then repeated for each of the other grid points, including those which are actually inside the molecule, until the potential has been calculated for every position.

At this stage it would be possible to print out the individual potential values as a table of numbers for detailed study. The findings could then be used as input for further computations, but studying a printed data table would be a rather clumsy way of displaying the results, and a much better method is to create a three-dimensional computer plot showing a contour surface surrounding the molecule. This contour surface defines a single user-selected value of the electrostatic potential, and the final picture usually shows the molecule together with something looking rather like a child's balloon!

Program GRID works in very much the same way, but the objective is to obtain chemically specific information about the molecule. An electrostatic potential does not normally allow one to differentiate between favorable binding sites for a primary or a secondary or a tertiary amine cation, or tetramethyl ammonium or

pyridinium or a sodium cation, and the GRID method is an attempt to compute analogous potentials which do have some chemical specificity. The generic name "*target*" is given to the molecule (or group of molecules) being studied by GRID, and the object used to measure the potential at each point is called a "*probe*". The individual potential values are called "*GRID values*" and the final computer plot is called a "*GRID map*". Many different probes can be used on the same target one after the other, and each probe represents a specific chemical group so that chemically specific information can be accumulated about the way in which the target might interact favorably with other molecules.

The GRID method differs in three critical ways from traditional programs which just display electrostatic potentials:

1. GRID probes are often anisometric.
2. The target "responds" when the probe is moved around it from place to place.
3. It is assumed that both the target and the probe are immersed in water.

These differences must now be considered in more detail.

1.4.1 GRID Probes Are Anisometric

Most GRID probes are anisometric because each probe represents an atom or a small group of atoms. For example a carbonyl oxygen probe is one oxygen atom with a couple of sp2 lone pairs. It has a size and a polarizability and an electrostatic charge, and each lone pair can accept one hydrogen bond. The center of the oxygen is placed at the first grid point, and a check is then made for unacceptably bad close contacts. If none is found the program then searches for nearby hydrogen-bond donor atoms on the target, and a list of those donors is made and sorted. Target atoms are rejected from the list if their donor hydrogens are pointing the wrong way, and the probe is then rotated (keeping its oxygen fixed at the grid point) so that its lone pairs will be oriented until they make the best possible hydrogen bonds to nearby target atoms. When this has been done the GRID force field is used to compute a GRID value for that particular probe at that particular point, and the whole process is repeated systematically until the potential for carbonyl oxygen is known for every grid point on the map.

An aromatic sp2 hydroxy probe differs in several ways from carbonyl oxygen. The oxygen atom of the hydroxy is placed at the grid point as before, but the probe has a larger polarizability and makes hydrogen bonds of a different strength. It can accept only one hydrogen bond, but the oxygen is bonded to a hydrogen atom which can donate. If both the donor and acceptor hydrogen bonds are made simultaneously they will be mutually constrained towards the sp2 angle of 120°. The bond length from the oxygen to its hydrogen is about 1 Å, and the probe's donor hydrogen moves round the grid point at this distance when the probe is rotated. Figure 1.1 shows the target with an sp2 hydroxy probe placed at the first point, ready for the computation to begin.

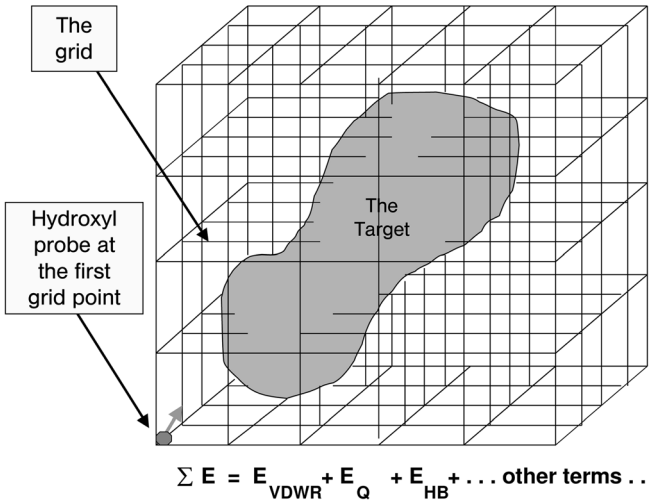

Figure 1.1. The set up for GRID. See text Section 1.4.1.

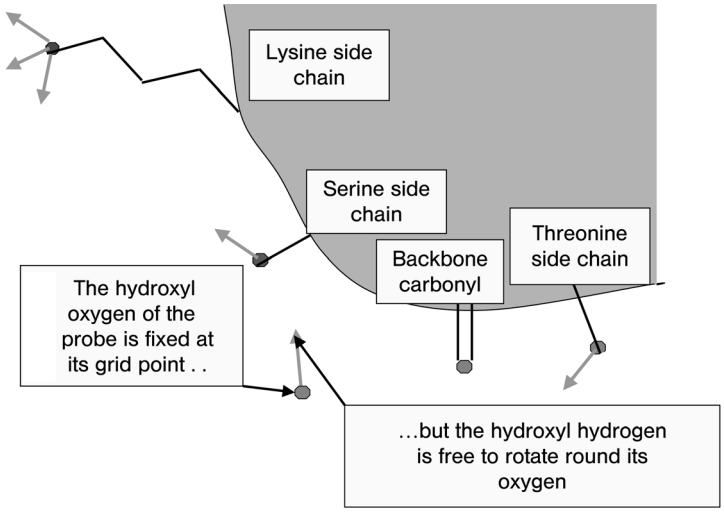

Figure 1.2. The initial orientation of an sp2 hydroxy probe at its GRID point.

The sp2 carboxyl oxygen probe differs from both sp2 carbonyl and sp2 hydroxy, having a much greater polarizability and much greater negative charge than either. The sp3 aliphatic hydroxy probe is distinguished by making its hydrogen bonds at the sp3 angle of 109° instead of 120°, and by accepting at two lone pairs instead of just one. "Multi-atom probes" can also be used, such as aromatic carboxylate which represents the anion of a complete benzoic acid molecule. This multi-atom probe has two sp2 carboxy oxygens both bonded to the carboxy carbon

which is bonded to the aromatic ring. Each oxygen has a couple of lone pairs, all appropriately oriented and of appropriate strength. Both oxygens are deprotonated, and they both have a substantial negative charge which is partly counterbalanced by a modest positive charge on the carboxy carbon, so the whole probe has an overall charge of –1. Its oxygens are both identical, and one of them is fixed as usual at the grid point. The whole multi-atom probe is then rotated to find all the orientations in which it can make good hydrogen bonds to the target, and good electrostatic interactions, while avoiding steric clashes. The chosen oxygen always stays on its grid point, and the GRID potential for that point is computed when the best orientation of the whole multiatom probe has been established.

