

4.1.6 AUTODOCK

AUTODOCK is a suite of C programs used to simulate interactions between small flexible ligands and macromolecules of known structure (Morris *et al.*, 1998). Docking is achieved through a search of conformational space using a Lamarckian genetic algorithm coupled with energy assessments using a method based on the AMBER force field. The combination of these two functions produces a family of molecular coordinates detailing possible docked ligand conformations which can then be used as a starting point for theoretical ligand design and study. Confidence in the docked conformation is represented by an energy value based on both quantum and molecular mechanical modelling of atomic forces.

4.1.7 Genetic Algorithms in AUTODOCK

Genetic algorithms represent a family of powerful mathematical functions derived from concepts based on the language of molecular genetics. They represent an effective method for the optimisation of systems with large numbers of related variables, and have recently supplanted older Monte Carlo simulated annealing algorithms traditionally used in *in silico* molecular binding experiments (Morris *et al.*, 1996). Using mathematical concepts designed to simulate the conditions influencing biological evolution, genetic algorithms are able to search conformational space by "mutating" a ligand in order to find its lowest energy conformation in the "environment" of a fixed protein. Searches driven by this energy funnelling have been shown to provide a good indication as to the optimum protein-ligand interaction and therefore the structure most likely to be found *in vivo* (Hetenyi *et al.* 2002).

The genetic algorithm used in AUTODOCK defines a ligand's "chromosome" as having seven standard genes accounting for the ligand's Cartesian coordinates, and four variables specifying its orientation. Covalent structure is kept intact however additional genes representing flexible torsion angles can be defined if needed. Once the genes have been defined the genetic algorithm starts by creating a population of random individuals confined within a user specified box also containing the protein. For each individual the three translation genes (x,y,z) are given a random value between the minimum and

maximum of the search area, the four genes describing the orientation given a random quaternion consisting of a unit vector and rotation angle, whilst the torsion angle genes (if any) are given random values between -180° and 180° . These gene values are then converted into a corresponding phenotype that enables the assessment of each individual's "fitness" measured by interactions both within the ligand and between the ligand and the protein. This assessment is followed by a selection procedure that decides which individuals will be allowed to progress into the next refinement round based on:

$$n_o = \frac{f_w - f_i}{f_w - \langle f \rangle} \quad f_w \neq \langle f \rangle$$

where n_o is the number of offspring to be allocated to the individual, f_i is the fitness of the individual (energy of the ligand), f_w is the fitness of the worst individual (highest energy individual) and $\langle f \rangle$ is the mean fitness of the population. As f_w will always be higher than both, f_i and $\langle f \rangle$ (except for when $f_w = f_i$), individuals with energy values lower than the mean will always have a larger numerator and hence be allowed at least one offspring. In this way the more successful individuals will be allowed to carry their "genes" into the next generation whilst the least successful will eventually become "extinct". If $f_w = \langle f \rangle$ then the population is assumed to have converged and docking is terminated.

Once reproducing individuals have been selected the next generation is created by two functions imitating biological "crossover" and "mutation" events. In the crossover event equivalent genes are swapped between a small set of randomly picked chromosomes where, for instance, a one point crossover between chromosome AB and ab result in chromosomes Ab and aB. Once this has been achieved a random mutation based on a Cauchy distribution is used to generate mutations on genes picked randomly from a further subset of chromosomes. Each generation consists of approximately fifty individuals whose genotypes are translated back into phenotypes for fitness evaluation. Both crossovers and mutations are characteristics of Darwinian evolution as applied to Mendelian genetics, where selection is performed on the phenotypic level based on the translated fitness of the available genotypes. The genetic algorithm utilised in

AUTODOCK also performs a limited local search based on the energy phenotype of each resulting chromosome, followed by reverse-transcription of the optimised phenotype back into the genome in much the same way as hypothesised by the discredited Lamarckian evolutionary theory. Translational step sizes of 0.2Å along with orientation and torsional step sizes of 5° are applied to a user defined proportion of each new generation, followed by an energy evaluation of the new conformation. If this conformation is evaluated as having lower interaction energy than its predecessor, the local search is repeated until the energy increases or a set number of steps have been performed (figure 4.5).

