MEETING REPORT

Meeting on binding sites: Characterizing and satisfying steric and chemical restraints. University of York, 28–30 March 1993

It seems intuitively sensible that if the structure of a protein is known, then it should be relatively easy to design small molecules to fit into a binding site in that protein. Experience has shown us that this is not a straightforward process. Initial algorithms attempted only to fit ligands sterically to a rigid protein. Work has progressed significantly in this area in the past few years, and the Molecular Graphics Society felt it was timely to try to bring together the various groups that have contributed to this progress to share their ideas and propagate discussion. The principal interest from a commercial point of view comes from the pharmaceutical industry which is trying to design ligands to fit into protein binding sites for their eventual use as drugs. Indeed, over 40% of the attendees came from that industry, with a similar number from academia. Over 130 people from 9 countries attended what turned out to be an extremely exciting and thought-provoking meeting.

Peter Goodford (University of Oxford) opened the meeting with an update of his GRID program, which defines the best place to position probe atoms in a binding site. It can also be used around a small molecule to indicate the geometry of interactions that might be expected. When used to probe a static macromolecule, contours are generated to help visualize the preferred locations of the probe atoms. By using a modest estimate of the number of possible ways to describe just the electrostatic component of the interaction energy, Goodford produced a figure of 24,192 different options. He then went on to describe various applications of GRID, including some work on analyzing probe interactions with variable nucleotide triplets in DNA. He found, for example, that an amide probe splits the triplets into five clusters, using principal component analysis, and concluded that different probes can discriminate between different triplet sequences. His final comment emphasized that GRID does not address the entropic component of free energy.

Martin Karplus (Harvard University) described his protocol for designing ligands, the first step of which involved Multiple Copy Simulation Search (MCSS). This explores the protein for functional group binding sites by saturating the site with multiple copies of ligands. The protein feels a mean force exerted by the ligands whereas each ligand experiences only the presence of the protein. Optimal positions for the functional groups can be found in this way. The current methodology uses a static protein, but it is hoped that future work will incorporate a flexible protein by calculating a molecular dynamics (MD) trajectory. Candidate ligands can then be constructed using his Automatic Connecting and Clustering algorithm and minimized. Karplus exemplified this approach with the docking of a peptide inhibitor into HIV protease. Using N-methyl-acetamide as probe, he found positions for the peptide bonds of the inhibitor. Other probes were used to define side chain positions. The resultant peptide coordinates were made available to Alex Wlodawer, who reanalyzed his X-ray data and discovered a major occupancy orientation for the peptide that fitted Karplus' binding mode to within ~1Å root-mean-square (rms).

Where the structure of the macromolecular target is unknown, a pharmacophore has first to be determined. A fast new approach to pharmacophore mapping, DISCO, was described by ***Yvonne Martin*** (Abbott Laboratories). Clique-detection methods such as the Bron–Kerbosh algorithm as implemented by Brint and Willett are used to define a bioactive conformation and supply a superposition rule for each active compound. All low energy conformations are generated and optimized before ALADDIN is used at run time to define points, whether real atoms or receptor points, to be considered for superposition. When applied to four D2 agonists of different structure, DISCO took 10 seconds on a VAX 6310 to propose a pharmacophore. Typically, DISCO is run several times to compare alternative pharmacophore maps.

An approach was described by ***Andrew Leach*** (University of Southampton) which considers the six degrees of orientational freedom, the conformational flexibility of the ligand and the conformational degrees of freedom of the protein's side chains. Translational degrees of freedom are considered using a 2.5 Å grid with 22 orientations defined by equipartitioning the surface of a sphere. Conformational flexibility of the ligand and protein side chains is restricted to discrete conformational states; the ligand to previously calculated low energy conformations and each side chain to rotameric states extracted by analysis of the Brookhaven database. The number of possible combinations of all these

degrees of freedom can be very large (10^{40} or more). However, the sequential application of two searching algorithms (Dead-End Elimination and A*) allows the global energy minimum to be located, as well as structures within a specified energy difference. Leach used the example of phosphocholine in the McPC603 antibody to illustrate the method. Using a rigid receptor, he did not obtain a fit to experiment better than 4.4 Å rms, but using a flexible receptor, the fit improved to 2.2 Å rms.

