Chapter 1

An Introduction to Molecular Modeling and Computer-aided Drug Design

1.1 Molecular Modeling and Computational Chemistry

The definition currently accepted of what molecular modeling is, can be stated as this: “molecular modeling is anything that requires the use of a computer to paint, describe or evaluate any aspect of the properties of the structure of a molecule” (Pensak, 1989). Methods used in the molecular modeling arena regard automatic structure generation, analysis of three-dimensional (3D) databases, construction of protein models by techniques based on sequence homology, diversity analysis, docking of ligands or continuum methods. Thus, today molecular modeling is regarded as a field concerned with the use of all sort of different strategies to model and to deduce information of a system at the atomic level. On the other hand, this discipline includes all methodologies used in computational chemistry, like computation of the energy of a molecular system, energy minimization, Monte Carlo methods or molecular dynamics. In other words, it is possible to conclude that computational chemistry is the nucleus of molecular modeling. Identification of biomolecular moieties involved in the interaction with a specific receptor permits to understand the molecular mechanism responsible of its specific biological activity. In turn, this knowledge is aimed at designing new active molecules that can be successfully used as drugs. Due to the fact that simulation accuracy is limited to the precision of the constructed models, when it is possible, computational simulations have to be compared with experimental results to confirm model accuracy and to modify them if necessary, in order to obtain better representations of the system.

1.2 Quantum Mechanics and Molecular Mechanics

There are two different approaches to compute the energy of a molecule. First, quantum mechanics, a procedure based on first principles. In this approach, nuclei are arranged in the space and the corresponding electrons are spread all over the system in a continous electronic density and computed by solving the Schroedinger equation. For biomolecules this process can be done within the Born-Oppenheimer approximation, and for most of the purposes the Hartree-Fock self consistent field is the most appropiate procedure to compute the electronic density and the energy of the system. When chemical reactions do not need to be simulated, classical mechanics can describe the behavior of a biomolecular system. This mathematical model is known as molecular mechanics, and can be used to compute the energy of systems containing a large number of atoms, such as molecules or complex systems of biochemical and biomedical interest. In contrast to quantum mechanics, molecular mechanics ignore electrons and compute the energy of a system only as a function of the nuclear positions. Then, it is possible to take into account in an implicit way the electronic component of the system by adequate parameterization of the potential energy function. The set of equations and parameters which define the potential surface of a molecule is called *force field*.

1.3 Force Fields

In molecular mechanics the electrons and nuclei of the atoms are not explicitly included in the calculations. Molecular mechanics considers a molecule to be a collection of masses interacting one with each other through harmonic forces. Thus, the atoms in molecules are treated as ball of different sizes and flavors joined together by springs of variable strength and equilibrium distances (bonds). This simplification allows using molecular mechanics as a fast computational model that can be applied to molecules of any size.

In the course of a calculation the total energy is minimized with respect to the atomic coordinates, and consists in a sum of different contributions that compute the deviations from equilibrium of bond lengths, angles and torsions plus non-bonded interactions :

$$E_{tot} = E_{str} + E_{bend} + E_{tors} + E_{vdw} + E_{elec} + \ldots \quad (1.1)$$

where E_{tot} is the total energy of the molecule, E_{str} is the bond-stretching energy term, E_{bend} is the angle-bending energy term, E_{tors} is the torsional energy term, E_{vdw} is the van der Waals energy term, and E_{elec} is the electrostatic energy term.

In the first term in Eq. 1.1 describes the energy change as a bond stretches and contracts from its ideal unstrained length. It is assumed that the interatomic forces are harmonic so the bond-stretching energy term can be described by a simple quadratic function given in Eq. 1.2

$$E_{str} = \frac{1}{2} k_b (b - b_0)^2 \quad (1.2)$$

where k_b is the bond-stretching force constant, b_0 is the unstrained bond length, and b is the actual bond length.

Also for angle bonding a simple harmonic, spring-like representation is employed. The expression describing the angle-bending term is shown in Eq. 1.3

$$E_{bend} = \frac{1}{2} k_\theta (\theta - \theta_0)^2 \quad (1.3)$$

where k_q is the angle-bending force constant, θ_0 is the equilibrium value for the bond angle θ, and θ is the actual value for θ.

A common expression for the dihedral potential energy term is a cosine series as Eq. 1.4

$$E_{tors} = \frac{1}{2} k_j (1 + \cos(n\varphi - \varphi_0)) \tag{1.4}$$

where k_j is the torsional barrier, φ is the actual torsional angle, n is the periodicity (number of energy minima with a full cycle), and φ_0 is the reference torsional angle.

The van der Waals interactions between not directly connected atoms are usually represented by a Lennard-Jones potential (Eq. 1.5).

$$E_{vdW} = \sum \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} \tag{1.5}$$

where A_{ij} is the repulsive term coefficient. B_{ij} is the attractive term coefficient and r_{ij} is the distance between the atoms *i* and *j*. In order to describe the electrostatic forces an additional term with the Coulomb interaction is used (Eq. 1.6).

$$E_{elec} = \frac{1}{\varepsilon} \frac{Q_1 Q_2}{r_{ij}} \tag{1.6}$$

where ε is the dielectric constant, and Q_1 and Q_2 are atomic charges of interacting atoms and r_{ij} is the interatomic distance.