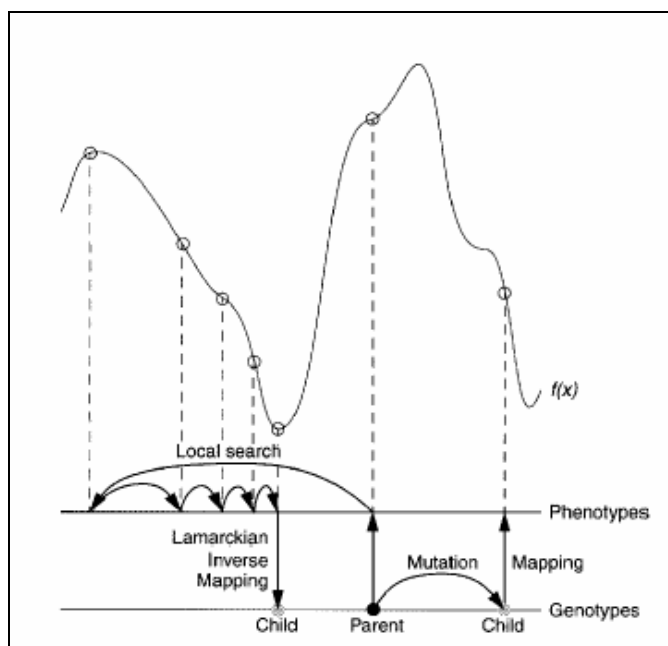


Figure 4.5 Comparison between Darwinian (right) and Lamarckian local search (LG-LS) (left) algorithms. In strict Darwinian system mutations perform a "blind" phenotypic jump that can either be accepted or rejected based on the phenotypic fitness. In the LG-LS algorithm inverse mapping allows a certain level of localised optimisation (Figure from Morris et al, 1998).

4.1.8 Energy assessments within AUTODOCK.

An efficient and accurate energy assessment of new ligand conformations is as important to the success of a docking simulation as the power of the search algorithm.

AUTODOCK uses a variation on the AMBER'95 force field (Cornell *et al*, 1995) with terms empirically determined by linear regression analysis from a set of protein-ligand complexes with known binding constants (Morris *et al*, 1996, 1998). Free energy (Gibbs, ΔG) is represented by a master equation, broken down into six terms to model dispersion/repulsion, hydrogen bonding, electrostatic interactions, deviation from covalent geometry, internal ligand torsional constraints, and desolvation effects:

$$\Delta G = \Delta G_{vdw} + \Delta G_{hbond} + \Delta G_{elec} + \Delta G_{conform} + \Delta G_{tor} + \Delta G_{sol}$$

Energy calculations are performed by placing the receptor protein inside a user defined three dimensional grid and allowing various probes to systematically visit every grid point. Each probe consists of an atom type found in the ligand whose dispersion/repulsion energies and, if appropriate, hydrogen-bonding energies are calculated for all receptor atoms within 8Å of each grid point. This information, along with electrostatic-potentials calculated using a probe with a single positive charge, is then stored in separate grid files which can later be sampled by the main auto-docking routine. Desolvation energy values are obtained by calculating burial of apolar atoms (hydrophobic effect), whilst conformation and torsion energies are calculated according to AMBER'95 force field terms. Final energy values predicted by these methods have been tested both by the developers of AUTODOCK (Morris *et al*, 1998) and by other groups (Wang *et al*, 2003), and found to be in accordance with experimentally derived values, although inevitable limitations are encountered when ligands contain more than about eight torsion angles.