Two presentations on the application of genetic algorithms to the design of novel ligands followed. In the first, ***Scott Dixon*** (SmithKline Beecham), ***Jeff Blaney*** (Chiron Corporation), and ***Dave Weininger*** (Daylight Chemical Information Systems) introduced the concept of genetic algorithms and suggested that evolution can be regarded as an optimization of a complex function. In the case of molecular design, they propose the chemical graph of the molecule as the genome. The four elements of genetic algorithms were discussed; reproduction, crossover, mutation, and scoring. SMILES strings were used to describe the molecular population, thus making it fairly easy to produce new molecules. Chemical "reasonableness" was checked, and initially the algorithm was evaluated by using similarity to a target molecule as the scoring function. A better score was rewarded by an increased probability of that molecule being represented in the next generation. The initial seed molecule contained two heavy atoms, e.g., methylamine, and within 40 generations the target dopamine was achieved. Another feature had to be included when considering evolving a molecule to fit an active site. Each new molecule had to be docked into the active site in order for the score (a representation of intermolecular energy) to be evaluated. The docking was accomplished using a combined distance geometry/ligand sphere-matching method into a rigid binding site. Dihydrofolate reductase (DHFR) was used as a test case, and initially, acyclic polyamines that fit sterically into every "nook and cranny" evolved. To take account of entropic factors, cyclic molecules were then scored higher, and examples of the sort of molecules that evolved were illustrated.

In the second presentation on genetic algorithms, ***Robbie Glen*** (Wellcome Research Laboratories) considered constraints other than just an enzyme active site. He included physical properties such as log P, and influenced cyclicity by using an atom-to-bond count ratio. Mutations such as element change, ring forming or breaking, fragment or ring addition, and methylene insertion were allowed, together with bond or molecule rotation. Glen used a fragment database to achieve some of the mutations described. DHFR was used to illustrate his algorithm, and a video recording of an evolution sequence was displayed. He showed plots of the scoring function being optimized over succeeding generations, and indicated that the evolution could start from simple molecules such as ethane or from a lead compound such as trimethoprim.

Hans-Joachim Bohm (BASF) described LUDI, a fragment-based method for building ligands into binding sites. Initially, probe interaction sites are defined in what the user considers to be the site of interest, using a cutoff radius. These sites can be objects such as hydrogen-bonding vectors or lipophilic points derived from an analysis of interactions in the Cambridge Structural Database. Fragments from a library containing ~1000 entries are then fitted to these sites. A scoring function that includes the number and quality of hydrogen bonds and the hydrophobic contact area between ligand and protein is calculated to help the user select fragments on which to focus for the next round. During this next phase, other fragments can be positioned in bonding orientations relative to the first ones in order to build molecules. Examples given included trypsin, where using a 4.5 Å cutoff, 900 fragments took 120 seconds on a Silicon Graphics 4D35, and HIV protease, where using a 6 Å cutoff, 1100 fragments took 80 seconds. Recent advances include the use of the Fine Chemicals Directory and improved methods for prioritization of the hit lists. In a cautionary remark, Bohm indicated that for HIV protease, LUDI was unable to identify the best P3′ substituent for peptide-like inhibitors due to the need for side chain adjustment to accommodate the ligand.

A different method of growing ligands to fit a binding site was described by ***Andrew Miranker*** (Oxford Centre for Molecular Sciences), and uses ligand perturbation space (LPS). In this method, the binding site is saturated with functional groups, e.g., sp^3 carbons or acetate. An additional attribute is associated with the atomic properties of the functional group atoms, namely occupancy—a continuous function with values between 0 and 1. During a high-temperature Monte Carlo simulation, atoms in different fragments are allowed to bond only via carbon atoms, and thus attenuation of partial charges should not present a problem. Thus, using novel functions to describe intermolecular interactions, molecules can be grown within a binding site. Indeed, if a bond is formed, the atoms are encouraged to exist (occupancy increases). Miranker illustrated the method with the immunosuppressant FK506/ binding protein complex, for which an X-ray structure is available. Some peculiar geometries were obtained due to the omission of 1,4 interactions. The simulations were carried out in the presence and absence of the ligand in order to grow both optimized ligand structures and completely novel ones. Future work involves the incorporation of 1,4 interactions, the continuous perturbation of hybridization states, and the relief of strain in the molecules generated.