The equilibrium values of these bond lengths and bond angles are the corresponding force constants used in the potential energy function defined in the force field and define a set known as *force field parameters.* Each deviation from these equilibrium values will result in increasing total energy of the molecule. So, the total energy is a measure of intramolecular strain relative to a hypothetical molecule with an ideal geometry of equilibrium. By itself the total energy has no strict physical meaning, but differences in total energy between two different conformations of the same molecule can be compared.

1.4 Energy-Minimizing Procedures

Energy minimization methods can be divided into different classes depending on the order of the derivative used for locating a minmum on the energy surface. Zero order methods are those that only use the energy function to identify regions of low energy through a grid search procedure. The most well known method of this kind is the SIMPLEX method. Within first-derivative techniques, there are several procedures like the steepest descent method or the conjugate gradient method that make use of the gradient of the function. Second-derivative methods, like the Newton-Raphson algorithm make use of the hessian to locate minima. In the present study only first-derivative methods have been used and will be briefly described.

1.4.1 Steepest Descent Method

In the steepest descent method, the minimizer computes numerically the first derivative of the energy function to find a minimum. The energy is calculated for the initial geometry and then again after one of the atoms has been moved in a small increment in one of the directions of the coordinate system. This process is repeated for all atoms which finally are moved to a new position downhill on the energy surface. The procedure stops when a predetermined threshold condition is fulfilled. The optimization process is slow near the minimum, and consequently, the steepest descent method is often used for structures far from the minimum as a first, rough and introductory run followed by a subsequent minimization employing a more advanced algorithm like the conjugate gradient.

1.4.2 Conjugate Gradient Method

The conjugate gradient algorithm accumulates the information about the function from one iteration to the next. With this proceeding the reverse of the progress made in an earlier iteration can be avoided. For each minimization step the gradient is calculated and used as additional information for computing the new direction vector of the minimization procedure. Thus, each successive step refines the direction towards the minimum. The computational effort and the storage requirements are greater than for steepest descent, but conjugate gradients is the method of choice for larger systems. The greater total computational expense and the longer time per iteration is more than compensated by the more efficient convergence to the minimum achieved by conjugate gradients.

As a summary, the choice of the minimization method depends on two factors: the size of the system and the current state of the optimization. For structures far from minimum, as a general rule, the steepest descent method is often the best minimizer to use for the 100-1000 iterations. The minimization can be completed to convergence with conjugate gradients.

There are several ways in molecular minimization to define convergence criteria. In non-gradient minimizers only the increments in the energy and/or the coordinates can be taken to judge the quality of the actual geometry of the molecular system. In all gradient minimizers, however, atomic gradients are used for this purpose. The best procedure in this respect is to calculate the root mean square gradients of the forces on each atom of a molecule.

The value chosen as a maximum derivative will depend on the objective of the minimization. If a simple relaxation of a strained molecule is desired, a rough convergence criterion like a maximum derivative of 0.1 kcal $mol^{-1}Å^{-1}$ is sufficient while for other cases convergence to a maximum derivative less than 0.001 kcal $mol^{-1}Å^{-1}$ is required to find a final minimum.

1.5 Use of charges and solvents

Molecular mechanics calculations are traditionally carried out in vacuum conditions $\varepsilon = 1$. The investigation of molecules containing charges and dipoles however requires the consideration of solvent effects; otherwise conformations are influenced by strong electrostatic interactions. Force fields try to maximize the attractive electrostatic interaction, resulting in energetically strongly preferred but unrealistic low-energy conformations of the molecule. This can be prevented by employing the corresponding solvent dielectric constant. For example, in water ε amounts to 80.

The strength of the electrostatic interaction decreases slowly with r^{-1}. Therefore, in some cases the dielectric constant is chosen to be distance-dependent in order to decrease more rapidly, avoiding the need to consider atoms far away from each other, simulating the effect of displacement of solvent molecules in course of the approach of a ligand molecule to a macromolecular surface.

1.5.1 Periodic Boundary Conditions

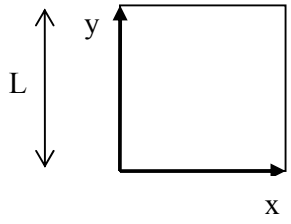
A more realistic approach is to use the solvent explicitly. This is done by soaking the molecule in a box of solvent molecules. This method has the disadvantage of requiring additional computational effort. Periodic Boundary Conditions (PBC) are normally employed to model the bulk solvent. In infinite PBC, the simulation box is infinitely replicated in all directions to form a

lattice. In practice, most molecular dynamics (MD) simulations evaluate potentials using some cutoff scheme for computational efficiency. In these cutoff schemes, each particle interacts with the nearest images of the other N-1 particles (minimum-image convention), or only with those minimum images contained in a sphere of radius R_{cutoff} centered at the particle. The use of cutoff methods, however, has been shown to introduce significant errors and artificial behavior in simulation (Bader et al., 1992, Schreiber et al., 1992, York et al. 1994).

The total Coulomb energy of a system of N particles in a cubic box of size L and their infinite replicas in PBC is given by

$$U = \frac{1}{2}\sum_{n}^{'}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{q_i q_j}{r_{ij,n}} \tag{1.7}$$

where q_i is the charge of particle *i*. The cell-coordinate vector is $n = (n1, n2, n3) = n_1 Lx + n_2 Ly + n_3 Lz$, where *x, y, z* are the cartesian coordinate unit vectors.