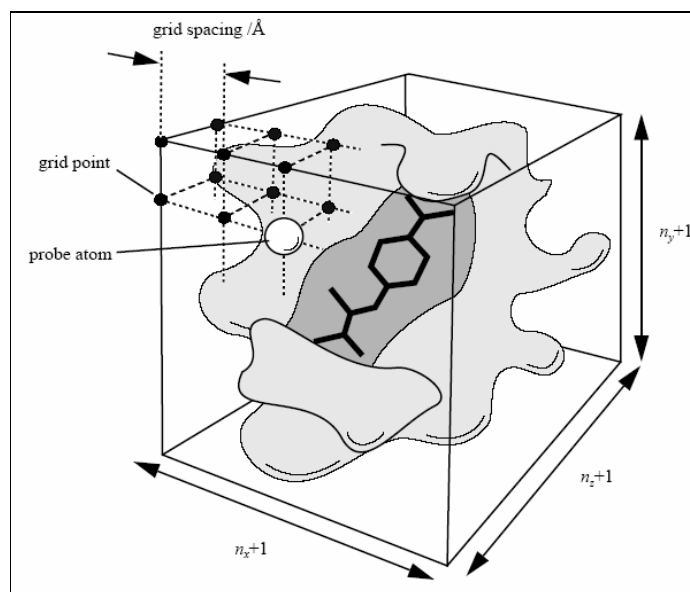


Figure 4.6 Diagrammatic representation of grids produced by AUTOGRID, a subroutine of AUTODOCK. Grids are constructed around the protein and probes corresponding to each atom type in the ligand are given energy values that can later be sampled by the main docking program if ligand atoms are found near these points (Morris 1998).

4.1.9 Autodocking with SAP

Docking was performed on a single SAP protomer from coordinates refined for subunit B of the 1.9 \AA acetate structure described above. This structure was chosen because the small acetate ligand in the binding site was thought less likely to interfere with the positions of surrounding amino acids and waters, although dockings with SAP protomers solved in the presence of larger ligands were also examined. Prior to docking polar hydrogen atoms were added to the structure and unnecessary water molecules deleted. An area of 36 \AA x 36 \AA x 36 \AA with grid point spacing of 0.375 \AA and grid centre just in front of the two calcium atoms, was used as the search space for the ligand and definition of energy grid maps (figure 4.7).

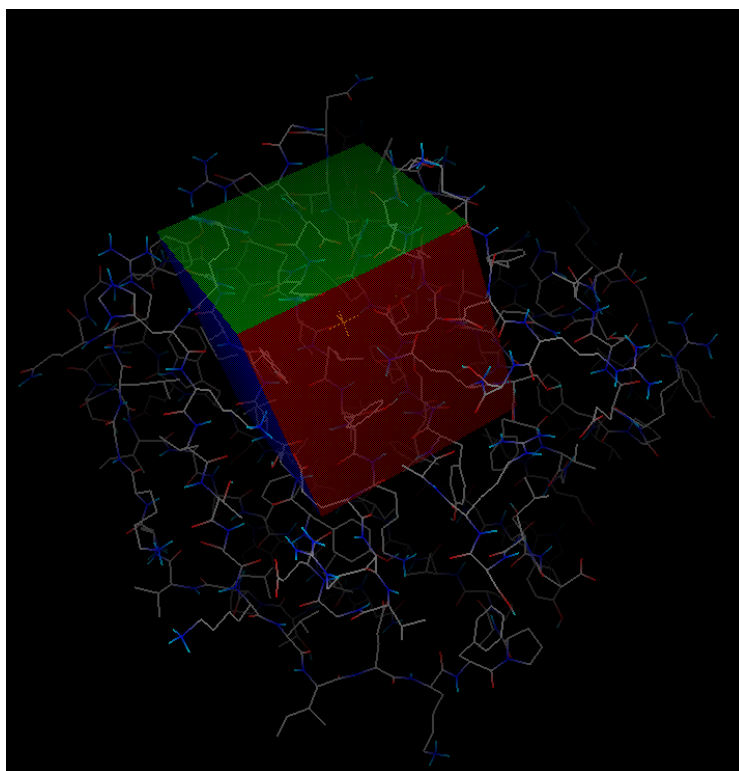


Figure 4.7 Image generating using ADT (Autodock tools), showing SAP protomer after addition of polar hydrogens and deletion of water molecules. The docking search box is $36\text{\AA} \times 36\text{\AA} \times 36\text{\AA}$ and centred on SAP's calcium binding site.

