DEVELOPMENT AND APPLICATION OF MODELLING TECHNIQUES IN DRUG DESIGN

A THESIS SUBMITTED TO CARDIFF UNIVERSITY FOR THE DEGREE OF
PHILOSOPHIAE DOCTOR

NICOLA ZONTA

WELSH SCHOOL OF PHARMACY
CARDIFF UNIVERSITY

SUPERVISOR: Doctor Andrea Brancale

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Declaration

This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree.

Signed.....(candidate)
Date.....3/1/10.....

Statement 1

This thesis is being submitted in partial fulfillment of the requirements for the degree of PhD.

Signed.....(candidate)
Date.....3/1/10.....

Statement 2

This thesis is the result of my own independent work/investigations, except where otherwise stated. Other sources are acknowledged by explicit references.

Signed.....(candidate)
Date.....3/1/10.....

Statement 3

I hereby give consent for my thesis, if accepted, to be available for photocopying and interlibrary loans after expiry of a bar on access previously approved by the Graduate Development Committee.

Signed.....(candidate)
Date.....3/1/10.....

Abstract

Structure-based drug design is a creative process that displays several features that make it closer to human reasoning than to machine automation. However, very often the user intervention is limited to the preparation of the input and analysis of the output of a computer simulation. In some cases, allowing human intervention directly in the process could improve the quality of the results by applying the researcher intuition directly into the simulation.

Haptic technology has proven to be a useful method to interact in realtime with a virtual environment, enriching the user's experience and allowing for a more natural and direct interaction with three-dimensional systems.

Reported in this thesis is the design and implementation of a user-driven computer program for structure-based drug design based on haptic technology and characterised by a trimodal feedback system and its application alongside more traditional approaches to drug design projects in the anticancer and antiviral area.

The software proved to be very useful in several projects, validating the applicability of haptic technology to drug design. The results were in good agreement with those obtained by traditional techniques. Moreover the approach resulted in the identification of novel HCV inhibitors and a putative inhibitor of the dimerisation of EGFR which resulted active in vitro tests.

Contents

Chapter 1

Molecular modelling and drug design

Molecular modelling is the science (or art) of representing molecular structures numerically and simulating their behavior with the equations of quantum and classical physics. Computational chemistry programs allow scientists to generate and present molecular data including geometries (bond lengths, bond angles, torsion angles), energies (heat of formation, activation energy, etc.), electronic properties (moments, charges, ionisation potential, electron affinity), spectroscopic properties (vibrational modes, chemical shifts) and bulk properties (volumes, surface areas, diffusion, viscosity, etc.)[1, 2]. It is a rather new branch of theoretical chemistry and is evolving very fast, following the evolution of computing hardware that provides the computational power to carry out more and more complex calculations. The impact of molecular modelling on the process of drug discovery is already significant, and destined to become even more important[3, 4]. In the age of human genome being deciphered and 3-dimensional structures of proteins being made available in an increasing number, drugs may now be designed in advance guessing possible 3-dimensional shape of the targets. In drug design, potential compounds may be conceptualised for the performance of required functions based on essential characteristics (pharmacophores), including idealised structural and physical properties. Molecular models of com-

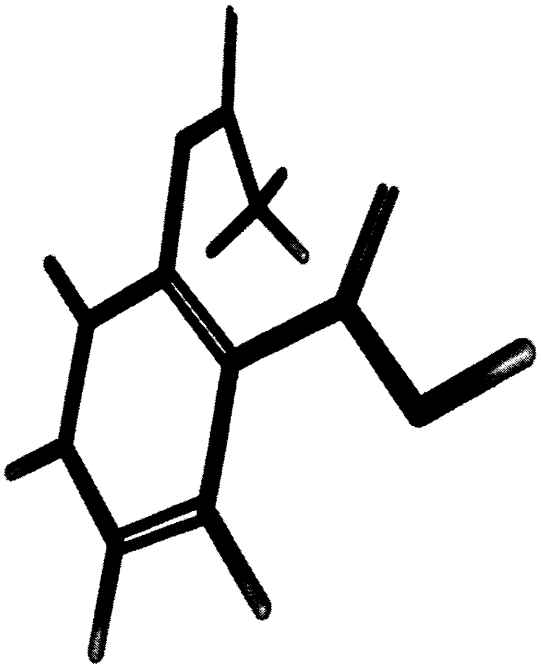
pounds can be built, and virtual tests may be run to assess their suitability before an expensive synthesis attempt is made. Virtual experiments are cheaper, faster, and safer than real experiments, and the data can help scientists to eliminate compounds that would not perform the required function. These are now essential tools for data management and cost-effective development processes in pharmaceutical Research and Developement[5].
Computer aided drug design approaches are traditionally divided into two major categories[6]: **ligand based** and **structure based**. Ligand based approaches rely on chemometric functions to compute various descriptors from the 2D or 3D structure of a set of molecules and use statistical methods to build correlation models to predict biological properties (such as activity) of compounds. Structure based approaches are based on the key-lock metaphor for ligand-target interaction.
This project focuses on the latter approach.

1.1 Molecular mechanics

The term molecular mechanics refers to the application of Newtonian mechanics to the modelling of molecular systems. In order to lower its complexity, the physical-mathematical description of molecules in a molecular mechanics simulation adopts several simplifications. Some of the most important are:

- molecules are treated as balls-and-springs systems
- bonds are unbreakable
- electrons are not considered as separate entities
- atoms are treated as solid spheres with fixed radius (typically the Van der Waals radius) and charge
- atoms differ among themselves for atom-type (depending on atomic number, hybridisation and chemical properties)

A molecule is therefore reduced to a set of atom types and a bonding list (Figure 1.1.



```
@<TRIPOS>MOLECULE
NONAME
21 21 0 0 0
SMALL
USER_CHARGES

@<TRIPOS>ATOM
  1 C         1.3750    0.7012    0.4866 C.ar  1 <1>   0.0862
  2 C         0.4248    1.4304   -0.2349 C.ar  1 <1>   0.0825
  3 O         0.5773    2.8137   -0.4648 O.3   1 <1>  -0.2325
  4 C        -0.7685    0.8436   -0.6684 C.ar  1 <1>  -0.1500
  5 H        -1.5144    1.4326   -1.1950 H     1 <1>   0.1500
  6 C        -0.9999   -0.5058   -0.4119 C.ar  1 <1>  -0.1500
  7 H        -1.9214   -0.9742   -0.7490 H     1 <1>   0.1500
  8 C        -0.0529   -1.2543    0.2876 C.ar  1 <1>  -0.1500
  9 H        -0.2360   -2.3072    0.4893 H     1 <1>   0.1500
 10 C         1.1277   -0.6545    0.7380 C.ar  1 <1>  -0.1500
 11 H         1.8476   -1.2605    1.2844 H     1 <1>   0.1500
 12 C         2.6135    1.3288    0.9459 C.2   1 <1>   0.6338
 13 O         3.1238    2.3390    0.4986 O.2   1 <1>  -0.5700
 14 O         3.1746    0.6610    1.9738 O.3   1 <1>  -0.6500
 15 H         3.9777    1.1747    2.2826 H     1 <1>   0.5000
 16 C         1.0925    3.3169   -1.6580 C.2   1 <1>   0.6590
 17 C         1.6199    2.3073   -2.6386 C.3   1 <1>   0.0610
 18 H         0.8095    1.6843   -3.0230 H     1 <1>   0.0000
 19 H         2.4099    1.7022   -2.1892 H     1 <1>   0.0000
 20 H         2.0617    2.8405   -3.4869 H     1 <1>   0.0000
 21 O         1.0958    4.5227   -1.8906 O.2   1 <1>  -0.5700
@<TRIPOS>BOND
  1   1   2  ar
  2   1  10  ar
  3   1  12  1
  4   2   3  1
  5   2   4  ar
  6   3  16  1
  7   4   5  1
  8   4   6  ar
  9   6   7  1
 10   6   8  ar
 11   8   9  1
 12   8  10  ar
 13  10  11  1
 14  12  13  2
 15  12  14  1
 16  14  15  1
 17  16  17  1
 18  16  21  2
 19  17  18  1
 20  17  19  1
 21  17  20  1
```

Figure 1.1: Examples of parameterisation for aspirin in TRIPOS mol2 format

Although most molecular mechanics application consider each atom as a separate entity, variations on this theme are possible; for example, many simulations have historically used a "united-atom" representation in which methyl and methylene groups were represented as a single particle, and large protein systems are commonly simulated using a "bead" model that assigns two to four particles per amino acid.

Sets of mathematical equations (forcefields) have been developed to compute various properties from this description of a molecular system. Inside the wide environment of molecular mechanics several specific applications can be performed. The

following sections describe some of the most important ones in computer aided drug design.

1.1.1 Forcefields and scoring functions

In a molecular mechanics environment all electrons are neglected and molecular properties cannot be calculated using quantum mechanics approaches. Sets of mathematical equations (forcefields) have been developed to compute various properties from atomic coordinates and atom types. The derivation of a forcefield can be empirical, trying to fit the equations to experimental or calculated data, or *ab initio*, deriving equations from classical physics, but very often results from the combined application of the two approaches. Many of the forcefields currently in use can be interpreted in terms of a fairly simple four-component picture of the inter- and intra-molecular forces within the system. Energetic penalties are associated with the deviation of bonds and angles away from their "reference" or "equilibrium" values, there is a function that describes how the energy changes as bonds are rotated, and finally the forcefield contains terms that describe the interaction between non-bonded parts of the system, usually in terms of electrostatic and van der Waals interactions. More sophisticated forcefields may have additional terms, but they invariably contain these four components[1]. Examples of forcefield functions are reported in Figure 1.2.

Some existing forcefields usually do not account for electronic polarization of the environment, an effect that can significantly reduce electrostatic interactions of partial atomic charges. This problem was addressed by developing polarizable force fields [7] or using a macroscopic dielectric constant. However, application of a single value of dielectric constant is questionable in highly heterogeneous environments such as proteins or biological membranes [8].

Like forcefields, scoring functions are mathematical functions that relate atomic coordinates to energy values[9]. They are usually developed for applications such as molecular docking and are parameterised to reproduce the interactions of intermolecular complexes. Scoring functions are usually much simpler than forcefield functions,

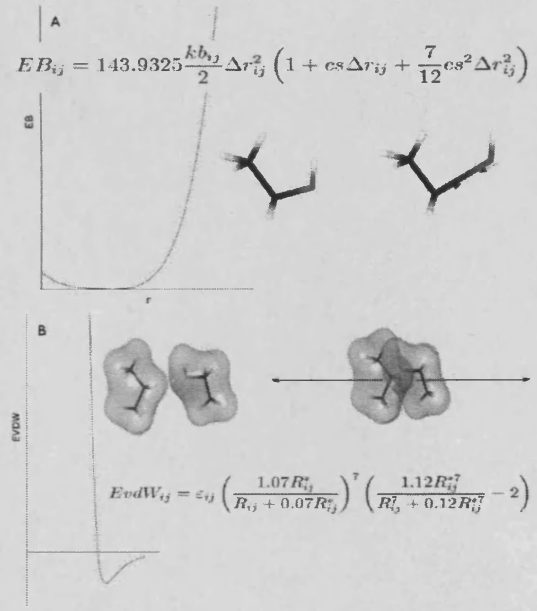


Figure 1.2: Examples of forcefield functions: A: MMFF94 Bond stretching energy. B: MMFF94 Van Der Waals interactions

as they are the speed determining point in, for instance, a docking algorithm. They usually focus on nonbonded interactions, and many of them consider bonded energies constant. Some scoring functions include empirical terms such as hydrogen bond energies or an estimate of the entropic loss of ligands upon binding or solvation effect. There are three major kinds of scoring functions, according to the way they are derived[1]:

- **Forcefield based**: these functions are based on a forcefield. The main drawback is that forcefields are not designed to account for entropic or solvation effects and therefore this effects cannot be accounted for.
- **Empirical**: they are designed to reproduce experimental affinity data. The parameters of the various energy contributions are obtained from a regression analysis using a training set of crystal structures of complexes and the relative experimental affinity data. They usually perform well and they are the most widely used, but they also show a strong dependence from the training set which might limit their applicability in some cases.
- **Knowledge-based**: they are trained to reproduce experimental structural data instead of affinites. They are based on the assumption that atom configurations that are often seen in crystallographic structures are energetically favoured.

1.1.2 Energy minimisation

Molecular mechanics forcefields are usually parameterised directly to reproduce potential energies. Using both ab-initio quantum mechanics-derived and empirical functions and parameters, forcefields compute both bonded and non-bonded interaction energies, thus providing the total potential energy of a molecular system as a mathematical function of its inner variables (usually atom coordinates). The analytical mathematical study of the energy function (usually known to computational chemists as energy surface or energy landscape) provides important insight on the properties of the conformational states of a particular molecule or a system of molecules. A

minimum point in the energy function (local minimum) is associated with a stable state of the system. Energy minimisation finds its application in several molecular mechanics and dynamics experiments, such as in generating good quality conformations of a given molecule or in refining molecular docking results or homology models. Various standard mathematical optimisation methods can be applied to find a local minimum of the energy landscape. The general idea common to most of this minimisation algorithms is walking downhill on the energy surface from a starting point until a minimum point is reached. Obviously this method can just find one minimum point, the one nearest to the starting point. Sometimes, however, it is important to find the global minimum (the minimum point with the lowest value), for instance to predict the most stable conformation of a molecule under given conditions. When this cannot be spotted with a systematic approach (as is usually the case in molecular modelling) a stochastic process is used to generate starting points for several minimisation runs.

Mathematically, a minimum point is one where the first derivative of the function with respect to each of the variables is zero and the second derivatives are all positive (equation 1.1).

$$\frac{\partial f}{\partial x_i} = 0; \frac{\partial^2 f}{\partial x_i^2} > 0 \tag{1.1}$$

The first partial derivative can be used to decide which way the function is going downhill for a particular variable. This, however, assumes that the function is derivable, which sometimes is not the case. Depending on the characteristics of the function, derivative methods (such as the steepest descent or the conjugate gradients method) or non-derivative ones (such as the simplex method) are used.

1.1.3 Docking simulation

In molecular docking simulations modelling techniques are used to predict how any macromolecule (typically a protein) interacts with other molecules (may be other proteins, nucleic acids or small drug-like molecules)[10, 11]. Small molecule-protein docking (usually known as protein-ligand docking) is the most frequently used as the

smaller dimensions of the system allow for a higher accuracy of the mathematical description and a higher reliability of the model. Moreover it has dramatically important applications in medicinal chemistry, where it is used as the base of structure based drug design approaches[12]. Protein-protein docking is still at its beginning, as its high requirements in terms of computational power has long made it too slow or too inaccurate [13]. However it is developing fast and has already scored some impressive results.

In a docking simulation a search algorithm is used to find the global minimum of a scoring function. The computational description of this search (the so-called docking problem) is usually unsuitable to be treated with deterministic algorithms because its complexity grows exponentially with the number of variables and is beyond the computational power of common hardware. The approaches used to solve this problem can be broadly classified as *fragment-based* methods, *stochastic optimisation* methods, or as *multiconformer docking* approaches [14].

- *Fragment based* approaches consist in splitting the ligand into different rigid fragments and perform an incremental construction of the solution. The docking of single fragments is much easier of the one of a full molecule, as they can be treated as rigid bodies. An approach of this kind is employed by FlexX [15].

- *Stochastic optimisation* methods imply stochastic (random) steps in the search to speed up the process. They cannot guarantee to find the best solution, but are used to find a good solution in a reasonable time. Examples of this class of search algorithms are the genetic algorithm used in GOLD [16] and autodock [17], Monte Carlo minimisation in GLIDE [18] and simulated annealing, evolutionary programming, and tabu search in PRO_LEADS [19].

- *Multiconformer docking* is performed by pre-generating a library of low-energy conformations for the input molecule and docking them as rigid bodies. This is

the approach exploited by FRED (OpenEye).

1.1.4 Virtual screening

Virtual screening experiments are based on the application of molecular docking techniques to large numbers of molecules on the same target. A common computer can dock a medium sized molecule with reasonable accuracy in few minutes and more powerful computers can further reduce this time. It has become common practice for pharmaceutical industries and universities to screen hundreds of thousands or even millions of drug-like candidates in this way to find a drug lead for a particular target. Each molecule-target interaction is scored and the best-scoring candidates are chosen for further investigation *in silico* or for synthesis and *in vitro* testing. Compared to traditional *in-vitro* high-throughput screenings, virtual screenings are orders of magnitude faster and less expensive[20, 21, 22, 23, 24].

A virtual screening process can be divided in three phases:

1. *Docking phase*: the binding pose for each molecule of the database is generated with molecular docking.

2. *Scoring phase*: every molecule in the database is scored with one or more scoring functions and the molecules are sorted according to their predicted affinity.

3. *Visual inspection*: the best-ranking molecules are inspected by the user and the most promising ones are selected.

1.1.5 *De novo* drug design

De novo drug design approaches use structural information on a specific target to build ligands which interact with the active site by means of hydrogen bonds, electrostatic and/or hydrophobic interactions. Information on the binding mode of known ligands, from protein-ligand complex structures, may also be used as a guide in the

process. There are three basic classes of computational approaches to *de novo* drug design:

- *Fragment placement* determines favourable binding positions for single rigid fragments, which aim to capture the basic molecular interactions in the active site. GRID is an example of an application of this approach: its algorithm generates a regular grid inside the active side and then places functional group probes (such as hydroxyl or methyl groups) at the vertices, scoring the interaction potential at each vertex and locating the best placements for each fragment.

- *Connection methods* attempt to place small molecules in the active site aiming to match site points (derived by rules or using the fragment placement method) where they would have favourable interactions. Alternatively, fragment connection methods search for scaffolds which are suitable to connect isolated fragments which have favourable interactions with regions of the active site.

- *Sequential growth* consists in the construction of a ligand, step by step, within an active site. This methodology is useful to generate new potential leads or optimising a known ligand by modifications introduced in its structure.

1.2 Molecular dynamics

While Molecular Mechanics is concerned with atomic interactions, three dimensional geometries and energies in a given moment, Molecular Dynamics (MD) aims to predict the evolution of a molecular system over time [25].

In MD simulations atoms and molecules are allowed to interact for a period of time under known laws of physics, generating a trajectory for the motion of the atoms. By solving Newton's equations of classical mechanics a MD simulation is capable of generating successive configurations of a system of molecules and propagating the

positions and velocities forward in time. It mimics the way molecules explore their conformational space in reality, being particularly suited for the conformational study of proteins or other large molecules[1].

The collection of forces F_i that act on each atom i of the system in study cause the latter to change in a way which follows Newton's second law of motion (Equation 1.2).

$$F_i = m_i a_i \tag{1.2}$$

Equation 1.2 can be rearraged, given that the acceleration is the second derivative of the displacement s_i with respect to time t_i yielding Equation 1.3

$$\frac{\delta^2 s_i}{\delta t^2} = \frac{F_i}{m_i} \tag{1.3}$$

Integration of equation 1.3 gives the displacement of each particle as a function of time, u_i being the initial velocity and a_i the acceleration caused by F_{ij} given by equation 1.4.

$$s_i = u_i t + \frac{1}{2} a_i t^2 + c \tag{1.4}$$

This is the key equation for MD simulations, allowing the calculations of the position of each atom in the system as a function of time.

Although molecular dynamics can be carried out *in vacuo*, unlike most molecular mechanics calculation it usually uses explicit solvent approaches. A box filled with water molecules is generated around the molecule to study, and periodic boundary conditions are applied (i.e. the box is surrounded by identical copies of itself, so whatever leaves it from one side enters back from the opposite). A typical MD system is shown in picture 1.3. Because molecular systems generally consist of a vast number of particles, it is impossible to compute the properties of such complex systems analytically; MD simulation circumvents this problem by using numerical methods.

An MD simulation usually involves 3 steps:

- **Energy minimisation**: the system is brought to a minimum of the energy landscape.

- **Solvent equilibration**: the system is allowed to evolve for a certain time with restrains on atomic positions. This allows the solvent to lose memory of its initial state.

- **Molecular Dynamics**: the proper molecular dynamics step.

The span of time that is simulated depends on various factor, the most importants being the computational power used and the time the simulation is allowed to run. Hower normal dynamics rarely covers more than a few tens of nanoseconds.

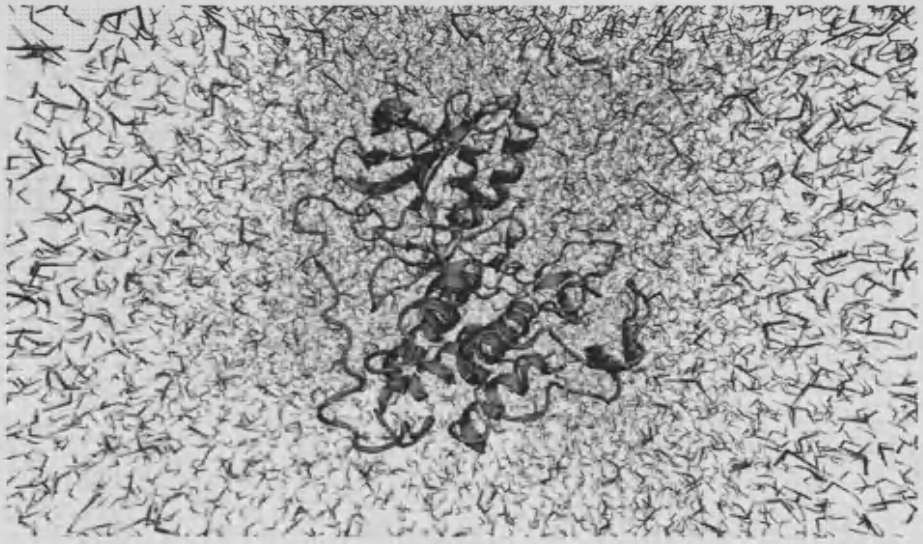


Figure 1.3: Example of a Molecular Dynamics system: Protein Kinase A in water

Chapter 2

Haptic technology in drug design

Drug design is an extremely creative process and displays several features that make it closer to human reasoning than to machine automation. The traditional way to tackle it, however, is often a black-box approach, with the user analysing structures proposed by a computer algorithm without any chance of intervention on the process of their generation other than the setting of parameters at the beginning. This often leads to a number of "false positives", with the computer wasting calculation time and display space on structures that are not chemically accessible, or that could be easily discarded by an average chemist for other reasons. Being able to avoid these wastes not only would save time, but would also allow a more complete exploration of the chemical or conformational space during a simulation, enriching it with high-quality results.

Current advances in hardware technology make it possible to think of approaching the problem from a different angle, putting the reins of drug design back into human hands. Ergonomics is the science of designing a process or a tool to fit the user. It is becoming a keyword in many aspects of life, from production tools to the entertainment industry. This project is aimed to the construction of an ergonomic Molecular Mechanics environment for interactive, user-driven de novo drug design and ligand optimisation, characterised by a trimodal feedback system. The goal is to create an environment where the user plays an active role in the simulation, and is able to steer

it in real time to bring his/her own contribution to the generation of the final result.

2.1 Haptic technology

Haptic, from the Greek $\alpha\varphi\eta$ (Haphe), means "pertaining to the sense of touch". A haptic stimulus is for the sense of touch what a visual stimulus is for the sense of sight. Haptic technology refers to technology which interfaces the user via the sense of touch by applying forces, vibrations and/or motions to him/her. This mechanical stimulation is used to create haptic virtual objects that can be touched and felt through an haptic device, just like visual virtual objects can be seen through a monitor.

Haptics is gaining widespread acceptance as a key part of virtual (or augmented) reality systems, adding the sense of touch to previously visual-only solutions. Most of these solutions use stylus-based haptic rendering, where the user interfaces to the virtual world via a tool or stylus, giving a form of interaction that is realistic and computationally affordable on today's hardware. Examples of successful implementation of haptic-based computer simulation include tools for the training of surgeons who can perform on virual patients or virtual sculpting tools for rapid prototyping.

2.2 Human-computer interactions

Human beings probe their environment via the five senses. Engaging the user's senses is the key to a good human-computer interface. Sight and hearing have traditionally been exploited, and touch has recently been added to augmented reality simulations. These three senses define three major communication channels (Picture 2.1) between mind and CPU. Each channel can be exploited for input to the computer (using sensors) or output to the user (using actuators).

- **Visual Channel**: This is probabily the most intuitive and widely used one. Computers have used monitors since their birth to convey visual feedback, and recently cameras and other devices have been employed as visual input devices.

- **Auditory Channel**: The use of speakers to play sounds is not recent. Microphones can also be used as auditory sensors. System sound alerts and game sound effects are examples of auditory output. Videogames use surround audio systems to help the player locate the position of a virtual object that cannot be directly seen.
- **Kinesthetic Channel**: This channel is linked to the sense of touch. Touch screens and even mice are examples of kinesthetic sensors. Haptic feedback is a recent application, and is usually provided by a haptic device.

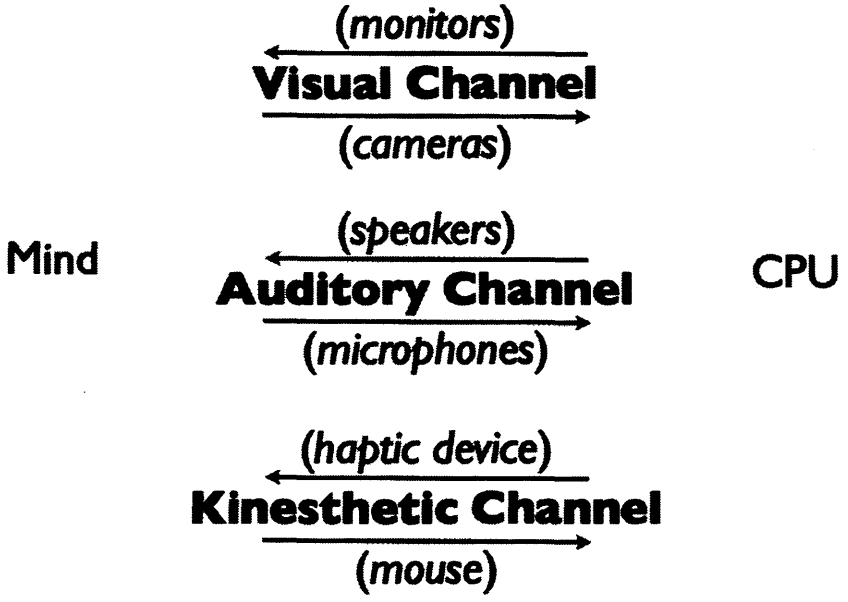


Figure 2.1: Human-computer interactions: the three main channels of communication between mind and CPU

Multimodal (interacting via more than one sensory channel) virtual environments have been demonstrated to convey a better sense of presence and a richer experience than single-sensory technologies. It has also been shown that response times of users to multisensorial stimuli are significatively shorter than the corresponding stimuli alone. Mental processing times of a trimodal signal (haptic, audio and visual) are

shorter than those associated to the processing of any bimodal or unimodal signals [26].

2.3 Previous applications to molecular systems

The possibility of using haptic technology to interact with chemical environments or simulations has been described before [27]. In previously developed software packages like the Chemical Force Feedback system [28] and HAMStER [29] the force feedback was successfully applied to different chemical systems (usually interactions between small molecules). However the way the problem was tackled in these applications and the need for computer power that was not available in the past made their extension to larger systems and their applicability to drug design problematic. In particular these systems relied on the use of pre-calculated potential grids and, as pointed out by Wollacott, this would limit the use of haptics on large biological systems like proteins [29].

Chapter 3

Aims

Molecular modelling and drug design are creative processes that display several features that make them closer to human reasoning than to machine automation. However, very often the user intervention is limited to the preparation of the input and analysis of the output of a computer simulation[30]. In some cases, allowing human intervention directly in the process could improve the quality of the results by applying the researcher intuition directly into the simulation. This project is aimed at the implementation of an interactive drug design environment based on haptic technology and its application alongside more traditional computational approaches to several drug design projects in the antiviral and anticancer fields. Besides comparing the approaches on general bases the goal is to test the performance of the interactive approach on actual research projects and understand if their combination with traditional approaches can be beneficial in terms of improving the quality of the designed structures.

Chapter 4

Implementation of a Trimodal Feedback Drug Design Environment

4.1 Zodiac

Zodiac is an open source molecular visualiser (www.zeden.org). It relies on openbabel (www.openbabel.org) for molecular files input and output and some chemical prediction features, openGL for the 3D rendering and Trolltech's QT4 for the GUI. At the time the present project was started zodiac was able to open a molecular file, load the atomic coordinates and display the molecule on the screen. Some options to tune the graphical output were possible, and bonds could be displayed as lines or cylinders, while atoms could be optionally rendered as space-filling CPK spheres.

Besides optimising the code for speed and reliability, zodiac needed the implementation of a series of visual as well as audio feedbacks along with the addition of the haptic code and the minimisation engine.

4.2 Core calculation cycle

A conceptual scheme of the implementation is shown in Figure 4.1. Algorithm 1 shows the pseudocode for the main calculation cycle.

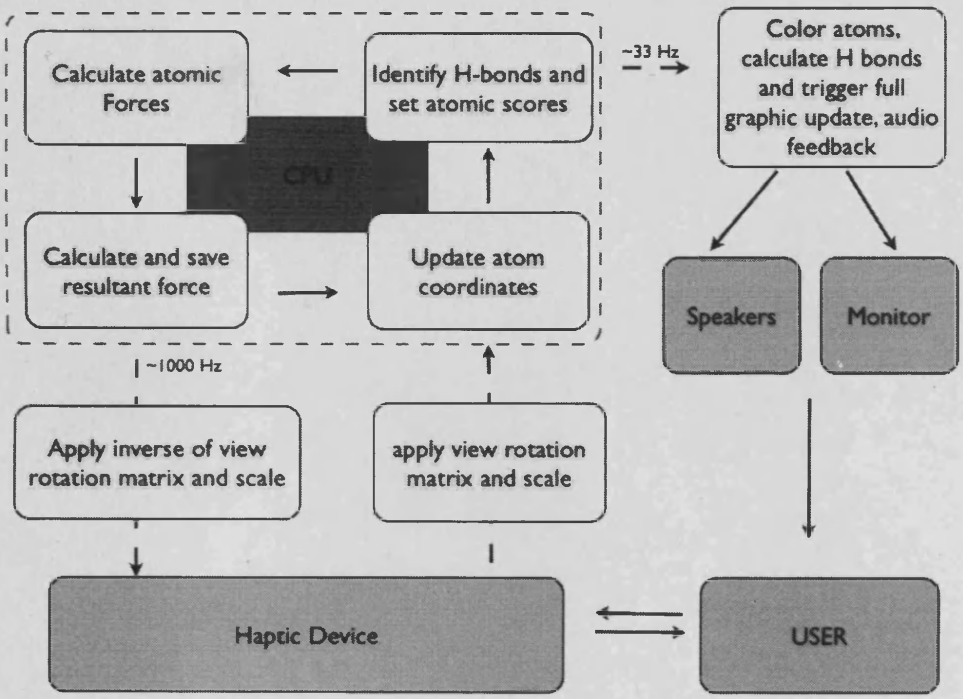


Figure 4.1: Conceptual scheme of the implementation: the core cycle is shown along with the three interaction channels between user and CPU

Inside the molecular environment, a set of atoms (typically a small molecule) are defined as a probe (ligand). All the remaining atoms form the environment. The molecular scene can be rotated via the use of a virtual trackball. The rotation is stored into a matrix. For every frame of the simulation, the movements of the haptic device are translated into coordinate shifts of the probe. The inverse of the rotation matrix is applied to this translations to make sure that the reference system of the probe remains the same as the user's, despite the rotations of the view. A forcefield

Algorithm 1 Calculation cycle

```
UpdateLigandPositionWithHapticPointer ()
ComputeInternalInteractions()
for i = 1 to ligandatoms do
  GetNeighbourList (atom_i)
  CalculateLigandProteinInteractions (atom_i)
end for
ComputeTotalForce ()
ApplyForces ()
```

calculates interactions internal to the probe and between probe and environment and sets a force on each atom of the probe accordingly. Each atom of the probe is then moved proportionally to its force. The total force on the probe is computed, smoothed (see section 4.6), rotated using the view rotation matrix and set on the device. Visual and audio feedbacks are also calculated. The total energy of the system, the interaction energy and the inner energy of the probe are calculated. If the energy values and the coordinates of the atoms satisfy the threshold and clustering limits set by the user, the pose is saved in molecular database for further analysis at the end of the simulation along with the three corrisponding energy values. The GUI features a button for the user to force the saving of a pose he/she likes.

4.3 Visual, auditory and haptic feedback

The resulting force on each of the ligand's atoms is applied to the user's hand via the haptic device. The user has therefore a direct feedback on the degrees of freedom he/she can control (at least for translations. It is theoretically possible to do the same for torsions applying torques to the user, but that involves the use of a much more expensive device, which hasn't been used in this project yet). At any moment the user can decide to let the simulation evolve or can oppose his/her own force and guide it in the direction he/she prefers.

To improve the experience of the user the visual feedback has also been improved. Zodiac is capable of exploiting 3D projection and visualisation devices to give a full 3D experience. Hydrogen bonds are also detected and marked in real time, and shown with lines. The opacity of the line is proportional to the "strength" of the bond (in terms of deviation from optimal distance and angle, as defined in chemscore). It is therefore possible to see lines fade in and fade out as the corresponding hydrogen bonds are formed or broken. Atoms are usually coloured according to their atomic number. However it is possible to choose an energy-based colour coding. This will colour atoms with different shades of colour according to their predicted interaction energy. While the simulation runs atoms will change their shade of colour as they move towards a more favoured position, or will turn a different colour if they are forced into a clash or another energy-disfavoured interaction. It is also possible to define water accessible surfaces (Connolly surfaces) to define the active site shape, and clipping planes that allow to work with ligands in narrow pockets.
Audio feedback is also provided in the form of sound alerts that mark when certain favourable conditions set by the user are met. Quantitative information can also be encoded in the audio by modulating the volume of a soundtrack.