(a)

n(-1,1) x	n(0,1) x	n(1,1) x
n(-1,0) x	n(0,0) x	n(1,0) x
n(-1,-1) x	n(0,-1) x	n(1,-1) x

(b)

Figure 1.1. In a 2D system (a) the unit cell coordinates and (b) a 3 x 3 periodic lattice built from unit cells.

1.5.2 Ewald Summation Techniques

In most MD simulations, the long-range interactions (Coulomb interactions) are the most time consuming. Ewald summation was introduced (Ewald, 1921) as a technique to sum the long-range interactions between infinite particles and all their infinite periodic images efficiently. Long-range interactions are evaluated as sums that converge extremely slow. The trick when calculating the Ewald sum is to convert the summation of the potential energy into two series, each of which converges much more rapidly and a constant term (Eq. 1.8). This is done by considering each charge to be surrounded by a neutralizing charge distribution of equal magnitude but of opposite sign as shown in Figure 1.2. A Gaussian charge distribution is commonly used.

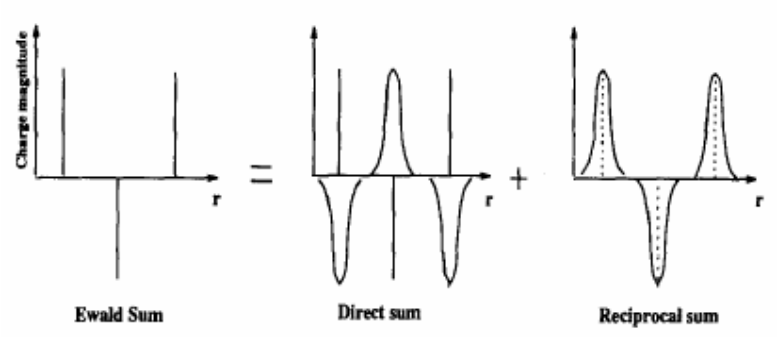


Figure 1.2. In a 2D system (a) the unit cell coordinates and (b) a 3 x 3 periodic lattice built from unit cells. Figure adapted from Toukmaji et al. (1996).

The sum over point charges is now converted to a sum of the interactions between the charges plus the neutralizing distributions. This part is the real space sum U^r.(Eq. 1.9). A second charge distribution is added to the system which exactly counteracts the first neutralizing distribution. This summation is performed in the reciprocal space and is termed U^m (Eq. 1.10). The self-term U^0 is a correction term that cancels out the interaction of each of the introduced artificial Gaussian counter-charges with itself (Eq. 1.11).

$$U_{Ewald} = U^r + U^m + U^0 \tag{1.8}$$

$$U^r = \frac{1}{2}\sum_{i,j}^{N'}\sum_{n} q_i q_j \frac{erfc(\alpha r_{ij,n})}{r_{ij,n}} \tag{1.9}$$

$$U^m = \frac{1}{2\pi V}\sum_{i,j}^{N'} q_i q_j \sum_{m\neq 0} \frac{\exp(-(\pi m/\alpha)^2 + 2\pi i m\cdot(r_i - r_j))}{m^2} \tag{1.10}$$

$$U^o = \frac{-\alpha}{\sqrt{\pi}}\sum_{i=1}^{N} q_i^2 \tag{1.11}$$

In these equations *V* is the volume of the simulation box, *m=(l,j,k)* is a reciprocal-space vector, and *n=(n1, n2, n3)* is the cell coordinator vector. The complimentary error function decreases monotonically as *x* increases and is defined by

$$erfc(x) = 1 - erf(x) = 1 - (2/\sqrt{\pi})\int_0^x e^{-u^2} du \tag{1.12}$$

The dipole term includes the effects of the total dipole moment of the unit cell, the shape of the macroscopic lattice, and the dielectric constant of the surrounding medium.

1.5.3 Particle Mesh Ewald

The Particle-Mesh Ewald method (PME) divides the potential energy into Ewald's standard direct and reciprocal sums and uses the conventional Gaussian charges distributions (Darden et al., 1993). The direct sum in Eq. 1.9, is evaluated explicitly using cutoffs while the reciprocal sum (Eq. 1.10) is approximated using Fast Fourier Transform (FFT) with convolutions on a grid where charges are interpolated in the grid points. Furthermore, PME does not interpolate but rather evaluates the forces by analytically differentiating the energies, thus reducing memory requirements substantially (Figure 1.3).