A multiatom probe usually finds pairs of minima which would correspond in this example to the two oxygens of the carboxylate group. Of course the computation for a multi-atom probe takes somewhat longer than the map for a simpler probe, but the force field was written explicitly for the program and so GRID computations are never particularly time consuming.

1.4.2 The Target "Responds" to the Probe

Figure 1.2 shows in more detail how an aromatic sp2 hydroxy probe might be placed on its grid point at the start of a cycle of computation. In this figure the hydrogen of the probe happens to be pointing by chance towards a nearby serine residue of the target. The orientation of the probe is completely random at this early stage of the job, but with a slight readjustment GRID can make the probe's hydrogen point directly at the serine's side chain sp3 hydroxy oxygen. A hydrogen bond could then be formed and that would be quite a good arrangement, but a better one is shown in Fig. 1.3. Program GRID has to search and find the better alternative, and must do three things to make this happen:

1. GRID has to rotate the probe, while keeping its oxygen firmly anchored at the grid point, until its hydrogen is redirected towards the nearby backbone sp2 carbonyl oxygen as shown in Fig. 1.3.
2. The probe then has to spin about an imaginary sp2–sp2 axis (A in Fig. 1.3) which links it to the backbone carbonyl oxygen, until the probe's own lone pair points as directly as possible towards the sp3 hydroxy oxygen of the serine.
3. The sp3 hydroxy group of the serine must finally spin about bond B which links it to the to the rest of the protein, until its hydrogen points as well as possible towards the probe's lone pair.

This rotation of the serine oxygen is called the "response of the target to the probe", and finding the best response is often a much more complicated job than it appears in Fig. 1.3. There are usually many different hydrogen bonding groups on or near the surface of the target, reasonably close to the probe, as shown in Fig. 1.4, and they must all be taken into account.

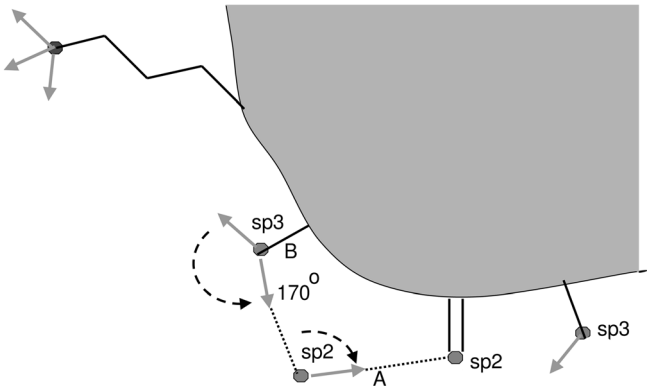

Figure 1.3. Rotational adjustments of the probe. See text.

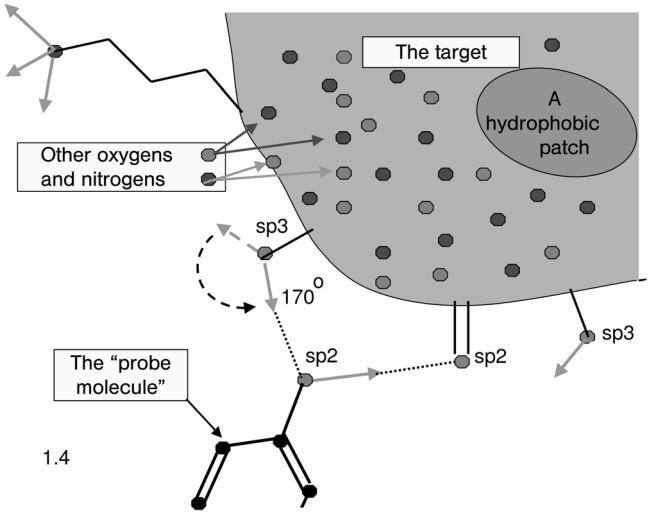

Figure 1.4. The final position of the probe showing additional features of the binding site. See text.

Methods are provided in GRID so that the user can adjust the size of the "response." For instance he could prevent the serine hydroxy from rotating on its axis, if he knew that it was already making another strong hydrogen bond which would be broken if the probe interacted as described above. There is also a lysine side chain shown near the top of Figs. 1.2, 1.3 and 1.4, but the NH_3^+ group of that lysine cannot reach the probe at its grid point as things are shown in the figures. However resetting one of the directives would allow the lysine side chain to swing down, and perhaps make a useful hydrogen-bond interaction with the probe. The directives are always set by default so that things like this do not happen, and long side chains like the lysine do not normally search around during a regular GRID

run, unless the user has made a positive decision to release them. That kind of decision can only be taken after a thorough study of the binding site of the protein. The user must understand some or all of its properties, and this enhancement of the user's understanding was a major objective when program GRID was being written.

1.4.3 The Target is Immersed in Water

The concept of electrical potentials was developed by physicists in the 19th century, and they quite naturally took a vacuum as their reference state. The dielectric constant of a vacuum is 1.0 by definition, and many of the early experiments on electrostatics were made in air which has a dielectric constant very close to unity. However biological systems are full of water, and biologists must invoke a dielectric constant of up to 80 in order to make traditional electrostatic calculations. It is therefore hardly surprising that MIF computations in biological systems tend to give unstable results, when such a large dielectric correction factor must be used.

The GRID force field was designed on a more appropriate basis for biology. It is assumed a priori that the environment surrounding the target has a bulk dielectric of 80, and that the dielectric diminishes towards 4 in the deep center of a large globular macromolecule. These are the default values which were used in calibrating the MIF, but of course each user can alter them to any reasonable alternative during his own GRID runs. It has been reported [3] that a value between 10 and 20 gives results which agree better with experiments on small molecules.

Some years ago a large oil company wanted to use program GRID for calculations on zeolites. These are minerals, and it was first necessary to calibrate several elements such as silicon which had not previously been used in GRID runs. Preliminary computations were then started, but the results from zeolites were misleading. A bulk dielectric of 80 would clearly be inappropriate in this case, because zeolites are used at approximately 300 °C for oil refining and are therefore completely dry. However it was impossible to find any dielectric values which yielded satisfactory results for zeolites, and this seems to demonstrate that one should not expect a single MIF to work for every system. Each force field should be calibrated for the job in hand, and much more sophisticated methods are needed if one wishes to study all 100 elements in all experimental conditions. GRID and its force field must be restricted to the wet biological environment for which they were calibrated.

1.5 The GRID Force Field

The target is always prepared and checked by an associated program called GRIN which is used before the actual GRID run begins, and perhaps the most important job of GRIN is the amalgamation of every nonpolar hydrogen atom of the

target with its neighboring heavy atom to give an "extended atom". Consider, for example, a very small target $H_3C–CH_3$ consisting of one ethane molecule. GRIN will represent this by two extended methyl atoms instead of two carbons plus six hydrogens, and this condensation of eight real atoms into a pair of extended atoms allows the GRID programme to run much faster. Of course there is some loss of accuracy, but real targets for GRID are usually much more complicated than ethane. Real targets usually have conformationally flexible side chains, and it is very easy to place too much emphasis on the exact hydrogen coordinates of a biological macromolecule when those hydrogens have not even been observed by the X-ray crystallographer.