4.3.1 Implementation of visual feedback

Backbone

Computing higher-level graphical representations from a set of atom coordinates is crucial to outline the 3D organisation of macromolecules. Protein backbone trace is a very useful guide in understanding folding and spatial organisation. An adaptation of the algorithm described by Pristle[31] has been implemented in Zodiac to calculate backbone traces. For each residue of a target protein the midpoint of the two peptidic bonds with the preceding and following amino acids are stored in a list. When an amino acid is not forming a peptidic bond, the coordinates of the α nitrogen or of

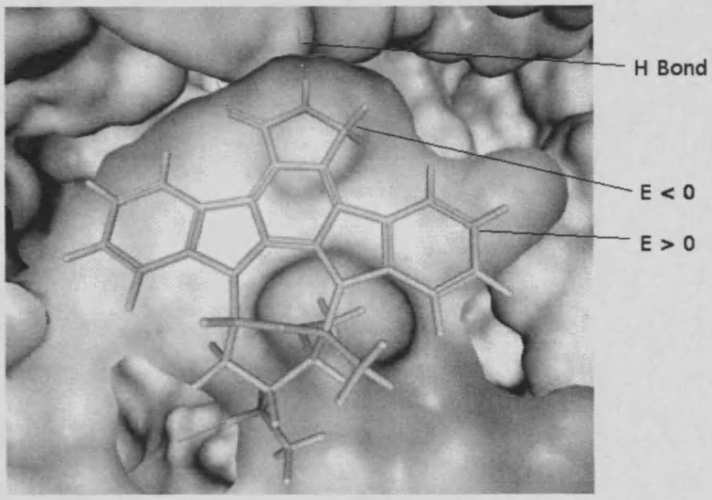


Figure 4.2: Visual feedback: the per-atom color-coded scores are shown, along with H-bond depiction and clipping planes

VISUAL	
Shape and structures	backbone trace, molecular surfaces
H Bonds	H bonds highlighting
Scores	per atom color coded scores
total energy	energy value display
Potential fields	potential isosurfaces
3D space and depth	stereo display
AUDIO	
total energy	soundtrack volume
choice of a new good pose	sound alert
HAPTIC	
total force on the ligand	force on the haptic device

Table 4.1: Multisensorial feedback

the carbon of the α carboxylic group are used. Intermediate points are interpolated between points in the list iteratively using the following approach (see Figure 4.3):

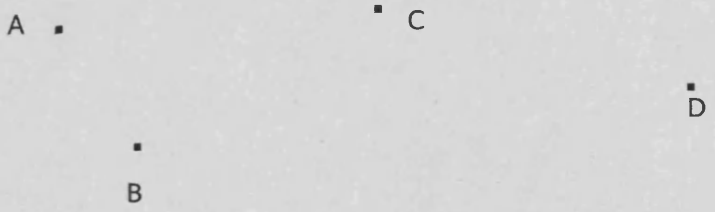
- The penultimate point from the list of the preceding residue and the second point from the list of the following residue are added to the list. If these points cannot be found they are extrapolated from the coordinates of the first two and last two points of the list respectively. These points are used as guides to make sure that the junction between one the backbone line of one residue and the following is smooth.

- For each three consecutive points ABC in the list, two points I and L are interpolated, so that A, B, C, I and L all stand on the same circle and the arc AI is equal to IB and BL is equal to LC. If A B and C are colinear, then I and L are simply the midpoints of AB and BC respectively.

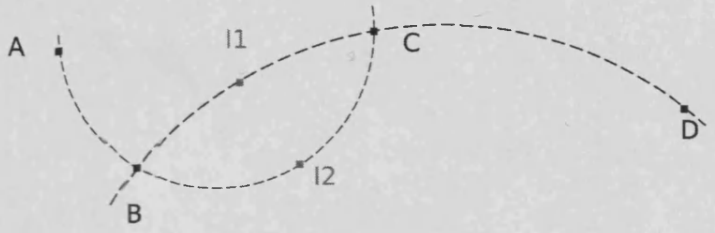
- If the list has more than four points, between all pairs of points but the first pair and the last pair two points will be generated (if the points are A B C D, between B and C there will be a point on the circle ABC (I1) and one on BCD (I2)). The midpoint between these two interpolated point is added to the list in the appropriate position.

- the first and last points are removed from the list.

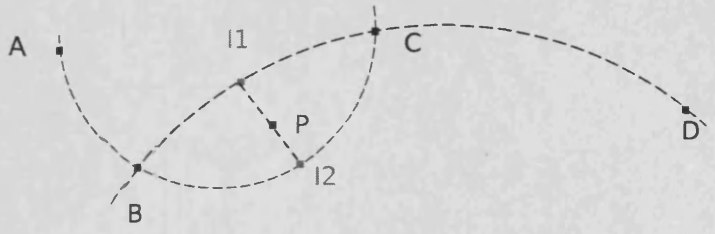
Each cycle generates $k-1$ new points where k is the number of points in the list. The number of total points is $2^n + 1$ where n is the number of smoothing cycles. 3 or 4 cycles are usually sufficient to generate the impression of a continuous curved line. The backbone is either rendered as a line connecting each point in the residues lists, or as a tube (Figure 4.4). For tube representation for each point in each list a "modified circle" is computed with the equation:
$X = bsin(\alpha) - bcsin^3(\alpha)$
$Y = acos(\alpha)$
$Z = 0$
Where the Z axis is aligned along the backbone line and X gives the direction of the



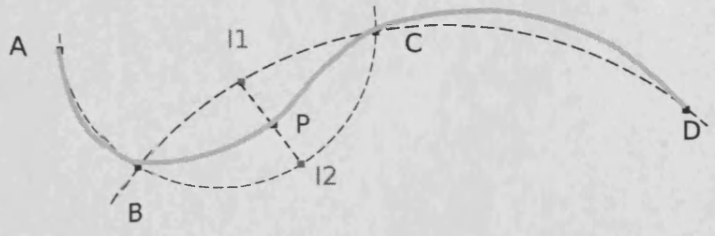
(a) Four consecutive points in a list A B C D



(b) Circles for ABC and BCD are identified. I1 and I2 are the midpoints of the BC arcs on the two circles.



(c) P, midpoint of I1I2 is added to the list.



(d) Iteration of this process gives a list of points that form a smooth line still passing for A B C and D

Figure 4.3: Backbone points interpolation

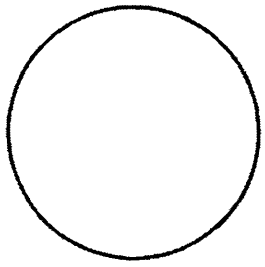
C=O bond in the backbone, with $\alpha \in [0, 2\pi[$. For Helices and β-sheets a = 0.3 b = 1 and c = 0.9 are used by default, to give the impression of a flat ribbon, while for loops a = b = 0.3 and c = 0 yeld a circle (Figure 4.5) a, b and c values for loops, α-helices and β-sheets can be easily customised through the preferences menu.

(a) Line (b) Tube

Figure 4.4: Backbone representations

Surfaces

The perception of the three-dimensional structure of a macromolecule is very difficult without the aid of a graphical representation that helps the user understand its shape. The implementation of such representations is therefore vital for any molecular modelling software that aims to a successful interaction with the user. The generation of two different types of surfaces was implemented in Zodiac for this project:



(a) loops section: a = 0.3 b = 0.3 c = 0.0

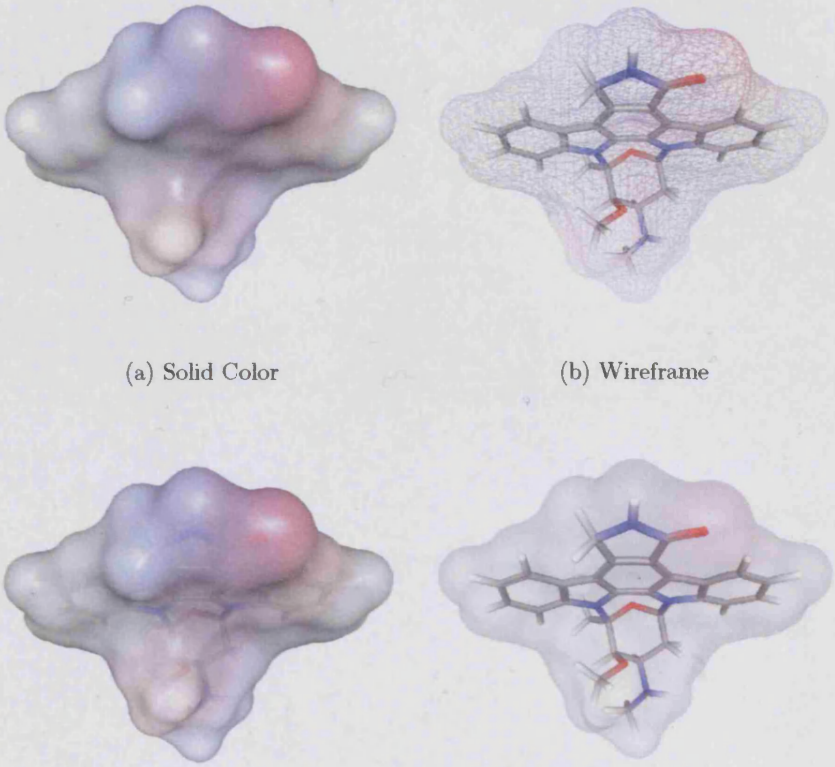
(b) helices and sheets section: a = 0.3 b = 1.0 c = 0.9

Figure 4.5: Backbone sections

- **Chemscore clash map:** This surface is very fast to generate and gives a general idea of the shape of a set of atoms. The target molecule is enclosed into a grid, with a resolution set by the user, 4 angstrom bigger than the molecule in each dimension, on each side. For each node of the grid the chemscore clash value of a probe atom is calculated and stored. Tetracoordinate atoms are not considered for this calculation, as they should not contribute to the surface for geometrical reasons. A Marching Cubes algorithm is then run on the grid with a threshold value of 0. For each cube of the grid this routine analyses the 8 values at the corners. Unless all of them are either bigger or smaller than the threshold value, the region where the threshold is met is estimated by means of a look-up table, and one or more triangles are computed on the surface that splits the cube into a "bigger than threshold" and "smaller than threshold" area. The resulting set of triangles from each cube form an isosurface where the clash term of chemscore is approximately null, which gives a good idea of the shape of the molecule.

- **Solvent accessible surface:** Also known as Connolly surface, is more precise and delimitates the portion of space that is not accessible to a molecule of the solvent (generally water). This is done in two steps. First a grid is built similarly to the one used for the chemscore clash maps. On each node of the grid the distance from the nearest atom (distance from the center minus the Van der Waals radius of the atom) of the molecule is calculated (tetracoordinated atoms are not considered). An isosurface of value 1.4 is generated with the same marching cubes algorithm described earlier. This surface represents the positions of the center of a sphere of radius 1.4 angstrom that is rolled onto the molecule. A second isosurface is generated of points that are 1.4 angstroms away from the first surface, and less than 1.4 angstroms from an atom of the protein.

Each surface can be visualised either come solid (or transparent) surface or as a wireframe (Figure 4.6).



(a) Solid Color

(b) Wireframe

(c) alpha = 0.8

(d) alpha = 0.2

Figure 4.6: Surface display modes

Potential maps

Potential maps allow to calculate the potential (energy that a probe atom would have if was placed in a particular region of space). A map can be derived from any energy function and is usually rendered by means of one or more isosurfaces. Just like molecular surfaces, potential isosurfaces can be transparent, solid color or wireframe. Lipophilic maps help identify hydrophobic pockets and are useful when working with lipophilic ligands or fragments, Hydrogen bond potentials can help direct the design and positioning of hydrogen bond donor or acceptor moieties. Electrostatic maps are useful when working with charged groups and exploring the possibility for the formation of salt bridges.

An example of potential map is shown in Figure 4.7

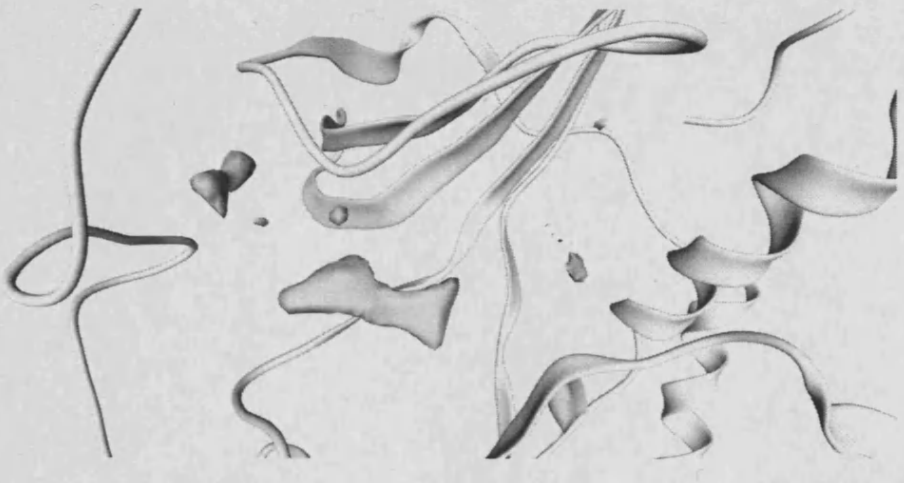


Figure 4.7: Potential maps

Managing transparency

Transparent surfaces are very useful because they allow the user to see which atom is closest to a particular region, which is vital when probing a chemical environment. OpenGL uses an RGBA (red green blue alpha) system for the representation of colors. Setting the alpha value to less then 1. and enabling alpha blending allows for

the display of transparent objects. However, due to the way alpha blending works, transparent objects need to be drawn on the screen in the correct order, from the furthest away from the screen to the nearest. This order depends not only from the coordinates of each object, but also from the rotation of the viewport, and needs to be recomputed every time the scene is rotated. To do this in a time-efficient way a pigeon hole approach was used in Zodiac. After every non transparent object has been drawn, a rapid check among each object flagged as "transparent" is done, to find the distances from the screen at which the region containing transparent objects starts and ends. This region is sliced into N subregions of fixed width, or NMAX subrregions of appropriate width if N is greater than NMAX. Each transparent triangle is assigned to the appropriate region, according to the distance of its center from the screen. Each region is then rendered, starting from the furthest away from the screen.

Stereo display

The perception of depth is very important when working with a three-dimensional environment such as a molecular modelling simulation. To convey the impression of the third dimension on a bidimensional surface such as a monitor or a projection screen it is necessary to bring two different images on each eye. One popular way to do this is using different light for the two images (different polarisation is often used). Special glasses are worn that can block one of the two images on each eye. The two images can be projected on the same screen or can be composed into an interlaced image where odd and even rows (or columns) of pixels originate from the two different sources.

To exploit 3D system the graphic scene needs to be rendered twice. The scene for the left eye is rendered by translating the camera to the left and rotating slightly to the right, and the opposite is done for the right eye. For stereo-enabled graphic cards (such as the NVIDIA Quadro series) four buffers are used (back and front left, back and front right). The two different scenes are drawn on the left and right back buffers. For interlaced stereo systems, a single image needs to be constructed. With

the use of the stencil buffer first alternated rows (or columns) are filled with one of the pictures, then the remaining rows (or columns) are filled with the other.

4.3.2 Implementation of auditory feedback

While the user is engaged in the simulation the total energy of the complex is printed out on screen. Following this number as it changes over time can be distracting because the user is forced to look away from the molecule he/she is working with and also keeping track of the changes of a nuber that is updated 30 times a second is not an easy task. To help convey the energy information in a more natural way two kinds of auditory feedback has been implemented.

In the first one a simple sound alert is played whenever a new pose is saved (see chapter 4). This gives direct information about if a particular region of the binding pocket is yielding good results.

The second one provides more precise information about the absolute value of the interaction energy and his variation over time. The user selects one or more audio files (.mp3 .wav and any other format that is supported by the underlying system is accepted) to be used as a soundtrack. The files are played one after the other while the simulation runs, and the volume is linked to the relative value of the interaction energy with respect to the highest interaction value found up to that moment (See Equation 4.1).

$$volume = \begin{cases} 1 & if E < E_{best} \\ 0 & if E > E_{best} + tolerance \\ \dfrac{10^{\frac{E_{best}+tolerance-E}{tolerance}}}{10} & if E_{best} < E < E_{best} + tolerance \end{cases} \tag{4.1}$$

Where volume = 0 is no sound and volume = 1 is the loudest volume the current settings of the underlying operative system can output. The exponential function

allows for bigger changes in volume when the energy is close to the best one observed.

4.3.3 Implementation of haptic feedback

The setting of the feedback force on the user hand is highly dependent on the device and the haptic library that are being used. When the value for a new force vector is generated from the computation cycle, it is stored in a variable. The haptic code that is responsible for the actual setting of the force to the device runs on a separate looping thread, using each time the latest available value for the force. This allows for a smoother feedback (see chapter 4.7) and an easier implementation of different haptics libraries. Three different haptic libraries have been implemented:

- **gHaptics** is the library provided by Novint. It is compatible with the Falcon device only, and only available for windows.
- **OpenHaptics** is developed by SensAble. It is cross-platform and supports the whole range of sensable devices.
- **HAPI** is an open source library, part of the H3DAPI library, developed and distributed by senseGraphics. It acts as a layer between the code and a series of different proprietary libraries, including gHaptics and Ghost (another of sensable's SDKs). It is therefore able to drive Novint and Sensable devices among many others.

4.4 Degrees of freedom

The number of degrees of freedom of a problem is the minimum number of independent variables that are necessary for its mathematical description. The scoring of the interaction of a single rigid molecular fragment and a receptor depends on six parameters (three coordinates for the translational vector and three angles for rotations to position the ligand with respect to the protein). Accounting for internal flexibility of

ligand and protein results in a higher number of degrees of freedom.
In this application two different models were implemented:

6 DoF model

This is the simplest possible model. Both the ligand and the protein are considered as rigid bodies. This simplification allows not only to reduce the number of degrees of freedom to 6, but also allows the internal energy of ligand and protein to be considered constant during binding, thus reducing the computation of the binding energy to the determination of the interaction energy alone, neglecting the change in the internal energy in the binding partners. The main drawback of this type of model is that binding affinity is highly conformation-dependent and molecules in general simply cannot be considered completely rigid. This approach, though, is still useful in fragment-based methods, as molecular fragment such as fused aromatic rings or carboxylic groups can be treated succesfully as rigid entities.

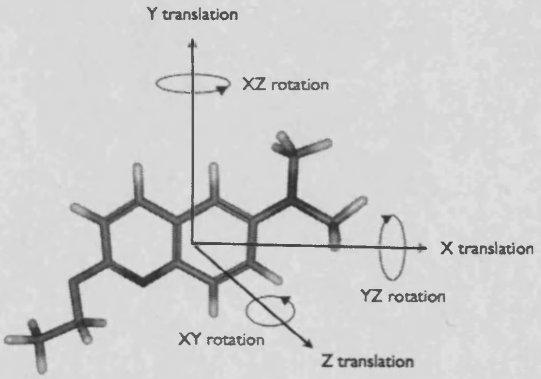


Figure 4.8: 6 degrees of freedom model

3N DoF model

This is the most accurate possible model. Each atom of the system is allowed to

move freely (the number of degrees of freedom is three times the number of atoms, as each of the three cartesian coordinates of each atom is independent from the others). This way though, internal energies cannot assumed to be constant any more, and contribution such as bond stretchings energy or bond torsion energies must be computed. This is the way molecular systems are usually treated during energy minimisation and in molecular dynamics simulations.

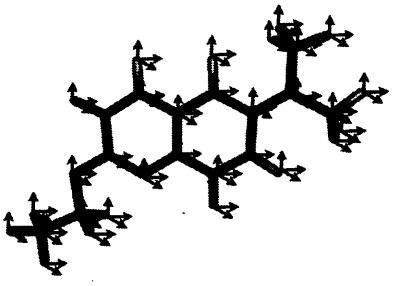


Figure 4.9: 3N degrees of freedom model

A third model is possible:

6+r DoF model

This is an intermediate approach between the two previous models. Bond lengths and angles are assumed to be constant, while rotations around single bonds (usually only for ligand rotatable bonds) are allowed. This brings the number of degrees of freedom to six plus the number of rotatable bonds in the system. The computation of internal energies can be simplified, since the contribution from bond stretchings and angle bendings are constant. Other terms such as torsional energies of the rotatable bonds must still be computed. This model is the one most docking programs imple-

ment as the simplification of fixing bonds stretching and bendings is a slight one and allows for a drastic increase in the speed of the algorithm. A drawback of this sort of model is that it is more difficult to implement, as it requires a kinematic chain to be built for the molecule and inverse kinematics application to have it correctly respond to an outer force, which in the particular case of this implementation would render it probably slower than the previous, more precise one.

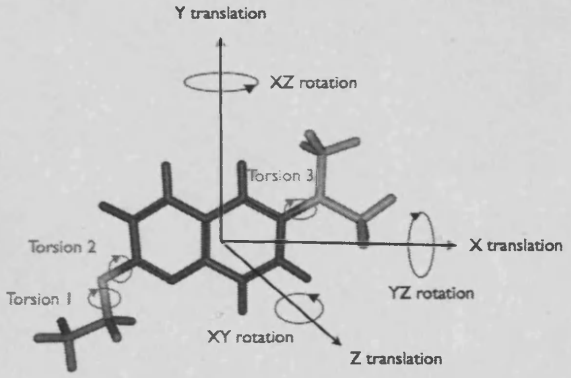


Figure 4.10: 6+r degrees of freedom model

4.5 Minimisation

Depending on the type of haptic device, the user is able to directly control 3 or 6 degrees of freedom. As the model has at least 6, and usually more, it is necessary to find a way to control the others. A double input such as haptic-keyboard or haptic-mouse could be used in the 6 dof model, but it is clearly impossible to let the user control all the degrees of freedom in an interactive way in a more accurate model. Energy refinement techniques have been employed to overcome this problem. Three

or six "major" degrees of freedom (the three components of the translation vector and optionally the three rotation angles applied to the ligand) are linked to the haptic pointer position, and are therefore controlled directly by the user, while the CPU controls the others. A steepest descent-like method was chosen, computing at each step the first derivative of the energy with respect to each degree of freedom and taking for each a step proportional to the opposite of the derivative. Usually in molecular mechanics minimisations these steps are biased to improve the speed of the algorithm in finding the local minimum. In this application however minimisation does not have to be too fast because otherwise it would appear unnatural. The goal is achieving a system capable of reacting at the user's inputs in realtime, slowly enough to let him or her the time to understand in which direction it is evolving and have an idea of the energy landscape in a particular region. Physically, our minimisation is equal to the system moving with a speed proportional to the force to which it is subjected. It is import to point out that this is not molecular dynamics. Dynamics employs Newton's law of motion, in which a force causes an acceleration. In molecular dynamics concepts like kinetic energy and inertial forces have a significance and allows to simulate the actual movement of molecules. In this implementation the system evolves according to Aristotle's laws of motion, which states that a force causes a velocity. Therefore an atom in motion subjeted to no force will stop its motion immediately, as opposed as continuing with constant speed.

4.6 Viscous forces

The calibration of the scaling constants that translate predicted forces to the actual one that is applied to the user is not trivial. The parameters to take into account are different and sometimes difficult to evaluate: if the force is not strong enough the user might not feel the weaker interactions, and is also allowed to move atoms in closer proximity, where the van der waals function is steeper and therefore more difficult to control, making the simulation unstable. If, on the other hand, the force is too strong it might damage the device, and makes it difficult for the user to control the

simulation (which will converge to the closest strong interaction, such as a salt bridge, and trap the device there). The size of the virtual working space is also important. The haptic working space is mapped to a cube in the virtual space. If this cube is too big, a small change in the device's position will result in a big change in the ligand's coordinates, and therefore a big change in its environment and an unexpected change in the force applied to the device. This in turn will cause an even bigger movement of the user's hand and start a vicious circle. If on the other hand the mapped space is too small it becomes hard to explore the binding site effectively. One problem that often arised during the parameterisation was the interaction with narrow pockets. If the clash with one wall of the pocket is strong enough to send the ligand to clash with the opposite one before the user has time to react and oppose his force to stop the motion, the ligand will likely start to bounce from one end to the other, even if the space in the middle is big enough to host it. Besides reducing the virtual working space and increasing the force scaling constant, a "viscous force approach" was used. This consists in considering not only the present value, but also the recent past ones when the force is calculated. Whenever a new value for the total force is obtained, it is pushed into a queue. If the size of the queue is bigger than a fixed value, the least recent value is popped out and the force to apply is then calculated from the mean of all the values in the queue. This allow for forces to change more gradually and lets the user more time to react to sudden changes (Figure 4.11). Another approach has been implemented, which does not require a queue and is faster. The force applied to the device is computed as shown in Equation 4.2, where $Flast$ is the last force that has been set on the device and $Fnew$ is the latest computed force.

$$F = (1 - K) * Flast + K * Fnew \tag{4.2}$$

4.7 Asynchronous multithreading

In order to be perceived as continuous a set of haptic stimuli needs a very high frequency of update (around 1KHz as opposed to 30Hz for visual stimuli). This is due

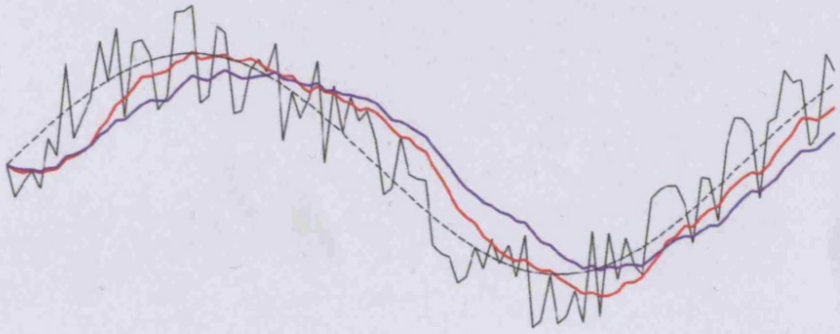


Figure 4.11: Viscous Forces: black continuous line derives from black dashed line with random noise. Red line represents the corrisponding values of the queue approach (size of the queue = 10), blue line refers to the values of the second algorithm (K = 10)

to the fact that the sense of touch is far more efficient than sight in distinguishing two separate impulses. This high requirement is traditionally a challange in haptic programming, requiring the calculations of the force to be carried out in less then one millisecond for each update. In the present implementation the task of computing all the atomic interactions involved in molecular recognition and setting the corrisponding forces in this time window appeared impossible on a single processor. The solution that was adopted involves the use of asynchronous threads. The main thread computes the interactions, setting the forces on each atom and moving them accordingly. Every 30 ms a visual update is triggered on a second thread. This assures that the the video output looks smooth, avoiding useless computations. A third thread is exclusively dedicated to the haptic feedback, feeding the device with the latest data computed by the main thread. This allows the haptic feedback to feel smooth even when the complexity of the system slows the computation thread below 1kHz. Since the time required by the haptic thread computation is independent of the number of atoms, the haptic update should continue to be adequate even with complex systems.

4.8 Implementation of the forcefield

A molecular mechanics environment is characterised by one (or more) set of equations that describe how it behaves and evolves. This sets of equations are called forcefields. Many different forcefields have been developed and described in the literature, each one representing a simplification of the laws of physics and chemistry aimed to the reproduction of particular molecular characteristcs. The forcefield chosen for the Molecular Mechanics environment is MMFF94s.

MMFF94 is one of the most frequently-used force fields in molecular mechanics. Developed by Thomas Halgren at Merck ([32, 33, 34, 35, 36]), it is an all-atom general-purpose force field, meaning that its parameterisation is meant to account for its application to both small molecules and biomacromolecules. Its s (static) version has slightly modified parameters that are meant to reproduce "averaged" conformations. The main difference is in the conformation of the nitrogen in amines, which should be tetragonal but is characterised by a low inversion barrier, and is therefore considered to be planar (average between the two opposite conformations). The general equation for the energy calculation is made up of seven components (Equation 4.3). The first five are bonded contributions, and are used in intramolecolar energy calculations, while the last two correspond to nonbonded interactions, and are applied to both inter and intramolecular calculations alike.

All the equations are parameterised to give energy predictions in kCal.

$$E = \sum EB_{ij} + \sum EA_{ijk} + \sum EBA_{ijk} + \sum EOOP_{ijk;l} + \sum ET_{ijkl} + \sum EvdW_{ij} + \sum EQ_{ij} \tag{4.3}$$

STRETCHING ENERGY

EB_{ij} (Equation 4.4) is the energy associated with the stretching of the bond between atoms i and j. It is calculated with a quartic function from Δr_{ij}, the deviation of the interatomic distance from a reference value (depending in turn on the two atom types).

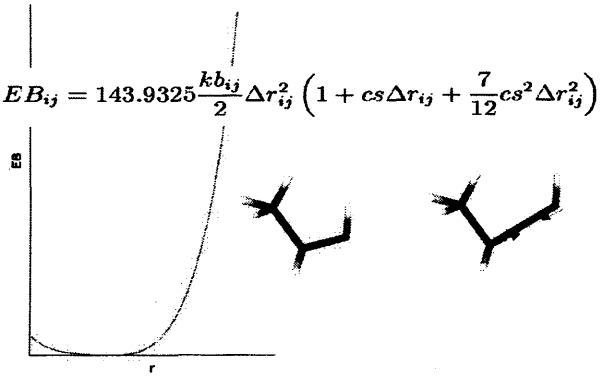


Figure 4.12: MMFF: Bond stretching function

$$EB_{ij} = 143.9325\frac{kb_{ij}}{2}\Delta r_{ij}^2\left(1 + cs\Delta r_{ij} + \frac{7}{12}cs^2\Delta r_{ij}^2\right) \tag{4.4}$$

kb_{ij} is a constant that depends on the atom types involved in the bond, while cs (cubic strenght) is a constant value.

BENDING ENERGY

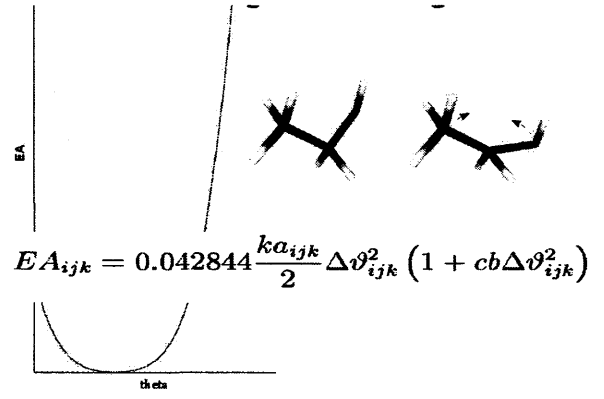


Figure 4.13: MMFF: Angle bending function

EA_{ijk} This term refers to the energy associated with binding angle bendings (Equation 4.5).

$$EA_{ijk} = 0.042844 \frac{ka_{ijk}}{2} \Delta\vartheta_{ijk}^2 \left(1 + cb\Delta\vartheta_{ijk}^2\right) \tag{4.5}$$

where $\Delta\vartheta_{ijk}$ ís the difference between the angle formed from atoms ijk and a reference value, ka_{ijk} is the force constant depending on the atom types involved and cb a constant term.

STRETCH-BEND INTERACTIONS

$$EBA_{ijk} = 2.51210 \left(kba_{ijk}\Delta r_{ij} + kba_{kji}\Delta r_{kj}\right) \Delta\vartheta_{ijk}$$

Figure 4.14: MMFF: Stretching-bending function

This term (Equation 4.6) relates the stretching of two bonds that share an atom j to the angle ϑ_{ijk} they form.

$$EBA_{ijk} = 2.51210 \left(kba_{ijk}\Delta r_{ij} + kba_{kji}\Delta r_{kj}\right) \Delta\vartheta_{ijk} \tag{4.6}$$

OUT OF PLANE BENDING ENERGY

This term (Equation 4.7) is used for penalising non-planar tricoordinate atoms (like aromatic carbons for instance).

$$EOOP_{ijk;l} = 0.043844 \frac{koop_{ijk;l}}{2} \chi_{ijk;l}^2 \tag{4.7}$$

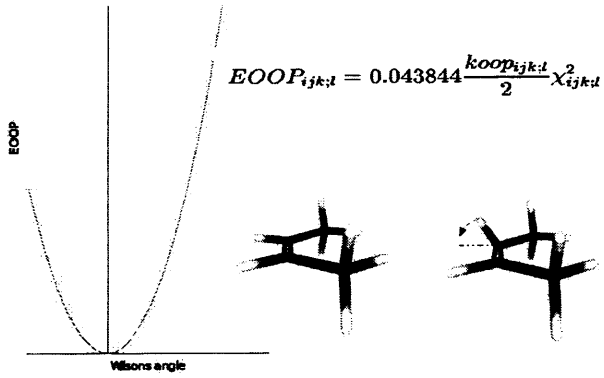


Figure 4.15: MMFF: Out of plane bending function

$\chi_{ijk;l}$ is the Wilson angle between $j-l$ bond and $i-j-k$ plane.

TORSIONAL ENERGY

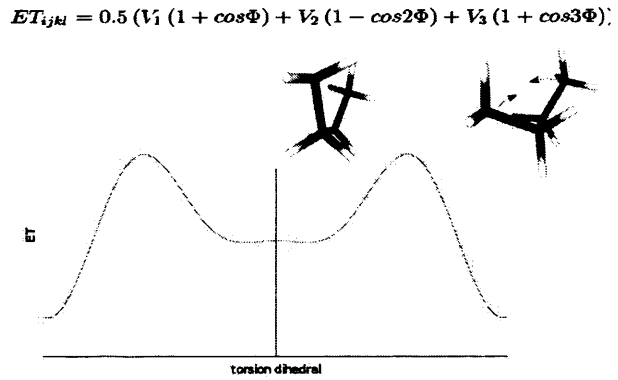


Figure 4.16: MMFF: Torsional energy function

This term estimates the energy associated with torsion around bonds. It exploits an empiric periodic equation of the dihedral angle defined by four atoms bound to one another to estimate the torsional contribution of the bond between central two (Equation 4.8). The three constants V_1, V_2, V_3 depend on the atom types of the four

atoms.

$$ET_{ijkl} = 0.5\,(V_1\,(1 + cos\Phi) + V_2\,(1 - cos2\Phi) + V_3\,(1 + cos3\Phi)) \tag{4.8}$$

VAN DER WAALS ENERGY

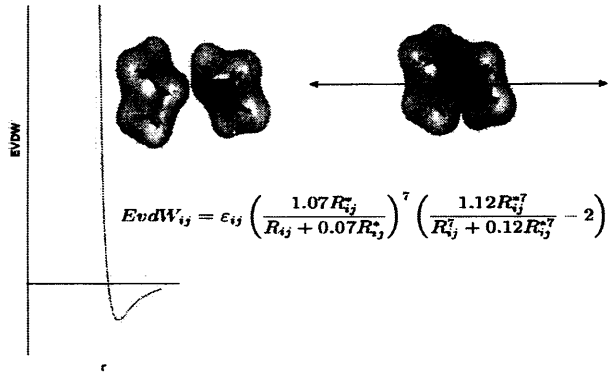


Figure 4.17: MMFF: Van der Waals interaction function

The equation that computes Van der Waals interactions is usually made up of two parts: an attractive and a repulsive one, both depending on negative powers of the interatomic distance (generally -14,-7 -12,-6 or -8,-4). The difference between the two exponents causes the repulsive component to prevail at short distances, while as the interatomic distance increases the attractive component becomes stronger. An equation of this kind is named Leonnard-Jones equation. In MMFF the Van der Waals term is computed with the -14/-7 Leonnard-Jones-like function (Equation 4.9).

$$EvdW_{ij} = \varepsilon_{ij}\left(\frac{1.07R^*_{ij}}{R_{ij}+0.07R^*_{ij}}\right)^7\left(\frac{1.12R^{*7}_{ij}}{R^7_{ij}+0.12R^{*7}_{ij}}-2\right) \tag{4.9}$$

ELECTROSTATIC INTERACTIONS

The electrostatic interaction between two atoms is computed with a Coulomb-like equation (Equation 4.10). δ is a small positive constant, that avoids an infinite result when the interatomic distance becomes 0.

$$EQ_{ij} = 332.0716\frac{q_i q_j}{D\,(R_{ij}+\delta)} \tag{4.10}$$

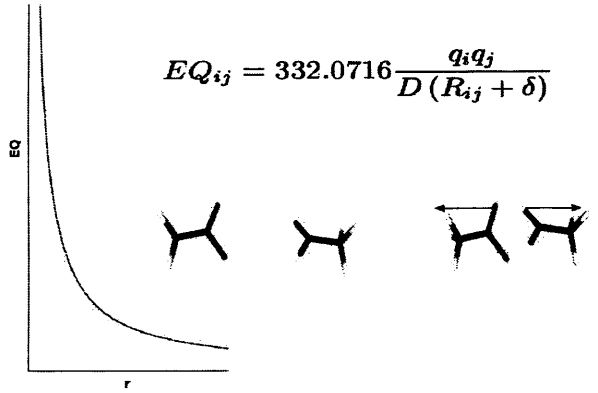


Figure 4.18: MMFF: Electrostatic interaction function

4.9 Implementation of the scoring functions

MMFF functions are not optimised to reproduce interaction energies in implicit-solvent approaches like the one reported here. Hydrogen bonds are considered non-directional and electrostatic in nature, and some of the main interactions that are considered to drive ligand-protein binding, such as hydrophobic interactions, are not considered. For these reasons a hybrid model was built, which exploits both MMFF and an empirical scoring function, in addition to the purely force field based one. In this model bonded and nonbonded energy contributions internal to the ligand are computed by MMFF, while nonbonded interaction energies between the ligand and the target are estimated with an empirical scoring function. A large number of scoring functions have been developed for docking purposes, differing for parametrisation and complexity. Two of the most widely used empirical scoring functions (PLP and Chemscore) were implemented in the software.