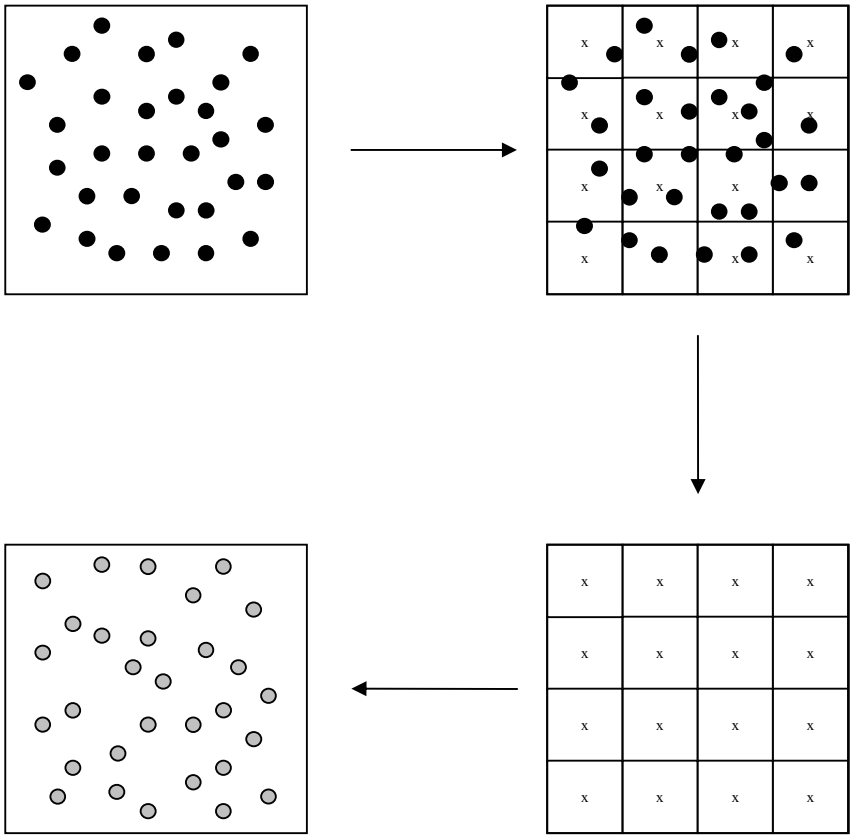


Figure 1.3. A 2D schematic of particle-mesh technique used in most Fourier-based methods. (a) A system of charged particles. (b) The charges are interpolated on a 2D grid. (c) Using FFT, the potential and forces are calculated at grid points. (d) Interpolate forces back to particles and update coordinates.

1.6 Conformational Analysis of a Peptide: Multiple Minima Problem

Conformational analysis is the characterization of the structures that a molecule can adopt and how these influence its properties. A key component of a conformational analysis is the conformational search, the object of which is to identify the preferred conformations of a molecule, i.e. those conformations that determine its behavior. This usually requires the characterization of conformations that are a minimum on the potential energy surface. For a peptide, due to its high conformational flexibility in solution, there are a so large number of minima on the energy surface that is impractical to characterize them all. This is known as the multiple minima problem (Gibson and Scheraga, 1988) and it is the main difficulty to structurally characterize a peptide. Specifically, most of the peptides exist in physiological conditions as a mixture of interchangeable conformations with similar energies populated according to the Boltzmann distribution. It is important to remember that the statistical weights of the different conformations involve also entropic contributions. Solvation effects may also be important, and various schemes are now available for calculating the solvation free energy of a conformation, that may be added as an additional term to the intramolecular energy. Under such circumstances, it is often assumed that the native (i.e. naturally occurring) conformation is the one with the very lowest value of energy. This conformation is usually referred to as the global minimum. Although the global minimum exhibits the lowest energy value, it may not be highly populated because of the contribution of the vibrational entropy to the statistical weight of each structure. Moreover, the global minimum may not be the active (i.e. the functional) structure. In this case, it may be even necessary for a molecule to adopt more than one conformation. For example, a substrate might bind in one conformation to an enzyme and then adopt a different conformation prior to reaction is produced. Indeed, in some cases it is possible that the active conformation does not correspond to any minimum on the energy surface of the isolated molecule.

Computational methods for the exploration of the conformational space of a peptide started about thirty years ago (Scheraga, 1968). From then different strategies have been described and reviewed (Howard and Kollman, 1988; van Gunsteren & Berendsen, 1990; Leach, 1991; Scheraga, 1992) and, although many efforts have been devoted, this field of research still remains open. Conformational search methods can be divided into the following categories: systematic search algorithms, model-building methods, random approaches, distance geometry and MD. Independently from the strategy selected, four key elements are needed to carry out the exploration of a peptide conformational space. The first consists of employing a peptide model description based on classic mechanics, i.e. a force field that permits to calculate the energy of a determined conformation. The second is to find a method capable of generating different

conformations, in order to explore all the low energy regions of the conformational space. The third key element consists of minimizing the different conformations, whereas the fourth and last element is to find a convergence criterion to assess if the conformational space has been sufficiently explored.

A conformational search method that has shown to be particularly effective for the exploration of the conformational space of peptides is the iterative simulated annealing (Filizola et al., 1997). The method has been used in the present thesis work for the conformational analysis of the farnesyltransferase inhibitor analogs described in chapter 2. The simulated annealing method was first described in 1983 (Kirkpatrick et al., 1983). This method is based on the similarity that exists between locating the global minimum of the potential energy function of a molecule and the slow cooling required to obtain a perfect crystal. In fact, crystal growing will probably be perfect if the system is cooled very slowly by reaching the thermodynamic equilibrium when passing through restrained regions of the phase space. Application of this concept to the exploration of the conformational space can be translated in terms of starting the simulation at a sufficiently high temperature and subsequently decreasing it gradually until the system is frozen in the global minimum. All the studies carried out using the simulated annealing method have demonstrated that although the cooling scheme is not sufficiently slow to find the global minimum, it is capable to find local energy minima of the regions explored. This means that simulated annealing combined with a searching strategy, which permits to cross different potential energy barriers and to reach the low energy regions, is a very efficient method to explore the conformational space.