Programme GRIN also checks the target for errors, and the GRID run then begins. The GRID energies are usually computed pairwise between the probe at its grid point and each extended atom of the target, one by one. Recent releases of the program include more terms, but early versions used only three energy components for each pairwise energy E_{PAIR}:

$$E_{PAIR} = E_{LJ} + E_Q + E_{HB} \quad (1)$$

1.5.1 The Lennard-Jones Term

The E_{LJ} term in Eq. (1) is the well-known "Lennard-Jones energy", and is computed as the sum of two terms:

$$E_{LJ} = (Ad^i - Bd^j)\ F \quad (2)$$

in which $i = -12$, $j = -6$ and F=1. A and B are positive constants which are chosen so that E_{LJ} will be calculated in kcal mol^{-1}, and d is the distance between the probe at its grid point and the extended atom of the target. The first term Ad^i is always positive, and represents the repulsion of the atoms for each other if they are unacceptably close together. The second term $-Bd^j$ is negative and measures their induction and dispersion attractions for each other.

1.5.2 The Electrostatic Term

E_Q is an electrostatic term computed as $E_Q = q_1q_2/\mathrm{d}D$ where q_1 is the charge of the probe, q_2 the charge of the extended target atom, and D is the dielectric constant value to be used when their pairwise electrostatic interaction is calculated. Computing D is a slow business [2] because a square-root calculation is always required, and many atom pairs must be studied in a GRID run. D must be estimated individually for each pair, and extensive tests have shown that acceptable results cannot be obtained reliably unless all pairwise values of D and E_Q are worked out. Of course it is very tempting to ignore E_Q if the interacting atoms are

more than 20 (or 30 or 40) Å apart, but this attractive short cut is unacceptable because it often gives rise to significant errors. The method finally adopted [2] to compute D is based on classical electrostatics [4] with the important additional assumption that one is dealing with a system of two homogeneous phases separated by a flat planar surface. It is easy to construct models in which the assumption of a flat surface can give rather misleading results, but in practice this does not seem to happen very often and the general approach for calculating D seems a reasonable approximation.

1.5.3 The Hydrogen Bond Term

E_{HB} is a hydrogen bond term [5–7] which is used only when one of the interacting atoms can donate a hydrogen bond and the other can accept. Equation (2) is again used but the constants A and B now have values which depend on the chemical nature of the interacting atoms, and the function F depends on their hybridisation and the relative positions of the interacting atoms and their bonded neighbors. E_{HB} and E_{LJ} both define relatively short-range effects, and are set to zero if the interacting atoms are more than a few Angstroms apart.

1.5.4 The Other Terms

The E_{LJ}, E_Q and E_{HB} functions are very simple, but they are also very well known which gives them one particularly important advantage: everybody understands them and can judge and criticise them for themselves. Moreover GRID displays by default the individual E_{LJ}, E_Q and E_{HB} and dielectric D values for every pairwise interaction, and so the source of any suspected error can usually be discovered very easily. After careful analysis it may then turn out that GRID did not make the suspected error, and that the user's worries were misplaced. Irrespective of the final outcome, the user may gain an enhanced understanding of the system by checking things like this for himself, and this enhanced comprehension was a major objective when the program GRID was first being written.

As mentioned above, there is an ongoing policy to search continually for the worst features in the current release of GRID, and then to take account of them. Later versions of the programme therefore include many extra terms which were not present in the original E_{PAIR} function. For instance:

1. When there is a rather close contact between a target and a probe atom, the computed E_{LJ} value may be strongly positive, suggesting mutual repulsion. However, if E_{HB} is simultaneously negative the atoms may actually be close together because they are making a hydrogen bond to each other, and GRID must detect when this happens and then allow E_{HB} to override E_{LJ}.
2. An adjustment must be made for the effect of an atom's charge upon the strength of the hydrogen bonds which it makes. For example hydrogen atoms bonded to an aliphatic carbon do not normally participate in hydro-

gen bonding. However GRID must take special account of the alpha carbon atom at the N-terminal of a protein chain, because this carbon sometimes donates a hydrogen bond as it can pick up positive charge from the nearby cationic N-terminal nitrogen.

3. Metals now receive special attention in GRID according to their hardness or softness.
4. Some water molecules in a biological system appear to make four tetrahedral sp3 hydrogen bonds. Others donate two hydrogen bonds but accept only one, making these three interactions in roughly the same plane. GRID must therefore be able to deal with both the flat and the tetrahedral arrangements.
5. The input programme GRIN always checks the overall electrostatic charge of the target, and expects nucleic acids to be surrounded by a cloud of counterions which maintain overall electroneutrality. However the ions were not mobile in early releases of GRID, and GRID maps of DNA were therefore full of holes which surrounded each counterion. They looked rather like a Swiss cheese, but GRID Probes can now nudge the counterions out of the way and thus generate a GRID map without misleading holes.
6. Some water molecules near a target may be so strongly bound that they almost behave like a part of the target itself. GRID must therefore treat each of these waters in a way which depends on its particular environment. For example, a water already bound to two carboxy groups would normally be donating a hydrogen bond to each carboxy oxygen, and would be much more likely to accept a hydrogen bond from the probe than to donate. On the other hand, a water already accepting from a couple of arginine guanidinium NH_2 groups would be most likely to donate to the probe. GRID must therefore be able to examine each water of the target and take its local environment into account.
7. The force field now incorporates entropy terms. For instance a lysine side chain of the target can adopt only one or two conformations when fully extended towards a distant probe, but can assume more conformations to reach a probe which is nearer, and GRID must be able to allow for this.
8. There is now a "hydrophobic probe" which detects hydrophobic regions on the surface of the target, and this probe must also take account of entropy.

The list of interesting special cases could be extended indefinitely. GRID will deal automatically with some of them, and directives are provided so the user can decide how to deal with others.

1.6 Nomenclature

Several different types of oxygen atom have already been mentioned including sp2 carbonyl oxygen, sp2 carboxyl oxygen, sp2 hydroxy, sp3 hydroxy, phosphate oxygens, aliphatic ether oxygen and furan oxygens. All of these have the same chemical symbol O, but each type of oxygen has its own specific properties. One needs to have a straightforward list of this detailed information for each kind of atom, showing all its characteristics on one line across the page or visual display screen. In order to reduce the amount of data on each line we therefore decided to give each atom a "*Type*" number which would specify its hybridisation and other electronic properties. Carbonyl oxygen for example, is a Type 8 oxygen.

Many different ways of tabulating the necessary information have been proposed by scientists working in various international agencies, national laboratories, universities and companies. The system adopted for GRID is based on the methods of the Protein Data Bank (PDB). Their nomenclature was agreed after extensive international discussions between workers in many fields, and it divides atoms into two distinct categories: "ATOM" and "HETATM".