4.9.1 PLP

PLP [37] (Piecewise Linear Potential) is an extremely simple function. It is a heavy-atom-only function (hydrogen atoms are not considered). Heavy atoms are assigned one of the following atom types:

	Donor	Acceptor	Both	Non polar
Donor	Repulsive	H-bond	H-bond	Steric
Acceptor	H-bond	Repulsive	H-bond	Steric
Both	H-bond	H-bond	H-bond	Steric
Nonpolar	Steric	Steric	Steric	Steric

Table 4.2: PLP interaction types

- H-bond Donor
- H-bond Acceptor
- H-bond Donor/Acceptor
- Nonpolar atom

An interaction type is assigned to each protein-ligand atom couple according to Table 4.2.

The intensity of the interaction is computed with the simple function shown in Figure 4.19.

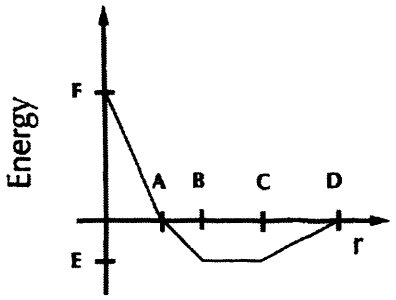


Figure 4.19: PLP energy function

A, B, C, D, E and F parameters for each interaction type are shown in Table 4.3. A, B, C and D value are expressed in Angstrom, E and F in kcal/mol.

	A	B	C	D	E	F
H-bond	2.3	2.6	3.1	3.4	-2.0	15.0
Steric	0.93	r_1+r_2	1.25 B	1.5B	-0.4	15.0
Repulsion	3.4	-	-	-	0.0	15.0

Table 4.3: PLP parameters for the energy function

4.9.2 Chemscore

Chemscore is an all-atom scoring function [38]. We implemented its version optimised for docking as described in [16]. The interaction energy in the binding of a ligand to a protein is estimated with Equation 4.11:

$$\Delta G_{binding} = \Delta G_{hbond} S_{hbond} + \Delta G_{metal} S_{metal} + \Delta G_{lipo} S_{lipo} + E_{clash} \tag{4.11}$$

where S_{hbond}, S_{metal} and S_{lipo} terms all make use of a block function, f, described in Equation 4.12

$$f(x, x_1, x_2) = \begin{cases} 1 & x \leq x_1 \\ \dfrac{(x_2 - x)}{(x_2 - x_1)} & x_1 < x \leq x_2 \\ 0 & x > x_2 \end{cases} \tag{4.12}$$

where x is a running variable, and x_1 and x_2 are constants controlling the fall-off of f. The hydrogen-bond term is calculated for each complementary combination of donor and acceptor on ligand and protein. It consists of a distance and an angle-dependent part (Eqation 4.13).

$$S_{hbond} = \Sigma_{DA} f(\Delta r_{DA}, \Delta r_1, \Delta r_2) f(\Delta \alpha_{DA}, \Delta \alpha_1, \Delta \alpha_2) \tag{4.13}$$

with $\Delta r_{DA} = |r_{DA} - r_0|$ and $\Delta \alpha_{DA} = |\alpha_{DA} - \alpha_0|$, where r_{DA} is H bond distance and α_{DA} the H bond angle (D-H A angle). r_0 and α_0 are the ideal hydrogen bond distance and angle. The metal (Equation 4.14) and lipo (Equation 4.15) terms share the same

functional form and both depend on the interatomic distance of the metal-acceptor or nonpolar-nonpolar pair.

$$S_{metal} = \Sigma_{MA} f(r_{MA}, r_{m,1}, r_{m,2}) \tag{4.14}$$

$$S_{lipo} = \Sigma_{LL} f(r_{LL}, r_{l,1}, r_{l,2}) \tag{4.15}$$

The clash term (Equation 4.16) is summed over all non-hydrogen protein-ligand atom pairs.

$$E_{clash} = \Sigma \epsilon_{clash}(r, r_{clash}) \tag{4.16}$$

where r is the interatomic distance and r_{clash} the clash distance. The clash energy for each atom pair depends on the nature of the protein and ligand atom; it is 0 for $r > r_{clash}$, and for $r \leq r_{clash}$ gets the value shown in Equation 4.17:

$$\epsilon(r, r_{clash}) = \begin{cases} \dfrac{20}{\Delta G_{hbond}} \dfrac{(r_{clash} - r)}{r_{clash}} & \text{donor-acceptor pairs} \\ \dfrac{20}{\Delta G_{metal}} \dfrac{(r_{clash} - r)}{r_{clash}} & \text{metal-acceptor pairs} \\ 1 + 4\dfrac{(r_{clash} - r)}{r_{clash}} & \text{all other non-H pairs} \end{cases} \tag{4.17}$$

4.10 Haptic devices

The market of haptic devices has until recently been a very small and specialised one. Haptic devices have therefore traditionally been expensive tools. In the recent past, however, hardware companies have started to release low-end devices aimed particularly to enhance video game experience. The implementation of this project was started on such a device, the Novint Falcon, and was then extended to use a more professional device, SensAble's Phantom Omni.

4.10.1 Novint Falcon

Novint Technologies first released the Falcon (Figure 4.20) device in 2006. It is a USB device with a ping pong ball-sized spherical grip, connected via three arms to a weighted base. The base contains sensors and motors attached to each arm. The sensors can triangulate the coordinates of the grip in space, and the motors can act together to apply a force to it with desired direction and intensity. The grip can only be translated in space, but not rotated. The Falcon is therefore a three degrees of freedom (3-DoF) device. Novint also released a software developing kit and drivers for the device. Both are windows-specific, and, according to Novint's developer center, support for other operative systems is not scheduled at the moment. Nevertheless there are unofficial open source projects for the development of falcon's drivers for linux and macosx.

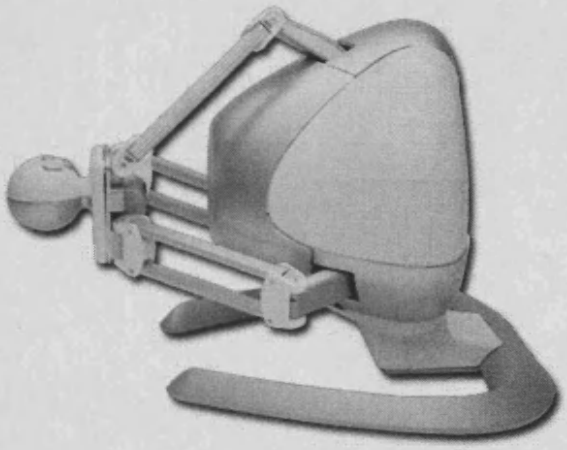


Figure 4.20: Novint falcon device

4.10.2 SensAble Phantom Omni

SensAble Technologies is the leader in 3D Touch force-feedback technology. The Phantom Omni (Figure 4.21) is the most affordable of a wide range of devices, that

differ for precision, maximum appliable force intensity and size of workspace. The grip is a pen-shaped stylus, attached via a segmented arm to the base. There is a range of sensors capable of reconstructing the position and the orientation of the pen in 3D space (six degrees of freedom). The motors can apply a directional force, but not a torque (3-DoF force feedback). Data are transferred to the computer via FireWire. SensAble supports Windows, Linux and MacOsX, releasing versions of the drivers and the SDK for each platform. However the linux driver is outdated, not supported on 64 bit systems nor on the newer versions of the linux kernel. Likewise, the Mac version is compiled for PowerPC processors and therefore runs slower on the new Intel Macs. The OpenHaptics library (one of Sensable's SDKs, which despite its name is not open source) is compatible with the whole range of Sensable's haptics. Therefore the same implementation should be able to be driven with more professional SensAble tools.

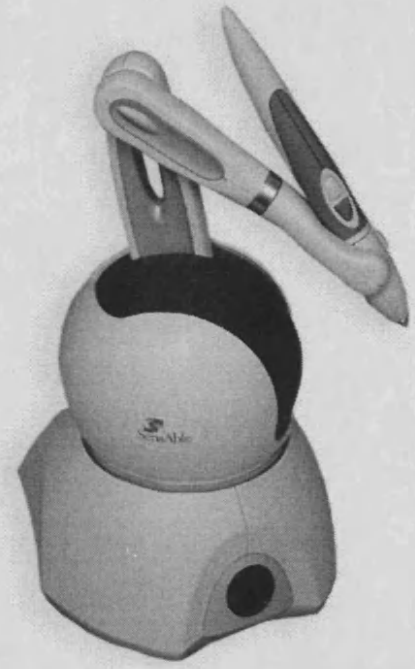


Figure 4.21: Sensable phantom omni device

4.10.3 Nintendo Wii Remote

Nintendo has recently released a gaming console (Nintendo Wii) based on a new idea of ergonomic controllers. Instead of the traditional joystick input, the console tracks the position of the controllers (Figure 4.22) as they are moved in space. Interestingly all the technology that allows this resides in the controller itself. Each controller is equipped with a 3-axis accelerometer, that is able to track the direction of the force applied to the device. When the device is kept still it can track gravity, and therefore calculate its orientation. The accelerometer alone is unable to track rotations along horizontal planes, because they don't change the relative position of the gravity vector. Therefore the controller also features an infra-red camera. By using 2 sources of infra-red light (IR LEDs are the first choice, but normal candles are excellent IR light sources as well) it is possible to track this rotations and also the distance of the device from the sources. The camera can detect up to 4 infra-red signals. By using four infra-red sources with different intensity it is theoretically possible to derive the position and the orientation of the device using the camera alone. The Wii Remote is obviously not designed to work on a computer, but exploiting a bug in a videogame various drivers have been reverse-engineered and allow the exchange of data between the device and a computer via Bluetooth. The Wii Remote is equipped with a rumble pack. Tuning the frequency of the rumbling and exploiting a tactile illusion it is possible to have a pseudo-directional haptic feedback from it.

4.11 Speed measurements

Speed tests were carried out under Microsoft Windows XP Professional 64 bit on a Viglen Genie featuring an Intel Xeon E5335 processor, 2 Gb RAM memory and ATI RADEON X1550 and under Ubuntu Feisty Fawn 64 bit on a Mac Pro 2.66 GHz featuring two Dual-Core Intel Xeon processors, 2 Gb RAM memory and a NVIDIA GeForce 7300 GT (256MB) graphic card. As a model of fragment based drug design the interactions of quinazoline with the protein kinase A active site (PDB code 1Q8U)

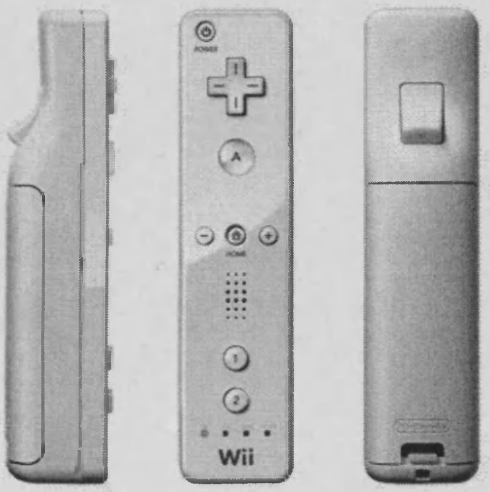


Figure 4.22: Nintendo Wii Remote

were modelled. With water and any non-protein atoms removed and hydrogen atoms added, the protein system counts as many as 5051 atoms. Quinazoline is a rigid fused ring system of 16 atoms. The speed of the simulation depends on the relative position of the fragment to the protein, and has its minimum when the fragment is inside the ATP binding cleft, surrounded by protein atoms. The haptic update constantly remained above 1KHz, while the computation thread update remained over 200 Hz for the fully flexible model.

4.12 Future development: protein flexibility

The current implementation, similarly to most docking approaches used at present, treats the target macromolecule as a rigid body, and just explores the flexibility of the ligand (semi-flexible approach). While this is often accepted as a reasonable simplification, it is true that many software companies are moving towards the inclusion of protein flexibility exploration (fully-flexible approach) in molecular docking simulations. This has been done to different degrees, from the simple exploration of rotamers

of predefined amino acid side chains to the prediction of backbone flexibility using short molecular dynamics runs. Similar approaches could be used to include protein flexibility into the haptic-driven simulation. Atoms of the side chains of preselected residues could be treated in the same way as the atoms of the ligand, allowing them to move freely, and constraining the position of the α carbon. Simulations of this kind has been shown to produce an high number of unrealistic conformations, so it would be probabily be necessary to restrain the original coordinates of the atoms to a certain extent. Theoretically the same principle can be extended to backbone atoms. The complexity of the system however increases combinatorially with the number of ligand atoms, and therefore this method would probabily not be effective for more than a few flexible side chains without rewriting the core of the computation. One possible way to gain computational power would be the parallelisation of the computation code, that would allow to effectively exploit clusters of CPUs or even GPUs. Another feasable approach would be the use of libraries of rotamers. Before the simulation is started libraries of low energy rotamers could be calculated for each amino acid. While it is running a separate looping thread could pick random amino acids and check if one of the rotamers from its library would have a lower energy and if that is the case set that conformation as the current one. It would be also possible to allow for fully flexible proteins by using a simplified force field similar to the ones used for molecular dynamics. The whole protein would be allowed to evolve in a solvent-implicit dynamics simulation, while the ligand and its interactions with the protein would be treated by a more precise forcefield. The minimisation of the protein wouldn't need to be as fast as it is now, because it is not involved in the generation of haptic feedback. Since protein atoms would no longer be fixed in space (characteristic on which neighbours lists currently rely) minor changes would be necessary to the neighbours list generation algorithm to account for that. Apart from this the force generation code should run exactly as it is, as long as the access to protein atoms coordinates and forces are made thread-safe (which should be the case as soon as the code is parallelised).

Chapter 5

Design of Anticancer Drugs Targetting the EGFR Pathway

5.1 EGFR

EGFR (epidermal growth factor receptor) is the best-studied member of the HER (human epidermal growth factor receptor) family. Some EGFR inhibitors are already on the market as anticancer drugs[39, 40, 41, 42, 43]. The aim of this section was to use various modelling techniques for an *in silico* characterisation of the binding modes of some current drugs to use for the developement of new scaffolds for EGFR inhibitors. Various molecular modelling techniques were combined to design lead molecules for *in vitro* testing.

5.1.1 Eukaryotic protein kinases

Eukaryotic Protein Kinases (EPKs) are a class of enzymes that catalyse the transfer of a phosphate group from a donor nucleotide (ATP or GTP) onto a peptidic substrate. This reaction (named phosphorylation) is the most frequent and well-studied post-translational protein modification *in vivo* and is opposed to dephosphorylation

(catalysed by protein phosphatases)[44]. The balance between phosphorylation and dephosphorylation (Figure 5.1) of target proteins has proven to be a key mechanism in the regulation of several cellular processes, such as cell cycle progression, growth and apoptosis. For this and other reasons kinases are currently the second most important target in pharmacological therapy [45] after GPCRs.

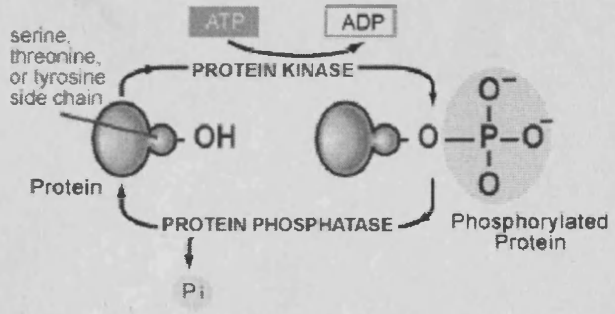


Figure 5.1: Phosphorylation and dephosphorylation

The human kinome [46, 47] counts more than 500 members, divided into seven subfamilies. The map of the human kinome is shown in Figure 5.2. According to the their substrate specificity EPKs are classified into two major groups:

- **Ser-Thr kinases** transfer the phosphate onto an aliphatic hydroxyl group (serine and threonine side chains)
- **Tyr kinases** target an aromatic hydroxyl group (tyrosine side chain)

Tyr kinases can be both solubile enzymes (non-receptor tyrosine kinases or NRTKs) or membrane-bound receptors (receptor tyrosine kinases or RTKs).

5.1.2 RTKs

RTKs are transmembrane proteins. They consist of 3 superdomains: a transmembrane region and one on each side of the membrane. The extracellular portion presents

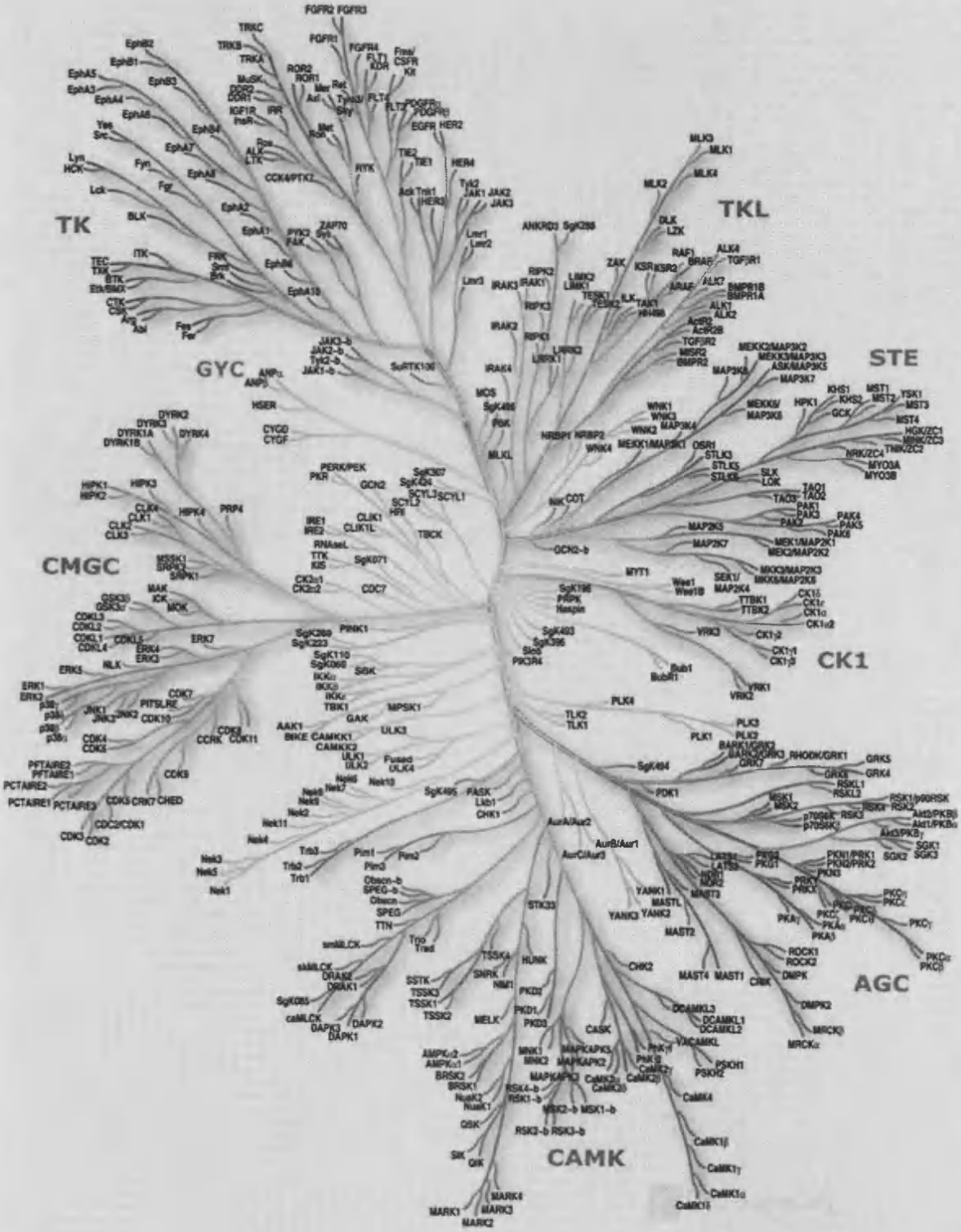


Figure 5.2: Map of Human Kinome (http://kinase.com/human/kinome)

a ligand-binding domain and an extracellular juxtamembrane region connecting it to the transmembrane domain. The intracellar portion is divided into intracellular juxtamembrain region, kinase domain and a flexible C-term tail.
When a growth factor acts upon a specific cell surface receptor harboring a tyrosine kinase activity, cell fate decisions are affected through a complex succession of signalling events. The large family of growth factor receptor tyrosine kinases may be considered as allosteric enzymes, whose regulatory portion binds the allosteric regulator, a ligand growth factor of relatively low molecular weight, and consequently undergoes large-scale conformational changes leading to receptor dimerisation. As a result, the catalytic kinase portion, which faces the intracellular milieu, undergoes activation and transfers phosphate groups to the tyrosine residues of nearby molecules, including dimer partners and substrate proteins participating in signal transfer [48].

5.1.3 The HER family

Among all RTKs, the HER (ErbB) family (also called the type I RTKs) is considered the prototypic founder sub-group of the RTK super-family, which includes 18 other small sub-groups of related receptors.

Two out of the four HER proteins, namely HER-1 (also called epidermal growth factor receptor; EGFR) and HER-4, are autonomous; when bound by a ligand growth factor they undergo dimerisation and generate intracellular signals culminating in cell proliferation, migration or differentiation. The other two receptors are non-autonomous: HER2 (also called HER2 in human and Neu in rodents) binds no soluble ligand, but acts as a preferred partner in heterodimeric complexes with other, ligand-bound HERs. On the other hand, HER-3 cannot generate signals in isolation because the kinase function of this receptor is impaired. Nevertheless, in the context of a heterodimer, primarily with HER2, HER3 can mount potent intracellular signals. Interestingly, HER1, -3 and -4 can each bind several distinct ligand molecules, which, according to genetic evidence, are partially redundant in their function. [49]

5.1.4 Structure of the kinase domain

Kinase domains of the EPKs are highly conserved in sequence and structure. They fold into a a bilobate structure, the N-term portion forming a lighter lobe (N-lobe) and the C-term portion forming a heavier subdomain, called C-lobe. The motif connecting the two subdomains is named hinge region (Figure 5.3).

The N-lobe is folded mainly in antiparallel β-sheets, with the only exception of a conserved α-helix (α-C helix). The C-lobe is composed of several α-helices. Other important motifs are the Gly-rich loop (or phosphate-binding loop or P-loop) and the catalytic loop.

ATP binds in a deep cleft between the two lobes, the phosphate tail forms several interactions with the Gly-rich loop and other conserved amino acids, and the γ-phosphate is positioned near the catalytic aspartate residue on the catalytic loop (Figure 5.4). The peptidic substrate positions itself in a pocket between the activation loop and the C-α helix.

5.1.5 HER kinases activation

The HER family of receptors activate a multi-layered network mediating crucial pathways leading to cell proliferation, differentiation, migration and altered metabolism (Figure 5.5), in response to activation of the receptor by the epidermal growth factor (EGF), transforming growth factor-α (TGF-α), and several other related peptide growth factors.

Activation of the ligand-binding domain by the extracellular ligand leads to homodimerisation or heterodimerisation among the members of the HER family. The structural basis for ligand-induced dimerisation of the extracellular region of EGFR family members is now well understood [51, 50]. This dimerisation (Figure 5.6) results in the phosphorylation of tyrosine residues in the C-terminal tail segments, which serve as docking sites for signaling molecules that contain SH2 or PTB (phosphotyrosine binding) domains and are responsible for onward transmission of the signal

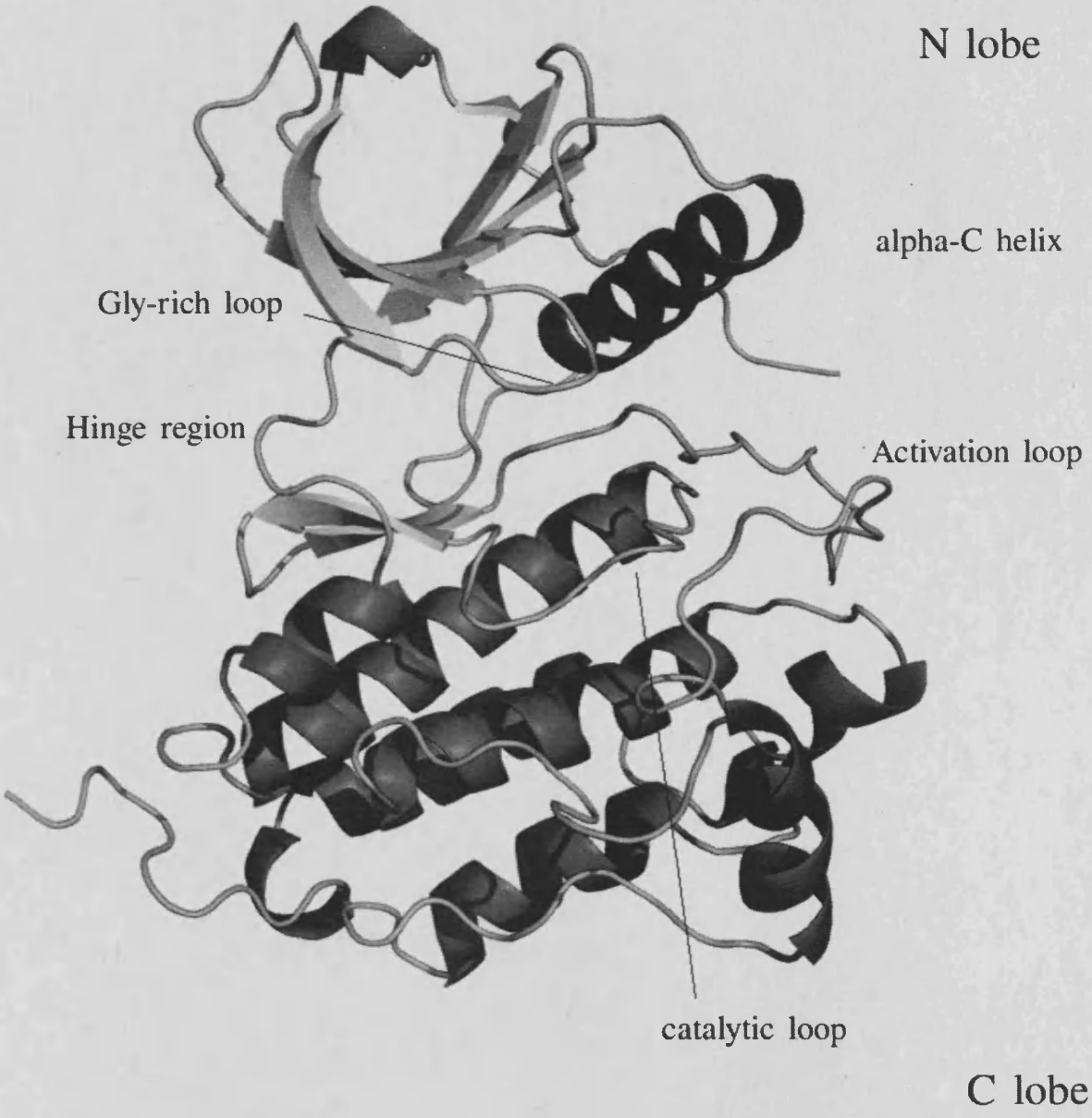


Figure 5.3: Structure of kinase domain

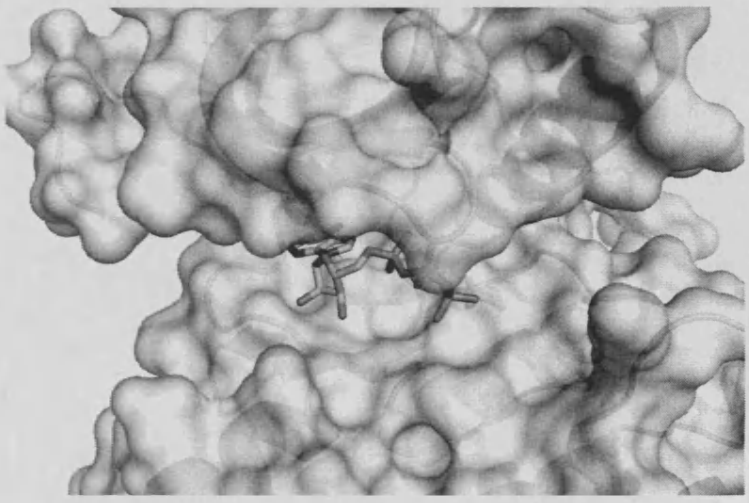


Figure 5.4: ATP binding pose

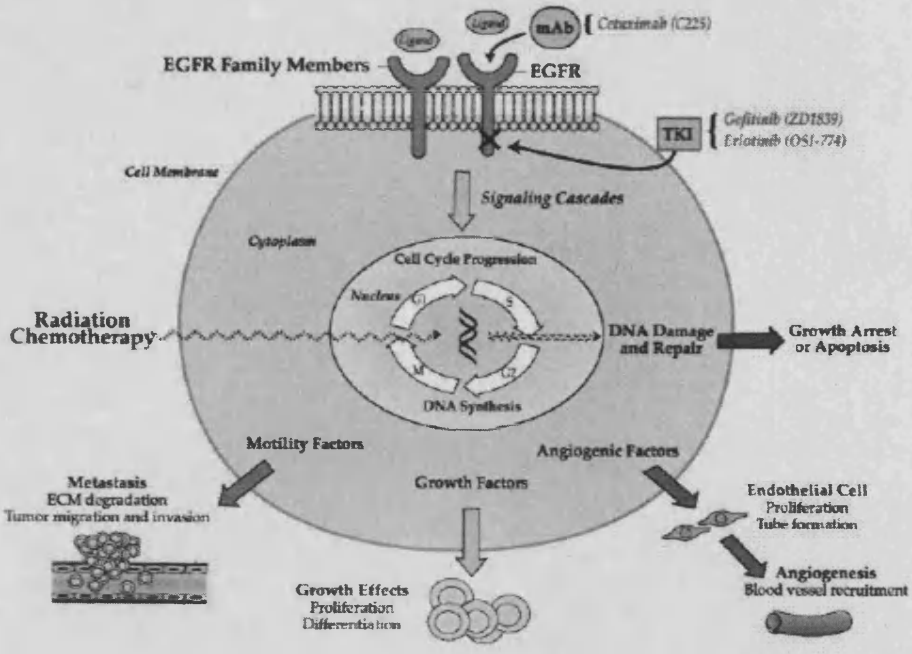


Figure 5.5: EGFR signalling pathway[50]

[52]. EGFR, HER2, and HER4 contain catalytically competent kinase domains and can form heterodimeric pairs with each other. HER3 contains an inactive kinase domain, but it can pair with and activate the other members of the family.

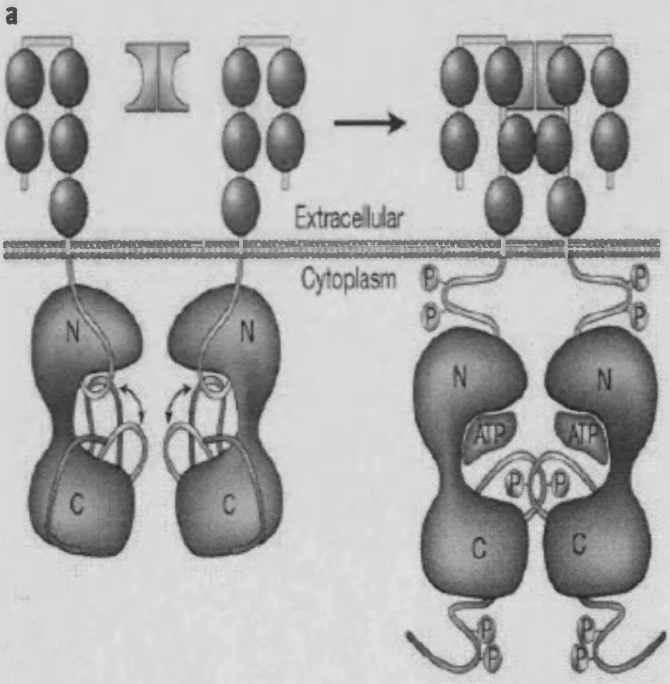


Figure 5.6: EGFR dimerisation[50]

Structural events taking place in the intracellular region after extracellular dimerisation and resulting in kinase domain activation are much less clear. It has been found that the EGFR kinase domain can be activated by increasing its local concentration or by mutating a leucine (L834R) in the activation loop, the phosphorylation of which is not required for activation. The crystal structure of the EGFR tyrosine kinase domain was first published in 2002 both alone and in complex with an inhibitor (erlotinib) [53], and displayed several unique features. While structural studies of the kinase domains from other RTKs (and protein kinases in general) have shown them to be catalytically inactive until an activation event such as ligand-induced dimerisation causes them to become autophosphorylated within the activation loop and

subsequently become locked into a characteristic active comformation, members of the EGFR family appear to be unique in not requiring activation loop phosphorylation for their activity. The crystal structure of an unphosphorylated form of the EGFR kinase domain was found to be constitutively locked in the active configuration. In a recent study[54], it is suggested that the kinase domain of the EGFR is regulated by a direct allosteric stabilisation of the active conformation. This suggests that the kinase domain is intrinsically autoinhibited, and an intermolecular interaction promotes its activation. Using further mutational analysis and crystallography it was demonstrated that the autoinhibited conformation of the EGFR kinase domain resembles that of Src kinases [55] and cyclin-dependent kinases (CDKs) [56].

EGFR activation results from the formation of an asymmetric dimer (Figure 5.7) in which the C-terminal lobe of one kinase domain plays a role analogous to that of cyclin in activated CDK/cyclin complexes. The CDK/cyclin-like complex formed by two kinase domains thus explains the activation of EGFR-family receptors by homo- or heterodimerization. An activating heterozygous mutation in the activation loop of the kinase domain of EGFR, L834R (also identified as Leu858 in an alternative numbering of the human EGFR sequence), has been found frequently in lung cancer patients (accounting for 41% of EGFR mutations in these cancers [57, 58]). An adjacent leucine residue (Leu837) is mutated to glutamine in some lung cancer patients [59]. These residues are either partially (Leu834) or completely (Leu837) surface exposed and do not play an obvious structural role in the active conformation. In contrast, the importance of these two residues is obvious upon consideration of the Src/CDK-like inactive conformation because both residues pack against hydrophobic sidechains in the interior of the kinase domain. Replacement of either leucine by a polar sidechain is likely to destabilise the Src-like inactive conformation and thereby activate the kinase domain. Interestingly, Erlotinib is incompatible with the Src/CDK-like inactive conformation, which may explain why activating mutations potentiate the effectiveness of the closely related drug Gefitinib [57, 58].