Under such circumstances, a protocol was developed in our group for the exploration of the conformational space of peptides, based on the simulated annealing method but performed in an iterative fashion (Filizola et al., 1997). The proposed strategy is schematically shown in Figure 1.4. The method has demonstrated to be particularly efficient in the case of large peptide sequences. This procedure requires a starting structure, generally in an extended conformation, that is subjected to an annealing process and then minimized. This conformation constitutes the starting structure for a new cycle of simulated annealing after the molecule is abruptly heated. Heating is fast in order to force the molecule to jump to a different region of the conformational space. At this point the structure is slowly cooled and then minimized. The structure is subsequently stored on a file and used as the starting conformation for a new cycle of simulated annealing. In this way, a library of low energy conformations is generated. The procedure is repeated until a certain convergence criterion is fulfilled.

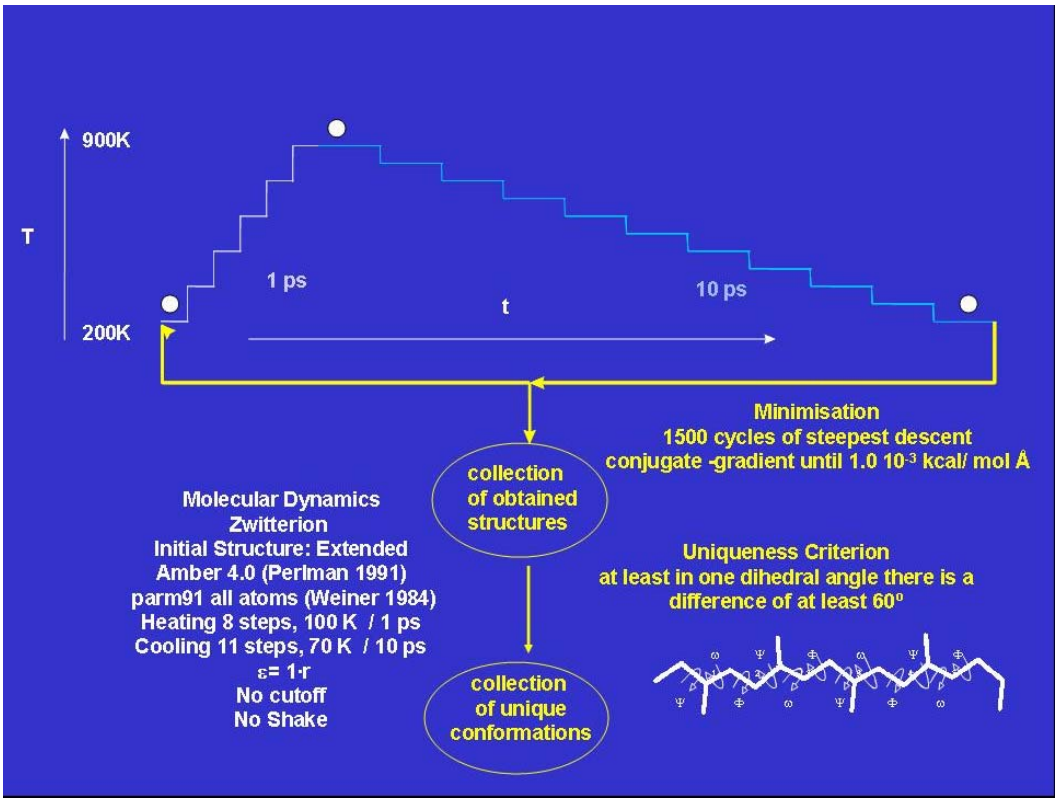


Figure 1.4. Schematic diagram of the iterative simulated annealing protocol used in the present work.

1.7 Assessment of the Bioactive Conformation

The computational methods used for the exploration of a peptide conformational space constitute an alternative procedure to experimental approaches to determine the bioactive conformation. Computational methods are very useful when diverse structure-activity studies about different analogs of the wild peptide are available. One of the basic assumptions in indirect methods is that the bioactive conformation is one of the thermodynamically accessible conformations of the peptide. From an energetic point of view, it is a reasonable assumption to consider that the bioactive conformation is one that exhibits high thermodynamical probability to exist in solution. The more populated this conformation is, the lower configurational entropy loss will occur at the binding step. This is supported by the fact that conformationally constrained analogs exhibit higher affinity for the receptor than linear peptides. This is probably due to the reduced set of conformations accessible to them and to the reduced entropy loss in the complex formation.

Exploration of the conformational space of a peptide leads to characterize the set of low energy conformations, which have high probability of being populated at physiological conditions. There is really no reason to think that the lowest energy conformation is the bioactive form. In fact, it is possible that another conformation is capable of establishing a better interaction with the receptor, acquiring a higher statistic weight in the receptor environment. In addition, it is hard to think that conformational analysis on only one peptide might be sufficient to determine its bioactive conformation.

Alternatively, in comparative conformational analysis the bioactive conformation is found to be one of the low energy conformations common to all the active peptide analogs studied. This approximation is based on the idea that the bioactive conformation is related to the biological activity that the peptide exerts. This implies that the peptide primary sequence contains all the functional groups needed to determine the biological answer, whereas the bioactive conformation corresponds to the spatial arrangement that favors the interaction of these functional groups with the receptor. If various related peptides interact with the same receptor with a similar activity, it is possible to infer that they adopt the same conformation.