1.6.1 "ATOM" Records

ATOM records are used to specify molecules which occur frequently in biological systems. These are called the "known molecules" and include amino acids, heme, cofactors, some of the "unnatural" amino acids used by medicinal chemists and a variety of molecules of general interest to GRID users. Here is a Protein Data Bank ATOM record:

```
ATOM    234 NZ  LYS   28   21.361  29.854  65.530   1.00 81.36
```

and some aspects of this record require a brief explanation:

1. The PDB nomenclature uses the first six characters on a line in order to define different kinds of record, and this is an "ATOM" record.
2. The PDB nomenclature defines the sequence in which the atoms of a protein must be specified, and this happens to be the 234th atom of its protein.
3. The abbreviated name of each amino acid is also defined, and LYS is the abbreviated name for this amino acid which is lysine.
4. The name of each atom of an amino acid is defined, and this is "nitrogen zeta" of the lysine. The abbreviated name for this atom is defined as NZ.
5. This lysine is the 28th amino acid residue along the protein chain.

The next three numbers are the x, y, and z coordinates of this particular atom, measured in Angstrom using orthogonal reference axes. They are essential input data for any MIF, but the last two numbers 1.00 and 81.36 which represent the

relative occupancy and isotropic temperature factor of the atom in the protein crystal, are not used by GRID.

A valuable characteristic of PDB format is that so much relevant information can be condensed into about 60 characters at the start of a line. The rest of the line can then be used by GRID for all the other data which are needed to specify the properties of an atom. The input program GRIN prepares all this data automatically, and writes it at the end of the line after the first 60 characters.

1.6.2 "HETATM" Records

Of course GRID users also need to study all sorts of molecules as well as proteins, and PDB format provides "HETATM" records as an easy way for doing this. If the user wanted for some special reason to define his protein structure using HETATM records, the same atom in the same molecule would appear like this :

```
HETATM   25 N3+ MOL    1    21.361   29.854   65.530   1.00 81.36
```

Notice the similarities and differences between the ATOM and HETATM records:

1. All the numbers and symbols are lined up in the same columns as before, but the record now begins with the word HETATM instead of ATOM. This name HETATM is an abbreviation for "heteroatom".
2. The protein is no longer being treated as a string of amino acids, but as a single molecule which can be given any molecule number (in this case 1).
3. The molecule is now called MOL instead of using the specific name LYS which was required by PDB format for lysine in a protein. GRID can accept any three-letter name for a molecule when HETATMs are being used, except names such as LYS which are reserved for "known molecules".
4. The sequence of ATOM records in a molecule is specified by PDB format. However HETATM records can be listed in any sequence, and the nitrogen has been moved to the 25th row of the new HETATM file. There is nothing special about its new position, and the user could just as easily have moved it to the first or last row of the file, or left it where it was.
5. There is a convention for HETATM names in GRID. They indicate the structure of the atom, so this N3+ nitrogen is a HETATM with three bonded hydrogens and it is positively charged as indicated by the + sign. Note in particular that 3 is the count of bonded hydrogens, and not the hybridisation which by coincidence is also sp3 in this case. Hybridisation and other electronic properties are defined by the type of an atom which is determined by the input programme GRIN when it prepares the data.
6. The x, y, z coordinates are unchanged, and the last two numbers 1.00 and 81.36 have not been altered. They will be overwritten by GRIN when it prepares the input data.

1.7 Calibrating the GRID Force Field

It is often convenient to think of drugs and proteins in terms of their chemical formulae and three-dimensional structures. However, an alternative interpretation is to regard the structure as nothing more than a set of frictionless rods and levers which transmit forces from one part of the system to another. This is the philosophy which underlies GRID, and it puts the main emphasis onto thermodynamics rather than structure. However it does raise a number of problems:

1. When the thermodynamic viewpoint has been adopted, it is the free energy of the system rather than the chemical structure of the molecules which needs the most careful study. Free energies can be most conveniently computed for reversible equilibria, and so the results from GRID should apply, strictly speaking, only to equilibrium systems. GRID has been found in practice to give useful predictions [8–11], but it is not easy to estimate the size of any errors caused by deviations from equilibrium.
2. Some free energy changes in biology are almost vanishingly small, while others may be greater by several orders of magnitude. The biggest changes often correspond to covalent reactions which break the "rods and levers", and these can completely swamp the weaker effects. We therefore decided to study only the ground state at body temperature, and so the GRID force field is not applicable to ligands which bind covalently to their receptor. In many cases this may just be another way of saying that GRID predictions are restricted to reversible equilibrium systems.
3. It is the differences and not the similarities between one drug molecule and another which are important, and the calibration of the GRID force field must be sensitive enough to differentiate between similar yet different atoms. For instance it would have been easy to assign the same parameters to the oxygen of an aliphatic ether and the oxygen of furan, but GRID would not have been able to differentiate between those two kinds of oxygen atom if this had been done. We therefore decided *not* to restrict the number of atom types in the force field, and we always welcome suggestions from GRID users, although the calibration of a new atom type is a nontrivial job which may take some considerable time. However, more than 10 different types of oxygen atom and 20 types of nitrogen have now been calibrated for GRID as a result of this policy.
4. One of the earliest decisions was to calibrate the GRID force field whenever possible by using experimental measurements rather than theoretical computations, and calorimetric measurements were therefore needed for the initial calibration in order to differentiate the enthalpic and entropic contributions to the overall free energy. However, only a very little calorimetric data was readily available at that time, about well characterised biological systems in which the structures of the interacting ligand and macromolecule were both known, and so a different approach was initially needed.

Fortunately several other kinds of experimental data were available for calibrating the GRID force field. Crystallographic measurements provided values for the Van der Waals radii of many atoms in all sorts of molecules, and corresponding but shorter radii were estimated for atoms making hydrogen bonds to each other. Many experimental determinations of atomic polarizabilities have been reported, and these were used together with the number of outer-orbital electrons in an atom to predict its Lennard-Jones interaction energy E_{LJ}. The observed structure of a molecule allows one to determine the bond order and hybridisation of its atoms, and hence to predict the maximum number of hydrogen bonds which each atom can donate and accept. Atomic charges can, in principle, be deduced from accurate X-ray data, but relatively few X-ray observations are precise enough for this job and theoretical methods were therefore used to estimate atomic charges. This only left the hydrogen-bond strength as an undetermined variable to be fitted to the observations.

All the necessary data was collected together in a file called GRUB which is revised whenever a new version of the program GRID is released. The first part of the GRUB file contains data values for ATOMS in known molecules (The natural amino acids, heme, cofactors, etc.). For example, there is an entry for the NZ ATOM of lysine. The second part of GRUB has individual HETATM values, and so it has an entry for N3+.