A literature search did not yield any parameters for the treatment of calcium atoms within AUTODOCK's energy scoring algorithm, so these were derived by considering a number of force field and molecular dynamic models of proteins containing metal ions (Marchand *et al*, 1998, Cates *et al*, 2002, Hu *et al*, 2003), coupled by consultation with Dr. Garrett Morris and Dr. Ruth Huey of the Scripps institute. As a result the inter-nuclear separation value (R_{ii} or R_{eqm}) was set at 2.4\AA , approximately twice the atomic radii expected for a calcium atom bound within a protein, whilst the Lennard Jones 12-6 well depth (ϵ_{sij}) was set at 0.15kcal/mol (figure 4.8) as a model for the Van der Waals interactions. Charges on the calciums were set at +2.

Ligand parameter files in PDBQ format were generated and energy minimised using the online PRODRG2 server (Schuettelkopf *et al*, 2004). These parameters were then loaded into the subroutines AUTOTORS and AUTOGRID in order to assign flexible torsion angles and calculate grid files ready for docking.

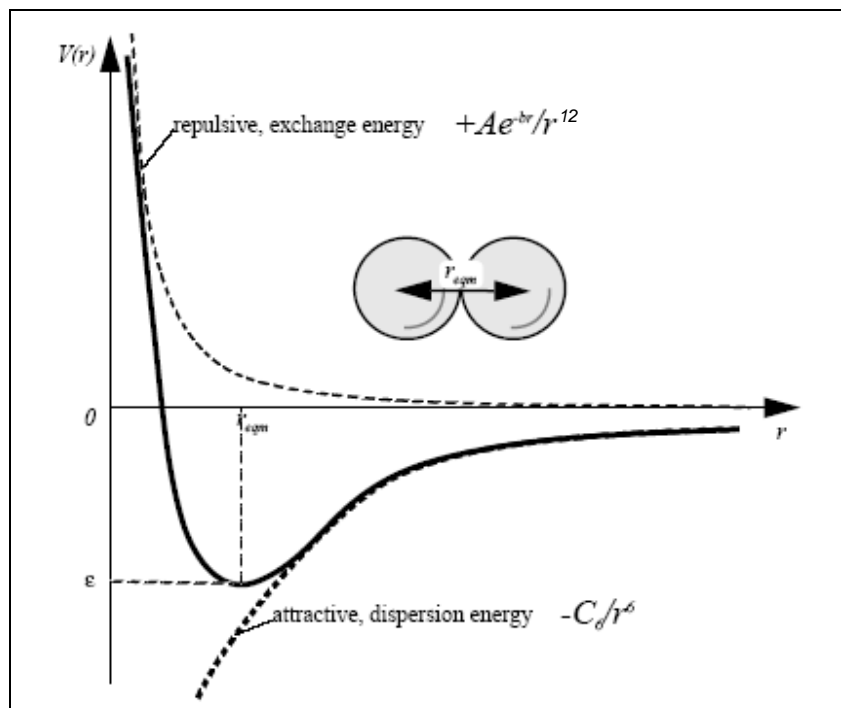


Figure 4.8 Lennard Jones 12-6 potential modelling Van der Waals interactions based on the atomic radii (r_{eqm}) and well depth (ϵ) (figure from Morris *et al*, 2003).

4.2.4 Autodocking results

Subunit B of the refined SAP/acetate structure was chosen for *in silico* docking experiments due to particularly clear electron density around the active site (see figure 4.14 above). All water molecules were deleted except for the two participating in the calcium coordination sphere of the second calcium atom. Initial docking experiments were carried out with a single acetate molecule to see if the crystal structure could be reproduced. Cluster analysis of the resulting docked ligands showed all ten conformations located in the binding site, with an RMSD of 0.2\AA between the docked conformations. Alignment between the lowest energy docked conformation and the crystallographic ligand position showed an RMSD of 1.08\AA (see figure 4.15), well within the "success" criteria of 2.4\AA described in the literature (Cecchini, *et al*, 2004). Despite this success AUTODOCK did seem to consistently position the acetate's carboxylate group closer to the calciums than expected (see discussion below).