It is also important to include inactive analogs in the comparison, although in this case data have to be interpreted cautiously. Considering that complex formation occurs due to a number of peptide-receptor interactions, the lack of only one of them leading to a decrease of the activity is of difficult interpretation. It is possible that inactive analogs still retain capability to adopt the bioactive conformation although the complex is less stable due to the lack of one interaction. Furthermore, it is also possible that a residue favors the bioactive conformation candidate that is included in the low energy conformation set of an inactive analog and this could lead to wrong conclusions.

Comparative conformational analysis can be performed if two conditions are met: i) the exploration of the conformational space is thorough and most of the thermodynamically accessible conformations of the peptide and its analogs are characterized; ii) the set of analogs has to be sufficient large to reach to only one solution capable to explain all the structure-activity results of the peptide. Selection of an exploration strategy depends very much on peptide size. In the case of small peptides, different exploration methods have been successfully described, whereas for larger peptides the strategy selection is more difficult. In this case, the simulated annealing method in an iterative fashion has demonstrated to be particularly efficient (see chapter 2).

It is also convenient that the set of selected analogs have a sequence as similar as possible. In the case of active analogs it is better to select the smaller sequence responsible for their activity. For the active and inactive analogs, it is convenient to select those that have precise residue substitutions because the interpretation of the results may be much easier, due to the

dependence of only one interaction type. Indeed, from a computational point of view it is convenient to select the smaller set of analogs (higher information with the less number of analogs) and it is better to select analogs with small size that allow a more thorough exploration of the conformational space, although it is not always possible.

1.8 Folding Studies on Peptides

Although obtaining the whole folding mechanism of proteins through MD has remained elusive until now (Duan et al. 1998a, Duan et al. 1998b), folding studies of peptides through MD are within the reach for currently available computational power. The reversible folding of peptides through MD has been described in past years (Daura et al., 1998, Daura et al., 1999, Daura et al., 2001). Given that secondary structural motifs formation seems to occur in the nanosecond timescale (Hummer et al., 2001**),** a 10-residue peptide is expected to fold during a 5 to 10 ns trajectory. Chapter 3 and 4 in the present work describe the folding of a 10-residue peptide, substance P, under different solvent conditions by MD and NMR spectroscopic experiments. Although peptides could seem good candidates to study folding events, they present several disadvantages, being one example the multiplicity of attainable conformations in solution. Although they have a dominant backbone conformation in the simulation they also visit other conformations along the MD trajectory. This constitutes an ensemble of conformations that allows the study of conformational transitions but makes impossible to identify a unique conformation of the peptide in solution. Ensemble behavior is difficult to evaluate. Traditionally the tendency has been to assess predictive success by using the root-mean-square-deviation (RMSD) from an X-ray structure. This is a continuous scalar value and thus, the definition of the *native* and *unfolded* states becomes a subjective matter. Duan et al. (1998a) classified the structures by a clustering algorithm and used this classification to study transitions in order to follow the evolution of the folding event. We agree with this strategy and based on it we have developed an objective classification algorithm based on information theory (see Chapter 3). The method is suited for assessing the evolution of the folding process and the study of transitions between different groups of conformations. Furthermore, it is possible to know how far we have gone in the folding process by looking at the rate of appearance of new conformations, as the longer the trajectory the less new conformations appear. The method is suited for peptides and proteins of any size and could be easily automated for the study on parallel of different peptides or proteins.

1.9 Computer-aided Drug Design

Computer-aided drug design, often called *structure based design* involves using the biochemical information of ligand-receptor interaction in order to postulate ligand refinements. For

example, if we know the binding site the steric complementarity of the ligand could be improved to increase the affinity for its receptor. Indeed, using the crystal structure of the complex we can target regions of the ligand that fit poorly within the active site and postulate chemical modifications that lower the energetic potential by making more negative the Van der Waals terms, thus improving complementarity with the receptor. In a similar fashion, functional groups on the ligand can be changed in order to augment electrostatic complementarity with the receptor.

When a target is selected for the design of new lead compounds three different situations can be faced regarding the amount of information of the system that is available: 1) the structure of the receptor is well known and the bioactive conformation of the ligand is or is not known, 2) only the bioactive conformation of the ligand is known and 3) the target structure and the bioactive conformation of the ligand are unknown (Figure 1.5).

The best possible starting point is an X-ray crystal structure of the target site. If the molecular model of the binding site is precise enough, one can apply docking algorithms that simulate the binding of drugs to the respective receptor site, like Autodock (Morris et al., 1998). In a first step the program creates a negative image of the target site through the use of several atom probes that determine affinity potentials for each atom type in the substrate molecule at different points in a grid, place the putative ligands into the site and finally they evaluate the quality of the fit. The program will try a set of different conformers of the ligand in order to obtain the best disposition of the atoms of the molecule for maximizing the scoring function that quantifies ligand-receptor interaction.

A different strategy for obtaining new lead compounds through rational drug design is the *de novo* design of ligands with the use of a builder program, like Ligbuilder (Wang et al., 2000). This program also determines the shape and the electrostatic properties of the binding site cavity through the use of several atom probes and then it combines from a library of chemical fragments those that better fill the cavity based on steric and electrostatic complementarity.