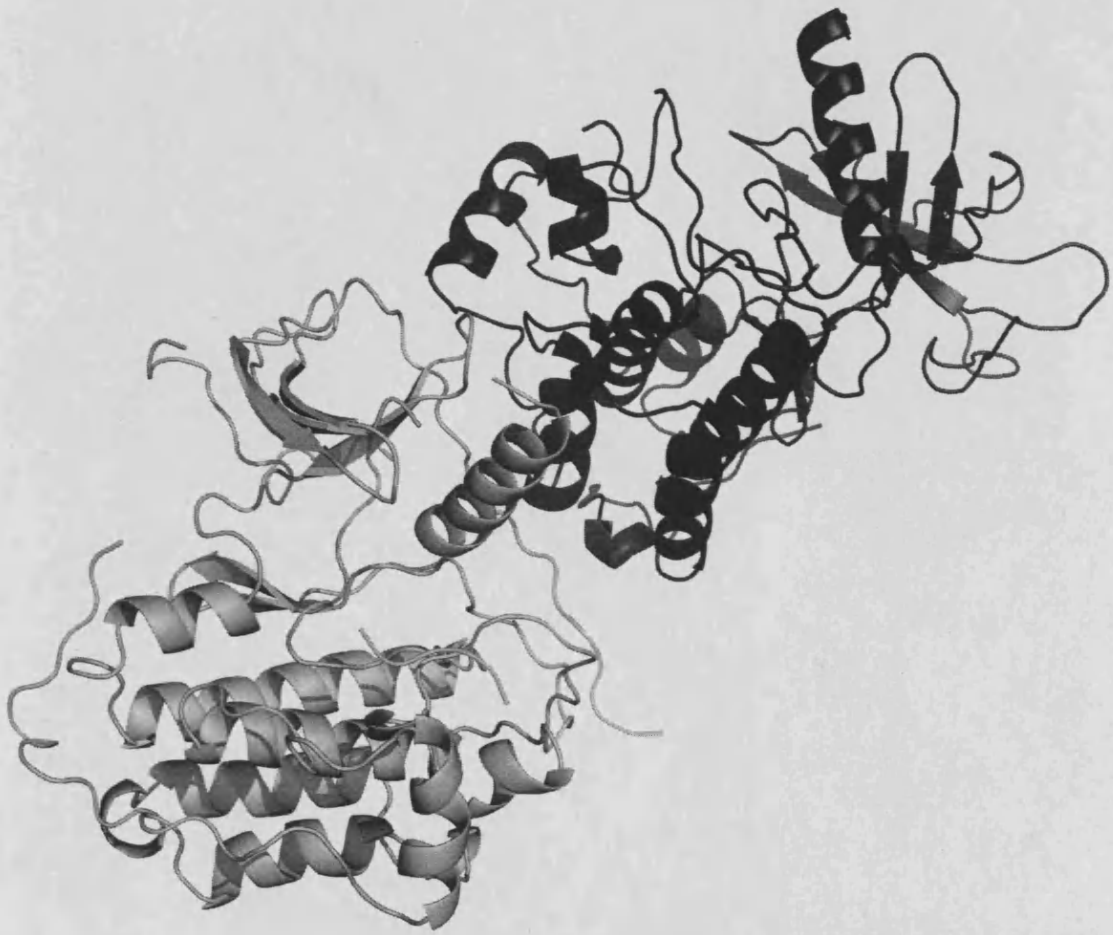


Figure 5.7: EGFR kinase domain asymmetric dimer

5.1.6 Anti-EGFR therapies in clinical use

Current anti-EGFR chemotherapy follows two main approaches: tyrosine kinase inhibitors (TKIs) and monoclonal antibodies (mABs). Both classes have reached an advanced stage of clinical developement and each approach has its strengths and weaknesses.

- **TKIs**: Tyrosine kinase inhibitors are a class of small molecule drugs that compete with ATP for the binding to a tyrosine kinase active site. There are several example of this class of compounds among EGFR inhibitors currently on market or in advanced clinical trial. The major problem of this class of drugs is selectivity. There are about 518 known protein kinases in the human proteome [46] and all of them share a higly conserved ATP-binding pocket. It has been proven that kinase inhibitors can cross-inhibit even loosely related kinases, making selectivity a real challenge with this class of drugs [60].

- **mABs**: Monoclonal antibodies directly inhibit EGFR outside the cell by binding to the extracellular domain and preventing ligand binding. Antibodies have some important advantages over small molecules: being proteins their metabolitic pathways are well understood and completely safe and they exibit high affinity and selectivity. On the other hand, however, they also exibit the typical drawbacks of biotechnological drugs: they are expensive and difficult to manifacture and store, have a short shelf-life due to denaturation and degradation, they are not suitable for oral administration, they're often immunogenic and susceptible to enzymatic degradation in vivo.

5.2 Aims

Focus of this part is to employ the haptic-based approach alongside traditional molecular modelling techniques to the design and optimisation of anti-EGFR drugs, on the one hand to test its performance on a real drug design project and on the other to

achieve the design of structures of interest. Both lead optimisation of existing structures (from drugs currently in the market, in clinical trials or inhibitors described in the literature) and de novo design approaches have been used. The target is the inhibition of the kinase activity of EGFR and other members of his family, both in a traditional way with molecules that compete with ATP for the active site and targeting a protein-protein interaction essential for the activation of the enzyme.

5.3 Validation of docking algorithm

At the time when this project was started the Protein Data Bank [61] hosted 7 different structures of the EGFR kinase domain, shown in Table 5.3.

PDB Code	description	resolution (Å)
1M17	human EGFR with Erlotinib	2.6
1M14	human EGFR (apoenzyme)	2.6
2J5E	human EGFR with 13-JAB	3.1
2J5F	human EGFR i with 34-JAB	3.0
2GS6	human EGFR with an ATP-analog-peptide conjugate	2.8
2GS7	human EGFR (apoenzyme)	2.8
1XKK	human EGFR with Lapatinib	2.4

Table 5.1: EGFR kinase domain crystals in the Protein Data Bank

Several small molecules have been co-crystallised with the EGFR kinase domain. Quinazoline derivatives, in particular, are well represented. This class of compounds have been chosen to validate the docking protocol. The analysis of the crystals showed that quinazoline is involved in a water bridge with the active site (probabily with the side chain of Thr766). That water molecule was therefore kept in every docking simulation. The docking algorithm was able to retrieve the crystallographic pose for all of

the tested molecules (with the exception of lapatinib) in their own crystal structure and in cross-docking experiments on 1M17 (co-crystal with erlotinib). The reasons why lapatinib's pose could not be retrieved are discussed in section 5.5. Quinazoline was docked as well to validate a fragment-based approach and his crystallographic pose was successfully retrieved. Figure 5.8 shows some of the results.

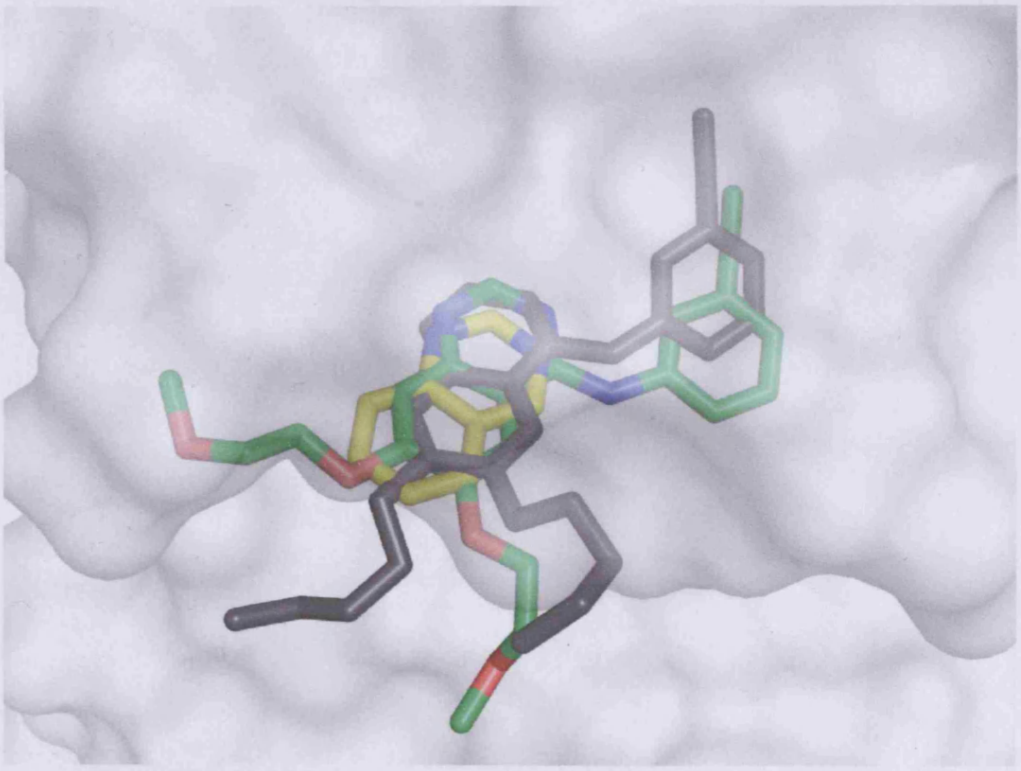


Figure 5.8: Quinazoline (yellow) and Erlotinib (green) docked poses compared to Erlotinib crystallographic pose (black)

5.4 Investigating EGFR plasticity

The understanding of the conformational balance of a protein is essential in the design of inhibitors. Crystal structures of the EGFR kinase domain have been determined in the literature in two distinct crystal lattices. In one the kinase domain is unphosphorylated but adopts an active conformation, either in a ligand-free form or in complex with the cancer drug Erlotinib (Tarceva) [53]. The EGFR kinase domain in a second crystal form [62] is bound to the drug Lapatinib (Tykerb), or obtained through a mutation [54] and is in an inactive conformation that resembles closely that of inactive Src-family kinases and cyclin-dependent kinases (CDKs).

To elucidate the structurale rationale for the activation state of the enzyme, 20 different structures from the Protein Data Bank were superimposed (Figure 5.9) to analyse the main differences.

In the inactive conformation the axis of the α-C Helix resulted to be rotated, allowing the activation loop to fold back and close the access to the peptide binding site.

5.5 Analysis of different inhibition modes

Among quinazoline analogues several compounds show remarkable features that outline different inhibition modes. Four candidates were chosen and analysed to provide structural insight for their binding mode. Figure 5.10 shows the structures of the four molecules.

Three of the compounds (Erlotinib, Gefitinib and Lapatinib) are crystallised in their binding mode, while the forth has a very similar analogue crystallised in the Protein Data Bank (PDB code 2J5E), and docking experiments confirmed that it is very likely to mantain the same binding mode. In all four compounds the quinazoline moiety appears to bind to the hinge region in the same way, with the interactions outlined in the previous section.

Figure 5.9: Superimposition of 20 crystallographic structures: the active ones are shown in red, the inactive in blue. The shift of the activation loop and the rotation of the α-C helix are clearly visible

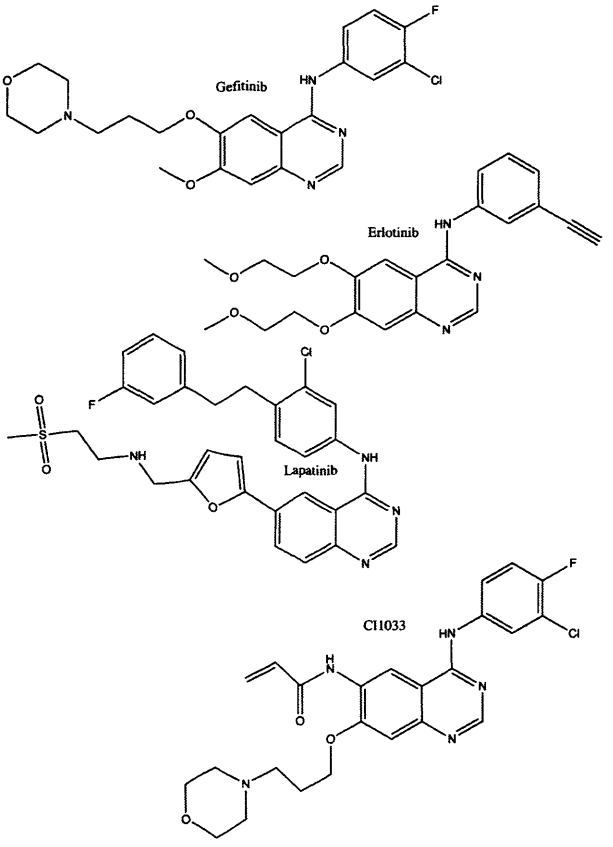


Figure 5.10: Structures of EGFR inhibitors used in the study

Some important pharmacokinetic properties for these compounds are shown in Table 5.2.

Inhibitor	K_i (nM)	Complex Half-life (min)
Erlotinib	0.7	less than 10
Gefitinib	0.4	less than 10
Lapatinib	3.0	300
CI1033	30 (IC_{50})	∞

Table 5.2: Pharmacokinetic properties of some quinazoline-derivated TKIs

CI1033 is an irreversible inhibitor and in its binding pose the Michael acceptor moiety is pointing towards a cysteine residue (Cys797) in the active site. It is capable of covalently binding to the sulphur and thus its complex half life is virtually infinite. Erlotinib and Gefitinib are both on the market as anticancer drugs (Tarceva and Iressa). They show comparable inhibition properties. Lapatinib, on the other hand, exhibits a much higher complex half-life.

It has been reported that Lapatinib is selective for the inactive form of EGFR kinase domain and can somehow lock the enzyme into the inactive conformation [62]. In fact it is the only one inhibitor so far co-crystalised with the enzyme in the inactive conformations.

Docking experiments were used to find a rationale for the difference in Lapatinib's binding mode from the ones of the other inhibitors. Each molecule was docked into the receptor in both its active (PDB code 1M17) and inactive (PDB code 1XKK) conformation (Figure 5.11).

Erlotinib, Gefitinib and CI1033 were successfully docked in the protein in both conformational states. Lapatinib docking, unfortunately failed for both the active and the inactive crystal. This originates from a limitation of the docking approach, as the inactive crystal (1XKK) is actually a co-crystal with Lapatinib itself. The

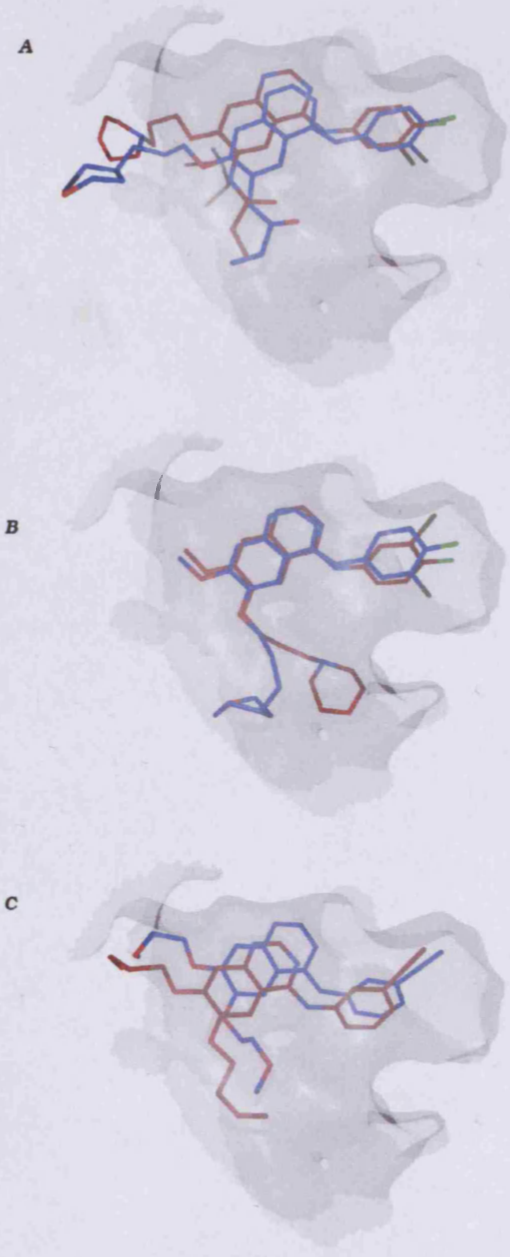


Figure 5.11: Cross docking experiments on the active and the inactive conformation: CI1033 (A), Gefitinib (B) and Erlotinib (C) dock in the same position in both crystals. Blue poses are obtained on the inactive conformation of the enzyme, red poses refer to the active conformation

molecule's dimensions and flexibility make it a very difficult example of ligand for molecular docking.
A simple superimposition of the crystallographic pose to both enzyme structures followed by energy minimisation was therefore used, and its results revealed a structural rationale for Lapatinib's selectivity: In the inactive conformation a small subpocket is present that accomodates the phenyl moiety of the drug. In the active one, however, the α-C helix shifts towards the ATP binding region and closes that pocket. Figure 5.12 shows the enzyme surface and its interactions with the drug. Blue regions are the ones that are predicted to negatively interact with the drug. In the active conformation the steric clash of the phenyl ring with the protein is clearly visible. Developing inhibitors able to bind in this region and stabilise the inactive conformation could be an interesting way to improve anticancer potency and selectivity.

5.6 Design of a selective TKI drug

Tyrosine Kinase inhibitors are designed to bind in the active site and to compete with ATP for its binding pocket. The general pharmacophore of TKIs [63], with some exceptions, can be reduced to: an aromatic group with up to three hydrogen bond donor-acceptor moieties, to mimic adenine and bind to the hinge region backbone. a second aromatic group directed to an hydrophobic poket (sometimes known as the kinase selectivity poket) that controls selectivity. a linker between the two aromatic groups.

The first group is very often either quinazoline, pyrido-pyridine or pyrido-pyrimidine.

As the adenine-binding pocket is highly conserved among kinases, it was decided to keep quinazoline as a binding fragment and focus the design on the kinase selectivity pocket, following a fragment based approach was therefore chosen. A library of 206 fragment-like molecules (obtained from Enamine) was virtually screened. The binding site was restricted to the kinase selectivity pocket and the best scoring fragments were eye-inspected for their interactions with the enzyme.

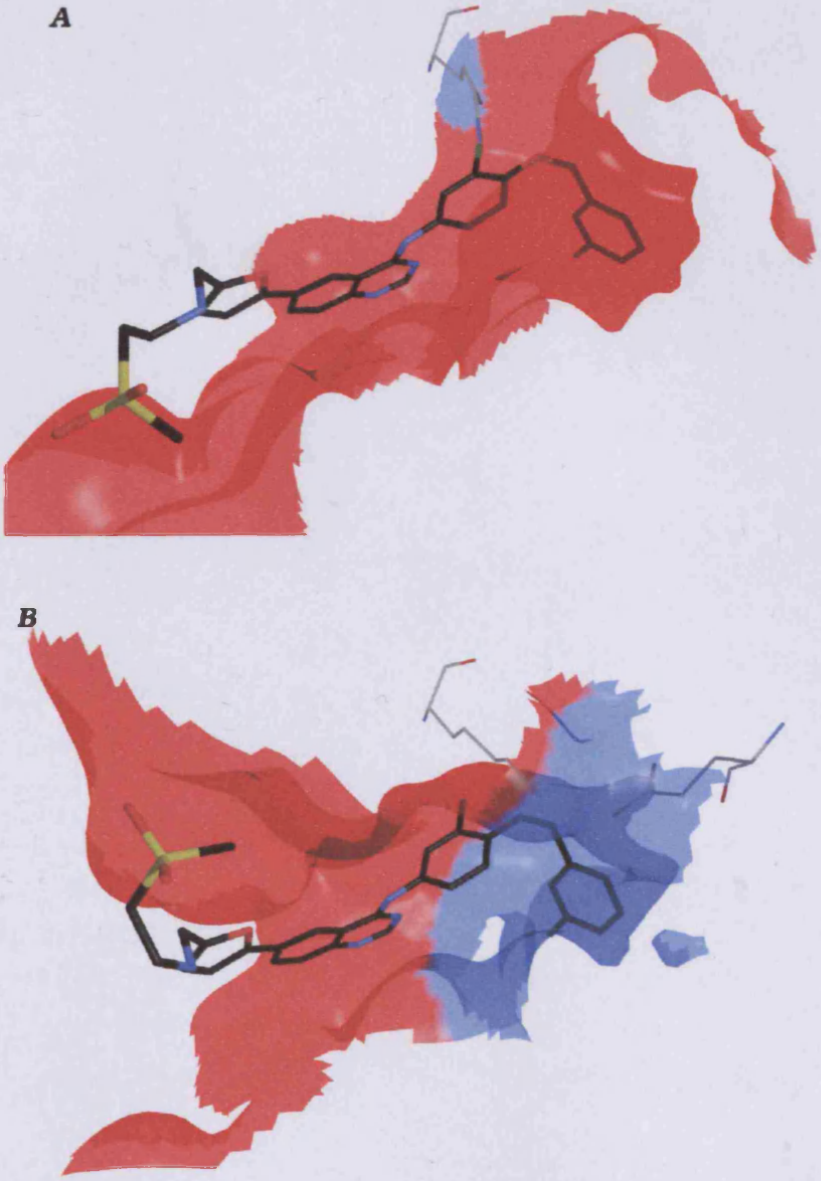


Figure 5.12: Lapatinib binding: red regions represent favorable binding, blue regions represent clashes and repulsions. Lapatinib's binding poses in the inactive conformation (A), and in the active one (B) are shown

6-Hydroxyl-dihydroquinolinone (Figure 5.13) was be a promising group to bind to this pocket and was selected for further in silico investigation.

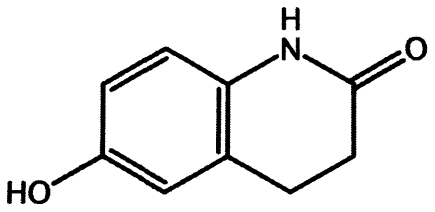


Figure 5.13: Lead Fragment: 6-hydroxy-dihydroquinoline-2(1H)-one

The haptic-based approach was then applied to refine the lead fragment. The fragment was used to probe the forces in the selectivity pocket, and per atom scoring was used to predict which portion of the fragment were essential for binding. As the amide group was not predicted to be crucial for the interaction with the molecule it was eliminated in successive runs. The absence of the bulky carbonyl group seemed to allow for an even better orientation of the hydroxyl for its hydrogen bond (with the backbone of A743). An unsaturation was also introduced to avoid prochiral centers and boost hydrophobicity. Hydroxynaphtalene was therefore chosen as a fragment and its pose was confirmed by docking experiments.

The following step was estabilishing a good linker bridge between the two aromatic portions, that was able to allow both fragments to bind with the correct orientation. A set of analogues, differing just for the linker bridge, were built and docked into the active site. Ethylene (Figure 5.14) seemed to provide the right distance and angle to allow the two aromatic portions to mantain the binding modes each showed when docked alone.

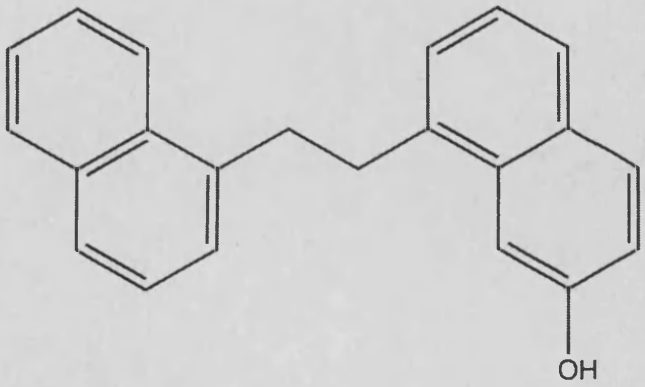


Figure 5.14: Lead design: ethylene bridge

Figure 5.15: Lead design: proposed lead. The fragment binding in the selectivity poket forms an hydrogen bond with the backbone of A743

5.7 Design of a EGFR allosteric inhibitor

As data about Lapatinib binding mode have pointed out, control over the active/inactive conformational balance of EGFR and possibly other members of the HER family could be the key to a selective and potent inhibition of this family of receptors. It has been demonstrated that EGFR kinase domain is capable of auto-regulating its activation state *in trans* [53]. The present model for EGFR activation considers the kinase domain to be autoinhibited and constrained most of the time in the inactive conformation. Upon ligand activation the extracellular domains of two HER members dimerise, so that their kinase domains are brought at a close distance. This allows the formation of a kinase domain dimer. The C-lobe of one domain transactivates the other by binding to the α-C helix region of the other and stabilising the active conformation. Mutations of some key amino acids in the interaction interface has been proved to prevent the kinase activity and lock EGFR in an inactive form[53]. This interaction is peculiar to this family of receptor. Being able to target this interface[64] and disrupt the interaction between the two kinase residues with a small drug-like molecule would therefore be the first step towards a new class of selective and potent anti-EGFR drugs.

5.7.1 Choice of the target

An analysis of the general characteristics of the two interface surfaces was carried out in order to choose which one was the most attractive for the inhibitor binding (Figure 5.16). The C lobe one was chosen as the most promising one. The main reasons for this choices are illustrated below:

- Its structure is conserved in each EGFR crystal structure, and its shape is well defined, while the N lobe one appears to be highly flexible.
- The molecular surfaces presents some clefts and pseudopockets that could fit a small molecule.

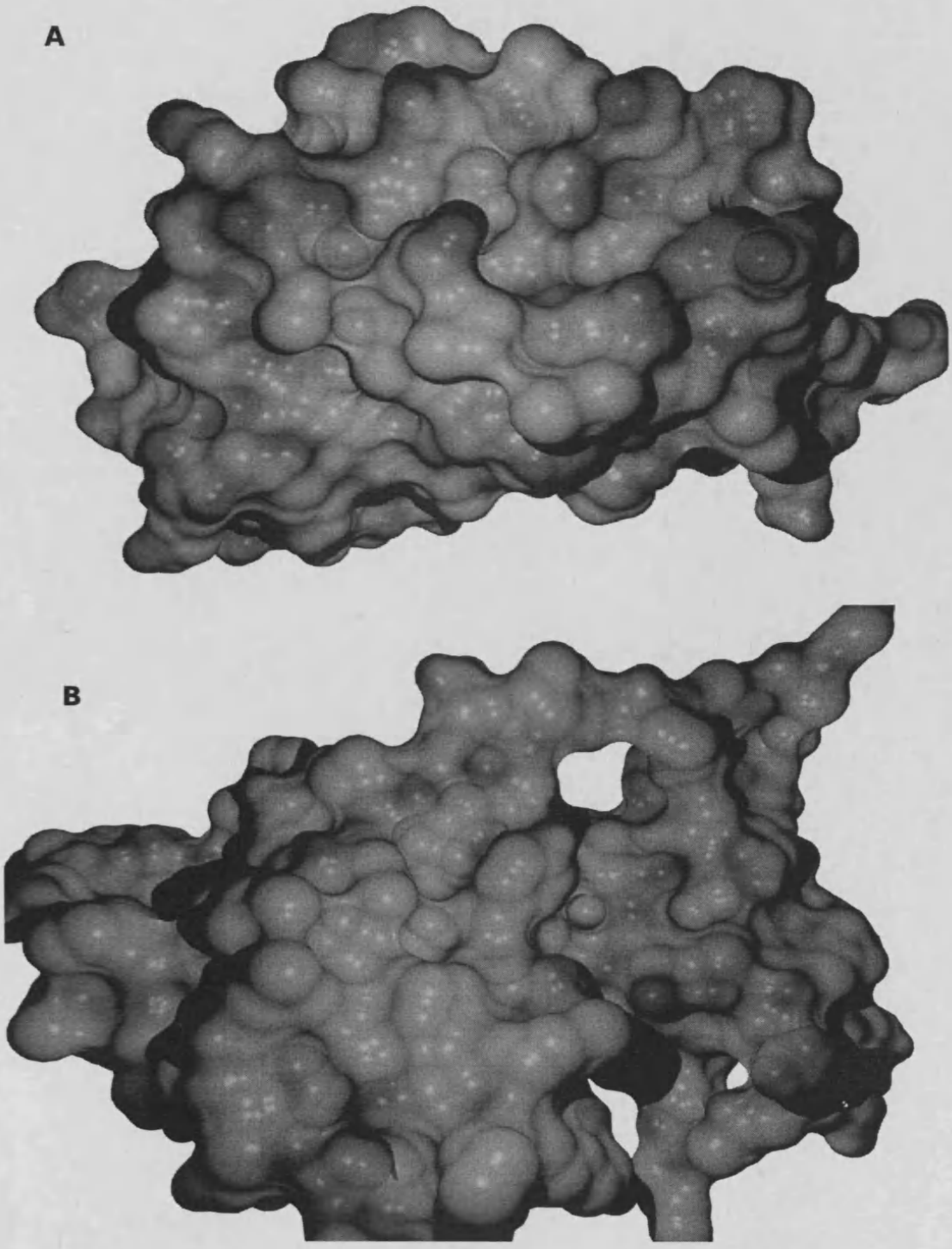


Figure 5.16: Interaction Surfaces: C lobe (top) was preferred over N lobe (bottom)

- The sequence of this region is more conserved among the HER family members. An inhibitor targetting this surface is therefore more likely to be cross-active.
- It has the same conformation in both the active and the inactive form of the kinase domain, which should in theory allow the molecule's binding to be independent form the activation balance of the receptor.
- The binding of a small molecule to this region is highly unlikely to affect the active/inactive balance of the kinase itself, while it is reasonable to think that a molecule targeting the N lobe surface might disrupt the protein-protein interaction while still keeping the kinase in its active conformation.

5.7.2 Virtual screening approach

1XKK was selected as target structure. In the PDB file Met945 side chain is not present, and was added using 2GS7 (V924R mutant) as a template. A large database from Enamine comprising 18143 structures was virtual-screened, docking each molecule on the C lobe interaction surface of the crystal. Among the top scoring molecules the one in Figure 5.17 appears particularly promising. Figure 5.18 shows the predicted binding pose of the molecule.

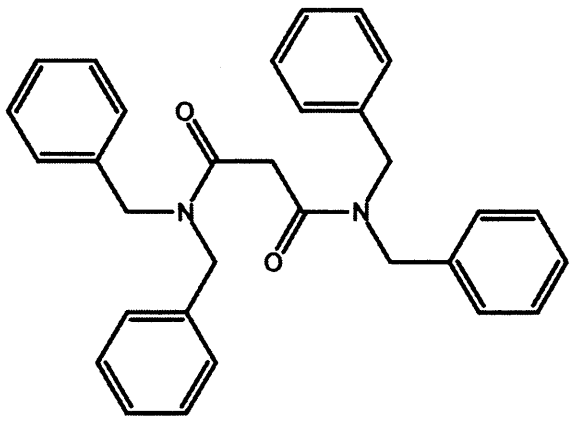


Figure 5.17: Structure of lead

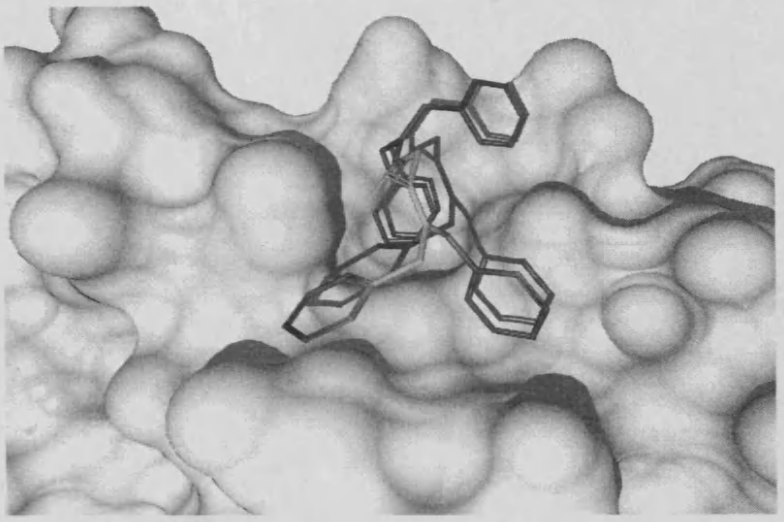


Figure 5.18: Predicted binding mode

Even though it is not binding directly to the interacting residues of the interface, superimposition of this binding mode to the EGFR kinase dimer (Figure 5.19) shows that this molecule could clash with the second monomer thus interfering with the complex.

5.7.3 In vitro tests

The compound was synthesised and *in vitro* test were carried out for the compound on four different cancer cell lines. Results are shown in Figure 5.20.

From these preliminary results the molecule appears to be active on some of the cell lines in the low micromolar range. This is result is very important considering that the inhibition of the dimerisation of the kinase domains of EGFR is not a validated target.

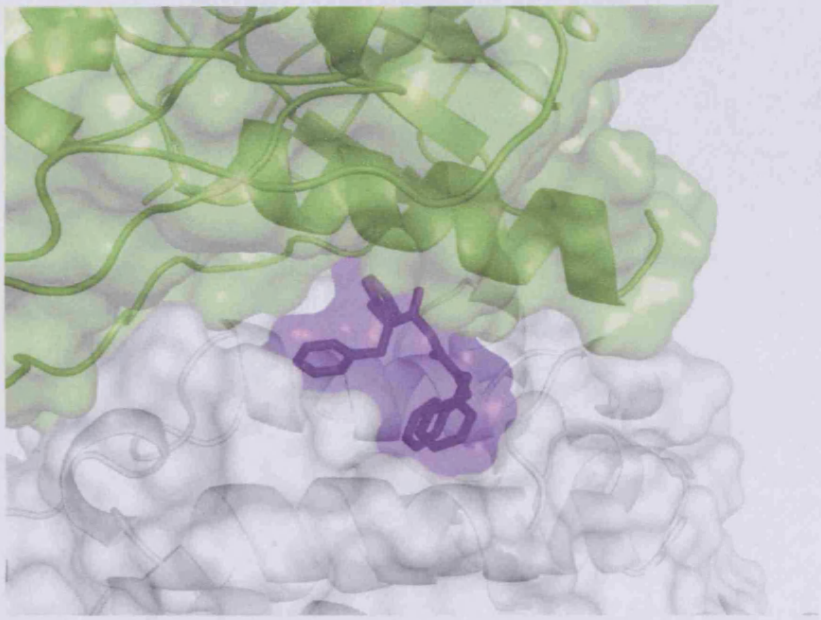


Figure 5.19: Inhibitor in the dimer structure

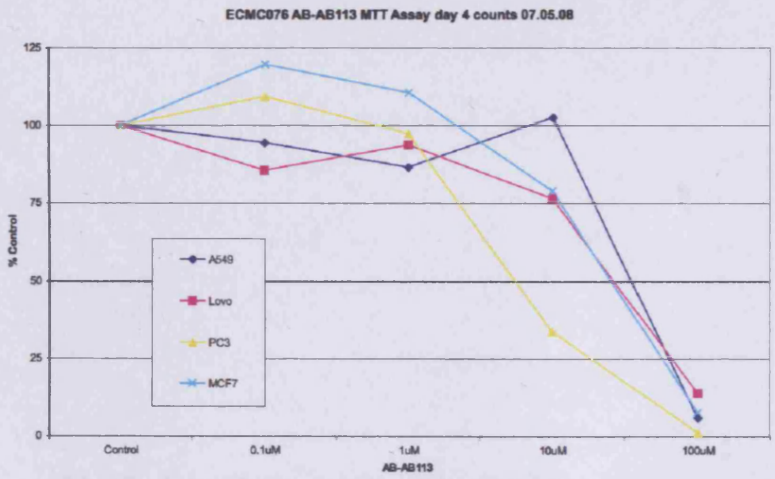


Figure 5.20: In vitro activity data

5.7.4 Considerations on other members of the family

EGFR activation pathway is partially redundant. Selective inhibitors *in vivo* have to face the compensation by other HER receptors of the impaired EGFR kinase activity. There is growing interest in the development of molecules capable of inhibiting different related kinases. Lapatinib is one of these dual inhibitors, being able to inhibit both HER1 and HER2[65]. Since different members of the HER family are thought to be able to form heterodimers with one another, it is reasonable to suppose that the interacting surfaces are fairly conserved in at least some of the kinases. These interfaces were studied to predict if the selected molecule could be able to bind to other EGFR family members and act as a selective multi kinase inhibitor.