1.7.1 Checking the Calibration

Very many crystallographic observations have been reported on ligands bound to macromolecules. The structure of these complexes is usually measured to within a fraction of an Angstrom, and the formation of the crystals is determined by free energy. We therefore decided to use these readily available, crystallographically observed, ligand-macromolecule structures in order to check and refine the GRID force field after the initial calibration. In the absence of appropriate calorimetric measurements one cannot know whether enthalpy and entropy each make their appropriate contributions to the overall energy values computed by GRID. However as crystal structures were used to check and refine the force field, it seemed reasonable to hope that GRID would be able to predict the location of favorable binding sites, and this is indeed the case [8,11].

1.7.2 Checking Datafile GRUB

Users may often want to edit their copy of GRUB, or copy it from one directory to another, and of course mistakes may be made. It is therefore easy for errors to creep into the stored parameters, and so various kinds of check are made:

1. The input programme GRIN always analyses the input data before each new GRID computation. It checks both the structure of the target and the integrity of datafile GRUB, and warns the user about any doubtful features

in either file. In particular, it always reports the overall electrostatic charge of the target, because MIF computations can give very misleading results if the total charge of the whole system is significantly different from zero.

2. We frequently check GRID maps prepared from high resolution X-ray structures, because significant calibration errors in datafile GRUB would cause a systematic bias in the output. Many mistakes in the datafile were corrected in this way when GRUB was first being prepared, but such changes are not required so often now.
3. Regular users would quickly detect errors in the output from GRID by visual inspection of their maps, and we have had valuable feedback from users for many years. They would let us know very quickly if probes were being systematically predicted in the wrong place. In the 1986 release of GRUB, for example, something was wrong with the amino acid histidine and it was a great help to learn about this from a user.
4. Another kind of check comes from people who use the results generated by GRID as input data for further computations. For instance, we learned in 1997 that GRID was giving statistically biased results for compounds which contained acetylenic carbon atoms, and we had not been aware of this until a GRID user informed us. A reappraisal showed that relatively little information about acetylenes had been available when the GRID force field was first being parametrised, and a slight adjustment brought acetylene into line with the rest of the calibration data once we knew about the bias. This shows that the statistical analysis of GRID results can make an important contribution to the improvement of the force field.

Particular emphasis must be placed on the importance of checking all the input data before beginning any MIF computation. Of course error checking is not a satisfying job, but more problems seem to occur because of input errors than for any other reason. Program GRIN always checks the input thoroughly by default, and this helps to diminish the workload, but some users ignore error messages or try to save a little time by altering directive LEVL to a low value which turns off checking altogether. This is not recommended, and for our own research we never set LEVL in programme GRIN below the default value of 3.

1.8 The Output from GRID

The GRID method was explicitly designed in order to get selective information about binding sites, and the output can be used in two quite distinct ways:

1. To prepare GRID maps which are intuitively easy to understand, and can therefore provide a focal point for discussions between people with backgrounds in different fields of science, and indeed for people with little formal scientific training.
2. To generate matrices of numerical data which can be analysed statistically.

This article is not the place in which to consider statistical methods in detail, but the use of GRID maps to interpret interesting features of molecular structures will now be described.

1.8.1 GRID Maps from Macromolecules

Figure 1.5 shows an amphipathic alpha helix whose structure as part of a large globular protein was observed by X-ray crystallography. No hydrogens are displayed, and the helix has been separated from the rest of the protein in order to have an uncluttered figure. The side of the helix which faces towards the bottom of the page contains alanine, leucine and similar hydrophobic amino acids. The opposite side of the helix (the top side in the figure) has a prominent lysine side chain and other polar groups. This type of amphipathic helix has often been observed floating on the outside surface of globular proteins, with its hydrophobic amino acids facing towards the globular centre and its polar side chains in the surrounding water phase.

The flexible side chain of the lysine (CH_2–CH_2–CH_2–CH_2–NH_3^+) is displayed in an all-trans conformation in Fig. 1.5 because it was assigned all-trans coordinates by the X-ray crystallographer. An all-trans structure like this is often reported when the atoms of a side chain are in such vigorous dynamic thermal motion that they cannot be detected by X-ray methods. Arbitrary all-trans coordinates are then assigned by default, because the crystallographer knows that the amino acid is lysine from the DNA sequence although he cannot observe the side chain atoms himself.

The GRID map in Fig. 1.5 was deliberately prepared in order to demonstrate how easy it is to obtain misleading results when inappropriate directives are thoughtlessly used for an MIF computation. GRID would never normally generate such a deceptive map, and a special set up was needed in order to force it to prepare Fig. 1.5 at all. The blue sphere marks the terminal N3+ group of the lysine side chain, and GRID was deliberately used on the implausible and unrealistic assumption that the helix and its side chains were all completely rigid. There are three energy minima (colored red in the figure) corresponding to the three hydrogen atoms of the cationic nitrogen, and these minima misleadingly suggest that an incoming ligand would be able to make particularly favorable interactions in these three highly localised positions. This must be an incorrect conclusion if the side chain is actually sweeping backwards and forwards in vigorous motion across a relatively wide region.

Figure 1.6 shows what happens when slightly more appropriate settings are used for the GRID run. Torsional rotation of the sp3 amino group is now allowed round the terminal CH_2–NH_3^+ bond of the side chain, and so a halo is generated. The halo in Fig. 1.6 is not a uniform ring because there would be eclipsing between the CH_2 and NH_3^+ hydrogens at some torsion angles, and eclipsing would be energetically unfavorable. This description of the NH_3^+ interactions might be reasonable if the methylene groups of the side chain were buried within the bulk

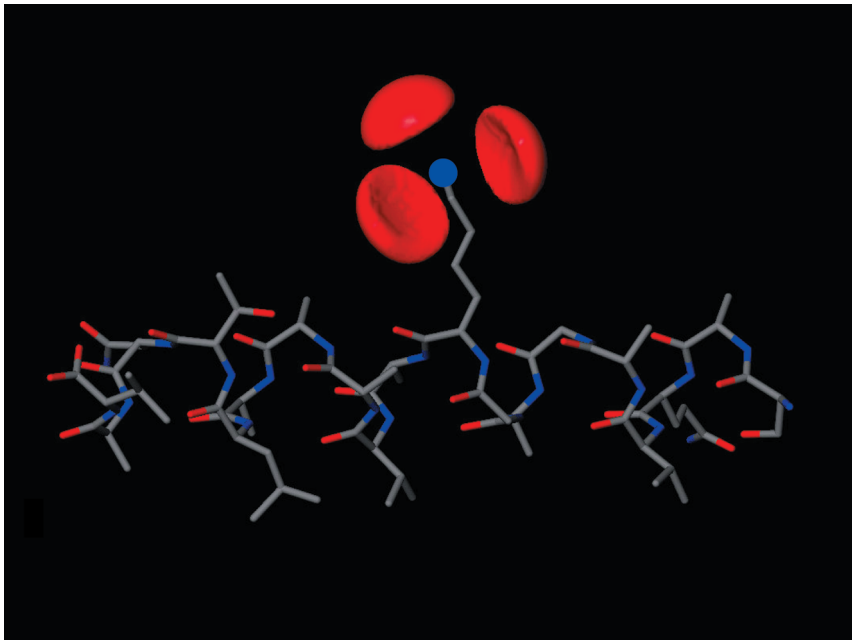

Figure 1.5. An alpha helix with a lysine side chain. The terminal NH_3^+ group of the lysine is marked with a blue sphere. GRID was deliberately misused to prepare this figure on the unrealistic assumption that the helix and its side chains were all completely rigid. This GRID map is therefore misleading, and this figure demonstrates how important it is to use MIF programs with great care because it is very easy to obtain deceptive results by misusing any MIF program. See text.