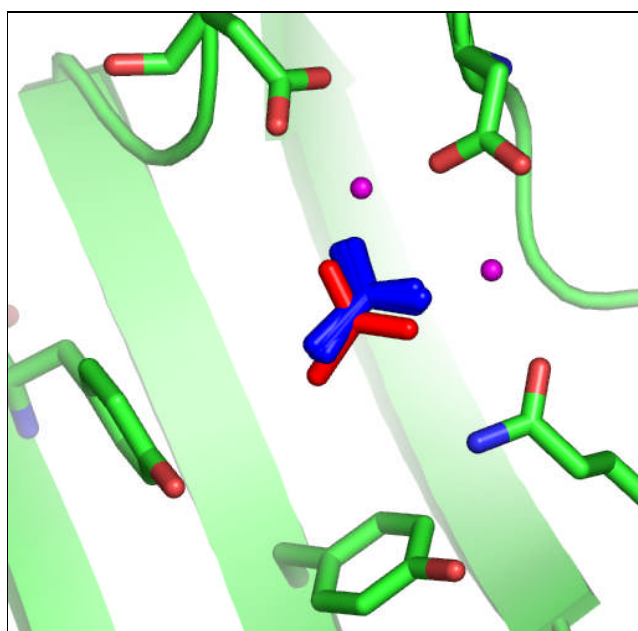


Figure 4.15 Acetate conformations generated by AUTODOCK (blue) compared to the experimentally derived ligand position (red). Calcium atoms are pink.

A second set of docking simulations were conducted using the same SAP protomer but this time using the ligand MO β DG. This dicyclic sugar (see figure 4.16) also interacts with the calciums via a carboxylate group, but in addition has a methyl group known to fit into the SAP hydrophobic pocket (Thompson *et al*, 2002). Interestingly energy minimisation prior to docking using PRODRG (Schuttelkopf *et al*, 2004) suggested that the lowest internal energy state for this ligand was with both rings in a *boat* conformation, contrary to the dual *chair* conformation seen in the crystallographic structure. Subsequent energy calculations in AUTODOCK seemed to agree as all docked configurations were likewise in the dual *boat* conformation. Auto-docking gave a final overall free energy estimation of -6.31 kcal/mol, with an internal energy of +1.4 kcal/mol for the ligand, and an RMSD of 1.45Å when compared to the crystal structure. To examine this further a second set of docking experiment were performed with MO β DG treated as a rigid body locked in the dual chair conformation. This resulted in a much closer corroboration with the crystallographic results (RMSD of 0.71Å), and an even lower calculated free energy (-7.92 kcal/mol), although treatment of the ligand as a rigid body prevented the incorporation of internal ligand energy in the energy calculations.

Again, in both cases, AUTODOCK seemed to displace the docked ligand position towards the calcium atoms (figure 4.16).