The structure of HER2, HER3 and HER4 have not been solved yet. Homology models based on the EGFR structure have therefore been built from the sequences of the three kinases and compared to 1XKK. The sequences have been aligned using MOE 2007.09. Ten models for each were predicted using the same sofware and the best scoring one was chosen in each case and inspected for deviations in terms of inter-residue diheadral angles and bond lengths. No adjustments to the structures were needed.

In particular the amino acids surrounding the predicted binding pose (less 4.5 Å away from any of the ligand's atoms) were investigated. Table 5.3 shows the differences in the sequences in the area of interest:

HER3 appears to be less related to EGFR than the other two. This protein, unlike the rest of the family is catalytically inactive. However it is thought to be able to activate the others in trans by forming an asymmetric dimer with their kinase domain. The C lobe interface is in fact well conserved in all the generated homology models. The putative binding site for the lead molecule (Figures 5.21, 5.22 and 5.23)

	HER2	**HER3**	**HER4**
residue identity with EGFR	78.9 %	59.2 %	78.5 %
D942	D	D	D
V943	V	V	V
M945	M	M	M
I946	I	V	V
K949	K	K	K
P959	P	P	P
K960	R	T	K
E963	E	E	E
L964	L	L	L
I966	S	N	A
E967	E	E	E
F968	F	F	F
K970	R	R	R
M971	M	M	M
Y978	F	Y	Y

Table 5.3: Comparison of the binding site residues: conserved residues are shown in black, conservative mutations in blue and non conservative mutations in red

is very similar and all the non conservative mutations are far enough from the ligand to predict it as a possible multi HER kinase inhibitor.

Modelling results are very incouraging as they predict a possible cross-activity of a molecule targetted to the interface on all the members of the HER family.

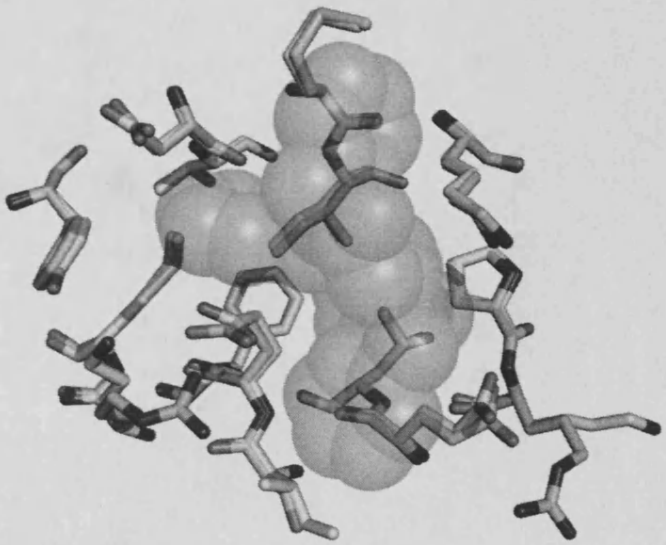


Figure 5.21: Her2 homology model: HER2 is shown in green and EGFR in yellow for comparison

5.7.5 Lead optimisation

The predected pose for the putative interface inhibitor was used as a starting point for a lead optimisation project aimed at the design of more diverse structures which would retain and possibly increase the inhibitory potential.

This project was carried out by a project student using the haptic-based approach described earlier. An analysis of the binding pose revealed that three out of the four benzene rings fit in hydrophobic pockets on the surface of the enzyme, while the forth seems to have limited interactions with the protein. Starting from an analysis of

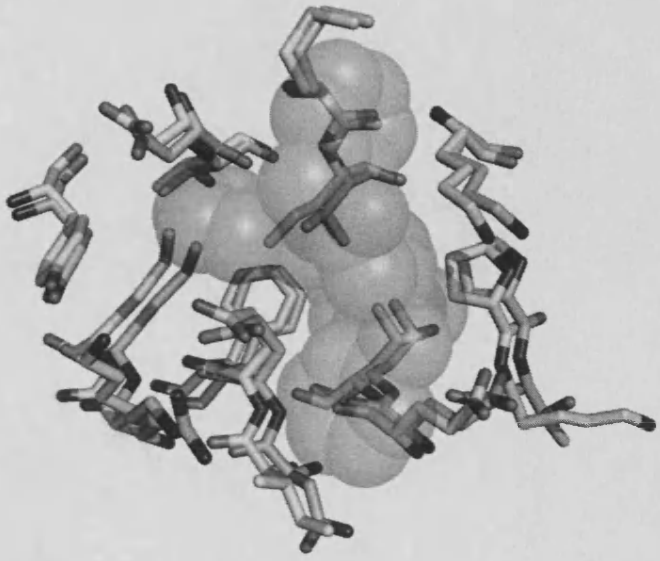


Figure 5.22: Her3 homology model: HER3 is shown in green and EGFR in yellow for comparison

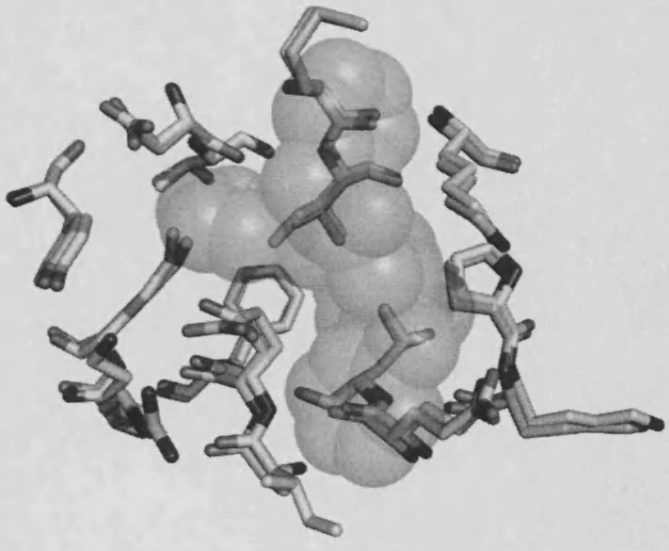


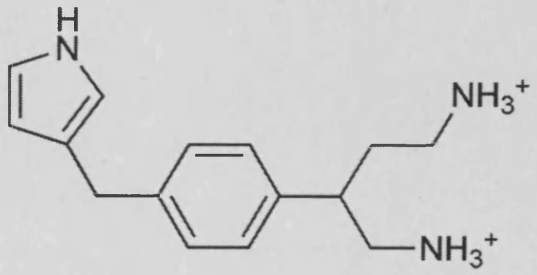
Figure 5.23: Her4 homology model: HER4 is shown in green and EGFR in yellow for comparison

these three pockets and of the aminoacids surrounding them, a library of fragments was derived, and each one was tested with the haptic device for interaction with the protein. The more favourable fragments were grown into bigger portions and retested. The process was repeated several times till the single fragments merged into molecules, which were finally docked to confirm their poses. Four among the resulting compounds, whose poses are all confirmed by docking, appeared to be particularly significant and are shown in Figures 5.24, 5.25, 5.26, 5.27.

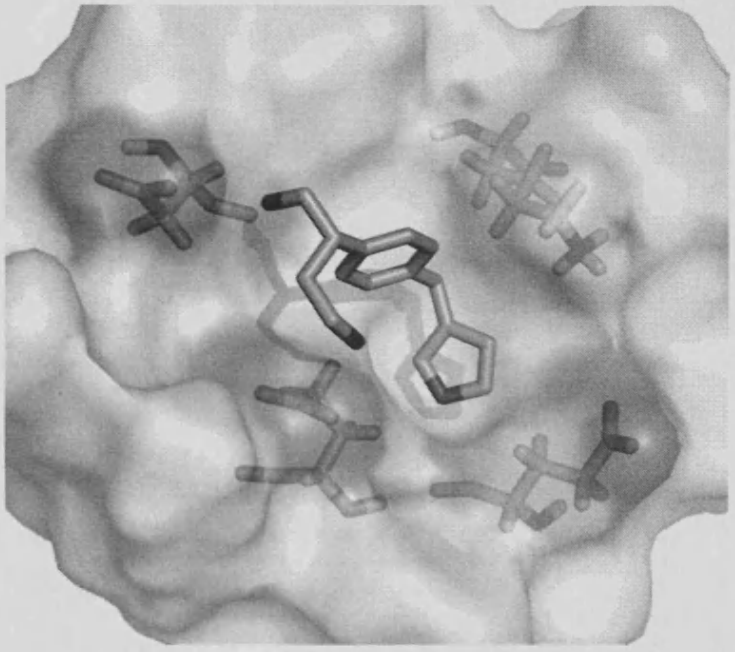
5.7.6 Peptidomimetics approach

Drug design often takes natural ligands as templates, and surface inhibitor design is no exception [66]. As an alternative route to design an inhibitor to target EGFR kinase domain dimerisation interface the interacting residues of the two monomers were considered. Residues 916-926 form a small α helix, that in the dimer interacts tightly with the surface on the N lobe.

The synthesis and in vitro testing of the activity of this portion could provide insight on viability of the dimerisation interface as a target.
Molecular dynamics simulations were performed to predict the stability of the helix in solution. A peptide with sequence: CTIDVYMIMVKCW, corrisponding to AAs 915-927 was allowed to evolve for 20 ns from an α helix structure. As shown in Figure 5.28, the Tryptophan residue at the C term resulted to be very flexible and its side chain movements were thought to compromise the stability of the secondary structure. A second dynamics was run mutating the tryptophan to alanine, which improved the stability of the helix in the simulation. The deleted form (CTIDVYMIMVKC) was also investigated, and appeared to be more stable than the other two, and was therefore chosen for the second step. A complex between the helix and the kinase monomer was generated, and allowed to evolve for 5 ns in a molecular dynamics simulation. As shown in figure 5.29 the complex remained stable and all the main interactions between the protein and the peptide were conserved for the whole duration of the simulation.

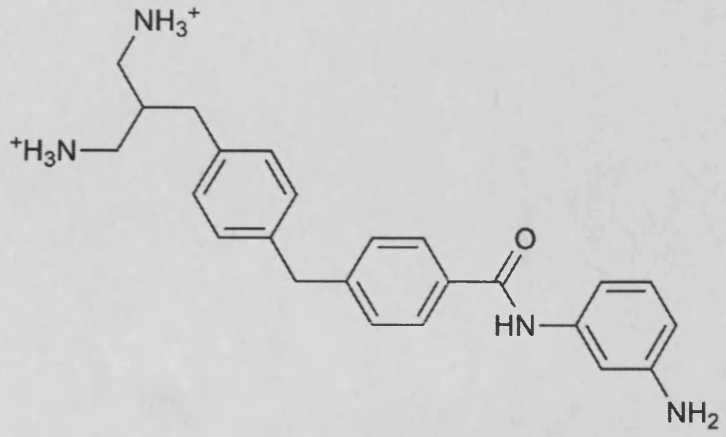


(a) Compound S1: structure

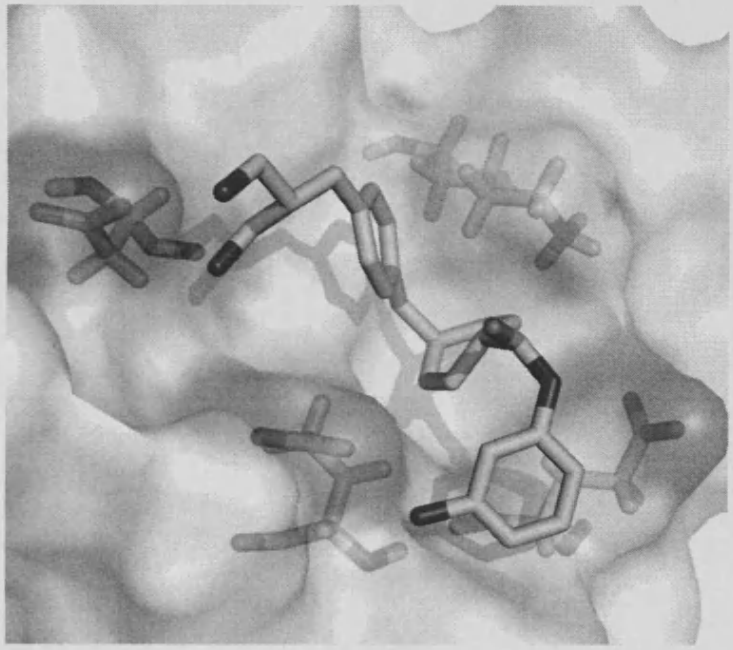


(b) Compound S1: docked pose

Figure 5.24: Fragment based lead optimisation: compound S1

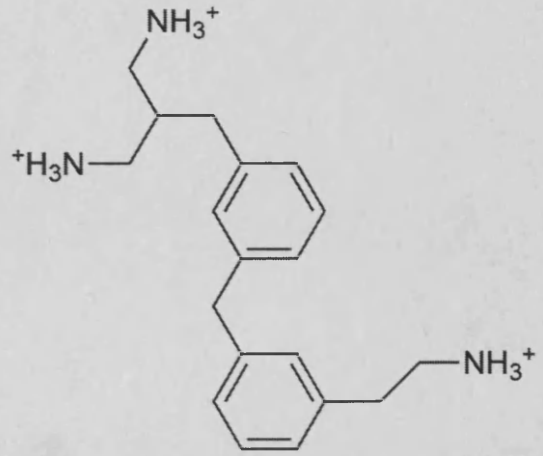


(a) Compound S2: structure

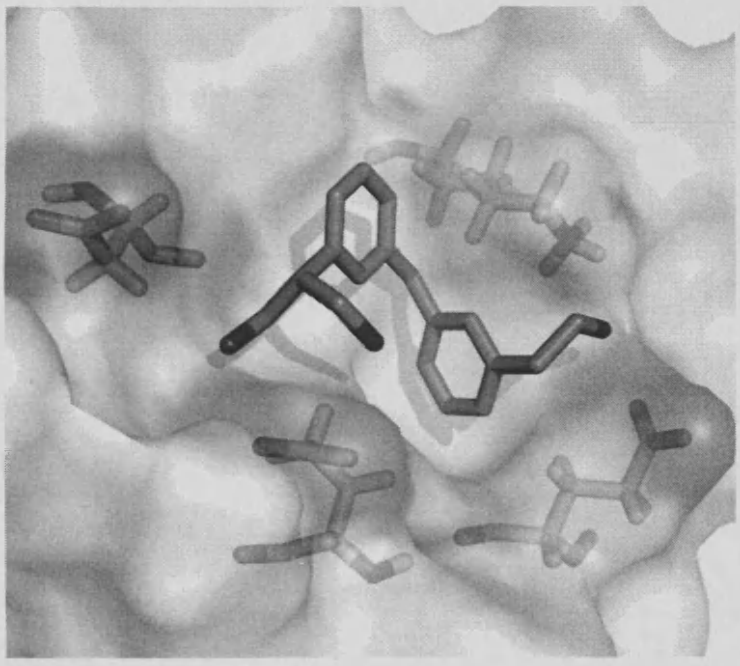


(b) Compound S2: docked pose

Figure 5.25: Fragment based lead optimisation: compound S2

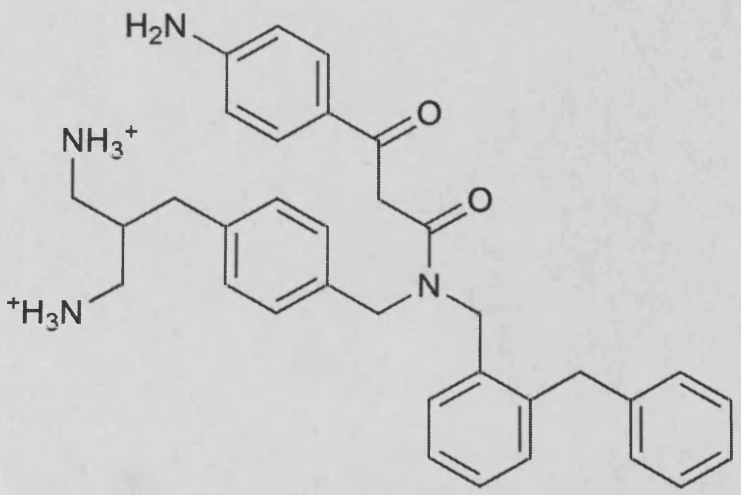


(a) Compound S3: structure

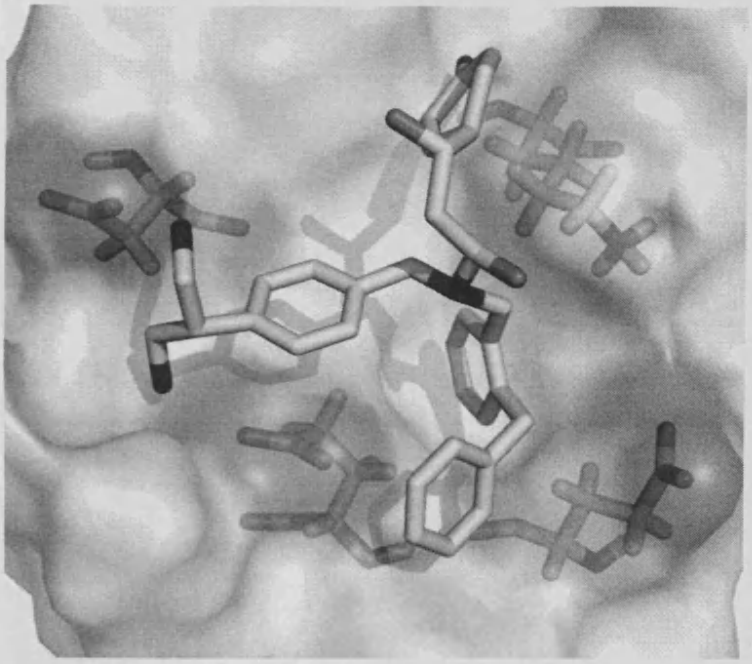


(b) Compound S3: docked pose

Figure 5.26: Fragment based lead optimisation: compound S3



(a) Compound S4: structure



(b) Compound S4: docked pose

Figure 5.27: Fragment based lead optimisation: compound S4

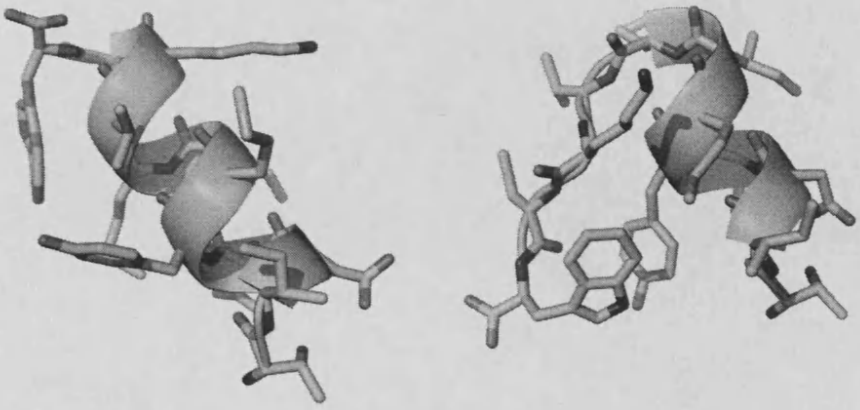


Figure 5.28: Helix stability studies: structure of the helix at the beginning and after 4 ns

5.8 Conclusions

The haptic based approach was applied to the development of inhibitors of the EGFR pathway together with other more traditional drug design approaches. Both in the case of active site inhibitors and allosteric inhibitors the approach was considerably useful in the process of lead identification and optimisation allowing a direct intervention of the user and a rapid test of hypotheses. The application resulted in a molecule that shows biological activity in in vitro tests.

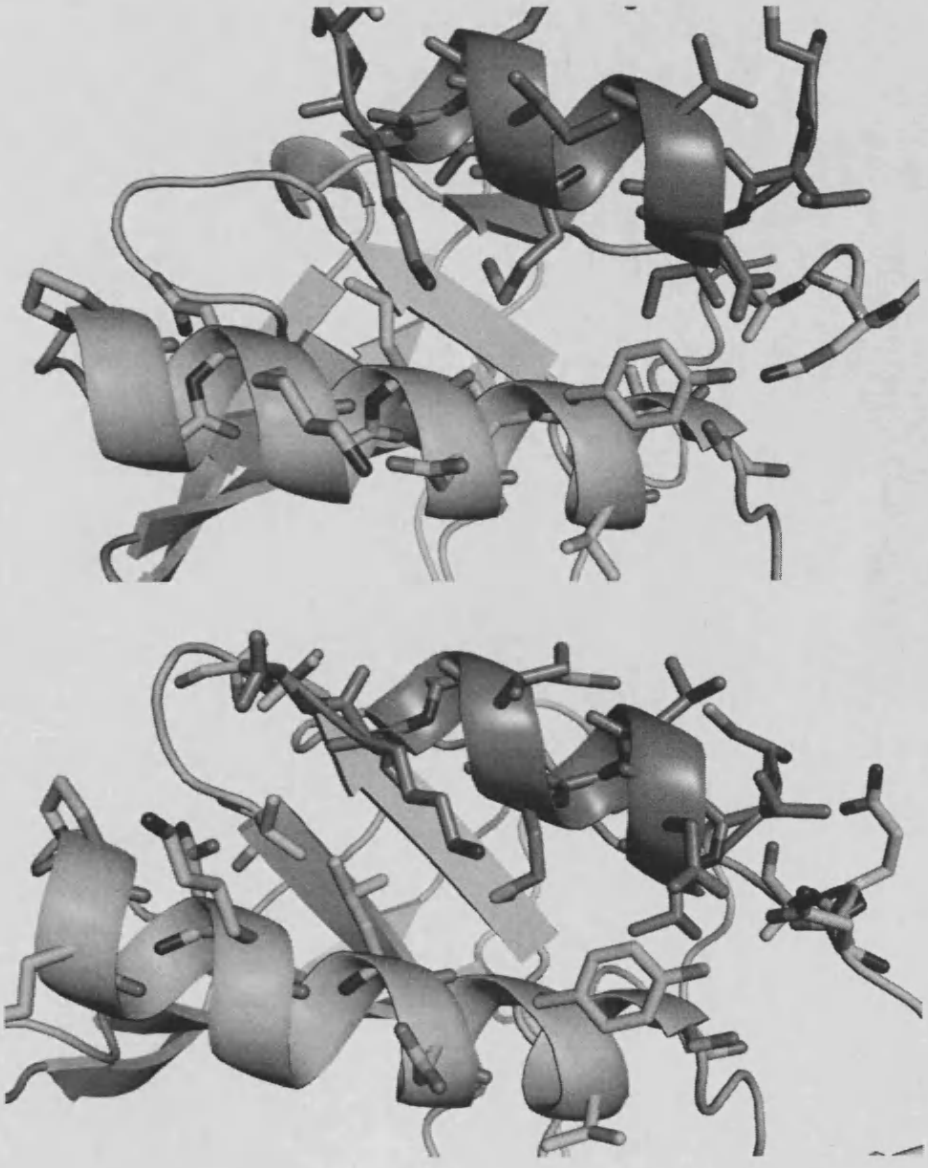


Figure 5.29: Complex stability studies: structure of the complex at the beginning and after 5 ns

Chapter 6

Lead Optimisation of an Anti-HCV Polymerase Inhibitor

6.1 Hepatitis C virus

Since its discovery in 1989, hepatitis C virus (HCV) has been recognised as a major cause of chronic liver disease and affects approximately 200 million people worldwide at the present time[67, 68], representing nearly 3% of the world's population. Persistent infection with HCV is associated with the development of chronic hepatitis, hepatic steatosis, cirrhosis, and hepatocellular carcinoma. In general, chronic hepatitis C infection is relatively asymptomatic and have few, if any, clinical manifestations prior to the development of cirrhosis.

The approved therapy for HCV infection shows only limited efficacy and no effective vaccine has been developed.[69, 70]

Liver disease associated with HCV infection is the leading indicator for liver transplantation in the United States.

6.1.1 Genome organisation

HCV is a small, enveloped RNA virus belonging to the Hepacivirus genus of the Flaviviridae family, which also includes several classical flaviviruses, including Dengue virus and Yellow Fever virus, as well as pestiviruses, such as Bovine Viral Diarrhea virus and the unassigned GB viruses [9,10]. The HCV genome consists of an RNA molecule, of approximately 9.6 kb, that contains a large open reading frame flanked by structured 5' and 3' non-translated regions (NTRs).

Viral proteins are translated as a polyprotein precursor via an internal ribosome entry site (IRES) located in the 5' NTR. The polyprotein undergoes a complex series of co- and post-translational cleavage events catalyzed by both host and viral proteases to yield the individual HCV proteins. The 10 HCV proteins are organized in the polyprotein in the order: NH2-C-E1-E2-p7-NS2-NS3-NS4A-NS4B-NS5A-NS5B-COOH[69]. An additional HCV protein, F (for frameshift protein) or ARFP (for alternate reading frame protein), generated by an overlapping reading frame in the core (C) protein coding sequence, has been proposed[71].

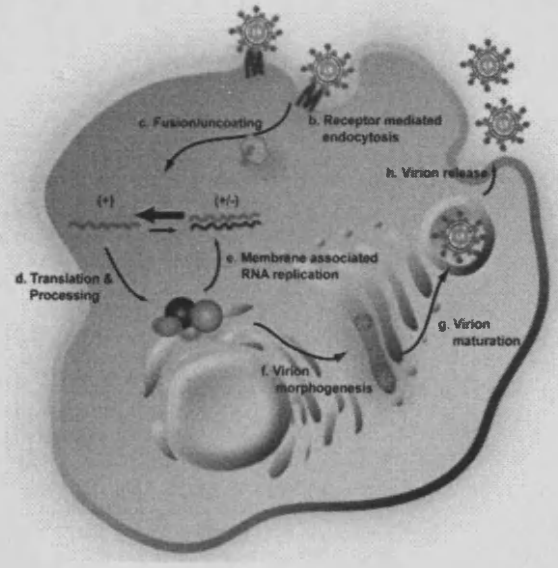


Figure 6.1: HCV life cycle. [72]

6.2 NS5B

NS5B is a 68-kDa protein with a conserved sequence motif characteristic of viral RNA-dependent RNA polymerase (RdRp), including a hallmark GDD motif that produces catalytic activity[69]. NS5B is a tail-anchored protein and its C-terminal 21-aa region forms an alpha-helical transmembrane domain, which is dispensable for polymerase activity in vitro but is responsible for post-translational targeting to the cytoplasmic side of the ER [67]. Analysis of the crystal structure of NS5B has revealed that the HCV RdRp resembles a right hand and contains fingers, palm, and thumb subdomains, similar to other template-dependent polymerases. Unlike the more open structures of other template-dependent DNA polymerases, such as the Klenow Fragment and the human immunodeficiency virus 1 reverse transcriptase, the HCV RdRp has a fully encircled active site through extensive interactions between the fingers and thumb subdomains, resulting in a protein that predominantly exists in a "closed" conformation.

The success of antiviral therapies based on chemotherapeutic agents targeting viral polymerases has prompted intense efforts to develop inhibitors of HCV NS5B. Studies with HIV reverse transcriptase validate the clinical utility of two distinct classes of viral polymerase inhibitors, nucleoside and non-nucleoside inhibitors. Nucleoside inhibitors function as competitive substrate analogs that prevent RNA chain elongation when incorporated by the viral enzyme, resulting in premature chain termination. HIV reverse transcriptase non-nucleoside inhibitors bind to a site residing outside the enzyme active site and inhibit catalysis by an allosteric mechanism. Several putative allosteric binding sites on the surface of HCV NS5B have been suggested based on recent structural studies, and several chemical classes of NS5B non-nucleoside inhibitors have been described. [73].

6.3 The ProTide approach

The first example of aryloxy phosphoramidate was reported in 1992 [74] as prodrugs of AZT. Further development of the approach resulted in ProTides containing alkyl or haloalkyl esters in place of the aryl moiety and structure activity relationship studies of the amino acid side chain [75, 76].

The two masking groups represented by the aminoacid ester and thee aryl moiety increase the lipophilicity of the molecule allowing the nucleotide monophosphate on passive diffusion into the cell. The activation results in the release of the free monophosphate, bypassing the first step of phosphorylation[77].

The phosophoramidate approach was later successfully applied to a number of different nucleotide analogues greatly improving their antiviral or anticancer activity.

6.4 Aims

2'-Methyl guanosine (Figure 6.2 is a known inhibitor of NS5B polymerase activity [78]. The protide approach has been successfully applied to this drug to bypass the first phosphorylation step dramatically increasing the availability of the monophosphate form inside the cell. This project is focused on the optimisation of the parent drug and the exploration of its structure-activity relationships, and aims to the discovery of derivatives with a better selectivity and activity profile and enhanced pharmacodynamic properties [79, 80].

6.5 Model building and refining

HCV NS5B is one of the most challenging enzymes for docking programs [81]. Although crystallised many times under different conditions, there is no reliable structure of the complex with a template RNA strand to date, which is essential for structure-based approaches to the design of nucleoside analogue inhibitors. Its high

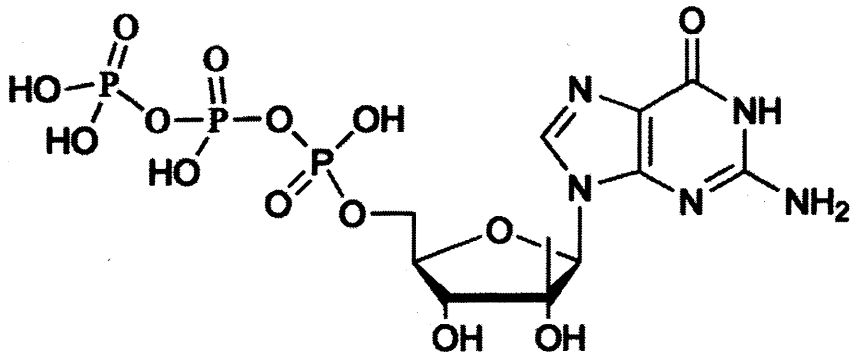


Figure 6.2: Structure of the lead compound

degree of flexibility (common to all the polymerases) and the presence of not completely understood allosteric regulation mechanisms make the task of building a model even more challenging. On the positive hand, however, the literature on the field is copious and provides a relevant amount of experimental data that can be used for validation purposes. In particular, Merck has published a paper [73] describing the construction and validation of a model for NS5B nucleoside analogue inhibitors. An approach similar to what they describe has been followed initially to build a model.

A crystal structure of NS5B (PDB code 1GX6) and one of the bacteriophage ϕ 6 RNA-dependent RNA polymerase (PDB code 1H10) were superposed. 1H10 is crystalised in the initiation complex with an incoming NTP, a primer, various metal ions and a template strand. These were copied into the NS5B active site and then energy minimised to correct the geometries (using the MOE implementation of MMFF94x). Preliminary docking studies on the model confirmed its ability to dock active GTP inhibitors in the pocket. However 2'C-Me-GTP, the lead structure, although correctly forming the base paring interactions, appeared to be too far from

the protein to justify the evidence for its inactivity on the S282T mutant. The region close to the 2'C appeared to be very open, while it is believed that region is closely interacting with the ligand, selecting ribonuclotides over deoxy-ribonuclotides. A structural refinement stage was therefore necessary before the model could be used. This was done by running a short molecular dynamics simulation on the protein-ligand complex. GROMACS was used to run the simulation for 4 ns of simulated time. The trajectory was analysed, monitoring the distance between the C' Methyl and Ser282 as a parameter. The result showed that the distance was shortening, while the overall structure of the enzyme and the base pairings remained constant. The snapshot with the minimum distance was chosen, and further minimised with MMFF94x.

Docking tests confirmed the minimised model to be in general more predictive than the non minimised one.

6.5.1 Model for uracil binding substrates

The template strand in the model is a poly-dC oligonucleotide. While this is appropriate to model the binding of 2'C-Me-GTP and most of its derivative, some modifications on the purine base required the investigation of possible paring with a different base. Therefore a model with a uracil-based template was also built. This was done by mutating the last cytidine base (the one paring with the inhibitor) to uracil, minimising the structure (in its apo form) and then minimising again with 6-Me ATP in the binding site, to get optimal orientation for the base paring. The building of the other two possible models (with purines as template) was not necessary, since the investigation of possible base mutants was not extended to pyrimidines.

6.5.2 Model validation

The docking approach was validated in its ability to retrieve the pose of the lead molecule. Docking results of the triphosphates were rather poor. This is probabily

because the triphosphate chain is charged and highly flexible, therefore able to adapt itself to the pocket and to form very strong interactions that can prevail on the base paring and lead to mis-score bad poses. The docking of nuclosides, besides being much faster (around 6-fold), yielded much better results. The sugar, however, appeared to be very flexible and its position was very variable from one run to the other. The docking of nucleotides (monophosphated nucleosides) improved the results, locking the 5' C in its correct position and therefore constraining the orientation of the whole sugar.

6.6 SAR

Figure 6.3 shows the main positions that has been explored for derivatives. Due to the high degree of approximation of the model and the non-trivial correlation between binding and activity, the binding affinity of the modelled compounds had to be studied in a semi-quantitative if not qualitative fashon. Poses were analysed focusing on the mantainment of a correct base pair alignment and a good position of the sugar in the active site. The number and the orientation of hydrogen bonds with the protein and the presence of other weak interactions was also considered.

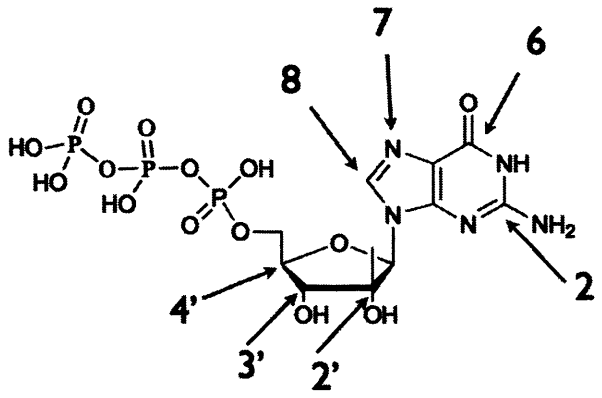


Figure 6.3: Structure activity relationship.