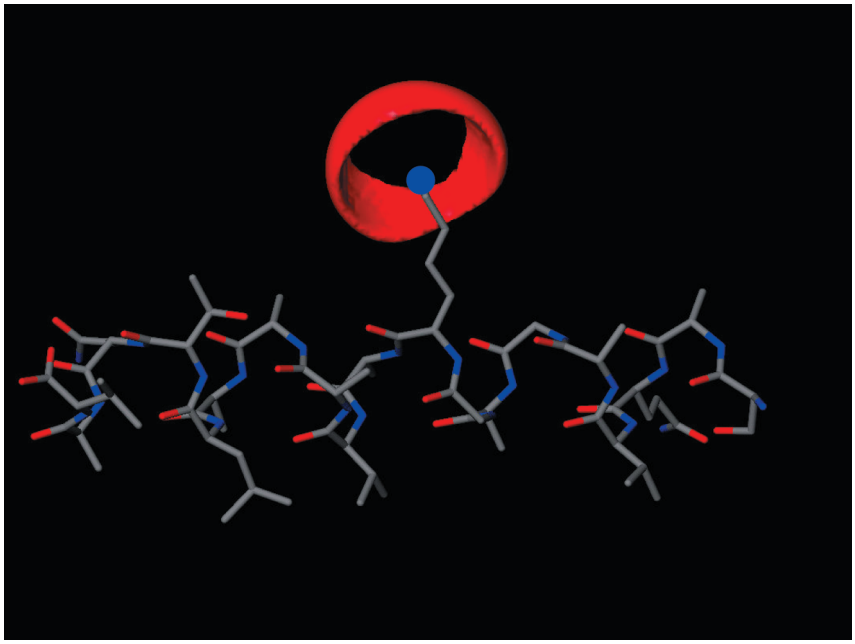

Figure 1.6. The same helix, but torsional rotation of the terminal side chain bond is now permitted, and this map is slightly more realistic than Fig. 1.5. See text.

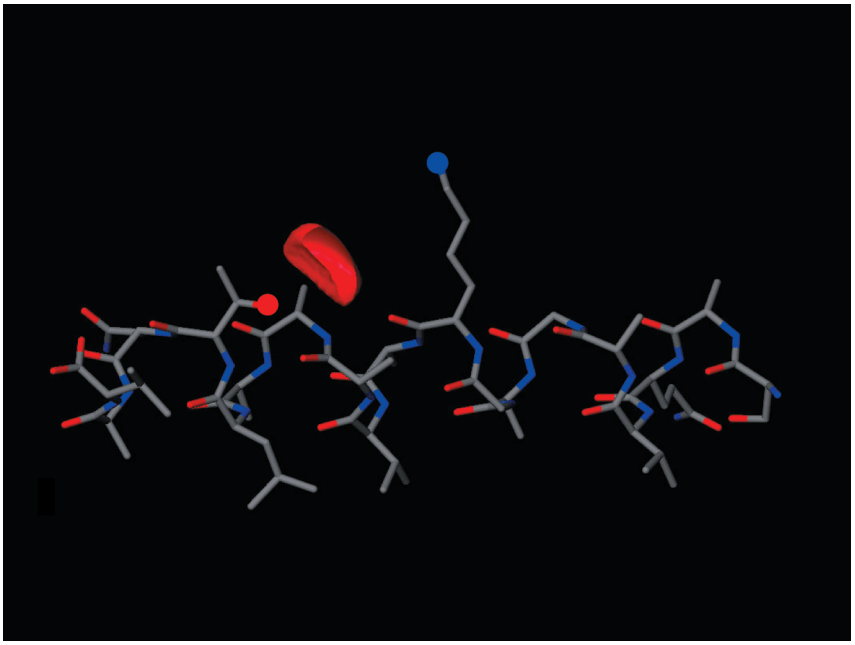

Figure 1.7. The same helix when the whole side chain is allowed to move freely. GRID now detects a favorable binding site where the hydroxy group (red sphere) of a threonine side chain and the terminal NH_3^+ group (blue sphere) of the lysine can both interact simultaneously with the probe. See text.

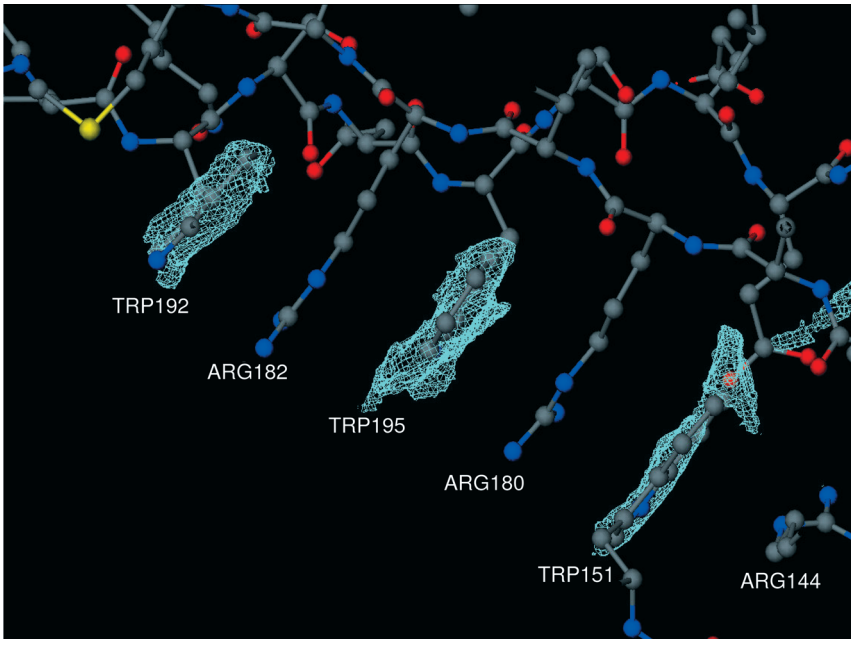

Figure 1.8. Some residues near the surface of another protein. GRID is used to elucidate why the polar arginine and nonpolar tryptophan side chains pack so closely together. See text.

of the protein, as they often are in some proteins because methylene is a hydrophobic moiety. Movement of the buried methylenes would then be restricted, and the most significant torsional rotation in the lysine side chain might be around the CH_2–NH_3^+ bond. However, the methylene groups of the lysine in Fig. 1.6 are not deeply buried, and the side chain with its nitrogen is actually free to move over a wide region, so Fig. 1.6 is nearly as misleading as Fig. 1.5.