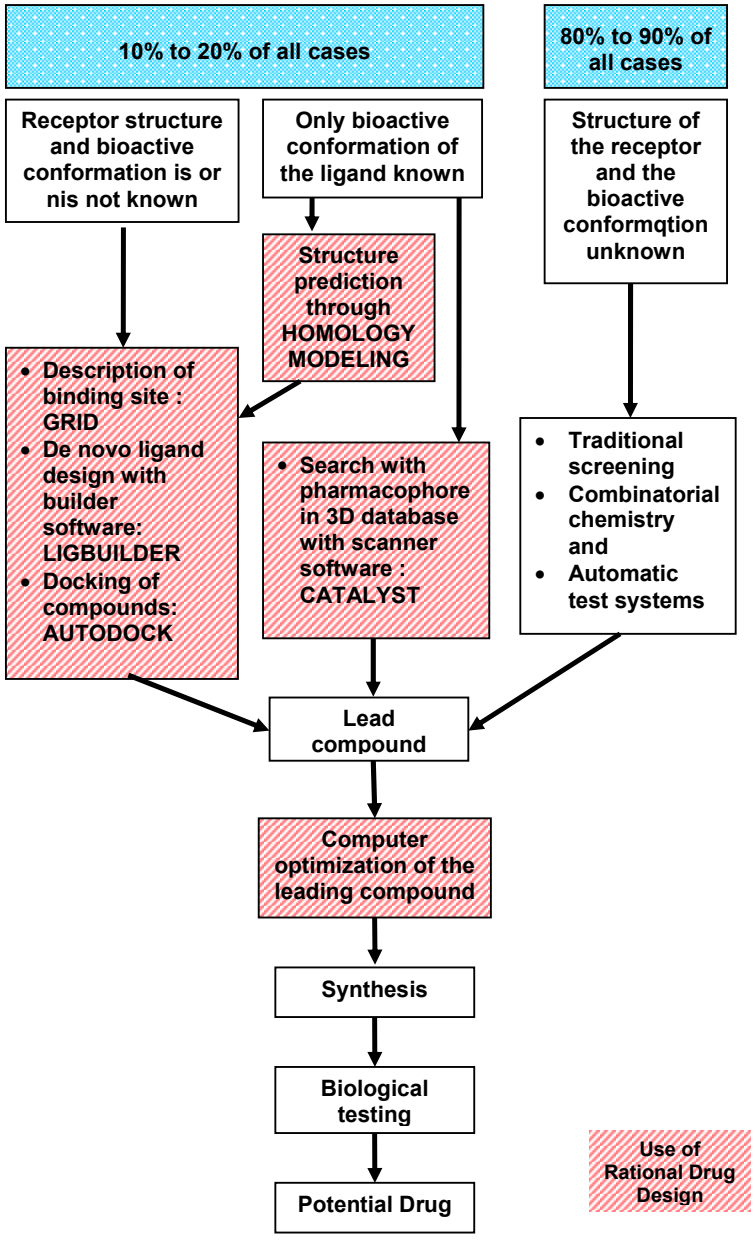


Figure 1.5. Overview of different strategies used for the search of new potential drugs.

Even if the structure of the binding site of the receptor is unknown computational methods may assist in predicting its 3D structure by comparing the chemical and physical properties of drugs that are known to act at a specific site. Moreover, if the amino acid sequence of the receptor site is known, one can try to predict the structure of the unknown site. This can either be done from scratch or by using a known structure of a related protein as template. If about 25 to 30 % of the amino acid residues are identical in two proteins, one may assume that the 3D structure of these two proteins is very similar. The technique used for this approach is called *homology modeling*: the folding pattern of the template protein is maintained and the side chain atoms of the template protein are replaced by the side chain atoms of the unknown protein. Basically, the 3D structure of a protein is represented by the 3D organization of the backbone atoms. The side chain atoms, which are different for all 20 amino acids, define the specific interactions with ligands or other protein domains. Replacing the side chains while maintaining the backbone allows to keep the general structure of the protein and to evaluate the specific properties of the unknown protein with respect to ligand interactions.

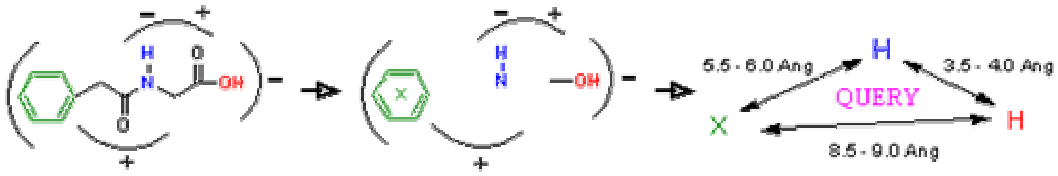


Figure 1.6. Pharmacophore definition from a known bioactive conformation.

If there is not a described structure for the receptor but the bioactive conformation of the ligand is known a pharmacophore can be derived from it and new lead compounds can be sought in 3D-databases with the aid of a scanner program like *Catalyst*^{TM} (2001). A pharmacophore is a group of chemical functions of known relative spatial disposition (Figure 1.6). The pharmacophore is translated into a query to the 3D database and a group of hits will be identified (Figure 1.7). An example of this strategy for the search of new leads is described in Chapter 3. *Catalyst*^{TM} superimposes for each molecule in the databases all the existing conformers giving as a result the group of molecules fulfilling the pharmacophore requirements and the conformers that better fit the pharmacophore descriptors. The conformational space of each molecule is determined with the use of a *Poling function* that improves the efficacy of the sampling. The penalty function used for minimization is modified to force similar conformers away from each other, thus reducing the number of redundancies and as a consequence decreasing the time needed for a wide exploration of the conformational space (Smellie et al., 1995).