A procedure for growing flexible ligands into semiflexible binding sites was described by ***Mill Lambert*** (Glaxo Inc). It is a generalization of the Gibson and Scheraga build-up method that predicted the structures of oligopeptides. Most of the atoms in the protein are held fixed, and the surface area terms used to estimate the solvation free energies are omitted from the calculation until the final stages of the growth process. Side chains deemed to be flexible are grown one flexible bond at a time at each cycle of growth. The program searches for low-energy side chain conformations, and the lowest energy minima after energy minimization are kept for the next cycle of growth. When several hundred low-energy conformations of the protein, differing only in the conformations of certain side chains, are obtained by this procedure, a selected rigid functional group from the ligand is then docked into each of these structures by one of several methods. Again, the lowest energy structures, after energy minimization, for the functional group/protein complex are kept for the next cycle of growth. This involves adding ligand atoms to the complex one rotatable bond at a time. This is repeated until all the ligand atoms are in place. In the case of porcine phos-

pholipase A2 bound to a phospholipid inhibitor, the rigid functional group that was initially docked was the phosphate group. In total, 3449 Monte Carlo steps were used, followed by partial minimization and clustering. The rest of the inhibitor was grown in stages as indicated, keeping the best 100 structures at each stage. Structures where the growing methylene chain was extending into solvent were rejected as the chain got longer. In another example, 7300 benzamidines were initially docked around trypsin. In both cases, the eventual structure with the lowest energy turned out to be very similar to the X-ray structure.

Rene DesJarlais (SmithKline Beecham) gave a presentation illustrating a method of overcoming the restriction of using only shape complementarity in Kuntz's DOCK 1.0 program. Targeted-DOCK attempts to match certain atom types to selected positions in the binding site. The target sites can be chosen based on structural evidence, calculation, or inspection. The new algorithm is also faster. DesJarlais illustrated Targeted-DOCK with searches using thermolysin, both free (Brookhaven code 3tln) and complexed (5tmn). A speed-up factor of ~550 was obtained with 3tln, whereas 5tmn gave a factor of ~120. The 3tln search using DOCK 1.0 produced no correct binding modes (all were rotated through 180°), but Targeted-DOCK produced 2 correct ones. With 5tmn, DOCK 1.0 produced 2 correct positions, whereas Targeted-DOCK gave all correct ones. DesJarlais pointed out some of the limitations in the methodology, namely, that the search is nonexhaustive, single target points may not be sufficient, scoring is by shape alone, and both the ligand and the receptor are rigid.

Richard Lewis (Rhone-Poulenc Rorer) described a method of joining isolated functional groups that may have been positioned in a binding site that could also be used with a pharmacophore model. His program, TORSION, builds strings of chemical fragments linked by rotatable bonds. The values of the torsion angles about the rotatable bonds are found analytically if there are six or fewer torsions. The method requires user specification of the position and orientation of the chemical groups to be joined, but otherwise is fully automatic. The scope and duration of the search can be controlled by varying the number (up to three) and nature of chemical fragments used in the search. Fragments can be simple, such as sp^3 carbon, nitrogen, oxygen, or sulphur; extended, such as ortho, meta, and para disubstituted benzene; or "skew," such as the peptide group. Suggestions that arise from the program can be modified to reduce conformational energies, to enhance synthetic accessibility, and to incorporate other ideas that may improve the characteristics of the molecule.

The CLIX suite of programs for finding novel protein ligands was described by ***Peter Colman*** (CSIRO). It searches the Cambridge Structural Database (CSD) for small molecules that have both chemical and geometric complementarity to a defined binding site on a protein of known three-dimensional structure. The algorithm is described in five steps. First, all substituent chemical groups of all potential candidate molecules selected by the user from the CSD are identified. GRID is used to locate likely binding positions for a variety of chemical groups. Each candidate molecule is oriented into the binding site in such a way as to maximize the number of chemical groups overlaying with the sites identified by GRID. Each orientation is scored depending on the chemical and geometric complementarity to the binding site. Finally, the highest scoring orientations are inspected on a graphics workstation. CLIX succeeded in identifying, with high score, known ligands in their correct binding orientations in influenza virus neuraminidase and influenza virus haemagglutinin. For example, the rms fit to the crystal structure for sialic acid in neuraminidase was 0.4 Å.