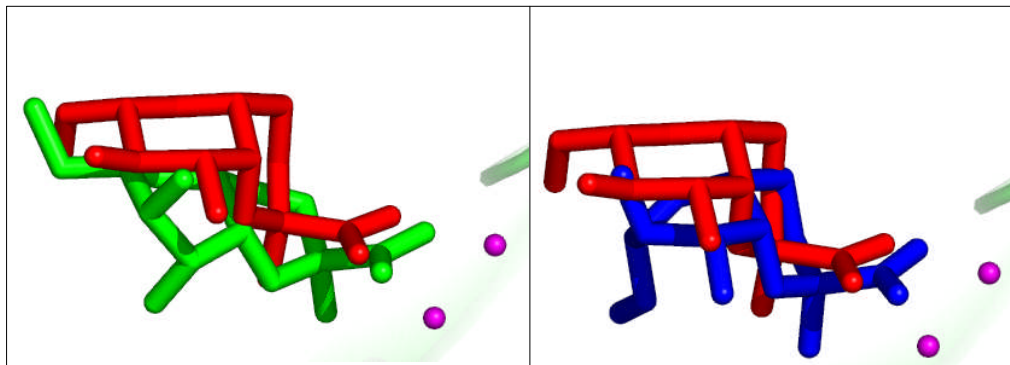


Figure 4.16 MOBDG coordinates obtained from the crystal structure (red) compared to the energy minimised docked structure in the dual boat conformation (green), and docking in the dual chair conformation (blue) after treating the ligand as a rigid body. As before the calcium atoms are shown in pink.

A third set of docking experiments were conducted using phospho-ethanolamine (PE) as the SAP ligand. This molecule interacts with the calciums via a phosphate rather than carboxylate group, and also has a two carbon hydrophobic group that fits into SAP's hydrophobic pocket (Pye, 1999). Docking of the energy minimised ligand coordinates produced a group of ten structures with an RMSD of 0.7\AA , and with an average RMSD of 1.06\AA from the crystallographic coordinates. As with the previous two ligands this difference was mostly due to the docked conformations being approximately 1\AA closer to the calcium atoms than in the crystal structures, however AUTODOCK also seemed to have problems recognising and docking the hydrophobic component of PE into the hydrophobic pocket (see figure 4.17 and discussion below). Free energy estimations of the docked conformations ranged from -8.02 to -7.11kcal/mol .

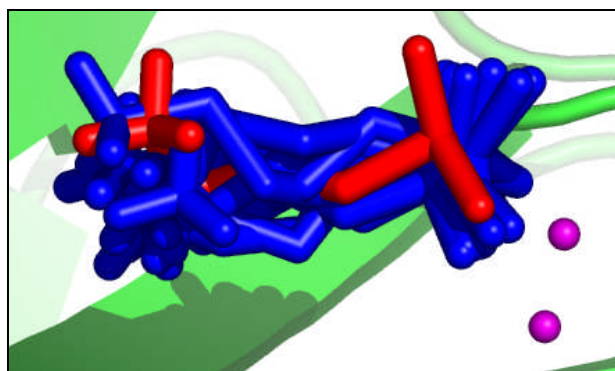


Figure 4.17 Position of ten docked PE positions (blue) compared to ligand position obtained using crystallography (red).

AUTODOCK recognised the calcium binding site in all docking simulations attempted thus far, however did seem to consistently place docked conformations approximately 1Å closer to the calciums than observed in the crystal structures. To see if this displacement was due to AUTODOCK treating the protein as a rigid body, the docking protomer was aligned with an equivalent subunit from the published SAP/MOBDG structure (1GYK.pdb, Thompson *et al*, 2002) using the CCP4 program PDBSET (CCP4, 1994). This alignment showed only minor displacements of the protein side chains surrounding the calcium binding site (RMSD = 0.3Å) (figure 4.18), suggesting that variation in ligand position was more likely due to errors in AUTODOCK's energy scoring function. This conclusion was not entirely surprising considering the lack of accurate metal parameters within the current version of AUTODOCK, and indeed had also been reported in auto-docking simulations using various ligands of the zinc containing matrix metalloproteinases (Hu *et al*, 2003). In these zinc experiments docking was improved by varying formal charges on the metal atoms using simplex optimisation of the calculated free binding energies. However, AUTODOCK uses the Gasteiger-Marsili method for calculating atomic charges (Gasteiger *et al*, 1980, reviewed by Leach, 2001), and calculates electro-negativity based on the occupancy and valency of interacting electron orbitals. As a result metal charges can only be accurately assigned by taking into account charges on all surrounding atoms, and hence any manual variation of atomic charge must also be reflected in the charges on neighbouring atoms within the energy scoring grid. As this cannot be adequately achieved in the current program, adopting formal charges as an initial approximation is

probably the safest option until a newer version of AUTODOCK incorporating calcium parameters is released.