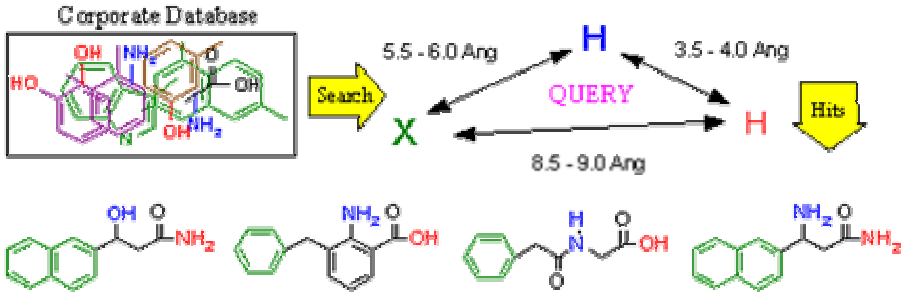


Figure 1.7. Results obtained in a 3D database search.

Database search programs have inherent strengths. To begin, the user has complete control over the query specifications. This allows for the retrieval of structures that meet the requirements of the pharmacophore and have a better opportunity to complement the receptor. Secondly, because these programs utilize a database of known compounds, synthetic feasibility is not an issue. In addition, these programs are usually highly optimized for speed, which allows for the rapid determination of potential binding ligands. Finally, since compounds are retrieved that mirror the query, no scoring functions are required. The assumption is that the 3D structure stored in the database is representative of biological reality. Although this can be true of small molecules, larger structures are often too flexible for the assumption to hold true.

1.10 References to chapter 1

Bader, J. and Chandler, D. *J. Phys. Chem.,* 96, 6423, (1992).

Catalyst^{TM}, Molecular Simulations, Inc. San Diego, CA, (2001).

Darden,T., York, D. and Pederson, L. *J. Chem. Phys.,* 98, 10089, (1993).

Daura, X., Gademann, K., Schäfer, H., Jaun, B., Seebach, D. and van Gunsteren, W.F. *J. Am. Chem. Soc.,* 123, 2393-2404, (2001).

Daura, X., Jaun, B., Seebach, D., van Gunsteren, W.F. and Mark, A.E. *J. Mol Biol.,* 280, 925-932, (1998).

Daura, X., van Gunsteren, W.F. and Mark, A.E. *Prot. Struct. Funct. Genet.* 34, 269-280, (1999)

Duan, Y. and Kollman, P.A. *Science,* 282, 740-744, (1998).

Duan, Y., Wang, L. and Kollman, P.A. *Proc. Natl. Acad. Sci. USA.* 95(17), 9897-902, (1998).

Ewald, P. *Ann. Phys.,* 64, 253, (1921).

Filizola, M. *PhD Thesis. Seconda Università degli Studi di Napoli*, (1999).

Filizola, M., Centeno, N.B., Perez, J.J. *J. Peptide Sci.,* 3, 85-92, (1997)

Gibson, K.D. and Scheraga, H.A. In *"Structure & Expression: Vol. 1: From Proteins to Ribosomes"*, Eds. R.H. Sarma & M.H. Sarma, Adenine Press, Guilderland, N.Y., p. 67-94, (1988).

Höltje, H.D. and Folkers, G. *Molecular Modeling. Basic Principles and Applications. Series: Methods and Principles in Medicinal Chemistry;* Ed. Mannhold, R., Kubinyi, H. and Timmerman. Vol. 5. VCH Publishers, Inc. New York, NY, (1997).

Howard, A.E. and Kollman, P.A. *J. Med. Chem.,* 31, 1669-1675, (1988).

Hummer, G., García, A.E. and Garde, S. *Prot. Struct. Funct. Genet.,* 42, 77-84, (2001).

Kirkpatrick, S., Gelatt, C.D. and Vecchi, M.P. *Science*, 220, 671-80, (1983).

Leach. A. In *Reviews in Computational Chemistry*; Eds. Lipkowitz, K. and Boyd, D. Vol. 2, pp. 1-47. VCH Publishers, (1991).

Morris, G. M., Goodsell, D. S., Halliday, R.S., Huey, R., Hart, W. E., Belew, R. K. and Olson, A. J. *J. Comput. Chem.,* 19, 1639-1662, (1998).

Pensak D.A. *Pure & Appl. Chem.,* 61, 601, (1989).

Scheraga, H.A. Scheraga. *Adv. Phys. Org. Chem.,* 6, 103-184, (1968).

Scheraga, H.A. *Int. J. Quantum Chem.,* 42, 1529-1536, (1992).

Schreiber, H. and Steinhauser, O. *Biochemistry*, 31, 5856, (1992).

Smellie, A., Teig, S.L. and Towbin, P. *J. Comp. Chem.*, 16, 171-187, (1995).

Toukmaji, A.Y. and Board. J.A. *Computer Physics Communications,* 95, 73-92, (1996).

van Gunsteren, W.F. and Berendsen, H.J.C. *Angew. Chem. Int. Ed. Engl.*, 29, 992-1023, (1990).

Wang RX, Gao Y, and Lai L.H. *J. Mol. Model.,* 6, 498-516, (2000).

York, D. Wlodawer, A.,Pedersen, L and Darden, T. *Proc. Natl. Acad. Sci. USA,* 91, 8715, (1994).