Michael Eisen (Harvard University) presented a description of HOOK, a program for joining functional groups together in a binding site. The functional groups are positioned in the binding site using MCSS with 150–200 copies of each one. HOOK then uses a fragment database of ~50,000 rigid structures to join the functional groups. A scoring function is used to assess the linking groups, and the output is in the form of a spreadsheet. The method tends to be biased towards larger molecules, but this bias can be reduced by filtering the output based on size. The output can also be filtered on overlap score, and other properties such as accessible surface area are being incorporated. Eisen presented an application using chloramphenicol acetyl transferase.

Dick Cramer (Tripos Associates) was not on the original list of speakers, but there was so much interest in his poster on POSIT that he agreed to give a presentation at short notice. POSIT is a Monte Carlo method which makes use of a fragment library, although new molecules not composed of these fragments can also be generated. POSIT can be used with a receptor structure, a CoMFA, or any overlay of structures. It is written in Sybyl Programming Language and can be run in different modes, where existing molecules can be optimized or completely new ones created. It also supports an interactive design mode. Various moves are allowed which effectively mutate the ligand. Scoring is performed using van der Waals and Coulombic terms on a precomputed grid. In preliminary performance trials trying to optimize methotrexate in DHFR, dozens of novel and energetically reasonable structures are generated per hour.

Flemming Jorgensen (Royal Danish School of Pharmacy) described applications of the GRID, CLIX, and LUDI programs to the metalloenzymes phospholipase A2 and C, which catalyze the hydrolysis of phospholipids in different positions. Using GRID he found the most favorable interaction sites for probes resembling the functional groups in the phospholipid molecule. In this way, he proposed a binding orientation for the natural substrate, phosphatidylcholine, by placing phosphate, choline, and diacylglycerol in the positions indicated by GRID, aided by CLIX. Attempts to use LUDI were frustrated by problems in defining interaction sites to metals.

Joe Moon (The Upjohn Company) provided an update of an attempt to generalize his GROW program. The original program used 50,000 low-energy conformations of the 20 naturally occurring amino acids. A seed residue was docked into the binding site and other residues were grown sequentially, the lowest energy ones being kept at each stage. The scoring function uses a modified AMBER potential which considers van der Waals and Coulombic interactions, as well as internal strain energy and desolvation of both ligand and receptor. Examples involving Rhizopus pepsin and renin were used to illustrate the program. Moon described recent work to generalize the procedure for connecting frag-

ments together, enabling it to construct nonpeptide ligands. He uses a fragment library of ~120 organic fragments with the connectors defined. The increased size of the structure space considered required the use of a more efficient simulated annealing algorithm for the optimization of ligand fit. Currently, there is no capability for ring closure.

The final presentation of the meeting was given by ***Val Gillet*** (University of Leeds) on the SPROUT program. Target sites are first identified in the binding site by whatever method the user prefers. Then a set of carbon fragments is used to generate a skeleton for joining the target sites. Various fragment joining rules are in place and heuristics are used to score and prune the structures produced. The example used to illustrate the program was methotrexate in DHFR, where the target sites identified corresponded to atoms in methotrexate. A limitation of the method is that points are used as target sites, and so directionality in fitting to the binding site could be improved. Future developments in the program include the ability to substitute heteroatoms to improve interaction with the binding site, clustering of the output, and ranking of the output by synthetic feasibility.

In conclusion, the field of designing molecules to fit a binding site has developed significantly over the past few years. Earlier limitations that were highlighted, including the use of static ligand and protein, steric fits only, and the use of points to represent interaction sites are now being addressed. Completely different methods to tackle the problem are being developed such as MCSS, ligand perturbation space, and genetic algorithms, where the molecules that are suggested are not constrained by some sort of database. This is a rapidly expanding and very exciting area and the meeting reflected this.

It became obvious during the meeting that if you want to work in this area, you need crystal structures of DHFR! In addition to the formal talks, there was an excellent poster session which was complemented by demonstrations of some of the software discussed in the meeting, including LUDI and POSIT. The meeting was organized by the Molecular Graphics Society and special thanks are due to ***Rod Hubbard*** and his colleagues at York for putting together such a stimulating program.

Tony Slater and Dave Timms
Zeneca Pharmaceuticals, Alderley Park, Macclesfield Cheshire, SK10 4TG, UK