Further docking experiments were performed with the remaining SAP ligands, and also with SAP's homologue CRP, using a subunit from the deposited 2.5Å 1B09 structure (Thompson *et al*, 1999).

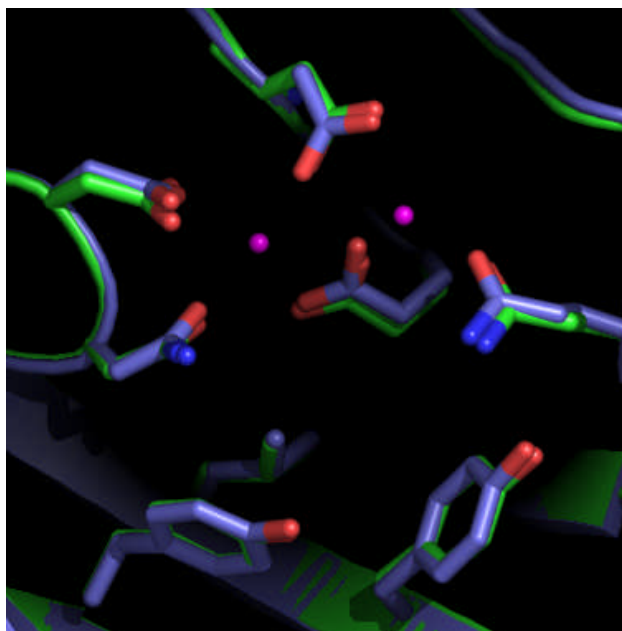


Figure 4.18 Alignment of a SAP subunit from the 1.88Å SAP/acetate structure used for autodocking (green), with a subunit from the 2.2Å SAP/MOBDG structure 1GYK.pdb (blue).

Docking between SAP and NADPro seemed to give a large internal energy value for the ligand, eclipsing the intermolecular interaction energy and resulting in a positive ΔG . This was interesting as the docked ligand conformation was very similar to that seen in the crystal structure (RMSD 1.16Å) suggesting that AUTODOCK was having trouble calculating the internal energies on ring structures. Removing this term by setting the ligand as a rigid body in its energy minimised conformation prior to docking seemed to resolve this problem, however this caused a slight increase in the RMSD by about 0.3Å compared with the crystal structure. AUTODOCK seemed unable to dock NALPro into the hydrophobic pocket unless NALPro was also treated as a rigid ligand, in which case it

found the crystallographic conformation four out of ten times, but with the highest positional accuracy of any of the ligands tested.

Results for all simulations are summarised in table 4.4 and 4.5 below, including a comparison between AUTODOCK's calculated free energy and the empirically derived free binding energy obtained using ITC. Ligand conformations closest to those observed in the crystal structures were chosen on occasions when AUTODOCK generated widely differing solutions (see discussion).