6.6.1 2' substitutions

NS5B is naturally able to select nucleotides over deoxy-nucleotides. It is thought that this is due the 2'OH donating an hydrogen bond to Asp225 side chain[82]. However -F derivatives have been reported to have strong activity[82], suggesting a possible role for hydrogen bond acceptors as well in this position. Docking studies confirm the binding of 2'-F derivatives.

The structures of the modelled compounds are reported in Figure 6.4 and the predicted affinity summarised in Table 6.1.

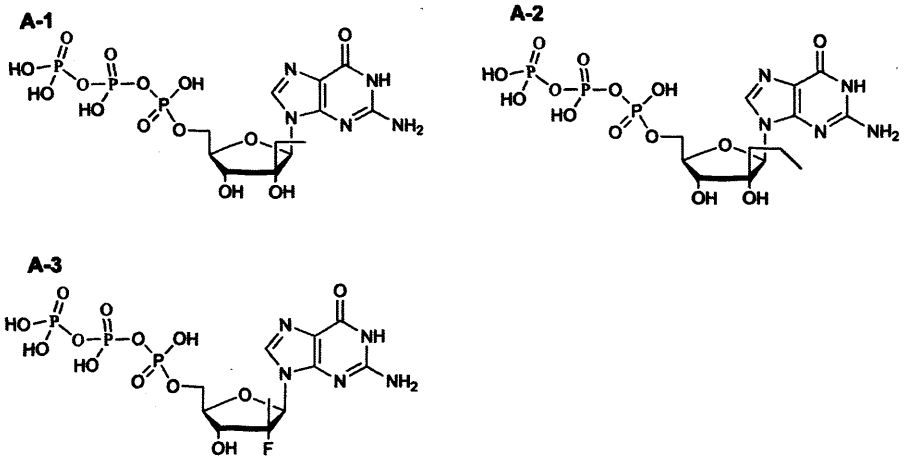


Figure 6.4: 2' modifications

Compound	Score
A-1	V
A-2	X
A-3	V

Table 6.1: predicted affinity: X weak binding or not binding in the conventional pose, V binding in the conventional pose, VV strong binding in the conventional pose

As expected the fluoro derivative (compound A-1) shows a promising interaction profile (Figure 6.5). The methyl group in the 2' position of the parent drug causes steric interference and doesn't allow the next nucleotide to bind correctly after the drug is incorporated. However extending this substituent causes a worse binding of the drug to the protein. 2'Ethyl derivatives (compound A-2, docking pose in Figure 6.6) are predicted to be less active (as confirmed by in vitro studies) and the propyl derivative (compound A-3) completely prevents binding (and is inactive in vitro).

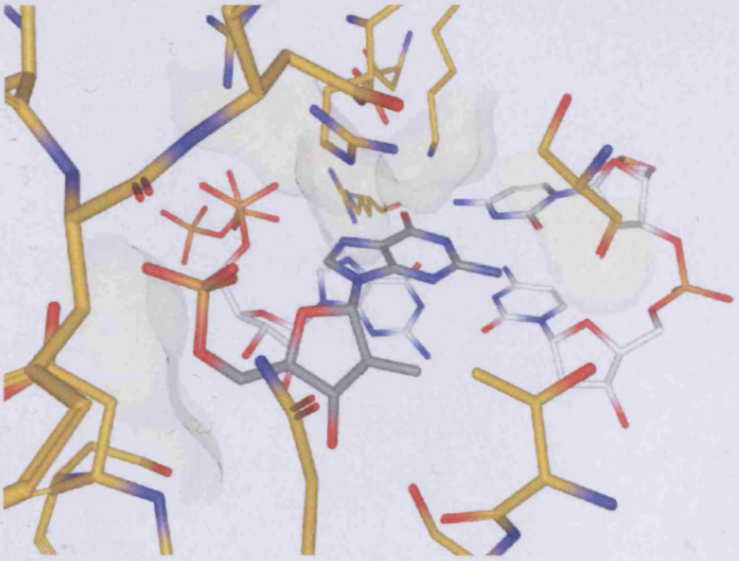


Figure 6.5: Docking results of the 2'F derivative: the binding pose of the parent molecule is conserved, and the fluorine atom in 2' accepts an hydrogen bond

6.6.2 3' substitutions

The structures of the modelled compounds are reported in Figure 6.7 and the predicted affinity summarised in Table 6.2.

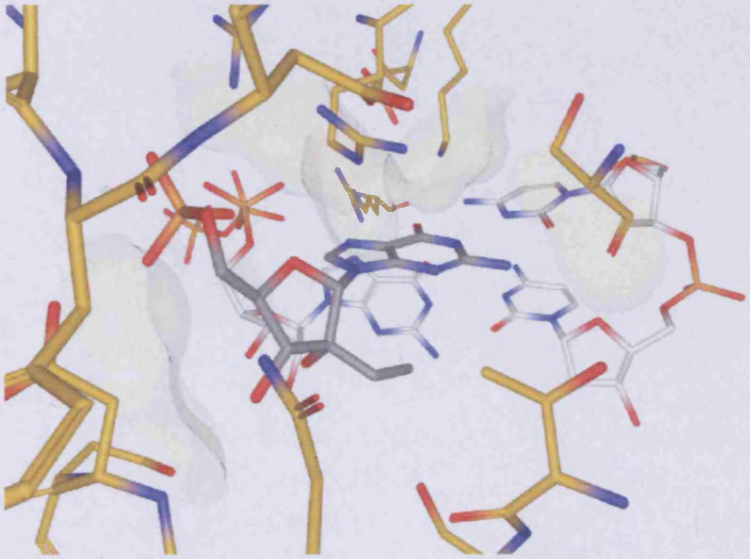


Figure 6.6: Docking results of the 2'ethyl derivative: the binding pose is similar to that of the parent molecule, but steric interactions of the ethyl group cause a shift in the position of the sugar

Compound	Score
B-1	V
B-2	VV
B-3	VV
B-4	VV
B-5	V
B-6	V
B-7	V
B-8	V
B-9	V
B-10	V

Table 6.2: predicted affinity: X weak binding or not binding in the conventional pose, V binding in the conventional pose, VV strong binding in the conventional pose

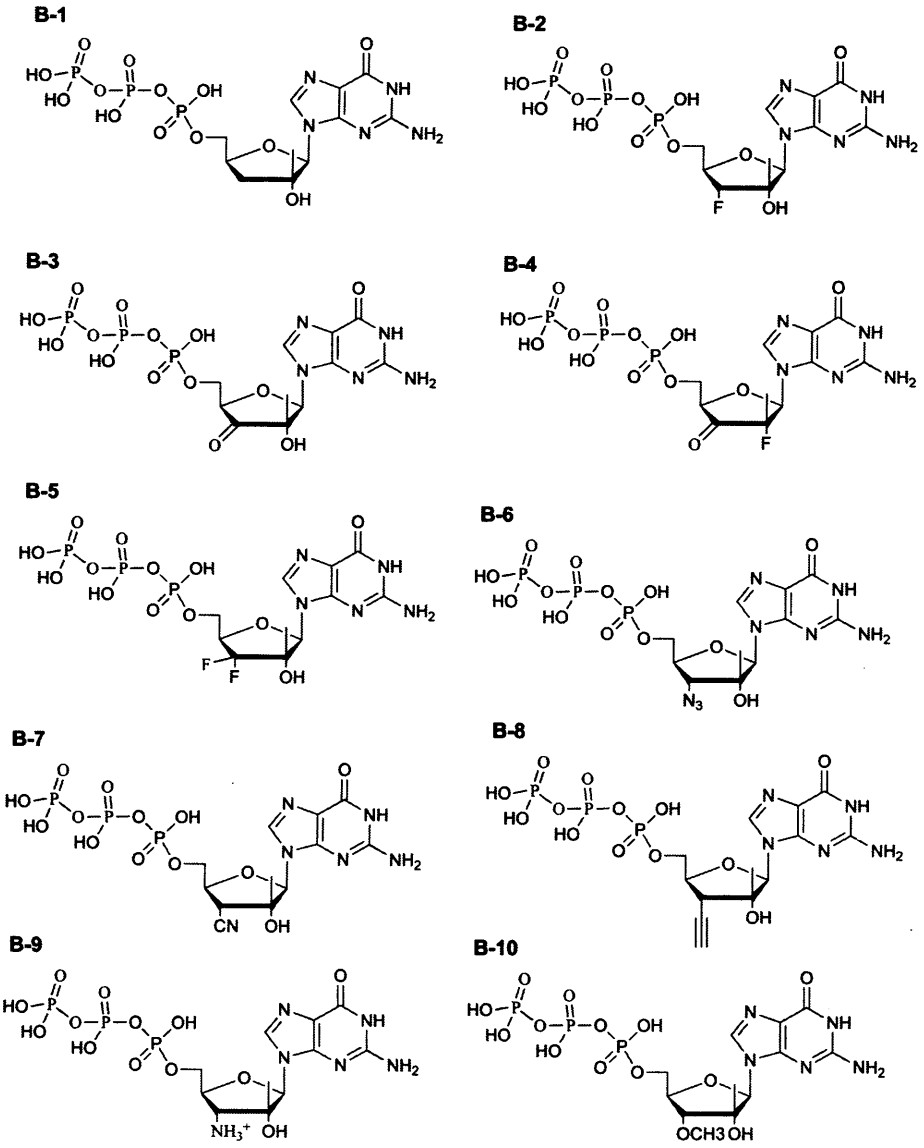


Figure 6.7: 3' modifications

The 3' OH is involved in an hydrogen bond with the enzyme. 3'-Deoxy nucleotides are a common strategy in polymerase inhibition, because they act as chain terminators since they do not provide a free OH for the condensation of the incoming nucleotide. The docking of the 3'-deoxy derivative (compound B-1) showed that the drug still fits in the active site, but the orientation of the sugar appears to be much more variable (Figure 6.8). The loss of the hydrogen bond seems to cause instability in the sugar and the binding is therefore probabily less effective. This unfortunately can not be confirmed with activity in vitro tests, as the 3'deoxy nucleotide is acting with a different mechanism (chain termination) and is therefore not directly comparable in terms of activity data. The keto derivative (compound B-3) shows a particularly promising binding pose. The conformation of the sugar is fixed and binding geometries appear to be optimal. This molecule is involved in keto-enol tautomerism that extends to the 2' OH. Substituting the 2' OH (with fluorine for instance, compound B-4) might be a good strategy to improve the binding affinity of this class of derivatives.

Many other substitutions appear to be tolerated here, including -F (compound B-2, Figure 6.9), di-F (compound B-5, Figure 6.10), 3'-azido (compound B-6), 3'-CN (compound B-7), 3'methoxy (compound B-10).

6.6.3 4' substitutions

The structures of the modelled compounds are reported in Figure 6.11 and the predicted affinity summarised in Table 6.3.

Compound	Score
C-1	VV

Table 6.3: predicted affinity: X weak binding or not binding in the conventional pose, V binding in the conventional pose, VV strong binding in the conventional pose

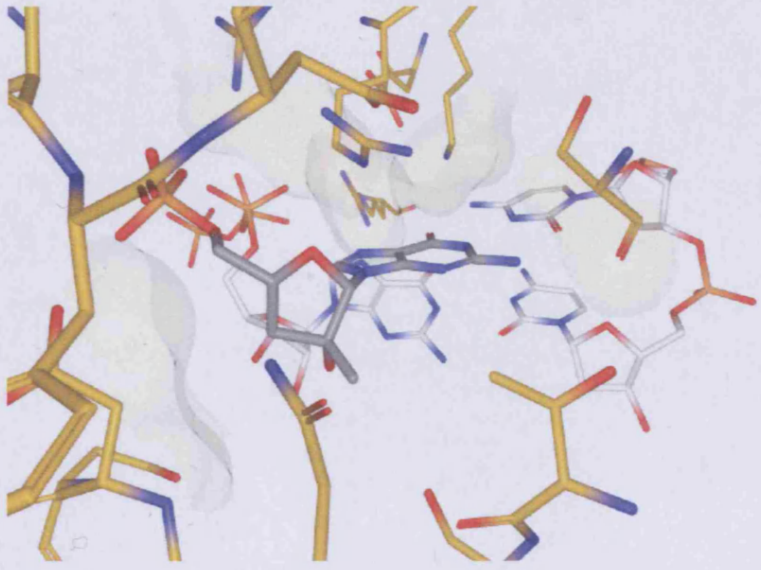


Figure 6.8: Docking results of the 3'deoxy derivative: the binding pose is similar to that of the parent molecule, but the position of the sugar is much less defined

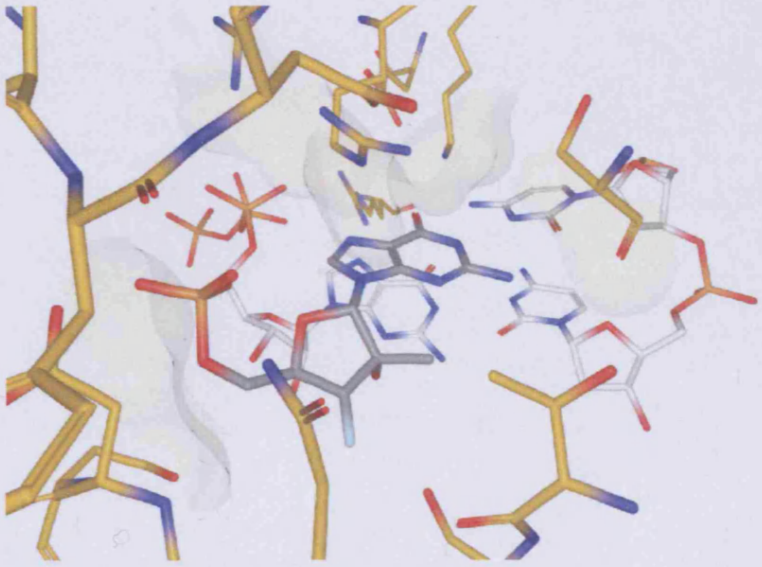


Figure 6.9: Docking results of the 3'fluoro derivative: the binding pose of the parent molecule is conserved

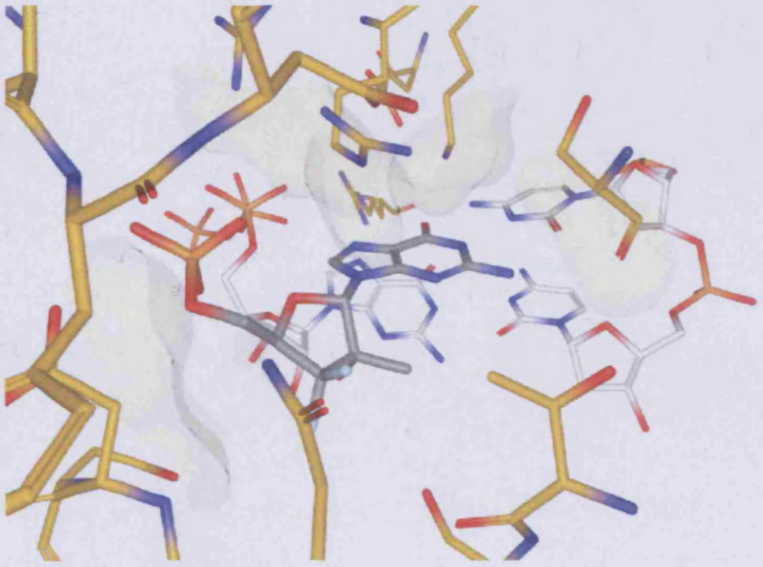


Figure 6.10: Docking results of the 3'gem fluoro derivative: the binding pose of the parent molecule is conserved

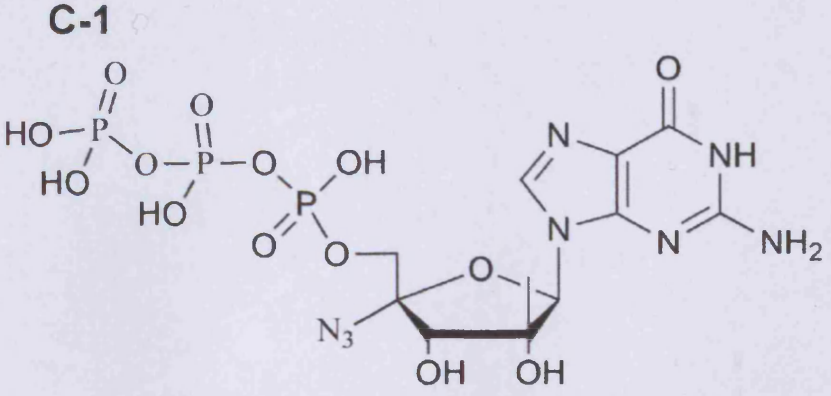


Figure 6.11: 4' modifications

Merck has published a series of 4' Azido derivatives that has showed activity against HCV NS5B[82]. To test the predictability of the model, the 4' Azido derivative (compound C-1) of 2'Me guanosine monophosphate was docked (Figure 6.12). The position of the base and the sugar are those expected for the parent nucleotide, and the azido moiety is accomodated inside a pocket of the active site.

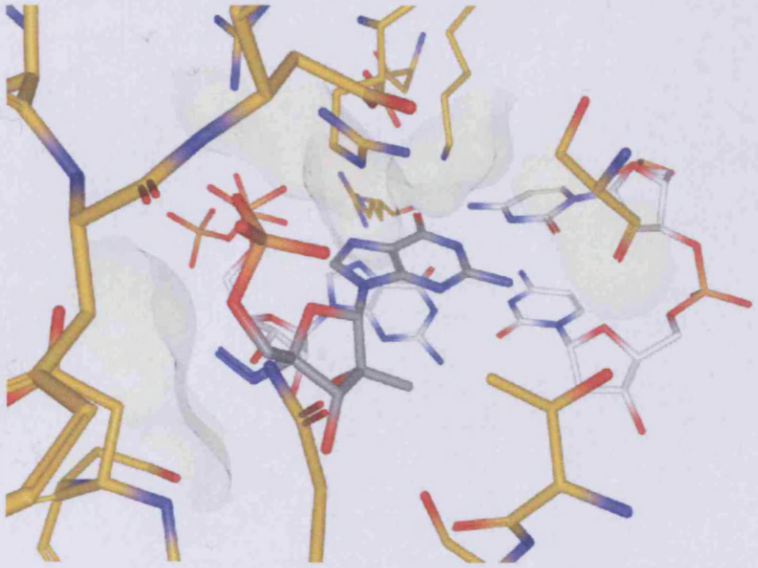


Figure 6.12: Docking results of the 4'azido derivative: the binding pose of the parent molecule is conserved, the azido group is accomodated in a subpocket of the active site

6.6.4 2 substitutions

The structures of the modelled compounds are reported in Figure 6.13 and 6.14 and the predicted affinity summarised in Table 6.4.

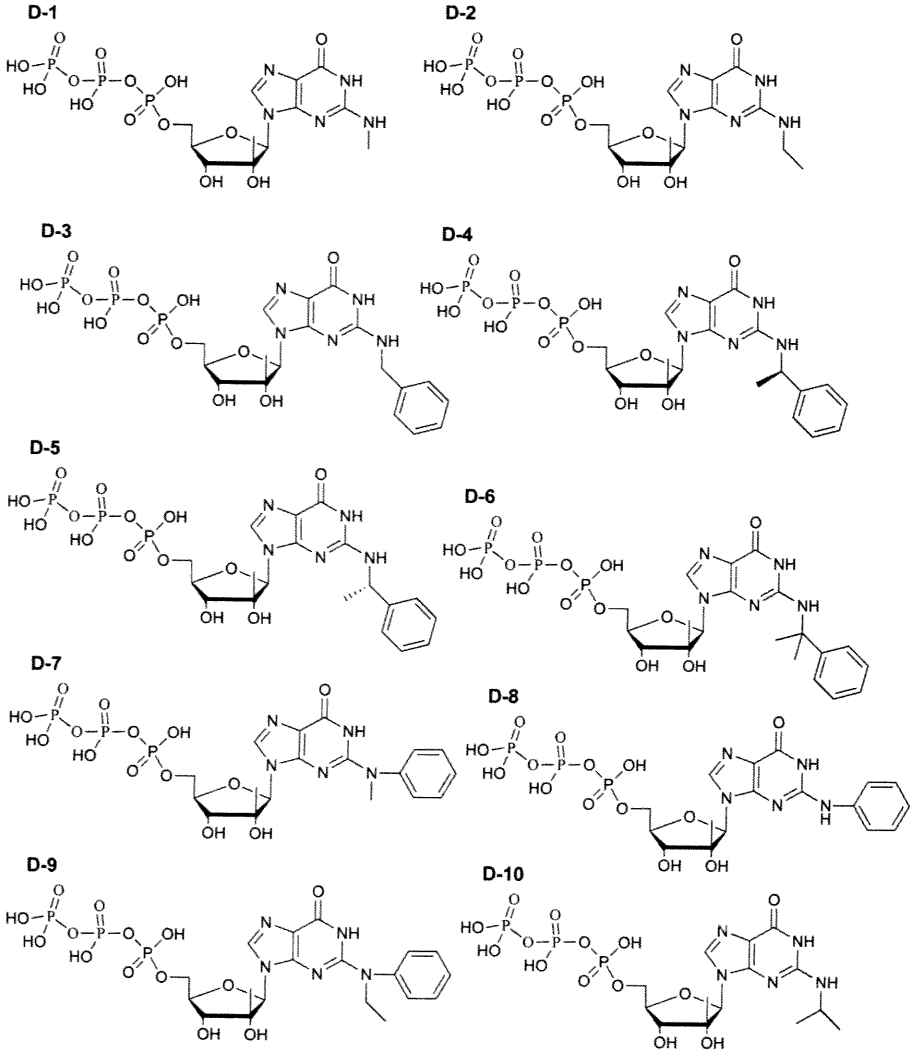


Figure 6.13: 2 modifications

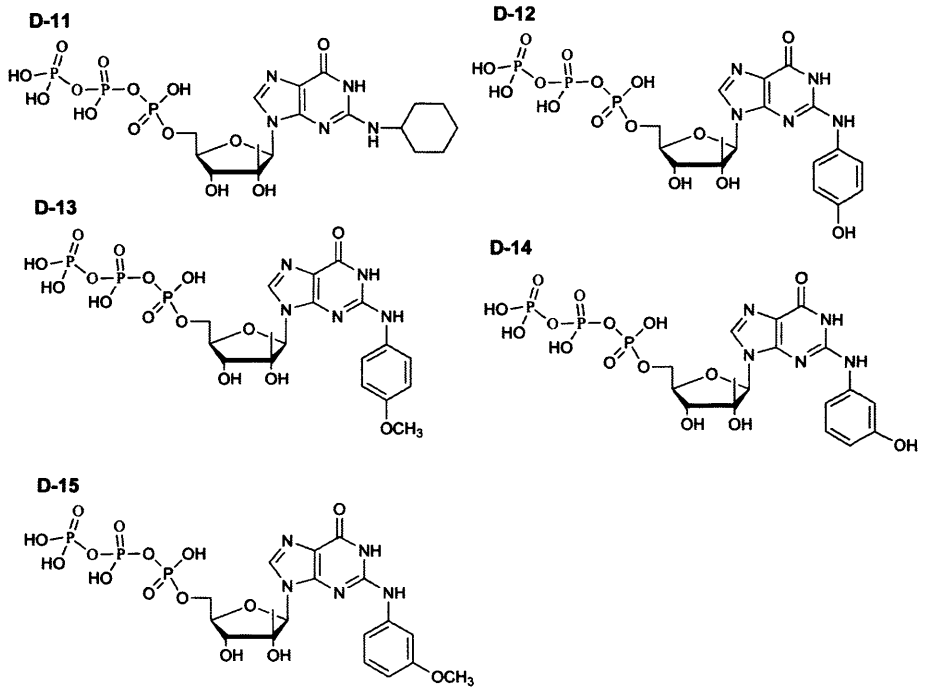


Figure 6.14: 2 modifications - follows

Compound	Score
D-1	V
D-2	V
D-3	VV
D-4	VV
D-5	V
D-6	V
D-7	VV
D-8	V
D-9	VV
D-10	V
D-11	V
D-12	V
D-13	V
D-14	VV
D-15	VV

Table 6.4: predicted affinity: X weak binding or not binding in the conventional pose, V binding in the conventional pose, VV strong binding in the conventional pose

The nitrogen in the 2 position is involved in an hydrogen bond which is essential for the base pairing. Substitutions which don't affect the hydrogen bonding capabilites of the nitrogen are tolerated, even with bulky substituents. A good example is the 2N benzyl derivative (compound D-2, Figure 6.15).

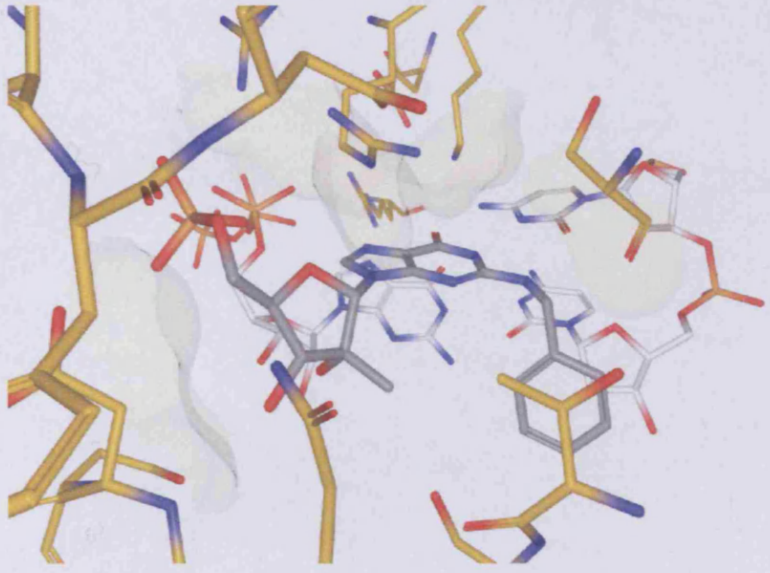


Figure 6.15: Docking results of the 2 benzyl derivative: the binding pose of the parent molecule is conserved

6.6.5 6 substitutions

The structures of the modelled compounds are reported in Figure 6.16 and the predicted affinity summarised in Table 6.5.

The 6 keto moiety is involved in the base paring as an hydrogen bond acceptor. Its replacement with a primary amine gives an active compound. Although it cannot

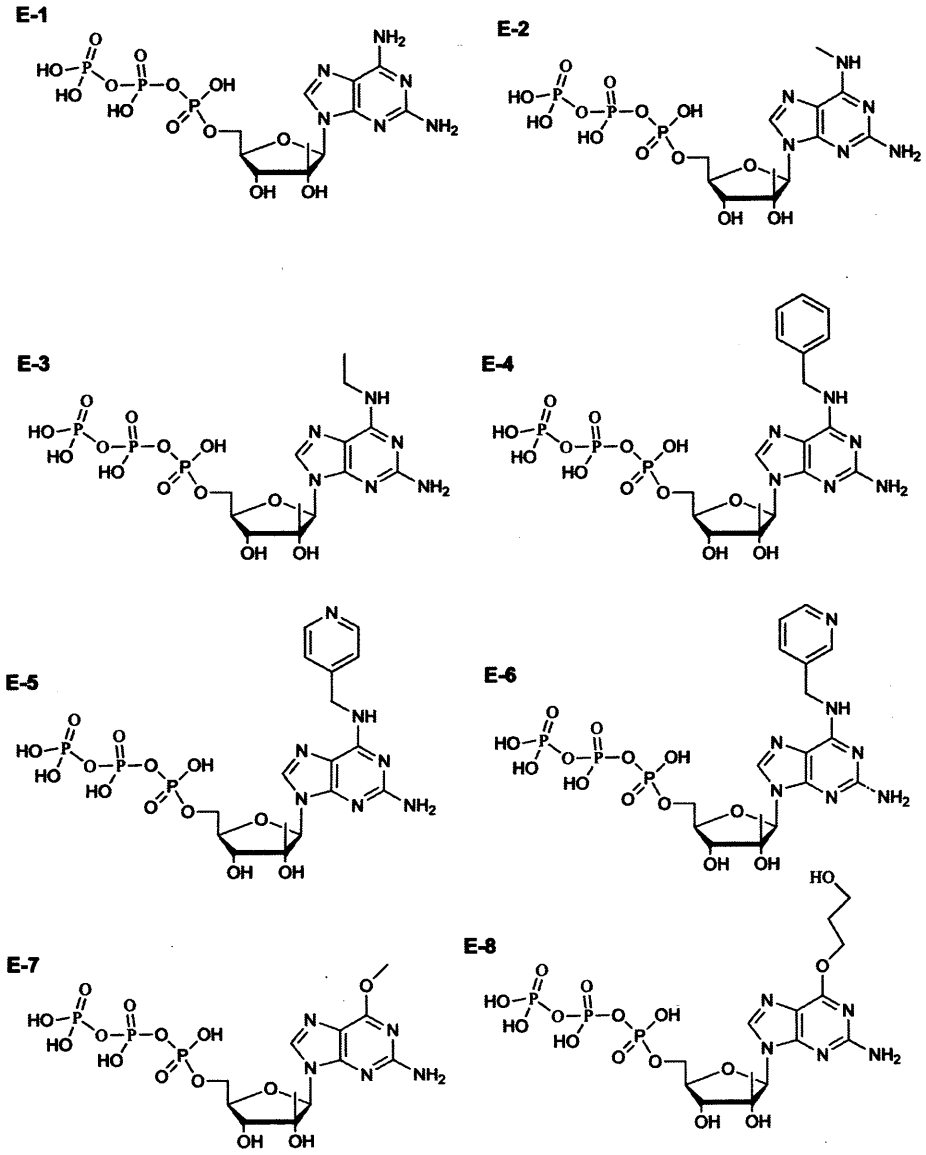


Figure 6.16: 6 modifications

Compound	Score
E-1	VV
E-2	VV
E-3	VV
E-4	VV
E-5	VV
E-6	VV
E-7	V
E-8	V

Table 6.5: predicted affinity: X weak binding or not binding in the conventional pose, V binding in the conventional pose, VV strong binding in the conventional pose

be docked using the conventional model, this compound is successfully docked in the model bearing Uracil in the template strand. In fact the compound shows the same interactions pattern as adenine (plus an additional hydrogen bond from the amino group in 2). However, docking studies also predict a second pose (Figure 6.17), with similar binding energy. In this second pose the three hydrogen bonds with uracil are conserved, but the coordinates of the nitrogens in 2 and 6 are inverted. Whether this second pose is biologically meaningful or is just a docking artifact is not clear, but could be investigated by comparing the biological activities of different 2-N and 6-N substituted derivatives.

6.6.6 7-8 substitutions

The structures of the modelled compounds are reported in Figure 6.21 and the predicted affinity summarised in Table 6.6..

7-deaza compounds (Figure 6.22) have been reported in the literature as NS5B inhibitors. They usually show a better activity compared to their nitrogen counter-

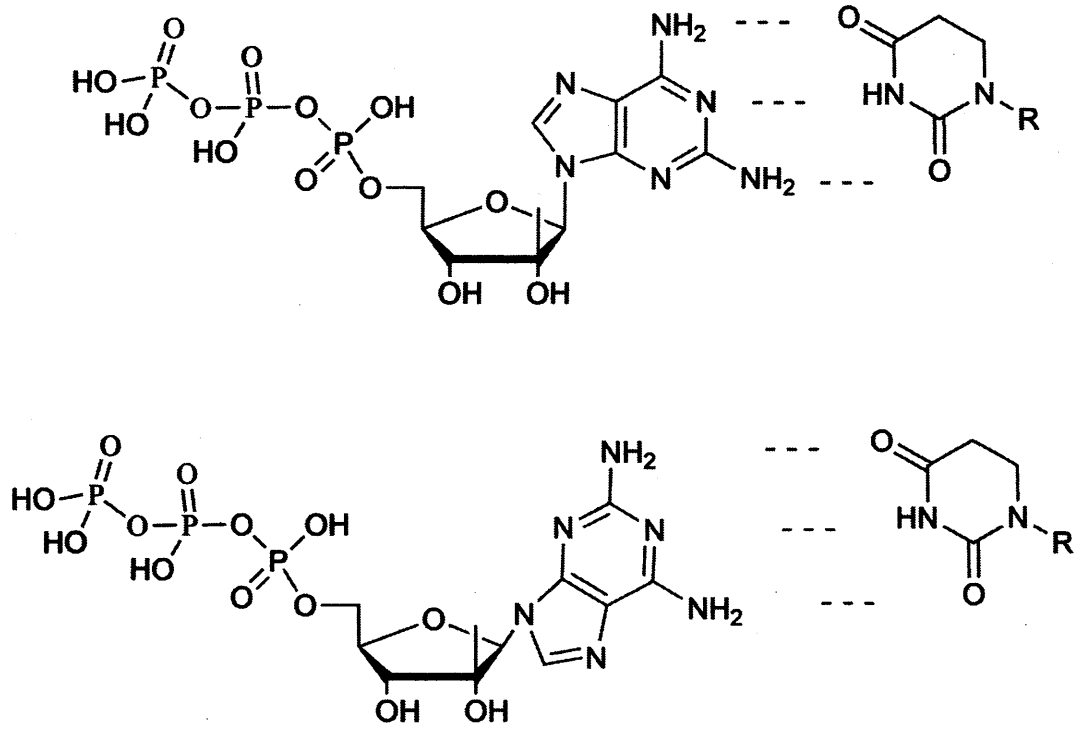


Figure 6.17: 6-amino derivative poses. 6 amino compounds show two different binding modes.