Programs GRIN and GRID always search the target for any small parts which can move freely. However the user must set a dedicated directive (called MOVE) if he wants this feature to be used in his computations. This directive never allows total flexibility because the whole structure of the target might then unravel. Domains are always treated as rigid units, and when MOVE has been set the helix backbone is still treated as a rigid domain. However side chains can now move [12], and the resulting GRID map is shown in Fig. 1.7.

A small red sphere now marks the hydroxy group of a nearby threonine which is on the same side of the helix as the NH_3^+ of the lysine, and directive MOVE alerts GRID to the proximity of these two polar groups. GRID then tests whether the flexible lysine side chain could reach far enough in the direction of the threonine, so that a probe might be able to interact with both the NH_3^+ and the threonine hydroxy group at the same time. In this example the geometry is acceptable, but the torsional flexibility of the lysine would be restrained by its interaction with the probe, and an entropic allowance must be made for this. However the enthalpic benefit of two good hydrogen bonds outweighs any entropic penalty for torsional restraint, and so this is a particularly appropriate place for an incoming ligand.

The red region in Fig. 1.7 shows where a probe would be located when interacting with both the threonine hydroxy and the lysine NH_3^+ groups. However, it was necessary to contour this GRID map at a slightly more negative energy level than Fig. 1.6 in order to show the result clearly, and Fig. 1.7 was therefore contoured at an energy roughly corresponding to a pair of hydrogen bonds. At this energy level there is no blurring due to the weaker interactions which the NH_3^+ group would make as it searched through wide regions alone at the end of its side chain, and the absolute minimum near the threonine is unmistakable in the GRID map.

A completely different application of GRID is illustrated in Fig. 1.8 which shows several amino acids in the cytokine binding region of the human protein gp130. The surface of the protein faces towards the bottom left corner of the figure, and there is a sandwich structure in this part of the macromolecule where alternate tryptophan (TRP) and arginine (ARG) side chains lie one above each other like slices in a loaf of bread. The close-packed TRP–ARG relationship is unexpected because arginine is one of the most polar amino acids, while the side chain of tryptophan consists almost entirely of nonpolar hydrocarbon groups.

GRID maps suggest an interpretation of this structure. Tryptophan has two aromatic rings with nine CH and CH_2 groups but only one nitrogen which can make one hydrogen bond and no more. Arginine is very polar because it has a permanent cationic charge and three nitrogen atoms in a guanidinium group which can donate up to five hydrogen bonds. The cationic charge of guanidinium tends to

increase the strength of its donated bonds, but arginine would never accept a hydrogen bond from a tryptophan side chain. Moreover a glance at Fig. 1.8 shows that the observed geometry does not permit the hydrogen bonding of either ARG180 or ARG182 to the nearby tryptophan rings, and it is most surprising that such very polar residues should be squashed between the hydrophobic tryptophans in such an apparently unfavorable position. Arginines also have long flexible side chains which are frequently found in vigorous motion, like the lysine in Fig. 1.7, but the X-ray findings from gp130 show that the side chain methylene groups of ARG180 and ARG182 are not moving much more than the adjacent main-chain alpha-carbon atoms. This is another surprising feature of the observed structure, because there must be an entropic penalty when these flexible side chains are so firmly pinned down.

The PDB structure for this protein (PDB Reference: 1BQU) was therefore edited to remove the side chains of TRP192, TRP195 and TRP151, and thus make way for GRID probes to explore the volume normally occupied by these bulky moieties. The edited file was then used to prepare a target for GRID, and the contours in Fig 1.8 were generated using the hydrophobic probe on this target. The tryptophan side chains are shown in their observed positions (although the map itself was generated *when they were absent*), and the contours show that GRID predicts a hydrophobic region roughly surrounding each tryptophan ring. Further examination shows that the extended methylene chains of arginines 180 and 182 are almost ideally arranged for making hydrophobic interactions with the tryptophans, although 144 has a less favorable crumpled conformation.

However, it is not only the arginine methylene groups which generated the hydrophobic contours. GRID also predicts that the top and bottom faces of the arginine guanidinium groups can make favorable hydrophobic interactions, because the hydrogen bonds of guanidinium are so very firmly constrained to the plane of the guanidinium system. This is particularly well shown by the contours surrounding TRP195, which extend well beyond the reach of the methylene groups in the side chains of ARG180 and ARG182.

One must conclude that hydrophobic interactions may stabilise the multilayer TRP–ARG sandwich of gp130, in spite of the different character of these two amino acids, and in spite of the entropic penalty mentioned above. However the gp130 crystals themselves came from a solution which contained glycerol molecules and sulfate ions, and both of these components were trapped in the crystals where they may have helped to stabilise the observed protein structure. It would not be altogether surprising if some alternative conformation or conformations of gp130 may also occur *in vivo*, if those somewhat unphysiological substances are not present in the human body in sufficient concentrations to stabilise the structure as observed.

1.8.2
GRID Maps from a Small Molecule

Leucine is an amino acid, and is one of the building blocks of proteins. It has a nitrogen atom which is shown as a blue sphere in Fig. 1.9, a pair of oxygens both shown in red, and a small cluster of hydrophobic groups shown as yellow spheres towards the left of the figure. Hydrogen atoms are not displayed in order to keep the picture as clear and simple as possible, but the N3+ nitrogen (blue) has three bonded hydrogens and is therefore cationic. The carboxy oxygens have no bonded hydrogens and are negatively charged, so the molecule taken as a whole is electrically neutral.