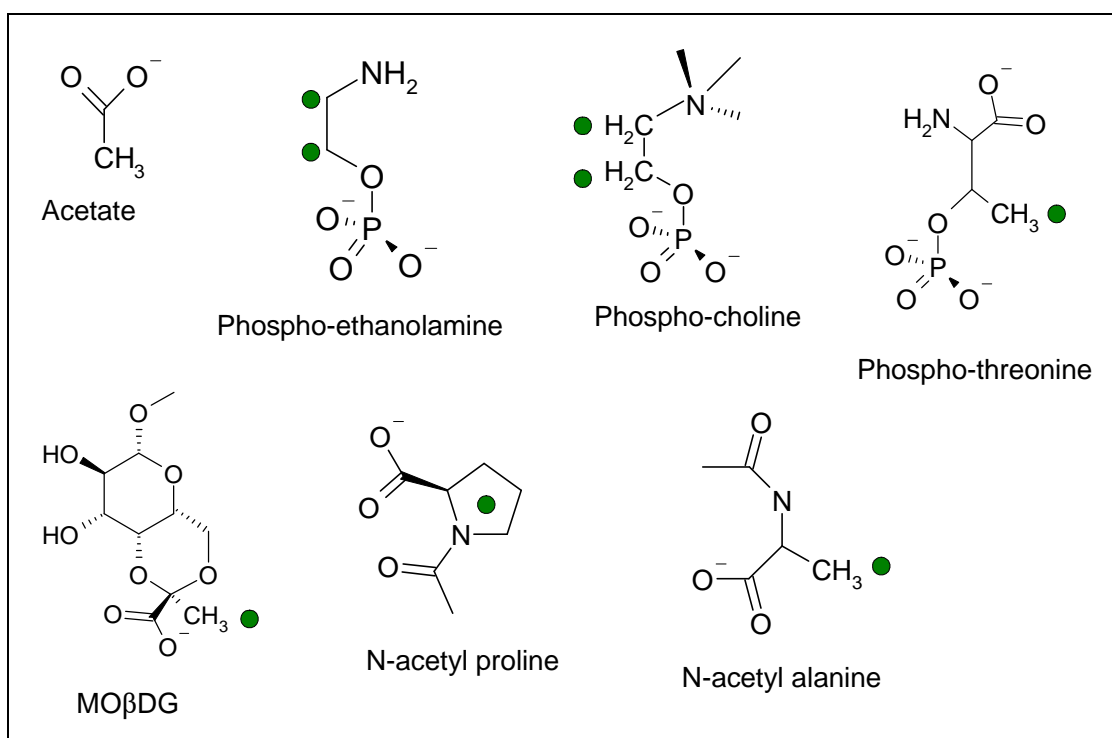


Figure 4.19 SAP and CRP ligands used in calorimetry and docking experiments. Green circle denotes groups thought to bind into hydrophobic pocket.

Ligand	ΔG_{exp} (kcal/mol)	ΔG_{prd} (kcal/mol)	Error ($\Delta G_{\text{prd}} - \Delta G_{\text{exp}}$) (kcal/mol)	RMSD Å	Crystal Structure Reference
NADPro (rigid)	-6.12	-7.28	1.16	1.22	Purvis, 2002
PE	-6.07	-7.02	0.32	1.06	Pye, 2000
MOBDG	-5.82	-7.08	1.26	1.45	Thompson, 1997
MOBDG (rigid)	-5.82	-7.30	1.48	0.71	Thompson, 1997
PT	-4.67	-6.91	2.24	1.54	Chapter 2
NALPro (rigid)	-4.76	-6.62	1.86	0.13	Chapter 2
NADAla	-4.00	-7.32	3.32	N/A	Unsolved
NALAla	-3.41	-6.10	2.64	N/A	Unsolved
PC	N/A	-7.10	N/A	1.47	Thompson, 2000
Acetate	N/A	-3.57	N/A	1.08	Emsley, 1994

Table 4.4 SAP/ligand free energy of binding calculated by AUTODOCK (ΔG_{prd}) compared to free energy measured by ITC (ΔG_{exp}), and compared to crystallographic ligand positions (RMSD).

Ligand	ΔG_{exp} (kcal/mol)	ΔG_{prd} (kcal/mol)	Error ($\Delta G_{\text{prd}} - \Delta G_{\text{exp}}$) (kcal/mol)	RMSD Å	Crystal Structure Reference
PC	-7.74	-8.13	0.39	1.47	Thompson, 1999
PE	-6.09	-7.02	0.93	1.51	Pye, 2000
MOBDG	N/A	-8.22	N/A	N/A	Unsolved
MOBDG (rigid)	N/A	-5.02	N/A	N/A	Unsolved
PT	N/A	-6.47	N/A	N/A	Unsolved
NALPro	N/A	-6.37	N/A	N/A	Unsolved
NADPro	N/A	-6.05	N/A	N/A	Unsolved
NALAla	N/A	-5.83	N/A	N/A	Unsolved
NADAla	N/A	-5.61	N/A	N/A	Unsolved
Acetate	N/A	-3.84	N/A	N/A	Unsolved

Table 4.5 CRP/ligand binding data showing experimental and predicted free energies along with RMSD's where crystal structure is available.