Compound	Score
F-1	VV
F-2	VV
F-3	V
F-4	V
F-5	VV
F-6	VV
F-7	X
F-8	X

Table 6.6: predicted affinity: X weak binding or not binding in the conventional pose, V binding in the conventional pose, VV strong binding in the conventional pose

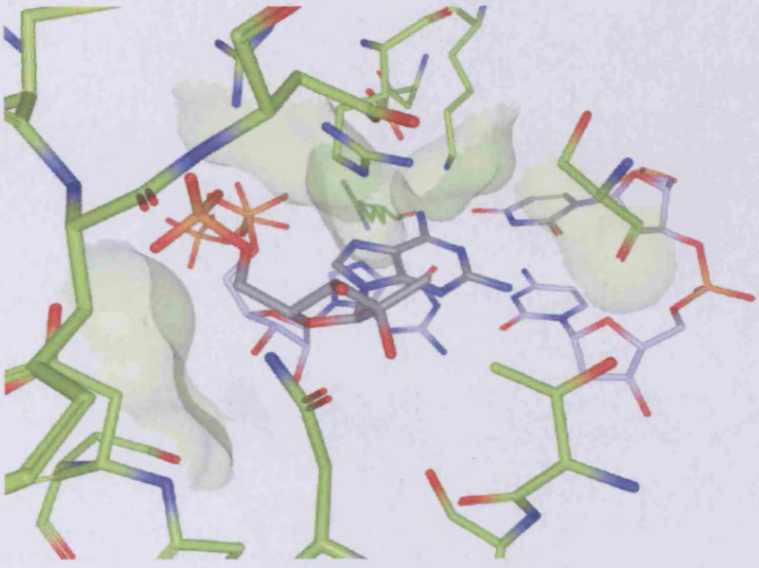


Figure 6.18: Docking results of the 6 amino derivative: direct pose

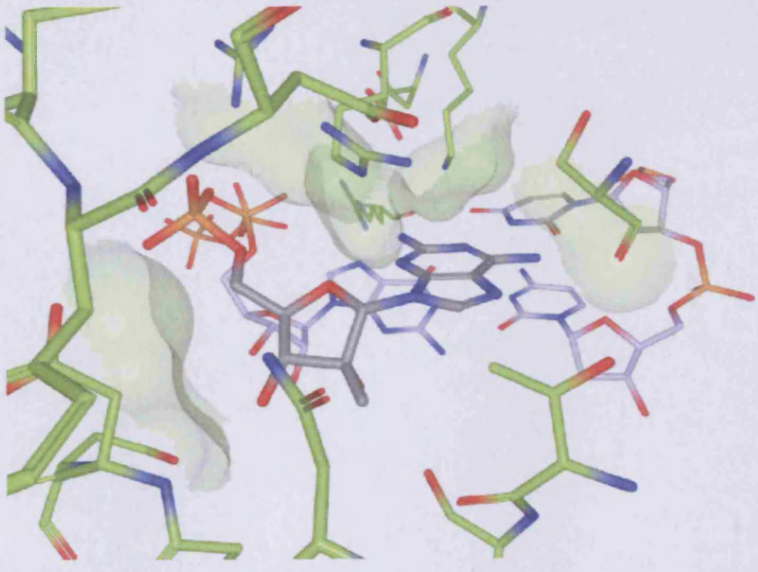


Figure 6.19: Docking results of the 6 amino derivative: inverted pose

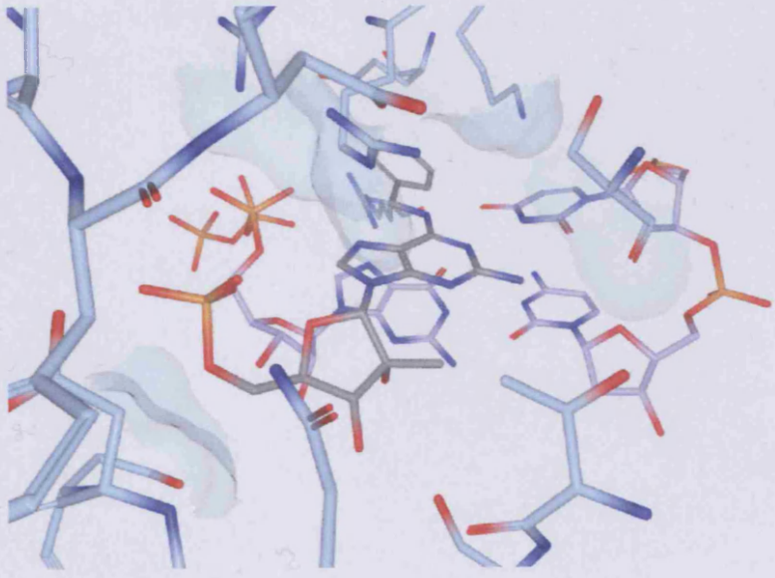


Figure 6.20: Docking results of the 6 benzyl derivative: the binding pose of the parent molecule is conserved

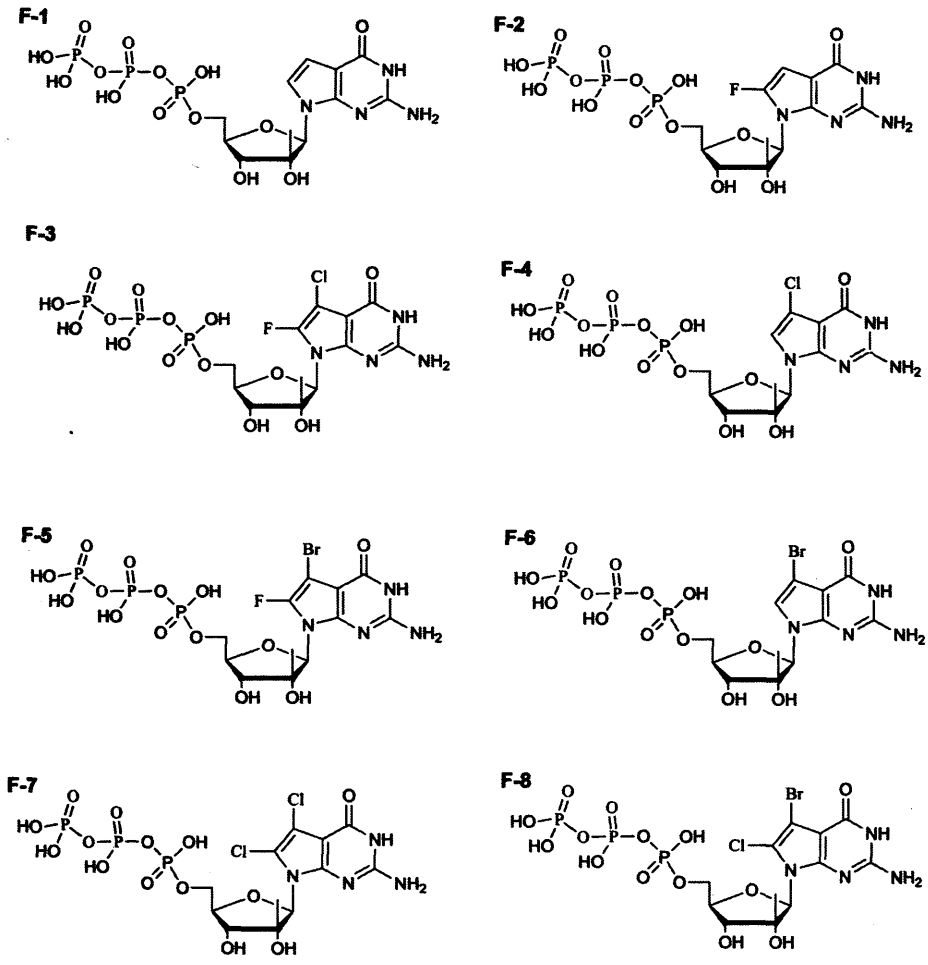


Figure 6.21: 7-8 modifications

parts. The analysis of the electrostatic potential in the region of binding provided a rationale for it. The nitrogen in 7 is pointing towards a negatively charged region. The CH in the deaza derivative is polarised, with a partial positive charge on the hydrogen. Therefore CH is favoured for electronic (as well as steric) reasons over N in the 7 position. To exploit this feature, some 7C substituted compounds has been designed and docked. In particular, 7-CBr has shown promising results. The steric interaction of bromine appears to be optimal, and a electron withdrawing group in 8 should be able to induce a positive charge on the bromine atom. The most promising compound of this series appeared to be 7-Br, 8-F (compound F-5).

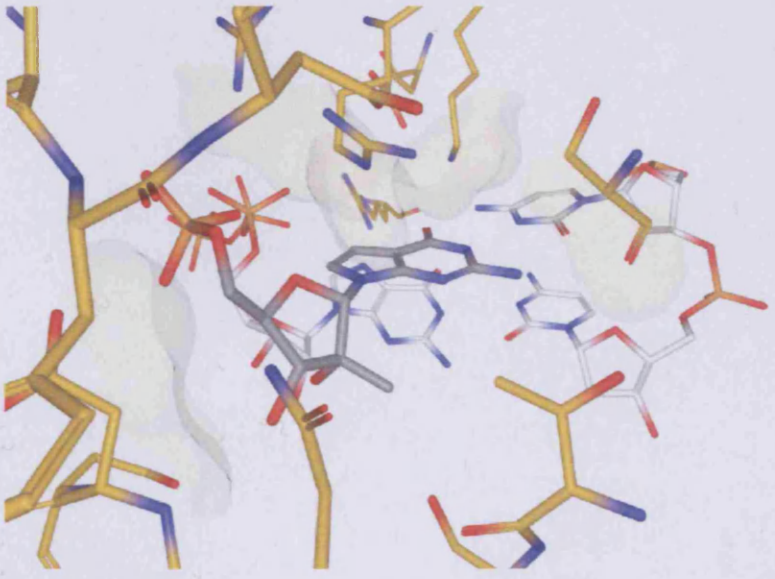


Figure 6.22: Docking results of the 7 deaza derivative: the binding pose of the parent molecule is conserved

6.7 Conclusions

The project on HCV polymerase resulted in the construction of a model which showed good agreement with biological data. It has been used to direct the synthesis of

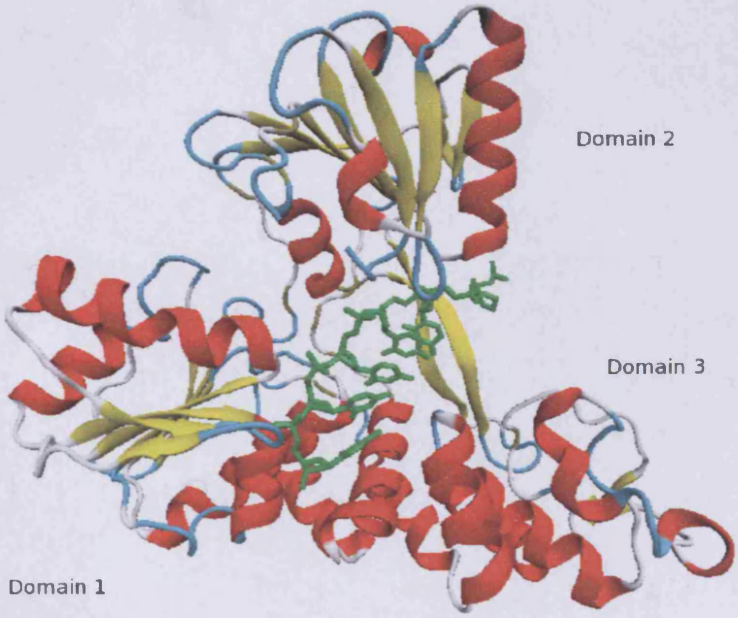


Figure 7.1: Structure of the NS3 enzyme

potentially important as therapeutic agents [91, 98, 99, 100, 101]. The necessity of the unwinding activity mediated by the NTPase/helicase enzymes in the viral life cycle also has recently been reported in knock-out experiments that demonstrated unambiguously that the switch-off of the helicase activity abolishes the virus propagation of bovine diarrhea virus (BVDV) and of dengue fever virus (DENV)[102, 103]. Inhibition of the helicase enzyme has the potential not only to terminate the proliferation of the virus but also to indirectly stimulate a cellular antiviral ds RNA sensing machinery because of the expected build up of intracellular dsRNA intermediates[104, 105, 106]. Recently, a preclinical proof of concept for helicase inhibitors as antiviral agents has been obtained for Herpes simplex virus HSV[98, 99, 100, 101]. Using high-throughput screening (HTS) and optimization of the screening hits resulted in compounds that inhibited HSV growth in cell culture with little cytotoxicity and which were orally active in an animal model of HSV[107, 108]. Over the past decade, significant progress has been made in the development of selective helicase inhibitors as antiviral and anti-cancer drugs[109].

7.1.2 Aims

The standard treatment for chronic HCV infection so far is pegylated IFN- plus ribavirin[110]. Although the mechanism of action of these drugs is debated, with both antiviral and immunostimulatory mechanisms being implicated, the sustained response rates are far from ideal. Moreover, there is substantial associated toxicity and the likelihood of success depends on viral and host factors that are often beyond the control of patients and physicians, in addition to the high cost [111, 112]. Clearly, more effective and less toxic treatment regimens are needed. Among hepatitis C viral proteins, NS3 helicase is a valuable but yet largely unexplored target. NS3 helicase inhibition is a promising strategy because it has the potential not only to terminate the proliferation of the virus but also to indirectly stimulate a cellular antiviral response against double-stranded RNA [104, 105, 106, 113]. Moreover despite the presence on the market and in clinical trials of several anti-NS5B inhibitors[114], there are no marketed NS3 inhibitors. The development of such molecules would be greatly

beneficial from a combined treatment perspective[115], The NS3 helicase enzyme has multiple functional domains that present multiple potential mechanisms of action for the design of an inhibitor.

7.2 Previous work

A modelling project previously carried out in the lab has identified a lead molecule for inhibiting the HCV NS3 helicase activity [116]. The chemical structure of the molecule is shown in Figure 7.2 and its putative interactions with the RNA binding site are shown in Figure 7.3.

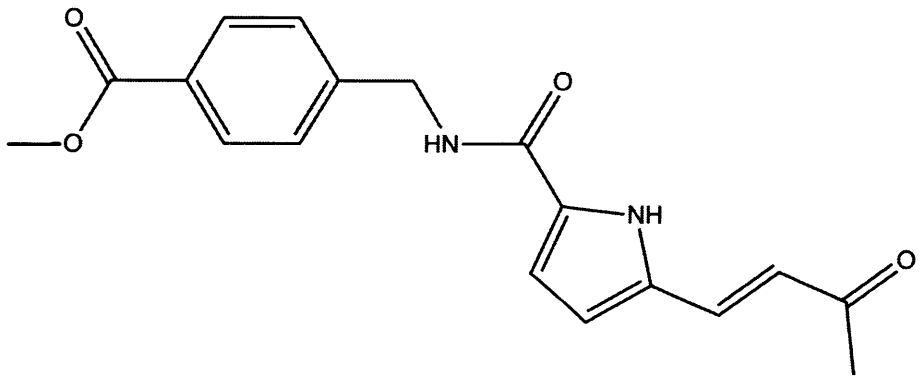


Figure 7.2: Structure of the lead compound

The molecule has a michael acceptor group capable of forming a covalent bond with Cys431, irreversibly blocking the catalytic function of the enzyme. The downside of such a reactive group is the potential of aspecific reactivity, which leads to the high toxicity of the compound. The aim of the optimisation is therefore substituting the reactive moiety with a less reactive group or a reversible binder while keeping or even increasing the activity of the whole molecule by possibly extending it to exploit further contacts with the active site.

Figure 7.3: Interactions of the lead compound

7.3 Methods

The in silico optimisation process was done using the fragment-based, haptic-driven approach described earlier. This is an ideal application because the process of lead optimisation is highly human-biased and the active side, being open and solvent exposed, represents a difficult target for docking and traditional de novo approaches. Every compound derived from the fragment based approach was validated with molecular docking using PLANTS (see appendices for the exact parameterisation).

7.4 Hot spots identification

The human-directed approach chosen for the development necessitates a good knowledge of the active site in terms of potential binding sites. The chemical identity and of the residues of the pocket was inspected. Potential maps were generated to identify areas more likely to be involved in hydrogen bonds, salt bridges and hydrophobic interactions. The haptic device was then used to probe the pocket with a negativly and successively positively charged atom. All the information gathered with these three approaches about the chemical properties and the forces of the different areas of the pocket were combined with the information from previous docking experiments

on the active lead to identify a series of hot spots (i.e. residues that can be used to form interactions with the growing molecule) for the fragment-based approach.

Arg393: this residue closes the pocket around the double stranded RNA and forms a salt bridge with the phosphate chain. This is also the putative binding parner of the carboxylate in the lead molecule.

Cys431: the putative target of the michael acceptor. This residue can be used for covalent binding.

Arg481: this second arginine is buried in the pocket and can provide a second strong interaction to orient a ligand inside the pocket.

Thr295, Thr433: The two hydroxyl groups are close together and appear in good orientation to accomodate a group capable of forming two hydrogen bonds (such as -OH or =O).

7.5 Fragments

This section summarises the most interesting fragments that were selected for the linking phase. All the poses shown have been identified with the haptic driven approach.

benzoic acid (Figure 7.4): This fragment is also present in the lead molecule. In its predicted pose it forms a salt bridge with Arg393. The ring is involved in hydrophobic interactions with Ala413, Arg393 and Tyr391.

thiophenol (Figure 7.5): The thiol has a potential for forming a disulfide bond with cystein 431. Being less reactive than the michael acceptor it is a valid candidate for the reduction of the toxicity of the compounds.

amide (Figure 7.6): the oxygen acts as a hydrogen bond acceptor, while the NH2 donates another hydrogen bond. The distance between the two groups is ideal to contact bouth Thr295 and Thr433.

Figure 7.4: benzoic acid: binding pose chosen with the haptic approach. The carboxylic group is engaged in a salt bridge with Arg393.

Figure 7.5: thiophenol: binding pose chosen with the haptic approach. The thiol group is in close proximity to Cys431.

Figure 7.6: amide: binding pose chosen with the haptic approach. The carbonyl and the amidic nitrogen are involved in hydrogen bonds with Thr295 and Thr433.

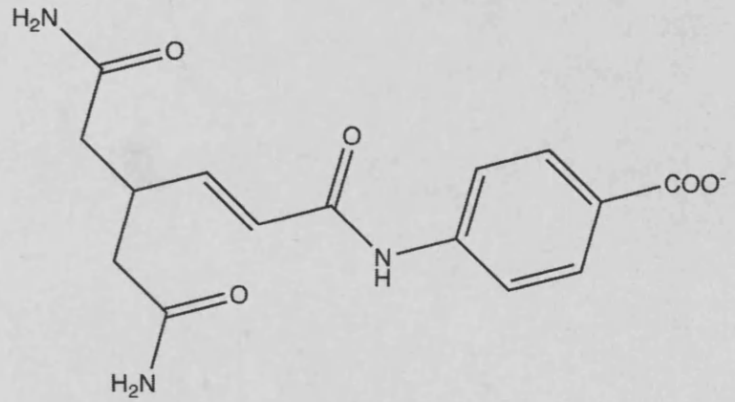
4-carboxy imidazole (Figure 7.7): The carboxylic group forms an hydrogen bond with Arg481. The two nitrogens on the ring form two hydrogen bonds with Thr295 and Thr433.

Figure 7.7: 4-carboxy imidazole: binding pose chosen with the haptic approach. The carboxyilic group interacts with Arg481. The nitrogens on the ring are involved in hydrogen bonds with Thr295 and Thr433.

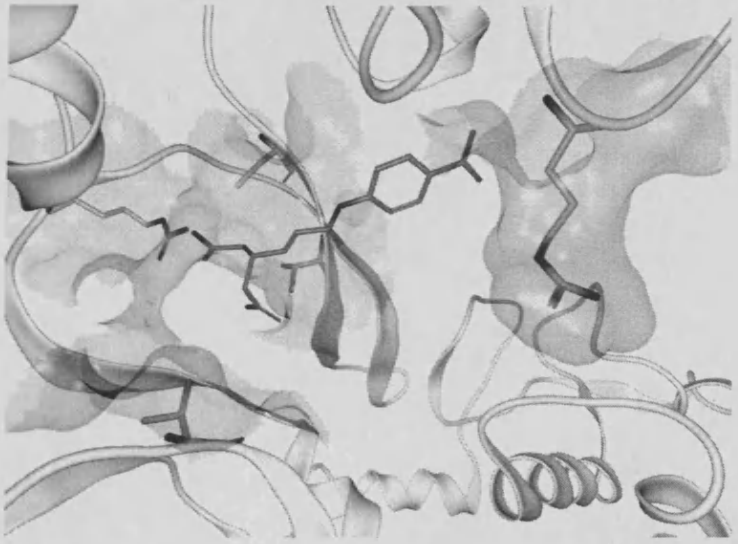
7.6 Lead molecules

After the generation of poses for each fragments, lead molecules were generated by linking various fragments in the active site trying to maintain orientations and interactions with the enzyme. Each resulting molecule was docked with PLANTS and the poses were confirmed.

Figures 7.8 to 7.17 show the poses of the various leads inside the active site.

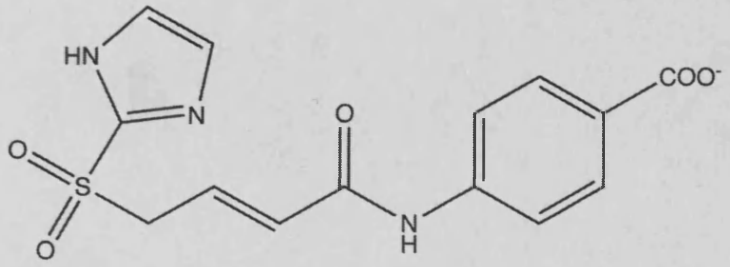


(a) molecular structure

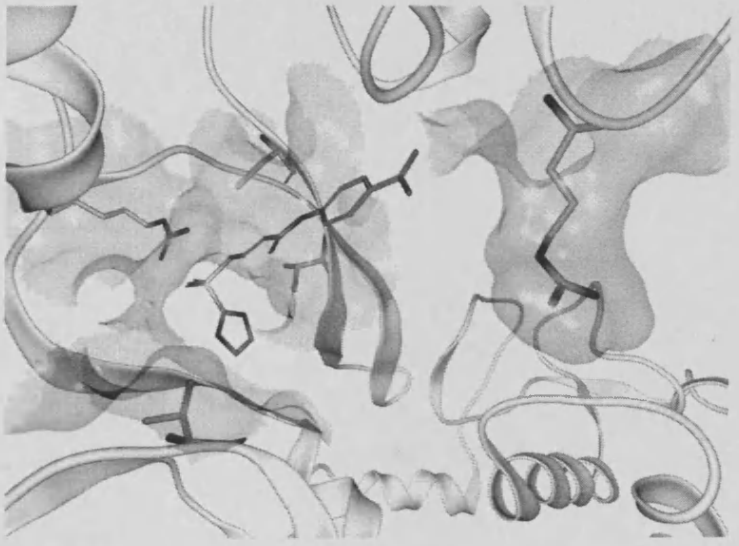


(b) pose within the active site

Figure 7.8: Compound H1. Structure and interactions with the enzyme

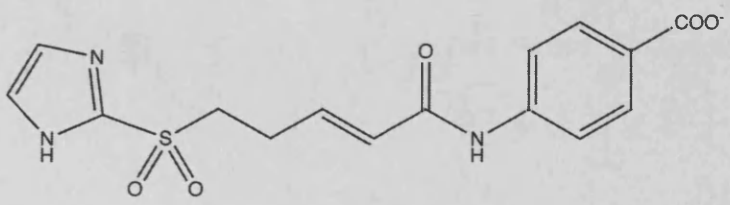


(a) molecular structure

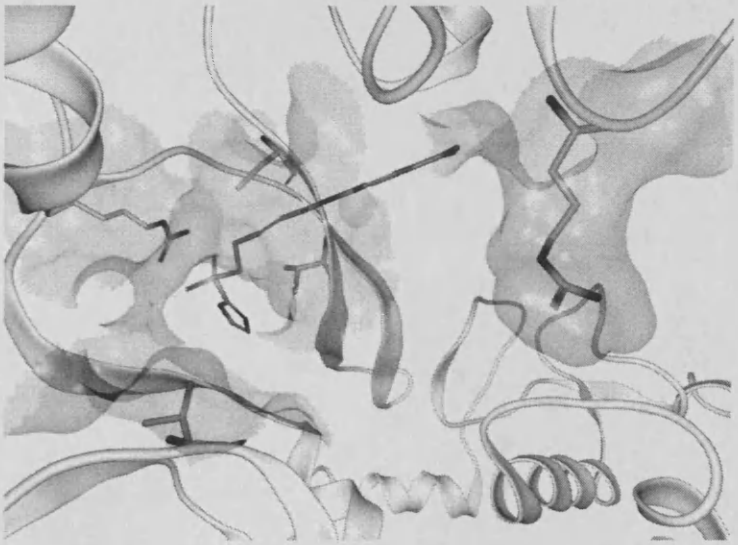


(b) pose within the active site

Figure 7.9: Compound H2. Structure and interactions with the enzyme

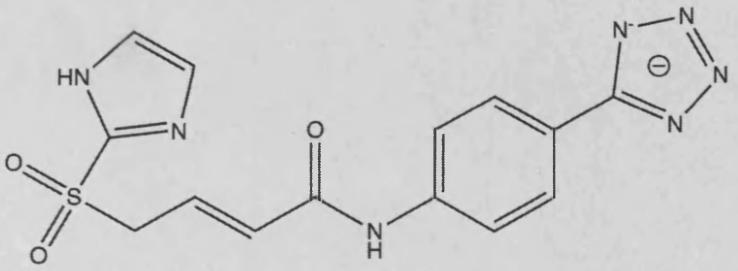


(a) molecular structure



(b) pose within the active site

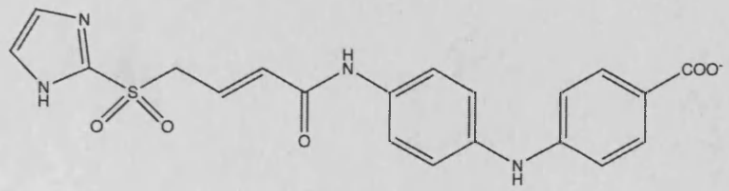
Figure 7.10: Compound H3. Structure and interactions with the enzyme



(a) molecular structure

(b) pose within the active site

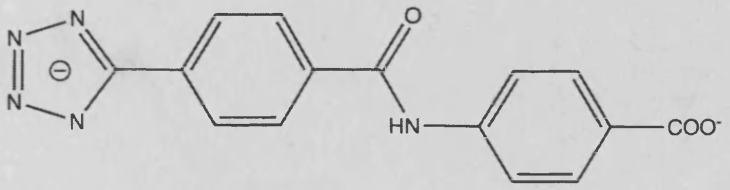
Figure 7.11: Compound H4. Structure and interactions with the enzyme



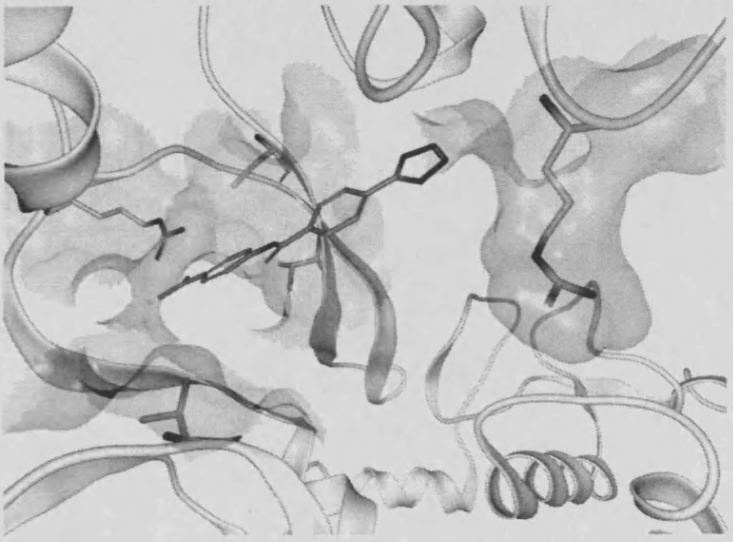
(a) molecular structure

(b) pose within the active site

Figure 7.12: Compound H5. Structure and interactions with the enzyme

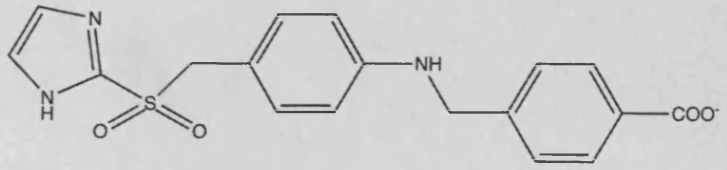


(a) molecular structure



(b) pose within the active site

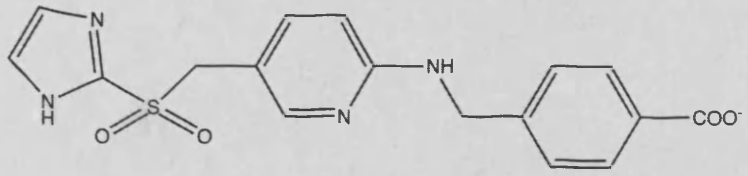
Figure 7.13: Compound H6. Structure and interactions with the enzyme



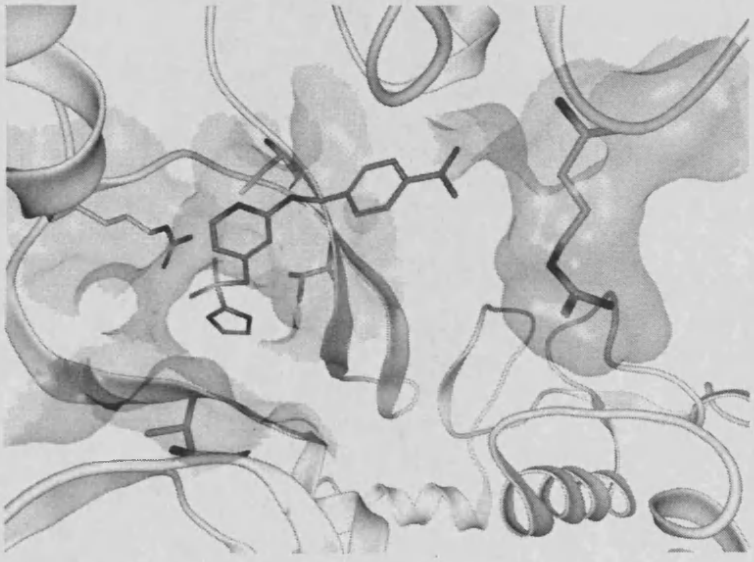
(a) molecular structure

(b) pose within the active site

Figure 7.14: Compound H7. Structure and interactions with the enzyme

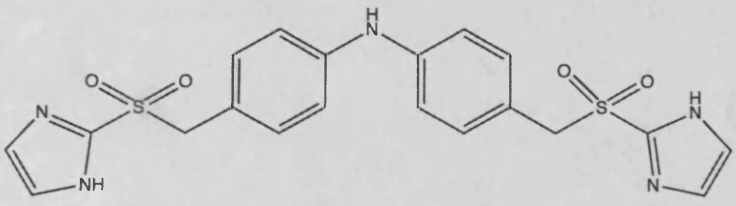


(a) molecular structure

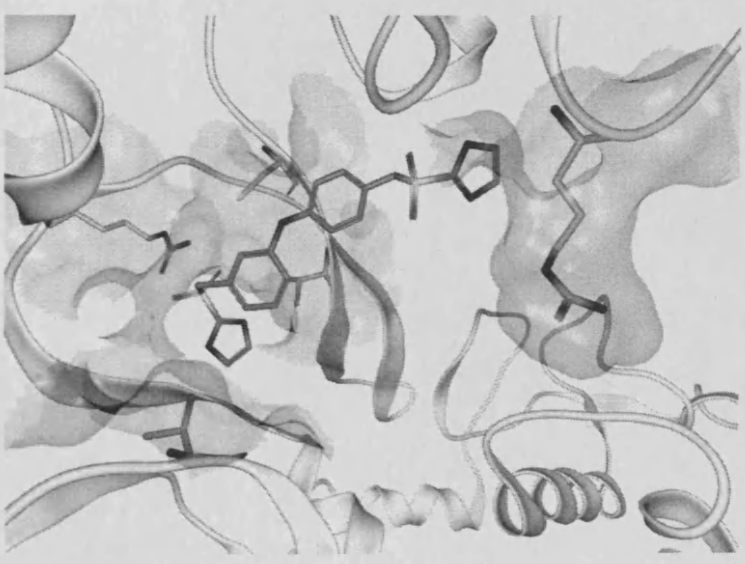


(b) pose within the active site

Figure 7.15: Compound H8. Structure and interactions with the enzyme

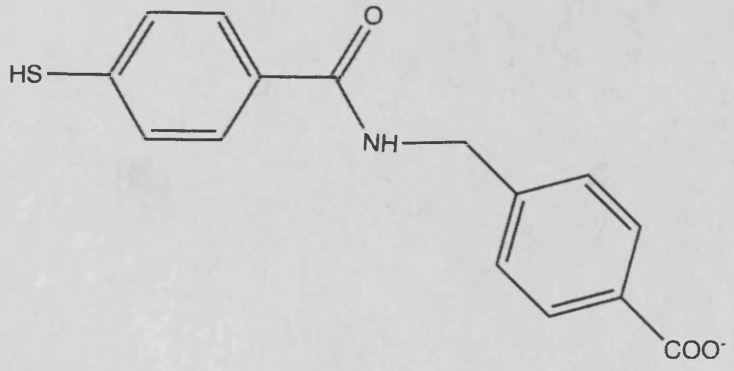


(a) molecular structure

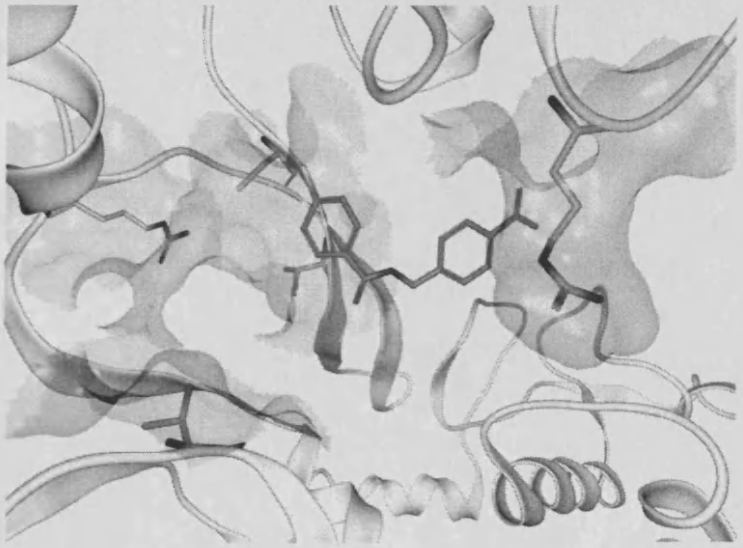


(b) pose within the active site

Figure 7.16: Compound H9. Structure and interactions with the enzyme



(a) molecular structure



(b) pose within the active site

Figure 7.17: Compound H10. Structure and interactions with the enzyme

Compound H10 (Figure 7.17) was synthesised and tested in vitro on the HCV helicase enzyme. It resulted to be an active inhibitor, with an IC_{50} of 320 nM and a good toxicity profile.

7.7 Conclusions

The application of the haptic approach to the design of HCV NS3 inhibitors demonstrated its applicability to de novo drug design projects. The fragment-based nature of the de novo approach makes it ideal to be treated with the application of haptic technology because of the lack of inner flexibility of ligands. This offers the possibility of directly controlling all the degrees of freedom of a fragment via the haptic stylus. The results were consistently in perfect agreement with docking predictions and one of the resulting hits showed activity in *in vitro* tests.

Chapter 8

Conclusions

This project has resulted in a fully working realtime drug design environment based on haptic technology and trimodal feedback. It has demonstrated the applicability of haptics to drug design and the feasability of a human steered approach to molecular modelling. The software has been made publicly availible on the internet and is free of charge for commercial and non-commercial use. The implementation of different haptic devices has been tested, ranging from the extremely low-end nintendo wii remote to more professional devices.

The haptic-based software has then been used alongside more conventional approaches showing its applicability and verifying in practice the importance of the direct human contribution in the process of drug design both in the antiviral and anticancer fields.

The limitations of the approach have also emerged: the control of the inner degrees of freedom of a molecule remains unsatisfactory in real time, and make working on complete molecules unpractical in most cases. Though still useful for the rapid testing of hypotheses on complete molecules, the approach is certainly more suitable for fragment-based solutions.

In the field of fragment-based de novo drug design the haptic approached has been demonstrated to be useful and consistently in accord with molecular docking

prediction. It was also possible to validate it with some in vitro results on both antiviral and anticancer drug design.