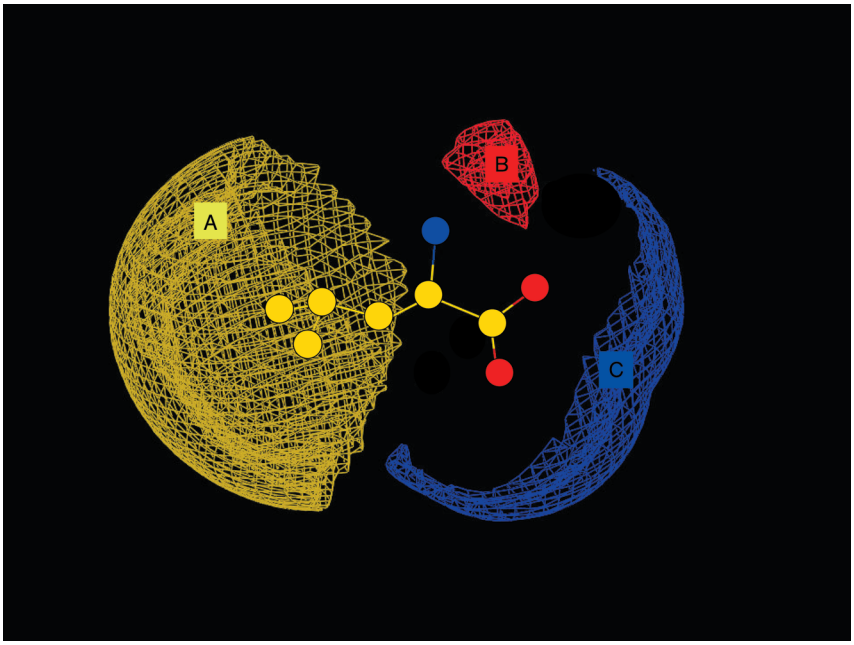

Figure 1.9. A molecule of leucine with GRID maps for a hydrophobic probe (A, yellow); a multiatom cis-amide probe (B, red); and an sp3 NH_3^+ probe (C, blue). See text.

This very simple target was chosen in order to demonstrate the selectivity of the GRID method. The yellow contours (A) were generated using the hydrophobic probe, and they show that one part of the amino acid is nonpolar and very hydrophobic. Binding clefts on biological macromolecules often expose a hydrophobic surface, and it is very important to detect the hydrophobic surfaces of ligands if one wishes to design high affinity molecules.

The blue contours (C) in Fig. 1.9 were generated by the N3+ cationic amine probe which makes good hydrogen bonds to the (red) carboxy oxygen atoms of the target. When generating these blue contours GRID takes account of the fact that

both the interacting atoms are charged, and can therefore make a particularly strong hydrogen bond to each other. GRID also takes account of the local dielectric, and the electrostatic attraction between the target's oxygens and the cationic probe is therefore attenuated towards the right of the figure, because this is where the probe would be most exposed to the higher dielectric of the surrounding bulk water.

It is important to give careful consideration to apparently small details in a GRID map. The carboxy group in Fig. 1.9 may appear to be symmetrical, but the blue contours are stronger round the oxygen at the bottom of the figure. This difference between the oxygens may be caused by two quite distinct influences:

1. The upper oxygen is closer to the cationic nitrogen N3+ of the leucine, and so the cationic N3+ probe may experience an unfavorable electrostatic repulsion when it is close to the leucine nitrogen's cationic charge.
2. The bottom oxygen in the figure is partly shielded from bulk water by the hydrophobic moiety (yellow spheres) of the target, and so its dielectric environment may tend to favor the electrostatic attraction of a cationic probe.

The red contours (B) at the top of Fig. 1.9 were generated by a multiatom amide probe CO.NH which was arranged cis so that its hydrogen and oxygen are both on the same side of the CN axis. This multiatom probe therefore detects regions where it can donate a hydrogen bond from its nitrogen and can accept at its carbonyl oxygen. It sits between the N3+ group and a carboxy oxygen of the target, and is at a slightly awkward angle because the hydrogen bonding atoms of the target do not line up perfectly with those of the multiatom probe.

There are many other selective probes which can be used to elucidate the properties of a target. For instance the "amphipathic probe" finds boundary surfaces where a part of the target with polar characteristics touches neighboring hydrophobic regions. It would draw attention to a boundary of this type which runs up the middle of Fig. 1.9 where it separates the polar nitrogen and oxygen atoms on the right from the hydrophobic carbons towards the left of the figure. Each additional probe provides qualitatively different information, and competing research groups may reach interestingly different conclusions when studying the same set of molecules simply because they select different probes for their investigations.

1.9 Conclusions

The design of molecular interaction fields has been a matter of continuously improving approximations, and no force field is perfect. It is therefore critically important to choose the right MIF for each particular job, and the GRID force field was explicitly designed for use with the GRID method. This is an approach which generates selective information about binding sites on proteins, therapeutic agents and other important biological molecules of known structure. The output is intuitively easy to understand, and can provide a focal point for discussions be-

tween people with backgrounds in different fields of science. Results from GRID can also be analysed statistically.

In this chapter some emphasis has been placed on the subjective influences which can often modify the results of force field computations, and examples of subjective decision making have been provided. However the need to make decisions can enhance the GRID user's intuitive understanding of noncovalent interactions between molecules, and this enhancement was a prime objective when the GRID method was first being devised. Figures 1.5, 1.6 and 1.7 show how important it is to have a proper understanding of the system under investigation, if one wants to obtain meaningful results.

The GRID force field has always been calibrated as far as possible by studying experimental measurements, and the calibration is then checked by studying how well GRID predicts observed crystal structures. Crystal packing is determined by free energy considerations rather than by enthalpy alone, and recent versions of the force field include entropic terms. GRID can detect the hydrophobic binding regions which are so important when high-affinity ligands are being designed, and it can also detect sites for the polar groups which determine ligand selectivity. The GRID method is being systematically extended, and new versions are issued from time to time.

Acknowledgment

To print a list of the teachers and colleagues who have helped me would occupy too much space here. My debt to all of them is immense, and is only matched by my gratitude to each one for their generosity, and to my wife for all her support and encouragement.

References

1 M.L. Huggins, L.Pauling *Z. Kryst.* **1934**, *87*, 205–238.
2 P. J. Goodford *J. Med. Chem.* **1985**, *28*, 849–857.
3 G. Cruciani, K. A. Watson *J. Med. Chem.* **1994**, *37*, 2589–2601.
4 L. D. Landau, E. M. Lifshitz *Course of Theoretical Physics*, Englis (ed.), Pergamon Press, Oxford, **1960**, Vol. 8.
5 D. N. A. Boobbyer, P. J. Goodford, P. M. McWhinnie, R. C. Wade *J. Med. Chem.* **1989**, *36*, 1083.
6 R. C. Wade, J. Clark, P. J. Goodford *J. Med. Chem.* **1993**, *36*, 140–147.
7 R. C. Wade, P. J. Goodford *J. Med. Chem.* **1993**, *36*, 148–156.
8 M. von Itzstein et al. *Nature (London)* **1993**, *363*, 418–423.
9 P. J. Goodford *J. Chemometrics* **1996**, *10*, 107–117.
10 T. Langer *Quant. Struct.-Act. Relat.* **1996**, *15*, 469–474.
11 A. Berglund; M. C. De Rosa, S. Wold *J. Comput-Aided. Mol. Design* **1997**, *11*, 601–612.
12 P. J. Goodford **1998** Rational Molecular Design in Drug Research, in *Alfred Benzon Symposium 42*, Munksgaard, Copenhagen, Liljefors (ed.), **1998**, 215–226.