The application of the haptic-based approach lead to the design of a molecule with anti-HCV activity, acting on the helicase enzyme and several other candidates which has not been tested yet.

The in silico characterisation of the binding modes of a large numbers of structurally related compounds was carried out using different approaches and helped direct the synthesis of future analogues.

The design of an active molecule putatively acting on an unvalidated anticancer drug was also achieved, targetting the EGFR kinase dimerisation protein-protein interaction. The application of the haptic based approach was essential for the developing of this lead molecule into a family of structurally different compound.

Future developments of the approach include the introduction of protein flexibility. This will significantly increase the computational load of the simulation and will need a re-structuring of the code in order to make it run in parallel on multiple CPUs. The introduction of protein flexibility should render the simulation more realistic and allow for study of a wider range of phenomena.

Appendix A

Materials and Methods

A.1 Molecular docking and virtual screening

All docking experiments and virtual screenings were carried out using PLANTS (Protein Ligand ANT System) v 1.07 [117]. Plants performance for docking and virtual screening on a model kinase has previously been investigated at the University of Padova [118] and it outperformed most commercial softwares, with an enrichment ratio around 7.0 for the top 10% of the test database.

Plants search algorithm is based on Marco Dorigo's Ant Colony Optimisation algorithm [119]. A set of virtual ants explore the conformational space laying trails of pheromone to mark good solutions. The pheromone attracts more ants, which in turn lay more pheromone. An evaporation operator is used to weaken abandoned trails. Solutions are refined using a non-derivative minimisation algorithm (simplex algorithm).

The pseudocode of the search algorithm is shown in Algorithm 2.

The default parameterisation was used, with sigma = 1.0 for virtual screening and sigma = 4.0 for single docking (See Appendices) . In both cases a Lennard-Jones potential was used for the ligand inner clash energy and chemplp was used as the scoring function.

Algorithm 2 PLANTS search algorithm

```
InitialiseParametersAndPheromones()
for i = 1 to iterations do
  for j = 1 to ants do
    s_j ← ConstructSolution()
    s*_j ← LocalSearch(s_j)
    M ← M ∪ s*_j
  end for
  s^ib ← GetBestSolution()
  s^ib ← RefinementLocalSearch(s^ib, 0.2)
  M ← M ∪ s^ib
  UpdatePheromones(s^ib)
  if diversificationCriteriaMet then
    ApplySearchDiversification()
  end if
end for
return best solution found, M
```

For virtual screening the five different poses generated for each ligand were rescored using the MOE scoring function (scoring.svl). This has been shown to increase the performance of the virtual screening[118].

A.2 Molecular dynamics

All Molecular Dynamics simulations were carried out using GROMACS[120] v3.3.1. Energy Minimisation was done with a steepest descent algorithm. The exact parameters are shown in the appendices (Table B.3).

The solvent relaxation phase was allowed to last for 20 ps using a restrain on atomic positions under NPT conditions (constant number of particles, pressure and temperature). The exact parameters are shown in the appendices (Table B.4).

The system was then allowed to evolve without restraints under NPT conditions. When not otherwise stated this phase was allowed to last for 5 ns. The exact parameters are shown in the appendices (Table B.5).

A.3 Visualisation and graphics

Molecular Dynamics trajectories were analysed using VMD [121]. All other simulations were analysed in Zodiac and MOE. Ray-traced pictures for this thesis were generated using Zodiac, MOE, Pymol [122] and POVRay [123].

Appendix B

Parameters

B.1 Molecular docking and virtual screening

This section includes all the parameters used for PLANTS runs.

aco_ants	20
aco_evap	0.25
aco_sigma	5.0
flip_amide_bonds	0
flip_planar_n	1
cluster_structures	5
cluster_rmsd	2.0

Table B.1: Parameters for plants for single docking

aco_ants	20
aco_evap	0.25
aco_sigma	1.0
flip_amide_bonds	0
flip_planar_n	1
cluster_structures	5
cluster_rmsd	2.0

Table B.2: Parameters for plants for virtual screening

B.2 Molecular dynamics

This section includes all the parameters used for GROMACS runs.

define	-DFLEXIBLE
constraints	none
integrator	steep
dt	0.002
nsteps	2000
nstlist	5
ns_type	grid
rlist	0.9
coulombtype	PME
rcoulomb	0.9
rvdw	1.4
fourierspacing	0.12
fourier_nx	0
fourier_ny	0
fourier_nz	0
pme_order	4
ewald_rtol	1e-5
emtol	1000.0
emstep	0.01

Table B.3: Parameters for the energy minimisation .mdp file

define	-DPOSRES
constraints	all-bonds
integrator	md
dt	0.002
nsteps	10000
nstenergy	10
nstlist	5
ns_type	grid
rlist	0.9
coulombtype	PME
rcoulomb	0.9
rvdw	1.4
fourierspacing	0.12
fourier_nx	0
fourier_ny	0
fourier_nz	0
pme_order	4
ewald_rtol	1e-5
optimize_fft	yes
Tcoupl	berendsen
tau_t	0.1 0.1
tc-grps	protein sol
ref_t	300 300
Pcoupl	berendsen
tau_p	0.5
compressibility	4.5e-5
ref_p	1.0
gen_vel	yes
gen_temp	300
gen_seed	173529

Table B.4: Parameters for the solvent relaxation .mdp file

constraints	all-bonds
integrator	md
dt	0.002
nsteps	2500000
ns_type	grid
rlist	0.9
coulombtype	PME
rcoulomb	0.9
rvdw	1.4
fourierspacing	0.12
fourier_nx	0
fourier_ny	0
fourier_nz	0
pme_order	4
ewald_rtol	1e-5
optimize_fft	yes
Tcoupl	berendsen
tau_t	0.1 0.1
tc-grps	protein sol
ref_t	300 300
Pcoupl	berendsen
tau_p	0.5
compressibility	4.5e-5
ref_p	1.0
gen_vel	yes
gen_temp	300
gen_seed	173529

Table B.5: Parameters for the molecular dynamics .mdp file

Bibliography

[1] A R Leach. *Molecular Modelling principles and applications.* 2001.

[2] M. Hashida. [In-silico prediction of pharmacokinetic properties]. *Yakugaku Zasshi*, 125:853–861, Nov 2005.

[3] B. Waszkowycz. Towards improving compound selection in structure-based virtual screening. *Drug Discov. Today*, 13:219–226, Mar 2008.

[4] G. Nray-Szab. [Methods of computer-assisted molecular modeling]. *Acta Pharm Hung*, 68:7–13, Jan 1998.

[5] G. Ferenczy. [Structure-based drug design]. *Acta Pharm Hung*, 68:21–31, Jan 1998.

[6] S. G. Dahl and I. Sylte. Molecular modelling of drug targets: the past, the present and the future. *Basic Clin. Pharmacol. Toxicol.*, 96:151–155, Mar 2005.

[7] J. W. Ponder and D. A. Case. Force fields for protein simulations. *Adv. Protein Chem.*, 66:27–85, 2003.

[8] C. N. Schutz and A. Warshel. What are the dielectric "constants" of proteins and how to validate electrostatic models? *Proteins*, 44:400–417, Sep 2001.

[9] J. R. Tame. Scoring functions–the first 100 years. *J. Comput. Aided Mol. Des.*, 19:445–451, Jun 2005.

[10] Inbal Halperin, Buyong Ma, Haim Wolfson, and Ruth Nussinov. Principles of docking: An overview of search algorithms and a guide to scoring functions. *Proteins*, 47(4):409–43, 2002.

[11] R. C. Glen and S. C. Allen. Ligand-protein docking: cancer research at the interface between biology and chemistry. *Curr. Med. Chem.*, 10:763–767, May 2003.

[12] R. T. Kroemer. Structure-based drug design: docking and scoring. *Curr. Protein Pept. Sci.*, 8:312–328, Aug 2007.

[13] Alexandre M. J. J. Bonvin. Flexible protein-protein docking. *Current Opinion in Structural Biology*, 16(2):194–200, April 2006.

[14] R D Taylor, P J Jewsbury, and J W Essex. A review of protein-small molecule docking methods. *J Comput Aided Mol Des*, 16(3):151–166, Mar 2002. Comparative Study.

[15] M Rarey, B Kramer, T Lengauer, and G Klebe. A fast flexible docking method using an incremental construction algorithm. *J Mol Biol*, 261(3):470–489, Aug 1996.

[16] Marcel L Verdonk, Jason C Cole, Michael J Hartshorn, Christopher W Murray, and Richard D Taylor. Improved protein-ligand docking using GOLD. *Proteins*, 52(4):609–23, 2003.

[17] D S Goodsell, G M Morris, and A J Olson. Automated docking of flexible ligands: applications of AutoDock. *J Mol Recognit*, 9(1):1–5, Jan 1996.

[18] Thomas A Halgren, Robert B Murphy, Richard A Friesner, Hege S Beard, Leah L Frye, W Thomas Pollard, and Jay L Banks. Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J Med Chem*, 47(7):1750–9, 2004.

[19] C A Baxter, C W Murray, D E Clark, D R Westhead, and M D Eldridge. Flexible docking using Tabu search and an empirical estimate of binding affinity. *Proteins*, 33(3):367–382, Nov 1998.

[20] C. M. Song, S. J. Lim, and J. C. Tong. Recent advances in computer-aided drug design. *Brief. Bioinformatics*, 10:579–591, Sep 2009.

[21] S. A. Shaikh, T. Jain, G. Sandhu, N. Latha, and B. Jayaram. From drug target to leads–sketching a physicochemical pathway for lead molecule design in silico. *Curr. Pharm. Des.*, 13:3454–3470, 2007.

[22] C. A. Taft, V. B. Da Silva, and C. H. Da Silva. Current topics in computer-aided drug design. *J Pharm Sci*, 97:1089–1098, Mar 2008.

[23] D. B. Kitchen, H. Decornez, J. R. Furr, and J. Bajorath. Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat Rev Drug Discov*, 3:935–949, Nov 2004.

[24] A. N. Jain. Virtual screening in lead discovery and optimization. *Curr Opin Drug Discov Devel*, 7:396–403, Jul 2004.

[25] A. Liwo, C. Czaplewski, S. Odziej, and H. A. Scheraga. Computational techniques for efficient conformational sampling of proteins. *Curr. Opin. Struct. Biol.*, 18:134–139, Apr 2008.

[26] D. Hecht and M. Reiner. Multimodal virtual envonments: response times, attention and presence. *Presence*, 15:515–523, 2006.

[27] E. F. Meyer, S. M. Swanson, and J. A. Williams. Molecular modelling and drug design. *Pharmacol. Ther.*, 85:113–121, Mar 2000.

[28] P. B. Persson and M. D. Cooper. Designing and evaluating a haptic system for biomolecular education. *IEEE virtual rReality Conference*, pages 171–178, 2007.

[29] A. M. Wollacott and K. M. Merz. Haptic applications for molecular structure manipulation. *J. Mol. Graph. Model.*, 25:801–805, Mar 2007.

[30] N. Zonta, I. J. Grimstead, N. J. Avis, and A. Brancale. Accessible haptic technology for drug design applications. *J Mol Model*, 15:193–196, Feb 2009.

[31] J. P." "Pristle. Ribbon: a stereo cartoon drawing program for proteins. *J Appl Cryst*, 21:572–576, 1988.

[32] Thomas Halgren. Merk Molecular Force Field. I. Basis, Form, Scope, Parameterization, and Performance of MMFF94. *J Comp Chem*, 17(5-6):490–519, 1995.

[33] Thomas Halgren. Merk Molecular Force Field. II. MMFF94 van der Waals and Electrostatic Parameters for Intermolecular Interactions. *J Comp Chem*, 17(5-6):520–552, 1995.

[34] Thomas Halgren. Merk Molecular Force Field. III. Molecular Geometries and Vibrational Frequencies for MMFF94. *J Comp Chem*, 17(5-6):553–586, 1995.

[35] Thomas Halgren. Merk Molecular Force Field. IV. Conformational Energies and Geometries for MMFF94. *J Comp Chem*, 17(5-6):587–615, 1995.

[36] Thomas Halgren. Merk Molecular Force Field. V. Extension of MMFF94 Using Experimental Data, Additional Computational Data, and Empirical Rules. *J Comp Chem*, 17(5-6):616–641, 1995.

[37] Daniel K Gehlhaar, Gennady M Verkhivker, Paul A rejto, Christopher J Sherman, David B Fogel, Lawrence J Fogel, and Stephan T Freer. Molecular recognition of the inhibitor ag-1343 by hiv-1 protease: conformationally flexible docking by evolutionary programming. *Chemistry and Biology*, 2(4):317–324, 1995.

[38] MD Eldrige, CW Murray, TR Auton, GV Paolini, and RP Mee. Empirical scoring functions: 1. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J Comput-Aided Mol Des*, 1(14):731–751, 1997.

[39] F. R. Hirsch, M. Varella-Garcia, and F. Cappuzzo. Predictive value of EGFR and HER2 overexpression in advanced non-small-cell lung cancer. *Oncogene*, 28 Suppl 1:S32–37, Aug 2009.

[40] H. Yamashiro and M. Toi. [Anti-HER2 antibody and small-molecule tyrosine kinase inhibitor therapy targeting EGFR signal pathway]. *Gan To Kagaku Ryoho*, 36:1067–1071, Jul 2009.

[41] B. Zahorowska, P. J. Crowe, and J. L. Yang. Combined therapies for cancer: a review of EGFR-targeted monotherapy and combination treatment with other drugs. *J. Cancer Res. Clin. Oncol.*, 135:1137–1148, Sep 2009.

[42] J. Betowski and E. Lowicka. EGF receptor as a drug target in arterial hypertension. *Mini Rev Med Chem*, 9:526–538, May 2009.

[43] M. Y. Homs, E. E. Voest, and P. D. Siersema. Emerging drugs for esophageal cancer. *Expert Opin Emerg Drugs*, 14:329–339, Jun 2009.

[44] Diego Miranda-Saavedra and Geoffrey J Barton. Classification and functional annotation of eukaryotic protein kinases. *Proteins*, 68(4):893–914, Sep 2007.

[45] Philip Cohen. Protein kinases–the major drug targets of the twenty-first century? *Nat Rev Drug Discov*, 1(4):309–315, Apr 2002.

[46] G Manning, D B Whyte, R Martinez, T Hunter, and S Sudarsanam. The protein kinase complement of the human genome. *Science*, 298(5600):1912–1934, Dec 2002.

[47] Sam A Johnson and Tony Hunter. Kinomics: methods for deciphering the kinome. *Nat Methods*, 2(1):17–25, Jan 2005.

[48] Ilaria Rebay. Keeping the receptor tyrosine kinase signaling pathway in check: lessons from Drosophila. *Dev Biol*, 251(1):1–17, Nov 2002.

[49] Erez M Bublil and Yosef Yarden. The EGF receptor family: spearheading a merger of signaling and therapeutics. *Curr Opin Cell Biol*, 19(2):124–134, Apr 2007.

[50] Kathryn M Ferguson, Mitchell B Berger, Jeannine M Mendrola, Hyun Soo Cho, Daniel J Leahy, and Mark A Lemmon. EGF activates its receptor by removing interactions that autoinhibit ectodomain dimerization. *Mol Cell*, 11(2):507–517, Feb 2003. In Vitro.

[51] Hyun-Soo Cho and Daniel J Leahy. Structure of the extracellular region of HER3 reveals an interdomain tether. *Science*, 297(5585):1330–1333, Aug 2002.

[52] Y Yarden and M X Sliwkowski. Untangling the ErbB signalling network. *Nat Rev Mol Cell Biol*, 2(2):127–137, Feb 2001.

[53] Jennifer Stamos, Mark X Sliwkowski, and Charles Eigenbrot. Structure of the epidermal growth factor receptor kinase domain alone and in complex with a 4-anilinoquinazoline inhibitor. *J Biol Chem*, 277(48):46265–46272, Nov 2002.

[54] Xuewu Zhang, Jodi Gureasko, Kui Shen, Philip A Cole, and John Kuriyan. An allosteric mechanism for activation of the kinase domain of epidermal growth factor receptor. *Cell*, 125(6):1137–49, 2006.

[55] T Schindler, F Sicheri, A Pico, A Gazit, A Levitzki, and J Kuriyan. Crystal structure of Hck in complex with a Src family-selective tyrosine kinase inhibitor. *Mol Cell*, 3(5):639–648, May 1999.

[56] H L De Bondt, J Rosenblatt, J Jancarik, H D Jones, D O Morgan, and S H Kim. Crystal structure of cyclin-dependent kinase 2. *Nature*, 363(6430):595–602, Jun 1993.

[57] Thomas J Lynch, Daphne W Bell, Raffaella Sordella, Sarada Gurubhagavatula, Ross A Okimoto, Brian W Brannigan, Patricia L Harris, Sara M Haserlat, Jeffrey G Supko, Frank G Haluska, David N Louis, David C Christiani, Jeff Settleman, and Daniel A Haber. Activating mutations in the epidermal growth

factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *N Engl J Med*, 350(21):2129–2139, May 2004.

[58] J Guillermo Paez, Pasi A Janne, Jeffrey C Lee, Sean Tracy, Heidi Greulich, Stacey Gabriel, Paula Herman, Frederic J Kaye, Neal Lindeman, Titus J Boggon, Katsuhiko Naoki, Hidefumi Sasaki, Yoshitaka Fujii, Michael J Eck, William R Sellers, Bruce E Johnson, and Matthew Meyerson. EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. *Science*, 304(5676):1497–1500, Jun 2004.

[59] Hisayuki Shigematsu, Li Lin, Takao Takahashi, Masaharu Nomura, Makoto Suzuki, Ignacio I Wistuba, Kwun M Fong, Huei Lee, Shinichi Toyooka, Nobuyoshi Shimizu, Takehiko Fujisawa, Ziding Feng, Jack A Roth, Joachim Herz, John D Minna, and Adi F Gazdar. Clinical and biological features associated with epidermal growth factor receptor gene mutations in lung cancers. *J Natl Cancer Inst*, 97(5):339–346, Mar 2005.

[60] Peter L Toogood. The kinome is not enough. *Chem Biol*, 12(10):1057–1058, Oct 2005. Comment.

[61] H M Berman, J Westbrook, Z Feng, G Gilliland, T N Bhat, H Weissig, I N Shindyalov, and P E Bourne. The Protein Data Bank. *Nucleic Acids Res*, 28(1):235–242, Jan 2000.

[62] Edgar R Wood, Anne T Truesdale, Octerloney B McDonald, Derek Yuan, Anne Hassell, Scott H Dickerson, Byron Ellis, Christopher Pennisi, Earnest Horne, Karen Lackey, Krystal J Alligood, David W Rusnak, Tona M Gilmer, and Lisa Shewchuk. A unique structure for epidermal growth factor receptor bound to GW572016 (Lapatinib): relationships among protein conformation, inhibitor off-rate, and receptor activity in tumor cells. *Cancer Res*, 64(18):6652–6659, Sep 2004.

[63] P Traxler and P Furet. Strategies toward the design of novel and selective protein tyrosine kinase inhibitors. *Pharmacol Ther*, 82(2-3):195–206, May 1999.

[64] J. Zeng. Mini-review: computational structure-based design of inhibitors that target protein surfaces. *Comb. Chem. High Throughput Screen.*, 3:355–362, Oct 2000.

[65] L. A. McHugh, A. E. Sayan, J. Mejlvang, T. R. Griffiths, Y. Sun, M. M. Manson, E. Tulchinsky, J. K. Mellon, and M. Kriajevska. Lapatinib, a dual inhibitor of ErbB-1/-2 receptors, enhances effects of combination chemotherapy in bladder cancer cells. *Int. J. Oncol.*, 34:1155–1163, Apr 2009.

[66] Laurel O Sillerud and Richard S Larson. Design and structure of peptide and peptidomimetic antagonists of protein-protein interaction. *Curr Protein Pept Sci*, 6(2):151–169, Apr 2005.

[67] T. Suzuki, K. Ishii, H. Aizaki, and T. Wakita. Hepatitis C viral life cycle. *Adv. Drug Deliv. Rev.*, 59:1200–1212, Oct 2007.

[68] G. Garg and P. Kar. Management of HCV infection: current issues and future options. *Trop Gastroenterol*, 30:11–18, 2009.

[69] T.L. Tellinghuisen and C.M. Rice. Interaction between hepatitis C virus proteins and host cell factors. *Curr. Opin. Microbiol.*, 5:419–427, Aug 2002.

[70] B. Kronenberger and S. Zeuzem. Current and future treatment options for HCV. *Ann Hepatol*, 8:103–112, 2009.

[71] A. Varaklioti, N. Vassilaki, U. Georgopoulou, and P. Mavromara. Alternate translation occurs within the core coding region of the hepatitis C viral genome. *J. Biol. Chem.*, 277:17713–17721, May 2002.

[72] T.L. Tellinghuisen, M.J. Evans, T. von Hahn, S. You, and C.M. Rice. Studying hepatitis C virus: making the best of a bad virus. *J. Virol.*, 81:8853–8867, Sep 2007.

[73] G. Migliaccio, J.E. Tomassini, S.S. Carroll, L. Tomei, S. Altamura, B. Bhat, L. Bartholomew, M.R. Bosserman, A. Ceccacci, L.F. Colwell, R. Cortese,

R. De Francesco, A.B. Eldrup, K.L. Getty, X.S. Hou, R.L. LaFemina, S.W. Ludmerer, M. MacCoss, D.R. McMasters, M.W. Stahlhut, D.B. Olsen, D.J. Hazuda, and O.A. Flores. Characterization of resistance to non-obligate chain-terminating ribonucleoside analogs that inhibit hepatitis C virus replication in vitro. *J. Biol. Chem.*, 278:49164–49170, 2003.

[74] C. McGuigan, R. N. Pathirana, N. Mahmood, K. G. Devine, and A. J. Hay. Aryl phosphate derivatives of AZT retain activity against HIV1 in cell lines which are resistant to the action of AZT. *Antiviral Res.*, 17:311–321, Apr 1992.

[75] K. G. Devine, C. McGuigan, T. J. O'Connor, S. R. Nicholls, and D. Kinchington. Novel phosphate derivatives of zidovudine as anti-HIV compounds. *AIDS*, 4:371–373, Apr 1990.

[76] C. McGuigan, K. G. Devine, T. J. O'Connor, and D. Kinchington. Synthesis and anti-HIV activity of some haloalkyl phosphoramidate derivatives of 3'-azido-3'-deoxythymidine (AZT): potent activity of the trichloroethyl methoxyalaninyl compound. *Antiviral Res.*, 15:255–263, 1991.

[77] D. Cahard, C. McGuigan, and J. Balzarini. Aryloxy phosphoramidate triesters as pro-tides. *Mini Rev Med Chem*, 4:371–381, May 2004.

[78] S. S. Carroll and D. B. Olsen. Nucleoside analog inhibitors of hepatitis C virus replication. *Infect Disord Drug Targets*, 6:17–29, Mar 2006.

[79] S. J. Hecker, K. R. Reddy, P. D. van Poelje, Z. Sun, W. Huang, V. Varkhedkar, M. V. Reddy, J. M. Fujitaki, D. B. Olsen, K. A. Koeplinger, S. H. Boyer, D. L. Linemeyer, M. MacCoss, and M. D. Erion. Liver-targeted prodrugs of 2'-C-methyladenosine for therapy of hepatitis C virus infection. *J. Med. Chem.*, 50:3891–3896, Aug 2007.

[80] A. Lahm, A. Yagnik, A. Tramontano, and U. Koch. Hepatitis C virus proteins as targets for drug development: the role of bioinformatics and modelling. *Curr Drug Targets*, 3:281–296, Aug 2002.

[81] C.R. Corbeil, P. Englebienne, C.G. Yannopoulos, L. Chan, S.K. Das, D. Bilimoria, L. L'heureux, and N. Moitessier. Docking ligands into flexible and solvated macromolecules. 2. Development and application of fitted 1.5 to the virtual screening of potential HCV polymerase inhibitors. *J Chem Inf Model*, 48:902–909, 2008.

[82] K. "Klumpp, G. Kalayanov, Han Ma, S. Le Pogam, V Leveque, W. Jiang, N. Inocencio, A. De Witte, S. Rajyaguru, E. Tai, S. Chanda, M. R. Irwin, C. Sund, A. Winqist, T. Maltseva, S. Eriksson, E. Usova, M. Smith, A. Alker, I. Najera, N. Cammack, J. A. Martin, N. Gunnar Johansson, and D. B." Smith. 2'-deoxy-4'-azido nucleoside analogs are highly potent inhibitors of hepatitis C virus replication despite the lack of 2'-alpha-hydroxyl groups. *J Biol Chem*, 283(4):2167–75, 2008.

[83] L. E. Bird, H. S. Subramanya, and D. B. Wigley. Helicases: a unifying structural theme? *Curr. Opin. Struct. Biol.*, 8:14–18, Feb 1998.

[84] H. S. Cho, N. C. Ha, L. W. Kang, K. M. Chung, S. H. Back, S. K. Jang, and B. H. Oh. Crystal structure of RNA helicase from genotype 1b hepatitis C virus. A feasible mechanism of unwinding duplex RNA. *J. Biol. Chem.*, 273:15045–15052, Jun 1998.

[85] H. S. Cho, N. C. Ha, L. W. Kang, K. M. Chung, S. H. Back, S. K. Jang, and B. H. Oh. Crystal structure of RNA helicase from genotype 1b hepatitis C virus. A feasible mechanism of unwinding duplex RNA. *J. Biol. Chem.*, 273:15045–15052, Jun 1998.

[86] S. G. Mackintosh, J. Z. Lu, J. B. Jordan, M. K. Harrison, B. Sikora, S. D. Sharma, C. E. Cameron, K. D. Raney, and J. Sakon. Structural and biological identification of residues on the surface of NS3 helicase required for optimal replication of the hepatitis C virus. *J. Biol. Chem.*, 281:3528–3535, Feb 2006.

[87] J. L. Kim, K. A. Morgenstern, J. P. Griffith, M. D. Dwyer, J. A. Thomson, M. A. Murcko, C. Lin, and P. R. Caron. Hepatitis C virus NS3 RNA helicase

domain with a bound oligonucleotide: the crystal structure provides insights into the mode of unwinding. *Structure*, 6:89–100, Jan 1998.

[88] P. Soultanas, M. S. Dillingham, S. S. Velankar, and D. B. Wigley. DNA binding mediates conformational changes and metal ion coordination in the active site of PcrA helicase. *J. Mol. Biol.*, 290:137–148, Jul 1999.

[89] S. S. Velankar, P. Soultanas, M. S. Dillingham, H. S. Subramanya, and D. B. Wigley. Crystal structures of complexes of PcrA DNA helicase with a DNA substrate indicate an inchworm mechanism. *Cell*, 97:75–84, Apr 1999.

[90] D. A. Bernstein, M. C. Zittel, and J. L. Keck. High-resolution structure of the E.coli RecQ helicase catalytic core. *EMBO J.*, 22:4910–4921, Oct 2003.

[91] G. Kadar and A. L. Haenni. Virus-encoded RNA helicases. *J. Virol.*, 71:2583–2590, Apr 1997.

[92] K. Theis, P. J. Chen, M. Skorvaga, B. Van Houten, and C. Kisker. Crystal structure of UvrB, a DNA helicase adapted for nucleotide excision repair. *EMBO J.*, 18:6899–6907, Dec 1999.

[93] A. E. Gorbalenya and E. V. Koonin. Viral proteins containing the purine NTP-binding sequence pattern. *Nucleic Acids Res.*, 17:8413–8440, Nov 1989.

[94] N. Yao, T. Hesson, M. Cable, Z. Hong, A. D. Kwong, H. V. Le, and P. C. Weber. Structure of the hepatitis C virus RNA helicase domain. *Nat. Struct. Biol.*, 4:463–467, Jun 1997.

[95] J. E. Walker, M. Saraste, M. J. Runswick, and N. J. Gay. Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *EMBO J.*, 1:945–951, 1982.

[96] P. Borowski, O. Mueller, A. Niebuhr, M. Kalitzky, L. H. Hwang, H. Schmitz, M. A. Siwecka, and T. Kulikowsk. ATP-binding domain of NTPase/helicase

as a target for hepatitis C antiviral therapy. *Acta Biochim. Pol.*, 47:173–180, 2000.

[97] G. Maga, S. Gemma, C. Fattorusso, G. A. Locatelli, S. Butini, M. Persico, G. Kukreja, M. P. Romano, L. Chiasserini, L. Savini, E. Novellino, V. Nacci, S. Spadari, and G. Campiani. Specific targeting of hepatitis C virus NS3 RNA helicase. Discovery of the potent and selective competitive nucleotide-mimicking inhibitor QU663. *Biochemistry*, 44:9637–9644, Jul 2005.

[98] G. Kleymann, R. Fischer, U. A. Betz, M. Hendrix, W. Bender, U. Schneider, G. Handke, P. Eckenberg, G. Hewlett, V. Pevzner, J. Baumeister, O. Weber, K. Henninger, J. Keldenich, A. Jensen, J. Kolb, U. Bach, A. Popp, J. Mben, I. Frappa, D. Haebich, O. Lockhoff, and H. Rbsamen-Waigmann. New helicase-primase inhibitors as drug candidates for the treatment of herpes simplex disease. *Nat. Med.*, 8:392–398, Apr 2002.

[99] G. Kleymann, R. Fischer, U. A. Betz, M. Hendrix, W. Bender, U. Schneider, G. Handke, P. Eckenberg, G. Hewlett, V. Pevzner, J. Baumeister, O. Weber, K. Henninger, J. Keldenich, A. Jensen, J. Kolb, U. Bach, A. Popp, J. Mben, I. Frappa, D. Haebich, O. Lockhoff, and H. Rbsamen-Waigmann. New helicase-primase inhibitors as drug candidates for the treatment of herpes simplex disease. *Nat. Med.*, 8:392–398, Apr 2002.

[100] P. S. Jones. Strategies for antiviral drug discovery. *Antivir. Chem. Chemother.*, 9:283–302, Jul 1998.

[101] D. N. Frick. Helicases as antiviral drug targets. *Drug News Perspect.*, 16:355–362, 2003.

[102] B. Gu, C. Liu, J. Lin-Goerke, D. R. Maley, L. L. Gutshall, C. A. Feltenberger, and A. M. Del Vecchio. The RNA helicase and nucleotide triphosphatase activities of the bovine viral diarrhea virus NS3 protein are essential for viral replication. *J. Virol.*, 74:1794–1800, Feb 2000.

[103] A. E. Matusan, M. J. Pryor, A. D. Davidson, and P. J. Wright. Mutagenesis of the Dengue virus type 2 NS3 protein within and outside helicase motifs: effects on enzyme activity and virus replication. *J. Virol.*, 75:9633–9643, Oct 2001.

[104] P. Borowski, S. Schalinski, and H. Schmitz. Nucleotide triphosphatase/helicase of hepatitis C virus as a target for antiviral therapy. *Antiviral Res.*, 55:397–412, Sep 2002.

[105] C. E. Samuel. Antiviral actions of interferons. *Clin. Microbiol. Rev.*, 14:778–809, Oct 2001.

[106] M. Yoneyama, M. Kikuchi, T. Natsukawa, N. Shinobu, T. Imaizumi, M. Miyagishi, K. Taira, S. Akira, and T. Fujita. The RNA helicase RIG-I has an essential function in double-stranded RNA-induced innate antiviral responses. *Nat. Immunol.*, 5:730–737, Jul 2004.

[107] J. J. Crute, I. R. Lehman, J. Gambino, T. F. Yang, P. Medveczky, M. Medveczky, N. N. Khan, C. Mulder, J. Monroe, and G. E. Wright. Inhibition of herpes simplex virus type 1 helicase-primase by (dichloroanilino)purines and -pyrimidines. *J. Med. Chem.*, 38:1820–1825, May 1995.

[108] U. A. Betz, R. Fischer, G. Kleymann, M. Hendrix, and H. Rbsamen-Waigmann. Potent in vivo antiviral activity of the herpes simplex virus primase-helicase inhibitor BAY 57-1293. *Antimicrob. Agents Chemother.*, 46:1766–1772, Jun 2002.

[109] X. G. Xi. Helicases as antiviral and anticancer drug targets. *Curr. Med. Chem.*, 14:883–915, 2007.

[110] F. C. Spector, L. Liang, H. Giordano, M. Sivaraja, and M. G. Peterson. Inhibition of herpes simplex virus replication by a 2-amino thiazole via interactions with the helicase component of the UL5-UL8-UL52 complex. *J. Virol.*, 72:6979–6987, Sep 1998.

[111] J. M. Pawlotsky. Mechanisms of antiviral treatment efficacy and failure in chronic hepatitis C. *Antiviral Res.*, 59:1–11, Jun 2003.

[112] F. V. Chisari. Unscrambling hepatitis C virus-host interactions. *Nature*, 436:930–932, Aug 2005.

[113] E. C. Borden, G. C. Sen, G. Uze, R. H. Silverman, R. M. Ransohoff, G. R. Foster, and G. R. Stark. Interferons at age 50: past, current and future impact on biomedicine. *Nat Rev Drug Discov*, 6:975–990, Dec 2007.

[114] J. R. Burton and G. T. Everson. HCV NS5B polymerase inhibitors. *Clin Liver Dis*, 13:453–465, Aug 2009.

[115] J. R. Burton and G. T. Everson. HCV NS5B polymerase inhibitors. *Clin Liver Dis*, 13:453–465, Aug 2009.

[116] S. Kandil, S. Biondaro, D. Vlachakis, A. C. Cummins, A. Coluccia, C. Berry, P. Leyssen, J. Neyts, and A. Brancale. Discovery of a novel HCV helicase inhibitor by a de novo drug design approach. *Bioorg. Med. Chem. Lett.*, 19:2935–2937, Jun 2009.

[117] O. Korb, T. Stützle, and T. E. Exner. Plants: Application of ant colony optimization to structure-based drug design. In M. Dorigo, L. M. Gambardella, M. Birattari, A. Martinoli, R. Poli, and T. Stützle, editors, *Ant Colony Optimization and Swarm Intelligence, 5th International Workshop, ANTS 2006*, volume 4150 of *Lecture Notes in Computer Science*, pages 247–258. Springer, Berlin, 2006.

[118] N. Zonta, G. Cozza, A. Gianoncelli, O. Korb, T. E. Exner, F. Meggio, G. Zagotto, and S. Moro. Scouting Novel Protein Kinase A (PKA) Inhibitors by Using a Consensus Docking-Based Virtual Screening Approach. *Lett Drug Des Discov*, 6:327–336, Aug 2009.

[119] M Dorigo, G Di Caro, and L M Gambardella. Ant algorithms for discrete optimization. *Artif Life*, 5(2):137–172, Spring 1999.

[120] David Van Der Spoel, Erik Lindahl, Berk Hess, Gerrit Groenhof, Alan E Mark, and Herman J C Berendsen. GROMACS: fast, flexible, and free. *J Comput Chem*, 26(16):1701–1718, Dec 2005.

[121] William Humphrey, Andrew Dalke, and Klaus Schulten. VMD – Visual Molecular Dynamics. *Journal of Molecular Graphics*, 14:33–38, 1996.

[122] W.L. DeLano. The PyMOL Molecular Graphics System. http://www.pymol.org, 2002.

[123] POVRay Team. The Persistence of Vision Raytracer. http://www.povray.org, 2000.